

Subspace Leakage Error Randomized Benchmarking of Mølmer-Sørensen Gates

R. T. Sutherland,¹ A. C. Hughes,¹ J. P. Marceaux,¹ H. M. Knaack,¹ C. M. Löschnauer,¹ and R. Srinivas^{1,2,*}

¹*Oxford Ionics (an IonQ company), Oxford, OX5 1PF, UK*

²*Department of Physics, University of Oxford, Oxford, OX1 3PU, UK*

(Dated: October 13, 2025)

We demonstrate a new technique that adapts single-qubit randomized benchmarking to two-qubit Mølmer-Sørensen gates. We use the controllable gate phase to generate Cliffords that act on a two-state subspace, enabling benchmarking of two-qubit gates without single-qubit operations. In addition to quantifying the gate infidelity, the protocol provides valuable information about the type of error by distinguishing between those that conserve the two-state subspace and those that result in leakage out of it. We demonstrate the protocol for calibrating and validating all-electronic maximally entangling gates in a trapped-ion quantum computer, achieving a two-qubit gate error of $2.6(2) \times 10^{-4}$.

I. INTRODUCTION

One of the most important metrics describing a quantum computer is the fidelity of two-qubit entangling operations [1]. While two-qubit gates can already be performed below the quantum error correction (QEC) threshold in many physical platforms [2–5], further gate fidelity improvements are still highly beneficial, as they allow for reduced QEC overheads, additional near-term applications, and increased engineering margin during QC scale-up. To that end, we need methods that estimate gate errors *and* methods that characterize the noise mechanisms underlying the errors. Such methods also need to scale easily to lower error rates as two-qubit gate fidelities continue to improve.

The popular approach to gate quality characterization is through randomized benchmarks – a family that includes, among others, Clifford randomized benchmarking (RB) [6–8], character RB [9], cross-entropy benchmarking [10], mirror circuit benchmarks [11], and quantum volume [12]. Clifford RB is particularly efficient as it comprises a series of random Clifford operators that form a two-design on the subspace they act on, and can be simulated classically. Clifford RB can be used to measure arbitrarily small errors by increasing the sequence length, and measurement at different sequence lengths can also eliminate state preparation and measurement (SPAM) errors. However, as Clifford RB depolarises all error channels by design, it provides no information about the error sources. Further, the required Clifford operations are composed of both two-qubit and addressed single-qubit gates, making it challenging to disentangle errors associated with two-qubit rotations, single-qubit rotations, and individual addressing. Symmetric-subspace benchmarking [13] eliminates the need for individual addressing, but still requires single-qubit rotations. This requirement poses a limitation, especially as two-qubit gate error rates continue to improve [14].

On the other hand, many techniques have been developed to characterize the nature of errors in two-qubit gates. Historically, partial state tomography (PST) was used in the early entanglement demonstrations [15]. While PST is sensitive to SPAM errors, and – as it only contains a single entangling gate – has high statistical uncertainty per shot, separate measurements of the populations and coherences help distinguish between different error mechanisms. In general, tomographic approaches such as direct fidelity estimation [16], process tomography [17], and gate set tomography [18] can offer even greater insight into error processes. However, they require a larger number of measurements and have a more complex data analysis procedure. It would be useful to combine the scalability, SPAM insensitivity, and simplicity of Clifford RB, with the ability to pinpoint the origins of error offered by tomographic techniques.

In this paper, we propose and demonstrate one such method. Our technique is specifically aimed at characterizing two-qubit Mølmer-Sørensen gates [19, 20] in trapped-ion systems, and it uses the tunable MS gate phase degree of freedom ϕ_{MS} as the singular free parameter in each circuit decomposition. Varying ϕ_{MS} allows us to adapt single-qubit RB to generate Cliffords on a two-state subspace $\mathcal{S}_{\text{RB}}$ of the full two-qubit subspace $\mathcal{S}_{2Q}$. Measuring the rate of population transfer, both within $\mathcal{S}_{\text{RB}}$ and between $\mathcal{S}_{\text{RB}}$ and other subspaces $\mathcal{S}_{\text{leak}} \subset \mathcal{S}_{2Q}$, gives us an estimation of the average two-qubit gate fidelity and provides useful information for characterizing errors. Hence, we call the technique subspace leakage error randomized benchmarking (SLERB).

This paper is structured as follows. Section II describes the theory of SLERB using two approaches. One approach describes the technique by examining the relationship between certain population transfer matrix elements and the average gate infidelity (Sec. II B), and the other examines the technique’s group-theoretical properties (Sec. II C). In Section III, we perform numerical simulations of SLERB sequences to verify that it faithfully reproduces true gate errors. Finally, in Sec. IV, we demonstrate SLERB for calibrating and characterizing

* r.srinivas@oxionics.com

all-electronic MS gates on trapped-ion qubits, measuring an error rate of $2.6(2) \times 10^{-4}$.

II. THEORY

A. General overview

1. Fidelity estimation techniques

In the context of characterizing single- and two-qubit gates, Clifford RB has become a ubiquitous, industry-standard technique. Clifford RB “twirls” coherent gate noise into uniform depolarization noise that can be simply characterized. The fundamental mathematical reasoning is that sampling from a unitary two-design allows one to approximate the Haar measure [21]. The experimental implementation involves constructing physical gate sequences that ideally compile to representations of elements of the Clifford group on n -qubits. A circuit is constructed by selecting a random number of l Cliffords plus an $(l + 1)$ th inversion operation, and the whole sequence is then run after preparing a given initial state.

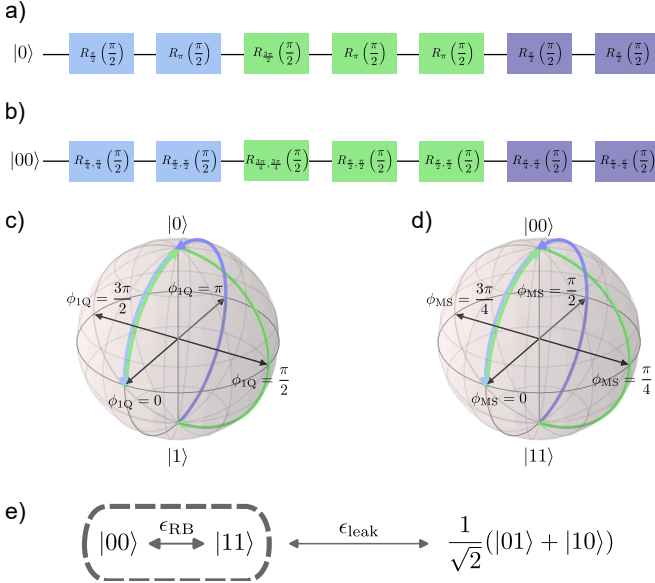


FIG. 1. Conversion of single-qubit to two-qubit sequences. SLERB sequences are generated from (a) sequences of single-qubit rotations ($R_\phi(\pi/2)$, $\phi = \phi_{1Q}$) (a), which are converted to (b) sequences of two-qubit MS gates, ($R_{\phi,\phi}(\pi/2)$, $\phi = \phi_{MS} = \phi_{1Q}/2$). Each pulse ideally rotates the states by an angle $\pi/2$ within their respective two-state subspaces, as shown in the Bloch spheres for (c) single and (d) two qubits. Each rotation is color-matched to its respective Clifford in (a) and (b), and corresponds to a specific axis of the Bloch sphere specified by ϕ_{1Q} or ϕ_{MS} . (e) Overview of population transfers. SU2 errors keep the states within the $\{|00\rangle, |11\rangle\}$ subspace, with a corresponding error rate ϵ_{RB} , while leakage errors result in populations in the $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ state with a corresponding error rate ϵ_{leak} .

The resulting success probability P_{survival} of measuring the input state at the output can be shown to decay according to a single exponential in the length l of the Clifford sequence

$$P_{\text{survival}} = A + Bq_{\text{RB}}^l. \quad (1)$$

Here, q_{RB} is an experimentally observed decay parameter, and $A, B \leq 1$ are constants that encode the effects of state preparation and measurement (SPAM) errors as well as any errors in the final inverting Clifford. Under certain assumptions, there is a strong relationship between q_{RB} and the fidelity of the gates used to generate the Clifford sequence [22].

Clifford RB has been generalized to address non-standard noise effects such as leakage. Importantly for our work, Ref. [23] generalized single-qubit Clifford RB to include the effects of leakage from the qubit manifold to neighboring states. In the case of only one leakage process, the resulting success probability is expected to decay according to a double exponential model

$$P_{\text{survival}} = A + Bq_{\text{RB}}^l + Cq_{\text{leak}}^l. \quad (2)$$

The decay in the case of leakage to many states can have more exponential terms – one for each type of leakage process. Similar work [24] introduced a “Blind RB” protocol that isolates single-exponential decays by taking linear combinations of P_{survival} and the bit flip probability P_{flip} , as

$$\begin{aligned} P_{\text{survival}} + P_{\text{flip}} &= 2(A + Cq_{\text{leak}}^l), \\ P_{\text{survival}} - P_{\text{flip}} &= 2Bq_{\text{RB}}^l. \end{aligned} \quad (3)$$

The technique we introduce is analogous to these protocols applied to a single-qubit subspace of a two-qubit gate.

Standard Clifford RB has also been modified to address particular experimental or platform-specific desiderata. Especially relevant for this work, Ref. [13] derived a protocol that implements an effective qutrit Clifford RB protocol [25] on the symmetric subspace (here, symmetric refers to qubit exchange) of a two-qubit Hilbert space in trapped ions. This protocol removes a number of experimental confounders, notably that antisymmetric operations in some trapped ion systems require transport, which might introduce new sources of error that are independent of the performance of the two-qubit gate.

2. Subspace leakage error randomized benchmarking

Figure 1 shows how our protocol builds upon single-qubit randomized benchmarking (1QRB), where a sequence of operations sampled at random from the set of single-qubit Cliffords $\mathcal{C}_{1Q}$ is used to estimate the error of single-qubit gates. Each $\hat{C}_{1Q} \in \mathcal{C}_{1Q}$ can be decomposed into a series of single-qubit gates driving $R_{\phi_{1Q}}(\pi/2)$ rotations around the x , y , $-x$, and $-y$ axes on the

Bloch sphere [26], with the rotation axis defined by ϕ_{1Q} . This sequence of single-qubit rotations is converted to two-qubit gates, driving $R_{\phi_{MS}, 1, \phi_{MS}, 2}(\pi/2)$ rotations, except the rotation axis is set by $\phi_{MS} = \phi_{1Q}/2$. Figure 1(d) shows how these pulses can be visualized on the $\mathcal{S}_{RB} \equiv \{|00\rangle, |11\rangle\}$ Bloch sphere.

We implement our benchmarking protocol following the approach of single-qubit RB [6]. Explicitly, we identify 24 MS pulse sequences that compile to $\mathcal{C}_{1Q}$ when acting on $\mathcal{S}_{RB}$. We select a sequence of l Cliffords at random from the set of 24 Cliffords and then calculate the inverse of the sequence. We then implement the physical pulse sequence for the $l + 1$ Clifford gates, including the final inverse operation, on an initial state of $|00\rangle$. In this way, we produce a final state $\rho^{(l)}$ according to the map

$$\rho^{(l)} = G_{l+1} \circ G_l \circ \dots \circ G_1 [|00\rangle\langle 00|], \quad (4)$$

where G_k is a channel representation of the k th noisy Clifford. Finally, we measure $\rho^{(l)}$ in the computational basis. We repeat the procedure to collect a number of shots s from a particular compilation, and we further run the protocol for a number r of different randomizations in the sequences. In this way, we collect $s \times r$ random categorical variables supported on the measurement outcome space $\{00, 01, 10, 11\}$. This protocol is a direct generalization of single-qubit RB. From this perspective, we can directly deploy standard approaches to calculate an effective SU(2) error rate on the embedded subspace of the MS gate used in SLERB. However, it is desirable to extend our results to estimate gate errors averaged over $\mathcal{S}_{2Q}$.

3. Clifford decompositions

A ‘standard’ two-qubit Clifford RB circuit comprises two-qubit Clifford operators $\mathcal{C}_{2Q}$. However, as generating $\mathcal{C}_{2Q}$ on a trapped-ion QC involves many physical operations in addition to two-qubit gates (addressed single-qubit gates, transport, cooling etc.), this creates challenges for directly estimating the two-qubit gate error rate. Ref. [13] partially addressed this issue with an RB variant they referred to as ‘subspace randomized benchmarking’ (SRB). The technique uses benchmarking sequences comprising symmetric (with respect to qubit exchange) $d = 3$ Cliffords $\mathcal{C}_+$. However, SRB still requires single-qubit gates in addition to two-qubit gates.

With this in mind, we focus on sequences that decompose into only the unitaries naturally generated by the Mølmer-Sørensen (MS) gate [19, 20]. Up to a global phase, this can be expressed as:

$$\hat{U}_{2Q}(\theta_{MS}, \phi_{MS}) = \exp\left(-\frac{i\theta_{MS}}{2}\hat{\sigma}_{\phi_{MS}, 1}\hat{\sigma}_{\phi_{MS}, 2}\right), \quad (5)$$

where $\hat{\sigma}_{\phi_{MS}, j} \equiv \hat{\sigma}_{x,j} \cos(\phi_{MS}) + \hat{\sigma}_{y,j} \sin(\phi_{MS})$ is a Pauli operator acting on qubit j . We choose ϕ_{MS} to be the free parameter to generate our Clifford decompositions;

this choice is convenient for MS gate experiments because ϕ_{MS} is the phase of the gate fields relative to the qubit and is relatively straightforward to adjust. However, this method can also be applied to $\propto \hat{\sigma}_{z,1}\hat{\sigma}_{z,2}$ gates by adding collective single-qubit $\pi/2$ -rotations before and after the gate.

The ‘even’ $\mathcal{S}_{RB} \equiv \{|00\rangle, |11\rangle\}$ and ‘odd’ $\mathcal{S}_{leak} \equiv \{|01\rangle, |10\rangle\}$ subspaces are invariant under $\hat{U}_{2Q}$; note that $\mathcal{S}_{RB} + \mathcal{S}_{leak} = \mathcal{S}_{2Q}$, where $\mathcal{S}_{2Q}$ is the total two-qubit Hilbert space. In the following, it will be convenient to further divide $\mathcal{S}_{leak}$ into symmetric $\mathcal{S}_{leak,+} \equiv \{2^{-1/2}(|01\rangle + |10\rangle)\}$ and antisymmetric $\mathcal{S}_{leak,-} \equiv \{2^{-1/2}(|01\rangle - |10\rangle)\}$ single-state subspaces. The action of $\hat{\sigma}_{\phi_{MS}, 1}\hat{\sigma}_{\phi_{MS}, 2}$ on $\mathcal{S}_{RB}$ is:

$$\begin{aligned} \hat{\sigma}_{\phi_{MS}, 1}\hat{\sigma}_{\phi_{MS}, 2} \Big|_{\mathcal{S}_{RB}} &= e^{-2i\phi_{MS}} |11\rangle\langle 00| + e^{2i\phi_{MS}} |00\rangle\langle 11|, \\ &\equiv \hat{\sigma}_{2\phi_{MS}, RB}. \end{aligned} \quad (6)$$

Thus, $\hat{U}_{2Q}$ is isomorphic to a single-qubit Pauli $\hat{\sigma}_{2\phi_{MS}, RB}$ operator when acting on $\mathcal{S}_{RB}$. Similarly, the collective Pauli- z operator:

$$\frac{1}{2}(\hat{\sigma}_{z,1} + \hat{\sigma}_{z,2}) \Big|_{\mathcal{S}_{RB}} \equiv \hat{\sigma}_{z, RB}, \quad (7)$$

is isomorphic to a single-qubit Pauli- z operator when acting on $\mathcal{S}_{RB}$. Since we can make effective $\propto \hat{\sigma}_{x, MS}$ and $\propto \hat{\sigma}_{y, MS}$ rotations on $\mathcal{S}_{RB}$, we can generate Euler rotations, and, therefore, any SU(2) operation on $\mathcal{S}_{RB}$. Hence, we can generate any $\hat{C}_{1Q}$ on $\mathcal{S}_{RB}$, and implement single-qubit RB on $\mathcal{S}_{RB}$ with only $\hat{U}_{2Q}$ operations.

