


## Fast Bayesian Tomography of a Two-Qubit Gate Set in Silicon

T.J. Evans,<sup>1</sup> W. Huang,<sup>2</sup> J. Yoneda<sup>1,2</sup>, R. Harper<sup>1</sup>, T. Tanttu,<sup>2</sup> K.W. Chan<sup>1,2</sup>, F.E. Hudson<sup>1,2</sup>,  
K.M. Itoh,<sup>3</sup> A. Saraiva<sup>1,2</sup>, C.H. Yang<sup>1,2</sup>, A.S. Dzurak,<sup>2</sup> and S.D. Bartlett<sup>1,\*</sup>

<sup>1</sup>Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, Sydney 2006, Australia

<sup>2</sup>School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, NSW 1466, Australia

<sup>3</sup>School of Fundamental Science and Technology, Keio University, Yokohama 223-8521, Japan

 (Received 8 September 2021; revised 29 September 2021; accepted 2 February 2022; published 25 February 2022)

Benchmarking and characterizing quantum states and logic gates is essential in the development of devices for quantum computing. We introduce a Bayesian approach to self-consistent process tomography, called fast Bayesian tomography (FBT), and experimentally demonstrate its performance in characterizing a two-qubit gate set on a silicon-based spin qubit device. FBT is built on an adaptive self-consistent linearization that is robust to model approximation errors. Our method offers several advantages over other self-consistent tomographic methods. Most notably, FBT can leverage prior information from randomized benchmarking (or other characterization measurements), and can be performed in real time, providing continuously updated estimates of full process matrices while data are acquired.

DOI: [10.1103/PhysRevApplied.17.024068](https://doi.org/10.1103/PhysRevApplied.17.024068)

### I. INTRODUCTION

In developing technology and devices for fault-tolerant quantum computing [1], the diagnosis and mitigation of errors in quantum gates is critical. A now standard approach to assessing the performance of single- and two-qubit gates for quantum processors is to use randomized benchmarking (RB) [2–5] as well as variants built on this approach [5–11]. The success of RB and related schemes comes from their efficiency for small systems as well as their robustness to state, preparation, and measurement (SPAM) errors. However, a key limitation is their inability, in general, to provide full diagnostic information about the types of errors. This information is essential for these errors to be mitigated or eliminated.

When full characterization of a gate or process is required, tomography remains the gold standard. Similar to RB, there are many methods to perform such characterization, including tomography for unitary gates [12–15] and for general channels [16]. An important step in the development of process tomography was the recognition that the operations used in the SPAM steps were also faulty, and required simultaneous characterization with the gates themselves. Process tomography that is capable of this simultaneous characterization is said to be *self-consistent* [17]. Currently available implementations of self-consistent process tomography such as gate-set tomography (GST) [16] require the tailoring of specific

sequences for specific gate sets. Typically, the number of sequences for two-qubit characterization is very large, with commensurate computational requirements. The nonlinear nature of tomographic reconstruction makes it difficult for nonexpert users to understand the simplifying assumptions and technical insights required to make the numerical methods reliable in practice [18], which may explain its limited adoption for characterizing two or more qubits. Another limitation of all such systems is their inability to incorporate additional data or to update (without a full rerun of the nonlinear, nonconvex solvers) when more data become available. Other tomography methods addressing some of these issues include Bayesian approaches [19–21], which provide the ability to encode prior information and so reduce the required data. However, this efficiency is usually traded off for computational cost due to the need for extensive sampling methods.

In this paper we present a Bayesian method for self-consistent process tomography, called *fast Bayesian tomography* (FBT). Our approach is able to extract tomographic information about a whole gate set from *arbitrary* sequences of gates. Our goal with FBT was to create a tomography method that was operationally simple and efficient in both the experimental and computational cost. The computational efficiency is achieved by defining a statistically optimal linearization of the self-consistent objective function. Moreover, the Bayesian model can be continuously updated as data become available, providing ongoing improvement to the linearization. The data from standard RB experiments can be used to bootstrap the model. As we

\*stephen.bartlett@sydney.edu.au

demonstrate, reinterpreting the RB data can already give good initial estimates of the process matrices of each of the gates, together with related uncertainties. Where required, such matrices and the uncertainties in their values can be updated with information from experiments as further data are extracted from the device.

To illustrate FBT and its performance, we use it to characterize a two-qubit gate set using the spins of electrons in a silicon-metal-oxide-semiconductor (SiMOS) quantum dot device. Spin qubits in silicon are among the most promising semiconductor architectures for scalable quantum information processing due to their compatibility with modern semiconductor manufacturing processes [22], nanoscale footprint, and capability for hosting highly coherent qubits [23,24]. High-performance single-qubit gates are well characterized and have reached charge noise limits [25,26]. Moreover, two-qubit gates have recently demonstrated fidelities exceeding 99% [27]. Although some experiments have attempted to probe the type of noise affecting two-qubit gates [28,29], the limiting factors remain unclear. Identifying the key noise sources and failure modes for two-qubit gates lies at the heart of the development of semiconductor spin qubits, and it is here that a fast, flexible approach to self-consistent process tomography to characterize the noise is expected to be invaluable.

Specifically, we reconstruct a two-qubit native and non-standard gate set in a silicon spin qubit device, as demonstrated in Ref. [30], in a self-consistent manner. We bootstrap data from an RB characterization of the gate set, followed by additional randomized gate sequences. The results are high-accuracy estimates of the process matrices describing the gates, together with robust quantification of the uncertainties (error bars) on these estimates, highlighting the ability of the protocol to characterize nonstandard gate sets as well as to leverage benchmarking data. The processing of data takes less time than the data acquisition itself, demonstrating that FBT can be performed in real time during an experiment. From the reconstructed process matrices we can infer SPAM-free Bell-state metrics [27], as opposed to directly performing state tomography [31]. We infer Bell-state fidelities in the range 94.6%–98.3%, comparable to those reported in Ref. [27], and concurrences from 88.3% to 92.0%, the highest reported in silicon quantum dots. We find the individual two-qubit controlled rotation (CROT) gate fidelities are between 96.1% and 97.4%, consistent with the gate fidelity reported in a previous experiment using the same device [30]. Finally, to illustrate the flexibility of FBT for data that were not tailored specifically for tomography, we show that data from an RB characterization of the device can be repurposed and can yield high-precision estimates of the gate set.

Section II presents an overview of the FBT protocol and a summary of the experimental demonstration. Section III

then presents a more detailed description and analysis of the FBT protocol.

## II. RESULTS

### A. FBT protocol

In this section we outline the FBT protocol and demonstrate its performance on a two-qubit gate set in a SiMOS device. FBT provides a means of performing process tomography over a primitive gate set in a self-consistent manner. Moreover, it is efficient in its use of resources, both experimental and computational, and operationally simple, requiring no tailoring of complex experimental settings.

In an experiment involving a sequence of gates, the measurement outcomes are governed by Born's rule, and this provides us with the functional relationship between the gate parameters and our data. Tomography is the procedure to invert this relationship to solve for the gate parameters given the measurement outcomes. All tomographic methods suffer from the central problem that this function is a high-dimensional polynomial, of potentially high degree (for repeated gates). Solving for the gates requires a nonlinear optimization that is typically intractable without simplifying assumptions, leading to large nonlinear, nonconvex optimizations that risk local-minima traps and can lead to long difficult computations [18]. For this reason, the initial proposal of self-consistent methods [17] relied on linearization to reduce the computational demand. However, linearizing around the identity imports a number of assumptions that can lead to poor performance.

The Bayesian approach proposed in this paper gives us the ability to address this problem, and to do so in a way that is efficient in experimental and computational resources. First, we show how to use prior information that we may already have about the gates, such as average gate fidelities from RB experiments, to provide an improved initial linearization. Second, the Bayesian approach allows us to update our estimates of the gate set in real time (*online*), as data become available. This feature allows not only for data collection to be halted when reconstructions are of a desired accuracy, but also for additional data to be incorporated into the estimates without the need to rerun the entire computation.

Third, and most importantly, it provides us with an optimal linearization. Rather than linearizing the model about the identity, FBT instead uses the prior distribution as a starting point. We believe that this observation, albeit simple, has a significant impact on the utility and accuracy of the protocol. The online nature of the protocol further supplements this feature as the mean is iteratively updated, improving the linearization every time additional measurements are performed. By allowing linearization around increasingly better estimates we believe that this

also means that gate sets of arbitrary (e.g., lower) fidelity can be reconstructed, something many characterization methods such as RB can struggle to do. Moreover, the Bayesian model provides an accurate quantification of the approximation error incurred by the linearization process. As a result, we can take this error into account during the reconstruction. By knowing our estimate of the error we can prevent the model overfitting to this error.

The Bayesian model uses a multivariate Gaussian prior over all of the selected channel parameters. Each gate in the set is decomposed into the ideal implementation of a unitary gate  $G$  followed by a gate-dependent noise channel  $\Lambda_G$ , such that  $\tilde{G} = \Lambda_G G$ . We are interested in reconstructing  $\Lambda_G$  for each  $G$ , quantified by a mean  $\bar{\Lambda}_G$  and a covariance  $\Gamma_G$ . In principle, one would desire a prior that is constrained to be a physical quantum channel. A Gaussian distribution does not necessarily guarantee a physical quantum channel and requires us to project onto one, but this limitation is far outweighed by the utility and robustness of the Gaussian model.

With non-Bayesian methods where reconstruction occurs after data acquisition, such reconstruction can often take many tens of hours for full two-qubit reconstruction of a gate set. By comparison the computational costs of FBT occur mainly as the data are initially processed and the model error needs to be calculated, and then quickly reduces to fast calculations that can be done as data are generated (we discuss this in more detail in Sec. II D). This computation includes the calculations necessary to constrain the estimations to physical priors (see later discussion in Sec. B 6). The complexity of quantum sampling will, of course, depend on the accuracy required. By using prior data (such as RB data), we reduce the number of quantum circuits that are required. As we show below, FBT achieved very high accuracy results with fewer runs than we estimate is required for other gate-set tomographic protocols. However, if our priors are incorrect or there are unfortunate runs of random sequences then it may be that more data is required for satisfactory convergence.

Full details of the FBT protocol are given in Sec. III. Here, we demonstrate FBT on a gate set for a two-qubit device based on electron spins in silicon.

