
Supplementary information

Precision tomography of a three-qubit donor quantum processor in silicon

In the format provided by the
authors and unedited

SUPPLEMENTARY INFORMATION:

Precision tomography of a three-qubit electron-nuclear quantum processor in silicon

Mateusz T. Mądzik,^{1,2,*} Serwan Asaad,^{1,2,†} Akram Youssry,^{3,4} Benjamin Joecker,^{1,2} Kenneth M. Rudinger,⁵ Erik Nielsen,⁵ Kevin C. Young,⁵ Timothy J. Proctor,⁵ Andrew D. Baczewski,⁶ Vivien Schmitt,^{1,2,‡} Fay E. Hudson,¹ Kohei M. Itoh,⁷ Alexander M. Jakob,^{8,2} Brett C. Johnson,^{8,2} David N. Jamieson,^{8,2} Andrew S. Dzurak,¹ Christopher Ferrie,³ Robin Blume-Kohout,⁵ and Andrea Morello^{1,2,§}

¹*School of Electrical Engineering and Telecommunications, UNSW Sydney, Sydney, NSW 2052, Australia*

²*Centre for Quantum Computation and Communication Technology*

³*Centre for Quantum Software and Information,
University of Technology Sydney, Ultimo, NSW 2007, Australia*

⁴*Department of Electronics and Communication Engineering,
Faculty of Engineering, Ain Shams University, Cairo, Egypt*

⁵*Quantum Performance Laboratory, Sandia National Laboratories,
Albuquerque, NM 87185 and Livermore, CA 94550, USA*

⁶*Center for Computing Research, Sandia National Laboratories, Albuquerque, NM 87185, USA*

⁷*School of Fundamental Science and Technology, Keio University, Kohoku-ku, Yokohama, Japan*

⁸*School of Physics, University of Melbourne, Melbourne, VIC 3010, Australia*

CONTENTS

S1. Absence of direct interaction between the nuclear spin qubits	2
S2. CNOT truth tables	3
S3. Bell state tomography ancillary data	4
S4. Single-qubit gate set tomography	5
S5. Frequency recalibration during GST	6
S6. Expansion of GHZ echo circuit	7
S7. High-fidelity measurement of three-qubit states by mapping and readout on the nuclear spins	8
S8. Comparison of GST model fits	10
S9. Reduced metrics of gate error	11
A. Relationship between generator infidelity and entanglement infidelity	13
B. Relationship between the total error and the diamond error	15
S10. Detailed analysis of the $X_{\pi/2}$ gate on Q2	17
S11. Rationalizing two-qubit errors on one-qubit gates	19
A. Validation of entangling errors	19
B. The direct entangling interaction between nuclear spins is too weak	19
C. Accounting for electron-mediated internuclear entanglement	20
D. Analysis of the impact of the electron-mediated interaction on one-qubit gates	21
E. Leakage of the ESR carrier signal leads to coherent entangling errors	23
References	26

* These two authors contributed equally; Currently at QuTech, Delft University of Technology, 2628 CJ Delft, The Netherlands.

† These two authors contributed equally; Currently at Center for Quantum Devices, Niels Bohr Institute, University of Copenhagen, and Microsoft Quantum Lab Copenhagen, Copenhagen, Denmark.

‡ Currently at Univ. Grenoble Alpes, Grenoble INP, CEA, IRIG-PHELIQS, F-38000 Grenoble, France.

§ To whom correspondence should be addressed; a.morello@unsw.edu.au

S1. ABSENCE OF DIRECT INTERACTION BETWEEN THE NUCLEAR SPIN QUBITS

In this section we provide experimental evidence for the absence of a direct interaction between the nuclear spin qubits Q1 and Q2. First, we remove the outermost electron from the 2P cluster and measure the resonance frequencies of Q1 and Q2. We apply RF pulses at very low power (-21 dBm at the source) to minimise power broadening of the NMR resonances. Supplementary Figure S1a shows that the two qubits have identical resonance frequencies, which are consistent with our estimate of the external magnetic field induced by the permanent magnet board ($\nu_{^{31}\text{P}^+} = \gamma_n B_0$, $\gamma_n = 17.23 \text{ MHz/T}$, $B_0 \approx 1.328 \text{ T}$). These results are consistent with the remaining two electrons being in a perfect $S = 0$ singlet state, whereby $A_1 = A_2 = 0$. They are also consistent with having removed all electrons from the 2P cluster, but we consider this interpretation less likely, based on the anomalously low value of the electron spin relaxation time T_{1e} (see Extended Data Fig. 2). We can also deduce that the direct dipole-dipole interaction between the two nuclear spins is negligible in comparison to the broadening ($\sim 1 \text{ kHz}$) of the NMR resonance peaks.

Next, we investigate the NMR spectrum of both Q1 and Q2 with all electrons present on 2P cluster. We measure the frequency response of a target qubit (Q1 - Supplementary Figure S1b; Q2 - Supplementary Figure S1c) with the spectator qubit (Q2 - Supplementary Figure S1b; Q1 - Supplementary Figure S1c) prepared in either $|\downarrow\rangle$ or $|\uparrow\rangle$ state. We observe no detectable resonance frequency shift due to a coupling between the nuclei, which in this case might be mediated by the shared electron.

These experiments corroborate the analysis in Section S11A, wherein it is shown that no plausible value of inter-nuclear interaction can explain the presence of the weight-2 entangling errors unveiled by GST.

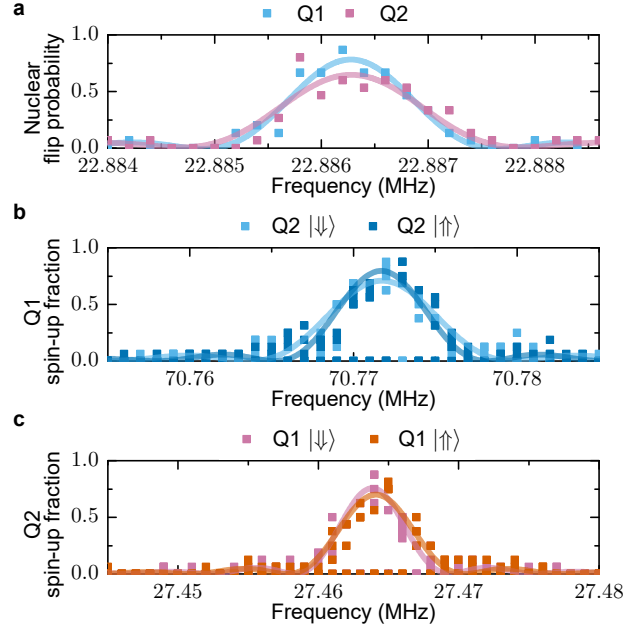


Figure S1. **a**, NMR spectrum of Q1 and Q2 with the third electron removed. The data was acquired at -21 dBm NMR power to minimise spectral broadening. The two remaining electrons are left in the magnetically inactive, $S = 0$ singlet state. The resonance frequencies of the two qubits are equal and consistent with the prediction for the ionized ^{31}P in the estimated magnetic field produced by the permanent magnet board ($B_0 \approx 1.328 \text{ T}$). **b**, NMR spectrum of Q1 (**c**, Q2) with all three electrons present on the 2P cluster. Q2 (**c**, Q1) was prepared in either $|\downarrow\rangle$ or $|\uparrow\rangle$. In both cases the Q1 (**c**, Q2) resonant frequency remains the same, indicating the absence of significant coupling between the two nuclei.

S2. CNOT TRUTH TABLES

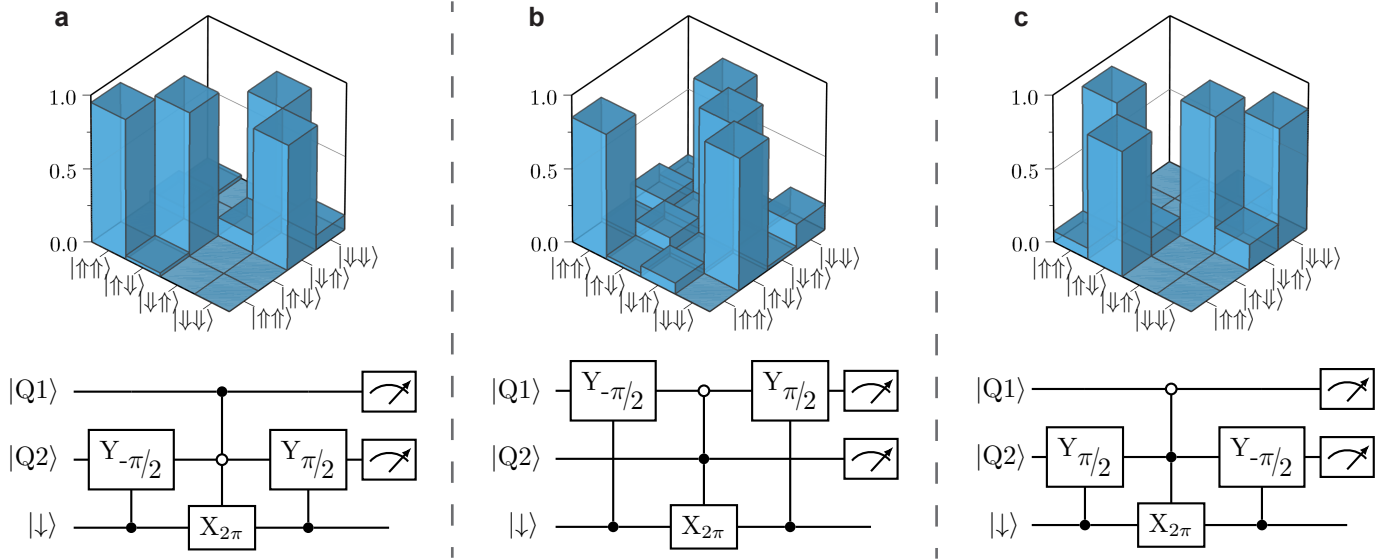


Figure S2. Experimental CNOT truth-tables. The qubits, encoded on the nuclear spins, have been prepared in all four eigenstates, with the electron spin in the $|\downarrow\rangle$ state. We adopt a notation for the computational basis consistent with the standard quantum information conventions, where $|\downarrow\rangle \equiv |1\rangle$ and $|\uparrow\rangle \equiv |0\rangle$. **a**, CNOT quantum logic gate, where Q1 serves as a control and Q2 as a target. **b**, CNOT quantum logic gate, where Q1 serves as a target and Q2 as a control. **c**, Zero-CNOT (zCNOT) quantum logic gate, where Q1 serves as a control and Q2 as a target. Here, by selecting a different ESR transition for the electron 2π -pulse and exchanging the single-qubit gates, the target qubit is inverted when the control qubit is in the $|0\rangle$ state.

S3. BELL STATE TOMOGRAPHY ANCILLARY DATA

The errors in the state preparation fidelity have been calculated using Monte Carlo bootstrap resampling. In this method, we assume a binomial distribution on the measured results and compute the possible measurement outcomes with regards to the experimental sampling (number of single shot readouts for state probability). We sample from the binomial distributions and reconstruct the density matrices 500 times. For each density matrix we calculate state preparation fidelity and concurrence. This allows for estimation of the average values, as well as assessment of the uncertainties in our results.

$$\rho_{\Phi^+} = \begin{pmatrix} 0.4802 + 0.0000i & 0.0600 + 0.0063i & -0.0333 + 0.0119i & 0.4722 - 0.0284i \\ 0.0600 - 0.0063i & 0.0089 + 0.0000i & -0.0054 + 0.0017i & 0.0567 - 0.0097i \\ -0.0333 - 0.0119i & -0.0054 - 0.0017i & 0.0096 + 0.0000i & -0.0219 - 0.0196i \\ 0.4722 + 0.0284i & 0.0567 + 0.0097i & -0.0219 + 0.0196i & 0.5013 + 0.0000i \end{pmatrix} \quad (1)$$

$$\rho_{\Phi^-} = \begin{pmatrix} 0.4880 + 0.0000i & 0.0002 + 0.0248i & -0.0433 + 0.0316i & -0.4804 - 0.0000i \\ 0.0002 - 0.0248i & 0.0079 + 0.0000i & 0.0070 + 0.0060i & -0.0012 + 0.0355i \\ -0.0433 - 0.0316i & 0.0070 - 0.0060i & 0.0125 + 0.0000i & 0.0481 + 0.0408i \\ -0.4804 + 0.0000i & -0.0012 - 0.0355i & 0.0481 - 0.0408i & 0.4916 + 0.0000i \end{pmatrix} \quad (2)$$

$$\rho_{\Psi^+} = \begin{pmatrix} 0.0055 + 0.0000i & 0.0365 + 0.0166i & 0.0464 + 0.0054i & 0.0017 - 0.0000i \\ 0.0365 - 0.0166i & 0.4461 + 0.0000i & 0.4664 - 0.0902i & 0.0121 - 0.0575i \\ 0.0464 - 0.0054i & 0.4664 + 0.0902i & 0.5301 + 0.0000i & 0.0103 - 0.0498i \\ 0.0017 + 0.0000i & 0.0121 + 0.0575i & 0.0103 + 0.0498i & 0.0184 + 0.0000i \end{pmatrix} \quad (3)$$

$$\rho_{\Psi^-} = \begin{pmatrix} 0.0328 + 0.0000i & -0.0724 + 0.0147i & 0.0784 - 0.0149i & 0.0104 - 0.0022i \\ -0.0724 - 0.0147i & 0.4730 + 0.0000i & -0.4728 + 0.0167i & -0.0138 + 0.0131i \\ 0.0784 + 0.0149i & -0.4728 - 0.0167i & 0.4847 + 0.0000i & 0.0236 - 0.0153i \\ 0.0104 + 0.0022i & -0.0138 - 0.0131i & 0.0236 + 0.0153i & 0.0095 + 0.0000i \end{pmatrix} \quad (4)$$

S4. SINGLE-QUBIT GATE SET TOMOGRAPHY

During the initial calibration stage and in the preparation for the two-qubit gate set tomography, we performed single-qubit GST on both qubits. The 1-qubit GST experiments consist of 448 circuits of length up to 14. We have used these GST error estimates to iteratively correct our control pulses for imperfect calibration, in order to obtain the highest gate fidelities. Supplementary Figure S3a shows an example of data from a single-qubit GST on both Q1 and Q2.