B. Transfer matrix

In a benchmarking sequence, an ‘ideal’ circuit comprising l random Cliffords will map the qubit onto some target ‘survival’ state $|\text{survival}_l\rangle \in \mathcal{S}_{RB}$, i.e., the state generated if no errors occurred. In this section, we are interested in the population transfer matrix that results specifically from *errors* in each Clifford in this series. In the experiment, we measure the average population vector after l Cliffords $\vec{P}_l \equiv (P_{\text{survival}, l}, P_{\text{flip}, l}, P_{\text{leak}, l})$. Here, $P_{\text{survival}, l}$ is the probability of measuring $|\text{survival}_l\rangle$ after the l Clifford sequence, $P_{\text{flip}, l}$ is the population of $\mathcal{S}_{RB} - \{|\text{survival}_l\rangle\}$ the other state in the RB subspace, and $P_{\text{leak}, l}$ is the population in $\mathcal{S}_{leak}$ the total leakage subspace. We define the transfer matrix T with the equation:

$$\vec{P}_{l+1} = (I + T)\vec{P}_l, \quad (8)$$

where $T = 0$ indicates no population transfer errors. We want to determine the matrix elements of T . Let $k \in \mathcal{K}$ be the set of all noise sources k present during the gate, all of which are assumed to be small. We can calculate the population transfer resulting from each noise mechanism k (in isolation), then add them to obtain the transfer

matrix resulting from all of $\mathcal{K}$:

$$T = \sum_{k \in \mathcal{K}} T_k. \quad (9)$$

For a given k the total Hamiltonian is:

$$\hat{H}_t = \hat{H}_g + \hat{H}_{e,k}, \quad (10)$$

where $\hat{H}_g$ is the ideal gate Hamiltonian and $\hat{H}_{e,k}$ is the Hamiltonian describing error source k . In the limit that $\hat{H}_{e,k}$ is small we can factor the total propagator as [27]:

$$\hat{U}_t = \hat{U}_g \tilde{U}_{e,k}, \quad (11)$$

where $\tilde{U}_{e,k}$ is the time-propagator of $\hat{H}_{e,k}$, which can be calculated perturbatively [28]. Since it is unitary, we can express the error unitary as:

$$\tilde{U}_{e,k} \simeq e^{-i\alpha_k \hat{Q}_k \otimes \hat{M}_k}, \quad (12)$$

for some real number α_k and some Hermitian operator $\hat{Q}_k \otimes \hat{M}_k$, which we assume can be written as a tensor product of $\hat{Q}_k$ acting on the qubits and $\hat{M}_k$ acting on $|m\rangle$.

After l Cliffords, the system will be in some mixed state:

$$\rho^{(l)} = \sum_j P_{j,l} P_m |\psi_{j,l}\rangle \langle \psi_{j,l}| \otimes |m\rangle \langle m|, \quad (13)$$

which we interpret (equivalently) as the system having probability $P_{j,l} P_m$ of being in the pure state $|\psi_{j,l}\rangle |m\rangle$, where $|\psi_{j,l}\rangle$ refers to the qubits and $|m\rangle$ is an index that runs over the (potentially multiple) external degrees-of-freedom in the system. We assume the qubits are not entangled to any external degree of freedom after l Cliffords. Eq. 13 depends both on the distribution (set of incoming states and their probabilities) of qubit states $\mathcal{D}_l \equiv \{|\psi_{j,l}\rangle, P_{j,l}\}$ and external degrees-of-freedom $\mathcal{M} \equiv \{|m\rangle, P_m\}$ that are being averaged over; all of the information in $\rho^{(l)}$ about the qubit states is contained in $\mathcal{D}$ and all the information about the external degrees-of-freedom are in $\mathcal{M}$. The reason we have included the index l in $|\psi_{j,l}\rangle$ is because target $|\text{survival}_l\rangle$ and $|\text{flip}_l\rangle$ can change after each gate. The *average* probability $P_{g,k}$ that the next $\hat{U}_g$ generates ‘the target state $|g_{j,l+1}\rangle$, conditioned on input state $|\psi_{j,l}\rangle$ ’, i.e. $|g_{j,l+1}\rangle = \hat{U}_g |\psi_{j,l}\rangle$ even when $|\psi_{j,l}\rangle \neq |\text{survival}_l\rangle$, is:

$$P_{g,k} \equiv \sum_{m',m} \sum_j P_j P_m \left| \langle \psi_{j,l} | \langle m' | e^{-i\alpha_k \hat{Q}_k \otimes \hat{M}_k} | \psi_{j,l} \rangle | m \rangle \right|^2 \simeq 1 - \lambda_{\hat{Q}_k}^2 \gamma_k, \quad (14)$$

where:

$$\gamma_k \equiv |\alpha_k|^2 \sum_m P_m \langle m | \hat{M}_k^2 | m \rangle, \quad (15)$$

quantifies the size of the noise channel, and

$$\lambda_{\hat{Q}_k}^2 \equiv \sum_j P_j \left(\langle \psi_{j,l} | \hat{Q}_k^2 | \psi_{j,l} \rangle - \left| \langle \psi_{j,l} | \hat{Q}_k | \psi_{j,l} \rangle \right|^2 \right), \quad (16)$$

is the variance of $\hat{Q}_k$, averaged over $\mathcal{D}$. Eq. 17 is important because it shows we can factor out $E_{g,k}$'s dependence on $\mathcal{D}$. While this tells us the probability $E_{g,k} \equiv 1 - P_{g,k}$ that a gate error results in population transfer to some state other than $|g_{j,l}\rangle$:

$$E_{g,k} \simeq \lambda_{\hat{Q}_k}^2 \gamma_k, \quad (17)$$

we still need to know where this errant population will go before we can write an expression for T_k .

We divide the set of all noise sources $\mathcal{K} = \mathcal{K}_{\text{RB}} + \mathcal{K}_{\text{leak}}$ into two subsets: noise that conserves $\mathcal{S}_{\text{RB}}$, which we label $\mathcal{K}_{\text{RB}}$, and noise that does not, which we will label $\mathcal{K}_{\text{leak}}$. The classification of each k set by the structure of $\hat{Q}_k$; if needed for clarity, we indicate the category of $\hat{Q}_k$ with subscripts $\hat{Q}_{\text{RB},k}$ or $\hat{Q}_{\text{leak},k}$. It is possible that some noise mechanism results in a $\hat{Q}_k$ with components that fall into both categories. In this case, we can treat $\hat{Q}_k$ as two distinct noise mechanisms $\hat{Q}_{\text{RB},k}$ and $\hat{Q}_{\text{leak},k}$. We assume our error budget comprises only symmetric errors, see Sec. IID. If $|\psi_{j,l}\rangle \in \mathcal{S}_{\text{RB}}$, errors from $\mathcal{K}_{\text{RB}}$ will flip the state of $|g_{j,l+1}\rangle$ to the remaining state in $\mathcal{S}_{\text{RB}} - |g_{j,l+1}\rangle$. If $|\psi_{j,l}\rangle \in \mathcal{S}_{\text{leak}}$, then we assume $\mathcal{K}_{\text{RB}}$ does not affect $|g_{j,l+1}\rangle$. We can express every qubit operator resulting from $\mathcal{K}_{\text{RB}}$ as:

$$\hat{Q}_k \Big|_{\mathcal{S}_{\text{RB}}} = c_{i,k} \hat{I}_{\text{RB}} + c_{p,k} \vec{\sigma}_{\text{RB}} \cdot \hat{r}, \quad (18)$$

noting we can incorporate any normalization into $|\alpha_k|^2$ that is necessary to ensure $c_{i,k}^2 + c_{p,k}^2 = 1$. Since our gate set forms a two-design on $\mathcal{S}_{\text{RB}}$, we know the variances will be identical to the single-qubit Pauli operators averaged over $\text{SU}(2)$:

$$\begin{aligned} \langle \lambda_{\hat{Q}_k}^2 \rangle_{\mathcal{S}_{\text{RB}}} &= c_p^2 \langle \lambda_{\hat{Q}_k}^2 \rangle_{\mathcal{S}_{\text{RB}}} \\ &= \frac{2}{3} c_p^2. \end{aligned} \quad (19)$$

The only symmetric, $\mathcal{S}_{\text{RB}}$ conserving $\hat{Q}_{\text{RB}}$ that is $\propto \hat{I}_{\text{RB}}$ on $\mathcal{S}_{\text{RB}}$ is $\hat{\sigma}_{z,1} \hat{\sigma}_{z,2}$, while, for every other $\hat{Q}_{\text{RB}} \in \mathcal{Q}_{\text{RB}}$, $c_p^2 = 1$. Thus, $\propto \hat{\sigma}_{z,1} \hat{\sigma}_{z,2}$ error channels will be invisible to SLERB, and do not enter into our infidelity estimate; for the $\propto \hat{\sigma}_{\phi,1} \hat{\sigma}_{\phi,2}$ MS gates, error channels of this form are likely to be small. For every other symmetric $\hat{Q}_{\text{RB}}$, $c_p^2 = 1$, which we will assume from here on. We can now write an expression for the total rate of population transfer between the states in $\mathcal{S}_{\text{RB}}$:

$$\epsilon_{\text{RB}} = \frac{2}{3} \sum_{k \in \mathcal{K}_{\text{RB}}} \gamma_k, \quad (20)$$

giving the $\mathcal{S}_{\text{RB}}$ conserving component of the transfer matrix:

$$T_{\text{RB}} = \begin{bmatrix} -\epsilon_{\text{RB}} & \epsilon_{\text{RB}} & 0 \\ \epsilon_{\text{RB}} & -\epsilon_{\text{RB}} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (21)$$

Next, consider the set of two-qubit Pauli operators that couple $\mathcal{S}_{\text{RB}}$ and $\mathcal{S}_{\text{leak}}$: $\mathcal{Q}_{\text{leak}} \equiv \{(\hat{\sigma}_{\phi,1} + \hat{\sigma}_{\phi,2}), (\hat{\sigma}_{\phi,1}\hat{\sigma}_{z,2} + \hat{\sigma}_{z,1}\hat{\sigma}_{\phi,2})\}$. Each operator in $\mathcal{Q}_{\text{leak}}$ couples both the states in $\mathcal{S}_{\text{RB}}$ to the single-state subspace $\mathcal{S}_{\text{leak},+}$. The variance of either operator is $\langle \lambda_{\mathcal{Q}_{\text{leak}}}^2 \rangle_{\mathcal{S}_{\text{RB}}} = 2$ for every state in $\mathcal{S}_{\text{RB}}$, while $\langle \lambda_{\mathcal{Q}_{\text{leak}}}^2 \rangle_{\mathcal{S}_{\text{leak},+}} = 4$. This tells us the rate of population transfer from either state in $\mathcal{S}_{\text{RB}}$ to $|\text{leak},+\rangle$ is:

$$\epsilon_{\text{leak}} = 2 \sum_{k \in \mathcal{K}_{\text{leak}}} \gamma_k. \quad (22)$$

The rate of population transfer from $\mathcal{S}_{\text{leak}}$ back to $\mathcal{S}_{\text{RB}}$ is $2\epsilon_{\text{leak}}$. From symmetry, we know that when population leaks back from $\mathcal{S}_{\text{leak}}$ to $\mathcal{S}_{\text{RB}}$, it leaks equally to both states in $\mathcal{S}_{\text{RB}}$. The leakage component of the transfer matrix is then:

$$T_{\text{leak}} = \begin{bmatrix} -\epsilon_{\text{leak}} & 0 & \epsilon_{\text{leak}} \\ 0 & -\epsilon_{\text{leak}} & \epsilon_{\text{leak}} \\ \epsilon_{\text{leak}} & \epsilon_{\text{leak}} & -2\epsilon_{\text{leak}} \end{bmatrix}. \quad (23)$$

Together, this makes the total transfer matrix:

$$T = T_{\text{RB}} + T_{\text{leak}}. \quad (24)$$

As shown in the Supplementary Information, the general solution to this set of coupled differential equations is:

$$\begin{aligned} P_{\text{survival}} &= \frac{1}{3} + \frac{1}{2}(1 - 2\epsilon_{\text{RB}} - \epsilon_{\text{leak}})^l + \frac{1}{6}(1 - 3\epsilon_{\text{leak}})^l, \\ P_{\text{leak}} &= \frac{1}{3} - \frac{1}{3}(1 - 3\epsilon_{\text{leak}})^l, \\ P_{\text{flip}} &= \frac{1}{3} - \frac{1}{2}(1 - 2\epsilon_{\text{RB}} - \epsilon_{\text{leak}})^l + \frac{1}{6}(1 - 3\epsilon_{\text{leak}})^l. \end{aligned} \quad (25)$$

Eq. 25 is the general analytical expression used for fitting population decays in the absence of SPAM errors. When non-negligible SPAM errors are present, a more general expression derived in the Supplementary Information (Eq. S39) can be used.

1. Estimating average gate fidelities

Eq. 14 gives the average probability of generating target state $|g_{j,l+1}\rangle = \hat{C}_{l+1}|\psi_{j,l}\rangle$, given some noise source k and distribution of initial states $\mathcal{D} \equiv \{(|\psi_{j,l}\rangle, P_{j,l})\}$. Defining the error channel k as:

$$\mathcal{E}_k(|\psi\rangle) \equiv \sum_{m,m'} P_m \langle m' | \hat{U}_{e,k}(|\psi\rangle \langle \psi| \otimes |m\rangle \langle m|) \hat{U}_{e,k}^\dagger |m'\rangle, \quad (26)$$

encapsulates the dependence on the state of the external degrees of freedom. If $\mathcal{D}$ forms a two-design on $\mathcal{S}_{2Q}$, we can replace the sum over $\mathcal{D}$ in Eq. 14 with a Haar measure, after which the equation reduces to:

$$\mathcal{F}_{g,k} \equiv \int d\psi \langle \psi | \mathcal{E}_k(|\psi\rangle) | \psi \rangle, \quad (27)$$

equal to the average gate fidelity in the presence of noise mechanism k , averaged over $\mathcal{S}_{2Q}$. If we sum over $\mathcal{I}_{g,k} \equiv 1 - \mathcal{F}_{g,k}$ for every $k \in \mathcal{K}_{2Q}$, we get the total average gate infidelity:

$$\begin{aligned} \mathcal{I}_g &\simeq \sum_{k \in \mathcal{K}_{\text{RB}}} \langle \lambda_{\mathcal{Q}_k}^2 \rangle_{\mathcal{S}_{2Q}} |\alpha_k|^2 \sum_m P_m \langle m | \hat{M}_k^2 | m \rangle \\ &+ \sum_{k \in \mathcal{K}_{\text{leak}}} \langle \lambda_{\mathcal{Q}_k}^2 \rangle_{\mathcal{S}_{2Q}} |\alpha_k|^2 \sum_m P_m \langle m | \hat{M}_k^2 | m \rangle, \end{aligned} \quad (28)$$

where we have divided $\mathcal{K}$ into $\mathcal{K}_{\text{RB}}$ and $\mathcal{K}_{\text{leak}}$. Importantly, both terms are identical to their definitions in $\epsilon_{\text{RB(leak)}}$ —up to their dependence on $\mathcal{D}$, encapsulated entirely in the values of $\langle \lambda_{\mathcal{Q}_k}^2 \rangle_{\mathcal{D}}$. It follows that we can estimate the average fidelity over $\mathcal{S}_{2Q}$ by adding ϵ_{RB} and ϵ_{leak} , weighting them by the relative value of $\langle \lambda_{\mathcal{Q}_k}^2 \rangle$ during SLERB to the same value, averaged instead over $\mathcal{S}_{2Q}$. Crucially, every operator in $\mathcal{Q}_{\text{leak}}$ has the same $\langle \lambda_{\mathcal{Q}_{\text{leak}}}^2 \rangle_{\mathcal{S}_{\text{RB}}} = 2$ and the same $\langle \lambda_{\mathcal{Q}_{\text{leak}}}^2 \rangle_{\mathcal{S}_{2Q}} = 8/5$. To estimate how much $\mathcal{K}_{\text{leak}}$ contributes to the total error budget, we weight ϵ_{leak} by:

$$\frac{\langle \lambda_{\mathcal{Q}_{\text{leak}}}^2 \rangle_{\mathcal{S}_{2Q}}}{\langle \lambda_{\mathcal{Q}_{\text{leak}}}^2 \rangle_{\mathcal{S}_{\text{RB}}}} = \frac{4}{5}. \quad (29)$$

The symmetric Pauli operators in $\mathcal{K}_{\text{RB}}$, either have $\langle \lambda_{\mathcal{Q}_{\text{RB}}}^2 \rangle_{\mathcal{S}_{\text{RB}}} = 2/3$ or $\langle \lambda_{\mathcal{Q}_{\text{RB}}}^2 \rangle_{\mathcal{S}_{\text{RB}}} = 0$. A SLERB experiment only measures errors in the former category, all of which are isometric with the single-qubit Pauli operators on $\mathcal{S}_{\text{RB}}$. The operators in $\mathcal{Q}_{\text{RB}}$ that act as single-qubit $\hat{\sigma}_\phi$ both have the same $\mathcal{S}_{2Q}$ averaged variance $\langle \lambda_{\mathcal{Q}}^2 \rangle_{\mathcal{S}_{2Q}} = 4/5$, while $\frac{1}{2}(\hat{\sigma}_{z,1} + \hat{\sigma}_{z,2})|_{\mathcal{S}_{\text{RB}}} = \hat{\sigma}_z$ has an average variance of $\langle \lambda_{\frac{1}{2}(\hat{\sigma}_{z,1} + \hat{\sigma}_{z,2})}^2 \rangle_{\mathcal{S}_{2Q}} = 2/5$. To estimate the contribution of $\mathcal{K}_{\text{RB}}$ to $\mathcal{I}_g$, we take the larger of these two values $4/5$ and, again, weight ϵ_{RB} by ratio of the variance over $\mathcal{S}_{2Q}$ to the variance over $\mathcal{S}_{\text{RB}}$:

$$\frac{\langle \lambda_{\mathcal{Q}_{\text{RB}}}^2 \rangle_{\mathcal{S}_{2Q}}}{\langle \lambda_{\mathcal{Q}_{\text{RB}}}^2 \rangle_{\mathcal{S}_{\text{RB}}}} = \frac{6}{5}. \quad (30)$$

This choice means $\hat{Q}_{\text{RB},k} = \frac{1}{2}(\hat{\sigma}_{z,1} + \hat{\sigma}_{z,2})$ errors will be double-counted in SLERB, potentially leading to an over-estimation of $\mathcal{I}_g$. Together, this makes the SLERB-estimated average Clifford infidelity:

$$\mathcal{I}_{\text{Clifford}} = \frac{6}{5}\epsilon_{\text{RB}} + \frac{4}{5}\epsilon_{\text{leak}}, \quad (31)$$

or, dividing by 13/6—the average number of $\hat{U}_{2Q}$ per Clifford—we get:

$$\epsilon_{2Q} = \frac{6}{13} \mathcal{I}_{\text{Clifford}} \quad (32)$$

as our estimated average gate error.