### B. Device and gate implementation

Figure 1 shows a scanning electron microscope image of a SiMOS double quantum dot device, identical to the device used in this experiment. Qubits are defined in single electron quantum dots formed underneath gates G1 and G2. An external magnetic field  $B_0 = 1.42$  T creates a Zeeman splitting of  $E_Z = g\mu_B B_0 \approx 0.16$  meV, corresponding to an average electron spin resonance (ESR) frequency  $f = E_Z/h = 39.33$  GHz, where  $g$  is the electron  $g$ -factor,  $\mu_B$  is the Bohr magneton and  $h$  is Planck's

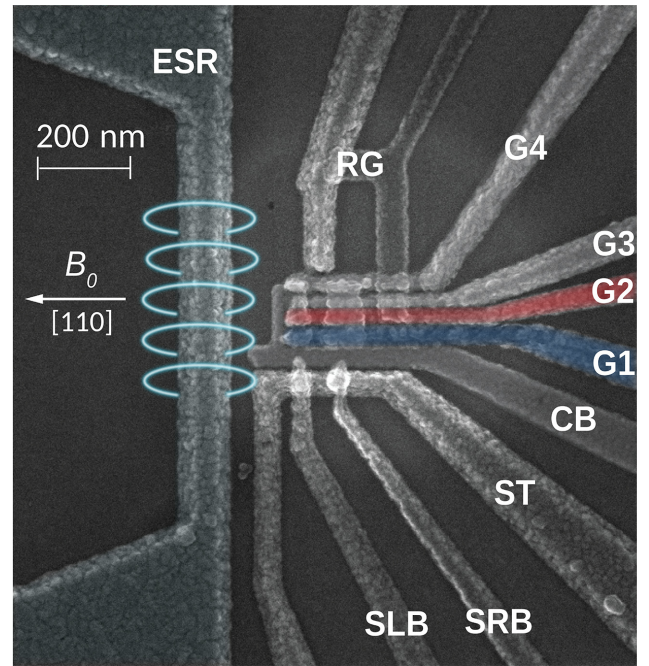


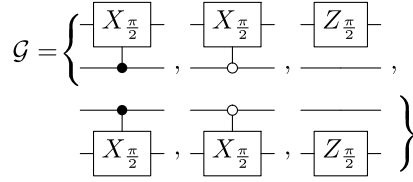
FIG. 1. False-color scanning electron microscope image of the device. Two quantum dots are formed underneath gates G1 (blue) and G2 (red). The gates CB, G3, and G4 form confinement barriers that laterally define the quantum dots. RG is the reservoir gate that supplies electrons to the quantum dots. The gate electrodes ST, SLB, and SRB define a single-electron transistor, designed to sense charge movement in the quantum dot region. An alternating current running through the ESR line generates an oscillating magnetic field  $B_1$  (light blue) to manipulate the electron spins. The direction of the external magnetic field  $B_0$  is indicated by the white arrow.

constant. The device is operated in a dilution refrigerator at an electron temperature of  $T_e \approx 180$  mK, allowing us to read the electron spin state sequentially via spin-dependent tunneling [32] and selectively load a  $|\downarrow\rangle$  electron for initialization. Single- and two-qubit control are enabled by a Zeeman energy difference of  $\delta E_Z/h = 13.42$  MHz and the exchange coupling  $J/h = 3.77$  MHz.

Two-qubit gates are commonly implemented in spin systems such as the  $\sqrt{\text{SWAP}}$  [33], the controlled phase [28,34,35] or the controlled rotation (CROT) [30,36]. In our experiment, the six primitive gates [Eq. (1)] for universal qubit control consist of variations of CROT gates and virtual- $Z$  gates. This choice of primitive gates allows us to control the qubits with constant exchange coupling, alleviating the requirement for high-bandwidth gate electrode voltage pulse and accurate synchronization between signal sources. High-fidelity one- and two-qubit gates were previously benchmarked on this device and reported in Ref. [30].

Our process tomographic characterization will be at the level of the device's native primitive ( $\pi/2$ -pulse) gate set.

The gate set we consider is



which we will denote by

$$\mathcal{G} = \{U_{1,\downarrow}^{\pi/2}, U_{1,\uparrow}^{\pi/2}, Z_1^{\pi/2}, U_{2,\downarrow}^{\pi/2}, U_{2,\uparrow}^{\pi/2}, Z_2^{\pi/2}\}, \quad (1)$$

respectively. See Ref. [30] for further details about this gate set.

As mentioned, the single qubit  $Z$  gates are implemented by a virtual change of reference. As a result, the fidelity of these gates can be expected to be significantly higher than the pulsed CROT-type gates. Despite their virtual nature, we still consider the  $Z$  gates as unknowns in the self-consistent protocol.

We also need to consider state preparations and measurements, and the noise associated with them. We assume that this can be modeled as a noise channel prior to the readout. The gauge degree of freedom between state preparation and measurement persists in this scenario; however, based on early characterization of this and similar devices it is reasonable to assume that the SPAM errors are dominated by the measurement. This is limited by the signal strength as the qubits being read out are far away from the charge detector. Furthermore, we ensure the high initialization fidelity by choosing a readout time of approximately 3.5 ms for both qubits which is sufficiently longer than the tunneling time of both ground and excited states.

### C. Randomized benchmarking

We envisage our FBT protocol being used immediately following an initial characterization of the device using RB, and in particular FBT is specifically designed to leverage the average gate fidelities from RB as prior information. We present the RB protocol used for the initial characterization of the device, following Ref. [30]. In this protocol we generate gate sequences of varying lengths  $L - 1$ , where all gates are randomly chosen from the two-qubit Clifford group. The final  $L$ th Clifford gate at the end of each sequence is randomly chosen from those that would ideally return the final state to  $|\uparrow\uparrow\rangle$ . The decay is fitted to the function  $P = A(1 - \frac{4}{3}r_C)^L + B$ , where the fitting parameters  $A$  and  $B$  absorb the SPAM errors, and  $r_C$  is the error per Clifford gate. The resonance frequencies are recalibrated after every two RB sequences.

From the decay, we extract an average Clifford gate fidelity of  $F_{\text{Clifford}} = 1 - r_C = 90.5\% \pm 1\%$ . With the estimated average primitive fidelity  $F_{\pi/2} = 98.2\% \pm 0.2\%$  we can update our prior distribution to be consistent with this

benchmark. Since RB is designed to eliminate all coherent error and to average out incoherent error, we can only learn so much from the single parameter it returns. Applied to the whole gate set, this will mean the corresponding prior estimates for each noise channel will be pure depolarizing channels. The benefit of the RB protocol is that these estimates are unaffected by SPAM errors, and this benefit will assist in separating the SPAM errors from gate errors in the subsequent experiment.

### D. Self-consistent process tomography

The initial characterization using RB is used to construct nontrivial prior distributions for all of our gates based on the average gate fidelities from RB (see Sec. III B). We then performed a series of tomography experiments. We measured 7140 settings consisting of sequences of primitive gates, with a frequency recalibration after every sequence. We randomly generated the set of tomographic settings of length  $L = 0$  (i.e., prepare and measure) to length  $L = 14$ . As we explain in Sec. III, FBT is not prescriptive as to the sequences that need to be performed, but we note that we have chosen shorter-length sequences for tomography compared with the longer sequences used in RB. The circuits contained sequences of randomly selected primitive gates in Eq. (1), and data acquisition took approximately 86 hours. We note that this experiment goes well beyond what would traditionally be considered as an informationally complete set of measurements.

Results are presented in Fig. 2. Between each setting, the measurement outcomes are used to update the posterior distribution of the gate sets. This update then improves the linear model for the subsequent data. Because we are using a linear approximation to the true nonlinear model, we end up with two sources of error: one from standard shot noise and the other from the linearization. The latter we refer to as the approximation error, and if this error is ignored (as in Refs. [17,37]) it can lead to overfitting, especially if the fitting function is linearized about the identity noise channel. The FBT protocol is robust to both of these error sources. Once our estimates are accurate enough, the approximation error can become insignificant relative to the shot noise, and so if certain conditions are met we can omit the approximation error from the update, which speeds up the inference significantly. The point at which we omit the approximation error is illustrated in Fig. 2. (We define these conditions more formally in Sec. III.) This means the posterior distribution has contracted around the “true” gate set such that our linearization performs as well as the exact nonlinear model, given the amount of shot noise in our measurements.

For this experiment, the data were analyzed after all measurements were taken. Nonetheless, we can simulate the real-time use of FBT, and note that it is capable of updating estimates faster than measurements can be taken,

