Experimental Bayesian Calibration of Trapped-Ion Entangling Operations

Lukas Gerster[1,*] Fernando Martínez-García,[2] Pavel Hrmo,[1,†] Martin W. van Mourik,[1] Benjamin Wilhelm,[1] Davide Vodola,[3] Markus Müller,[4,5] Rainer Blatt,[1,6] Philipp Schindler,[1] and Thomas Monz[1,7]

[1] *Institut für Experimentalphysik, Universität Innsbruck, Technikerstraße 25/4, Innsbruck 6020, Austria*

[2] *Department of Physics, Swansea University, Singleton Park, Swansea SA2 8PP, United Kingdom*

[3] *Dipartimento di Fisica e Astronomia Augusto Righi, Via Irnerio 46, Bologna, Italy*

[4] *Institute for Quantum Information, RWTH Aachen University, Aachen D-52056, Germany*

[5] *Peter Grünberg Institute, Theoretical Nanoelectronics, Forschungszentrum Jülich, Jülich D-52425, Germany*

[6] *Institut für Quantenoptik und Quanteninformation, Österreichische Akademie der Wissenschaften, Technikerstraße 21a, Innsbruck 6020, Austria*

[7] *AQT, Technikerstraße 17, Innsbruck 6020, Austria*

The performance of quantum gate operations is experimentally determined by how correctly operational parameters can be determined and set, and how stable these parameters can be maintained. In addition, gates acting on different sets of qubits require unique sets of control parameters. Thus, an efficient multidimensional parameter estimation procedure is crucial to calibrate even medium-sized quantum processors. Here, we develop and characterize an efficient calibration protocol to automatically estimate and adjust experimental parameters of the widely used two-qubit Mølmer-Sørensen entangling gate operation in a trapped-ion quantum information processor. The protocol exploits Bayesian parameter estimation methods that include a stopping criterion based on a desired gate infidelity. We experimentally demonstrate a tune-up procedure that leads to a residual median gate infidelity due to miscalibration of $1.3(1) \times 10^{-3}$, requiring 1200 ± 500 experimental cycles, while completing the entire gate calibration procedure in less than one minute, which provides a significant speedup over commonly used manual tune-up routines. This approach is applicable to other quantum information processor architectures with known or sufficiently characterized theoretical models.

I. INTRODUCTION

The development of quantum information processors has made rapid progress in recent years. The leading paradigm for quantum computation is the circuit model where local operations and two- or multiqubit entangling operations provide a gate set that allows for universal application of quantum circuits. New technological and theoretical developments in various quantum computing

---

*lukas.gerster@uibk.ac.at

†Present address: Trapped Ion Quantum Information Group, Institute for Quantum Electronics, ETH Zurich, 8093 Zurich, Switzerland.

platforms [1–6] have pushed the fidelities of the available entangling gate operations closer to the parameter regime needed for fault-tolerant quantum error correction [7–16]. Furthermore, even without quantum error correction there is the expectation that with moderately sized systems with upwards of about 50 physical qubits with sufficiently high fidelities quantum advantage can be observed [17–19]. However, achieving the necessary fidelities requires a precise calibration of the various classical experimental control parameters that determine the realized Hamiltonians that generate single-qubit and in particular two- or multiqubit entangling gate operations. Typically, such calibrations need to be performed by a highly trained operator who is familiar with the system. Furthermore, the control parameters are liable to drifts and will require some form of feedback to maintain the desired fidelity during the course of operation. It is thus highly desirable for a quantum computing platform to implement an automation procedure that can determine the optimal control parameters accurately, such that it can be operated by

an end user with the option for periodic recalibration [20–23]. As quantum computing platforms mature and move to remotely accessed services, such automated calibration routines will become indispensable to keep the machines at peak performance without the need for in-person maintenance, allowing the end user to focus on the algorithmic applications rather than calibration of the hardware. Ultimately, as the complexity of the control system will grow with the size of the qubit register, manual calibration of all couplings will no longer be feasible, and automation routines will underpin reliable long-term operation of the system.

At first sight, the problem of calibrating multiple control parameters would not appear difficult if their action on the quantum system could be independently measured and the parameter corrected accordingly. For example, Ramsey spectroscopy in both frequentist [24–26] and Bayesian [24,26–29] forms can be used to determine the mismatch between a qubit transition frequency and the driving field. Indeed, combined with a Rabi frequency measurement [30,31] to determine the applied field strength, single-qubit operations can be efficiently calibrated and traced in time using the minimum number of experimental measurements to correct for drifts [32,33]. However, two-qubit entangling gates often require a more complex combination of driving fields that can have combined effects, which cannot simply be measured individually without assessing the gate performance itself.

As a specific example, we consider the Mølmer-Sørensen (MS) gate [34,35], which is one of the leading implementations of entangling operations in trapped-ion systems. In the following we focus on the case of applying the MS gate on two qubits only, as circuits constructed from two-qubit gates rather than multiqubit gate operations are the most widely pursued approach to fulfill fault-tolerant circuit design properties. The MS gate requires a bichromatic driving field whose two frequency components are symmetrically detuned from the qubit transition of the two ions, naively yielding four independent control parameters (two frequencies and two intensities). However, because the fields are applied simultaneously, the total dynamical ac Stark shift arising from a multilevel atom needs to be compensated using either the frequency or intensity of both fields [36], and the relative phase of the fields starts to play a role [37]. This leads to these parameters being nonlinearly correlated and thus suggests that their calibration be carried out directly by measuring the gate action on a known input state rather than independently estimating each parameter. Such measurements are routinely used in "manual" parameter optimization, whereby the experimentalist has prior knowledge of the expected outcome of an imperfectly calibrated gate, and uses a scheme to iteratively measure and improve these outcomes while changing the experimental control parameters. In the context of the MS gate, this requires knowledge of the multidimensional parameter landscape describing the probability of inducing spin flips, i.e., changes in the electronic population, on the two ionic qubits as a function of the various control parameters. To evaluate the system response to the control parameters, we require an efficient characterization of the gate action. While the gold standard for such a procedure is process tomography [38], it is not without flaws in that it is inherently sensitive to state preparation and measurement (SPAM) errors, is prohibitively slow since the number of required measurements scales exponentially with the number of qubits [39], and can also be problematic when faced with systematic errors [40]. Therefore, recently, a number of alternative techniques have been developed that allow for either faster or more rigorous characterization of the gate performance, using randomized benchmarking [41,42], cycle benchmarking [43], gate set tomography [42,44], and adaptive methods based on Bayesian estimation [45,46]. However, in practice a full characterization of the gate relative to all experimental control parameters is not required to calibrate the gate. An often used practical method to enhance sensitivity to miscalibrated gate parameters is to instead apply a sequence of concatenated identical gates to a single input state and compare the measured electronic populations (spin excitations) to the expected values of the output states, e.g., of Bell states in the calibration of two-qubit gates. The trade-off is that this excitation landscape becomes increasingly complicated in the number N_g of gates, with many local minima, and thus requires a judicious choice of N_g according to the uncertainty on the control parameters. With a large amount of initial uncertainty, calibration begins with $N_g = 1$, but as this uncertainty diminishes, larger N_g values can be used to increase the calibration precision.

In this manuscript we demonstrate an approach to automate this calibration process using a Bayesian estimation technique to simultaneously determine and optimize key control parameters of the two-qubit MS gate Hamiltonian. The motivation behind selecting this approach over gradient descent least squares [47] or machine learning methods [48] is that it aims to reduce the required number of measurements to obtain an accurate parameter estimate while retaining the learnt information when switching between measurements using a different number of gates N_g. Importantly, the Bayesian approach presented here also provides an intrinsic measure of the uncertainty of the estimate, informing us about the progress of the estimation and thereby a quantitative criterion for when to stop the optimization routine once the optimization target has been reached. A requirement for such a Bayesian protocol to work is a precise knowledge of the underlying theoretical model and dominant imperfections. In our case, the Hamiltonian governing the MS gate operation depends on four control parameters, which

we introduce and discuss in detail below: *sideband detuning*, *center line detuning*, *Rabi frequency*, and *phase difference*. The action of this Hamiltonian and the associated noise processes are well understood and map accurately to experimental data, as we show in Sec. II D.

The manuscript is structured as follows. In Sec. II we review the underlying model of the Mølmer-Sørensen gate and introduce the experimentally relevant control parameters that need to be calibrated. In Sec. III we show how we can iteratively estimate the control parameters using a Bayesian protocol. In Sec. IV we describe the strategies for the choice of measurement setting, introduce the termination criterion of the optimization routine, and experimentally evaluate the performance of the algorithm. Section V provides conclusions and an outlook.

