New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
10 June 2016

REVISED
24 August 2016

ACCEPTED FOR PUBLICATION
16 September 2016

PUBLISHED
13 October 2016

PAPER

Optimization of a solid-state electron spin qubit using gate set tomography

Juan P Dehollain[1,4,5], Juha T Muhonen[1,6], Robin Blume-Kohout[2,3], Kenneth M Rudinger[2,3], John King Gamble[2,3], Erik Nielsen[2], Arne Laucht[1], Stephanie Simmons[1,7], Rachpon Kalra[1], Andrew S Dzurak[1] and Andrea Morello[1]

[1] Centre for Quantum Computation and Communication Technologies, School of Electrical Engineering and Telecommunications, UNSW Australia, Sydney, New South Wales 2052, Australia
[2] Sandia National Laboratories, Albuquerque, NM 87185, USA
[3] Center for Computing Research, Sandia National Laboratories, Albuquerque, NM 87185, USA
[4] Present address: QuTech & Kavli Institute of Nanoscience, TU Delft, 2628 CJ Delft, The Netherlands
[5] Author to whom any correspondence should be addressed.
[6] Present address: Center for Nanophotonics, FOM Institute AMOLF, 1098 XG, Amsterdam, The Netherlands
[7] Present address: Department of Physics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

E-mail: j.p.dehollainlorenzana@tudelft.nl, enielse@sandia.gov and a.morello@unsw.edu.au

Keywords: quantum computing, silicon, tomography

Supplementary material for this article is available online

Abstract

State of the art qubit systems are reaching the gate fidelities required for scalable quantum computation architectures. Further improvements in the fidelity of quantum gates demands characterization and benchmarking protocols that are efficient, reliable and extremely accurate. Ideally, a benchmarking protocol should also provide information on how to rectify residual errors. Gate set tomography (GST) is one such protocol designed to give detailed characterization of as-built qubits. We implemented GST on a high-fidelity electron-spin qubit confined by a single ^{31}P atom in ^{28}Si. The results reveal systematic errors that a randomized benchmarking analysis could measure but not identify, whereas GST indicated the need for improved calibration of the length of the control pulses. After introducing this modification, we measured a new benchmark average gate fidelity of 99.942(8)%, an improvement on the previous value of 99.90(2)%. Furthermore, GST revealed high levels of non-Markovian noise in the system, which will need to be understood and addressed when the qubit is used within a fault-tolerant quantum computation scheme.

1. Introduction

One of the main challenges in the physical implementation of a universal quantum computer lies in designing quantum bits that meet the exquisite operation accuracies demanded by fault-tolerant quantum codes. Sophisticated quantum error correction strategies [1–3] have driven required qubit tolerances down into the realm of experimental possibility; numerical evidence suggests that gate fidelities as low as 99% might be sufficient for fault-tolerant operation [4, 5]. Gate fidelities above this value have already been claimed by several qubit systems, including liquid-state NMR [6], atomic ions [7–9], superconducting qubits [10] and single spins in semiconductors [11–13]. However, all of these demonstrations have been achieved in single or few-qubit systems and it is likely that further optimization will be required in order to maintain the high fidelities above the fault tolerance threshold as the systems scale up. While problems with low-fidelity qubits can be discerned and addressed easily, improving high-fidelity qubits is more challenging since one must characterize the qubit operation to an ever-increasing degree of accuracy. Quantum Process Tomography (QPT) [14] has been a primary method for characterizing qubit gates. By preparing a set of input states, applying the gate to be evaluated to each state and measuring the output states via quantum state tomography, the operator (G) corresponding to the applied gate can be extracted. The problem with this method is that it assumes perfect state

preparation and measurement (SPAM); therefore, the accuracy in G is limited by the ratio of SPAM to gate errors [15, 16]. Most common quantum error correction codes require much higher fidelity on the qubit logic gates than on SPAM [4, 5]. The experimental push to increase gate fidelities without the need to improve as much in SPAM, is rendering QPT obsolete as a means to characterize qubit gates. Randomized benchmarking (RB) [17, 18] is an alternative protocol for assessing the performance of qubit gates. Random gate sequences are applied to the qubit and the measurement outcome is compared to the expected result to obtain an average gate fidelity. By observing the survival probability as the number of gates in the sequences are increased, we can extract an average gate fidelity which is independent of SPAM. The downside to this protocol is that it outputs a single benchmark for qubit gate performance, without providing further insight into qubit characteristics and the nature of the errors. In order to perform qubit optimization using RB, it is necessary to perform lengthy parametric sweeps of the average gate fidelity, in order to find the optimal set of qubit parameters that maximizes the gate performance [7, 10, 13].

Gate set tomography (GST) [19] is a tool for characterizing logic operations in a qubit system. By analysing carefully constructed experiments consisting of state preparation, quantum operation sequences, and measurements, it self-consistently characterizes the experimental system. GST operates with minimal assumptions about physical characteristics of the system; it outputs a set of logical gate operators—a gate set—that models the behaviour of the device. Characteristics of the system relevant to quantum information processing can be directly extracted from the gate set, such as rotation angles, relaxation and dephasing rates, and RB decay rates. By computing the goodness of a GST fit (i.e. how well the model fits the experimental data), one reveals any deviation in the behaviour of the device from an ideal qubit system. The protocol was conceptually conceived from the fundamental ideas of self-consistent QPT [20], from which we developed the techniques to implement its current capabilities. GST has been implemented in an ion-trap qubit [19], to prove that it is a practically feasible protocol; and more recently in a solid-state charge qubit [16], as a means to extract the process fidelity of the qubit gates.