C. Group-theoretical properties

We now consider process matrix representations of the benchmarking group $\mathcal{G}_{\text{RB}}$ generated by the sampling procedure. We derive the irreducible representations (irreps) of the benchmarking group, which correspond to invariant subspaces under the action of the group. This allows us to connect the effects of twirling over $\mathcal{G}_{\text{RB}}$ to the exponential decay of population in and between the irreps. Our experiment measures some of these decays and leaves others unmeasured. A *consistent* estimator [29] of gate fidelity would necessarily need to measure all these decays. Nevertheless, we average the measured decays to extrapolate the unmeasured decays, which produces an approximate fidelity estimator. This fidelity estimator makes different assumptions than the one in Eq. (31) yet yields remarkable agreement. Furthermore, our group-theoretical analysis reveals that some of the assumptions of the prior section are enforced by twirling over $\mathcal{G}_{\text{RB}}$.

Our protocol deploys a direct generalization of single-qubit RB with leakage on an invariant subspace of the MS gate. To establish the relationship, we begin by considering the unitaries that are generated by setting the MS phase to 0 and $\pi/4$:

$$\hat{M}_x \equiv \hat{U}_{2Q}(\frac{\pi}{2}, 0), \quad (33)$$

$$\hat{M}_y \equiv \hat{U}_{2Q}(\frac{\pi}{2}, \frac{\pi}{4}). \quad (34)$$

It can be shown (see Supplementary Information) that the target action of these gates is isomorphic to a block-diagonal decomposition of the action on the even-parity subspace $\mathcal{S}_{\text{RB}}$ and the odd-parity subspace $\mathcal{S}_{\text{leak}}$

$$\hat{M}_x \cong R_x^{00,11}(\frac{\pi}{2}) \oplus R_x^{01,10}(\frac{\pi}{2}), \quad (35)$$

$$\hat{M}_y \cong R_y^{00,11}(\frac{\pi}{2}) \oplus R_y^{01,10}(\frac{\pi}{2}), \quad (36)$$

where $R_x^{a,b}(\theta_{\text{MS}})$ and $R_y^{a,b}(\theta_{\text{MS}})$ are generalized SU(2) rotations about x and y over the subspace spanned by the states $|a\rangle, |b\rangle$. The benchmarking group generated by $\hat{M}_x$ and $\hat{M}_y$ acts on the even-parity subspace $\mathcal{S}_{\text{RB}}$ in the same way that the x and y Pauli rotations of single-qubit RB act on the associated qubit subspace. Thus, our gate set forms a two-design over this subspace, and any error sources that preserve $\mathcal{S}_{\text{RB}}$ and do not commute with the benchmarking group are completely depolarized on this subspace. However, the action of the benchmarking group on the odd-parity subspace $\mathcal{S}_{\text{leak}}$ is restricted to rotations about the x -axis; this does not completely depolarize noise on this subspace and complicates the irrep structure of the representation of the benchmarking group.

1. Decay model

To derive the expected decay forms for the benchmarking protocol, we begin by identifying the benchmarking group generated by $\hat{M}_x$ and $\hat{M}_y$

$$\mathcal{G}_{\text{RB}} = \langle \hat{M}_x, \hat{M}_y \rangle. \quad (37)$$

This group has 96 elements, and as can be seen from the character table, Tab. II, is not a separable product group of uncoupled action on $\mathcal{S}_{\text{RB}}$ and $\mathcal{S}_{\text{leak}}$. We consider a representation V of the group $\mathcal{G}_{\text{RB}}$ as process matrices acting on the Hilbert-Schmidt space of two-qubit vectorized density matrices.

Now, consider the twirl $\tilde{\Lambda}_{\mathcal{G}}$ of a noise channel Λ over the group $\mathcal{G}_{\text{RB}}$:

$$\tilde{\Lambda}_{\mathcal{G}} = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}_{\text{RB}}} V(g) \Lambda V(g)^{-1}. \quad (38)$$

Schur's lemma implies that the action of the twirled channel is isomorphic to a direct sum over actions on irreps of the representation [30]

$$\tilde{\Lambda}_{\mathcal{G}} \cong \bigoplus_{\nu} Q_{\nu} \otimes \mathbb{1}_{\nu}, \quad (39)$$

where ν indexes the irreps. To derive the expected decay signal for the benchmarking experiment, one must calculate the overlaps between the irreps and the given state preparation and measurement pair used in the experiment. For an initial state preparation of $|00\rangle$ or $|11\rangle$, the only irreps that have non-zero overlap are the trivial irrep and the irrep corresponding to depolarization on $\mathcal{S}_{\text{RB}}$, see Tab. V in the Supplementary Information; here, we show that the generic decay form of the success probability in the absence of SPAM errors consists of a constant offset and three exponential terms

$$P_{\text{survival}} = c_1 + c_2 q_{\text{leak},+}^l + c_3 q_{\text{leak},-}^l + \frac{1}{2} q_{\text{RB}}^l. \quad (40)$$

One exponential q_{RB}^l corresponds to depolarization in $\mathcal{S}_{\text{RB}}$, while the other two correspond to leakage between this subspace and $\mathcal{S}_{\text{leak}} = \mathcal{S}_{\text{leak},+} \cup \mathcal{S}_{\text{leak},-}$. This is the most specific model one can derive without assumptions about the noise in the system. Assuming, as in Sec. II B, that the errors are symmetric, we can derive the expected decay model (see the Supplementary Information):

$$P_{\text{survival}} = \frac{1}{3} + \frac{1}{2} q_{\text{RB}}^m + \frac{1}{6} q_{\text{leak},+}^m, \quad (41)$$

and similarly

$$P_{\text{flip}} = \frac{1}{3} - \frac{1}{2} q_{\text{RB}}^m + \frac{1}{6} q_{\text{leak},+}^l, \quad (42)$$

where q_{RB} is a decay term corresponding to errors in $\mathcal{S}_{\text{RB}}$ and $q_{\text{leak},+}$ is a decay term corresponding to leakage between $\mathcal{S}_{\text{RB}}$ and $\mathcal{S}_{\text{leak},+}$. This decay form is essentially the same as Eq. (25), derived from the transfer matrix model.

2. Extended fidelity estimator

As discussed in the Supplementary Information, the effective fidelity can be calculated given the knowledge of all subspace decays $q_{\nu,\mu}$. However, as shown in the previous section, SLERB only directly measures q_{RB} and $q_{leak,\pm}$. Our approach is therefore to report the decays on unmeasured subspaces by *extending* the measured decays to approximate the decays on unmeasured subspaces. One way to do this is by setting $q_{leak,+} = q_{leak,-}$ (pessimistic for symmetric errors), and averaging the measured decays to approximate the unmeasured ones. While not applicable to arbitrary adversarial noise, our analysis indicates these are conservative approximations for the anticipated noise sources, and numerical simulations in Sec. III show that the estimator closely approximates the true error rate for random unitary error channels. As shown in the Supplementary Information, setting all unmeasured terms to the uniform average of measured terms results in a process fidelity metric

$$\mathcal{F} = \frac{1 + 8q_{RB} + 7q_{leak,+}}{16}, \quad (43)$$

which can be converted to an average Clifford gate fidelity [31] and reported as

$$\bar{\mathcal{F}} = \frac{5 + 8q_{RB} + 7q_{leak,+}}{20}. \quad (44)$$

We can further convert to an expression for infidelity in terms of ϵ_{RB} and ϵ_{leak}

$$\bar{\mathcal{I}}_{\text{Clifford}} = \frac{4}{5}\epsilon_{RB} + \frac{29}{20}\epsilon_{leak}. \quad (45)$$

In practice, the Cliffords are compiled from native gate sequences. Thus, an “error per gate” metric would rescale the extended Clifford infidelity by the average number of gates per Clifford

$$\epsilon_{2Q} = \frac{6}{13}\bar{\mathcal{I}}_{\text{Clifford}}. \quad (46)$$

This estimator of Eq. (44) should be compared with Eq. (31). While these two estimators are not exactly equivalent, the small discrepancy between the two can be explained by differing assumptions. Eq. (31) is based on the assumptions of 1) incoherent population transfer, 2) leakage only within the symmetric subspace, 3) equal leakage and seepage rates, and 4) small errors. In contrast, Eq. (44) is derived by analyzing the decays on irreps that are enforced by twirling over $\mathcal{G}_{RB}$, and extending the measured decays to approximate unmeasured decays. Its assumptions are 1) twirling over $\mathcal{G}_{RB}$, 2) independent noise on each Clifford, 3) approximately unbiased noise. If one were to take a weighted, instead of uniform, average between the measured decays, then Eq. 31 could be recovered for a certain weighting.

The analysis in this section has shown that twirling over the benchmarking group $\mathcal{G}_{RB}$ enforces some of the

assumptions made in the prior section. Namely, we have shown that twirling over $\mathcal{G}_{RB}$ enforces the assumption of incoherent population transfer within $\mathcal{S}_{RB}$ and to/from $\mathcal{S}_{RB}$ and $\mathcal{S}_{leak}$. The analysis has also shown that twirling over $\mathcal{G}_{RB}$ combined with assumptions of leakage only in the symmetric space and equal leakage/seepage rates enforces the decay forms of Eq. 25.

Our derivations in this section also provide the necessary framework to construct a consistent estimator of gate fidelity with extensions to the current protocol. As we have said, we measure only two of the relevant decays required to construct a consistent fidelity estimate. One can extend the current protocol to measure all the subspace decays by including more state preparation and measurement pairs in the protocol. For example, if one were to prepare in $|01\rangle$ or $|10\rangle$, then the resulting RB signal would be sensitive to error processes that cause transfer between $\mathcal{S}_{leak,+}$ and $\mathcal{S}_{leak,-}$. In this case, the signal may contain oscillations due to residual coherence from partial twirling, but the theory can still be applied. Similarly, preparing superposition states between $\mathcal{S}_{leak}$ and $\mathcal{S}_{RB}$ allows us to measure decays associated with forward and backward mixing processes (Tab. III). Employing the mechanisms of character RB [9], filtered RB [22], or synthetic SPAM RB [32] would provide a general framework to analyze the resulting signals and extract a consistent estimator of $\mathcal{F}$.

D. MS gate error symmetry

While SLERB is a highly general technique, the analytical expressions for populations derived in the previous sections are based on the assumption that errors keep the populations within the symmetric subspace $\{|00\rangle, |11\rangle, \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)\}$. This assumption is well motivated by the physics of MS interactions; as each qubit is coupled to the same motional mode, the populations stay within the symmetric subspace even if each qubit couples to the MS drive field with a different strength. As a result, only a handful of physically plausible error channels have the potential to populate the antisymmetric subspace, notably photon scattering and differential qubit frequency shifts. We expect both of these to be negligible for the MS gate implementation in Sec. IV, as the gate involves no laser fields (eliminating photon scattering), and uses dynamical decoupling (suppressing qubit frequency shifts). In Sec. IIIB, we verify this reasoning through direct numerical simulations of SLERB sequences with additional error sources.

Finally, we note that SLERB offers two natural ways of verifying the symmetry assumption. The first is through the measurement of population asymptotes $\vec{P}_{l \rightarrow \infty}$. Eq. 25 shows that for $l \rightarrow \infty$, if $\epsilon_{leak} > 0$, the populations will reach an equilibrium state of $\vec{P}_{l \rightarrow \infty} \rightarrow (1/3, 1/3, 1/3)$. On the other hand, errors that drive population to $\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ will result in

$\vec{P}_{l \rightarrow \infty} \rightarrow (1/4, 1/4, 1/2)$. Thus, the presence of any such error would result in a measurable change in the asymptotes. Second, SLERB can measure such errors directly by modifying the sequence to act on the $|01\rangle, |10\rangle$ subspace instead. However, this subspace would require asymmetric MS interactions where ϕ_{MS} is different for each ion, see Supplementary Information.

III. NUMERICAL SIMULATIONS

We have performed extensive simulations of the protocol to verify the accuracy and precision of our fidelity estimator. These simulations have been written at three levels of abstraction: 1) state vector simulation of Clifford sequences with unitary errors acting on pure initial states, 2) direct simulation of the physical noise on the full MS dynamics that include both qubit and motional degrees of freedom, and 3) direct calculation of the twirled channel, which represents the infinite shot and randomization limit. Additionally, we use the full MS simulations to investigate the effect of common MS errors on the population dynamics for both the symmetric and antisymmetric subspaces.

A. Population transfer rates

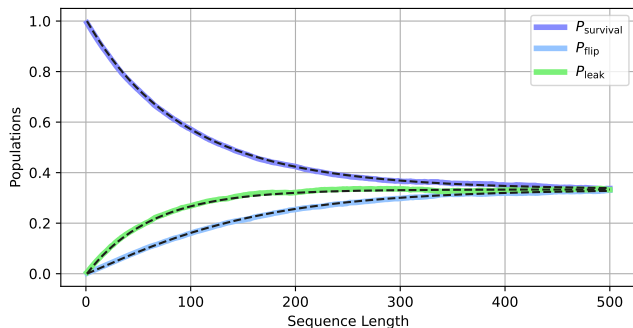


FIG. 2. Monte Carlo simulations (colored solid lines) of population transfer for a SLERB sequence of length $l = 500$, averaging over 5×10^3 random circuits of single-qubit Cliffords on $\mathcal{S}_{\text{RB}}$, all of which decompose into Eq. (5). For this run, we multiplied each Clifford by error unitaries representing two noise unitaries: one in $\mathcal{K}_{\text{RB}}$ and one in $\mathcal{K}_{\text{leak}}$. Using the results of Sec. II B, we calculate ϵ_{RB} and ϵ_{leak} analytically, then apply the resultant transfer matrix T to the initial population vector $\vec{P}_0 = (1, 0, 0)$ (black dashed lines). The result illustrates how the analytical expressions agree with numerical simulations.