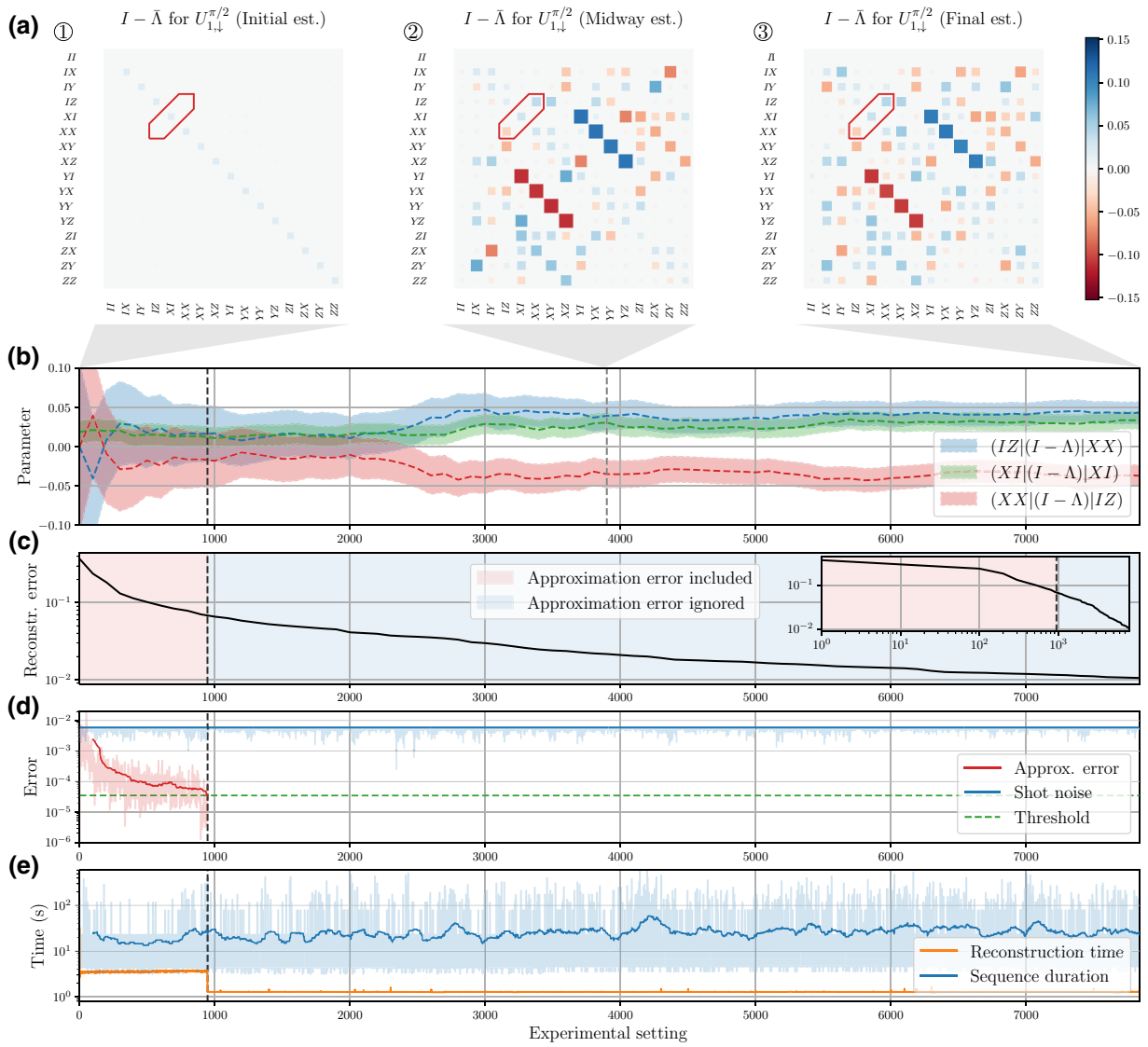


FIG. 2. High-level schematic of the FBT protocol and experiment, with the  $U_{1,\downarrow}^{\pi/2}$  gate used as an illustrative example. (a) The estimated noise channel  $\bar{\Lambda}$ , presented as a Hinton diagram of the noise residual  $(I - \bar{\Lambda})$ , for the  $U_{1,\downarrow}^{\pi/2}$  gate at three key stages during data acquisition: the initial estimate based only on the fidelity; the estimate after half of the total number of sequences; and the final estimate after all sequences have been measured. The noise residual  $(I - \bar{\Lambda})$  reveals greater, clearer detail than the Pauli transfer matrix itself. (b) The estimates for three parameters [indicated in (a)], with mean values and uncertainties, as they are updated throughout the experiment. The three parameters were chosen only for clarity of presentation, having nonzero, nonoverlapping mean values. (c) The expected reconstruction error, as determined by the trace of the covariance for the gate, throughout the experiment. Inset: as in the main plot, but using a logarithmic scale. (d) The magnitude of the two error processes (the approximation error and shot noise) throughout the experiment, characterized by the trace of their respective covariances. The approximation error is calculated as a moving average with window size of 100 (the darker line shows the moving average, the lighter lines the raw values). A threshold for approximation error, at two orders of magnitude less than the shot noise, indicates when it can be neglected. For ease of reference this point is indicated by a vertical dotted line in all graphs (b)–(e). (e) The speed of the FBT protocol compared to the time to perform each sequence in the experiment. The light blue lines show the raw time of execution of each sequence (which varies with the length of any particular sequence), the darker line the moving average.

despite not optimizing the code for speed. Figure 2(e) shows that, on average, each sequence takes approximately 26 s. While the approximation error is being sampled, FBT takes approximately 4 s to update for each sequence, and

this reduces to approximately 1 s once the approximation error becomes negligible. This affirms the potential for FBT to be used in real time while a device is being operated.

Figure 2 shows a high-level detail of the experiment and the corresponding reconstruction error throughout the experiment. The benefit of having a full Bayesian model for the gate set is that we always have a full covariance describing the uncertainty in the fits, including correlations between parameters as well as between separate gates. The reconstruction error shown in Fig. 2 is given by the mean square error of the estimate. The log scale inset plot shows the rate at which the fits improve throughout the experiment. As we expect, the mean square error approaches a central limit rate of approximately  $N_{\text{settings}}^{-1}$ , where  $N_{\text{settings}}$  is the number of experimental settings measured. The full set of tomographic results for the gate set can be found in Appendix C.

From the gate-set estimates we can extract metrics and benchmarks of interest. Gate dependent fidelities  $f(G)$ , unitarities  $u(G)$ , along with error bars, are all available to be extracted from the posterior distribution. In place of the unitarity we present the gate-dependent *incoherence* [25],  $\omega(G)$ , defined as

$$\omega(G) := \frac{d-1}{d} (1 - \sqrt{u(G)}). \quad (2)$$

The incoherence is bounded above by the infidelity, and gives a measure of how much of the gate infidelity can be attributed to incoherent errors and stochastic noise sources. Figure 3(a) presents a summary of the gate-dependent primitive infidelities and incoherences extracted by FBT.

### E. Bell-state tomography

Our FBT protocol can also be adapted to perform state tomography self-consistently. Although conceptually simpler than process tomography, state tomography generally requires the use of (noisy) gates to measure an informationally complete set of observables. Typically, state

tomography is performed under the assumption that the uncharacterized, noisy gates are ideal, or that all measurements are described by a single “measurement fidelity.” To perform state tomography self-consistently, it is necessary to perform tomography on the gate set first.

Having performed a characterization of the primitive gate set using FBT, we can directly infer the states that can be generated in our device. A common and important characterization of a quantum processor is the quality of the entangled states that can be prepared. Figure 3(b) presents the fidelities inferred from self-consistent state tomography for the four different Bell states  $\Phi^+$ ,  $\Phi^-$ ,  $\Psi^+$ ,  $\Psi^-$ , generated from the same circuits as in Ref. [30]. These fidelities range from 94.6% to 98.3%, with concurrences ranging from 88.3% to 92.0%. We note that these Bell-state fidelities and concurrences are larger than those presented in Ref. [30]. These improved metrics are not due to any performance improvement in the experiment; rather, they reflect improved statistical analysis as a result of using FBT.

### F. Measurement tomography

A common approach [25,30,35] to characterizing readout assignment errors is to prepare and measure the various basis states by using gates to rotate between the desired basis state and the computational basis. These assignment errors can be used to define a correction matrix which can be applied to measurement data. This approach, however, violates self-consistency as many of these measurements will also contain gate errors.

To complete the self-consistent characterization of our device, we can add the measurement noise channel to the gate set, treating it as another random variable in the Bayesian model. We model the SPAM as a single noise channel  $\Lambda_E$  experienced by the state prior to readout, as

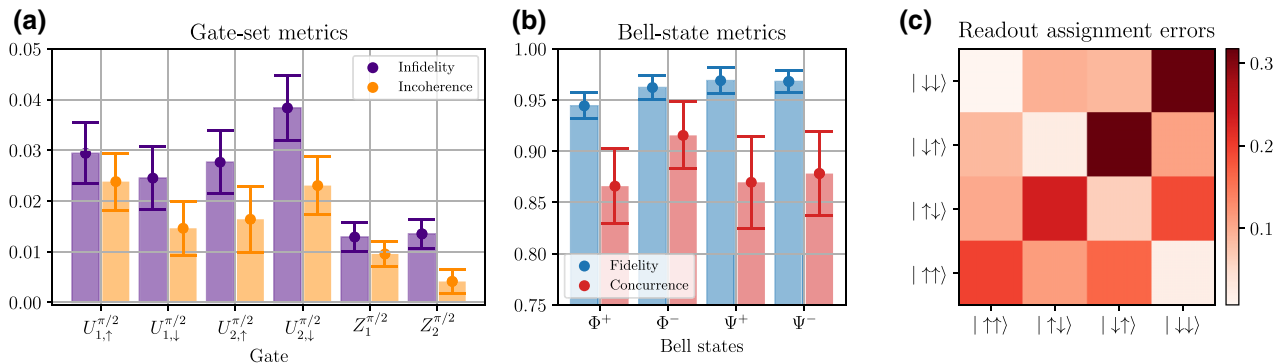


FIG. 3. Device benchmarks. (a) Gate infidelity and incoherence for each individual primitive gate, extracted from the final posterior distribution. All error bars are two standard deviations, which corresponds to a 95% credible interval. (b) Fidelity and concurrence for each of the four Bell states  $\Phi^+$ ,  $\Phi^-$ ,  $\Psi^+$ ,  $\Psi^-$ . Having tomographic estimates for the primitive gates allows us to infer SPAM-free estimates of these benchmarks, compared to the conservative estimates in Ref. [30] via state tomography. (c) Readout assignment matrix indicating assignment error rates for different preparations and measurements. The  $x$  axis represents the prepared two-qubit state and the  $y$  axis represents the corresponding measurement basis.

in this device the readout errors dominate errors in state initialization. By including this measurement noise into the model, we are able to separate SPAM errors from gate errors. This is aided by incorporating the RB inference into the prior. By providing the SPAM-free metric data into our model, we can more easily distinguish gate errors from SPAM errors in the subsequent experiments. From the full SPAM noise channel  $\Lambda_E$  we can infer the readout assignment error matrix for the spin states, shown in Fig. 3(c).

### G. Repurposing RB data

In the experiment reported above, we acquired new tomography data following the initial RB experiment in order to characterize the gate set. One of the key benefits of FBT is that it is able to extract tomographic information from any primitive gate sequence. Since the RB sequences can be decomposed from Clifford circuits into sequences of primitive gates, we can bootstrap this data and perform FBT over each primitive sequence as depicted in Fig. 4. We can extract more information about the gates by asking more of our RB data than fitting only the usual decay curve, and in particular we can extract information about the coherent action of the noise present in the gates beyond the decoherent components that RB otherwise measures.

In this RB data, we have access to 3430 measurement settings ranging from primitive length of 1 up to 1069 primitive gates, all sampled for 125 shots. Figure 4 shows the reconstruction of the  $U_{1,\uparrow}^{\pi/2}$  gate throughout the experiment, and we observe that we are able to reconstruct the process matrix to a moderate accuracy from the RB data alone. We note that, due to the longer sequences used in RB, a longer period of data acquisition is needed before the approximation error becomes negligible. Regardless, even when the approximation error is large, FBT remains robust to it. The fact that this was achieved with a standard RB dataset, from which a single average gate fidelity is usually all that is extracted, highlights the utility and flexibility of FBT [38].

## III. METHODS

Here we provide details of the FBT protocol and how it performs.

### A. Tomography and self-consistency

Consider a system consisting of  $n$  qubits. Let  $\mathcal{P} = \{P_i\}$  denote the set of all  $n$ -qubit normalized Pauli operators. Suppose we have a set of  $N_G$  gates  $\mathcal{G} = \{\tilde{G}_i\}$  that we can perform and that we are interested in characterizing.