The table in Supplementary Figure S3b compares the estimates of single qubit average gate fidelities obtained during these 1-qubit GST experiments, with those obtained as a subset of the 2-qubit GST experiments described in the main text. We choose to report average gate fidelities, rather than generator infidelities or entanglement infidelities, to facilitate easy comparison with other 1Q result available in the literature. The three datasets were acquired within 3 months from each other. Some variability in the gate fidelities across such a time span is entirely plausible, reflecting slow drifts in the sample caused e.g. by rearrangement of charges in the vicinity of the donors. Nevertheless, these experiments show that the 1-qubit gate fidelities extracted within 2-qubit GST are very close to those obtained by performing 1-qubit GST on each nucleus, and that the sample is remarkably stable over the course of many months.

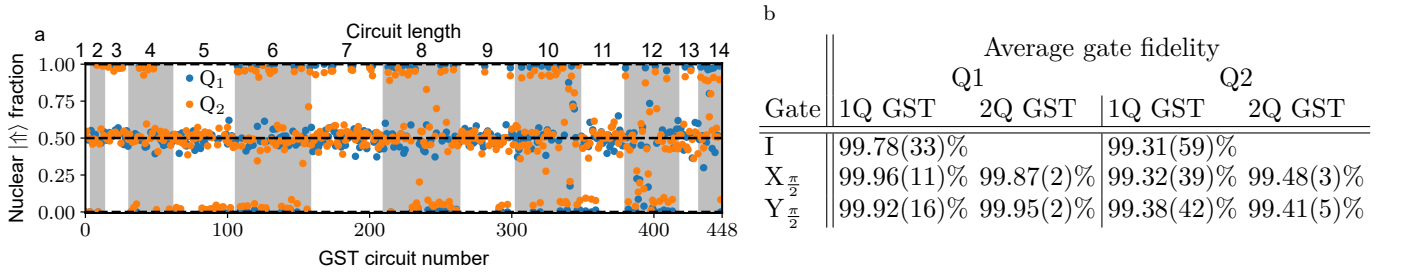


Figure S3. Single-qubit gate set tomography. **a**, Measured nuclear $|\uparrow\uparrow\rangle$ fraction of Q1 and Q2 for each of the 448 circuits on Q1 (blue) and Q2 (orange), sorted by circuit length. In the case of perfect gates and perfect measurements, the target $|\uparrow\uparrow\rangle$ fractions would be either 0, 0.5, or 1 (dashed lines), depending on the specific sequence. **b**, Average gate fidelities of the identity (I), $X_{\pi/2}$ and $Y_{\pi/2}$ single-qubit gates as obtained from the 1-qubit GST experiments shown in panel **a** (1Q), compared to the same quantities as obtained within the 2-qubit GST experiments described in the main text (2Q). For a fair comparison, we include in the 2Q fidelities only the component acting on the target qubit.

S5. FREQUENCY RECALIBRATION DURING GST

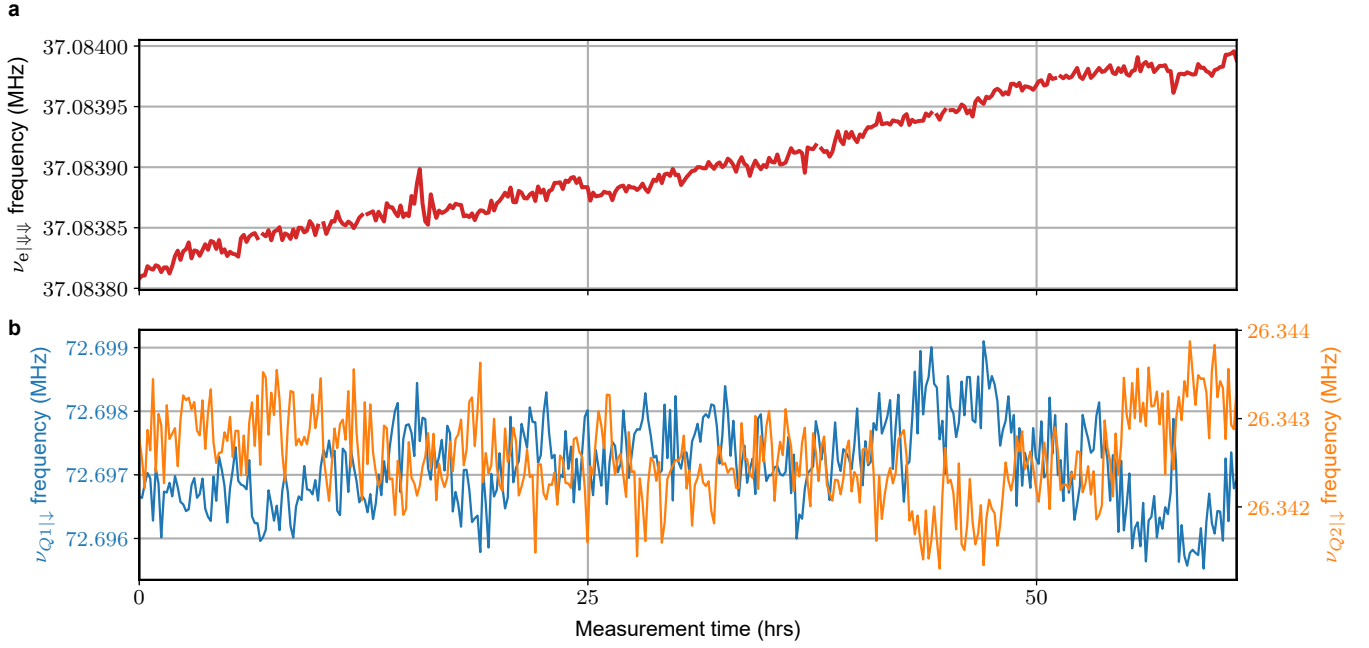


Figure S4. Electron $\nu_{e|\downarrow\downarrow}$ (a) and nuclear $\nu_{Q1|\downarrow}$ and $\nu_{Q2|\downarrow}$ (b) frequency drifts measured while performing two-qubit gate set tomography. The effects of the drifts were periodically cancelled by calibrating the frequencies every tenth GST circuit (see Methods for calibration routines). The electron spin resonance frequency $\nu_{e|\downarrow\downarrow}$ (a) experiences a fairly constant upwards drift, consistent with a steady increase of the magnetic field B_0 . This could be caused by the assembly of the permanent magnet board where the device resides in; in this particular board, some of the NdFeB magnets were purposely oriented in an opposing direction in order to reduce the magnetic field to < 1.4 T (limit set by the maximum frequency of our microwave source). This frustrated magnetic configuration could slowly relax to a lower-energy configuration, with a corresponding higher magnetic field. Conversely, the shifts of the NMR frequencies (b) do not appear to be dominated by a magnetic field drift. $\nu_{Q1|\downarrow}$ (blue) and $\nu_{Q2|\downarrow}$ (orange) show an evident anti-correlation. This is consistent with hyperfine-induced shifts, since the two hyperfine interactions A_1 and A_2 were observed to shift in opposite directions in response to applied gate voltages (Extended Data Fig. 2).

S6. EXPANSION OF GHZ ECHO CIRCUIT

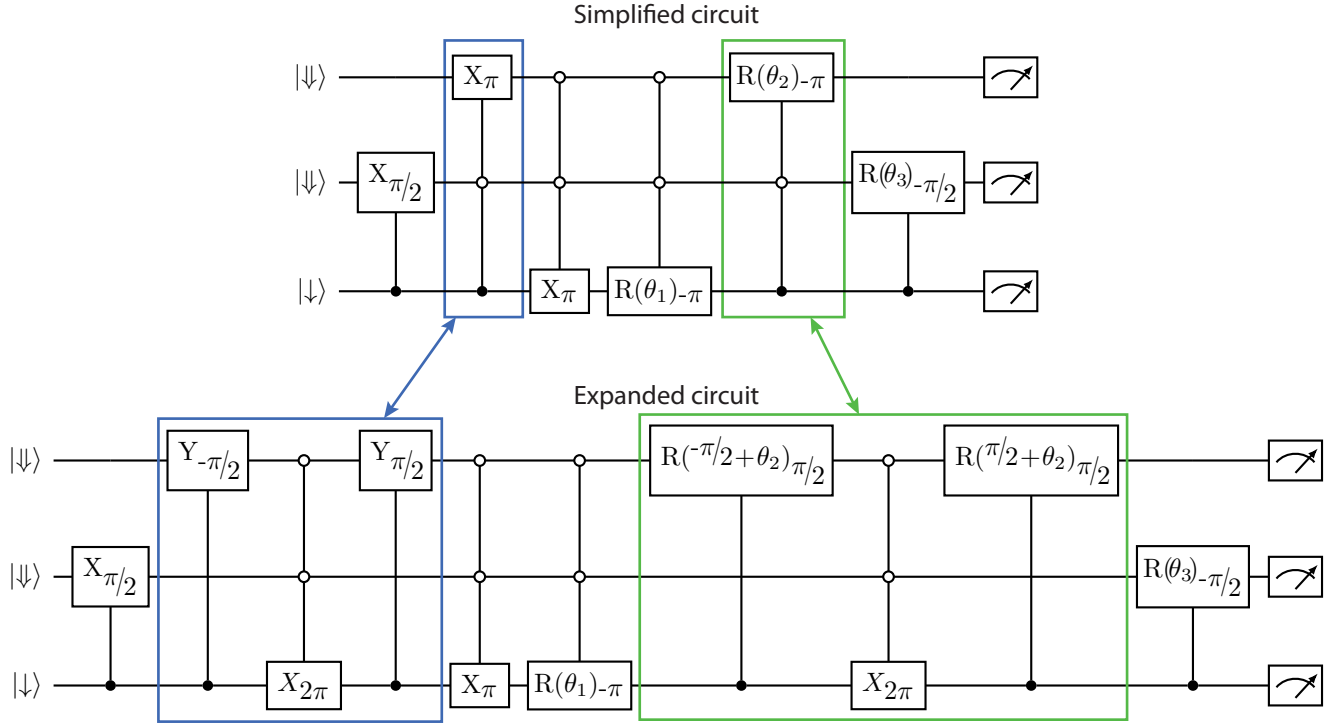


Figure S5. Expansion of the GHZ echo circuit. The circuit used to create a GHZ state and subsequently reverse the operations (top circuit and Fig. 4a in the main text) contains two nuclear operations on Q_1 that are conditional on the state of Q_2 (blue and green box). These operations are each composed of three pulses that are expanded in the bottom circuit.

S7. HIGH-FIDELITY MEASUREMENT OF THREE-QUBIT STATES BY MAPPING AND READOUT ON THE NUCLEAR SPINS

The measurements on the electron-nuclear three-qubit GHZ state (Fig. 4 in the main text) require extracting the populations of all eight basis states of the Hilbert space of the two nuclei and electron. One approach to do so is to first measure the electron spin, and then measure the nuclear spins by mapping their state onto the electron (see Methods). However, in this approach the readout fidelity is limited by the electron readout fidelity $\approx 80\%$, which would therefore limit the observable GHZ fidelity, unless the effect of SPAM errors were removed in post-processing.

In order to prove our ability to produce and measure a high-fidelity GHZ state without removing SPAM errors, we designed a method where specific qubit populations are swapped using conditional NMR or ESR π pulses, followed by high-fidelity ($> 99\%$) quantum nondemolition nuclear readout (see Methods). Each circuit is repeated three times, applying each of the following additional sets of swapping pulses prior to nuclear readout:

- A. No additional pulses,
- B. X_π at frequency $\nu_{Q2|\downarrow}$,
- C. X_π at frequency $\nu_{e|\downarrow\downarrow}$, then X_π at frequency $\nu_{Q1|\downarrow}$.