To verify the analytical transfer matrix results in Sec. II B, we use a Monte Carlo simulation that averages over random $\mathcal{S}_{\text{RB}}$ single-qubit Cliffords. Each Clifford is generated by numerically iterating through decompositions of Eq. (5) until we obtain 24 unique single-qubit

Cliffords on $\mathcal{S}_{\text{RB}}$. We apply two error unitaries:

$$\hat{U}_{\text{RB},e} = e^{-i\alpha_{\text{RB}}\hat{\sigma}_{x,1}\hat{\sigma}_{x,2}}, \quad (47)$$

as an example of an error that conserves $\mathcal{S}_{\text{RB}}$, and:

$$\hat{U}_{\text{leak},e} = e^{-i\alpha_{\text{leak}}(\hat{\sigma}_{x,1} + \hat{\sigma}_{x,2})}, \quad (48)$$

as a leakage error; to ensure our assumption that leakage is incoherent, we apply a random sign to $\pm\alpha_{\text{leak}}$ at every occurrence of $\hat{U}_{\text{leak},e}$. So if $\hat{C}_l$ is the l^{th} Clifford of a sequence, the unitary applied in the simulation is:

$$\hat{U}_{t,l} = \hat{U}_{2q,e}\hat{U}_{\text{leak},e}\hat{C}_l. \quad (49)$$

The unitary $\hat{U}_{\text{RB},e}$ represents a noise source in $\mathcal{K}_{\text{RB}}$, and the resulting population transfer rate within $\mathcal{S}_{\text{RB}}$ is:

$$\epsilon_{\text{RB}} = \frac{2}{3}|\alpha_{\text{RB}}|^2, \quad (50)$$

as discussed in Sec. II B. The unitary $\hat{U}_{\text{leak},e}$ is in $\mathcal{K}_{\text{leak}}$, and gives a leakage rate:

$$\epsilon_{\text{leak}} = 2|\alpha_{\text{leak}}|^2. \quad (51)$$

For the example in Fig. 2, we set $\alpha_{\text{RB}} = \alpha_{\text{leak}} = \pi/60$. Plugging these into our transfer matrix T , we show the results from our Monte-Carlo simulation for systems initialized to $\vec{P}_0 = (1, 0, 0)$ and averaged over 5×10^3 random trials of circuits up to length $l = 500$. We compare this to the transfer matrix model $\vec{P}_l = T^l \vec{P}_0$. The figure demonstrates that the two calculations converge, with no fitting parameters.

B. Full MS dynamics

We also perform full Hamiltonian-level simulations of SLERB sequences in QuTiP [33]. For each MS gate, we model the interaction as

$$\hat{H}_{\text{MS}} = \frac{\hbar\Omega_{\text{MS}}}{2} \sum_{n=1,2} [\hat{\sigma}_{x,n}\cos\phi_{\text{MS}} + \hat{\sigma}_{y,n}\sin\phi_{\text{MS}}] (\hat{a}e^{i\delta t} + \hat{a}^\dagger e^{-i\delta t}), \quad (52)$$

where $\hat{\sigma}_{i,n}$ is the Pauli operator for qubit n . The strength of the spin-motion coupling is described by Ω_{MS} , and δ is the detuning of the interaction from the motional mode of the trapped ions, described by the annihilation and creation operators $\hat{a}$ and $\hat{a}^\dagger$, respectively. We assume the ions' spins only couple to a single mode of motion; the initial state is $|00\rangle$, and the ground state of the motional mode. We model a two-loop Walsh-modulated gate by changing ϕ_{MS} by π halfway through the gate [34]. We set $\Omega_{\text{MS}}/2\pi = 1$, and thus $t_g = 1/\sqrt{2}$, and $\delta/2\pi = 2\sqrt{2}$. We generate the SLERB sequences in the exact same way as the experimental implementation (Sec. IV), using 50 random circuits for each sequence length.

We first investigate the effect of the gate Rabi frequency offset by setting $\Omega_{\text{MS}}/2\pi = (1 + 0.05)$. Fig. 3a) shows that, as expected for an error that only affects the geometric phase, there is no leakage. Thus, $P_{\text{leak}} = 0$, and the other two populations have an asymptote at $1/2$. Fig. 3b) instead presents simulation of a motional mode frequency offset $\delta/2\pi = (1 - 0.07) \times 2\sqrt{2}$. In contrast to the Rabi frequency offset, this error results in leakage as there is residual spin-motion entanglement at the end of each gate. Here, the asymptotes are instead $1/3$ for all three populations, as all states within the symmetric subspace are populated equally. We also observe the multi-exponential decay as we expect from Eq. 25. In the simulations, we reset the mode to the ground state to prevent any coherent buildup of errors. We discuss in the Supplementary Information the behavior when the motional mode is not reset.

Additionally, we perform individual simulations of common error mechanisms in laser-free MS gates to verify the error symmetry assumption in Sec. IID. The results are summarized in Table I. We find that all global errors – i.e., when both qubits experience an identical offset in a control parameter – keep populations within the symmetric subspace. Further, errors that only effect the geometric phase, but do not cause any residual spin-motion entanglement, keep populations within the $|00\rangle, |11\rangle$ subspace. This behaviour holds even when each ion has a different spin-motion coupling strength. Motional errors, such as mode frequency offsets, motional dephasing, or motional heating, can additionally populate the $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ state but still keep populations within the symmetric subspace. One exception is a differential qubit frequency shift, which populates the antisymmetric state $\frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$, but only if there is leakage to the $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ to begin with, making it a second-order effect. Similarly, if an additional carrier field is used for dynamical decoupling (see Sec. IV), a differential carrier Rabi frequency only results in population to the singlet state if the carrier does not commute with another interaction, such as the spin-dependent force or a global qubit frequency offset.

C. Random unitary error channels

Finally, we verify that our protocol performs well for random unitary noise even without the assumption of symmetric noise. These simulations were performed via direct calculation of the twirled process matrix associated with a unitary error. Explicitly, we construct a random unitary error $\hat{E}$ via

$$\hat{E} = \exp\left(i\vec{\theta} \cdot \mathbf{P}\right),$$

where $\vec{\theta}$ is a random variable and $\mathbf{P}$ is a basis of two-qubit Pauli matrices. We choose the distribution of $\vec{\theta}$ to be a multivariate Gaussian with uniform variance equal

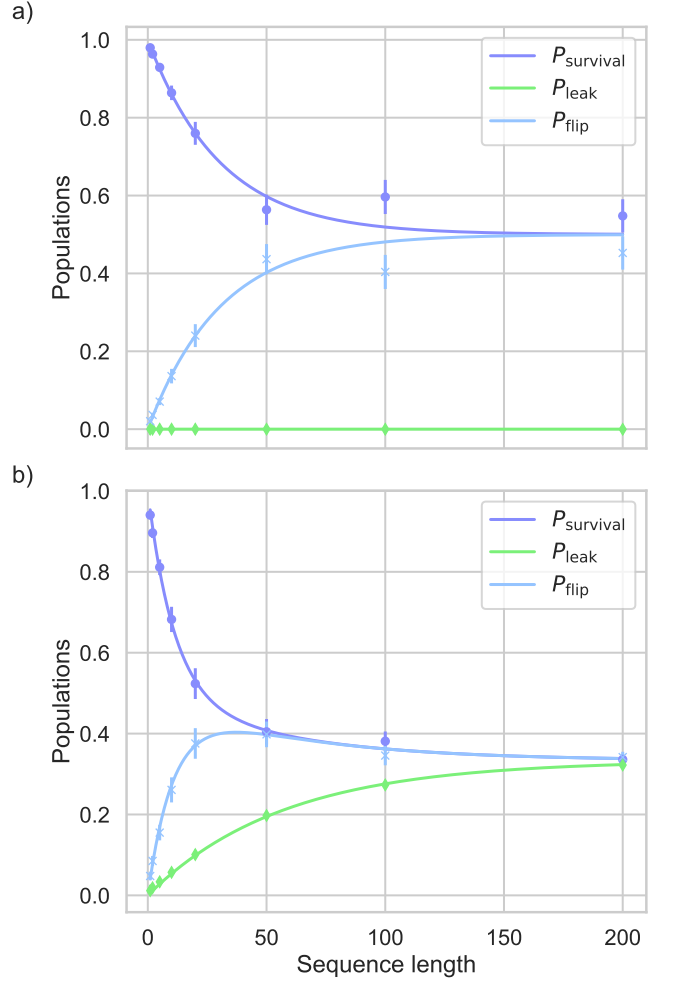


FIG. 3. Numerical simulations of SLERB sequences for (a) amplitude and (b) detuning errors. For each sequence length, we sample 50 random sequences; the error bars correspond to the standard deviation for each population from the 50 circuits. The solid lines correspond to fits to the data following Eq. 25. (a) For amplitude errors, there is no leakage and the states stay within the $|00\rangle, |11\rangle$ subspace, with asymptotes of $1/2$ for P_{survival} and P_{leak} . (b) For detuning errors, there is residual spin-motion entanglement, which results in leakage. The asymptotes are instead $1/3$ for all three populations.

to 0.01. This choice of distribution results in an expected fidelity in the range observed experimentally in Sec. IV. We then convert each unitary error into a process matrix representation $\Lambda = \hat{E}^* \otimes \hat{E}$ and calculate the twirled channel $\hat{\Lambda}_{\mathcal{G}}$ via Eq. (38), where $\mathcal{G}_{\text{RB}}$ is the entire benchmarking group of 96 elements. Calculation of the twirled channel requires averaging over the entire benchmarking group rather than a subset of 24 elements. In this way, we directly simulate the theoretical limit in the case of infinite shots and all randomizations. We then calculate P_{survival} , P_{flip} , and P_{leak} for varying l . Finally, we fit the resulting exponential decays to the model

$$P_{\text{survival}} + P_{\text{flip}} = A + Bq_{\text{leak}}^l, + \quad (53)$$

Errors	$ 00\rangle \updownarrow 11\rangle$	$ 00\rangle, 11\rangle \updownarrow 01\rangle + 10\rangle$	$ 00\rangle, 11\rangle \updownarrow 01\rangle - 10\rangle$	$ 01\rangle + 10\rangle \updownarrow 01\rangle - 10\rangle$
	$ 11\rangle$	$ 01\rangle + 10\rangle$	$ 01\rangle - 10\rangle$	$ 01\rangle - 10\rangle$
Global qubit frequency offset	✓	✓	✗	✗
Motional errors	✓	✓	✗	✗
Global Ω_{MS} offset	✓	✗	✗	✗
Differential Ω_{MS}	✓	✗	✗	✗
Global carrier phase offset, identical carrier Rabi frequencies	✓	✓	✗	✗
Differential qubit frequency shift	✓	✗	✗	✓
Global carrier phase offset, differential carrier Rabi frequencies	✓	✓	✓	✓

TABLE I. Population transfer dynamics of different laser-free MS gate errors. For each error, we mark the allowed population transfers with ✓, and those not allowed with ✗. In general, geometric phase errors keep the states within the $\{|00\rangle, |11\rangle\}$ subspace, while any motional errors result in populations of the $|01\rangle + |10\rangle$ state as well. Antisymmetric errors can result in populations in the $|10\rangle - |10\rangle$ state. However, the differential qubit frequency shifts only result in populations in the antisymmetric state if there is some leakage to begin with. In the absence of any leakage, differential qubit frequency shifts keep the populations within the $|00\rangle, |11\rangle$ subspace. Similarly, differential carrier Rabi frequencies only populate the antisymmetric state when the carrier does not commute with the gate, which causes leakage.

and

$$P_{\text{survival}} - P_{\text{flip}} = Cq_{\text{RB}}^l. \quad (54)$$

Note that the unitary simulations in this subsection do not respect the symmetric leakage assumption that has been made in the quantitative analysis in Sec. II and validated experimentally in Sec. IV. Hence, we would expect to see a second exponential term in $P_{\text{survival}} + P_{\text{flip}}$. However, we find that the symmetric model still provides a good estimate of gate fidelity, i.e., that fitting a single-exponential model to a double exponential decay still captures the dominant decay profile, resulting in relatively small approximation error.

We finally estimate the fidelity using Eqs. 31 and 44. We repeated this protocol 1000 times. The results are plotted in Fig. 4. We further calculate the relative error $e_{\mathcal{F}}$ between the true channel fidelity $\tilde{\mathcal{F}}$ and the estimated fidelity $\hat{\mathcal{F}}$:

$$e_{\mathcal{F}} = \frac{|\hat{\mathcal{F}} - \tilde{\mathcal{F}}|}{\tilde{\mathcal{F}}}. \quad (55)$$

We find that, for random unitary error channels in the simulated fidelity regime, the average relative error between each estimator and the truth is $16 \pm 12\%$ for the

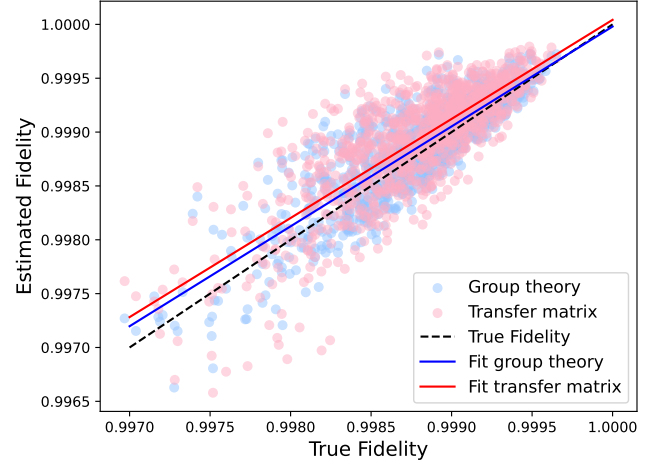


FIG. 4. Estimated vs true channel fidelity for 1000 random unitary error channels. Each point corresponds to a full simulation of a benchmarking experiment where a random unitary error channel was sampled, the numeric twirl of the resulting channel representation was calculated, and the signals were calculated and fit to extract an estimate of the fidelity. Transfer matrix points correspond to the fidelity estimator of Eq. 31, and group theory points correspond to Eq. 44. The lines represent fits of the two estimated distributions to a linear model.

group theory estimator and $22 \pm 15\%$ for the transfer matrix estimator.

IV. EXPERIMENTAL IMPLEMENTATION

We implement SLERB in a cryogenic ion trap setup similar to that described in Ref. [14]. Qubits are encoded in the $4S_{1/2}$ Zeeman sublevels of $^{40}\text{Ca}^+$ ions, each with a qubit frequency of $\omega_0 \approx 2\pi \times 240$ MHz. The MS gates are performed on the in-plane out-of-phase radial mode of a two-ion crystal at $\omega_m \approx 2\pi \times 3.5$ MHz by passing a bichromatic oscillating current $\omega_0 \pm (\omega_m + \delta)$ through an on-chip antenna which generates magnetic field gradients [35–37]. An additional carrier tone at ω_0 dynamically decouples from errors due to qubit frequency fluctuations [38].

The pulse sequence for an MS gate with phase ϕ_{MS} is shown in Fig. 5a). The sideband tones lead to a gate interaction with strength $\Omega_{\text{MS}} \approx 2\pi \times 6$ kHz, detuning $\delta \approx 2\pi \times 17$ kHz, and on/off ramp duration of ≈ 3 μs . The phase ϕ_{MS} of these tones is incremented by π halfway through the gate to achieve Walsh-1 modulation [34], reducing errors due to offsets in the gate mode frequency. The carrier tone has Rabi frequency $\Omega_c \approx 2\pi \times 40$ kHz and on/off ramp duration of ≈ 0.5 μs . The phase of the carrier tone is updated by π halfway through the gate to ensure minimal residual carrier rotation by the end of the gate. The total duration of the gate is 124 μs .