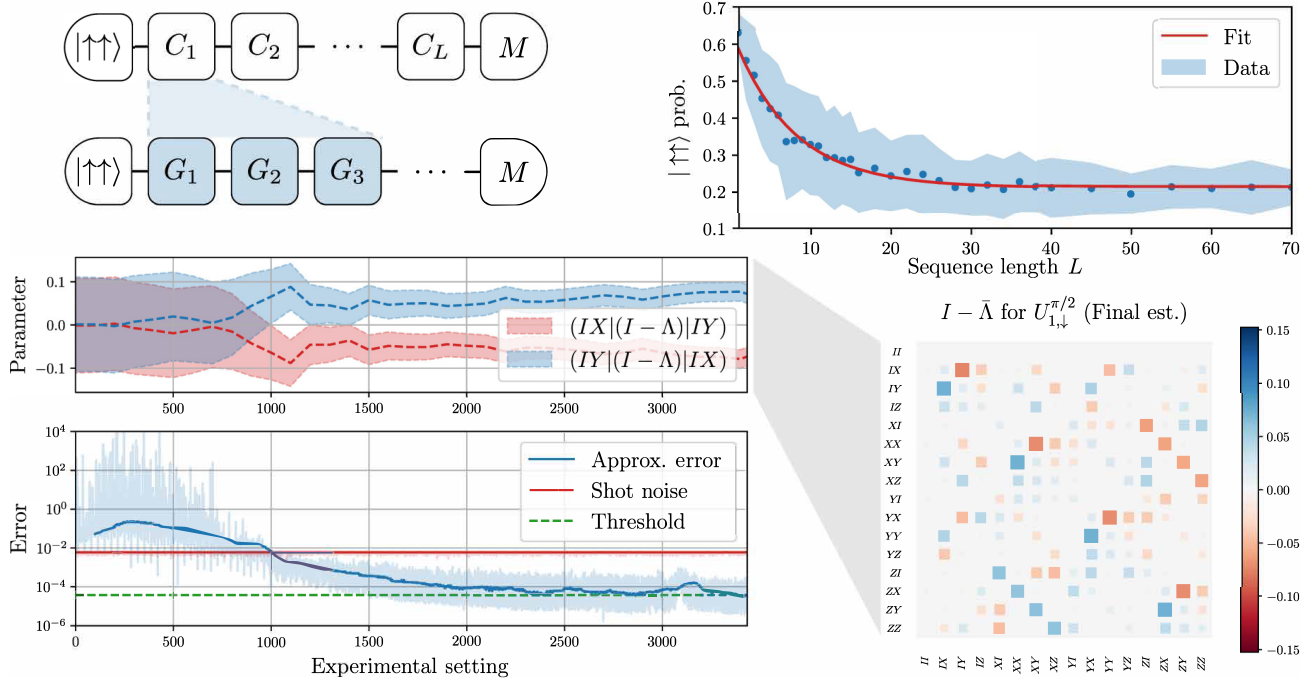


FIG. 4. Repurposing randomized benchmarking data for FBT. (a) The RB sequences can be repurposed as tomographic data for FBT by decomposing Clifford gates  $C_i$  into their primitive components  $G_i$ . (b) The RB decay curve for the projected state probability as a function of the number of Clifford gates in each sequence. (c) The iterative estimates for two parameters [selected as examples and indicated in (d)] throughout the RB experiment. (d) The final estimate for the  $U_{1,\downarrow}^{\pi/2}$  gate, presented as a Hinton diagram of the noise residual  $(I - \bar{\Lambda})$ . (e) The magnitudes of the two error processes. The approximation error is significantly larger in the repurposed RB sequences than the tomography data due to the use of much longer sequence lengths in RB.

Each  $\tilde{G}_i \in \mathcal{G}$  is formally described by a completely positive trace-preserving (CPTP) map. We will work in the Pauli transfer matrix (PTM) representation, which, as detailed in Ref. [14], has many useful features that we can take advantage of when parameterizing our channels. Standard quantum process tomography (QPT) attempts to reconstruct this matrix by performing an informationally complete set of measurements  $\{M_i\}$  and input preparations  $\{\rho_j\}$ . Given measurements of a gate  $\tilde{G}$  of the form

$$\begin{aligned} m_k &= \langle \langle M_i | \tilde{G} | \rho_j \rangle \rangle = \text{Tr}([M_i \otimes \mathbb{I}] \circ \tilde{G} \circ [\mathbb{I} \otimes \rho_j]) \\ &= \text{Tr}([\mathbb{I} \otimes \rho_j] \circ [M_i \otimes \mathbb{I}] \circ \tilde{G}) = \text{Tr}([M_i \otimes \rho_j] \circ \tilde{G}) \\ &= \langle \langle M_i \otimes \rho_j | \tilde{G} \rangle \rangle =: A_k | \tilde{G} \rangle, \end{aligned} \quad (3)$$

where we have used the cyclic property of the trace,  $\mathbb{I}$  is the appropriately normalized identity,  $| \tilde{G} \rangle \in \mathbb{C}^{d^2}$  denotes the  $\text{vec}(\tilde{G})$  operation using the row-major order convention and  $(i, j) \mapsto k$  is some labeling of the experimental settings.

If we stack our measurements such that

$$m := \begin{bmatrix} m_1 \\ \vdots \\ m_N \end{bmatrix}, \quad A := \begin{bmatrix} A_1 \\ \vdots \\ A_N \end{bmatrix}$$

then

$$m = A | \tilde{G} \rangle \quad (4)$$

defines a linear inverse problem in the entries of  $| \tilde{G} \rangle$ . The estimator for standard QPT is usually taken as the maximum likelihood, given by

$$\hat{G} = \underset{X \in \mathbb{C}^{d \otimes d}}{\text{argmin}} \left\| m - A | X \rangle \right\|_2^2 \quad (5)$$

$$= (A^\dagger A)^{-1} A^\dagger m. \quad (6)$$

where Eq. (6) makes use of the Moore-Penrose pseudoinverse.

The shortcomings of this approach have been well explored: with high probability the estimator will not be a CPTP map; the estimated fidelity can disagree with estimates from RB; the assumption is made that measurement errors are independent and identically distributed; etc. However, the most concerning shortcoming is the lack of *self-consistency*, introduced in Ref. [17]. This arises because the input states and measurements are generated by gates that have yet to be characterized, and which are themselves noisy.

Self-consistent process tomography addresses this problem by performing a simultaneous reconstruction of the entire gate set, including preparations and measurements.

Without loss of generality we can decompose each noisy gate  $\tilde{G}_i = \Lambda_i G_i$  into the ideal gate  $G_i$  followed by a noise channel  $\Lambda_i$ . Then each experimental setting is described by a sequence of  $N_k$  gates where

$$\begin{aligned} m_k &= E \prod_{i \in S_k} \tilde{G}_i | \rho_0 \rangle \rangle \\ &= E \prod_{i \in S_k} \Lambda_i G_i | \rho_0 \rangle \rangle \\ &=: \mathcal{A}_k(\lambda). \end{aligned} \quad (7)$$

Here,  $S_k : \mathbb{Z}_{N_k} \rightarrow \mathbb{Z}_{N_G}$  is a tuple of indices corresponding to which gates were performed in the  $k$ th measurement setting and  $\lambda := [\langle \langle \Lambda_1 |, \dots, \langle \langle \Lambda_{N_G} | \rangle \rangle]^\dagger$  contains all of the unknown parameters of the noise channels. The readout is a positive operator valued measure (POVM)  $\{E_1, \dots, E_M\}$  (often the set of  $n$ -qubit computational basis projective measurements) which can be combined in matrix form as

$$E := \begin{bmatrix} \langle \langle E_1 | \\ \vdots \\ \langle \langle E_M | \end{bmatrix}.$$

In this self-consistent picture, there is no longer a single gate for which we collect tomographic data by varying the inputs and outputs. Rather, each choice of experimental setting  $m_k$  contains tomographic data of *all* gates in that setting.

The tomographic reconstruction in this self-consistent scenario is much more challenging than just applying the simple formula of Eq. (6), with a nonlinear inversion of a high-degree (with many settings) polynomial required for Eq. (7). This problem of nonlinear inversion can be addressed by linearization. The approach used in Ref. [17] expresses each noise channel  $\Lambda_i = I + \mathcal{E}_i$  and linearizes Eq. (7) to obtain

$$\begin{aligned} m_k &\approx E \prod_{i \in S_k} G_i | \rho_0 \rangle \rangle \\ &+ \sum_{j=1}^{N_k} E \left[ \prod_{\substack{i \in S_k \\ i < j}} G_i \right] \mathcal{E}_j G_j \left[ \prod_{\substack{i \in S_k \\ i > j}} G_i \right] | \rho_0 \rangle \rangle \\ &=: m_{\text{ideal}} + A_k x \end{aligned} \quad (8)$$

where  $m_{\text{ideal}}$  is the ideal (noiseless) output and now  $x = [\langle \langle \mathcal{E}_1 |, \dots, \langle \langle \mathcal{E}_{N_G} | \rangle \rangle]^\dagger$ . This linear model yields a more efficient estimation of the gate set contained in  $x$ . Computational speed is crucial so that insights from tomography can be used for error mitigation [39] before the noise in the system drifts. However, by linearizing about the identity  $I$  for each channel  $\Lambda_i$ , this approach will only be accurate



for gates with high fidelity (i.e., noise channels close to identity). However, process tomography is predominantly employed to understand noise processes for the purpose of error mitigation, and demanding high-fidelity gates is prohibitive in many experimental settings. Ideally, process tomography should be suitable for use with gates of arbitrary fidelity for the purposes of early characterization and tuneup.

To linearize the nonlinear function Eq. (7) about a better location requires some *a priori* information about the noise in the gates. In the next section we will describe how FBT uses a Bayesian approach to construct a linear model that outperforms Eq. (8). Moreover, our approach gives rise to an online model that can be updated as data become available, improving both the computational speed and the accuracy of the model.

### B. Fast Bayesian tomography

In this subsection we formally introduce the FBT protocol. Appendix B contains an overview of the entire process as well as some of the constructions we use. Here we introduce the model and assumptions of the FBT protocol. Before we can define a linear model for FBT we need to construct a Bayesian structure for the random variables involved. Different prior models that reflect physical constraints have been previously proposed [19–21,40,41]. However, these approaches all lack scalability beyond single- or two-qubit processes as they rely on sampling methods such as sequential Monte Carlo or importance sampling. We instead propose the use of a Gaussian multivariate prior in order to pursue a conjugate posterior model that is computationally tractable and scalable. Although this prior lacks the ability to constrain our parameter space to physical (CPTP) maps, we will show in Sec. B6 how this prior can still be used effectively to reflect physical constraints.

Firstly, let each noise channel  $|\Lambda_i\rangle\rangle$  be distributed as a multivariate Gaussian  $|\Lambda_i\rangle\rangle \sim \mathcal{N}(|\bar{\Lambda}_i\rangle\rangle, \Gamma_i)$  with mean  $|\bar{\Lambda}_i\rangle\rangle$  and covariance  $\Gamma_i \in \mathbb{R}^{d^2 \times d^2}$ . In plain words, the mean  $\bar{\Lambda}_i$  represents our best guess *a priori* for the noise channel, and similarly, the covariance  $\Gamma_i$  encodes how much we trust the guess. We will also abuse notation and interchangeably refer to both  $\bar{\Lambda}_i$  and  $|\bar{\Lambda}_i\rangle\rangle$  as the mean channel depending on the context. We will also include SPAM noise channels  $\tilde{E} = E\bar{\Lambda}_E$  and  $\tilde{\rho}_0 = \Lambda_\rho|\rho_0\rangle\rangle$ .

One of the best parts of using Gaussian distributions is that we can efficiently sample from our prior. We can likewise efficiently project a sample onto the convex set of CPTP maps. We can combine these to inform a prior that will preclude channels that are very far from CPTP. The goal here is to envelope the region of CPTP maps contained within the larger space of  $d^2 \times d^2$  random matrices. We will build on this in the next section, where we

consider how we can incorporate prior information from benchmarking experiments, such as RB.