Each set of swapping pulses $\alpha \in [A, B, C]$ yields four measured nuclear state populations $[P(\downarrow\downarrow)_\alpha, P(\downarrow\uparrow)_\alpha, P(\uparrow\downarrow)_\alpha, P(\uparrow\uparrow)_\alpha]$, resulting in a total of twelve measured nuclear state populations. Each of the eight state populations $P(Q_1 Q_2 e)$ can be reconstructed from a minimum of five of the twelve measured state populations in ten different combinations, one example being

$$\begin{aligned}
 P(\uparrow\uparrow\uparrow) &= (P(\uparrow\uparrow)_A - P(\uparrow\downarrow)_B + P(\uparrow\downarrow)_C - P(\downarrow\downarrow)_B + P(\uparrow\uparrow)_C)/2, \\
 P(\uparrow\uparrow\downarrow) &= (P(\uparrow\uparrow)_A - P(\uparrow\uparrow)_B + P(\downarrow\downarrow)_C - P(\downarrow\uparrow)_B + P(\downarrow\uparrow)_C)/2, \\
 P(\uparrow\downarrow\uparrow) &= (P(\uparrow\downarrow)_A - P(\uparrow\uparrow)_B + P(\uparrow\uparrow)_C - P(\downarrow\downarrow)_B + P(\uparrow\downarrow)_C)/2, \\
 P(\uparrow\downarrow\downarrow) &= (P(\uparrow\downarrow)_A - P(\uparrow\downarrow)_B + P(\downarrow\uparrow)_C - P(\downarrow\uparrow)_B + P(\downarrow\downarrow)_C)/2, \\
 P(\downarrow\uparrow\uparrow) &= (P(\downarrow\uparrow)_A - P(\downarrow\downarrow)_B + P(\uparrow\downarrow)_C - P(\uparrow\downarrow)_B + P(\downarrow\uparrow)_C)/2, \\
 P(\downarrow\uparrow\downarrow) &= (P(\downarrow\uparrow)_A - P(\downarrow\uparrow)_B + P(\downarrow\downarrow)_C - P(\uparrow\uparrow)_B + P(\uparrow\uparrow)_C)/2, \\
 P(\downarrow\downarrow\uparrow) &= (P(\downarrow\downarrow)_A - P(\downarrow\uparrow)_B + P(\downarrow\uparrow)_C - P(\uparrow\downarrow)_B + P(\uparrow\downarrow)_C)/2, \\
 P(\downarrow\downarrow\downarrow) &= (P(\downarrow\downarrow)_A - P(\downarrow\downarrow)_B + P(\uparrow\uparrow)_C - P(\uparrow\uparrow)_B + P(\downarrow\downarrow)_C)/2.
 \end{aligned}$$

Each of the ten combinations produce slightly different state populations dependent on the individual measurement outcomes. The final state populations are then taken as the average of the ten possible state populations.

The validity of this state-readout technique is characterized by preparing each of the eight possible states, and then measuring the populations in all eight states. Figure S6 shows the measured populations, uncorrected for SPAM errors. The results show that for each of the eight initialized states, the measured population of the initialized state is near-unity. With this method we obtained the GHZ fidelity $\mathcal{F}_{\text{GHZ}} = 92.5(1.0)\%$ quoted in the main text, including the residual SPAM errors associated with nuclear readout.

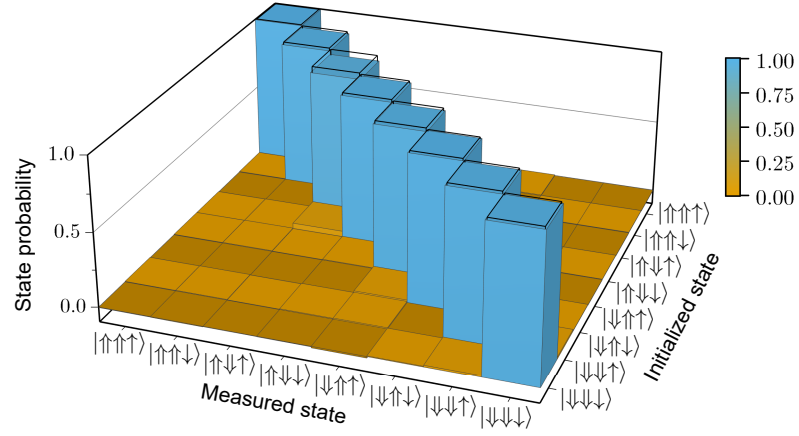


Figure S6. Extraction of eight electron-nuclear states by nuclear readout. Each row corresponds to the spin states of the two nuclei and electron being initialized into one of the eight possible eigenstates. Three sets of swapping pulses are then applied in distinct measurement instances, followed by nuclear readout. The resulting nuclear state populations are then combined to extract the eight state populations of the nuclei and electron. In all eight cases, the measured state population of the initialized state is close to unity, with small deviation arising from SPAM errors.

S8. COMPARISON OF GST MODEL FITS

The reduced model we analyzed in the main text was selected through a process that compared multiple candidate models (see Fig. S7). In all the candidate models we considered, the errors on gate G_i were described using an error generator $\mathbb{L}_i = \log(G_i \mathbb{G}_i^{-1})$, where $\mathbb{G}_i$ is the ideal target operation. Any 2-qubit $\mathbb{L}_i$ can be written as a linear combination of 240 elementary error generators of four types described in Ref. [1]: Hamiltonian (H), Pauli-stochastic (S), Pauli-correlation (C), and active (A).

- H generators, indexed by a single Pauli operator, cause coherent unitary errors (e.g., H_{ZZ} generates a coherent ZZ rotation).
- S generators, also indexed by a single Pauli, cause probabilistic Pauli errors (e.g. S_{IX} causes probabilistic X errors on Q2).
- C generators, indexed by two Paulis, transform Pauli-stochastic errors into stochastic errors that are not aligned with the Pauli basis.
- A generators, indexed by two Paulis, cause errors requiring feedback from the environment, including cooling (e.g. T_1 decay).

The largest model we considered allowed all 240 error generators on each of the 6 gates, and had a total of 1263 parameters (1440 gate + 63 SPAM - 240 gauge). We label this model “CPTP” because it can model almost all gate sets consisting of completely positive trace-preserving operations. We then repeatedly considered smaller (reduced) models, and compared each one to the next larger model using the technique described in the “Constructing and selecting reduced models” Methods section of the main text.

First we pinned all of the A and C generator rates of the CPTP model (1260 parameters) to zero. The resulting “H+S” model had an evidence ratio of $r = 1.1$, indicating that it was clearly preferable to the CPTP model. Next, we pinned all weight-2 S generator rates (correlated stochastic errors, 54 parameters) on all gates. The evidence ratio between this “H+S1” model and the H+S model was $r = 2.8$, and so again we preferred the smaller model. Pinning all of the weight-2 H generator rates to zero in a third “H1+S1” model was unacceptable, as this model had an evidence ratio of $r = 21.1$ when compared with the H+S1 model. By considering several models that added back different sets of weight-2 H rates, we found the final model analyzed in the main text. This model, labeled “H1+S1+ZZ**”, includes all the error rates of the H1+S1 model (all weight-1 H and S generators) along with the following weight-2 H rates:

- H_{ZZ} on all 6 gates
- $H_{\mathbb{G}[ZZ]}$ on each single-qubit gate G , e.g. H_{YZ} for the $X_{\pi/2} \otimes I$ gate.

With only these specific weight-2 H generators, the H1+S1+ZZ** model has 33 fewer physical parameters (43 fewer rates, of which 10 are gauge degrees of freedom) than the H+S1 model and an evidence ratio of $r = 3.4$, which we accept based on our $r < 5$ criterion. The log-likelihood and evidence ratios for the nested series of candidate models are shown in Fig. S7, along with diagrams showing the decrease in allowed error rates per gate as we consider smaller models.

Reduced Models

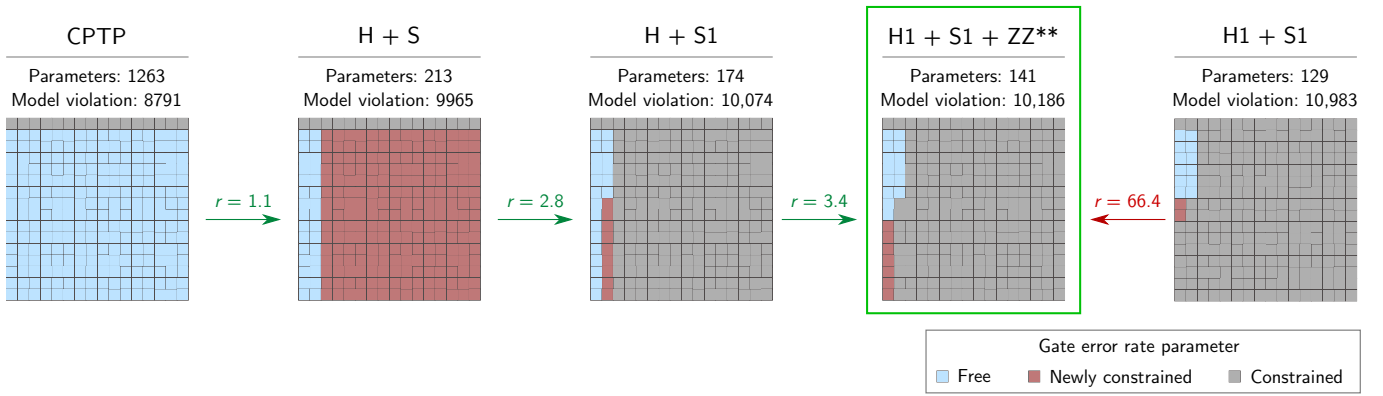


Figure S7. Comparison of candidate gate set models. Models are ordered left-to-right from the largest (most parameters) to the smallest. As explained in the text, the CPTP model allows gates to be arbitrary CPTP maps with all 240 error generator rates per gate. Moving rightward, more and more of these rates are pinned to zero, resulting in smaller models that necessarily provide worse fits to the data. For each model, a gate's allowed error generator rates are depicted as a set of blue boxes against a 15×16 grid of all the possible error generators for an arbitrary 2-qubit gate. Red boxes indicate the number of error generators that were pinned relative to the next larger model. Gray boxes indicate the number pinned to zero. For each model, the overall number (gates + SPAM - gauge) of model parameters N is given, along with the model violation, $2\Delta \log \mathcal{L}$ (lower = better fit to the data). The evidence ratio $r = \frac{2\Delta \log \mathcal{L}_i - 2\Delta \log \mathcal{L}_j}{N_j - N_i}$ is computed between each pair of models, and we prefer the smaller model when $r \leq 5$. A green rectangle highlights H1+S1+ZZ**, the best model among our candidates, which is used for the analysis in the main text.

S9. REDUCED METRICS OF GATE ERROR

Process matrices are the standard model for errors in quantum gates [2]. They provide a comprehensive description of all possible Markovian errors that can affect a gate's performance, and they can be used to predict the measurement outcome distribution for arbitrary quantum circuits. The price of this flexibility is that process matrices can be very complex. An arbitrary n -qubit process matrix is a dense $4^n \times 4^n$ matrix describing a completely positive, trace-preserving map on density matrices. One way to simplify the interpretation of process matrices is through the use of reduced metrics, such as average gate infidelity or the diamond error, that summarize the performance of a quantum gate with a single number. Because the impact of a gate's errors depends strongly on where that gate appears in a quantum circuit, there are many different reduced metrics. In this section, we discuss the most common metrics that appear in the literature, as well as their relation to the novel metrics we introduce in the main text.

Here, as in the main text, we represent a quantum gate by a process matrix $G = e^{\mathbb{L}}\mathbb{G}$, where $\mathbb{G}$ is the process matrix for the perfect unitary implementation of the gate and $\mathbb{L}$ is the error generator. As discussed in the Methods, the error generator is a weighted sum of Hamiltonian (H) and stochastic (S) terms. The rates of the Hamiltonian terms are labeled θ , and those of the stochastic terms are labeled ϵ . Here we define three of the most common reduced metrics of gate quality in terms of the gate's error generator $\mathbb{L}$:

1. The *entanglement infidelity* (ϵ_e):

$$\epsilon_e = 1 - \langle \varphi | (\mathbb{I} \otimes e^{\mathbb{L}}) (|\varphi\rangle\langle\varphi|) | \varphi \rangle, \quad (5)$$

where φ is any maximally entangled state over a doubled state space (here, a 4-qubit state space) and $\mathbb{I}$ is the identity operator.

2. The *average gate infidelity* ($\bar{\epsilon}$):

$$\bar{\epsilon} = 1 - \int d\psi \langle \psi | e^{\mathbb{L}} (|\psi\rangle\langle\psi|) | \psi \rangle, \quad (6)$$

where $d\psi$ is the Haar measure (here, over 2-qubit states).

3. The *diamond error* ($\epsilon_\diamond$):

$$\epsilon_\diamond = \frac{1}{2} \sup_{\rho} \| (e^{\mathbb{L}} \otimes \mathbb{I})(\rho) - \rho \|_1, \quad (7)$$

where ρ is a state over a doubled state space (here, a 4-qubit state space) and $\mathbb{I}$ is the identity operator. The supremum over states can be performed using a semidefinite program [3].

Both the entanglement infidelity (ϵ_e) and the average gate infidelity ($\bar{\epsilon}$) capture a gate's performance in a random context. As discussed below, these metrics are often used to report the results of randomized benchmarking [4]. The diamond error, on the other hand, characterizes worst-case performance. For any quantum circuit, the total variation distance between the observed and ideal probabilities of measurement outcomes is bounded above by the sum of diamond errors over all operations in the circuit [5]. For this reason, thresholds for error correcting codes are often stated in terms of the diamond error [6–8].

In the main text, we introduce two novel reduced metrics that can be computed simply in terms of the error generator rates:

4. The *generator infidelity* ($\hat{\epsilon}$)

$$\hat{\epsilon} = \epsilon_{\text{agg}} + \theta_{\text{agg}}^2, \quad (8)$$

where $\epsilon_{\text{agg}} = \sum_i \epsilon_i$ is the sum of the rates of the stochastic error generators, and $\theta_{\text{agg}} = \sqrt{\sum_i \theta_i^2}$ is the root sum square of the rates of the Hamiltonian error generators. As shown below in Sec. S9A, the generator infidelity is approximately equal to the entanglement infidelity.

5. The *total error* (ϵ_{tot})

$$\epsilon_{\text{tot}} = \epsilon_{\text{agg}} + \theta_{\text{agg}}, \quad (9)$$

where ϵ_{agg} and θ_{agg} are defined above. The total error captures worst-case performance, and, as shown in Sec. S9B, is closely related to the diamond error.