The experiment is implemented as follows. At the start

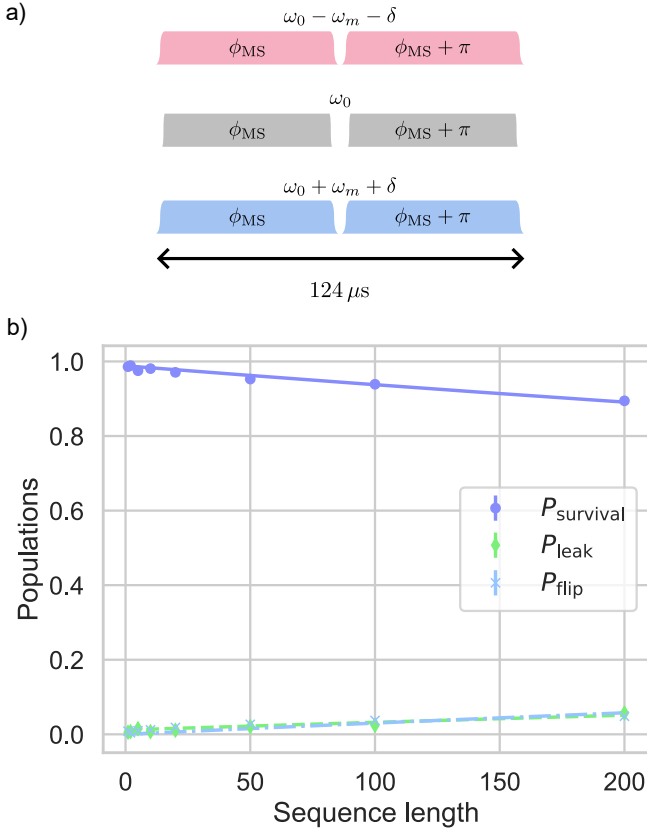


FIG. 5. SLERB implementation and results. (a) Pulse sequence for an individual MS gate with phase ϕ_{MS} . Each MS gate consists of three frequencies: blue and red sideband tones at $\omega_0 \pm (\omega_m + \delta)$, and a carrier at ω_0 for dynamical decoupling. We set the phases of all the tones to ϕ_{MS} in the first half of the gate, and to $\phi_{\text{MS}} + \pi$ in the second half of the gate. (b) SLERB results with up to 200 Cliffords. The populations P_{survival} (purple circles), P_{flip} (blue crosses), and P_{leak} (green diamonds) indicate the populations in the target state, in the incorrect state within the $\{|00\rangle, |11\rangle\}$ subspace, and the population that has leaked out of this subspace into the $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$ state, respectively. The measurements are performed with 50 randomizations and 50 shots each; error bars indicate the standard deviation of the populations across all the randomizations for each sequence length.

of every shot, the ions are initialized in $|00\rangle$ and cooled to close to the ground state in all 6 modes of motion, with an average phonon occupation $\bar{n} = 0.05(1)$ in the gate mode. We then perform a SLERB sequence, followed by a measurement in the computational basis. The target state of the sequence, either $|00\rangle$ or $|11\rangle$ with Pauli-frame randomisation, is assigned to P_{survival} , while the population in the other state within this subspace is P_{flip} . The population P_{leak} corresponds to measurements of either $|01\rangle$ or $|10\rangle$; our detection does not distinguish between $|01\rangle + |10\rangle$ or $|01\rangle - |10\rangle$.

Figure 5(b) shows the results for a SLERB sequence with up to 200 Cliffords (on average 433 two-qubit gates). We fit the data to Eq. S39, which modifies Eq. 25

to include SPAM errors, finding $\epsilon_{\text{RB}} = 3.2(3) \times 10^{-4}$, $\epsilon_{\text{leak}} = 2.2(3) \times 10^{-4}$, $\bar{\epsilon}_{\text{SPAM}} = 5.9(6) \times 10^{-3}$ and finally a two-qubit gate error $\epsilon_{2\text{Q}} = 2.6(2) \times 10^{-4}$, following Eq. 31. Using Eq. 45 instead, we obtain $\epsilon_{2\text{Q}} = 2.7(2) \times 10^{-4}$, which is statistically indistinguishable. The error bars correspond to the 68% confidence interval from non-parametric bootstrapping with 10,000 samples. These results are consistent with the error obtained by our team in Ref. [14] using partial state tomography.

A. Calibrations

Due to its ability to rapidly and directly estimate the magnitude and nature of errors, SLERB is an excellent tool for high-fidelity MS gate calibration. Figure 6 shows how SLERB can be used to find the optimal value of detuning δ and gate Rabi frequency Ω_{MS} . In each experiment, we apply a SLERB sequence of length $l = 100$ and record populations as before, with peaks of P_{survival} approximately indicating a local fidelity maximum. Following an initial coarse calibration, the high-fidelity MS gate in Fig. 5 was fine-tuned through such a series of single-parameter scans, with gradually increasing length to reduce statistical uncertainty.

These calibrations also demonstrate how SLERB enables the identification of errors and independent optimization of gate parameters. With the amplitude calibration (Fig. 6(b)), the main effect of miscalibrating Ω_{MS} is a geometric phase error. Ideally, the loop closure should be unaffected as the amplitude changes and hence should not have any effect on the leakage. Indeed, for these data, P_{leak} is roughly constant, and P_{survival} mainly decreases due to an increase in the P_{flip} . Small offsets in detuning δ are mainly expected to cause geometric phase errors (spin-motion entanglement errors are first-order suppressed by Walsh modulation, thus only appear at larger offsets). This matches the experimental results in Fig 6(a), where P_{leak} is approximately constant over ~ 0.5 kHz, and then increases as expected as δ becomes less negative.

B. Verifying asymptotes

As discussed in Sec. IID, the error symmetry assumption can be verified by measurements of population asymptotes. In Sec. IIIB, we have confirmed numerically that, in the presence of detuning errors, the populations should asymptote to $1/3$. We now confirm this experimentally by performing SLERB with δ intentionally offset by 2 kHz. The results are shown in Fig. 7. As expected, we observe that all three populations decay to an asymptote of $1/3$, confirming there no significant pumping channels into the antisymmetric state. We have also estimated through direct spectroscopy that differential qubit frequency shifts – the main potential source of

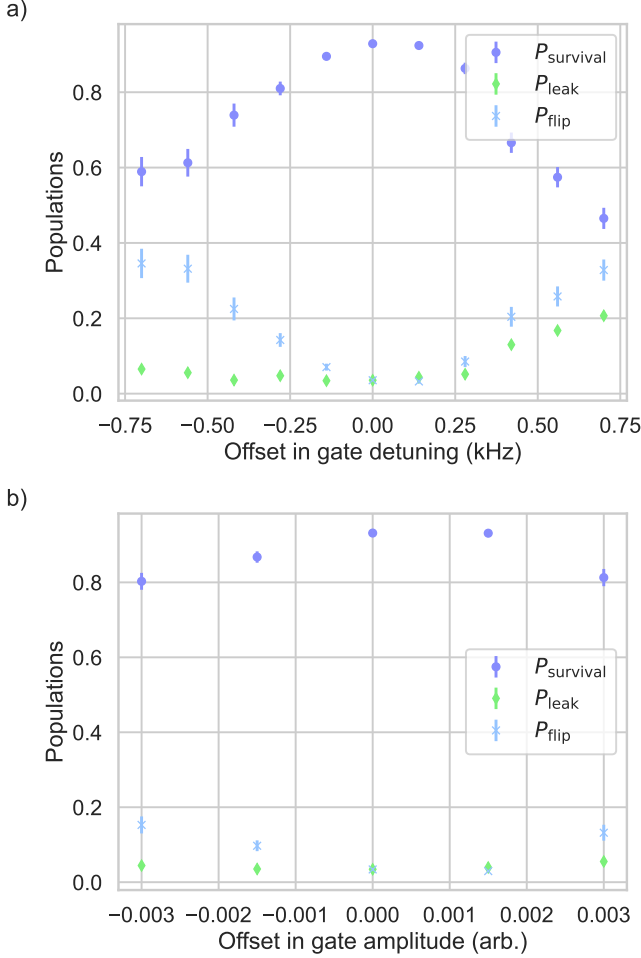


FIG. 6. Calibration data for the (a) gate detuning (δ) and the (b) gate amplitude (Ω_{MS}). For each of these datasets, we plot the populations versus the variation of the parameter for a fixed sequence length of 100 Cliffords with 50 randomizations and 50 shots each. Error bars indicate the 68% confidence interval for the average populations across the different randomizations. (a) As we vary the detuning of the MS interaction from the motional mode, we pick the detuning with the lowest value of P_{leak} to minimize any residual spin-motion entanglement. (b) As we vary the amplitude of the interaction, the leakage population P_{leak} is roughly constant; we set the amplitude to maximize P_{survival} .

antisymmetric errors – are smaller than 20 Hz, and thus should cause negligible gate errors ($< 10^{-5}$). Thus, the symmetric error assumption is validated both theoretically and experimentally in our system.

C. Applicability to laser-based MS gates

So far, we have focused on the use of SLERB for calibrating and benchmarking laser-free MS gates. However, SLERB can be performed with laser-based gates as well, as long as two considerations are kept in mind.

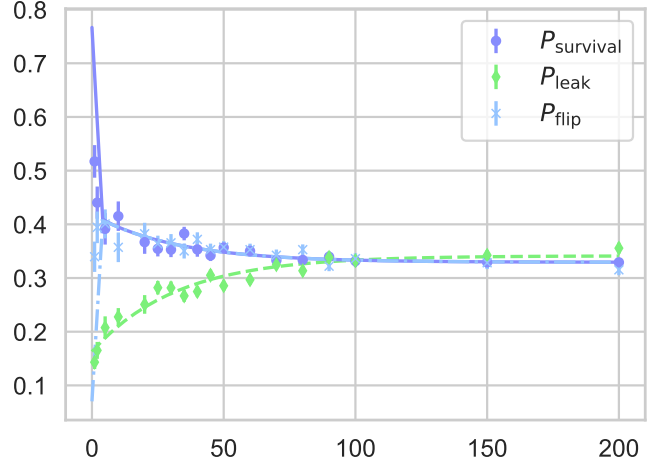


FIG. 7. SLERB sequence with a detuning error of 2 kHz. With this detuning offset, we observe leakage as there is residual spin-motion entanglement at the end of every gate. We fit the general decay model following Eq. 3, and find that the asymptote for P_{leak} is 0.341(9). The asymptotes for P_{survival} and P_{flip} are 0.329(5). These values are consistent with an asymptote of $1/3$ which corresponds to symmetric errors.

First, SLERB relies on the ability to perform many sequential two-qubit gates with phase coherence of the MS phase ϕ_{MS} . While this requirement is typically straightforward for electronic gates, the same level of coherence with laser-based approaches might be challenging. For example, if the spin phase changes with fluctuations of the optical phase, longer sequences might result in much larger errors. For two-photon stimulated Raman transitions [39, 40] at least, such fluctuations could be eliminated using phase-insensitive configurations that result in the spin phase being independent of the laser phase [41]. This configuration would result in the randomization of the motional phase of the MS interactions, which corresponds to the difference phase between the blue and red sidebands, but this phase has a negligible effect on the population decays, see Supplementary Information. Second, laser-based gates suffer from photon scattering or spontaneous emission, which can directly populate the antisymmetric subspace. Thus, additional experiments must be performed to estimate the effect of qubit decay and photon scattering on the total MS gate error (Sec. IID).

V. CONCLUSION

In conclusion, we have developed and demonstrated SLERB: a new technique for benchmarking two-qubit trapped-ion MS gates. We have validated its ability to estimate the magnitude and origin of MS gate errors through theoretical analysis and numerical simulations. Finally, we have demonstrated the use of SLERB for calibrating and characterizing high-fidelity laser-free MS

gates with low statistical uncertainty. SLERB's unique selling point is that it characterizes MS gates using only MS gates. This ability to measure errors simply and directly makes it highly valuable for large-scale QCs, which will benefit from low-complexity means of measuring component-level errors. We also anticipate that SLERB can be broadly applied to other QC platforms such as superconducting qubits, as long as they are able to implement $\hat{\sigma}_\phi \hat{\sigma}_\phi$ interactions. It can also be extended to $\hat{\sigma}_z \hat{\sigma}_z$ gates [42, 43], albeit with the inclusion of single-qubit rotations. The formalism developed here can also be adapted to analyze single-qubit logical randomized benchmarking over logical states $|\bar{0}\rangle$ and $|\bar{1}\rangle$ states as opposed to the physical ones considered here. The SLERB theory could be extended to more complex qubit states by incorporating leakage out of the qubit subspace alto-

gether [44]. Finally, the use of a single two-level subspace might also be beneficial for benchmarking multi-qubit interactions for more than two qubits at a time.

ACKNOWLEDGMENTS

We thank members of the Quantum Performance Laboratory, especially Jordan Hines and Yale Fan, for helpful discussions and feedback on the group theory analysis. We thank Jason Saied for insight into the basis for the irreps that appears in the supplement [45]. We thank Alex Kwiatkowski for insightful discussions. We thank Bryce Bjork for help setting up GAP code for the group-theoretical calculations. We thank the entire combined team at IonQ for their contributions to this work.

-
- [1] J. Preskill, Quantum computing in the NISQ era and beyond, *Quantum* **2**, 79 (2018).
 - [2] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, *Phys. Rev. A* **52**, R2493 (1995).
 - [3] A. M. Steane, Error correcting codes in quantum theory, *Phys. Rev. Lett.* **77**, 793 (1996).
 - [4] A. R. Calderbank and P. W. Shor, Good quantum error-correcting codes exist, *Phys. Rev. A* **54**, 1098 (1996).
 - [5] D. Gottesman, *Stabilizer codes and quantum error correction* (California Institute of Technology, 1997).
 - [6] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, *Phys. Rev. A* **77**, 012307 (2008).
 - [7] C. A. Ryan, M. Laforest, and R. Laflamme, Randomized benchmarking of single-and multi-qubit control in liquid-state nmr quantum information processing, *New J. Phys.* **11**, 013034 (2009).
 - [8] J. P. Gaebler, A. M. Meier, T. R. Tan, R. Bowler, Y. Lin, D. Hanneke, J. D. Jost, J. P. Home, E. Knill, D. Leibfried, and D. J. Wineland, Randomized Benchmarking of Multiqubit Gates, *Phys. Rev. Lett.* **108**, 260503 (2012).
 - [9] J. Helsen, X. Xue, L. M. K. Vandersypen, and S. Wehner, A new class of efficient randomized benchmarking protocols, *Npj Quantum Inf.* **5**, 71 (2019).
 - [10] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).
 - [11] T. Proctor, S. Seritan, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, Scalable Randomized Benchmarking of Quantum Computers Using Mirror Circuits, *Phys. Rev. Lett.* **129** (2022).
 - [12] A. Cross, L. Bishop, S. Sheldon, P. Nation, and J. Gambetta, Validating quantum computers using randomized model circuits, *Phys. Rev. A* **100**, 032328 (2019).
 - [13] C. Baldwin, B. Bjork, J. Gaebler, D. Hayes, and D. Stack, Subspace benchmarking high-fidelity entangling operations with trapped ions, *Phys. Rev. Res.* **2**, 013317 (2020).
 - [14] C. M. Löschnauer, J. M. Toba, A. C. Hughes, S. A. King, M. A. Weber, R. Srinivas, R. Matt, R. Nourshargh, D. T. C. Allcock, C. J. Ballance, C. Matthiesen, M. Malinowski, and T. P. Harty, Scalable, high-fidelity all-electronic control of trapped-ion qubits, *arXiv:2407.07694* (2024).
 - [15] C. A. Sackett, D. Kielpinski, B. E. King, C. Langer, V. Meyer, C. J. Myatt, M. Rowe, Q. Turchette, W. M. Itano, D. J. Wineland, and C. Monroe, Experimental entanglement of four particles, *Nature* **404**, 256 (2000).
 - [16] S. T. Flammia and Y.-K. Liu, Direct Fidelity Estimation from Few Pauli Measurements, *Phys. Rev. Lett.* **106** (2011).
 - [17] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, *J. Mod. Opt.* **44**, 2455–2467 (1997).
 - [18] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, Gate set tomography, *Quantum* **5**, 557 (2021).
 - [19] K. Mølmer and A. Sørensen, Multiparticle Entanglement of Hot Trapped Ions, *Phys. Rev. Lett.* **82**, 1835 (1999).
 - [20] A. Sørensen and K. Mølmer, Entanglement and quantum computation with ions in thermal motion, *Phys. Rev. A* **62**, 022311 (2000).
 - [21] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, *Phys. Rev. A* **80** (2009).
 - [22] J. Helsen, I. Roth, E. Onorati, A. H. Werner, and J. Eisert, General framework for randomized benchmarking, *PRX Quantum* **3** (2022).
 - [23] C. J. Wood and J. M. Gambetta, Quantification and characterization of leakage errors, *Phys. Rev. A* **97** (2018).
 - [24] R. W. Andrews, C. Jones, M. D. Reed, A. M. Jones, S. D. Ha, M. P. Jura, J. Kerckhoff, M. Levendorf, S. Meenehan, S. T. Merkel, A. Smith, B. Sun, A. J. Weinstein, M. T. Rakher, T. D. Ladd, and M. G. Borselli, Quantifying error and leakage in an encoded Si/SiGe triple-dot qubit, *Nat. Nanotechnol.* **14**, 747–750 (2019).
 - [25] A. Morvan, V. V. Ramasesh, M. Blok, J. M. Kreikebaum, K. O'Brien, L. Chen, B. K. Mitchell, R. K. Naik, D. Santiago, and I. Siddiqi, Qutrit Randomized Benchmarking, *Phys. Rev. Lett.* **126** (2021).