With prior knowledge about our gate set we have the ability to improve our model by linearizing Eq. (7) about the mean,  $\bar{\lambda}$ . This gives us the FBT model

$$\begin{aligned} m_k &\approx E\bar{\Lambda}_E \prod_{i \in S_k} \bar{\Lambda}_i G_i \bar{\Lambda}_\rho |\rho_0\rangle\rangle \\ &+ \sum_{j=1}^{N_k} E\bar{\Lambda}_E \left[ \prod_{\substack{i \in S_k \\ i < j}} \bar{\Lambda}_i G_i \right] \mathcal{E}_j G_j \left[ \prod_{\substack{i \in S_k \\ i > j}} \bar{\Lambda}_i G_i \right] \bar{\Lambda}_\rho |\rho_0\rangle\rangle \\ &\approx \bar{m}_k + \bar{A}_k x \\ &=: \bar{m}_k + \bar{A}_k x + e_k \end{aligned} \quad (9)$$

where  $\bar{m}_k$  is the expected output of the setting,  $\bar{A}_k$  is our linear model acting on the centralized variable  $x$ , and  $e_k$  is the combined finite statistic and model error (the noise in the tomographic measurement process). It is imperative in Bayesian inference to accurately quantify the noise in the measurement process as this informs the model of how strongly data should be weighted relative to the current prior information. We have two sources of error in Eq. (9): the measurement noise due to finite averaging of experiment outcomes, and the model error incurred by the use of an approximate model. We will treat both of these errors separately,  $e_k := \epsilon_k + \eta_k$  where  $\epsilon_k$  and  $\eta_k$  are the *shot noise* and *approximation error* [42], respectively.

The approximation error is defined as

$$\eta_k(x) := \mathcal{A}_k(x + \bar{\lambda}) - (\bar{m}_k + \bar{A}_k x), \quad (10)$$

which we assume to be mutually independent of the statistical shot noise  $\epsilon_k$ . The approximation error plays an important role when using a linearized model; omitting it can lead to overfitting. We can approximate both noise processes to be Gaussian as detailed in Appendix B and it can be efficiently computed or directly sampled via our prior. Having a Gaussian noise process and prior implies the posterior is also Gaussian, with mean and covariance that we can compute in closed form.

FBT is an online protocol, meaning the posterior distribution over our gate set can be updated as measurements are received. Throughout the protocol all we need to keep track of are our prior statistics  $(\bar{\lambda}, \Gamma_\lambda)$ . When a new measurement  $m_k$  is taken, we can compute our posterior distribution  $x|m_k$  which immediately becomes the prior for the following measurement. This means the approximate model is iteratively improved between every measurement setting. Appendix B gives the explicit construction of this Bayesian model.

### C. Initialization via benchmarking

One of the key benefits of having a Bayesian model is that we have the ability to include prior information, including benchmarks such as the average gate fidelity obtained from RB. Although these benchmarks are typically limited in their ability to diagnose detailed noise processes, they do provide tight error bounds on gate-set metrics. The majority of randomized benchmarking experiments are also robust to SPAM. This means that prior information extracted from benchmarking can inherit this SPAM-free property. Also, the more information we use prior to performing tomography, the better our approximate model will be, meaning less approximation error and a shorter time required before the approximation error can be ignored. Finally, the random sequences performed in RB are suitable tomographic settings for FBT, and can be reused as tomographic experiments, as we demonstrated in Sec. II G.

We briefly review the properties of data from standard RB [5], which yields an average fidelity over our gate set obtained by fitting an exponential decay curve to the survival probability of Clifford circuits of increasing length. Specifically, the average gate fidelity is

$$F_{\text{avg}}(\mathcal{G}) = \frac{1}{N_G} \sum_{G \in \mathcal{G}} f(G) \quad (11)$$

where

$$f(G) = \frac{\text{Tr}(\Lambda_G) - 1}{d^2}. \quad (12)$$

The estimate obtained by this procedure will contain uncertainty, which we take as our prior belief. That is, we now consider the average gate fidelity to be a random variable,  $f_{\text{avg}}$ , with some prior distribution. The distribution we choose to encode this average fidelity prior may vary on how accurate (i.e., how many sequences and shots) the RB experiment was, and we only require that this distribution can be efficiently sampled. Provided sufficient samples are taken, the distribution of  $f_{\text{avg}}$  should be well described by a univariate Gaussian distribution with mean  $\bar{f}$  and variance  $\sigma_f^2$ .

With this prior information from RB, we can now construct priors for each gate in our gate set. Given a sample of our gate set  $\mathcal{G} = \{G\}$  from the prior (which is not necessarily physical), we can take the projection onto the set of CPTP maps with a given average gate fidelity  $f$  by the semidefinite program

$$\begin{aligned} P_f(\mathcal{G}) = \operatorname{argmin}_{\{X_i\}} \sum_i \|\Lambda_i - X_i\| \\ \text{subject to } F_{\text{avg}}(\{X_i\}) = f, \\ \text{choi}(X_i) \geq 0. \end{aligned} \quad (13)$$

---

**Input:**  $\bar{x}, \Gamma_x, \bar{f}, \sigma_f^2$

- 1: #  $x \sim \mathcal{N}(\bar{x}, \Gamma_x)$  is the initial prior distribution
  - 2: #  $f_{\text{avg}} \sim \mathcal{N}(\bar{f}, \sigma_f^2)$  RB average fidelity prior distribution
  - 3:
  - 4: **for**  $i$  in  $1, \dots, N_{\text{samples}}$  **do**
  - 5:   Sample an average fidelity  $f$
  - 6:   Sample a gate set  $\mathcal{X}$
  - 7:   Store the projected sample  $P_f(\mathcal{X})$
  - 8: **end for**
  - 9: take  $(\bar{x}, \Gamma_x)$  to be the sample mean and covariance of the projected samples
- Output:**  $(\bar{x}, \Gamma_x)$
- 

Algorithm 1. RB UPDATE( $\bar{x}, \Gamma_x, \bar{f}, \sigma_f^2, N_{\text{samples}}$ ) randomized Benchmarking prior update

Algorithm 1 details how we can create a prior distribution that envelopes the set of CPTP maps such that the average fidelity statistics are consistent with results obtained via RB.

Algorithm 1 provides a mechanism to inject the average fidelity of a gate set into the FBT protocol. We should note that other benchmarks, such as unitarity [8], can be used similarly, by adapting the projection in Eq. (13) accordingly.

### D. Including SPAM errors

SPAM errors are a central issue in characterizing quantum devices. However, there is a distinction that should be made between *intrinsic SPAM errors* [16] (i.e., errors in readout and state initialization) and *gate SPAM errors* that result from the imperfect gate used when, for example, preparing a specific state. Self-consistency takes care of gate SPAM errors, since this is completely captured by the modeling the noise on the gates. The intrinsic SPAM errors can be modeled as a noise channel on each of the state and measurement (i.e.,  $\tilde{E} = E\Lambda_E$  and  $|\tilde{\rho}\rangle\rangle = \Lambda_\rho|\rho\rangle\rangle$ ). However, we note that there is a gauge degree of freedom between these two operators. Theoretically, these two operators can be distinguished by using ancilla qubits [43]. In such a setting, this could then be used to inform separate prior distributions on each of the noise channels  $\Lambda_\rho$  and  $\Lambda_E$ . This is, however, beyond the scope of this paper (as well as most other self-consistent tomography methods).

Measurement errors are often characterized by estimating error rates for all computational basis states and measurements [25,30,35], which implicitly assumes that the measurement errors are classical. Moreover, this approach requires the use of gates to initialize different input and output states, meaning the corresponding measurement fidelities will include gate noise, leading to unreliable estimates. By including a measurement noise channel in the underlying model  $\tilde{E} = E\Lambda_E$ , this additional noise channel is added to the inference problem and kept separate from gate errors

in our model. An advantage of this approach is that it allows for the reconstruction of a quantum noise channel on our measurement, provided we can make assumptions about the state preparation as in Sec. II F.

#### IV. DISCUSSION

Fast Bayesian tomography is an efficient, flexible method for self-consistent tomography, with many features that make it appealing for the rapid early characterization of quantum gates. It is designed to be used immediately following RB to provide immediate diagnostic information about gate performance and associated errors, allowing for rapid decisions on how gate fidelities can be improved. Gates that have initially low fidelity are not a barrier as in previous methods. The ability to incorporate prior information, such as the physics of state preparation and measurement, means that the gates are automatically constructed in a gauge familiar to those working with the device, without the need for any artificial gauge optimization. FBT is very flexible in terms of the data that is used; as we have shown, it can repurpose RB data, and it can be run on data from GST or other randomized sequences. When new data are available, they can be directly incorporated into FBT to provide further improved estimates. We believe that FBT fills an existing gap in the broad suite of characterization tools, complementing more prescriptive benchmarking methods such as GST.

Our demonstration using a two-qubit spin qubit device in silicon shows the utility of FBT. While this device was previously characterized using RB, we have shown how FBT can bootstrap such RB data to provide the fidelity, unitarity, and full process matrices for all primitive one- and two-qubit gates. The accuracy of our tomographic reconstructions is very high, with the approximation error due to our model becoming negligible. The Bell-state fidelities measured using FBT are higher than the conservative estimates in Ref. [30], because of the self-consistent nature of FBT. Although we do not explore it further here, the detailed diagnostic information about the two-qubit gates in this device may potentially be used to optimize the pulses and gate operating points such as to increase gate fidelities further.

#### ACKNOWLEDGMENTS

We thank Steve Flammia, Thomas Smith, Owen Dillon, Thaddeus Ladd, and Matthew Otten for helpful discussions. This work was supported by the Australian Research Council via EQUUS Project No. CE170100009, CQC2T Project No. CE170100012, and Laureate Fellowship Project No. FL190100167. R.H. acknowledges support from the Sydney Quantum Academy. We acknowledge support by the NSW Node of the Australian National Fabrication Facility. We also acknowledge support from

the U.S. Army Research Office Grant No. W911NF-17-1-0198. The views and conclusions contained in this paper are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government.

#### APPENDIX A: EXPERIMENTAL SETUP

The quantum dots are formed in a 900-nm-thick isotopically enriched  $^{28}\text{Si}$  epi-layer (residual  $^{29}\text{Si}$  concentration of 800 ppm [44]) on a natural silicon substrate. They are laterally confined by aluminum gate electrodes fabricated using multilayer gate stack technology. All measurements were performed in an Oxford Instruments wet dilution refrigerator with base temperature  $T_{\text{bath}} \approx 30$  mK and electron temperature  $T_e \approx 180$  mK. Battery-powered voltage sources (Stanford Research Systems, SIM928) and an arbitrary waveform generator (LeCroy ArbStudio, 1104 AWG) were used to generate dc voltages and voltage pulses, respectively, which were added through combiners with attenuation ratios 1 : 5 for dc voltages and 1 : 25 for voltage pulses. Low-pass filters were included for slow and fast lines (10 Hz to 80 MHz). An Agilent E8267D microwave vector signal generator was used to deliver ESR pulses to the on-chip microwave antenna after being attenuated at the 1.5 K stage (10 dB) and the 30 mK stage (3 dB). We used the internal arbitrary waveform generator of E8267D to perform single-sideband modulation (by mixing the carrier with the in-phase and quadrature signals) and generate the four different ESR-drive frequencies.