Here we reveal another layer in the capabilities of GST, by making use of its high-accuracy gate characterization to optimize the performance of a solid-state spin qubit. We first describe the physical system and the experimental methods used to perform a GST analysis of the gate fidelities. Analysing the information extracted by the GST protocol provides us with an opportunity to further optimize the qubit operation. We then complement the GST study with a new RB measurement, which highlights the improved gate fidelity obtained by applying the GST diagnostics. Finally, we discuss the current limitations to the accuracy and reliability of GST and propose future work to address these limitations.

2. Qubit description and operation

GST is *architecture-agnostic*, in that it directly characterizes the experimental system in the language of quantum information processing. Hence, to effectively interpret the GST results to help improve the experiment, it is necessary to understand the underlying physics, which we detail below.

The physical implementation of the qubit logic states—The qubit used in this study is the quantum two-level system formed by the spin-$\frac{1}{2}$ states of an electron bound to a ^{31}P donor, implanted [21] in isotopically purified ^{28}Si [22]. The fabrication and operation of the device has been described in great detail in references [23–27]. The spin energy states are split by an externally applied magnetic field $B_0 = 1.55$ T. The electron spin is coupled to the ^{31}P spin-$\frac{1}{2}$ nucleus via the hyperfine interaction $A = 98$ MHz, resulting in a two-spin, four-level system, whose eigenstates are the product states of the electron and nuclear spins. The relaxation rate of the nuclear spin is orders of magnitude smaller than the electron relaxation rate, allowing us to operate on a two-level electron-spin subsystem with the nuclear spin 'frozen' in an energy eigenstate. The qubit logic states $|1\rangle$ and $|0\rangle$ are then the eigenstates of the electron spin $|\uparrow\rangle$ and $|\downarrow\rangle$, respectively.

State preparation and measurement are performed via spin dependent tunnelling of the ^{31}P bound electron to and from a nearby single electron transistor (SET) [23, 24]. For this purpose, an aluminium gate stack is fabricated on top of an 8 nm SiO_2 layer, on the surface of the substrate above the donor. The substrate consists of a 1 μm epilayer of isotopically purified ^{28}Si with 800 ppm residual ^{29}Si concentration, grown on a natural silicon wafer [22]. The SET accumulates electrons from n^+ source-drain regions defined by phosphorus diffusion. The full device structure—as seen in figure 1—contains the SET, a set of gates (DG) used to control the electrochemical potential of the donor and an electron spin resonance (ESR) antenna used for qubit state manipulation [28]. The SET is very sensitive to changes in the electrostatic environment, providing high-fidelity detection of the charge state of the ^{31}P donor. Its electron island also acts as a reservoir to which the donor is tunnel coupled. The device is cooled down in a dilution refrigerator to an electron temperature $T_e \approx 100$ mK. At this temperature, the thermal broadening of the Fermi sea in the SET island (ΔE_F) is much smaller than the Zeeman splitting (E_Z) of the donor spin states. By tuning the donor spin electrochemical potentials ($\mu_{\uparrow,\downarrow}$) with