- [26] R. P. Feynman, F. L. Vernon Jr, and R. W. Hellwarth, Geometrical representation of the schrödinger equation for solving maser problems, *J. Appl. Phys* **28**, 49 (1957).
- [27] R. Blume-Kohout, M. P. da Silva, E. Nielsen, T. Proctor, K. Rudinger, M. Sarovar, and K. Young, A Taxonomy of Small Markovian Errors, *PRX Quantum* **3**, 020335 (2022).
- [28] R. T. Sutherland, Q. Yu, K. M. Beck, and H. Häffner, One- and two-qubit gate infidelities due to motional errors in trapped ions and electrons, *Phys. Rev. A* **105**, 022437 (2022).
- [29] A consistent estimator is one that converges in probability to the true value as the number of data points is increased to infinity.
- [30] J. Claes, E. Rieffel, and Z. Wang, Character Randomized Benchmarking for Non-Multiplicity-Free Groups With Applications to Subspace, Leakage, and Matchgate Randomized Benchmarking, *PRX Quantum* **2** (2021).
- [31] M. A. Nielsen, A simple formula for the average gate fidelity of a quantum dynamical operation, *Phys. Lett. A* **303**, 249 (2002).
- [32] Y. Fan, R. Murray, T. D. Ladd, K. Young, and R. Blume-Kohout, Randomized Benchmarking with Synthetic Quantum Circuits, *arXiv:2412.18578* (2024).
- [33] J. R. Johansson, P. D. Nation, and F. Nori, QuTiP 2: A Python framework for the dynamics of open quantum systems, *Comput. Phys. Commun.* **184**, 1234 (2013).
- [34] D. Hayes, S. M. Clark, S. Debnath, D. Hucul, I. V. Inlek, K. W. Lee, Q. Quraishi, and C. Monroe, Coherent Error Suppression in Multiqubit Entangling Gates, *Phys. Rev. Lett.* **109**, 020503 (2012).
- [35] D. J. Wineland, C. Monroe, W. M. Itano, D. Leibfried, B. E. King, and D. M. Meekhof, Experimental Issues in Coherent Quantum-State Manipulation of Trapped Atomic Ions, *J. Res. Natl. Inst. Stand. and Technol.* **103**, 259 (1998).
- [36] F. Mintert and C. Wunderlich, Ion-Trap Quantum Logic Using Long-Wavelength Radiation, *Phys. Rev. Lett.* **87**, 257904 (2001).
- [37] C. Ospelkaus, C. E. Langer, J. M. Amini, K. R. Brown, D. Leibfried, and D. J. Wineland, Trapped-Ion Quantum Logic Gates Based on Oscillating Magnetic Fields, *Phys. Rev. Lett.* **101**, 090502 (2008).
- [38] T. P. Harty, M. A. Sepiol, D. T. C. Allcock, C. J. Ballance, J. E. Tarlton, and D. M. Lucas, High-Fidelity Trapped-Ion Quantum Logic Using Near-Field Microwaves, *Phys. Rev. Lett.* **117**, 140501 (2016).
- [39] C. V. Raman and K. S. Krishnan, A New Type of Secondary Radiation, *Nature* **121**, 501 (1928).
- [40] D. J. Wineland, M. Barrett, J. Britton, J. Chiaverini, B. DeMarco, W. M. Itano, B. Jelenković, C. Langer, D. Leibfried, V. Meyer, T. Rosenband, and T. Schätz, Quantum information processing with trapped ions, *Phil. Trans. R. Soc. A* **361**, 1349 (2003).
- [41] P. J. Lee, K.-A. Brickman, L. Deslauriers, P. C. Haljan, L.-M. Duan, and C. Monroe, Phase control of trapped ion quantum gates, *J. Opt. B: Quantum Semiclass. Opt.* **7**, S371 (2005).
- [42] G. J. Milburn, S. Schneider, and D. F. V. James, Ion trap quantum computing with warm ions, *Fortschr. Phys.* **48**, 801 (2000).
- [43] D. Leibfried, B. DeMarco, V. Meyer, D. M. Lucas, M. Barrett, J. Britton, W. M. Itano, B. Jelenković, C. Langer, T. Rosenband, and D. J. Wineland, Experimental demonstration of a robust, high-fidelity geometric two ion-qubit phase gate, *Nature* **422**, 412 (2003).
- [44] Y.-H. Chen and C. H. Baldwin, Randomized benchmarking with leakage errors, *arXiv:2502.00154* (2025).
- [45] See supplemental material.
- [46] A. Hashim, L. B. Nguyen, N. Goss, B. Marinelli, R. K. Naik, T. Chistolini, J. Hines, J. Marceaux, Y. Kim, P. Gokhale, T. Tomesh, S. Chen, L. Jiang, S. Ferracin, K. Rudinger, T. Proctor, K. C. Young, I. Siddiqi, and R. Blume-Kohout, Practical Introduction to Benchmarking and Characterization of Quantum Computers, *PRX Quantum* **6** (2025).
- [47] M. A. Nielsen, The entanglement fidelity and quantum error correction, *arXiv:quant-ph/9606012* (1996).
- [48] J. P. Marceaux and B. Bjork, Mølmer-Sørensen groups and irrep projectors (2025), <https://doi.org/10.5281/zenodo.17227531>.
- [49] T. J. Proctor, A. Carignan-Dugas, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young, Direct Randomized Benchmarking for Multiqubit Devices, *Phys. Rev. Lett.* **123** (2019).

Supplementary Information

I. FIDELITY DEFINITIONS

We recall the standard definitions of average gate fidelity and process fidelity, which play a key role in this work. Below, we consider a CPTP channel Λ and assume the state space is of dimension d ($= 2^n$) where n is the number of qubits.

The average fidelity $\bar{\mathcal{F}}(\Lambda)$ to the identity is

$$\bar{\mathcal{F}}(\Lambda) \equiv \int d\psi \langle \psi | \Lambda(|\psi\rangle\langle\psi|) | \psi \rangle, \quad (\text{S1})$$

where $d\psi$ represents the uniform Haar measure on the state space. One may also define the average fidelity of a quantum channel to a target unitary U

$$\bar{\mathcal{F}}(\Lambda, U) \equiv \int d\psi \langle \psi | U^\dagger \Lambda(|\psi\rangle\langle\psi|) U | \psi \rangle. \quad (\text{S2})$$

The process (aka. entanglement) fidelity $\mathcal{F}(\Lambda)$ of a channel to the identity is formally defined

$$\mathcal{F} \equiv \langle \Psi | (\mathbb{1} \otimes \Lambda)(|\Psi\rangle\langle\Psi|) | \Psi \rangle, \quad (\text{S3})$$

where $|\Psi\rangle$ is *any* maximally entangled state. However, a much more useful formula can be derived [46]

$$\mathcal{F}(\Lambda) = \frac{\text{Tr}(\Lambda)}{d^2}, \quad (\text{S4})$$

where Λ is represented as a super-operator.

The Horodecki formula [31] connects entanglement fidelity and average fidelity by

$$\bar{\mathcal{F}}(\Lambda) = \frac{d\mathcal{F}(\Lambda) + 1}{d + 1}. \quad (\text{S5})$$

Converting to infidelity $\bar{\mathcal{I}} = 1 - \bar{\mathcal{F}}$ and $\mathcal{I} = 1 - \mathcal{F}$, the relationship looks like

$$\bar{\mathcal{I}} = \frac{d\mathcal{I}}{d + 1}. \quad (\text{S6})$$

A convenient formula for average fidelity is [31]

$$\bar{\mathcal{F}}(\Lambda) = \frac{\sum_j \text{Tr}(P_j \Lambda(P_j)) + d^2}{d^2(d + 1)}, \quad (\text{S7})$$

where P_j is a Pauli operator and the sum is over all Paulis.

Some argue [46] that process fidelity is more fundamental because it is “stable” in the sense that it composes well with growing state spaces [46, 47]. If Λ_1 acts on one sub-register and Λ_2 acts on another sub-register, then the fidelity of the joint action is the product of the fidelity of the action on each sub-register:

$$\mathcal{F}(\Lambda_1 \otimes \Lambda_2) = \mathcal{F}(\Lambda_1) \mathcal{F}(\Lambda_2). \quad (\text{S8})$$

This is not true for average gate fidelity.

II. GENERATORS OF THE BENCHMARKING GROUP

The MS gate implements the following unitary:

$$\hat{U}_{2Q}(\theta, \phi) = \exp\left(-\frac{i\theta_{\text{MS}}}{2} \hat{\sigma}_{\phi_{\text{MS}},1} \hat{\sigma}_{\phi_{\text{MS}},2}\right),$$

where

$$\hat{\sigma}_{\phi_{\text{MS}},j} \equiv \hat{\sigma}_{x,j} \cos(\phi_{\text{MS}}) + \hat{\sigma}_{y,j} \sin(\phi_{\text{MS}}).$$

Now,

$$\begin{aligned} \hat{\sigma}_{\phi_{\text{MS}},1} \hat{\sigma}_{\phi_{\text{MS}},2} &= \hat{\sigma}_{x,1} \hat{\sigma}_{x,2} \cos^2(\phi) + \hat{\sigma}_{y,1} \hat{\sigma}_{y,2} \sin^2(\phi) \\ &\quad + \frac{1}{2} \sin(2\phi) (\hat{\sigma}_{x,1} \hat{\sigma}_{y,2} + \hat{\sigma}_{y,1} \hat{\sigma}_{x,2}) \end{aligned}$$

In matrix form, this reduces to

$$\begin{bmatrix} 0 & 0 & 0 & e^{-2i\phi} \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ e^{2i\phi} & 0 & 0 & 0 \end{bmatrix}.$$

The matrix form of the corresponding unitary is thus

$$\hat{U}_{2Q} = \begin{bmatrix} \cos\left(\frac{\theta}{2}\right) & 0 & 0 & -ie^{-2i\phi} \sin\left(\frac{\theta}{2}\right) \\ 0 & \cos\left(\frac{\theta}{2}\right) & -i \sin\left(\frac{\theta}{2}\right) & 0 \\ 0 & -i \sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right) & 0 \\ -ie^{2i\phi} \sin\left(\frac{\theta}{2}\right) & 0 & 0 & \cos\left(\frac{\theta}{2}\right) \end{bmatrix}.$$

The compilations that generate the SLERB benchmarking group are

$$M_x \equiv \hat{U}_{2Q}(\pi/2, 0) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & -i \\ 0 & 1 & -i & 0 \\ 0 & -i & 1 & 0 \\ -i & 0 & 0 & 1 \end{bmatrix}.$$

and

$$M_y \equiv \hat{U}_{2Q}(\pi/2, \pi/4) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -i & 0 \\ 0 & -i & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

It follows that

$$M_x \cong R_x^{00,11}\left(\frac{\pi}{2}\right) \oplus R_x^{01,10}\left(\frac{\pi}{2}\right),$$

$$M_y \cong R_y^{00,11}\left(\frac{\pi}{2}\right) \oplus R_x^{01,10}\left(\frac{\pi}{2}\right).$$

III. CHARACTER TABLE OF THE BENCHMARKING GROUP

We used the GAP computer program to find the character table of the benchmarking group generated by $\langle \hat{M}_x, \hat{M}_y \rangle$. A character table provides all possible irreducible representations of a group. We use it to identify the irreps that appear in our particular representation of the group benchmarking group as process matrices. The character table is given in Tab. II

	1a	4a	4b	2a	12a	3a	8a	12b	8b	4c	4d	4e	4f	4g	6a	2b
X.1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
X.2	1	-1	-1	1	1	1	-1	1	-1	1	1	-1	1	-1	1	1
X.3	1	$-i$	i	-1	-1	1	$-i$	-1	i	1	-1	$-i$	-1	i	1	1
X.4	1	i	$-i$	-1	-1	1	i	-1	$-i$	1	-1	i	-1	$-i$	1	1
X.5	2	.	.	2	-1	-1	.	-1	.	2	2	.	2	.	-1	2
X.6	2	.	.	-2	1	-1	.	1	.	2	-2	.	-2	.	-1	2
X.7	2	$1-i$	$1+i$	.	$-i$	-1	.	i	.	.	$-2i$	$-1+i$	$2i$	$-1-i$	1	-2
X.8	2	$1+i$	$1-i$	.	i	-1	.	$-i$	.	.	$2i$	$-1-i$	$-2i$	$-1+i$	1	-2
X.9	2	$-1-i$	$-1+i$	.	i	-1	.	$-i$	.	.	$2i$	$1+i$	$-2i$	$-1-i$	1	-2
X.10	2	$-1+i$	$-1-i$	.	$-i$	-1	.	i	.	.	$-2i$	$1-i$	$2i$	$1+i$	1	-2
X.11	3	-1	-1	-1	.	.	1	.	1	-1	3	-1	3	-1	.	3
X.12	3	1	1	-1	.	.	-1	.	-1	-1	3	1	3	1	.	3
X.13	3	$-i$	i	1	.	.	i	.	$-i$	-1	-3	$-i$	-3	i	.	3
X.14	3	i	$-i$	1	.	.	$-i$	.	i	-1	-3	i	-3	$-i$	.	3
X.15	4	.	.	.	$-i$	1	.	$-i$	.	.	$-4i$	.	$4i$	.	-1	-4
X.16	4	.	.	.	i	1	.	i	.	.	$4i$	.	$-4i$	.	-1	-4

TABLE II. Character table of the benchmarking group

IV. IRREDUCIBLE REPRESENTATIONS OF THE BENCHMARKING GROUP

The benchmarking group G generated by M_x and M_y has 96 elements. Its character table was identified using the GAP programming language [48]. From this character table, we identify the irreps present in a process matrix representation of the benchmarking group. We find that there are 6 irreps present in the representation, some of which appear with multiplicity. Furthermore we calculate the projectors on the irreps in the in a Pauli Transfer Matrix (PTM) basis. These irreps and their characterization are listed in Tab. III. We further observe that there is a convenient non-canonical basis for the trivial irrep, listed in Tab. IV.

V. OVERLAPS OF IRREPS WITH COMPUTATIONAL BASIS STATES

The overlaps between the irreps and computational basis preparation and measurements are given in Tab. V.

VI. DERIVATION OF THE DECAY MODEL

To derive the decays associated with the experimental implementation of SLERB in the main text, we start by analyzing a slightly more general implementation than the one performed experimentally, and then we will argue that the experimental implementation is a good approximation. Namely, we will start by assuming that the physical operations are selected randomly from the entire benchmarking group of 96 elements rather than from the subset of 24 elements that were actually chosen experimentally. After deriving decay forms in this context, we will then argue that selecting from the subset of 24 elements instead is a good approximation to the general dynamics.

We consider the following experiment: a sequence of l Cliffords are selected at random from the 96-element benchmarking group, a final $l+1$ Clifford is selected as the inverse of the product of l prior gates, and the benchmarking sequence is run after preparing either $|00\rangle$ or $|11\rangle$. For concreteness, we will focus on preparing $|00\rangle$. The final state is then measured in the computational basis. To analyze the effects on outcome probabilities, we consider, as usual, gate-independent noise. That is to say, we assume that the noisy implementation $\mathcal{G}(g)$ of an element g is the benchmarking group G can be represented as a process matrix of the form

$$\mathcal{G}(g) = \Lambda V(g). \quad (\text{S9})$$

Now, to derive the expected decay forms, we begin with the definition of a twirl $\hat{\Lambda}_G$ of a noise channel Λ over a group G :

$$\hat{\Lambda}_G = \frac{1}{|G|} \sum_{g \in G} V(g) \Lambda V(g)^{-1}. \quad (\text{S10})$$

The key insight that we exploit is that matrix powers of a twirled noise channels are equivalent to arithmetic averages over random sequences,

$$\hat{\Lambda}_G^l = \frac{1}{|G|^l} \sum_{g_l \in G} \dots \sum_{g_1 \in G} V(g_{l+1}) \Lambda V(g_l) \Lambda \dots \Lambda V(g_1), \quad (\text{S11})$$

where g_{l+1} is the inverse of the product of the l prior group elements. We emphasize that this formal identification requires an approximation that the final inverse gate is perfect i.e. the error on the final Clifford gate is not twirled. In practice, the error on the final Clifford can be analyzed separately or else absorbed into measurement error.