#### APPENDIX B: BAYESIAN MODEL FOR FBT

##### 1. Overview

Overall FBT algorithm:

- (a) Create initial “reasonable guess” prior estimate of the gates, for example from fidelity estimates.
- (b) Update prior estimates using data such as that from randomized benchmarking (see Algorithm 1).
- (c) For each experimental setting, update gate estimates and covariances using Algorithm 2.

##### 2. Construction of the likelihood

We want to construct a Bayesian model that will give an updated estimate for our gate set given a new measurement  $m$ . In general,  $m$  could be a set of measurements, but as we are only interested in running the FBT protocol online, it suffices to consider the case where  $m$  is the measurement outcome from a single experimental setting. This amounts to determining the conditional probability distribution of  $x$  given  $m$ , known as the *posterior* distribution which we will

---

**Input:**  $m_k, \bar{x}, \Gamma_x, S_k$ 

- 1: #  $m_k$ , the data from current measurement
  - 2: #  $x \sim \mathcal{N}(\bar{x}, \Gamma_x)$  is the initial prior distribution
  - 3: #  $S_k$ , the current gate sequence
  - 4: Sample joint distribution of  $x, \eta_k$
  - 5: Compute noise statistics  $\Gamma_{\epsilon_k}$
  - 6: Compute linear model  $\bar{A}_k$  and  $\bar{m}_k$
  - 7:  $\Gamma_{e_k|x} \leftarrow \Gamma_{\epsilon_k} + \Gamma_{\eta_k} - \Gamma_{\eta_k} \Gamma_x^{-1} \Gamma_{x\eta}$
  - 8:  $L_{e|x} \leftarrow \sqrt{\Gamma_{e|x}^{-1}}$  via Cholesky decomposition
  - 9:  $L_x \leftarrow \sqrt{\Gamma_x^{-1}}$  via Cholesky decomposition
  - 10:  $\bar{B} \leftarrow \begin{bmatrix} L_{e|x}(\bar{A} + \Gamma_{\eta_k} \Gamma_x^{-1}) \\ L_x \end{bmatrix}$
  - 11:  $y \leftarrow \begin{bmatrix} m_k - \bar{m}_k - \bar{\eta}_k \\ 0 \end{bmatrix}$
  - 12:  $\Gamma_x \leftarrow \left( \Gamma_x^{-1} + \bar{A}_k^\dagger \Gamma_{e_k|x}^{-1} \bar{A}_k \right)^{-1}$
  - 13:  $\bar{x} \leftarrow (\bar{B}^T \bar{B})^{-1} \bar{B}^T y$
- Output:**
- $(\bar{x}, \Gamma_x)$
- 

Algorithm 2. FBT( $m_k, \bar{x}, \Gamma_x, S_k$ ) fast Bayesian tomography

denote as  $\pi(x|m)$ . By Bayes' theorem we have

$$\pi(x|m) \propto \pi(m|x)\pi(x) \quad (\text{B1})$$

where  $\pi(m|x)$  and  $\pi(x)$  are the likelihood and prior distributions, respectively. We begin by determining the likelihood for our approximate model.

Firstly, note that the linearization that we have in Eq. (9) is affine in our gates, and linear in  $x$  containing the parameters of the centralized random variables  $\{\mathcal{E}_i\}$ . We begin by assuming that our random variables  $m$ ,  $x$ , and  $e$  are jointly normal with probability density  $\pi(m, x, e)$ . Through repeated use of Bayes' theorem observe that

$$\begin{aligned} \pi(m, x, e) &= \pi(m|x, e) \pi(e|x) \pi(x) \\ &= \pi(m, e|x) \pi(x), \end{aligned}$$

hence

$$\pi(m, e|x) = \pi(m|x, e) \pi(e|x). \quad (\text{B2})$$

Using our model, the distribution of our data  $m$  given both the unknown gate parameters  $x$  and the noise  $e$  is completely determined. This means

$$\pi(m|x, e) = \delta(-\bar{m} - \bar{A}x - e). \quad (\text{B3})$$

Hence, combining Eqs. (B2) and (B3), we can marginalize over the noise to get our likelihood

$$\begin{aligned} \pi(m|x) &= \int \pi(m|x, e) \pi(e|x) de \\ &= \int \delta(-\bar{m} - \bar{A}x - e) \pi(e|x) de \\ &= \pi_{e|x}(-\bar{m} - \bar{A}x|x). \end{aligned}$$

The above distribution  $\pi_{e|x}(-\bar{m} - \bar{A}x|x)$  is the conditional distribution of  $e|x$ . Usually noise in estimation problems can be assumed to be mutually independent of the unknown parameters; however, the approximation error  $\eta$  depends explicitly on our prior.

Hence, we have the likelihood

$$m|x \sim \mathcal{N}(m - \bar{A}x - \bar{e}|x, \Gamma_{e|x}). \quad (\text{B4})$$

In the next subsection we will compute the statistics of this conditional noise process  $e|x = \epsilon|x + \eta|x$ .

### 3. Noise processes

In our model we have two separate noise processes: the shot noise and the approximation error. To compute the distribution of the noise consider the joint distribution

$$\begin{aligned} &\pi(x, \eta, \epsilon) \\ &\propto \exp \left( -\frac{1}{2} \begin{bmatrix} x - \bar{x} \\ \eta - \bar{\eta} \\ \epsilon - \bar{\epsilon} \end{bmatrix}^T \begin{bmatrix} \Gamma_x & \Gamma_{x\eta} & \Gamma_{x\epsilon} \\ \Gamma_{\eta x} & \Gamma_\eta & \Gamma_{\eta\epsilon} \\ \Gamma_{\epsilon x} & \Gamma_{\epsilon\eta} & \Gamma_\epsilon \end{bmatrix}^{-1} \begin{bmatrix} x - \bar{x} \\ \eta - \bar{\eta} \\ \epsilon - \bar{\epsilon} \end{bmatrix} \right) \end{aligned} \quad (\text{B5})$$

where each block in the joint covariance denotes the corresponding covariance matrix, that is,  $\Gamma_{\eta x} = \mathbb{E}[(\eta - \bar{\eta})(x - \bar{x})^T]$ . Recall that we are interested in determining the distribution of  $e|x = \epsilon|x + \eta|x$ .

Using standard identities for Gaussian conditional distributions [42], we can show

$$e|x \sim \mathcal{N}(\bar{e}|x, \Gamma_{e|x}) \quad (\text{B6a})$$

where

$$\bar{e}|x = \bar{\eta} + \bar{\epsilon} + (\Gamma_{\eta x} + \Gamma_{\epsilon x}) \Gamma_x^{-1} x \quad (\text{B6b})$$

and

$$\Gamma_{e|x} = \Gamma_\eta + \Gamma_\epsilon - (\Gamma_{\eta x} + \Gamma_{\epsilon x}) \Gamma_x^{-1} (\Gamma_{x\eta} + \Gamma_{x\epsilon}). \quad (\text{B6c})$$

Fortunately, all these distributions are easily and efficiently sampled due to being multivariate Gaussians. This can be achieved by sampling from the prior and directly computing the corresponding samples of  $\eta$  and  $\epsilon$  using the exact and approximate forward models  $\mathcal{A}(x)$  and  $\bar{A}x$  which we detail further in the next section. This gives us access to the full joint distribution  $\pi(x, \eta, \epsilon)$  in Eq. (B5) from which we can compute all of the necessary statistics for the likelihood.

We say that the shot noise *dominates* the approximation error [42] if

$$\text{Tr}(\Gamma_\epsilon) \gg \text{Tr}(\Gamma_\eta) + \|\bar{\eta}\|_2^2. \quad (\text{B7})$$

Initially, with little or no prior information, this is will not be the case, especially for long sequences such as we see

in Fig. 4(e). If Eq. (B7) is not satisfied, it is imperative that the approximation errors are quantified and accounted for. However, if the shot noise does dominate the approximation error, then the approximation is accurate enough for the approximation error to be neglected. This means that our prior is “close enough” to the true gate set that the linear model, relative the statistical shot noise, is as good as the nonlinear model.

The measurement noise  $\epsilon$  in our model comes from the nature of quantum measurement. Every setting requires  $N$  repetitions of the same experimental setting: initialize  $\rho_0$ , apply the sequence of gates from  $S$ , read out and collate the output of each run. In this way our data is inherently multinomial, where the bias probabilities are determined by the Born probabilities of the POVM operators  $p = [\text{Tr}(E_1\rho), \dots, \text{Tr}(E_M\rho)]$ . Using central limit theorem arguments, we can assume that the corresponding noise  $\epsilon$  is well approximated as a multivariate Gaussian random variable with mean zero, provided  $N$  is large enough and the device is not completely noiseless.

We make this assumption for the two-qubit gate-set results shown in Sec. II. This simplifies the computation of the noise as only the approximation error needs to be computed and we instead have  $e|x \approx \epsilon + \eta|x$ . The multinomial covariance is given by

$$\Gamma_{ij} = \begin{cases} Np_i(1-p_i) & i=j \\ -Np_i p_j & i \neq j. \end{cases} \quad (\text{B8})$$

To estimate the Born probabilities we can average the outputs and take  $p \approx m$ . This means that the noise process  $\epsilon \sim \mathcal{N}(0, \Gamma_\epsilon)$  where

$$\Gamma_\epsilon = \frac{\Gamma}{N} \quad (\text{B9})$$

from Eq. (B8).

Also, once we have the posterior has contracted sufficiently such that the shot noise dominates the approximation error we can omit it from the noise model completely. This also means that we no longer need to sample any noise statistics and  $e|x = \epsilon$ , significantly speeding up the protocol as shown in Fig. 2.

#### 4. Computing the noise statistics via sampling

In this section we will detail how the noise statistics can be directly estimated via sampling. Recall from Eq. (B6a) that we have

$$\bar{e}|x = \bar{\eta} + \bar{\epsilon} + (\Gamma_{\eta x} + \Gamma_{\epsilon x}) \Gamma_x^{-1} x$$

and

$$\Gamma_{e|x} = \Gamma_\eta + \Gamma_\epsilon - (\Gamma_{\eta x} + \Gamma_{\epsilon x}) \Gamma_x^{-1} (\Gamma_{x\eta} + \Gamma_{x\epsilon}).$$

As mentioned in Sec. B3, we can sample from the full joint distribution  $\pi(x, \epsilon, \eta)$  by sampling from our prior  $\pi(x)$  and

computing the corresponding noise processes given that sample.