In this work, five of the six gates we study are intended to implement single-qubit logic operations. In order to capture crosstalk effects, we model these single-qubit gates with *two-qubit* process matrices. Nonetheless, it can be useful to consider reduced metrics that quantify the performance of these gates when restricted to a single qubit (either the target qubit, or the spectator, which should undergo an identity operation). In the main text we use the following two restricted metrics:

6. The *single-qubit infidelity* ($\hat{\epsilon}^{(Q_j)}$):

$$\hat{\epsilon}^{(j)} = \epsilon_{\text{agg},(j)} + \theta_{\text{agg},(j)}^2, \quad (10)$$

where $\epsilon_{\text{agg},(j)} = \sum_{i,(Q_j)} \epsilon_i$ is the sum of the rates of the stochastic error generators with support solely on qubit Q_j , and $\theta_{\text{agg},(j)} = \sqrt{\sum_{i,(Q_j)} \theta_i^2}$ is root sum square of the rates of the Hamiltonian error generators that act solely on the qubit Q_j .

7. The *single-qubit average gate infidelity* ($\bar{\epsilon}^{(Q_j)}$):

$$\bar{\epsilon}^{(Q1)} = 1 - \frac{1}{2} \int d\psi \langle \psi | \text{tr}_{Q2} [e^{\mathbb{L}} (|\psi\rangle\langle\psi| \otimes \mathbb{I})] | \psi \rangle, \quad (11)$$

$$\bar{\epsilon}^{(Q2)} = 1 - \frac{1}{2} \int d\psi \langle \psi | \text{tr}_{Q1} [e^{\mathbb{L}} (\mathbb{I} \otimes |\psi\rangle\langle\psi|)] | \psi \rangle, \quad (12)$$

where $d\psi$ is the Haar measure over 1-qubit states, and tr_{Q_j} is a partial trace over the spectator qubit, Q_j . The single-qubit average gate infidelity is equal to the average gate infidelity of the following single-qubit quantum process:

- Initialize the system in a pure state of target qubit $|\psi\rangle\langle\psi|$ and a completely mixed state of the spectator ($\frac{1}{2}\mathbb{I}$).
- Apply the error channel ($e^{\mathbb{L}}$) to the two-qubit product state ($|\psi\rangle\langle\psi| \otimes \frac{1}{2}\mathbb{I}$).
- Trace out the spectator qubit.

The result is a mixed state (ρ), whose state infidelity to ψ can be computed simply as $1 - \langle \psi | \rho | \psi \rangle$. Averaging this quantity over Haar-random initial states yields the single-qubit average gate infidelity.

Each of the infidelities defined above has a corresponding *fidelity* equal to one minus the infidelity. In the literature, it is common to see the terms “infidelity” and “process infidelity” used to refer to either the entanglement infidelity or the average gate infidelity, often without specification. These two quantities are related to one another by a dimension-dependent proportionality factor [9], $\epsilon_e = \frac{2^n+1}{2^n}\bar{\epsilon}$, for an operation on n qubits. The average gate fidelity $\bar{\epsilon}$ is not a “stable” metric [10] – it does not compose nicely upon combination of multiple qubits – but it is nonetheless commonly used to report results from randomized benchmarking [4]. Randomized benchmarking has been broadly adopted to experimentally characterise the performance of quantum gates [11–15], so we have reported the on-target average gate fidelities to ease comparison of our results to previous work. Whenever “infidelity” appears unqualified in the main text, it refers to $\hat{\epsilon}$, as defined in Eq. 8.

In Supplementary Information S9 A, we show that the infidelity $\hat{\epsilon}$ is equal to the entanglement infidelity ϵ_e at lowest order in ϵ_{agg} and θ_{agg} . Unlike the average gate infidelity, the entanglement infidelity is a stable metric [10] – if two gates are performed on separate qubits in parallel, the entanglement fidelity of the composite layer is simply the product of the entanglement fidelities of the two individual gates. The stability of the entanglement infidelity partially motivated its adoption in the context of “cycle benchmarking,” which has been used, e.g., for estimating gate errors in a 10-qubit ion-trap quantum processor [16].

In Supplementary Information S9 B, we show that the total error ϵ_{tot} provides an upper bound on the diamond error for single-qubit process matrices. This bound must be relaxed slightly for multi-qubit processes.

The infidelity $\hat{\epsilon}$ and the total error ϵ_{tot} also share the appealing property of being equal to each other when $\theta_{\text{agg}} = 0$, i.e., in the absence of coherent Hamiltonian errors (“perfect gate calibration”). This property recalls a similar relationship between the diamond error and entanglement infidelity, which coincide for Pauli stochastic error channels [4].

In Table S1, we list the infidelity, the entanglement infidelity, the total error, and the diamond error for the six gates considered in the main text. For all gates, the infidelity and the entanglement infidelity are nearly equal, and the total error and the diamond distance agree to within 22% relative error.

Gate	Infidelity $\hat{\epsilon}$	Entanglement infidelity ϵ_e	Total error ϵ_{tot}	Diamond error $\epsilon_{\diamond}$
$X_{\frac{\pi}{2}} \otimes \mathbb{I}$	0.68(6)%	0.66(6)%	3.4(3)%	3.2(2)%
$Y_{\frac{\pi}{2}} \otimes \mathbb{I}$	0.75(6)%	0.73(7)%	3.4(3)%	3.6(4)%
$\mathbb{I} \otimes X_{\frac{\pi}{2}}$	2.86(7)%	2.78(7)%	6.9(3)%	6.3(3)%
$\mathbb{I} \otimes Y_{\frac{\pi}{2}}$	3.44(10)%	3.34(11)%	6.3(5)%	5.9(1.6)%
$\mathbb{I} \otimes Y_{-\frac{\pi}{2}}$	3.54(16)%	3.39(18)%	9.1(5)%	7.7(5)%
CZ	0.79(14)%	0.80(16)%	5.5(4)%	7.0(4)%

Table S1. Comparison between the various metrics used to summarize quantum gate performance. The infidelity ($\hat{\epsilon}$) is very close to the entanglement infidelity (ϵ_e) – all the relative differences, $|\hat{\epsilon} - \epsilon_e|/\epsilon_e$, are less than 5%. The 1σ confidence intervals show that the two quantities take values consistent with each other, indicating that the higher order terms discussed in the text are insignificant. The the total error (ϵ_{tot}) and the diamond error ($\epsilon_{\diamond}$)

show larger disagreement, but all relative differences, $|\epsilon_{\text{tot}} - \epsilon_{\diamond}|/\epsilon_{\diamond}$, are less than 22%.

A. Relationship between generator infidelity and entanglement infidelity

As part of our analysis within the main text, infidelities are computed for each gate in the H1+S1+ZZ** model. As stated in the “Aggregated error rates and metrics” Methods section, the infidelity we report, $\hat{\epsilon}$, is computed by summing the rates of the S generator rates and the squares of the rates of the H generator rates:

$$\hat{\epsilon} = \sum_i \epsilon_i + \sum_j \theta_j^2 = \epsilon_{\text{agg}} + \theta_{\text{agg}}^2, \quad (13)$$

where ϵ_i and θ_j run over all the S and H generator rates respectively. The infidelity ($\hat{\epsilon}$) is closely related to the entanglement infidelity (ϵ_e), defined in Eq. 5. This relationship can be seen by direct calculation. The entanglement infidelity [9] of a quantum process $\rho \rightarrow \mathcal{E}(\rho)$ is equal to:

$$\epsilon_e = 1 - \frac{1}{d^3} \sum_Q \text{tr}(Q\mathcal{E}(Q)) \quad (14)$$

where Q is a Pauli operator and d is the Hilbert space dimension. We express the quantum process using the error generator framework introduced in Ref. [1] as $\mathcal{E}(Q) = \exp(\mathbb{L})(Q)$ and we assume the Lindblad generator $\mathbb{L}$ contains only Hamiltonian (H) and stochastic (S) generators. Each of these generators is indexed by a single Pauli operator:

$$H_P(Q) = -i[P, Q] \quad (15)$$

$$S_P(Q) = PQP - Q \quad (16)$$

Expanding the error process to second order in the generator, we have:

$$\epsilon_e = 1 - \frac{1}{d^3} \sum_Q \text{tr}(Q \exp(L)(Q)) \quad (17)$$

$$\simeq 1 - \frac{1}{d^3} \sum_Q \text{tr} \left(Q \left[1 + \sum_P (\theta_P H_P + \epsilon_P S_P) + \frac{1}{2} \sum_{P,P'} (\theta_P H_P + \epsilon_P S_P) (\theta_{P'} H_{P'} + \epsilon_{P'} S_{P'}) \right] (Q) \right) \quad (18)$$

The terms in this sum evaluate to:

$$\frac{1}{d^3} \sum_Q \text{tr}(Q^2) = 1 \quad (19)$$

$$\frac{1}{d^3} \sum_{P,Q} \theta_P \text{tr}(Q H_P(Q)) = \frac{-i}{d^3} \sum_{P,Q} \theta_P \text{tr}(Q[P, Q]) = 0 \quad (20)$$

$$\frac{1}{d^3} \sum_{P,Q} \epsilon_P \text{tr}(Q S_P(Q)) = \frac{1}{d^3} \sum_{P,Q} \epsilon_P \text{tr}(QPQP - Q^2) = - \sum_P \epsilon_P \quad (21)$$

$$\frac{1}{2d^3} \sum_{Q,P,P'} \theta_P \theta_{P'} \text{tr}(Q H_P(H_{P'}(Q))) = -\frac{1}{2d^3} \sum_{Q,P,P'} \theta_P \theta_{P'} \text{tr}(Q[P, [P', Q]]) = - \sum_P \theta_P^2 \quad (22)$$

$$\frac{1}{2d^3} \sum_{Q,P,P'} \epsilon_P \theta_{P'} \text{tr}(Q S_P(H_{P'}(Q))) = \frac{1}{2d^3} \sum_{Q,P,P'} \epsilon_P \theta_{P'} \text{tr}(QP[P', Q]P - Q[P', Q]) = 0 \quad (23)$$

$$\frac{1}{2d^3} \sum_{Q,P,P'} \theta_P \epsilon_{P'} \text{tr}(Q H_P(S_{P'}(Q))) = \frac{-i}{2d^3} \sum_{Q,P,P'} \theta_P \epsilon_{P'} \text{tr}(Q[P, P'QP' - Q]) = 0 \quad (24)$$

$$\frac{1}{2d^3} \sum_{Q,P,P'} \epsilon_P \epsilon_{P'} \text{tr}(Q S_P(S_{P'}(Q))) = \frac{1}{2d^3} \sum_{Q,P,P'} \epsilon_P \epsilon_{P'} \text{tr}(QP(P'QP' - Q)P - Q(P'QP' - Q)) = O(\epsilon^2) \quad (25)$$

Including the terms that are lowest-order in ϵ and θ gives:

$$\epsilon_e \simeq \sum_P (\epsilon_P + \theta_P^2) = \hat{\epsilon} \quad (26)$$

In evaluating the traces above, we have used the following facts for d -dimensional Pauli matrices Q, P, P' :

- $\text{tr}(Q) = 0$ unless Q is the identity matrix, in which case the trace is equal to d .
- $\text{tr}(Q^2) = d$
- $\text{tr}(PP') = d$ if $P = P'$. Otherwise this trace vanishes.
- $\text{tr}(QPQP - Q^2) = -2d$ if $\{P, Q\} = 0$. Otherwise, this trace vanishes.
- $\text{tr}(QP'QP - P'P) = -2d$ if $P = P'$ and $\{P, Q\} = 0$. Otherwise, this trace vanishes.
- There are d^2 Pauli matrices, and any non-identity Pauli matrix commutes with exactly half of all Pauli matrices, and anti-commutes with the other half.

A comparison of $\hat{\epsilon}$ and ϵ_e for the gates discussed in the main text is shown in Table. [S1](#).

B. Relationship between the total error and the diamond error

In this section, we show that the total error (ϵ_{tot}) upper bounds the diamond error ($\epsilon_{\diamond}$) for single-qubit gates in the H+S model. For gates on multiple qubits, the diamond error can, in the worst case, exceed the total error by a factor that grows exponentially in the number of qubits.

The error generator for a gate in the H+S model is a sum of stochastic and Hamiltonian terms. S and H generators are each indexed by a single n -qubit Pauli matrix (P or Q):

$$\mathbb{L} = \sum_P \epsilon_P S_P + \sum_Q \theta_Q H_Q. \quad (27)$$

By the Trotter-Suzuki formula,

$$\exp(\mathbb{L}) = \lim_{t \rightarrow \infty} \left[\exp \left(\sum_Q \frac{\theta_Q}{t} H_Q \right) \prod_P \exp \left(\frac{\epsilon_P}{t} S_P \right) \right]^t \quad (28)$$

We can now compute the diamond error of the error map $\exp(\mathbb{L})$. By subadditivity of the diamond norm,

$$\epsilon_{\diamond}(\exp(\mathbb{L})) \leq \lim_{t \rightarrow \infty} t \left[\epsilon_{\diamond}(\exp(\sum_Q \theta_Q H_Q/t)) + \sum_P \epsilon_{\diamond}(\exp(\epsilon_P S_P/t)) \right] \quad (29)$$

We can bound the contributions of the stochastic terms:

$$\epsilon_{\diamond}(\exp(\epsilon_P S_P/t)) = \frac{1}{2} \sup_{\rho} \|\exp(\epsilon_P S_P/t)(\rho) - \rho\|_1 \quad (30)$$

$$= \frac{1}{2} \sup_{\rho} \|\exp(-\epsilon_P/t)\rho + (1 - \exp(-\epsilon_P/t))P\rho P - \rho\|_1 \quad (31)$$

$$= (1 - \exp(-\epsilon_P/t)) \frac{1}{2} \sup_{\rho} \|P\rho P - \rho\|_1 \quad (32)$$

$$= (1 - \exp(-\epsilon_P/t)) \quad (33)$$

$$\leq \epsilon_P/t, \quad (34)$$

where P is the Pauli operator associated with the S_P error generator. The supremum in Eq. 32 is equal to 2, and is achieved by $\rho = \frac{1}{2}(\mathbb{I} + Q)$ for any Pauli Q such that $\{P, Q\} = 0$.