Next, we employ Schur's lemma (see Fact 1 in [30]) to argue that the action of the twirled error channel is equivalent, up to a change of basis, to a block-diagonal

Irrep.	Projector	Multiplicity	$\dim(\mathcal{H}_\nu)$	Basis
Trivial	Π_0	3	1	$ 00\rangle\langle 00 + 11\rangle\langle 11 , \Psi^+\rangle\langle \Psi^+ , \Psi^-\rangle\langle \Psi^- ,$
SU(2) subspace	Π_c	1	3	$ 00\rangle\langle 11 , 11\rangle\langle 00 , 00\rangle\langle 00 - 11\rangle\langle 11 $
Symmetric-Antisymmetric transfer	$\Pi_{a,1}$	1	1	$ \Psi^-\rangle\langle \Psi^+ $
Antisymmetric-Symmetric transfer	$\Pi_{a,2}$	1	1	$ \Psi^+\rangle\langle \Psi^- $
Forward mixing processes	$\Pi_{b,1}$	2	2	$ \Psi^+\rangle\langle 00 , \Psi^+\rangle\langle 11 , 00\rangle\langle \Psi^- , 11\rangle\langle \Psi^- $
Backwards mixing processes	$\Pi_{b,2}$	2	2	$ \Psi^-\rangle\langle 00 , \Psi^-\rangle\langle 11 , 00\rangle\langle \Psi^+ , 11\rangle\langle \Psi^+ $

TABLE III. Irreps. of the SLERB benchmarking group with projector labels and properties.

Irrep.	Projector
Even-parity	$ 00\rangle\langle 00 + 11\rangle\langle 11 $
Symmetric odd-parity	$ \Psi^+\rangle\langle \Psi^+ $
Anti-symmetric odd-parity	$ \Psi^-\rangle\langle \Psi^- $

TABLE IV. A choice of basis for the the trivial irrep of multiplicity 3. Here $|\Psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$ and $|\Psi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$.

Π_0	$\begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}$
Π_c	$\begin{bmatrix} \frac{1}{2} & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -\frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}$
Π_a^1	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\ 0 & -\frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Π_a^2	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\ 0 & -\frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Π_b^1	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Π_b^2	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

TABLE V. Overlaps of projectors with computational basis state preparation and measurement pairs. Columns correspond to preparations in the order 00, 01, 10, 11, and rows correspond to measurements in the same order.

decomposition of actions on irreps of the benchmarking group

$$\frac{1}{|G|} \sum_{g \in G} V(g) \Lambda V(g)^{-1} \cong \bigoplus_{\nu} Q_{\nu} \otimes \mathbb{1}_{\nu}, \quad (\text{S12})$$

where ν indexes the irreps of the benchmarking group, Q_{ν} is a $n_{\nu} \times n_{\nu}$ matrix, n_{ν} is the multiplicity of the irrep in the benchmarking group, and $\mathbb{1}_{\nu}$ is an identity matrix of dimension $\dim(\mathcal{H}_{\nu})$ equal to the dimension of the irrep.

Thus,

$$\hat{\Lambda}_G^l \cong \bigoplus_{\nu} Q_{\nu}^l \otimes \mathbb{1}_{\nu}. \quad (\text{S13})$$

To analyze the dynamics of measured probabilities, we thus consider success probabilities, written in “super-ket” notation [18]:

$$\langle\langle a | \bigoplus_{\nu} Q_{\nu}^l \otimes \mathbb{1}_{\nu} | 00 \rangle\rangle = \sum_{\nu} \langle\langle a | Q_{\nu}^l \otimes \mathbb{1}_{\nu} | 00 \rangle\rangle,$$

where $\langle\langle a |$ is measurement effect in the computational basis. Now, to derive the contributions to success probabilities, we make the observation that each term $Q_{\nu} \otimes \mathbb{1}_{\nu}$ is contained in the support of the projector on the corresponding irrep. These were explicitly constructed for our benchmarking with the GAP language, see Tab. III. We then calculate the overlap between these projectors and the relevant state preparations and measurement effects in the same basis. These results are summarized in Tab. V. One may observe that the only irreps that have overlap when state preparations of 00/11 and measurement in the computational basis are the “SU(2) subspace irrep” and the trivial irrep. Because the SU(2) subspace irrep appears with multiplicity 1, its contribution is a single-exponential decay

$$\langle\langle 00 | Q_c^l \otimes \mathbb{1}_c | 00 \rangle\rangle = \frac{1}{2} q_{\text{RB}}^l,$$

where the factor of 1/2 comes from the overlap between the SU(2) irrep and this SPAM pair. Similarly, one can show

$$\langle\langle 11 | Q_c^l \otimes \mathbb{1}_c | 00 \rangle\rangle = -\frac{1}{2} q_{\text{RB}}^l.$$

The analysis of the trivial irrep is more subtle because it appears with multiplicity 3. The general form of the contributions follow as

$$\langle\langle a | Q_0^l \otimes \mathbb{1}_0 | 00 \rangle\rangle = c_1 q_{\text{leak},1}^l + c_2 q_{\text{leak},+}^l + c_3 q_{\text{leak},-}^l,$$

where c_j are constants that depend on the overlap of the eigen-vectors of Q_0 and the SPAM pair in question and q_j correspond to eigen-values. In the following section, $q_{\text{leak},\pm}^l$ correspond to leakage rates to the two subspaces $\mathcal{S}_{\text{leak}}^{\pm}$. This form can be further simplified by invoking

trace preservation, which manifests as a restriction of one eigenvalue of Q_0 to be 1. Hence:

$$\langle\langle a|Q_0^l \otimes \mathbb{1}_0|00\rangle\rangle = c_1 + c_2 q_{\text{leak},+}^l + c_3 q_{\text{leak},-}^l. \quad (\text{S14})$$

So that the general decay forms can be derived

$$\begin{aligned} \langle\langle 00|\hat{\Lambda}_G^l|00\rangle\rangle &= c_1 + c_2 q_{\text{leak},+}^l + c_3 q_{\text{leak},-}^l + \frac{1}{2} q_{\text{RB}}^l, \\ \langle\langle 11|\hat{\Lambda}_G^l|00\rangle\rangle &= d_1 + d_2 q_{\text{leak},+}^l + d_3 q_{\text{leak},-}^l - \frac{1}{2} q_{\text{RB}}^l. \end{aligned} \quad (\text{S15})$$

We emphasize that this decay model assumes no state preparation and measurement error and no error on the final inversion gate. Without further assumptions on the nature of the noise, this general form cannot be specified further. In the next section, we will show how an assumption of leakage only within the symmetric subspace and equal leakage/seepage results in the specific decay model we use in the paper.

The final step is to argue that the decay model above also accurately describes the dynamics of the experimental protocol in the main text – namely where a subset of 24 elements $\mathcal{T} \subset G_{RB}$ whose action on $\mathcal{S}_{RB}$ are selected to implement the 1-qubit Clifford group on $\mathcal{S}_{RB}$. There are two ways this can be argued: 1) via a notion of asymptotic twirling or 2) through simulations. To 1) we argue that selecting from the subset of Clifford $\mathcal{T}$ provides sufficient randomness that it asymptotically generates the same dynamics as would appear if one had sampled from the group of 96 elements. This will generally only be true after some initial sequence length is passed. This is a similar approach to that of direct RB [49]. To 2) we have observed through numeric simulations that the decay curves observed with direct unitary simulation agree to a large degree with those found by performing the direct twirl of the same noise channel.

VII. DECAY FORMS WITH SYMMETRIC LEAKAGE

The decay model of Eq. S27 represents the most specific model that can be derived with generic assumptions on the noise. However, as discussed in the main text, we expect that the dominant error channels in our system cause leakage to only the symmetric state $\mathcal{S}_{\text{leak}}^+$. Furthermore, we expect that, for the experimental regime in the main text, the dominant error mechanisms are unitary errors. This implies that the leakage rate from $\mathcal{S}_{RB}$ to $\mathcal{S}_{\text{leak}}^+$ and the seepage from $\mathcal{S}_{\text{leak}}^+$ to $\mathcal{S}_{RB}$ should be equal. We hence consider the effects on the dynamics under these assumptions. As will be shown in the next section, it turns out that these assumptions lead to a decay model that can be expressed as a continuous time Markov process. These derivations closely mirror analysis in Appendix C of Ref. [13], and we refer to reader also to Sec. V of Ref. [30] for related discussion.

Begin by selecting projectors for the isotopic components of the trivial irrep, as in Tab. IV, and converting them to trace-normalized superkets using the normalization convention of [13]:

$$|00\rangle\langle 00| + |11\rangle\langle 11| \mapsto \frac{|00\rangle\langle 00| + |11\rangle\langle 11|}{\sqrt{2}} \equiv |\mathbb{1}_{RB}\rangle\langle\mathbb{1}_{RB}|,$$

$$|\Psi^\pm\rangle\langle\Psi^\pm| \mapsto |\Psi^\pm\rangle\langle\Psi^\pm|.$$

We make the assumption that no leakage occurs from $\mathcal{S}_{RB}$ to $\mathcal{S}_{\text{leak}}^-$, ie., that

$$\langle\langle\Psi^-|\Lambda|\mathbb{1}_{RB}\rangle\rangle = 0. \quad (\text{S16})$$

Next, define the symmetric leakage rate

$$L = \langle\langle\Psi^+|\Lambda|\mathbb{1}_{RB}\rangle\rangle, \quad (\text{S17})$$

and the seepage rate

$$S = \langle\langle\mathbb{1}_{RB}|\Lambda|\Psi^+\rangle\rangle. \quad (\text{S18})$$

The next step is to observe that Q_0 can be decomposed in this basis for the trivial irrep

$$(Q_0 \otimes \mathbb{1}_0)_{\alpha\beta} = Q_{\alpha\beta} \langle\langle P_\alpha|\Lambda|P_j\rangle\rangle, \quad (\text{S19})$$

where P_j are projectors on isotopic components of the trivial irrep defined above. Based on our definitions of L and S we find

$$Q_{\alpha\beta} = \begin{bmatrix} 1 - \sqrt{2}L & S & 0 \\ L & 1 - S/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (\text{S20})$$

where the columns correspond to $\mathcal{S}_{RB}$, $\mathcal{S}_{\text{leak}}^+$, and $\mathcal{S}_{\text{leak}}^-$, resp., and similarly for the rows. See also Eq. C5 in [13]. It follows that there are two eigenvalues of 1 and a non-trivial eigenvalue of $\lambda_3 = 1 - L/\sqrt{2} - \sqrt{2}S$. These are associated with the eigenvalue projectors, in order:

$$\Pi_1 = \frac{1}{L + 2S} \begin{bmatrix} 2S & \sqrt{2}S & 0 \\ \sqrt{2}L & L & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (\text{S21})$$

$$\Pi_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{S22})$$

$$\Pi_3 = \frac{1}{L + 2S} \begin{bmatrix} L & -\sqrt{2}S & 0 \\ -\sqrt{2}L & 2S & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (\text{S23})$$

Using the overlaps between Π_j and $|00\rangle\langle 00|$ and $|11\rangle\langle 11|$ one arrives at

$$\langle\langle 00|Q_0^l \otimes \mathbb{1}_0|00\rangle\rangle = \frac{S}{L + 2S} + \frac{1}{2} \frac{L}{L + 2S} q_{\text{leak},+}^l \quad (\text{S24})$$

where $q_{\text{leak},+}^l$ is the non-trivial eigenvalue $\lambda_3 = 1 - L/\sqrt{2} - \sqrt{2}S$. Furthermore,

$$\langle\langle 11|Q_0^l \otimes \mathbb{1}_0|00\rangle\rangle = \frac{S}{L+2S} + \frac{1}{2} \frac{L}{L+2S} q_{\text{leak},+}^l. \quad (\text{S25})$$

In this way, we conclude that

$$\begin{aligned} \langle\langle 00|\hat{\Lambda}_G^l|00\rangle\rangle &= \frac{S}{L+2S} + \frac{1}{2} \frac{L}{L+2S} q_{\text{leak},+}^l + \frac{1}{2} q_{\text{RB}}^l, \\ \langle\langle 11|\hat{\Lambda}_G^l|00\rangle\rangle &= \frac{S}{L+2S} + \frac{1}{2} \frac{L}{L+2S} q_{\text{leak},+}^l - \frac{1}{2} q_{\text{RB}}^l. \end{aligned} \quad (\text{S26})$$

Finally, under the assumption that $L = S$, one can arrive at Eqs. 42 and 41:

$$\begin{aligned} \langle\langle 00|\hat{\Lambda}_G^l|00\rangle\rangle &= \frac{1}{3} + \frac{1}{6} q_{\text{leak},+}^l + \frac{1}{2} q_{\text{RB}}^l, \\ \langle\langle 11|\hat{\Lambda}_G^l|00\rangle\rangle &= \frac{1}{3} + \frac{1}{6} q_{\text{leak},+}^l - \frac{1}{2} q_{\text{RB}}^l. \end{aligned} \quad (\text{S27})$$

To connect to the Markov transition model presented in the next section, we expand $q_{\text{leak},+}^l$ under the assumption $L = S$ and write

$$\langle\langle 00|\hat{\Lambda}_G^l|00\rangle\rangle = \frac{1}{3} + \frac{1}{6} (1 - \frac{3}{\sqrt{2}}L)^l + \frac{1}{2} q_{\text{RB}}^l.$$

Making a change of variables $L/\sqrt{2} \mapsto \epsilon_{\text{leak}}$, one arrives at the expression

$$\langle\langle 00|\hat{\Lambda}_G^l|00\rangle\rangle = \frac{1}{3} + \frac{1}{6} (1 - 3\epsilon_{\text{leak}})^l + \frac{1}{2} q_{\text{RB}}^l,$$

which closely matches an independent derivation from the perspective of Markov transition processes.

VIII. A MARKOV TRANSITION MODEL

Under the assumption that there is no leakage to the anti-symmetric state, then the only relevant subspaces are $\mathcal{S}_{RB}$ and $\mathcal{S}_L^+$, which are both twirled over (recall $\mathcal{S}_L^+$ is 1-dimensional, so application of random phases will twirl over this space). Under these assumptions, the transfer of population between subspaces is completely incoherent. We write a Markov model that captures these dynamics, which results in a decay form that is very similar to that derived from representation-theoretical approaches.

Thus, let us write an effective Markov model for the transition of the population

$$M = \begin{bmatrix} 1 - \epsilon_{\text{RB}} - \epsilon_{\text{leak}} & \epsilon_{\text{leak}} & \epsilon_{\text{RB}} \\ \epsilon_{\text{leak}} & 1 - 2\epsilon_{\text{leak}} & \epsilon_{\text{leak}} \\ \epsilon_{\text{RB}} & \epsilon_{\text{RB}} & 1 - \epsilon_{\text{RB}} - \epsilon_{\text{leak}} \end{bmatrix}$$

where ϵ_{leak} is a leakage rate and ϵ_{RB} is a “spin flip rate” that corresponds to errors in $\mathcal{S}_{RB}$.

The eigenvalues of the transfer matrix are $1, 1 - 2\epsilon_{\text{RB}} - \epsilon_{\text{leak}}, 1 - 3\epsilon_{\text{RB}}$ and the normalized eigenvector projectors are

$$P_1 = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

$$P_2 = \frac{1}{2} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix},$$

$$P_3 = \frac{1}{6} \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}.$$

It follows that

$$M^p = \frac{1}{3} P_1 + \frac{1}{2} (1 - 2\epsilon_{\text{RB}} - \epsilon_{\text{leak}})^p P_2 + \frac{1}{6} (1 - 3\epsilon_{\text{leak}})^p P_3.$$

Hence, preparing in $|00\rangle$, and measuring $|00\rangle$ corresponds to

$$\langle\langle 00|M^p|00\rangle\rangle = \frac{1}{3} + \frac{1}{2} (1 - 2\epsilon_{\text{RB}} - \epsilon_{\text{leak}})^p + \frac{1}{6} (1 - 3\epsilon_{\text{leak}})^p, \quad (\text{S28})$$

and similarly for preparing in $|00\rangle$ measuring $|11\rangle$ and preparing in $|00\rangle$ measuring in the leaked state. This is essentially the exact dynamics that were derived from a representation-theoretical analysis of the problem, and indicate that a Markov description of population should be accurate under the assumption of symmetric leakage and equal rates of leakage and seepage.