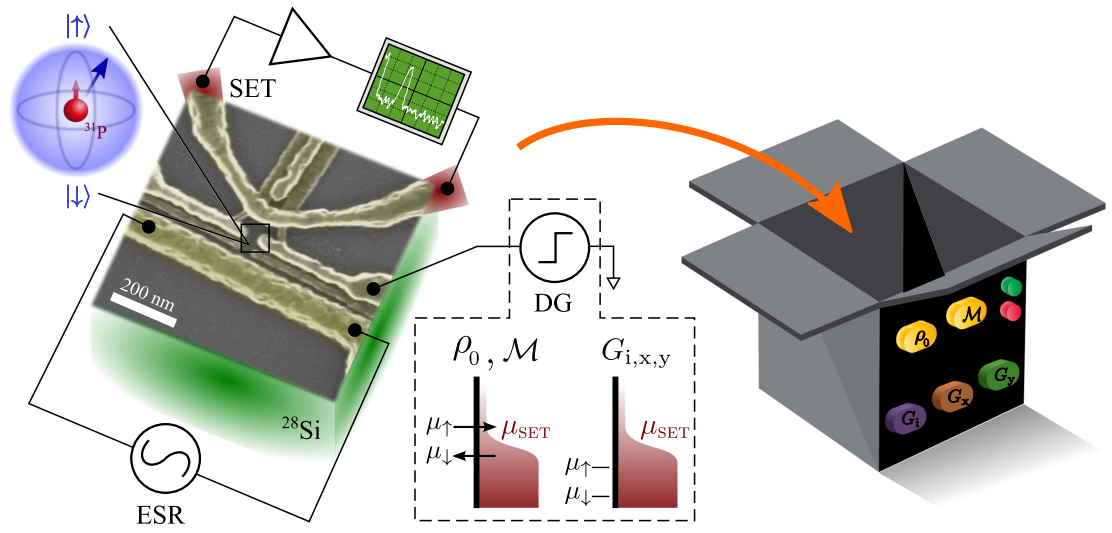


Figure 1. Diagram of qubit device and GST model of a qubit. SEM image of the on-chip gate structure of a device identical to the one used here. The aluminium gates have been false coloured for clarity. Depicted in red are the source-drain n^+ regions which connect the SET to the current measurement electronics. For initialization and measurement, the donor gates are pulsed such that $\mu_\uparrow > \mu_{SET} > \mu_\downarrow$, inducing spin-dependent tunnelling between the donor and SET. When applying a gate sequence, the DG are pulsed to higher voltage to prevent the donor electron from tunnelling to the SET. The inset diagram—zoomed from the approximate donor location—represents the Bloch sphere of the qubit, consisting on the spin of an electron confined by an implanted ^{31}P donor, with its nuclear spin frozen in an eigenstate. The GST model treats the qubit as a black box with buttons which allow to initialize (ρ_0), apply each gate in the gate set ($G_{i,x,y}$) and measure ($\mathcal{M}$) in the observable basis ($|\uparrow\rangle$ or $|\downarrow\rangle$).

respect to that of the SET island (μ_{SET}), such that $\mu_\uparrow > \mu_{SET} > \mu_\downarrow$, we restrict donor→island tunnelling to a spin-up electron, and island→donor tunnelling to spin-down electrons [24]. This allows us to perform single-shot readout and initialization with fidelities >98%.

The gate set—Logic gates are applied with ESR pulses. An oscillating magnetic field with amplitude B_1 and frequency ν, matching the qubit ESR frequency $\nu_0 = \gamma_e B_0 + A/2 \approx 43$ GHz (where $\gamma_e = 28$ GHz T^{-1} is the electron gyromagnetic ratio), will cause the spin qubit state to rotate coherently between $|\uparrow\rangle$ and $|\downarrow\rangle$. The frequency of rotation ν_1 and polar angle of the rotation axis θ can be extracted from the Rabi formula as

$$\nu_1 = \sqrt{(\nu_0 - \nu)^2 + \left(\frac{B_1}{2}\gamma_e\right)^2}, \tag{1}$$

$$\theta = \tan^{-1}\left(\frac{\frac{B_1}{2}\gamma_e}{\nu_0 - \nu}\right). \tag{2}$$

The x axis in the rotating frame of the qubit is defined by the phase of the first microwave pulse applied to it. Subsequent pulses can be phase-shifted by an angle φ_p to achieve rotations about an axis rotated by φ_p with respect to x. By controlling B_1, the pulse duration τ_p and φ_p, we can encode any arbitrary qubit state. The device contains an on-chip broadband (DC-50 GHz) antenna [28] used to transmit ESR pulses to the qubit. The antenna is connected to an Agilent E8267D vector signal generator. The ~43 GHz microwave signal is modulated by its internal dual arbitrary waveform generator, which allows precise and simultaneous control of B_1, τ_p and φ_p. For the experiments presented here, we use a fixed $B_1 \approx 12$ μT and calibrate τ_p and φ_p to apply the desired gate. For the purpose of GST we will characterize two active gates: G_x and G_y. G_x corresponds to a $\pi/2$ rotation on the x-axis of the Bloch sphere and is implemented by a pulse with $\tau_{\pi/2} = (4\nu_1)^{-1}$. G_y is a $\pi/2$ rotation on the y-axis of the Bloch sphere and is implemented by an identical pulse as G_x, but with a relative $\varphi_p = \pi/2$. Taken together these two gates are informationally complete, since they generate the single-qubit Clifford group. In addition to the active gates, we include the identity gate G_i, where no pulse is applied for the same duration $\tau_{\pi/2}$. This gate characterizes the behaviour of a qubit while it sits idle, waiting for other operations to finish in the quantum processor. The superoperators corresponding to each of these gates are displayed in table 1.

The decoherence rates—For the electron spin qubit, the free induction decay and Hahn echo decay times have been measured to be $T_2^* = 0.16$ ms and $T_2 = 1$ ms respectively [27]. Under constant driving, the qubit can maintain its coherence for up to $T_{1\rho} = 1.3$ s [29]. All of these dephasing times are shorter than the measured spin-lattice relaxation time $T_1 \approx 3$ s.

Table 1. Target superoperators for the experimental gate set in the Pauli basis, with ordering i, z, x, y.

G_i	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	G_x	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	G_y	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}$

3. Gate set tomography

GST [19] is a method for characterizing a set of quantum processes (gates), state preparation and measurement, simultaneously. GST requires no pre-calibration, and as such stands in contrast to state tomography, which requires pre-calibrated gates, and process tomography, which requires pre-calibrated SPAM. Furthermore, GST is able to obtain high-accuracy estimates efficiently, meaning that the number of experiments required to obtain a given accuracy, scales optimally with the desired accuracy. To use GST, one must perform a pre-determined set of experiments. Each experiment consists of (1) state preparation, (2) a sequence of gates, performed one after another, and (3) a measurement. Each gate sequence consists of three parts: (1) a short 'fiducial' gate sequence, followed by (2) a 'germ' sequence repeated some number of times, followed by (3) another short 'fiducial' sequence. Given a set of fiducial sequences, a set of germ sequences, and a list of maximum lengths (which dictate the number of times each germ is repeated), the set of all combinations of (*preparation fiducial*, *germ repeated to max-length*, *measurement fiducial*) gives the complete list of gate sequences required to run GST. Experiments for each gate sequence are repeated multiple times, and the resulting counts of measurement outcomes serve as input to the GST estimation algorithms. These algorithms find the best-fit gate set to the experimental data. Because the gate set is defined to contain only single-qubit operations, i.e. operations acting on a two-dimensional Hilbert state space, a gate set cannot capture effects due to additional Hilbert space dimensions. In particular, memory effects due to the environment, which are an example of what we refer to as 'non-Markovian noise', cannot be fit by *any* as-defined gate set. All physical systems will suffer from some degree of non-Markovian noise, and GST can detect this by assessing how well the best-fit gate set is able to reproduce the experimental data. The Pearson chi-squared test and the likelihood-ratio test are used to quantify the 'goodness-of-fit'.