II. EXPERIMENTAL SETUP AND MØLMER-SØRENSEN GATE

In this section we describe the trapped-ion setup that we use to generate entangling operations. We briefly review the physics of the MS gate and introduce the main control parameters that need to be calibrated.

A. Experimental setup

The experiments in this manuscript are performed on $^{40}\text{Ca}^+$ ions confined in a microstructured radio-frequency ion surface trap [49]. Qubits are encoded in the computational subspace formed by the $\left|4^2S_{1/2}, m_j = -1/2\right\rangle \equiv \left|g\right\rangle$ electronic ground state and the metastable excited $\left|3^2D_{5/2}, m_j = -1/2\right\rangle \equiv \left|e\right\rangle$ state. The ions form a crystal whose dynamics can be approximated by one-dimensional coupled harmonic oscillators [50,51]. These dynamics can then be decomposed into normal modes. Without loss of generality, we consider only the lowest frequency mode corresponding to the axial center-of-mass (c.m.) motion in our theoretical treatment, and use this mode to mediate the MS interaction. For any practical operating conditions, the frequency difference between the c.m. mode and any of the other modes of an N ion crystal is much larger than the Rabi frequency of the driving field [51], such that we can neglect the coupling to all other modes. For all experiments, the ions are initially Doppler cooled on the $4^2S_{1/2} \rightarrow 4^2P_{1/2}$ transition, followed by sideband cooling of the axial modes. State readout is performed by fluorescence detection with a photomultiplier tube [52].

B. Hamiltonian of the MS gate

The MS gate is a commonly used method to generate entanglement in trapped-ion systems by exploiting a common vibrational mode of the ions. The gate is based on the application of a force that is dependent on the internal state

of the ions. This force induces a periodic movement of the motional state of the ions in phase space. At the end of the gate the motion is returned to its original state, but with an accumulated relative phase in the internal states. This interaction can then be used to create entanglement between the qubits. The gate is experimentally implemented by the application of a bichromatic laser field that introduces four transition paths for a resonant two-photon process between the internal states, each of them going through a virtually excited intermediate state close to a motional sideband, as schematically depicted in Fig. 1(a). This process can be used to perform correlated spin flips mediated by the common vibrational modes of the ions in the trap. The MS gate has desirable properties for a trapped-ion entangling gate, such as the ability to create entanglement between two or more ions with a single operation, and being independent of the initial motional state to first order. This last property guarantees robust functioning of the gate even with imperfect ground-state cooling [34–36]. In combination with single-qubit rotations, the MS gate allows for universal quantum computation [53].

In the following, we review how this gate is implemented and the experimental control parameters that need to be calibrated in order to obtain a high-fidelity realization of the gate operation.

The desired action of the MS gate is, in the general case, an entangling operation acting on N ions of the form

$$U_{\text{MS}}^{\varphi}(\theta) = \exp(-i\theta S_\varphi^2), \tag{1}$$

where

$$S_\varphi = S_y \cos(\varphi) + S_x \sin(\varphi) \tag{2}$$

is the total spin operator in the direction defined by the angle φ and

$$S_\alpha = \frac{1}{2} \sum_{j=1}^{N} \sigma_{\alpha,j}, \quad \alpha = x, y, z, \tag{3}$$

with $\sigma_{x,j}$, $\sigma_{y,j}$, and $\sigma_{z,j}$ being the Pauli operators acting on the qubit encoded by the internal state of the jth ion. By choosing $\theta = \pi/2$ one can use this gate to map computational basis states of N qubits to maximally entangled states.

In order to derive the unitary evolution introduced by the MS gate, let us consider a system of N ions in a linear trap driven by a bichromatic laser of the two frequencies, ω_1 and ω_2. The Rabi frequency, $\Omega(t)$, is assumed to be equal for all ions and for both tones, and can be time dependent for a general pulse shape of the laser.
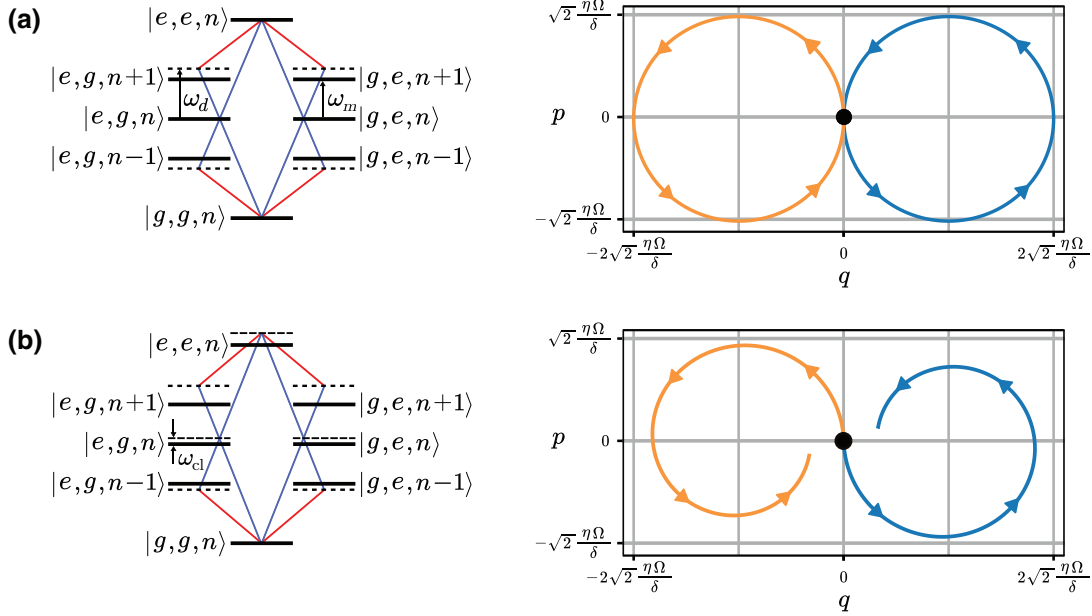
(a)

(b)

FIG. 1. (a) Left: energy level diagram for two ions with quantized center-of-mass vibrational mode of frequency ω_m interacting with a bichromatic laser of frequencies $\omega_1 = \omega_{eg} + \omega_d$ and $\omega_2 = \omega_{eg} - \omega_d$. The laser allows a resonant two-photon transition between $|g,g,n\rangle$ and $|e,e,n\rangle$. There are four different paths, each of them going through an intermediate virtual state separated by $\delta = \omega_m - \omega_d$ from the first motional sidebands ($|g,e,n\pm 1\rangle$, $|e,g,n\pm 1\rangle$). A similar diagram can be drawn for the two-photon transition between the states $|e,g,n\rangle$ and $|g,e,n\rangle$. Right: trajectory in phase space induced by the gate with constant Rabi frequency with the ions in the $+1$ eigenstate of S_φ (in orange) and the -1 eigenstate (in blue), with the black circle indicating the initial motional state. If we apply the gate for the correct amount of time, the trajectory returns to the initial point in phase space. The rotation introduced by the gate is given by the area enclosed by the trajectory. (b) Left: a source of error in the MS gate is when the frequencies of the laser are not centered around the carrier, but instead have a center line detuning $\omega_{cl} = \omega_{eg} - (\omega_1 + \omega_2)/2$, which we consider to be time independent. The larger the frequency ω_{cl} is, the less resonant the two-photon process is, which decreases the fidelity of the gate. Right: similar phase space plot as in the previous case but with $\omega_{cl} \neq 0$. The trajectories for this case are dependent on the initial motional state, where for this figure, we have considered it to be $|n = 0\rangle$.

This system may be described by the Hamiltonian

$$H = H_0 + H_{\text{int}},$$

$$H_0 = \sum_{j=1}^{N} \frac{\omega_{eg}(t)}{2}\sigma_{z,j} + \omega_m(a^\dagger a + 1/2),$$

$$H_{\text{int}} = \sum_{j=1}^{N} \frac{\Omega(t)}{2}(\sigma_j^+ + \sigma_j^-)$$

$$\times [e^{i(\vec{k}_1\vec{x}_j - \omega_1 t + \varphi)} + e^{i(\vec{k}_2\vec{x}_j - \omega_2 t + \varphi)} + \text{H.c.}], \quad (4)$$

where we consider $\hbar = 1$; $\omega_{eg}(t)$ is the transition frequency between the internal states $|e\rangle$ and $|g\rangle$, equal to a bare transition frequency $\omega_{eg,0}$ plus an ac Stark shift, $\omega_{ac}(t) \propto \Omega(t)^2$, due to the interaction of the laser with off-resonant atomic levels [36,54]; ω_m is the frequency of the c.m. mode, which defines the distance of the first motional sideband [Fig. 1(a)] from the carrier; $a^\dagger$ and a are the ladder operators related to the c.m. mode; σ_j^+ and σ_j^- are the Pauli operators acting on the internal states of the jth ion; φ is the phase of the two laser tones, which we consider

to be equal; $\vec{k}_1$ and $\vec{k}_2$ are the wavenumbers of each laser tone.