A given sample  $x_s$  from our prior defines a specific noise configuration in our gate set. Given that noise, we can compute the expected output  $p_s = \mathcal{A}(x)$  using the exact forward model. This defines the effective bias, namely the Born probabilities, for multinomial outputs we will measure upon repeated sampling. We sample from  $\text{Multinomial}(p_s, N)$  a number of times equal to the number of experiment samples and store the statistics of  $\epsilon$ . We use the same samples  $x_s$  and  $p_s$  to compute the corresponding sample of  $\eta_s$ . Computing the sample  $\eta_s = p_s - (\bar{m} + \bar{A}x_s)$ , we likewise add this to the samples of the full joint distribution. By stacking these samples of  $x, \epsilon, \eta$  we can empirically estimate the joint covariance and mean which we can then use in the FBT protocol for this experimental setting. With access to full joint distribution we can determine the necessary components for the posterior computation, namely  $\Gamma_\eta, \Gamma_\epsilon, \Gamma_{\eta x}, \Gamma_{\epsilon x}$  and  $\bar{\eta}$ .

#### 5. Posterior distribution

Now, by Eq. (B1), we can determine the posterior  $\pi(x|m)$  by applying the prior distribution. Since our prior  $\pi(x)$  is multivariate Gaussian, it is conjugate [42] to our likelihood. This means that the posterior is likewise a multivariate Gaussian which we will denote by  $x|m \sim \mathcal{N}(\bar{x}_{\text{post}}, \Gamma_{\text{post}})$ . The mean  $\bar{x}_{\text{post}}$  is the solution to

$$\underset{x}{\text{minimize}} \|L_{e|x}(m - \bar{m} - \bar{e}|x - \bar{A}x)\|_2^2 + \|L_x x\|_2^2 \quad (\text{B10})$$

and the covariance is

$$\Gamma_{x_{\text{post}}} = (\Gamma_x^{-1} + \bar{A}^T \Gamma_{e|x}^{-1} \bar{A})^{-1}, \quad (\text{B11})$$

where  $L_{e|x}$  and  $L_x$  are the Cholesky factors of  $L_{e|x}^T L_{e|x} = \Gamma_{e|x}^{-1}$  and  $L_x^T L_x = \Gamma_x^{-1}$ , respectively.

Importantly, Eq. (B10) has a global minimum that can be efficiently computed using many approaches, including gradient descent. However, we can actually compute the minimizer for Eq. (B10) in closed form. It is simple to show that the objective function Eq. (B10) is identical to

$$\underset{x}{\text{minimize}} \|\bar{B}x - y\|_2^2 \quad (\text{B12})$$

where

$$\bar{B} := \begin{bmatrix} L_{e|x}(\bar{A} + (\Gamma_{\eta x} + \Gamma_{\epsilon x}) \Gamma_x^{-1}) \\ L_x \end{bmatrix}, \quad (\text{B13})$$

$$y := \begin{bmatrix} m - \bar{m} - \bar{\eta}|x \\ 0 \end{bmatrix}. \quad (\text{B14})$$

This is now in the form of a linear least-squares objective where we can simply take

$$\bar{x}_{\text{post}} = (\bar{B}^T \bar{B})^{-1} \bar{B}^T y. \quad (\text{B15})$$

This defines the posterior which forms the basis of FBT. However, the posterior distribution encodes our knowledge of the gate set given the data; we must also select a relevant point estimate from this distribution. The *maximum a posteriori* estimator (the mode) is most commonly used; however, in general this will not be physical gate set, as desired. We will consider this in the next subsection.

## 6. Physical priors and estimates

One of the shortcomings of the Gaussian prior model we introduced above is that it does not necessarily guarantee the underlying random variables will be physical. The multivariate Gaussian priors are used to trade off a physically constrained prior for computational tractability. We can, however, use the convexity of the set of CPTP maps to our advantage to approximately update our prior to assign low probability to regions that are far from physical.

As motioned earlier, we use the PTM representation of channels. For a CPTP map a PTM  $\Lambda \in \mathbb{R}^{d^2 \times d^2}$  takes the form

$$\Lambda = \begin{bmatrix} 1 & \vec{0}^T \\ \vec{\tau} & \mathcal{U} \end{bmatrix}. \quad (\text{B16})$$

The vector  $\tau \in \mathbb{R}^{d^2-1}$  corresponds to the nonunit part of the channel and will be the zero vector for unital channels (mapping the identity to the identity). The top row of the PTM is  $[1, 0, \dots, 0]$  which corresponds to the trace-preservation condition of quantum processes. This means that we can trivially impose the trace-preserving constraint in our prior by fixing these parameters, while also reducing the number of parameters by  $d^2$  per gate. This means that these entries are completely determined in our prior and we only have to worry about complete positivity; however, this is not as visible a constraint for the PTM. Complete positivity of a channel is equivalent to ensuring the Choi matrix is positive semidefinite.

This means we need to solve a semidefinite program to constrain our results to physical gates. To project a sample gate from our prior, say  $\Lambda_i$ , onto the set CPTP gates we take

$$\hat{\Lambda}_i = \underset{\Lambda}{\operatorname{argmin}} \|\Lambda - \bar{\Lambda}_i\|_2 \quad \text{subject to } \operatorname{choi}(\Lambda) \geq 0, \quad (\text{B17})$$

which can be efficiently solved using methods in convex optimization [45]. The semidefinite programs in the results of this paper were computed using the CVX library [46]. We

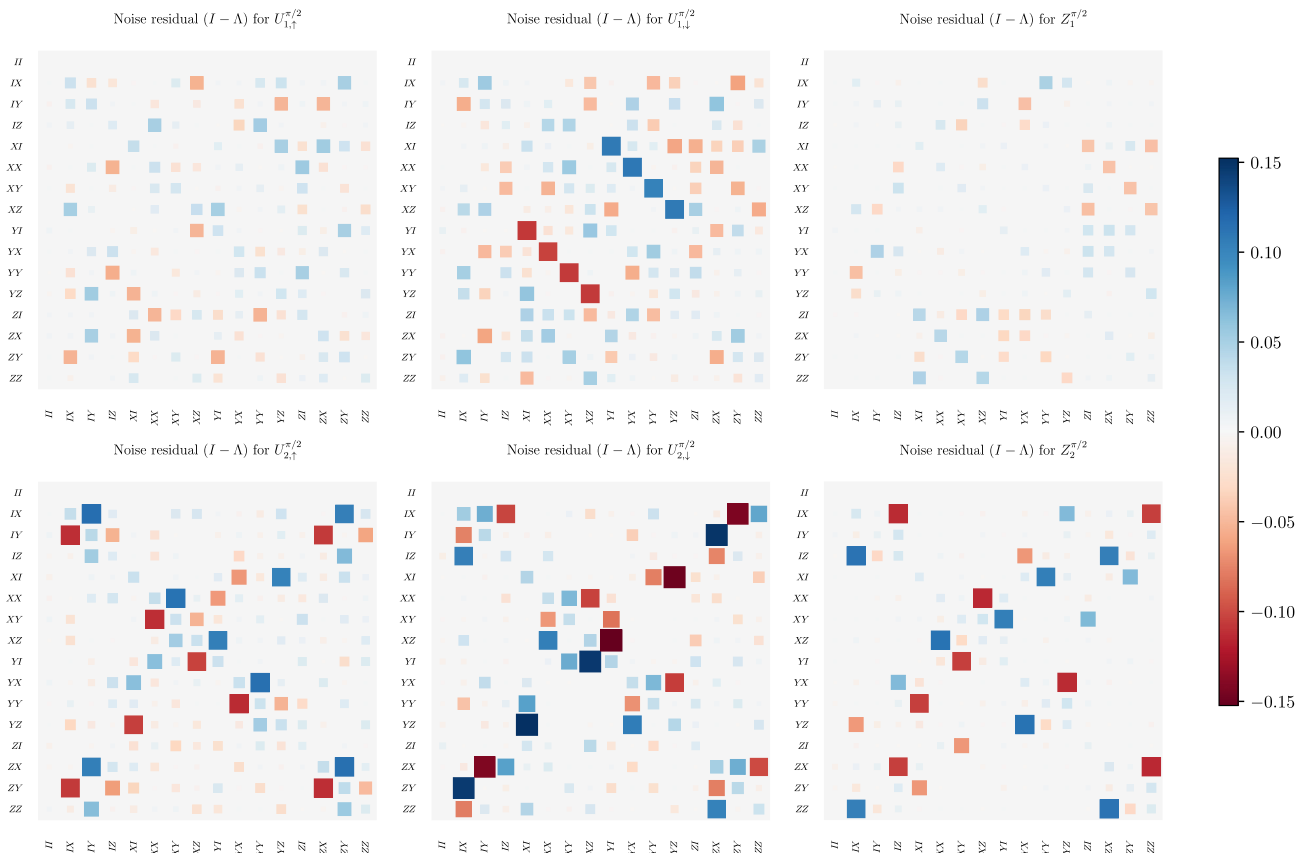


FIG. 5. Gate-set noise estimates presented as noise channel residuals  $(I - \bar{\Lambda})$ .