The Hamiltonian term, $\exp(\sum_Q \theta_Q H_Q/t)$ describes unitary evolution about an effective Hamiltonian $H = \sum_Q \theta_Q Q/t$, where Q is the Pauli operator corresponding to the H_Q error generator. In order to evaluate this term's contribution to Eq. 29, we rely on the well-known formula for the diamond error of a unitary channel:

$$\epsilon_{\diamond}(U) = \sqrt{1 - \min_{|\psi\rangle} \left(\langle \psi | U | \psi \rangle^2 \right)}, \quad (35)$$

Because of the Trotter expansion, the rotation induced by the effective Hamiltonian is very small. Expanding the unitary operator to second order in the effective Hamiltonian, we see:

$$\epsilon_{\diamond}(U) = \max_{|\psi\rangle} \left(\langle \psi | H^2 | \psi \rangle - \langle \psi | H | \psi \rangle^2 \right) \quad (36)$$

$$= \max_{|\psi\rangle} \text{Var}_{|\psi\rangle}(H) \quad (37)$$

$$= \frac{1}{2}(E_{\text{max}} - E_{\text{min}}) \quad (38)$$

The state that maximizes the variance of the Hamiltonian is an equal superposition of the eigenstates corresponding to the highest (E_{max}) and lowest (E_{min}) eigenvalues of the Hamiltonian. For a single qubit, eigenvalues of the effective Hamiltonian have magnitude $E = \sqrt{\sum_Q \theta_Q^2}/t = \theta_{\text{agg}}/t$. This results from the fact that all single-qubit Pauli matrices anti-commute with one another. For multiple qubits, it is possible for the Hamiltonian to be a sum of commuting terms. In this case, the extremal eigenvalue corresponds to the simultaneous +1 eigenstate of the Pauli operators

comprising the Hamiltonian, with magnitude $E = \sum_Q \theta_Q/t$. There are $2^n - 1$ traceless, mutually commuting Pauli operators on an n -qubit system, so the largest eigenvalue can be as high as $\sqrt{2^n - 1}\theta_{\text{agg}}/t$. The magnitude of the other extreme eigenvalue generally doesn't grow quite as quickly because it corresponds to a frustrated state. For local error channels (no crosstalk), the discrepancy will instead be proportional to $\sqrt{n}$. For few qubits, this correction is relatively small, and so θ_{agg} can nonetheless serve as a useful heuristic for aggregating the contributions from the Hamiltonian terms.

For a single qubit, we can combine the above results to see:

$$\epsilon_{\diamond}(\exp(\mathbb{L})) \leq \lim_{t \rightarrow \infty} t \left[\sum_Q \epsilon_{\diamond}(\exp(\theta H_Q/t) + \sum_P \epsilon_{\diamond}(\exp(\epsilon_P S_P/t)) \right] \quad (39)$$

$$\leq \lim_{t \rightarrow \infty} t \left[\theta_{\text{agg}}/t + \sum_P \epsilon_P/t \right] \quad (40)$$

$$= \epsilon_{\text{agg}} + \theta_{\text{agg}} \quad (41)$$

$$= \epsilon_{\text{tot}}(\exp(\mathbb{L})) \quad (42)$$

For two or more qubits, the total error no longer bounds the diamond error. Instead, the significantly weaker bound follows from the above discussion:

$$\epsilon_{\diamond}(\exp(\mathbb{L})) \leq \epsilon_{\text{agg}} + \sum_Q \theta_Q \quad (43)$$

This bound is achievable (at first order). For example, an n -qubit channel where each qubit is subject to small, single-qubit, coherent Z -rotations will saturate this bound. A comparison of ϵ_{tot} and $\epsilon_{\diamond}$ for the gates discussed in the main text is shown in Table. [S1](#).

S10. DETAILED ANALYSIS OF THE $X_{\pi/2}$ GATE ON Q2

The analysis presented in Fig. 3 utilized several distinct metrics of gate performance. In this section, we detail how these metrics emerge by taking a detailed look at one exemplary gate, the $X_{\pi/2}$ gate on qubit Q2. Recall that the errors on this gate (and all the others) are given by a Lindbladian generator ($\mathbb{L}$), so that if G and $\mathbb{G}$ are the noisy estimated and ideal $\mathbb{I} \otimes X_{\pi/2}$ gate, respectively, then $G = e^{\mathbb{L}}\mathbb{G}$. As in the main text, $\mathbb{L}$ is a weighted sum of particular (see S8) Hamiltonian (H) and Pauli-stochastic (S) elementary error generators:

$$\mathbb{L} = \sum_P \epsilon_P S_P + \sum_Q \theta_Q H_Q, \quad (44)$$

where P and Q are two-qubit Pauli operators and the sums range over allowed elementary generators. Rates ϵ_P and θ_Q are given by our GST estimate. This estimate possesses gauge degrees of freedom which we resolve by either 1) choosing a gauge that minimizes the sum-of-squared error rate estimates, or 2) computing linear combinations of the estimated error rates that are, in the limit of small errors, gauge invariant.

The first option is more straightforward, and leads to the results shown in panels (a) and (b) of Fig. 3. Choosing a sensible gauge gives meaning to ϵ_P and θ_Q , which are then categorized by their support and whether they are intrinsic (those that commute with the gate) or relational (those that don't, see Methods). The categorization of errors on the $\mathbb{I} \otimes X_{\pi/2}$ gate are shown in Fig. S8 and listed in Table S9, the former of which is an excerpt from Extended Data Fig. 8 and identical to a column of Fig. 3b except for the presence of labeled rates. For example, we see that H_{XI} , with rate $\theta_{XI} = 0.4\%$, is categorized as having support on Q1 (the bar is in the left "lane") and being an intrinsic error (XI commutes with the $\mathbb{I} \otimes X_{\pi/2}$ gate).

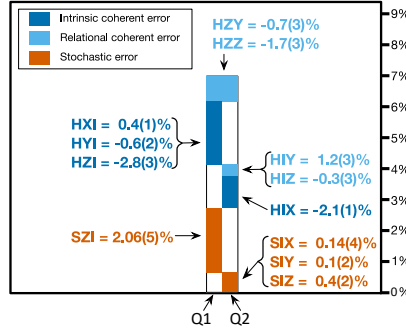


Figure S8. Distribution of error for the $X_{\pi/2}$ gate on Q2. Figure excerpted from Extended Data Fig. 8.

Elementary error generator	rate	type	support
H_{ZY}	-0.7%	relational	joint
H_{ZZ}	-1.7%	relational	joint
H_{XI}	0.4%	intrinsic	Q1
H_{YI}	-0.6%	intrinsic	Q1
H_{ZI}	-2.8%	intrinsic	Q1
H_{IY}	1.2%	relational	Q2
H_{IZ}	-0.3%	relational	Q2
H_{IX}	-2.1%	intrinsic	Q2
S_{ZI}	2.06%	intrinsic	Q1
S_{IX}	0.14%	intrinsic	Q2
S_{IY}	0.1%	intrinsic	Q2
S_{IZ}	0.4%	intrinsic	Q2

Figure S9. Elementary error rates of the $X_{\pi/2}$ gate on Q2

By aggregating these elementary rates we obtain the values in the cells of Fig. 3c. Stochastic errors add directly, and so the total stochastic error is $0.0206 + 0.0014 + 0.001 + 0.004 = 2.7\%$. Coherent errors add in quadrature, so aggregating the intrinsic H rates results in $\sqrt{0.004^2 + (-0.006)^2 + (-0.028)^2 + (-0.021)^2} = 3.6\%$. Similarly the relational H rates combine to give $\sqrt{(-0.007)^2 + (-0.017)^2 + 0.012^2 + (-0.003)^2} = 2.2\%$. The total coherent error (not shown in Fig. 3c), obtained by adding in quadrature all eight of the Hamiltonian error rates, is 4.2%.

Following the formulas for total error and infidelity given in S9 (and in Methods), we compute a total error of $0.027 + 0.042 = 6.9\%$ and an infidelity of $0.027 + 0.042^2 = 2.9\%$. Columns of Fig. 3c headed by (Q1) or (Q2) restrict the

aggregated error rates to those with the given support. For example, the total error for support (Q1) is found by summing the stochastic contribution, 2.06%, with the coherent contribution of $\sqrt{0.004^2 + (-0.006)^2 + (-0.028)^2} = 2.9\%$.

The second way we resolve the issue of gauge freedoms is by computing gauge-invariant error rates. These are linear combinations of the elementary error rates of one or more gates that are, to first order, unaffected by small gauge transformations (i.e., the gauge transformation matrix M is close to the identity). These gauge-invariant error rates are shown Fig. 3d. Gauge-invariant rates that are local to a single gate are placed on the nodes of the graph. These are identical to the intrinsic rates from before, and so we see that the $\mathbb{I} \otimes X_{\frac{\pi}{2}}$ node displays the same intrinsic stochastic rate (2.7%) and intrinsic coherent rate (3.6%) that were computed above.

Gauge-invariant rates that are linear combinations of elementary generators from two or more gates are called gauge-invariant relational error rates. In our analysis, all the gauge-invariant relational error rates were pairwise – they involved error generators from just 2 gates – and they are displayed as edges in Fig. 3d. These are a different type of error rate: they cannot be attributed to any one gate and have “units” different from those of the gauge-fixed on-gate relational errors of Fig. 3a,c (e.g., an error in the angle between two gates’ rotation axes). The largest relational error involving $\mathbb{I} \otimes X_{\frac{\pi}{2}}$ is a 2.7% coherent error relative to the $Y_{\frac{\pi}{2}} \otimes \mathbb{I}$ gate. This indicates that by different choices of the gauge, error can be moved between these gates.

S11. RATIONALIZING TWO-QUBIT ERRORS ON ONE-QUBIT GATES

In Section S11 A we discuss the evidence for the surprising weight-2 entangling errors that GST infers for our one-qubit gates, concluding that it is not caused by non-Markovian effects or outlier data points. This motivates a search for the physical origins of these errors. In Section S11 B we will conclude that the direct entangling interaction between ^{31}P nuclear spins is far too weak. Then, in Section S11 C we will account for the presence of an indirect electron-mediated entangling interaction between the nuclear spins through concatenated Schrieffer-Wolff transformations [17] that leave us with dressed nuclear spin qubit states. Finally, in Section S11 D we will consider the impact of driving fields in the doubly Schrieffer-Wolff transformed frame that defines our nuclear spin qubit states. We will find that plausible magnetic and electric fluctuations in the effective Hamiltonian, occurring due to the NMR drive, do not rationalize the GST-inferred coherent entangling errors. However, in Section S11 E, we invoke the leakage of microwaves near the ESR transition frequencies to rationalize the effective entangling interaction. The effect of this microwave leakage is qualitatively different from the types of errors studied in the context of fluctuations in the NMR drive, as they result in the accumulation of a geometric phase on the electron spin qubit that is conditioned on the nuclear spin state. The sizes of the observed errors are consistent with plausible levels of leakage in our experimental setup.

A. Validation of entangling errors

Analysis of the GST data using model selection (see “Construction of reduced model” in the Methods section of the main text) indicated strong evidence for weight-2 entangling H errors on the single-qubit gates. Our final 83-parameter model contained 2 entangling H errors on each of the 5 single-qubit gates. Eliminating them from the model decreased $2\log \mathcal{L}$ by 803. This evidence ratio of $r = 80$ constitutes compelling evidence that entangling errors on single-qubit gates are necessary to explain the data (we consider $r > 5$ significant). Eliminating H_{ZZ} and $H_{G[ZZ]}$ on any single gate yielded evidence ratios between 46 and 257, which constitutes compelling evidence for entangling errors on every gate.

We investigated whether this evidence could be caused by non-Markovian effects (e.g. slow drift or persistent environments), or whether it might be the result of outlier data points. The quality of the GST fit indicated a moderate degree of non-Markovianity – i.e., variations in the data that are inconsistent with any Markovian gate set – as follows. Relative to a “maximal model” constructed to fit each one of the 1592 circuit’s observed frequencies exactly, the full GST model displayed a total $2\Delta \log \mathcal{L}$ of 8793. Perfectly Markovian data would yield $2\Delta \log \mathcal{L} \approx 3 \times 1592 - 1263 = 3513$ (the number of free parameters in the data minus the number in the model), so these data are overdispersed by a factor of $\sqrt{8793/3513} \approx 1.6$. In principle this much overdispersion could create spurious effects at the scale we observed, but only if the deviations are concentrated on a small number of circuits. We did not observe such concentration. Instead, the observed deviations on almost all of the 1592 circuits were consistent with an overdispersed χ^2_3 distribution, indicating that the physical effects causing the deviations affected all circuits fairly uniformly. However, we observed between 6 and 15 clear outliers. Eliminating the largest 6 or 15 outliers and reanalyzing the data did not change conclusions.