We can describe the two laser frequencies ω_1 and ω_2 in terms of their symmetric, ω_d, and asymmetric, ω_a, detunings from the atomic transition frequency $\omega_{eg,0}$ (see Fig. 1) as $\omega_1 = \omega_{eg,0} + \omega_d + \omega_a$ and $\omega_2 = \omega_{eg,0} - \omega_d + \omega_a$. The asymmetric detuning as well as the ac Stark shift of the electronic levels create a detuning $\omega_{cl}(t) = \omega_{ac}(t) - \omega_a$ of the mean value of the bichromatic frequencies from the carrier [Fig. 1(b)], known as center line detuning. Having $\omega_{cl}(t) \neq 0$ breaks both the condition $\omega_1 + \omega_2 = 2\omega_{eg}(t)$ of the two-photon resonance and the symmetry between the four paths involved in the gate [see Fig. 1(b)]. Therefore, calibrating $\omega_{cl}(t)$ correctly is essential, and will be a central part of this work. With these definitions of ω_1 and ω_2, the interaction Hamiltonian H_{int} in Eq. (4) becomes

$$H_{\text{int}} = \sum_{j=1}^{N} \frac{\Omega(t)}{2}[e^{i(\vec{k}_1\vec{x}_j - (\omega_{eg,0} + \omega_d + \omega_a)t + \varphi)}$$

$$+ e^{i(\vec{k}_2\vec{x}_j - (\omega_{eg,0} - \omega_d + \omega_a)t + \varphi)} + \text{H.c.}](\sigma_j^+ + \sigma_j^-). \quad (5)$$

We can write $\vec{k}_i \vec{x} = \eta_i(a^\dagger + a)$ with $i = 1, 2$, where η_i is the Lamb-Dicke parameter of the ith tone, and since $\omega_d \ll \omega_{eg}$, we can assume that $\eta_1, \eta_2 \approx \eta$. We can simplify this Hamiltonian by assuming that we are in the Lamb-Dicke regime, $\eta \ll 1$, transforming to the interaction picture, introducing the sideband detuning, $\delta = \omega_m - \omega_d$, and defining

$$\Lambda(t) \equiv \int_0^t \omega_{cl}(t')dt' \qquad (6)$$

that describes an unwanted accumulated phase due to the existence of a center line detuning during the gate operation. As a result, we obtain the Hamiltonian

$$\begin{aligned}\hat{H} = &-\eta\Omega(t)(a^\dagger e^{i\delta t} + a e^{-i\delta t}) \\ &\times [S_y\cos(\varphi + \Lambda(t)) + S_x\sin(\varphi + \Lambda(t))], \quad (7)\end{aligned}$$

where we have used the rotating wave approximation to keep only the terms rotating with δ and ignore the other fast-rotating terms that go with $\omega_m + \omega_d$ or $\omega_{eg,0}$.

In the experiment, we can realize the desired entangling gate from the Hamiltonian in Eq. (7) by adjusting the gate time t_g, sideband detuning δ, center line detuning $\omega_{cl}(t)$, and phase φ. Let us consider first the case $\omega_{cl}(t) = 0$. Here the Hamiltonian can be integrated analytically to obtain the corresponding evolution operator

$$\hat{U}(t) = D[\gamma(t)S_\varphi]\exp[i\theta(t)S_\varphi^2], \qquad (8)$$

where

$$\gamma(t) = i\eta\int_0^t \Omega(t')e^{it'\delta}dt', \qquad (9)$$

$$\theta(t) = \eta^2\,\text{Im}\int_0^t dt'\int_0^{t'} dt''\,\Omega(t')\Omega(t'')e^{-i\delta(t''-t')}, \quad (10)$$

and $D(z) = \exp(za^\dagger - z^*a)$ is the displacement operator. Equation (8) allows us to choose the parameters of the gate in such a way that the result is the maximally entangling MS gate $U_{MS}^\varphi(\pi/2)$.

First, in order to obtain the required entanglement between the qubits, we need to ensure that the rotation angle $\theta(t_g)$ at the end of the application of the gate satisfies $\theta(t_g) = \pi/2$. In the case of a square pulse shape this condition implies that

$$(\eta\,\Omega)^2\,t_g/\delta = \pi/2. \qquad (11)$$

Then, in order to avoid any residual entanglement between the internal states of the ions and their motional state at the end of the pulse, the sideband detuning needs to be

chosen such that the quantity $\gamma(t)$ satisfies $\gamma(t_g) = 0$. This corresponds to closing a loop in phase space [as shown in the example in Fig. 1(a)], ensuring that the motional state returns to the initial state at the end of the gate. In the case of a square pulse shape the condition $\gamma(t_g) = 0$ implies that

$$|\delta| = 2\pi K/t_g \qquad (12)$$

for any integer K representing the number of loops in phase space the gate introduces [see Fig. 1(a)]. In the following we consider a single-loop gate, i.e., $K = 1$.

In the previous discussion, we considered the laser frequencies to be centered around the carrier, that is, $\omega_{cl}(t) = 0$ at all times. In the following we consider a time-independent center line detuning, $\omega_{cl}(t) = \omega_{cl}$, which we aim to set to zero. In the experiment, smooth ramps are used for the pulse shape instead of square pulses. This will lead to a changing ac Stark shift during the ramping and the error appearing from the approximation of a time-independent center line detuning will be given by a nonzero value of $\Lambda(t_g)$. This has the effect of accumulating a phase, $\Delta\varphi = \Lambda(t_g)$ for each application of each consecutive gate, as can be seen from Eq. (7). This cumulative phase shift can be detected by applying the MS gate more than once, and corrected by introducing a phase between consecutive gates.

C. Experimental control parameters

In order to calibrate the MS gate we need to set the gate parameters

$$\boldsymbol{\Theta} = (\Omega, \omega_{cl}, \delta, \Delta\varphi), \qquad (13)$$

where Ω is the Rabi frequency, ω_{cl} the center line detuning, δ the sideband detuning, and $\Delta\varphi$ the phase difference of the gate as close as possible to their optimal values

$$\boldsymbol{\Theta}_{opt} = \left(\Omega_{opt} = \frac{\pi}{\eta t_g}, \omega_{cl,opt} = 0, \delta_{opt} = \frac{2\pi}{t_g}, \Delta\varphi_{opt} = 0\right) \qquad (14)$$

as stated in the previous section, where the condition for Ω_{opt} is obtained from Eqs. (11) and (12). Since we do not have direct access to these parameters, we rely on changing the control parameters of the gate $\boldsymbol{\Xi}_c = (t_g, f_{cl}, f_{sb}, \phi)$, i.e., the gate time t_g, the common frequency $f_{cl} = (f_r + f_b)/2$ is the mean frequency between the red (f_r) and blue (f_b) tones of the laser, the difference frequency $2f_{sb} = f_b - f_r$ of the bichromatic laser field, and the difference in the common phases between consecutive gates, ϕ, of the two laser tones.

These control parameters $\boldsymbol{\Xi}_c$ can be used to calibrate the gate in the following way. By performing measurements of

the populations of the ions after the application of a gate sequence we can obtain estimates that we denote as

$$\overline{\mathbf{\Theta}} = (\overline{\Omega}, \overline{\omega}_{\text{cl}}, \overline{\delta}, \overline{\Delta\varphi}), \tag{15}$$

of the current parameters $\mathbf{\Theta}$. The control parameters $\mathbf{\Xi}_c$ can then be adjusted to set the parameters $\mathbf{\Theta}$ closer to $\mathbf{\Theta}_{\text{opt}}$ given the estimates of the parameters: The time of the gate will be corrected as

$$t_g \rightarrow t_g \frac{\Omega_{\text{opt}}}{\overline{\Omega}}, \tag{16}$$

where we choose to change the time of the gate instead of the laser power since it is easier to control in the experiment, while producing an equivalent correction.