The fiducial gate sequences and germ gate sequences, which are used to construct the final list of experiments as explained above, depend upon the ideal desired gates. In our case these gates, given in table 1, result in the six fiducial sequences

$$\{(\text{empty}),\ G_x,\ G_y,\ G_xG_x,\ G_xG_xG_x,\ G_yG_yG_y\}$$

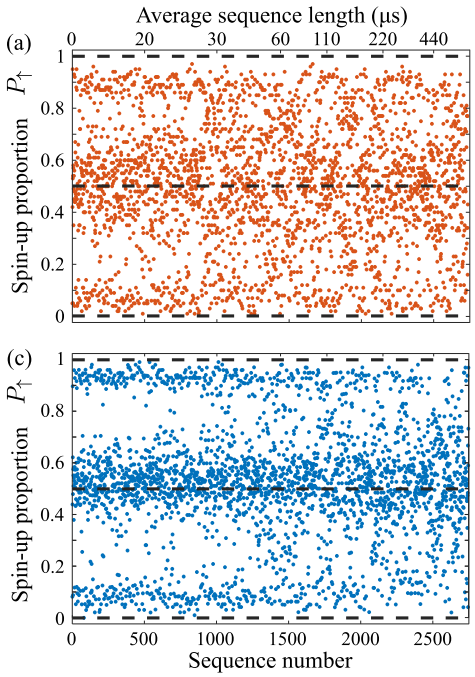
and eleven germ sequences

$$\{G_x,\ G_y,\ G_i,\ G_xG_y,\ G_xG_yG_i,\ G_xG_iG_y,\ G_xG_iG_i,\\ G_yG_iG_i,\ G_xG_xG_iG_y,\ G_xG_yG_yG_i,\ G_xG_xG_yG_xG_yG_y\}.$$

Details of how fiducial and germ sequences are computed can be found in the supplementary material of reference [30]. We used maximum lengths that were increasing powers of two from 1 to 256, which are chosen to include the longest sequences practical on our particular hardware given signal-to-noise and qubit decoherence considerations. The GST analysis was performed using the open-source pyGSTi code [32].

4. Optimizing the qubit operation with GST

Each cycle of initialization, gate sequence and measurement was repeated 100 times for each of the 2737 sequences constructed for GST. The number of $|\uparrow\rangle$ measurement outcomes was recorded for each sequence and the results were fed back to pyGSTi for analysis. Figure 2(a) shows a plot of the spin-up fraction $P_\uparrow$ for all the pulse sequences applied. For an ideal qubit, a sequence can have one of three possible $P_\uparrow$ outcomes: 0, 0.5, 1 (since the gates in our gate set consist of $\pi/2$ rotations). The high-precision of the GST protocol is obtained by designing sequences that amplify gate errors. This error amplification is evident from the scatter around the three $P_\uparrow$ values in the experimental dataset. Figure 2(b) shows a table with the estimated gates extracted from GST, highlighting on separate columns the rotation angle and axis implicit in these gate operators. Both G_x and G_y show rotation angles of 0.478π, which corresponds to a 4.4% under-rotation from the optimal 0.5π. Prior to the development of GST, we had optimized the qubit using the RB protocol [13]. RB returns a value for gate fidelity but does not provide any characterization of the gates. Therefore, qubit optimization is achieved by performing sweeps of intuitively chosen qubit operation parameters and searching for the parameter combination which yields the highest gate fidelity. In the RB study, we analysed gate fidelities for different pulse



(b) Without $\tau_{\pi/2}$ calibration

Gate	GST estimate				Rot. axis	Rot. angle
G_i	1	0	0	0	0	0.0027π
	0	0.992	−0.006	0.001	−0.061	
	0	0.012	0.993	−0.001	−0.028	
	0	0.001	0	1	−0.998	
G_x	1	0	0	0	0	0.4777π
	0	1	0.004	0.003	1	
	0	0.004	0.071	−0.995	0.003	
	0	−0.002	0.995	0.069	0.001	
G_y	1	0	0	0	0	0.4775π
	0	0.072	0.005	0.996	0.004	
	0	0.004	1	−0.002	1	
	0	−0.996	0.003	0.07	−0.001	

(d) With $\tau_{\pi/2}$ calibration

G_i	1	0	0	0	0	0.0059π
	0	0.979	−0.014	0.002	−0.081	
	0	0.025	0.986	−0.002	−0.029	
	0	0.004	−0.001	0.999	−0.996	
G_x	1	0	0	0	0	0.4999π
	0	1	0.002	−0.001	−1	
	0	0.003	0.003	−0.997	−0.002	
	0	−0.001	0.998	−0.002	−0.001	
G_y	1	0	0	0	0	0.5012π
	0	−0.002	0.003	0.996	0.001	
	0	0.004	1	0	1	
	0	−0.996	−0.001	−0.005	−0.002	