Finally, we looked for correlation between (1) the circuits that provided evidence for non-Markovianity, and (2) the circuits that provided evidence for entangling errors. For each circuit, we computed:

1. Its contribution to $2\Delta \log \mathcal{L}$ between (a) the maximal model and (b) the full GST model;
2. Its contribution to $2\Delta \log \mathcal{L}$ between (a) the GST model including entangling errors and (b) the GST model without entangling errors.

We observed no evidence at all of correlation.

In summary, we found no reason to suspect that the evidence of entangling errors on single-qubit gates is an artifact of outliers, data analysis, or systematic non-Markovian effects. This does not rule out such explanations, but it motivates searching for physical (Hamiltonian) causes in what follows.

B. The direct entangling interaction between nuclear spins is too weak

The only direct entangling interaction between the two nuclear spin qubits is the mutual interaction of their magnetic dipole moments through the magnetic fields that they generate. Due to the dipolar nature of these fields the strength of this interaction decays cubically with the distance between the nuclei, realizing a coupling of ≈ 1 Hz at 5 nm. For a one-qubit gate time of $\approx 10 \mu\text{s}$, this translates into an always-on ZZ error of $\approx 10 \mu\text{rad}$ in size. Instead, the coherent entangling errors are predicted to be ≈ 10 mrad, which would require an internuclear separation of 0.5 nm. This is approximately the lattice constant of silicon, meaning that the nuclear spins would have to be within only a couple

of lattice sites of each other to rationalize this direct interaction as the source of the coherent entangling errors on our one-qubit gates. Such proximity is inconsistent with the degree of controllability of the relative contact hyperfine interactions of the two nuclei (Extended Data Fig. 2), the weak anisotropic hyperfine interactions evident in the non-demolition nature of our readout (Extended Data Fig. 5), and the NMR spectra reported in Section S1. We can thus conclude that the direct interaction between the two nuclear spins is not the origin of these errors.

C. Accounting for electron-mediated internuclear entanglement

This still leaves the possibility of an indirect effective interaction between the two nuclear spins that is mediated by the electron that they share. To study the impact of such an effective interaction, we apply a sequence of two transformations to the static Hamiltonian for the two nuclear spins and their shared electronic spin. The first transformation will mix certain states with opposite nuclear spin parity and opposite electronic spin. This will give us an effective Hamiltonian in which there is a XX+YY coupling between the nuclear spins. We can then project onto the electron spin-down subspace of this Hamiltonian and apply a second transformation to reduce the remaining off-diagonal coupling even further. Identifying this doubly dressed basis will be critical to analyzing the impact of the electron-mediated effective entangling interaction in the presence of the driving fields that implement one-qubit gates in Section S11 D.

We first separate the static lab frame three-qubit Hamiltonian ($\hat{H}$) into the diagonal Zeeman term ($\hat{H}_0$) and a perturbation due to the contact hyperfine interaction ($\hat{V}$),

$$\hat{H} = \hat{H}_0 + \hat{V}, \quad (45a)$$

$$\hat{H}_0 = -\gamma_n B_0 (\hat{I}_{1,z} + \hat{I}_{2,z}) - \gamma_e B_0 \hat{S}_z, \text{ and} \quad (45b)$$

$$\hat{V} = (A_1 \hat{I}_{1,x} + A_2 \hat{I}_{2,x}) \hat{S}_x + (A_2 \hat{I}_{1,y} + A_2 \hat{I}_{2,y}) \hat{S}_y + (A_1 \hat{I}_{1,z} + A_2 \hat{I}_{2,z}) \hat{S}_z. \quad (45c)$$

The first transformation that we will apply is a Schrieffer-Wolff transformation to third order in $\hat{V}$ accompanied by a projection onto the electron spin-down subspace. The unitary component of this transformation is generated by a Hermitian operator of the form

$$\begin{aligned} \hat{G}_a = & \theta_1 \left[1 + \frac{\theta_2^2}{4} \right] (\hat{I}_{1,y} \hat{S}_x - \hat{I}_{1,x} \hat{S}_y) + \theta_2 \left[1 + \frac{\theta_1^2}{4} \right] (\hat{I}_{2,y} \hat{S}_x - \hat{I}_{2,x} \hat{S}_y) \dots \\ & \dots + \theta_1 \theta_2 (\hat{I}_{1,y} \hat{I}_{2,z} \hat{S}_x + \hat{I}_{1,z} \hat{I}_{2,y} \hat{S}_x - \hat{I}_{1,x} \hat{I}_{2,z} \hat{S}_y - \hat{I}_{1,z} \hat{I}_{2,x} \hat{S}_y), \end{aligned} \quad (46)$$

where

$$\theta_1 = \frac{A_1}{(\gamma_e - \gamma_n) B_0} = -2.55 \times 10^{-3} \text{ rad and} \quad (47a)$$

$$\theta_2 = \frac{A_2}{(\gamma_e - \gamma_n) B_0} = -2.42 \times 10^{-4} \text{ rad,} \quad (47b)$$

in which the numerical values for the coefficients are $A_1=95$ MHz, $A_2=9$ MHz, $\gamma_n=17.23$ MHz T⁻¹, $\gamma_e=-27.97$ GHz T⁻¹, and $B_0=1.33$ T.

This unitary transforms the bare eigenbasis of $\hat{H}_0$ into a dressed basis that incorporates the effects of $\hat{V}$ such that we can project out the electron spin-up subspace, leaving only the two nuclear spin qubits. From the form of $\hat{G}_a$, it is evident that this change of basis will mix states in which the electron and exactly one of the nuclear spins have both flipped. This dressed basis will be closer to the true qubit states than the eigenstates of $\hat{H}_0$, though the effective Hamiltonian will still not be strictly diagonal even in the absence of the driving fields that perform one- and two-qubit gates. Before writing down the effective Hamiltonian in the dressed basis, it is instructive to consider the form of the dressed ground state in terms of the bare eigenbasis. Up to a constant for normalization, the leading contributions are

$$|\uparrow\uparrow\downarrow'\rangle \sim |\uparrow\uparrow\downarrow\rangle + \theta_1 |\downarrow\uparrow\uparrow\rangle + \theta_2 |\uparrow\downarrow\uparrow\rangle + \mathcal{O}(\theta_1\theta_2). \quad (48)$$

This indicates that the dominant effect of the off-diagonal component of the contact hyperfine interaction is to weakly mix the bare ground state with *both* of the states with opposite nuclear parity *and* opposite electronic spin. This also suggests that the always-on nature of the contact hyperfine interaction means that our qubit states aren't *strictly* the up and down states of the individual nuclear spins, but they are to a very good approximation thanks to the smallness

of θ_1 and θ_2 evident in Eq. 47. It is this weak mixing that is at the heart of effective entangling interactions between the nuclear spins. While the effective Hamiltonian expanded to third order will only include second order terms from $\hat{G}_a$, we note that we included third order terms in Eq. 46 to illustrate that the next order of corrections will not introduce any new mixing among the bare eigenstates. Thus it will be safe to project the third order Hamiltonian onto the electron spin-down subspace and to concatenate a second transformation.

Projecting onto the electron spin-down subspace, the effective Hamiltonian in the dressed basis defined by $\hat{G}_a$ is

$$\hat{H}'_{\downarrow} = \hat{H}_{0,\downarrow} + \hat{V}_{eff,\downarrow} + \mathcal{O}(\hat{V}^4), \quad (49a)$$

$$\hat{H}_{0,\downarrow} = \left(-\gamma_n B_0 - \frac{A_1}{2} \left[1 - \frac{\theta_1}{2} - \frac{\theta_2^2}{4} \right] \right) \hat{I}_{1,z} + \left(-\gamma_n B_0 - \frac{A_2}{2} \left[1 - \frac{\theta_2}{2} - \frac{\theta_1^2}{4} \right] \right) \hat{I}_{2,z}, \text{ and} \quad (49b)$$

$$\hat{V}_{eff,\downarrow} = \left(\frac{A_1\theta_2 + A_2\theta_1}{4} \left[1 - \frac{\theta_1 + \theta_2}{4} \right] \right) (\hat{I}_{1,x}\hat{I}_{2,x} + \hat{I}_{1,y}\hat{I}_{2,y}) + \left(\frac{A_1\theta_2 + A_2\theta_1}{4} \frac{\theta_1 + \theta_2}{2} \right) \hat{I}_{1,z}\hat{I}_{2,z}, \quad (49c)$$

where we have shifted away terms proportional to the identity by redefining the zero of energy. We see that the change of basis and projection has renormalized the diagonal component of the contact hyperfine interaction experienced by either nuclear spin. We also note that there are now two always-on entangling interactions between the dressed nuclear spin qubit states:

1. A relatively strong XX+YY interaction, with strength ≈ 3 kHz.
2. A relatively weak ZZ interaction, with strength ≈ 4 Hz.

We note that the quoted strengths include factors of 2 from spin matrices.

On the timescale of an $\approx 10 \mu\text{s}$ one-qubit gate, these always-on interactions translate into a rotation of ≈ 30 mrad. While this rotation is the same order of magnitude as the ≈ 10 mrad coherent errors inferred by GST, most of this interaction is mediated by a term of the form $(\hat{I}_{1,x}\hat{I}_{2,x} + \hat{I}_{1,y}\hat{I}_{2,y})$ and the coefficient of the $\hat{I}_{1,z}\hat{I}_{2,z}$ term is three orders of magnitude too small to be consistent with the ZZ errors inferred by GST. The stronger XX+YY interaction seems to be at odds with the GST model that gives the best rationalization of our data. But these simplistic considerations do not account for the fact that the inferred two-qubit errors occur in the presence of driving RF fields that implement one-qubit gates.

To facilitate the analysis of the impact of driving fields on this effective interaction we need to reduce the remaining off-diagonal coupling due to the XX+YY interaction. The first Schrieffer-Wolff transformation mixed states with opposite nuclear parity *and* opposite electronic spin, whereas a second unitary that reduces this coupling will mix states with fixed electronic spin and the same nuclear spin parity. While one might naively expect the size of this mixing to be much smaller than the first transformation, the generator $\hat{G}_b$ will have comparable weights because the ratio of the off-diagonal coupling to the diagonal matrix elements separating the coupled states are roughly the same order of magnitude. In other words, while the effective XX+YY coupling is three orders of magnitude weaker than the contact hyperfine interaction, it is between states that are separated by an energy scale that is three orders of magnitude weaker than the electronic Zeeman splitting, namely the difference in the contact hyperfine interaction between the two nuclei.

The desired form of $\hat{G}_{b\downarrow}$ is

$$i\hat{G}_{b\downarrow} = i\theta_3 (\hat{I}_{1,y}\hat{I}_{2,x} - \hat{I}_{1,x}\hat{I}_{2,y}) \quad (50)$$

with the value of the mixing angle being

$$\theta_3 = \frac{(A_1\theta_2 + A_2\theta_1) \left(1 - \frac{\theta_1 + \theta_2}{4} \right)}{4 (E'_{\downarrow\uparrow} - E'_{\uparrow\downarrow})} \approx \frac{A_1\theta_2 + A_2\theta_1}{2(A_1 - A_2)} = 2.67 \times 10^{-4} \text{ rad}, \quad (51)$$

where $E'_{\downarrow\uparrow}$ and $E'_{\uparrow\downarrow}$ are the diagonal components of the effective Hamiltonian for the subscripted basis states. This second transformation eliminates the residual off-diagonal coupling in the effective Hamiltonian at leading order in the XX+YY interaction strength. The concatenation of the two transformations thus provides a description of the nuclear spin qubit states that incorporates dressing of the bare qubit basis states by a coherent electron-mediated internuclear interaction.

D. Analysis of the impact of the electron-mediated interaction on one-qubit gates

Finally, we consider the impact of the driving fields in terms of the doubly dressed Schrieffer-Wolff basis that defines our nuclear spin qubit states. This change of basis naturally includes the electron-mediated entangling interaction

between nuclear spins that might explain the presence of coherent entangling errors during our one-qubit gates. There are five distinct oscillatory contributions to the effective Hamiltonian expressed in this basis that might rationalize these errors.

1. Resonant magnetic drive of the nucleus on which one-qubit gates are targeted.
2. Off-resonant magnetic drive of the nucleus on which one-qubit gates are not targeted.
3. Off-resonant magnetic drive of the electronic spin that mediates the entangling interaction between the nuclei.
4. Electrical drive of both nuclear spins, through the modulation of their contact hyperfine couplings.
5. Electrical drive of the electron spin, through the modulation of its gyromagnetic ratio.

The first three contributions are due to the B_1 field that implements one-qubit NMR gates and the last two contributions are due to an errant electric component to that driving field, which is henceforth indicated by a parametric dependence on the electric potential (V). Each of these will contributions will have the form

$$\hat{H}_{drive}(t) = \Delta \hat{H}_{gen} \cos(\omega t - \phi), \quad (52)$$

where Δ is the drive amplitude in units of frequency, $\hat{H}_{gen}$ is a unitless operator through which the drive generates a rotation, ω is the frequency of the drive chosen to be resonant with driving exactly one of the nuclear spin flip transitions, and ϕ is the phase of the drive that determines the precise rotation being implemented. At linear order in the Schrieffer-Wolff generators, these terms will introduce coherent errors of the form

$$\hat{H}_{err}(t) = \Delta \left([i\hat{G}_{a\downarrow}, \hat{H}_{gen}] + [i\hat{G}_{b\downarrow}, \hat{H}_{gen}] \right) \cos(\omega t - \phi). \quad (53)$$

Here we have extended the subscript notation for $\hat{G}_{b\downarrow}$ to $\hat{G}_{a\downarrow}$ to represent the combined action of $\hat{G}_a$ and a projection onto the electron spin-down subspace. The effect of this projection is to discard terms that are proportional to $\hat{S}_x$ or $\hat{S}_y$, which lead to stochastic errors when tracing out the electron. It will be evident that terms that are higher order in the generators will be too small to plausibly rationalize the errors of interest, as we will be unable to find sufficiently strong errors at leading order. We provide the form for each of these errors in Table S2.