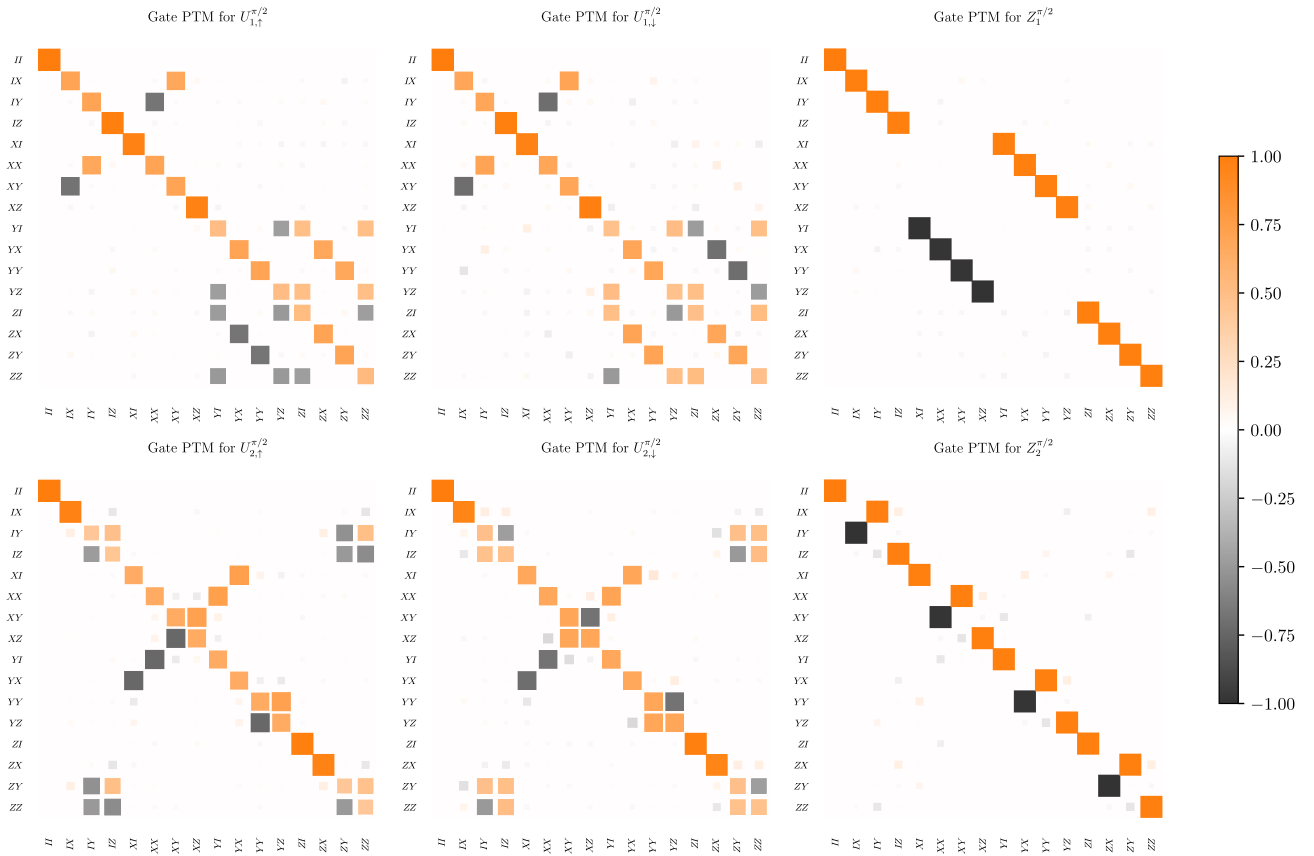


FIG. 6. Gate-set estimates presented as full gate Pauli transfer matrices.

use Eq. (B17) to project the maximum a posteriori estimate to a point estimate for FBT. Similar approaches for projected estimators have been studied in the context of state tomography [47].

### APPENDIX C: GATE-SET RESULTS

In this appendix we present the full results of all the channels from the gate-set tomography in Sec. II. For each gate in the gate set we present the noise residual,  $(I - \Lambda_G)$ , in Fig. 5 and the full gate PTM,  $\Lambda_G G$ , in Fig. 6.

- [1] D. Gottesman, Theory of fault-tolerant quantum computation, *Phys. Rev. A* **57**, 127 (1998).
- [2] 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).
- [3] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, *J. Opt. B: Quantum Semiclassical Opt.* **7**, S347 (2005).
- [4] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and Robust Randomized Benchmarking of Quantum Processes, *Phys. Rev. Lett.* **106**, 180504 (2011).

- [5] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, *Phys. Rev. A* **85**, 042311 (2012).
- [6] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, *Phys. Rev. A* **80**, 012304 (2009).
- [7] C. Granade, C. Ferrie, and D. G. Cory, Accelerated randomized benchmarking, *New J. Phys.* **17**, 013042 (2015).
- [8] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, Estimating the coherence of noise, *New J. Phys.* **17**, 113020 (2015).
- [9] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical Characterization of Quantum Devices Without Tomography, *Phys. Rev. Lett.* **107**, 210404 (2011).
- [10] S. T. Flammia and Y.-K. Liu, Direct Fidelity Estimation from few Pauli Measurements, *Phys. Rev. Lett.* **106**, 230501 (2011).
- [11] T. J. Proctor, A. Carignan-Dugas, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, Direct Randomized Benchmarking for Multiqubit Devices, *Phys. Rev. Lett.* **123**, 030503 (2019).
- [12] J. F. Poyatos, J. I. Cirac, and P. Zoller, Complete Characterization of a Quantum Process: The Two-Bit Quantum Gate, *Phys. Rev. Lett.* **78**, 390 (1997).
- [13] 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).

- [14] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki, Robust Extraction of Tomographic Information via Randomized Benchmarking, *Phys. Rev. X* **4**, 011050 (2014).
- [15] B. R. Johnson, M. P. da Silva, C. A. Ryan, S. Kimmel, J. M. Chow, and T. A. Ohki, Demonstration of robust quantum gate tomography via randomized benchmarking, *New J. Phys.* **17**, 113019 (2015).
- [16] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, *Nat. Commun.* **8**, 14485 (2017).
- [17] 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).
- [18] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, *Quantum* **5**, 557 (2021).
- [19] R. Blume-Kohout, Optimal, reliable estimation of quantum states, *New J. Phys.* **12**, 043034 (2010).
- [20] K. Schultz, Exponential families for bayesian quantum process tomography, *Phys. Rev. A* **100**, 062316 (2019).
- [21] C. Granade, J. Combes, and D. G. Cory, Practical bayesian tomography, *New J. Phys.* **18**, 033024 (2016).
- [22] A. M. J. Zwerver, *et al.*, Qubits made by advanced semiconductor manufacturing, (2021), arXiv e-prints, eid [ArXiv:2101.12650](https://arxiv.org/abs/2101.12650).
- [23] M. Veldhorst, J. C. C. Hwang, C. H. Yang, A. W. Leenstra, B. de Ronde, J. P. Dehollain, J. T. Muhonen, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak, An addressable quantum dot qubit with fault-tolerant control-fidelity, *Nat. Nanotechnol.* **9**, 981 (2014).
- [24] J. T. Muhonen, J. P. Dehollain, A. Laucht, F. E. Hudson, R. Kalra, T. Sekiguchi, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, and A. Morello, Storing quantum information for 30 s in a nanoelectronic device, *Nat. Nanotechnol.* **9**, 986 (2014).
- [25] C. H. Yang, K. W. Chan, R. Harper, W. Huang, T. Evans, J. C. C. Hwang, B. Hensen, A. Laucht, T. Tanttu, and F. E. Hudson, *et al.*, Silicon qubit fidelities approaching incoherent noise limits via pulse engineering, *Nat. Electron.* **2**, 151 (2019).
- [26] J. Yoneda, K. Takeda, T. Otsuka, T. Nakajima, M. R. Delbecq, G. Allison, T. Honda, T. Kodera, S. Oda, Y. Hoshi, N. Usami, K. M. Itoh, and S. Tarucha, A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%, *Nat. Nanotechnol.* **13**, 102 (2018).
- [27] X. Xue, M. Russ, N. Samkharadze, B. Undseth, A. Sammak, G. Scappucci, and L. M. K. Vandersypen, Computing with spin qubits at the surface code error threshold, *Nature* **601**, 343 (2022).
- [28] X. Xue, T. F. Watson, J. Helsen, D. R. Ward, D. E. Savage, M. G. Lagally, S. N. Coppersmith, M. A. Eriksson, S. Wehner, and L. M. K. Vandersypen, Benchmarking Gate Fidelities in a Si/SiGe Two-Qubit Device, *Phys. Rev. X* **9**, 021011 (2019).
- [29] J. M. Boter, X. Xue, T. Krähenmann, T. F. Watson, V. N. Premakumar, D. R. Ward, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, M. A. Eriksson, R. Joynt, and L. M. K. Vandersypen, Spatial noise correlations in a Si/SiGe two-qubit device from bell state coherences, *Phys. Rev. B* **101**, 235133 (2020).
- [30] 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, and K. M. Itoh, *et al.*, Fidelity benchmarks for two-qubit gates in silicon *Nature* **569**, 532 (2019).
- [31] R. C. C. Leon, C. H. Yang, J. C. C. Hwang, J. Camirand Lemyre, T. Tanttu, W. Huang, J. Y. Huang, F. E. Hudson, K. M. Itoh, and A. Laucht, *et al.*, Bell-state tomography in a silicon many-electron artificial molecule, *Nat. Commun.* **12**, 3228 (2021).
- [32] J. M. Elzerman, R. Hanson, L. H. Willems 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* **430**, 431 (2004).
- [33] A. J. Sigillito, M. J. Gullans, L. F. Edge, M. Borselli, and J. R. Petta, Coherent transfer of quantum information in a silicon double quantum dot using resonant SWAP gates, *npj Quantum Inf.* **5**, 1 (2019).
- [34] M. Veldhorst, C. H. Yang, J. 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* **526**, 410 (2015).
- [35] T. F. Watson, S. G. J. Philips, E. Kawakami, D. R. Ward, P. Scarlino, M. Veldhorst, D. E. Savage, M. G. Lagally, M. Friesen, and S. N. Coppersmith, *et al.*, A programmable two-qubit quantum processor in silicon, *Nature* **555**, 633 (2018).
- [36] 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).
- [37] Y. Gu, R. Mishra, B.-G. Englert, and H. K. Ng, Randomized linear gate-set tomography, *PRX Quantum* **2**, 030328 (2021).
- [38] We note that, in principle, these data could have been combined with the tomography data in Sec. IID. In fact, using FBT in this way would maximise what we can learn from RB data while supplementing it with shorter sequences to minimize the approximation error. We have not combined our data sets here due to the fact that the RB and tomography data were taken at different times (separated by approximately one week), and the experiment may have suffered some drift.
- [39] G. White, C. Hill, and L. Hollenberg, Performance Optimization for Drift-Robust Fidelity Improvement of Two-Qubit Gates, *Phys. Rev. Appl.* **15**, 014023 (2021).
- [40] I. A. Pogorelov, G. I. Struchalin, S. S. Straupe, I. V. Radchenko, K. S. Kravtsov, and S. P. Kulik, Experimental adaptive process tomography, *Phys. Rev. A* **95**, 012302 (2017).
- [41] O. Di Matteo, J. Gamble, C. Granade, K. Rudinger, and N. Wiebe, Operational, gauge-free quantum tomography, *Quantum* **4**, 364 (2020).
- [42] J. Kaipio and E. Somersalo, Statistical inverse problems: Discretization, model reduction and inverse crimes, *J. Comput. Appl. Math.* **198**, 493 (2007). special Issue: Applied Computational Inverse Problems



- [43] J. Lin, J. J. Wallman, I. Hincks, and R. Laflamme, Independent state and measurement characterization for quantum computers, *Phys. Rev. Res.* **3**, 033285 (2021).
- [44] K. M. Itoh and H. Watanabe, Isotope engineering of silicon and diamond for quantum computing and sensing applications, *MRS Commun.* **4**, 143 (2014).
- [45] M. Grant and S. Boyd, in *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, edited by V. Blondel, S. Boyd, and H. Kimura (Springer-Verlag Limited, 2008), p. 95, [http://stanford.edu/boyd/graph\\_dcp.html](http://stanford.edu/boyd/graph_dcp.html).
- [46] I. CVX Research, CVX: Matlab software for disciplined convex programming, version 2.0, <http://cvxr.com/cvx> (2012).
- [47] M. Guță, J. Kahn, R. Kueng, and J. A. Tropp, Fast state tomography with optimal error bounds, *J. Phys. A: Math. Theor.* **53**, 204001 (2020).