Figure 2. GST results. (a) Raw data points obtained after implementing each of the designed gate sequences and repeating them 100 times to extract the spin-up proportion $P_\uparrow$ for each sequence. We number the sequences from 0 to 2736 as shown in the bottom axis labels, and they increase in length as shown in the top axis labels. Dashed lines show target outcomes for an ideal qubit. (b) Post-processed GST results including the gate operators extracted from the data, and the rotation axis and angle implied by these operators. (c), (d) GST data and results after optimizing the pulse length calibration protocol to improve the $\tau_{\pi/2}$ accuracy.

shapes, ESR signal amplitudes and rise times of the pulses. We found a maximum Clifford gate fidelity $\mathcal{F}_G = 99.90(2)\%$ for square pulses, with a rise time of 100 ns and $B_1 = 12\ \mu$T (corresponding to $\tau_\pi = 3\ \mu$s). However, in that study we did not correctly account for the fact that the fixed rise times imply that the area under the time-dependent pulse amplitude—which determines the rotation—is not linear with pulse length. This effect is insignificant for long pulse lengths, but becomes more noticeable as τ_p becomes comparable to the rise time. This calibration protocol was designed to only calibrate τ_π and, for the rise time and pulse lengths used in our experiment, $\tau_\pi/2$ is 4.4% shorter in rotation than $\tau_{\pi/2}$, as identified by GST.

We corrected the issue by including a separate $\tau_{\pi/2}$ calibration step in the protocol. The data plot in figure 2(c)—taken after implementing the optimized calibration protocol—shows significantly less scatter in the data, a first indication that the gates are closer to the target gates. This is confirmed by the GST results in figure 2(d), now indicating G_x and G_y rotations within 0.7% of the target. One of the strengths of GST is that it supplies several figures of merit which provide information about the gates on different levels. Relevant to our gate optimization, the diamond norm ($\|\cdot\|_\diamond$) [31] provides a measure of distinguishability between two quantum processes. It is much more sensitive to coherent errors when compared to common measures of gate fidelity. GST extracts $\frac{1}{2}\|\cdot\|_\diamond$ between the experiment and target for each gate in the gateset. The values of $\frac{1}{2}\|\cdot\|_\diamond$ can range from 0 when the processes are completely indistinguishable, to 1 when the processes are maximally distinguishable. The results show a decrease in the $G_{x,y}$ average $\frac{1}{2}\|\cdot\|_\diamond$ from 0.036 before optimization, to 0.0034 after optimization. This order of magnitude improvement in $\frac{1}{2}\|\cdot\|_\diamond$ indicates that coherent errors in the gates were reduced by improving the pulse length accuracy.

Further details on the diamond norm, along with all the other figures of merit extracted from GST can be found in the full reports generated by pyGSTi, supplied in the supplementary material. Additionally, we have supplied the data files constructed from the experiments, along with the Python notebook used to generate the report. Instructions on how to use these files to generate the reports can be found in the pyGSTi project website [32].

To confirm the improvement in the gate calibration, we perform RB using the optimized calibration protocol. The RB protocol was implemented using the same Clifford gate set as in reference [13]. The protocol tests sequences with increasing number of Clifford gates N. To construct the sequences, a set of N Clifford gates is selected at random; a final state ($|\uparrow\rangle$ or $|\downarrow\rangle$) is also chosen at random and a final gate is added to the random gate sequence such that the spin is flipped to this final state. This sequence is repeated 200 times to compute $P_\uparrow$.

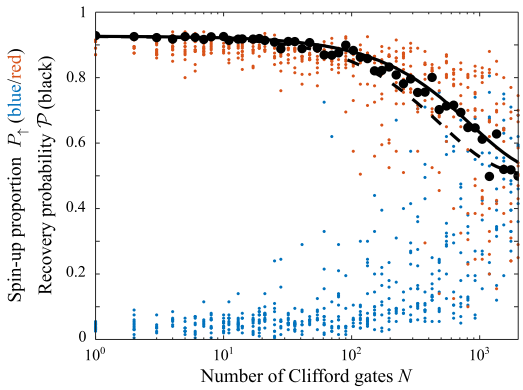


Figure 3. Randomized benchmarking with optimized pulse length calibration. Each of the small dots correspond to the $P_\uparrow$ extracted from 200 repetitions of a sequence; here, red(blue) dots correspond to sequences where the final state was chosen to be $|\uparrow\rangle$ ($|\downarrow\rangle$). Large black dots correspond to the overall correct recovery probability $\mathcal{P}$ as described in the main text. The solid line is a fit to the data using (3), yielding $C_0 = 0.4265(13)$ and $p = 0.99882(16)$, corresponding to $\mathcal{F}_G = 99.942(8)$. The fit is weighted with the inverse of the unbiased sample variance at each N. The dashed line uses $p = 0.998$, corresponding to the previously measured $\mathcal{F}_G = 99.9\%$ [13], scaled with the same C_0 for comparison.