The other corrections are implemented by subtracting the difference between the estimated and the ideal parameter values from the control parameter. In the case of the sideband detuning this is given by

$$f_{\text{sb}} \rightarrow f_{\text{sb}} - \overline{\delta} + \delta_{\text{opt}}, \tag{17}$$

where f_{sb} is the corresponding control parameter. As for the correction of the center line detuning, this is described by

$$f_{\text{cl}} \rightarrow f_{\text{cl}} - \overline{\omega}_{\text{cl}} + \omega_{\text{cl,opt}}. \tag{18}$$

Finally, the phase between consecutive gates is changed by

$$\phi \rightarrow \phi - \overline{\Delta\varphi} + \Delta\varphi_{\text{opt}}. \tag{19}$$

With this set of rules, we update the parameters of the MS gate to iteratively bring the parameters $\mathbf{\Theta}$ closer to $\mathbf{\Theta}_{\text{opt}}$.

D. Validating the MS model

Next, we aim to verify that the theoretical description of the gate action agrees with the experiment. Since we are focusing on the effects of systematic parameter miscalibration, we are neglecting other error sources such as finite motional and spin coherence times, laser amplitude noise, resonant carrier excitation, or unequal coupling strengths to the ions. As we target two qubit gates, all the following results presented are measured or calculated using a two ion crystal.

In order to validate the Hamiltonian in Eq. (7) as a model, we measure the outcome probabilities of a gate sequence while individually varying each control parameter. The remaining parameters are kept constant at their optimal values determined by manual calibration of the gate (Fig. 2).

The laser pulse is switched on adiabatically using a Blackman-like shape [55], with a 4 μs shape time at the beginning of the pulse to slowly increase the laser power to its maximum. The same shape is used in reverse during switch-off. The shaping ensures adiabaticity during switch-on of the laser, preventing unwanted excitation of the carrier transition. The gate time t_g is defined as the duration of the laser pulse of the full width at half maximum.

We compare the measurements to the expectation values obtained from numerically integrating the Hamiltonian using the QuTiP software package [56]. Comparing the measurement results to numerical simulations with the same parameters, we calculate reduced χ^2 values of < 2.5 by varying the center line detuning, sideband detuning, and phase, while the Rabi frequency scans for $N_g = 5$ and $N_g = 7$ gates have higher χ^2 values of up to 4. We attribute this to the ac Stark shift not being recompensated for these measurements as the actual laser power is varied instead of adjusting detuning and gate time to control the Rabi frequency Ω. While the χ^2 are larger than one, these results do indicate that Hamiltonian (7) captures the effects of the parameter miscalibrations and is not dominated by unmodeled error sources. This suggests that the Hamiltonian is a viable model to describe our experimental system, and the results from numerically integrating the dynamics can be used as a probability landscape to perform Bayesian inference on.

III. BAYESIAN INFERENCE

In order to estimate the values of the experimental control parameters that yield a high-fidelity MS gate implementation, we need to determine their relation to the gate parameters $\mathbf{\Theta}$. For this purpose, we apply a sequence of gates and perform a state readout to receive a measurement result. We are employing Bayesian inference for interpreting these measurements, since this framework allows us to straightforwardly incorporate information about our system obtained from measurement results using different gate sequences and control parameters $\mathbf{\Xi}_c$, experimental settings, or previous calibration measurements. Additionally, this approach allows one to quantify the uncertainty on the estimated parameters, allowing us to monitor the progress of the calibration procedure and terminate the algorithm once the estimates reach sufficient accuracy.

Bayes' theorem [57] prescribes how to estimate the probability distribution of a set of parameters $\mathbf{\Theta}$, given a prior distribution before the measurement $P(\mathbf{\Theta})$, a measurement with outcome m, and the likelihood of obtaining this measurement outcome, $P(m|\mathbf{\Theta})$, given a model of the process. The result is an updated posterior probability distribution

$$P(\mathbf{\Theta}|m) \propto P(m|\mathbf{\Theta})P(\mathbf{\Theta}) \tag{20}$$

up to a normalization factor [58]. If we perform a new measurement, we can apply Bayes' theorem again using the obtained posterior as a prior for the next measurement. This defines an iterative process (see Fig. 3) in which the
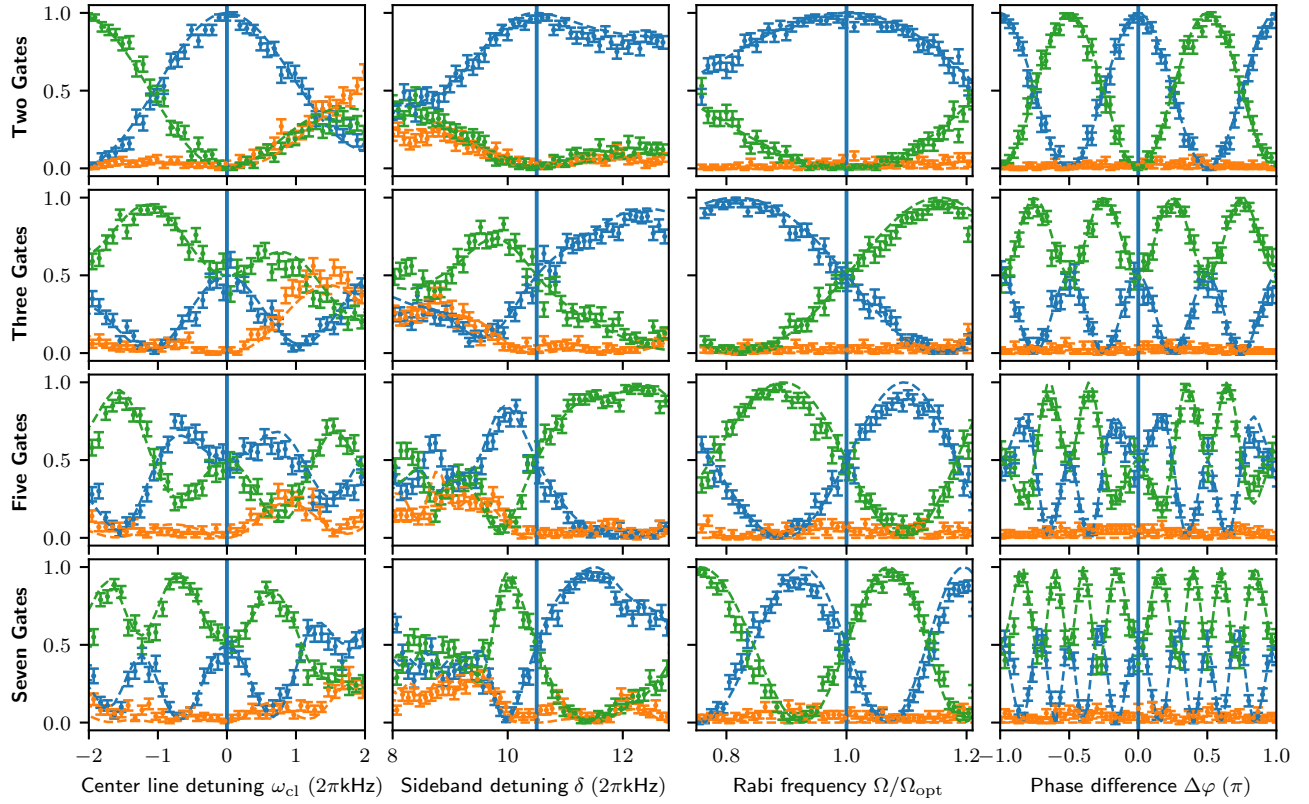
FIG. 2. Populations of $|g, g\rangle$ (green), of $|e, e\rangle$ (blue), and of $|e, g\rangle$ and $|g, e\rangle$ (orange) after the application of two, three, five, and seven gates with varying control parameters. For each scan, the other parameters have been held constant at the optimal value for the MS gate. The error bars are estimated from shot noise. Dashed lines indicate the numerical simulations. The vertical lines mark the target value for each parameter.

probability distribution after T measurements is given by

$$p(\boldsymbol{\Theta}|m_1, \ldots, m_T) \equiv p(\boldsymbol{\Theta}|m_{1:T}) \propto P(\boldsymbol{\Theta}) \prod_{t=1}^{T} P(m_t|\boldsymbol{\Theta}), \tag{21}$$

where we assume that all the measurement outcomes, m_t for $t = 1, \ldots, T$, are independent from each other. For an increasing number of measurements, we expect the degree of uncertainty of the parameters $\boldsymbol{\Theta}$ to decrease, and the process can be terminated if a desired limit of uncertainty for the estimates of $\boldsymbol{\Theta}$ is reached.

In order to calculate the posterior probability distribution of the parameters of the Mølmer-Sørensen gate using Eq. (20), we need to know the probability of obtaining any measurement outcome m for any given set of parameters $\boldsymbol{\Theta}$. These probabilities $P(m|\boldsymbol{\Theta})$ can be obtained by integrating the Hamiltonian in Eq. (7). However, for $\omega_{cl} \neq 0$, no analytical solution is known that results in a closed-form expression for these probabilities, thus requiring to numerically study the effects of this type of miscalibration [36]. As a consequence, we investigate a numerical treatment of the probability distributions. In the following

subsection we describe how the continuous probability distributions can be represented by discretized distributions instead. Afterwards, we explain how these discretized distributions can be used to implement a Bayesian inference. A visual representation of our protocol is given in Fig. 3.