Δ	$\hat{H}_{gen}$	$[i\hat{G}_{a\downarrow}, \hat{H}_{gen}]$	$[i\hat{G}_{b\downarrow}, \hat{H}_{gen}]$
$-\gamma_n B_1 \approx -25$ kHz	$\hat{I}_{1,x/y}$	0, $[i\hat{G}_a, \hat{H}_{gen}]$ strictly acts on $\hat{S}_{x/y}$	$\theta_3 \hat{I}_{1,z} \hat{I}_{2,x/y}$
$-\gamma_n B_1 \approx -25$ kHz	$\hat{I}_{2,x/y}$	0, $[i\hat{G}_a, \hat{H}_{gen}]$ strictly acts on $\hat{S}_{x/y}$	$\theta_3 \hat{I}_{1,x/y} \hat{I}_{2,z}$
$-\gamma_e B_1 \approx 40$ MHz	$\hat{S}_{x/y}$	$-\frac{\theta_1}{2} \hat{I}_{1,x/y} - \frac{\theta_2}{2} \hat{I}_{2,x/y} - \frac{\theta_1 \theta_2}{2} (\hat{I}_{1,x/y} \hat{I}_{2,z} + \hat{I}_{1,z} \hat{I}_{2,x/y})$	0, due to electron spin-down projection
$\delta A_1(V) \approx -100$ kHz	$\hat{I}_1 \cdot \hat{S}$	$\frac{\theta_1}{2} \hat{I}_{1,z} - \frac{\theta_2}{2} [1 + \frac{\theta_2}{2}] (\hat{I}_{1,x} \hat{I}_{2,x} + \hat{I}_{1,y} \hat{I}_{2,y}) - \frac{\theta_1 \theta_2}{4} (\hat{I}_{2,z} - 2\hat{I}_{1,z} \hat{I}_{2,z})$	$-\theta_3 (\hat{I}_{1,x} \hat{I}_{2,x} + \hat{I}_{1,y} \hat{I}_{2,y})$
$\delta A_2(V) \approx 70$ kHz	$\hat{I}_2 \cdot \hat{S}$	$\frac{\theta_2}{2} \hat{I}_{2,z} - \frac{\theta_1}{2} [1 + \frac{\theta_2}{2}] (\hat{I}_{1,x} \hat{I}_{2,x} + \hat{I}_{1,y} \hat{I}_{2,y}) - \frac{\theta_1 \theta_2}{4} (\hat{I}_{1,z} - 2\hat{I}_{1,z} \hat{I}_{2,z})$	$-\theta_3 (\hat{I}_{1,x} \hat{I}_{2,x} + \hat{I}_{1,y} \hat{I}_{2,y})$
$\delta \gamma_e(V) B_1 \approx 13$ Hz	$\hat{S}_{x/y}$	$-\frac{\theta_1}{2} \hat{I}_{1,x/y} - \frac{\theta_2}{2} \hat{I}_{2,x/y} - \frac{\theta_1 \theta_2}{2} (\hat{I}_{1,x/y} \hat{I}_{2,z} + \hat{I}_{1,z} \hat{I}_{2,x/y})$	0, due to electron spin-down projection
$\delta \gamma_e(V) B_0 \approx 12$ kHz	$\hat{S}_z$	$[i\hat{G}_a, \hat{H}_{gen}]$ strictly acts on $\hat{S}_{x/y}$	0, due to electron spin-down projection

Table S2. A summary of the forms of the error terms in Eq. 53 for each of the five mechanisms described at the beginning of Section S11 D. Estimates for the electric modulation of the hyperfine couplings and electron gyromagnetic ratio are taken from experiment, $\delta A_1(V) \approx -10 \text{ MHz V}^{-1} \times 10 \text{ mV} \approx -100 \text{ kHz}$, $\delta A_2(V) \approx 7 \text{ MHz V}^{-1} \times 10 \text{ mV} \approx 70 \text{ kHz}$, and $\delta \gamma_e(V) \approx -0.9 \text{ MHz V}^{-1} \text{ T}^{-1} \times 10 \text{ mV} \approx 9 \text{ kHz T}^{-1}$. We note that contributions that are third order in the generators are not indicated because they strictly renormalize errors that already occur at lower orders.

We next examine each entry of Table S2 to assess its impact on the execution of one-qubit gates. It is worth noting that the terms that resolve to zero because they act on $\hat{S}_{x/y}$ will still contribute to stochastic errors, but the focus of this assessment is on rationalizing the relatively strong coherent entangling errors. The only terms that will contribute appreciably to one-qubit gates are those that involves generators that flip exactly one of the nuclear spins. This is essential because of the oscillatory time dependence on all terms except for the term in the sixth row, which involves simultaneous electric and magnetic modulation resulting in an always-on term and a term that oscillates at twice the

driving frequency. For all other terms, those that flip exactly one nuclear spin will be constant in the rotating frame associated with driving a one-qubit gate on that particular nucleus.

We now summarize the strength and form of the coherent errors in Table S2. We will consider errors occurring during XI and IX gates for convenience and generalization to YI and IY gates is straightforward.

1. Applying an XI (IX) gate leads to an error of the form ZX (XZ) as is evident in the first two rows of Table S2. However, because the resonance frequencies of the two nuclei are different, this error will remain oscillatory in the rotating frame of the gate being applied. As the amplitude of this error is ≈ 2 Hz, this effect does not rationalize the errors inferred in the GST model.
2. However, the off-resonant magnetic drive on the *other* qubit when applying an XI (IX) gate will generate a XZ (ZX) error that is constant in the rotating frame of the gate being applied. However, over the course of an ≈ 10 μ s one-qubit gate, this will only lead to an XZ (ZX) rotation of ≈ 17 μ rad, which is still three orders of magnitude weaker than the plausible XZ (ZX) errors in the GST analysis.
3. Off-resonant magnetic drive of the electron that mediates the entangling interaction between nuclei gives rise to a number of errors evident in the third row of Table S2. The leading order contribution when driving an XI (IX) gate itself an XI (IX) “error”. Really, this is only an error in as far as it shifts the Rabi frequency and comparable physics has been predicted and observed in diamond NV centers [18, 19]. In spite of the smallness of θ_1 (θ_2), this shift is still 10s of kHz because of the largeness of γ_e and it is evident in our experiments. There is also a XZ (ZX) error proportional to $\theta_1\theta_2$. Over the course of an ≈ 10 μ s one-qubit gate, this will lead to an XZ (ZX) rotation of ≈ 31 μ rad, which is three orders of magnitude weaker than the plausible XZ (ZX) errors in the GST analysis.
4. Electric drive of the contact hyperfine couplings also gives rise to a number of errors, evident in the fourth and fifth rows of Table S2. However, each of them is rapidly oscillatory in the frame of any one-qubit X or Y gate. The amplitude of the aggregate XX+YY error is ≈ 19 Hz, so this effect does not rationalize the errors inferred in the GST model.
5. Electric drive of the electron gyromagnetic ratio gives rise to errors evident in rows six and seven of Table S2. Modulation of the coefficient on the oscillatory B_1 drive will give rise to a constant term and a term that oscillates at twice the drive frequency in a non-rotating frame. However the strength of the former will be exceptionally small, with XZ and ZX terms that have totally negligible amplitudes of ≈ 1 μ Hz. Modulation of the coefficient on the static B_0 coupling will strictly lead to stochastic errors.

In summary, we have shown that a microscopic model that accounts for all of the known physics in the NMR drive of our three-qubit system fails to rationalize the relatively large coherent entangling errors on one-qubit gates that were inferred by GST. The error mechanisms that come the closest to rationalizing these errors are due to an indirect electron-mediated entangling interaction between the nuclei, and the coupling of the magnetic drive to either the shared electron or the nucleus that isn’t being resonantly driven. However, the strength of the former (latter) mechanism is two (three) orders of magnitude weaker than the comparable errors that were inferred by GST.

E. Leakage of the ESR carrier signal leads to coherent entangling errors

In order to rationalize the observed ZZ errors, we instead need to turn to the leakage of microwaves near the ESR transition frequencies. The ESR pulses are driven by the modulation of a drive at a carrier frequency chosen to be off resonance relative to all four of the electron spin flip (ESR) transitions associated with the distinct nuclear spin basis states. When the ESR pulses are being driven, it is one of the sidebands generated by this modulation that is on resonance with one of these transitions. However, when these pulses aren’t being driven, the carrier frequency persists and some fraction of it will leak through the IQ-mixer.

The off-resonance drive provided by the leaked microwaves at the carrier frequency persist even during the NMR gates, which are produced by a separate signal generator. This leads to oscillations in the electron spin state, with a small amplitude proportional to the ratio of the intensity of the oscillating magnetic field produced by the carrier leakage and the detuning of the carrier frequency relative to any individual ESR transition. The frequency of these oscillations is determined by this same detuning, which is on the order of 10s of MHz allowing for the accumulation of many of these small oscillations over the course of an NMR gate. This detuning and the leakage magnetic field will differ for each of the individual nuclear spin basis states upon which the electron transitions are conditioned. Each basis state will then experience the accumulation of a distinct and spurious geometric phase due to these small but fast oscillations.

That these spurious phases depend on the support of the nuclear spin state on each of the relevant basis states during an NMR gate gives rise to the ZZ nature of the error that these leakage fields induce, rationalizing the coherent ZZ errors inferred by GST. Note also that the double-sideband modulation that is used to drive any individual ESR transition frequency introduces another source of off-resonance driving that introduces a comparable spurious accumulation of a geometric phase during the ESR pulse, i.e. the CZ gate, as well.

We next provide quantitative estimates of the size of errors due to leakage of the ESR carrier through the IQ-mixer. It is useful to begin by considering the impact of off-resonantly driving one of the ESR transitions for a fixed nuclear spin configuration. Without loss of generality, we consider this drive within the subspace spanned by $\{|\uparrow\uparrow\downarrow''\rangle, |\uparrow\uparrow\uparrow''\rangle\}$, where the $''$ notation indicates that these basis states are in the doubly Schrieffer-Wolff transformed frame. We provide the caveat that we are no longer considering projection onto the electron spin-down subspace prior to the second transformation and will thus require augmentation of the generating unitary to accommodate removal of the effective $XX+YY$ interaction on the electron spin-up space, as well.

It is then straightforward to adapt the solution to the off-resonant Rabi problem to the evolution of the dressed electron spin qubit states within this particular subspace. The effective rotating-frame Hamiltonian within this subspace is given as

$$H''_{\uparrow\uparrow} = \Delta_{\uparrow\uparrow} \hat{S}''_z + \Omega \hat{S}''_x, \quad (54)$$

where $\Delta_{\uparrow\uparrow}$ is the detuning of the carrier frequency relative to the ESR spin-flip transition frequency when the nuclei are in the $\uparrow\uparrow$ state (positive when the carrier frequency is less than the transition frequency), Ω is the effective Rabi frequency due to the coupling of the leakage field to the electron, and the notation $''$ indicates that the electron spin matrices are defined relative to the natural rotating frame in the doubly transformed basis. Ω is proportional to the effective B_1 field produced by the leaking carrier signal, which we will estimate based on the outcomes of our GST analysis. It is worth noting that because we are in the doubly Schrieffer-Wolff transformed frame Ω will be dressed by a nuclear-spin-dependent coupling of the leaked field to the nuclei beyond leading order. But this is negligible for the purposes of our estimates and we will assume that Ω is identical for all nuclear spin configurations.