For each N, 20 different random sequences are measured. From the data sets corresponding to each N, we can extract the overall probability of recovering the correct state $\mathcal{P} = 0.5(\bar{P}_\uparrow^{(\uparrow)} + (1 - \bar{P}_\uparrow^{(\downarrow)}))$, where $\bar{P}_\uparrow^{(\uparrow)}$ is the mean value of $P_\uparrow$ from sequences where the final state was chosen to be $|\uparrow\rangle$ ($|\downarrow\rangle$ for $\bar{P}_\uparrow^{(\downarrow)}$). $\mathcal{P}(N)$ can then be fitted [33] to

$$\mathcal{P} = C_0 p^N + 0.5, \tag{3}$$

where C_0 is a constant determined by SPAM errors and p determines the gate fidelity $\mathcal{F}_G = (1 + p)/2$. From the results shown in figure 3, we extract $\mathcal{F}_G = 99.942(8)\%$, setting a new gate fidelity benchmark for the ^{31}P electron-spin qubit.

5. Non-Markovian noise

The accuracy of GST relies greatly on the stability of the qubit over the timescale of the experiment. Essentially, GST assumes that the qubit is 'the same qubit' when each sequence is being applied. Any slow drift in the environment will reduce GST's ability to fit the data using a Markovian model, and thereby reduce the reliability of its estimates. While GST is able to detect and crudely quantify such non-Markovian noise (e.g. slow drift results in decreasing goodness-of-fit with increasing sequence length), it is as yet unable to assign meaningful error bars to account for this noise. An analysis of the goodness-of-fit from GST reveals that the experimental dataset violates the fitted Markovian model by up to 250 times the standard deviation returned by the fit (see supplementary GST reports for more details). This is a strong indicator that there are high levels of non-Markovian noise present in the system. As a consequence, we currently observe variabilities in the gate parameters between GST runs, which are larger than the error estimates. This is limiting our ability to optimize the qubit further. Apparent differences between the results in parameters that do not depend on the rotation angle (e.g. decoherence rates), are due to these variabilities induced by non-Markovian noise.

We attribute the majority of the non-Markovian noise to jumps on the order of 10 kHz in the qubit resonance frequency, which happen on timescales on the order of 10 min (figure 4). These jumps likely arise from single nuclear spin flips from either ^{29}Si or other ionized ^{31}P in the vicinity of the qubit. Recalling (1) and (2), a shift in the ESR frequency will cause deviations from the expected Rabi oscillation frequency ν_1 and will cause the instantaneous axis of Rabi rotation to lift away from the equator of the Bloch sphere, i.e. the polar angle θ of the rotation axis is $\neq 90^\circ$. However, the azimuthal angle φ is not affected by the detuning. Therefore, the resonance frequency jumps mainly affect the rotation angle. With the B_1 used in our experiments, a 10 kHz detuning will cause a $\sim$0.2% error in ν_1 and a $\sim$4% error in θ. This is well within the accuracy capabilities of GST.

While GST and RB are expected to agree to within their respective error bars on gates with Markovian errors, they respond very differently to the slow drift that causes non-Markovian behaviour in the system. Drift in the qubit resonance frequency produces coherent (unitary) errors in the gates, but ones that vary in time. RB is less sensitive to coherent errors than current criteria for fault tolerance [34, 35]. Large non-Markovian drifts in detuning frequency can cause the RB decay curve to become noticeably non-exponential [12, 33]; however, in the results presented here this effect is too subtle to observe. GST, on the other hand, is very sensitive to non-