A. Representation of probability distributions

We approximate the probability distributions using a so-called particle filter [59–62], where we replace the continuous probability function by a sum of N_p weighted particles:

$$p(\boldsymbol{\Theta})d\boldsymbol{\Theta} \approx \sum_{i=1}^{N_p} w^{(i)} \delta(\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}_i) d\boldsymbol{\Theta} \tag{22}$$

with $\delta(\boldsymbol{\Theta} - \widetilde{\boldsymbol{\Theta}}_i)$ being the Dirac delta function at the set of parameters $\widetilde{\boldsymbol{\Theta}}_i$. Each parameter set $\widetilde{\boldsymbol{\Theta}}_i$ is represented by a single particle with an associated weight $w^{(i)}$, and the set $\{\widetilde{\boldsymbol{\Theta}}_i, w^{(i)}\}_{i=1}^{N_p}$ forms the particle filter. The weights $w^{(i)}$ satisfy $\sum_{i=1}^{N_p} w^{(i)} = 1$, ensuring that the probability distribution is normalized.

The Bayesian inference requires an initial prior, and we must choose an initial set of particles to represent this prior.
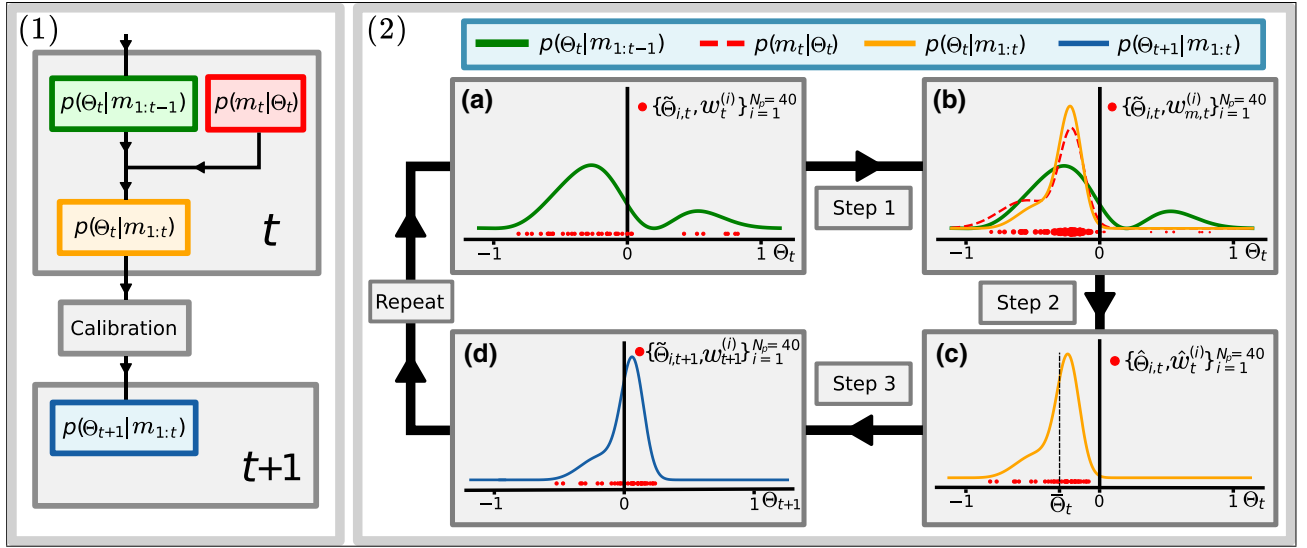
FIG. 3. (1) Representation of a cycle of the calibration process. The initial probability distribution at iteration t with $t-1$ measurements, $p(\Theta_t|m_{1:t-1})$, is updated by performing a new set of measurements, m_t, to obtain $p(\boldsymbol{\Theta}_t|m_{1:t})$. The estimates obtained from this probability distribution can then be used to calibrate the parameters, obtaining the initial probability distribution for iteration $t+1$. (2) Steps followed to perform an iteration of the particle filtering algorithm for a simplified case with only one parameter, Θ. In (a) we have particles $\{\widetilde{\Theta}_{i,t}, w_t^{(i)}\}$ (red dots) from the prior probability distribution $p(\Theta_t|m_{1:t-1})$ (green). A new set of measurements is performed to obtain the posterior probability distribution $p(\Theta_t|m_{1:t})$ (orange) in (b) by using the likelihood of the outcome obtained for the measurement, $p(m_t|\Theta_t)$ (red). The weights are updated following Eq. (23), where a larger weight is visually represented by a larger size of the particle. These weighted particles can be used to obtain the estimate of the parameter at this iteration, $\bar{\Theta}_t$, by using Eq. (24). In (c), a resampling process is performed to obtain particles with equal weights that approximate particles from the posterior probability distribution. The previously obtained value $\bar{\Theta}_t$ is also shown here. Finally, in (d), $\bar{\Theta}_t$ has been used to calibrate the parameter Θ, where for this example, we are assuming that the ideal value of Θ is zero; therefore, the correction applied is to change the value of the parameter by $-\bar{\Theta}_t$. The probability distribution $p(\Theta_t+1|m_{1:t})$ (blue) consequently has an estimate of 0.

We choose a known continuous probability density function as prior. We can then initialize the particle filter from that probability function by randomly sampling N_p times from the distribution and setting all the weights to $1/N_p$.

The density of this four-dimensional probability function describes the probability for a particular combination of Rabi frequency, Ω, center line detuning, ω_{cl}, sideband detuning, δ, and phase difference, $\Delta\varphi$, to describe the parameters of the entangling MS gate.

B. Bayesian update

The Bayesian update for the prior distribution, approximated by the particle filter $\{\widetilde{\boldsymbol{\Theta}}_i, w^{(i)}\}_{i=1}^{N_p}$, after performing a measurement [step 1 in Fig. 3(b)] is implemented by updating the weights as given by

$$w_m^{(i)} \propto P(m|\widetilde{\boldsymbol{\Theta}}_i)w^{(i)}. \tag{23}$$

The weights $w_m^{(i)}$ are normalized such that $\sum_{i=1}^{N_p} w_m^{(i)} = 1$, ensuring that the filter represents a valid probability distribution.

The probabilities $P(m|\widetilde{\boldsymbol{\Theta}}_i)$ required to perform the update are obtained by numerical integration of the Hamiltonian in Eq. (7) at each discrete sample point. Performing these numerical calculations in real time during an optimization run would, however, be prohibitively time consuming. The integration for a single point requires approximately 1 s on a CPU [63], which would result in a computation time of several hours to update the 10 000 particles that we use in our particle filter (see Sec. IV D).

Instead, we precompute the outcome probabilities of a single experimental shot on an equally spaced four-dimensional grid. The outcomes between those grid points can then be calculated using an interpolation function. We use a cubic spline interpolator [64], which allows us to better approximate the outcome probabilities without increasing the number of grid points required compared to using linear interpolation. For multiple gates, the underlying grid uses 21 points for $\Omega, \omega_{\mathrm{cl}}, \delta$ and 25 for $\Delta\varphi$. The ranges of points are $\Omega = (1 \pm 0.5)\,\Omega_{\mathrm{opt}}, \omega_{\mathrm{cl}} = 0 \pm 3.5 \times 2\pi\,\mathrm{kHz}, \delta = 10 \pm 5 \times 2\pi\,\mathrm{kHz}, \Delta\varphi = 0 \pm \pi$ for a gate duration of 100 μs. For the single gate grid, we use a wider spacing with $\omega_{\mathrm{cl}} = 0 \pm 7 \times 2\pi\,\mathrm{kHz}, \delta = 10 \pm 10 \times 2\pi\,\mathrm{kHz}$, and $\Omega = (1 \pm 0.5)\,\Omega_{\mathrm{opt}}$ is chosen identical to the previous grid, while the modeled outcomes are held

constant for all $\Delta\varphi$ since the phase difference has no influence in a single gate experiment.

The interpolation function can then be queried to receive the outcome probabilities at any set of parameters inside the region, with 10 000 particles taking approximately 100–500 ms to compute [65]. This computation is now faster than the time required to acquire the experimental data from an iteration (see Sec. IV C). We additionally introduce a small amount (1%) of depolarizing noise into our model to account for experimental errors, in particular SPAM errors. This is necessary as noise is always present, preventing the expectation values $P(|e,g\rangle)$ and $P(|g,e\rangle)$ from going to 0, while the perfect model predicts arbitrarily small expectation values. The added depolarizing noise limits how much the likelihood can be adjusted by a measurement of $|g,e\rangle$ or $|e,g\rangle$.