The leakage field will drive small oscillations on the Bloch sphere associated with the dressed electron spin qubit basis states with frequency $\Omega_{\text{eff},\uparrow\uparrow} = \sqrt{\Omega^2 + \Delta_{\uparrow\uparrow}^2}$. The amplitude of these oscillations will be such that the maximum probability of the electron spin being observed as $|\uparrow\rangle$ is $\Omega^2/\Omega_{\text{eff},\uparrow\uparrow}^2$. The angle of the axis of this rotation is defined by the usual effective magnetic field, $\vec{B}_{\text{eff}} = \Omega \hat{x} + \Delta_{\uparrow\uparrow} \hat{z}$ (see the arrows indicated in Fig. S10). Over the period of one of these oscillations, $1/\Omega_{\text{eff},\uparrow\uparrow}$, the electron spin qubit will accumulate a geometric Aharonov-Anandan phase [20] as it moves in a small circle on the Bloch sphere starting and ending at $|\uparrow\uparrow\downarrow''\rangle$. The associated geometric phase is given as

$$\beta_{\uparrow\uparrow} = \pi \left(1 - \frac{\Delta_{\uparrow\uparrow}}{\Omega_{\text{eff},\uparrow\uparrow}} \right). \quad (55)$$

We will find it convenient to introduce a tilde to indicate the total geometric phase accumulated over a time t ,

$$\widetilde{\beta_{\uparrow\uparrow}} = \beta_{\uparrow\uparrow} \Omega_{\text{eff},\uparrow\uparrow} t = \pi \Omega t \left(\frac{\Omega}{2\Delta_{\uparrow\uparrow}} + \mathcal{O} \left(\left[\frac{\Omega}{\Delta_{\uparrow\uparrow}} \right]^3 \right) \right). \quad (56)$$

This picture can be extended to the other nuclear spin basis states noting that each will accumulate distinct geometric phases, $\beta_{\downarrow\downarrow}$, $\beta_{\downarrow\uparrow}$, $\beta_{\uparrow\downarrow}$, and $\beta_{\uparrow\uparrow}$. In the rotating frame of an NMR pulse implementing an $X_{\pi/2} \otimes I$ gate, the effective Hamiltonian accounting for the geometric phase due to leakage is

$$H_{X_{\pi/2} \otimes I} = \frac{\pi}{4} \left(e^{i(\widetilde{\beta_{\downarrow\downarrow}} - \widetilde{\beta_{\uparrow\uparrow}})} |\downarrow\downarrow''\rangle \langle \uparrow\downarrow''| + e^{i(\widetilde{\beta_{\uparrow\uparrow}} - \widetilde{\beta_{\downarrow\downarrow}})} |\uparrow\uparrow''\rangle \langle \downarrow\uparrow''| + h.c. \right), \quad (57)$$

which can be decomposed into tensor products of Pauli matrices as

$$\begin{aligned} H_{X_{\pi/2} \otimes I} = & \frac{\pi}{8} \left(\text{Re} \{ e^{i(\widetilde{\beta_{\downarrow\downarrow}} - \widetilde{\beta_{\uparrow\uparrow}})} + e^{i(\widetilde{\beta_{\uparrow\uparrow}} - \widetilde{\beta_{\downarrow\downarrow}})} \} X \otimes I - \text{Im} \{ e^{i(\widetilde{\beta_{\downarrow\downarrow}} - \widetilde{\beta_{\uparrow\uparrow}})} - e^{i(\widetilde{\beta_{\uparrow\uparrow}} - \widetilde{\beta_{\downarrow\downarrow}})} \} Y \otimes I \dots \right. \\ & \left. \dots + \text{Re} \{ e^{i(\widetilde{\beta_{\downarrow\downarrow}} - \widetilde{\beta_{\uparrow\uparrow}})} - e^{i(\widetilde{\beta_{\uparrow\uparrow}} - \widetilde{\beta_{\downarrow\downarrow}})} \} X \otimes Z - \text{Im} \{ e^{i(\widetilde{\beta_{\downarrow\downarrow}} - \widetilde{\beta_{\uparrow\uparrow}})} + e^{i(\widetilde{\beta_{\uparrow\uparrow}} - \widetilde{\beta_{\downarrow\downarrow}})} \} Y \otimes Z \right), \end{aligned} \quad (58a)$$

being careful to note that we have factored out a common coefficient of $1/2$. This expression can be further simplified at first order in the geometric phases,

$$H_{X_{\pi/2} \otimes I} = \frac{\pi}{4} \left(X \otimes I - \frac{1}{2} \left(\widetilde{\beta_{\downarrow\downarrow}} - \widetilde{\beta_{\uparrow\downarrow}} - \widetilde{\beta_{\uparrow\uparrow}} + \widetilde{\beta_{\downarrow\uparrow}} \right) Y \otimes I - \frac{1}{2} \left(\widetilde{\beta_{\downarrow\downarrow}} - \widetilde{\beta_{\uparrow\downarrow}} + \widetilde{\beta_{\uparrow\uparrow}} - \widetilde{\beta_{\downarrow\uparrow}} \right) Y \otimes Z \right). \quad (59)$$

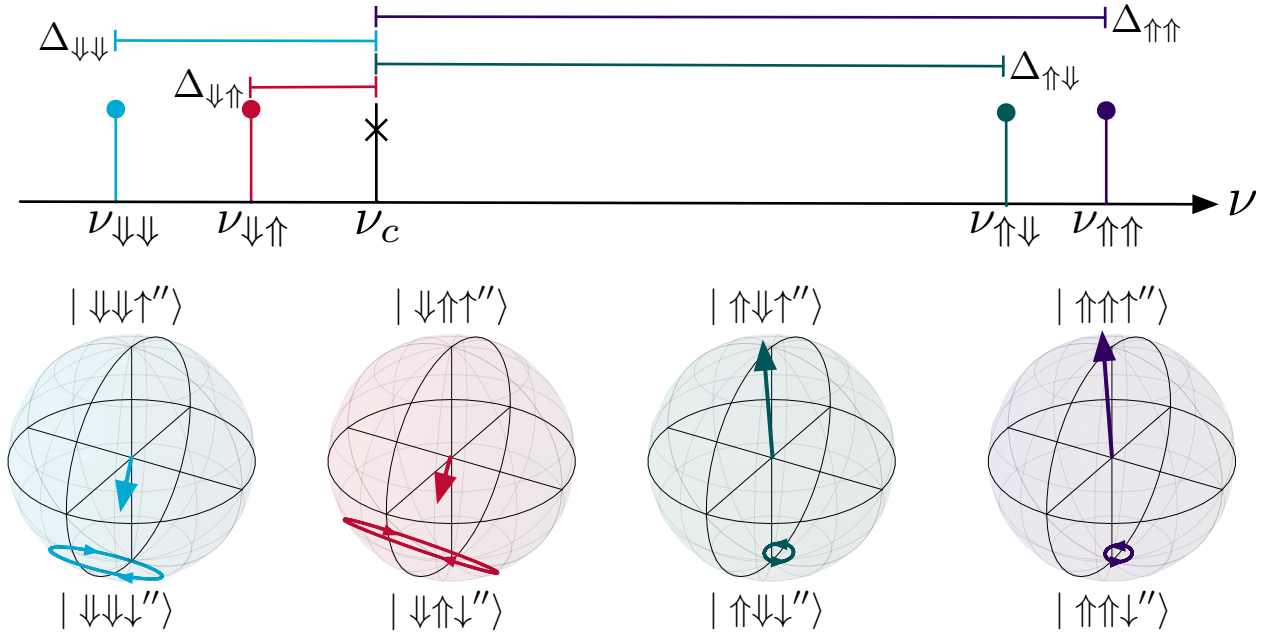


Figure S10. (Bottom panel) Bloch spheres for the electron spin qubit in each of the four nuclear spin basis states. The coloured vectors represent the effective driving field in the rotating frame, $\vec{B}_{eff}$, composed of the vector sum of the detuning and the driving field Ω , caused by the leakage of the carrier signal of the microwave source, at frequency ν_c , and the detunings Δ_{**} (where $**$ indicates any of the four nuclear orientations) from the individual ESR frequencies ESR spin-flip transitions ν_{**} (top panel). Starting from an electron spin in the $|\downarrow\rangle$ state (south pole of the Bloch sphere), the microwave leakage will lead to off-resonant driving of the electron spin through a trajectory indicated by the circles on each sphere, about an axis of rotation determined by $\vec{B}_{eff}$. These oscillations are quite small (≈ 1 mrad): for illustrative purposes, the value of Ω is drawn 100 times larger than the true experimental value. Even though the oscillations are 100 times smaller than illustrated, the associated geometric phases accumulate over the hundreds of cycles of the single-qubit NMR gates. While the effective Rabi frequency due to the leakage field is approximately the same for each nuclear spin basis state, the associated detunings vary over a range of 100 MHz (top). The basis state with the smallest detuning (in magnitude) will accumulate the largest geometric phase, in this case $|\downarrow\downarrow\downarrow\rangle$. The accumulation of different geometric phases across each nuclear spin basis states leads to coherent entangling errors on the nominally single-qubit gates. This figure was generated with the help of plotting routines in QuTiP [21].

This form of the rotating frame Hamiltonian makes it evident as to how leakage of the ESR carrier will lead to coherent errors on the nominally single-qubit gate. A unitary model for the gate itself is realized by the exponentiation of Eq. 59, leading to coherent single-qubit YI and ZI errors at first order in the geometric phase factors, as well as coherent YZ and ZZ errors. The ZI and ZZ errors arise due to the commutator between the XI term (rate ≈ 0.8) and the YI and YZ terms, appearing at first order in the geometric phases in the Zassenhaus formula. For a $\pi/2$ rotation the size of the ZI and ZZ errors will then be comparable to the YI and YZ errors, respectively, but they will be proportionally smaller for a rotation by a smaller angle (e.g., a rate $\ll 0.8$). It is straight forward to work out these same details for the other single-qubit gates, yielding two-qubit coherent errors of the form predicted by GST.

The carrier frequency used in our experiment was $\nu_c = 37\,100.4125$ MHz, and typical detunings from one set of measurements are $\Delta_{\downarrow\downarrow\downarrow} = -18.8225$ MHz, $\Delta_{\downarrow\downarrow\uparrow} = -11.6625$ MHz, $\Delta_{\uparrow\downarrow\downarrow} = 80.4375$ MHz, and $\Delta_{\uparrow\uparrow\uparrow} = 87.5475$ MHz. For these values, we can rationalize $\approx 1\%$ two-qubit coherent errors on a $20\,\mu\text{s}$ NMR gate with an effective leakage B_1 field of $5.8\,\mu\text{T}$. This translates to a 22 dB attenuation of the carrier relative to the 725 kHz ESR Rabi frequency driven with double-sideband modulation. This is a high but not implausible level of leakage, considering that we did not calibrate the IQ-mixer prior to the experiments.

For readers most accustomed to the standard picture of spin resonance, it is worth highlighting that this leakage has negligible effect on the electron spin state itself, as one can easily verify from a simple Rabi formula: the far-off-resonance drive on the electron resets to zero periodically and often. Conversely, the geometric phase imparted by electron on the nuclei does not reset – it *accumulates* with time. Since the imparted geometric phase differs for each of the four

ESR frequencies, due to the different values of Δ_{**} , it does not amount to an irrelevant global phase.

-
- [1] R. Blume-Kohout, M. P. da Silva, E. Nielsen, T. Proctor, K. Rudinger, M. Sarovar, and K. Young, A taxonomy of small markovian errors, arXiv preprint arXiv:2103.01928 (2021).
 - [2] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, *Quantum* **5**, 557 (2021).
 - [3] J. Watrous, Simpler semidefinite programs for completely bounded norms, *Chicago Journal of Theoretical Computer Science* **2013** (2013).
 - [4] E. Magesan, J. M. Gambetta, and J. Emerson, Characterizing quantum gates via randomized benchmarking, *Physical Review A* **85**, 042311 (2012).
 - [5] A. Y. Kitaev, Quantum computations: algorithms and error correction, *Russian Mathematical Surveys* **52**, 1191 (1997).
 - [6] P. Aliferis, F. Brito, D. P. DiVincenzo, J. Preskill, M. Steffen, and B. M. Terhal, Fault-tolerant computing with biased-noise superconducting qubits: a case study, *New Journal of Physics* **11**, 013061 (2009).
 - [7] P. Aliferis, D. Gottesman, and J. Preskill, Quantum accuracy threshold for concatenated distance-3 codes, *Quantum Inf. Comput.* **6**, 97 (2006).
 - [8] D. Aharonov and M. Ben-Or, Fault-Tolerant quantum computation with constant error rate, *SIAM J. Comput.* **38**, 1207 (2008).
 - [9] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, *Physics Letters A* **303**, 249 (2002).
 - [10] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Distance measures to compare real and ideal quantum processes, *Physical Review A* **71**, 062310 (2005).
 - [11] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, *et al.*, Superconducting quantum circuits at the surface code threshold for fault tolerance, *Nature* **508**, 500 (2014).
 - [12] M. Veldhorst, J. Hwang, C. Yang, A. Leenstra, B. de Ronde, J. Dehollain, J. Muhonen, F. Hudson, K. M. Itoh, A. Morello, *et al.*, An addressable quantum dot qubit with fault-tolerant control-fidelity, *Nature nanotechnology* **9**, 981 (2014).
 - [13] J. T. Muhonen, A. Laucht, S. Simmons, J. P. Dehollain, R. Kalra, F. E. Hudson, S. Freer, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, and A. Morello, Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking, *Journal of Physics: Condensed Matter* **27**, 154205 (2015).
 - [14] W. Huang, C. Yang, K. Chan, T. Tanttu, B. Hensen, R. Leon, M. Fogarty, J. Hwang, F. Hudson, K. M. Itoh, *et al.*, Fidelity benchmarks for two-qubit gates in silicon, *Nature* **569**, 532 (2019).
 - [15] X. Xue, T. Watson, J. Helsen, D. R. Ward, D. E. Savage, M. G. Lagally, S. N. Coppersmith, M. Eriksson, S. Wehner, and L. Vandersypen, Benchmarking gate fidelities in a Si/SiGe two-qubit device, *Physical Review X* **9**, 021011 (2019).
 - [16] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Characterizing large-scale quantum computers via cycle benchmarking, *Nature communications* **10**, 1 (2019).
 - [17] S. Bravyi, D. P. DiVincenzo, and D. Loss, Schrieffer–wolf transformation for quantum many-body systems, *Annals of physics* **326**, 2793 (2011).
 - [18] M. Chen, M. Hirose, P. Cappellaro, *et al.*, Measurement of transverse hyperfine interaction by forbidden transitions, *Physical Review B* **92**, 020101 (2015).
 - [19] S. Sangtawesin, C. McLellan, B. Myers, A. B. Jayich, D. Awschalom, and J. R. Petta, Hyperfine-enhanced gyromagnetic ratio of a nuclear spin in diamond, *New Journal of Physics* **18**, 083016 (2016).
 - [20] Y. Aharonov and J. Anandan, Phase change during a cyclic quantum evolution, *Physical Review Letters* **58**, 1593 (1987).
 - [21] J. R. Johansson, P. D. Nation, and F. Nori, Qutip: An open-source python framework for the dynamics of open quantum systems, *Computer Physics Communications* **183**, 1760 (2012).