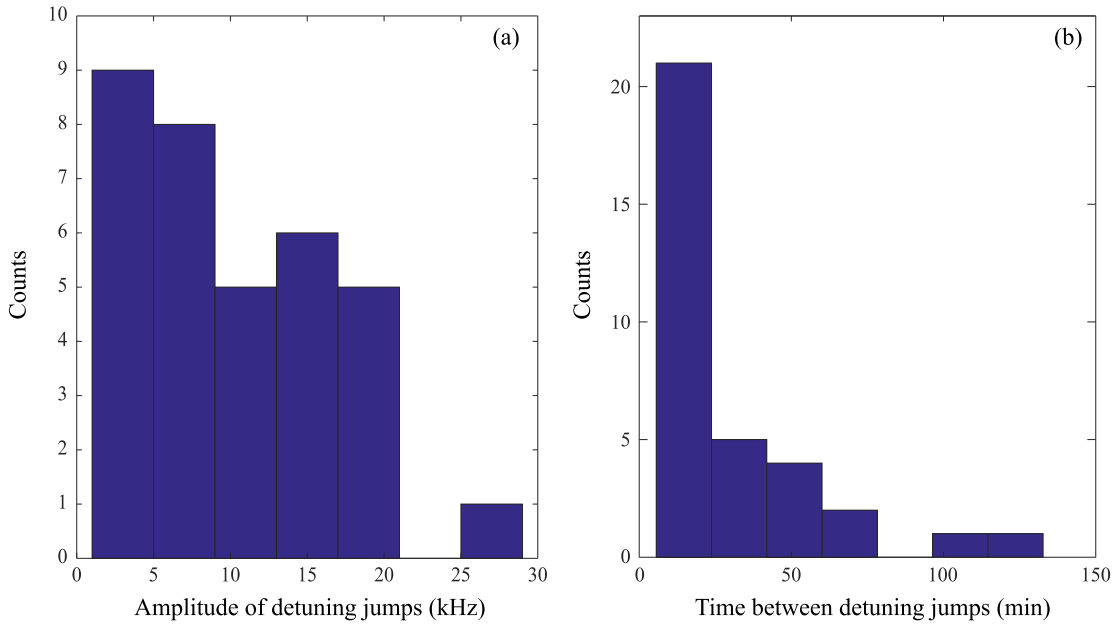


Figure 4. Statistical characterization of random jumps in the qubit resonance frequency. (a) Histogram of the amplitude of the observed frequency shifts; (b) histogram of the time interval between frequency jumps. This data is obtained from repeated resonance frequency calibrations over a period of ∼40 h. The calibration procedure is described in the main text. To obtain this dataset, a total of 791 calibrations were performed with 3 min intervals, and a total of 34 frequency jumps above the the threshold were recorded. The sampling rate and total length of the Ramsey measurement is set such that the frequency resolution of the calibration is 1 kHz and the maximum detuning detection is 100 kHz. The mean values of each dataset are: (a) 10 kHz and (b) 28 min. The Pearson correlation coefficient using the two datasets is −0.2(3), which indicates little correlation between the magnitude of frequency jumps and the interval between them.

Markovian noise—but has no mechanism for it. GST misclassifies this kind of non-Markovian noise (caused by slow drift) as stochastic noise. Therefore, while RB underestimates the total noise, GST overestimates the stochastic noise. For this reason, simulated RB using the GST estimated gate set from the optimized system (figure 2d), predicts an average Clifford gate infidelity of $1 - \mathcal{F}_G = 0.25(2)\%$. In contrast, the average Clifford gate infidelity observed in real RB experiments (figure 3) is $1 - \mathcal{F}_G = 0.058(8)\%$. Therefore, while GST fails to correctly predict RB, this is a direct consequence of the fact that GST is able to identify non-Markovian noise (although not to model it), and correctly warns that its presence compromises the accuracy of the results. Comparison of GST and RB results indicate that non-Markovian effects currently dominate Markovian stochastic noise in the system.

It has not been shown that quantum error correction can tolerate the same level of infidelity from Non-Markovian as from Markovian noise. Therefore, it is important to consider strategies for mitigating the effects of non-Markovian noise in order to use this qubit in a fault-tolerant setting. In all the experiments presented here, we monitor and calibrate the resonance frequency of the qubit by performing a Ramsey fringe experiment [36] to determine the detuning frequency. The calibration takes on average ∼1 min to complete and is performed every ∼20 min. Increasing the frequency with which the calibration is performed will unmanageably extend the total experiment duration. A different approach to minimize (but not eliminate) the impact of drift and/or non-Markovian noise is to interleave the 'shots' of each GST sequence [37]. By performing interleaving, the measurements are taken in 100 sequence sweeps with 1 single-shot per sequence (or, more feasibly, repeating $100/N$ sweeps and taking N shots for each sequence during each sweep). Interleaving would ensure that the data for each sequence are sampled from the full span of time for which the experiment runs. It does not eliminate non-Markovian behaviour (drift still has a significant impact on long sequences even with interleaving), but would result in a more reliable and meaningful estimate. However, this method is impractical with our current experimental setup, because the most time-consuming step in the experiment is loading a new sequence onto the arbitrary waveform generator, while repeating a measurement once a sequence is loaded is relatively much faster. Therefore, attempting to perform an adequate amount of interleaving would unmanageably increase the total duration of the experiment. Furthermore, this would not address the root of the problem: qubit drift over time that would become problematic when running real quantum circuits. Moving forward, an approach to correct this non-Markovian noise is to use dynamically corrected gates [38–40], where the gate sequence is interleaved with a dynamical decoupling sequence in order to suppress gate errors and decoherence effects from low-frequency noise sources. This approach has been successfully applied and verified to correct non-

Markovian noise using GST for a trapped-ion qubit [30], which leads us to believe that it would also be successful here. Another possible solution is to implement a Hamiltonian estimation protocol [41], which could potentially allow us to increase the speed and frequency of the detuning frequency calibration.

6. Conclusion

GST is a protocol designed to characterize and optimize qubit systems. By applying GST to the ^{31}P electron spin qubit in ^{28}Si, we were able to identify a 4.4% rotation error in some of the gates. We improved the calibration method to fix this error, which in turn improved the average gate fidelity of the qubit from 99.90(2)% to 99.942(8)%, measured via RB. Non-Markovian noise, originating from small jumps in the resonance frequency of the qubit, are detected by GST, and limit the performance of the qubit. The use of dynamically corrected gates should suppress the effects of non-Markovian noise, and should be first priority for future measurements. This work demonstrates that GST is capable of characterizing qubit gates to levels not previously accessible through any other experimental protocol. We envision that GST will become an increasingly important tool for validation and verification of quantum information hardware and protocols, as the community moves towards increasingly complex and high-fidelity gate operations.

Acknowledgments

We thank F Hudson for assistance in device fabrication and design, K Itoh for supplying the ^{28}Si epilayer, and J McCallum and D Jamieson for the ^{31}P donor implantation. This research was funded by the Australian Research Council Centre of Excellence for Quantum Computation and Communication Technology (project no. CE110001027) and the US Army Research Office (W911NF-13-1-0024). The authors acknowledge support from the Australian National Fabrication Facility. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the US Department of Energys National Nuclear Security Administration under contract DE-AC04-94AL85000. JKG gratefully acknowledges support from the Sandia National Laboratories Truman Fellowship Program, which is funded by the Laboratory Directed Research and Development (LDRD) program.