C. Parameter estimation

The particle filter allows us to straightforwardly extract statistical information from the distribution, most notably the mean and the variance of the probability distribution,

$$\overline{\mathbf{\Theta}} = \sum_{i=1}^{N_p} w_m^{(i)} \widetilde{\mathbf{\Theta}}_i, \tag{24}$$

$$\mathrm{Var}(\mathbf{\Theta}) = \sum_{i=1}^{N_p} w_m^{(i)} (\widetilde{\mathbf{\Theta}}_i - \overline{\mathbf{\Theta}})^2. \tag{25}$$

These estimates can be used during the algorithm to monitor progress and inform decisions about the next experimental setting (see Sec. IV B).

D. Resampling

We are expecting the variance of the probability distributions to decrease as information from more measurements is added, improving the estimates of the parameters. This poses a known problem for the particle filter as we repeatedly update the weights, since it leads to many particle weights going to zero while simultaneously having only a few particles with high weights. This situation causes both unnecessary computations to update the weights on particles that represent a small probability and thus contribute little to the estimates, and undersampling of the distribution around the particles with high probability, limiting the final precision of the estimate. We counteract this problem by a so-called resampling procedure [66,67], where we generate a new set of particles with equal weights to represent the same probability distribution (step 2 in Fig. 3):

$$\{\widetilde{\mathbf{\Theta}}_i, w_m^{(i)}\}_{i=1}^{N_p} \rightarrow \{\hat{\mathbf{\Theta}}_i, \hat{w}^{(i)} = 1/N_p\}_{i=1}^{N_p}. \tag{26}$$

We use the Liu-West algorithm [68] for this task, which first generates a new set of particles by randomly sampling with replacement from the old weighted particles. This already generates a set of equally weighted particles, but particles cannot explore new locations since they are up to this point duplicates of particles of the old set. To lift this degeneracy, every particle of the filter is moved towards the mean of the filter by a constant. A random perturbation is then applied to each particle, with the values of the perturbation sampled from a normal distribution. The covariance of this normal distribution is chosen such that the mean and covariance of the particle filter is preserved.

E. Feedback

The last step (step 3 in Fig. 3) remaining in our parameter estimation is to transform the posterior probability distribution into the prior distribution for the next iteration. We use the estimates from the particle filter [Eq. (25)] to calculate the required adjustments of our experimental control parameters for a perfect gate using Eqs. (16)–(19), given the current knowledge of our parameter estimates [Eq. (15)].

We also apply these corrections to the positions of our particles such that the expectation values of the new prior fulfill the relations for a perfect gate. These new values can now be used in the next iteration of the algorithm, thus improving the estimates of the parameters iteratively. An example of this behavior estimating the four MS gate parameters is shown in Fig. 4, where the probability distributions are mapped to the experimental control parameters t_g, f_{sb}, f_{cl}, and ϕ. At each update, 100 experimental shots are performed using the estimates for the control parameters by the previous iteration and the particle filter is then updated according to the measurement results. Over several iterations the probability distribution narrows, reducing the uncertainty on the estimates (shown in orange), and the changes to the control parameters become smaller between iterations.

IV. CALIBRATION ALGORITHM

We now proceed to use the Bayesian estimation framework introduced previously to calibrate all four key parameters that determine the performance of the Mølmer-Sørensen gate in our experiment. We introduce a stopping criterion that relates the parameters to gate infidelities. We describe the selection process of the measurement settings to improve convergence of the Bayesian parameter estimation protocol and investigate the effect on the experimental run time of the algorithm. The choice of measurement setting determines the amount of information that will be gained, and can have significant effects on the number of measurements required [27,46,69–71]. We finally experimentally verify the performance of the algorithm
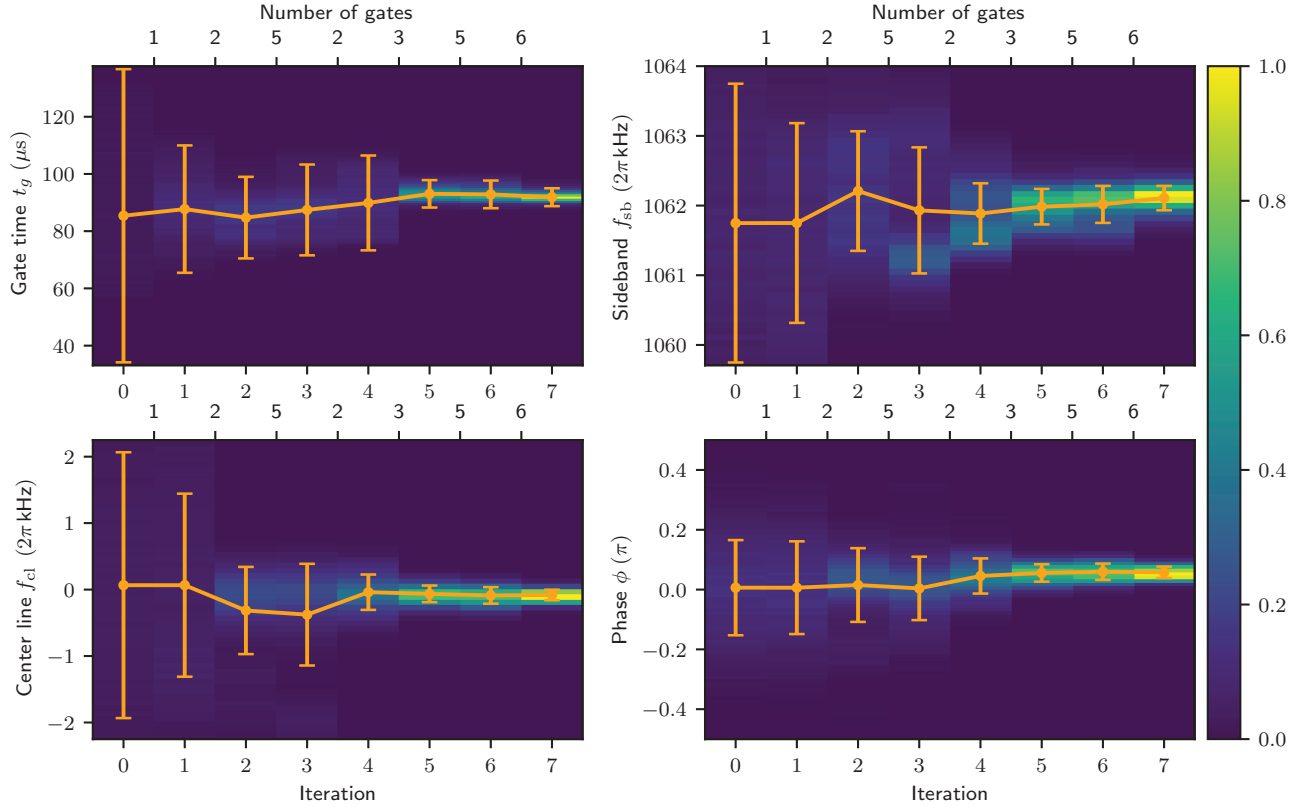
FIG. 4. Projections of the particle filter for the control parameters $\mathbf{\Xi}_c$ at each iteration of a single calibration run. The mean and standard deviation of the probability distribution of the control parameters calculated using the relations in Eqs. (16), (18), (17), and (19) is shown in orange. The probability density of each plot is normalized to its maximum. Each iteration uses the measurement outcomes of 100 repetitions of the experiment to update the probability density. As an initial prior, a Gaussian with widths of $\sigma_\Omega = 0.2\,\Omega_{\text{opt}}$ Rabi frequency uncertainty, $\sigma_{\omega_{\text{cl}}} = 2 \times 2\pi\,\text{kHz}$ center line uncertainty, $\sigma_\delta = 2 \times 2\pi\,\text{kHz}$ sideband uncertainty, and $\sigma_{\Delta\varphi} = 0.16\pi$ phase uncertainty is chosen, which corresponds to typical experimental uncertainties after preliminary calibration. The marks at the top indicate the length of the gate sequence used for the measurements, chosen according to the variance minimization strategy described in Sec. IV B.

by checking the consistency of the final gate parameters returned by the algorithm.