In this way we have shown

$$\begin{aligned} P_{\text{survival}} &= \frac{1}{3} + \frac{1}{2} (1 - 2\epsilon_{\text{RB}} - \epsilon_{\text{leak}})^l + \frac{1}{6} (1 - 3\epsilon_{\text{leak}})^l, \\ P_{\text{leak}} &= \frac{1}{3} - \frac{1}{3} (1 - 3\epsilon_{\text{leak}})^l, \\ P_{\text{flip}} &= \frac{1}{3} - \frac{1}{2} (1 - 2\epsilon_{\text{RB}} - \epsilon_{\text{leak}})^l + \frac{1}{6} (1 - 3\epsilon_{\text{leak}})^l, \end{aligned} \quad (\text{S29})$$

where P_{survival} is the probability of measuring the target final state ($|00\rangle$ or $|11\rangle$), P_{leak} is the probability of measuring population in the $|01\rangle$ or $|10\rangle$ states, and P_{flip} is the probability of measuring the non-target state within the $|00\rangle, |11\rangle$ subspace. Under these assumptions of symmetric leakage and equal rates of seepage and leakage, ϵ_{RB} can be interpreted as the error per Clifford within the two-state subspace, and ϵ_{leak} as a leakage error. We note that this model was derived in the absence of state preparation and measurement errors and assumes that the final inversion gate is perfect.

IX. QUANTIFYING RB FIDELITY UNDER PARTIAL TWIRLING

To derive the effective fidelity function, we analyze the decomposition of gate fidelity into its action on irreps.

As emphasized in the main text, the SLERB experiment only “fully twirls” over the subspace $\mathcal{S}_{RB}$, and the action of the benchmarking group on the entire space $\mathcal{S}_{2Q}$ does not represent a unitary 2-design. This will complicate our derivation of the fidelity function. To perform the analysis, we will consider the effects of twirling under the benchmarking group G_{RB} . This Section provides the general theory, and the next section specializes to the case of SLERB benchmarking.

We begin with the definition of the channel fidelity $\mathcal{F}$ of a noise channel Λ to the identity

$$\mathcal{F}(\Lambda) = \frac{1}{d^2} \text{Tr}[\Lambda]. \quad (\text{S30})$$

Next, we invoke the fact that the trace is invariant under twirling to write

$$\frac{1}{d^2} \text{Tr}[\Lambda] = \frac{1}{d^2} \text{Tr}[\hat{\Lambda}_G]. \quad (\text{S31})$$

where d is the dimension of the Hilbert space. Applying a general decomposition of twirled channel, one writes

$$\frac{1}{d^2} \text{Tr}[\hat{\Lambda}_G] = \frac{1}{d^2} \text{Tr} \left[\bigoplus_{\nu=1}^l Q_\nu \otimes \mathbb{1}_\nu \right]. \quad (\text{S32})$$

$$\mathcal{F}(\Lambda) = \frac{1 + 3q_{RB} + q_{\text{leak},+} + q_{\text{leak},-} + q_{a,1} + q_{a,2} + 2(q_{b,1} + q_{b,2} + q_{b,3} + q_{b,4})}{16}, \quad (\text{S35})$$

where $q_{RB}, q_{\text{leak},+}$ are the decays measured in the experiment of the main text, $q_{\text{leak},-}$ correspond to leakage to the anti-symmetric state, and the other terms correspond to unmeasured decays that will generally manifest with coherent effects (ie. oscillations).

In order to construct a consistent estimator of fidelity, it is necessary to measure all the decays. It would be possible to do so with an approach such as character or synthetic RB [32]. However, in our experiment the only decays that contribute to experimental observables are $q_{RB}, q_{\text{leak},+}$, and $q_{\text{leak},-}$. As we have said, we cannot resolve $q_{\text{leak},+}$ from $q_{\text{leak},-}$ and we expect that $q_{\text{leak},-} \gg q_{\text{leak},+}$. To construct an approximate estimator of gate fidelity, we set $q_{\text{leak},+} = q_{\text{leak},-}$ (which is likely overly pessimistic) and report the unmeasured decays as the average of the measured decays. This means setting all other terms equal to $(q_{RB} + q_{\text{leak},+})/2$.

$$\mathcal{F}(\Lambda) = \frac{1 + 8q_{RB} + 7q_{\text{leak},+}}{16}. \quad (\text{S36})$$

One can convert the channel fidelity above to an effective

It follows for properties of the trace and tensor products that

$$\frac{1}{d^2} \text{Tr}[\hat{\Lambda}_G] = \frac{1}{d^2} \sum_{\nu=1} \text{Tr}[Q_\nu] \text{Tr}[\mathbb{1}_\nu]. \quad (\text{S33})$$

Now, using that $\text{Tr}[\mathbb{1}_\nu] = \dim(\mathcal{H}_\nu)$, it follows that

$$\mathcal{F}(\Lambda) = \frac{1}{d^2} \sum_{\nu} \dim(\mathcal{H}_\nu) \sum_{\mu} q_{\nu,\mu}, \quad (\text{S34})$$

where $q_{\nu,\mu}$ is the μ th eigenvalue of the matrix Q_ν .

We see above that the fidelity of a noise channel can be calculated from the knowledge of $q_{\nu,\mu}$.

X. SLERB EXTENDED SUBSPACE FIDELITY

To derive Eqs. 43 and 44, we begin with the general decomposition of channel fidelity in terms of action on irreps, Eq. S34, reproduced below:

$$\mathcal{F}(\Lambda) = \frac{1}{d^2} \sum_{\nu} \dim(\mathcal{H}_\nu) \sum_{\mu} q_{\nu,\mu}.$$

In our case, the benchmarking group’s representation decomposes into 6 irreps that are listed in Tab. III. Hence, the general fidelity function for SLERB is

average gate fidelity via the Horodecki formula [31]

$$\bar{\mathcal{F}} = \frac{4\mathcal{F} + 1}{5},$$

to arrive at

$$\bar{\mathcal{F}} = \frac{5 + 8q_{RB} + 7q_{\text{leak},+}}{20}. \quad (\text{S37})$$

As emphasized in the main text, the formulas above represent an “extended subspace fidelity” and should not be conflated with a fully rigorous metric of two-qubit gate fidelity. These fidelity metrics are expected to accurately describe situations where the noise is not highly biased to particular decay contributions. We calculated the Q -matrix across the irreps for the dominant noise channels in laser-free MS gates discussed in the main text, and explicitly verified that no significant noise bias.

XI. STATE PREPARATION AND MEASUREMENT ERRORS

Without any state preparation and measurement (SPAM) errors, the measured populations follow

$$\begin{aligned}
P_{\text{survival}} &= \frac{1}{3} + \frac{1}{2}(1 - 2\epsilon_{\text{RB}} - \epsilon_{\text{leak}})^l + \frac{1}{6}(1 - 3\epsilon_{\text{leak}})^l, \\
P_{\text{leak}} &= \frac{1}{3} - \frac{1}{3}(1 - 3\epsilon_{\text{leak}})^l, \\
P_{\text{flip}} &= \frac{1}{3} - \frac{1}{2}(1 - 2\epsilon_{\text{RB}} - \epsilon_{\text{leak}})^l + \frac{1}{6}(1 - 3\epsilon_{\text{leak}})^l.
\end{aligned}$$

We perform Pauli-frame randomization, which averages any measurement errors of the states $|0\rangle$ and $|1\rangle$. In these experiments, we initialise both ions to $|00\rangle$. Thus, the measured populations for a sequence without any two-qubit gates are

$$\begin{aligned}
P_{00} &= (1 - \epsilon_{\text{SPAM},1})(1 - \epsilon_{\text{SPAM},2}), \\
P_{01 \text{ or } 10} &= \epsilon_{\text{SPAM},1}(1 - \epsilon_{\text{SPAM},2}) + (1 - \epsilon_{\text{SPAM},1})\epsilon_{\text{SPAM},2}, \\
P_{11} &= \epsilon_{\text{SPAM},1}\epsilon_{\text{SPAM},2},
\end{aligned}$$

where $\epsilon_{\text{SPAM},i}$ is the SPAM error for ion i i.e. the probability of measuring $|0\rangle$ when preparing $|1\rangle$. If $\epsilon_{\text{SPAM},i} \ll 1$, we simplify the equations above by keeping terms only first order in $\epsilon_{\text{SPAM},i}$. Defining the average state preparation error

$$\bar{\epsilon}_{\text{SPAM}} = \frac{1}{2}(\epsilon_{\text{SPAM},1} + \epsilon_{\text{SPAM},2}), \quad (\text{S38})$$

and using the initial populations to calculate the decays, we obtain

$$\begin{aligned}
P_{\text{survival}} &= \frac{1}{3}(1 - \bar{\epsilon}_{\text{SPAM}}) \\
&\quad + \frac{1}{2}(1 - 2\bar{\epsilon}_{\text{SPAM}})(1 - 2\epsilon_{\text{RB}} - \epsilon_{\text{leak}})^l \\
&\quad + \frac{1}{6}(1 - 4\bar{\epsilon}_{\text{SPAM}})(1 - 3\epsilon_{\text{leak}})^l, \\
P_{\text{leak}} &= \frac{1}{3}(1 + 2\bar{\epsilon}_{\text{SPAM}}) - \frac{1}{3}(1 - 4\bar{\epsilon}_{\text{SPAM}})(1 - 3\epsilon_{\text{leak}})^l, \\
&\quad (\text{S39})
\end{aligned}$$

$$\begin{aligned}
P_{\text{flip}} &= \frac{1}{3}(1 - \bar{\epsilon}_{\text{SPAM}}) \\
&\quad - \frac{1}{2}(1 - 2\bar{\epsilon}_{\text{SPAM}})(1 - 2\epsilon_{\text{RB}} - \epsilon_{\text{leak}})^l \\
&\quad + \frac{1}{6}(1 - 4\bar{\epsilon}_{\text{SPAM}})(1 - 3\epsilon_{\text{leak}})^l.
\end{aligned}$$

This derivation accounts for the starting states $|01\rangle$ and $|10\rangle$ having different asymptotes to $|00\rangle$ or $|11\rangle$. There will be an additional correction to the asymptotes due to measurement errors, despite Pauli-frame randomization, as we stay within the symmetric subspace. For simplicity, we ignore those here, as the main data is far from the asymptotes, and any correction would be a fraction of the measurement error, which is less than 10^{-4} .

XII. EFFECT OF MOTIONAL STATE

Errors in the MS interaction can result in residual spin-motion entanglement (SME) at the end of the gate duration. Without additional ground-state cooling, any subsequent two-qubit gates will be performed with this residual SME, which can result in non-Markovian errors due to a coherent build up of motional population. For the theory and experiments described in the main text, we do not consider this effect. Here, we perform simulations, again in QuTiP, to investigate this effect in two regimes: a fractional detuning error of 10^{-2} (Fig. 8), and 7×10^{-2} (Fig. 9). The first case corresponds to a two-qubit gate error of 2.5×10^{-4} , i.e. illustrates a scenario where the vast majority of the experimentally measured gate error is due to motional errors. For each of these cases, we perform simulations where we either keep the motional state unchanged after every gate sequence, or reset it to the ground state after every two-qubit gate. For the smaller error, the motional state has a negligible effect on the fidelity estimate. Thus, we conclude that our experimental fidelity estimation is not significantly biased by coherent motional population buildup. For the larger error, this effect is still small, but we observe the data deviating from the model as shown in Fig. 9a.

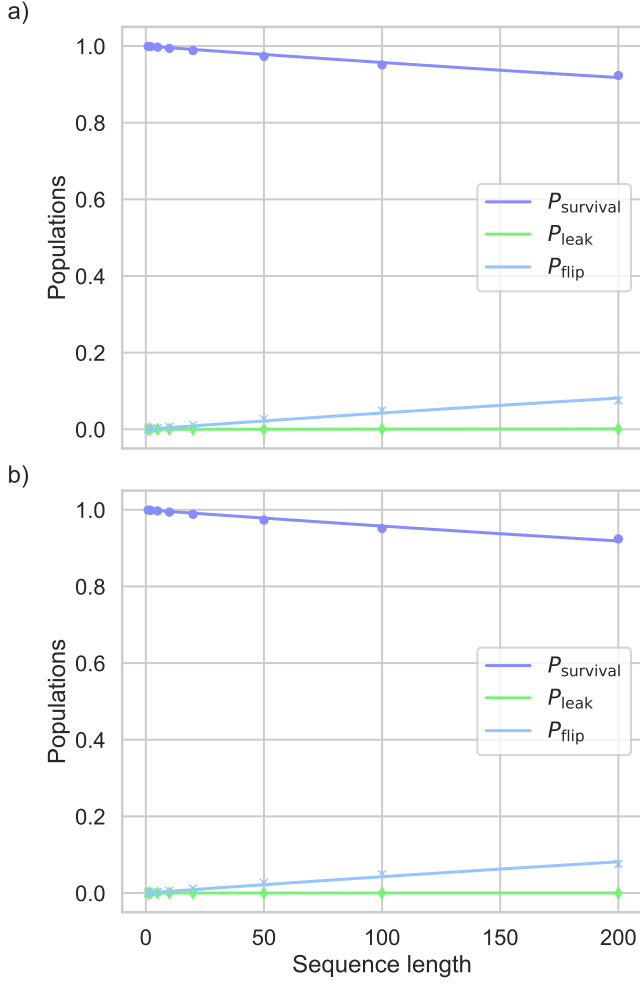


FIG. 8. Numerical simulations of SLERB sequences for a fractional detuning error of 10^{-2} with a) the motional state unchanged or b) reset to the ground state after every two-qubit gate. For each sequence length, we sample 50 random sequences; the error bars correspond to the 68% confidence interval for the average populations from all the randomisations. The solid lines correspond to fits to the data following Eq. 25. For this error, the motional phase has a negligible effect on the dynamics or the estimated two-qubit gate error of 2.5×10^{-4} .

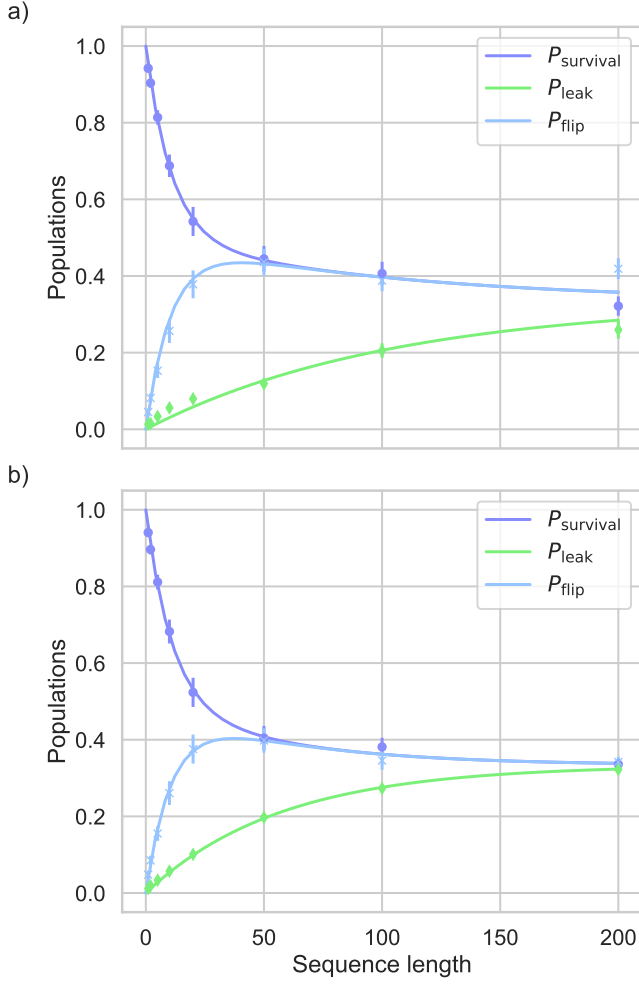


FIG. 9. Numerical simulations of SLERB sequences for a fractional detuning error of 7×10^{-2} with a) the motional state unchanged or b) reset to the ground state after every two-qubit gate. For each sequence length, we sample 50 random sequences; the error bars correspond to the 68% confidence interval for the average populations from all the randomisations. The solid lines correspond to fits to the data following Eq. 25. In contrast to Fig. 8, the motional state has a more significant effect on the dynamics and deviates slightly from the model. However, both sets of data result in an estimated two-qubit gate error of 2.5×10^{-2} .