References

[1] Shor P W 1995 *Phys. Rev.* A **52** R2493–6
[2] Steane A 1996 *Proc. R. Soc.* A **452** 2551–77
[3] Raussendorf R 2012 *Phil. Trans. R. Soc.* A **370** 4541–65
[4] Knill E 2005 *Nature* **434** 39–44
[5] Fowler A G, Mariantoni M, Martinis J M and Cleland A N 2012 *Phys. Rev.* A **86** 032324
[6] Ryan C A, Laforest M and Laflamme R 2009 *New J. Phys.* **11** 013034
[7] Olmschenk S, Chicireanu R, Nelson K D and Porto J V 2010 *New J. Phys.* **12** 113007
[8] Brown K R, Wilson A C, Colombe Y, Ospelkaus C, Meier A M, Knill E, Leibfried D and Wineland D J 2011 *Phys. Rev.* A **84** 030303(R)
[9] Harty T P, Allcock D T C, Ballance C J, Guidoni L, Janacek H A, Linke N M, Stacey D N and Lucas D M 2014 *Phys. Rev. Lett.* **113** 220501
[10] Barends R *et al* 2014 *Nature* **508** 500–3
[11] Dolde F *et al* 2014 *Nat. Commun.* **5** 3371
[12] Veldhorst M *et al* 2014 *Nat. Nanotechnol.* **9** 981–5
[13] Muhonen J T *et al* 2015 *J. Phys.:Condens. Mater* **27** 154205
[14] Chuang I L and Nielsen M A 1997 *J. Mod. Opt.* **44** 2455–67
[15] Chow J M, Gambetta J M, Tornberg L, Koch J, Bishop L S, Houck A A, Johnson B R, Frunzio L, Girvin S M and Schoelkopf R J 2009 *Phys. Rev. Lett.* **102** 090502
[16] Kim D *et al* 2015 *Nat. Nanotechnol.* **10** 243–7
[17] Emerson J, Alicki R and Życzkowski K 2005 *J. Opt. B: Quantum Semiclass. Opt.* **7** S347
[18] Knill E, Leibfried D, Reichle R, Britton J, Blakestad R, Jost J, Langer C, Ozeri R, Seidelin S and Wineland D 2008 *Phys. Rev.* A **77** 012307
[19] Blume-Kohout R, Gamble J K, Nielsen E, Mizrahi J, Sterk J D and Maunz P 2013 Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit arXiv:1310.4492 [quant-ph]
[20] Merkel S T, Gambetta J M, Smolin J A, Poletto S, Córcoles A D, Johnson B R, Ryan C A and Steffen M 2013 *Phys. Rev.* A **87** 062119
[21] van Donkelaar J *et al* 2015 *J. Phys.: Condens. Mater* **27** 154204
[22] Itoh K M and Watanabe H 2014 *MRS Commun.* **4** 143–57
[23] Morello A, Escott C, Huebl H, Willems van Beveren L, Hollenberg L, Jamieson D, Dzurak A and Clark R 2009 *Phys. Rev.* B **80** 081307
[24] Morello A *et al* 2010 *Nature* **467** 687–91
[25] Pla J J, Tan K Y, Dehollain J P, Lim W H, Morton J J L, Jamieson D N, Dzurak A S and Morello A 2012 *Nature* **489** 541–5
[26] Pla J J, Tan K Y, Dehollain J P, Lim W H, Morton J J L, Zwanenburg F A, Jamieson D N, Dzurak A S and Morello A 2013 *Nature* **496** 334–8
[27] Muhonen J T *et al* 2014 *Nat. Nanotechnol.* **9** 986–91
[28] Dehollain J P, Pla J J, Siew E, Tan K Y, Dzurak A S and Morello A 2013 *Nanotechnology* **24** 015202
[29] Laucht A *et al* 2016 A dressed spin qubit in silicon arXiv:1603.04800

[30] Blume-Kohout R, Gamble J K, Nielsen E, Rudinger K, Mizrahi J, Fortier K and Maunz P 2016 Certifying qubit operations below the fault tolerance threshold arXiv:1605.07674 [quant-ph]
[31] Kitaev A Y 1997 *Russ. Math. Surv.* **52** 1191
[32] Nielsen E, Rudinger K M, Gamble J K and Blume-Kohout R pyGSTi (python GateSet Tomography implementation) http://pygsti.info
[33] Fogarty M A, Veldhorst M, Harper R, Yang C H, Bartlett S D, Flammia S T and Dzurak A S 2015 *Phys. Rev.* A **92** 022326
[34] Wallman J J and Flammia S T 2014 *New J. Phys.* **16** 103032
[35] Sanders Y R, Wallman J J and Sanders B C 2015 *New J. Phys.* **18** 012002
[36] Ramsey N F 1950 *Phys. Rev.* **78** 695–9
[37] van Enk S J and Blume-Kohout R 2013 *New J. Phys.* **15** 025024
[38] Liu G Q, Po H C, Du J, Liu R B and Pan X Y 2013 *Nat. Commun.* **4** 2254
[39] Rong X, Geng J, Wang Z, Zhang Q, Ju C, Shi F, Duan C K and Du J 2014 *Phys. Rev. Lett.* **112** 050503
[40] Zhang J, Souza A M, Brandao F D and Suter D 2014 *Phys. Rev. Lett.* **112** 050502
[41] Shulman M D, Harvey S P, Nichol J M, Bartlett S D, Doherty A C, Umansky V and Yacoby A 2014 *Nat. Commun.* **5** 5156