A. Stopping criterion for the algorithm

We want to determine the parameters of the gate with sufficient accuracy to perform operations with a target fidelity. The calibration procedure thus requires a stopping criterion to decide when the routine has reached the necessary accuracy. Entangling gates are usually not quantified by the uncertainty in their control parameters, but by their fidelity [72] compared to the ideal expected output state. By averaging fidelities obtainable for various input states and unitary gate sequences, a more relevant quantity for performing an algorithm, the average gate error, can be extracted [73–75]. The fidelity of any real gate is limited by noise processes. It is thus sufficient to determine the parameters to an accuracy such that the present stochastic noise is the limiting factor of the gate fidelity. In order to relate miscalibrated parameters to gate fidelity, we

simulate randomized cycle benchmarking [43] with misset parameters. We consider the effect of the parameters individually, with the resulting infidelities shown in Fig. 5. For each parameter, we can then define a threshold T_θ that defines an acceptable region in which a miscalibration is no longer expected to significantly influence the gate performance. In our case we choose a threshold of 2×10^{-3} for the infidelity caused by a miscalibration of that single parameter, leading to a choice of thresholds on the individual parameters of $T_\Omega = 0.02\,\Omega_{\text{opt}}$, $T_{\omega_{\text{cl}}} = 150 \times 2\pi\,\text{Hz}$, $T_\delta = 200 \times 2\pi\,\text{Hz}$, $T_{\Delta\varphi} = 0.028\pi$. We proceed to run the calibration algorithm until the particle filter converges in all four dimensions to an uncertainty below the thresholds. Assuming that the parameters are normally distributed with the standard deviation of the distributions equal to the thresholds, we expect from simulation a median infidelity of approximately 5×10^{-3}, but we expect that correlations between the errors in the parameters can significantly affect the expected infidelity (see Sec. IV E).
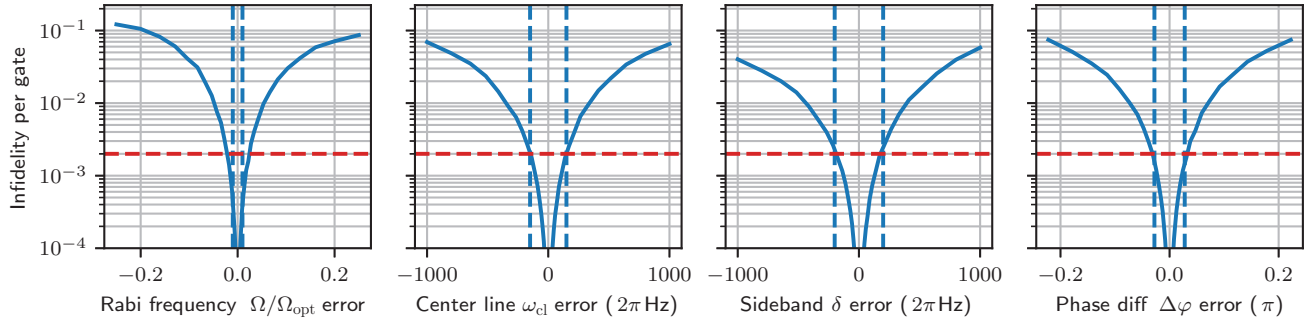
FIG. 5. Expected randomized benchmarking infidelity as a function of individual parameter miscalibrations for a $100\,\mu s$ gate. These curves inform the choice of target thresholds (blue dashed lines); an infidelity of 2×10^{-3} (red dashed line) for each parameter has been chosen as a threshold. The fixed parameters are kept at $\Omega = \Omega_{\mathrm{opt}}$, $\omega_{\mathrm{cl}} = 0$, $\delta = 10 \times 2\pi\,\mathrm{kHz}$, $\Delta\phi = 0$.

B. Selection of measurement settings

We have a choice of the sequence of gates for which we want to perform a measurement. Different gate sequences affect the control parameters differently. As an extreme example, a single gate has no dependency on the phase difference $\Delta\varphi$; thus, a measurement of a single gate does not provide any information about that parameter. We thus require a strategy for selecting suitable measurement settings that ideally maximize the amount of information gained from performing a certain measurement.

We restrict our considerations to sequences of MS gates without any local operations. Besides the number of gates N_g we can also intentionally introduce a phase difference $\Delta\varphi_{\mathrm{target}}$ between the gates. We restrict the phase difference settings to either $\Delta\varphi_{\mathrm{target}} = 0$ or $\Delta\varphi_{\mathrm{target}} = \pm\pi/4$, corresponding to either consecutive $U_{\mathrm{MS}}^0(\pi/2)$ gates or a sequence of $U_{\mathrm{MS}}^0(\pi/2)\, U_{\mathrm{MS}}^{\pm\pi/4}(\pi/2) \cdots \mathrm{MS}_{\pm(N_g-1)\pi/4}$ $(\pi/2)$ gates. We choose these two types of sequences as they change which parameters the sequence is most sensitive to (Fig. 6).

Applying more gates increases the sensitivity of the sequence as the peaks of the likelihood function $P(m|\Theta)$ become narrower, resulting in an increased first-order sensitivity to miscalibrations. At the same time, the peaks of the likelihood function become more closely spaced, which for a wide prior can lead to a multimodal posterior distribution that causes slow convergence. Additionally, the effect of unmodeled error sources such as decoherence increases with additional gates, limiting the total length of the sequence used in the optimization. We thus need to choose the number of gates N_g we apply, as well as the targeted phase difference between consecutive gates.

These gate sequences form a set of possible measurement settings. We work with two approaches for choosing from this set, which we present in the following subsections.

1. Variance minimization strategy

To maximize the information gained on our parameters from a measurement, we aim at performing a measurement that decreases the variance of our posterior distribution as much as possible. For each measurement setting s, we can predict the posterior estimate $\overline{\Theta}_{m_j,s}$ and variance $\mathrm{Var}(\Theta_{m_j,s})$ for each measurement outcome m_j by first calculating the weights for each possible outcome of the measurement with Eq. (23) using the corresponding
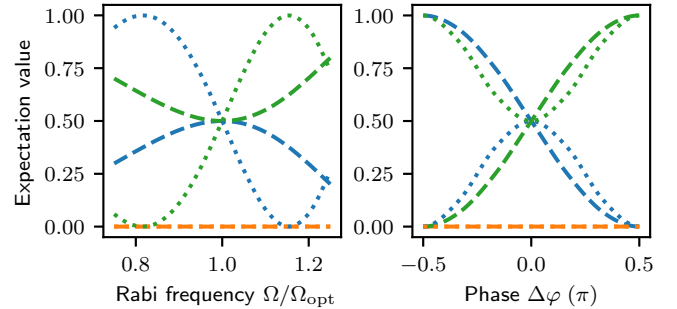

FIG. 6. Expected outcome probabilities of $P(|g,g\rangle)$ (green), $P(|e,e\rangle)$ (blue), and $P(|e,g\rangle) + P(|g,e\rangle)$ (orange) for a $3\,U_{\mathrm{MS}}^0(\pi/2)$ gate sequence (dotted) and for a $U_{\mathrm{MS}}^0(\pi/2)U_{\mathrm{MS}}^{\pi/4}(\pi/2)$ sequence (dashed). The first gate sequence is first-order insensitive to phase miscalibrations as the local minimum of the outcome probabilities around zero leads to only small variations in the likelihood of possible outcomes, and its symmetry around zero does not allow one to discriminate the sign of the miscalibration. Similar arguments can be used to see that this sequence is first-order sensitive to Rabi frequency miscalibrations. For the second gate sequence, the relation is inverted, being sensitive to phase miscalibrations at the cost of first-order insensitivity to the Rabi frequency. The qualitative behavior of the sensitivity of the center line matches the sensitivity of the phase difference, while the behavior of the sideband detuning matches the Rabi frequency.

probabilities $P_s(m_j|\mathbf{\Theta}_i)$:

$$\overline{\mathbf{\Theta}}_{m_j,s} = \sum_{i=1}^{N_p} \mathbf{\Theta}_i P_s(m_j|\mathbf{\Theta}_i)w^{(i)}, \qquad (27)$$

$$\text{Var}(\mathbf{\Theta}_{m_j,s}) = \sum_{i=1}^{N_p} (\mathbf{\Theta}_i - \overline{\mathbf{\Theta}}_{m_j,s})^2 P_s(m_j|\mathbf{\Theta}_i)w^{(i)}. \qquad (28)$$

We can estimate the total probability $P_s(m_j)$ of the outcome m_j for setting s by averaging over the particle filter

$$P_s(m_j) = \sum_{i=1}^{N_p} P_s(m_j|\mathbf{\Theta}_i)w^{(i)}. \qquad (29)$$

By weighing the variance with the expected probability for that outcome we can calculate the most likely variance $\text{Var}(\mathbf{\Theta}_s)$ for a given measurement setting given the current knowledge before actually measuring by averaging over all possible measurement outcomes m_j,

$$\text{Var}(\mathbf{\Theta}_s) = \sum_j \text{Var}(\mathbf{\Theta}_{m_j,s})P_s(m_j). \qquad (30)$$

We aim to minimize our variance to improve the estimate of all our parameters, which is not straightforward for four separate variances for the parameters, as different measurement settings do not affect all parameters equally. We thus employ a heuristic to calculate a score for each measurement setting for which we normalize each variance with its target threshold T_θ defined in Sec. IV A and sum up all the normalized variances to get a score X_s for the measurement setting,

$$X_s = \sum_{\theta \in \mathbf{\Theta}} \frac{\text{Var}(\theta_s)}{T_\theta^2}. \qquad (31)$$

We then apply the measurement setting with the lowest score to the experiment. The process is illustrated for one dimension in Fig. 7.

We need to calculate the outcome probabilities at each particle location for each measurement setting, which is computationally expensive. We interpolate probabilities from precomputed simulations to speedup this calculation, but the interpolation remains the most time-intensive operation in our classical computations. While calculating the probabilities is an unavoidable step for performing the Bayesian update of the particle filter as discussed in Sec. III B, predicting the variances requires running this calculation for each measurement setting under consideration. This significantly increases the total classical computational overhead.

2. Thresholded strategy

In order to reduce the computational cost of finding the next measurement setting we present a simple heuristic as an alternative to the variance minimization strategy described above. For this, we want to look at which settings are considered optimal by the variance minimization strategy. In practice, this means selecting the setting that uses the most gates while avoiding a multimodal posterior distribution. These multimodal posteriors are caused by the likelihood function having two maxima over the range of the prior. To avoid them, ideally the chosen measurement settings would yield outcome probabilities that have only one likely parameter value. While it is impossible to simultaneously fulfill this for all parameters, we can still attempt to limit the amount of possible maxima of the likelihood function with the right choice of measurement setting. To accomplish this, we want to impose a maximal requirement on the width of the prior before using a specific measurement settings.

We examine the expected outcomes of the model restricted to one dimension for each parameter and for each experimental setting, while the other parameters are set to their optimal values. We then choose the local extremum closest but not equal to the optimal value in the outcome probabilities, and define a threshold as the distance of the extremum to the center for each measurement setting. While this ignores possible correlations of the probability distribution between the parameters, we have found in practice that these thresholds are already sufficient to avoid multimodal posteriors.

We split the settings into first-order Rabi frequency Ω-sensitive settings and into first-order phase difference $\Delta\varphi$-sensitive settings depending on whether the $|e,e\rangle$ and $|g,g\rangle$ populations have a local extremum exactly at the optimal parameter value or a linear slope in either phase or Rabi frequency, respectively (see Fig. 6).

These thresholds are now used to define the measurement strategy that will be used by the calibration algorithm: on each iteration we alternate between selecting from either Rabi frequency or phase first-order-sensitive settings, comparing the defined thresholds to the current variances to ensure that the variance of both Rabi frequency Ω and phase difference $\Delta\varphi$ are being improved. We now compare the variance of our prior with the thresholds, and discard any setting for which the variance exceeds a threshold in any parameter. Among the remaining settings we pick the setting with the highest number of MS gates (for an example, see Fig. 7). This gives us a way to select measurement settings while avoiding the computational overhead incurred by the variance minimization strategy, with the time needed to evaluate the strategy ($\ll 1$ ms) being negligible compared to the time required to update the particle filter (Sec. III B).
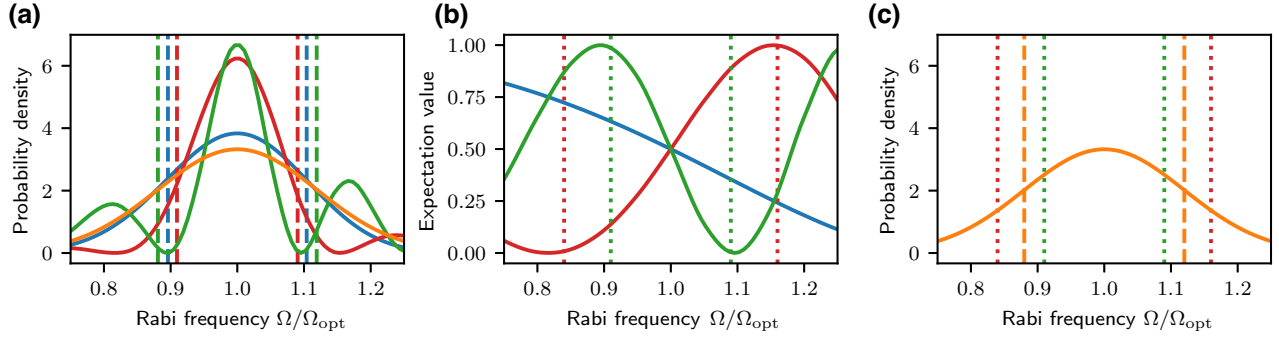
FIG. 7. Illustration of the decision-making process on which gate sequence to use in the next measurements. Plot (a) illustrates the variance minimization strategy for the prior given in orange and the expected posterior probability distributions calculated using the expected outcomes. We compare the standard deviations (dashed lines) of the expected posteriors for each setting under consideration ($N_g = 1$ blue, $N_g = 3$ red, $N_g = 5$ green with $\Delta\varphi_{target} = 0$) to select the smallest for the next measurement setting. In this example the deviation for three gates is the narrowest and is thus chosen. Plot (b) shows the expected outcome probability $P(|g,g\rangle)$ for the $|g,g\rangle$ state according to the model for each setting ($N_g = 1$ blue, $N_g = 3$ red, $N_g = 5$ green with $\Delta\varphi_{target} = 0$) depending on the Rabi frequency Ω. In this case the outcome probability for $P(|e,e\rangle) = 1 - P(|g,g\rangle)$ with population in the remaining states being zero. The dotted lines show the thresholds used for the thresholded strategy. In (c) these thresholds are used to evaluate the same prior. We only consider the standard deviation of the prior (dashed lines), with the width of the prior being larger than the five-gate threshold (green) but below the three-gate threshold (red), leading to the three-gate setting being chosen. The single gate setting does not have any thresholds, being the fallback option if no other setting passes the thresholds.

C. Run time of the algorithm

A single experimental shot only takes approximately 10 ms, while computing the particle filter update requires approximately 100–500 ms. To balance the classical computational effort for updating the particle filter and the latency in updating the control system with the duration of the experiment, we run 100 experimental cycles with the same measurement setting before updating the particle filter and choosing the next experimental setting. The variance minimization strategy additionally needs time (approximately 0.6–3 s) for selection of the next measurement setting, while the time required to evaluate the thresholded strategy is negligible.

In Fig. 8(a) we show the number of individual experiments required to reach the target thresholds, giving a comparable number of experimental shots for the variance minimization strategy (1200 ± 500 shots) and the thresholded strategy (1100 ± 500 shots). Figures 8(b) and 8(c) display the measurement settings used during these calibration runs. The variance minimization strategy focuses on either the phase or the Rabi frequency sensitive setting, trying to reduce the largest variance component before switching. The thresholded strategy instead alternates between the settings of different sensitivities. Both settings increase the number of gates on later iterations as the variances decrease. While the order of which measurement setting is chosen is different between the strategies, the number of experimental shots required is comparable.

From this we can conclude that the thresholded strategy is a good heuristic to select measurement settings compared to the variance minimization strategy.

Taking the computational overhead into account, the thresholded strategy has the advantage, completing with average run time of 41 ± 17 s compared to an average of 60 ± 26 s for the variance optimization. Averaged over both strategies we find that we require 1200 ± 500 experimental shots to reach our target thresholds. A strict quantitative comparison to all other possible optimization strategies is difficult to define. Using our Bayesian algorithm, we achieve a significant speedup compared to manual iteration over one-dimensional (1D) parameter scans as defined in Ref. [25]. Simulations of 1D parameter scans achieve worse residual infidelities, while requiring significantly more measurements. This comparison is fully discussed in the Appendix.

D. Capture range

Any Bayesian procedure requires a choice of an initial prior, where in our case we consider normal distributions. It is reasonable that if the true values are unlikely given our initial prior, that is, if the true values are too far away from the center of our initial prior, then the routine will fail to find them. In this section we study the dependence of the failure rate of our algorithm on this initial distance in parameter space. We also study the dependence of the success rate of the algorithm on the number of particles in the particle filter, since a larger number of particles allows for a more accurate sampling and, therefore, a better approximation of the probability distributions under study. This increase in the number of particles comes with the cost of a linear growth in the classical computing time required by the algorithm.
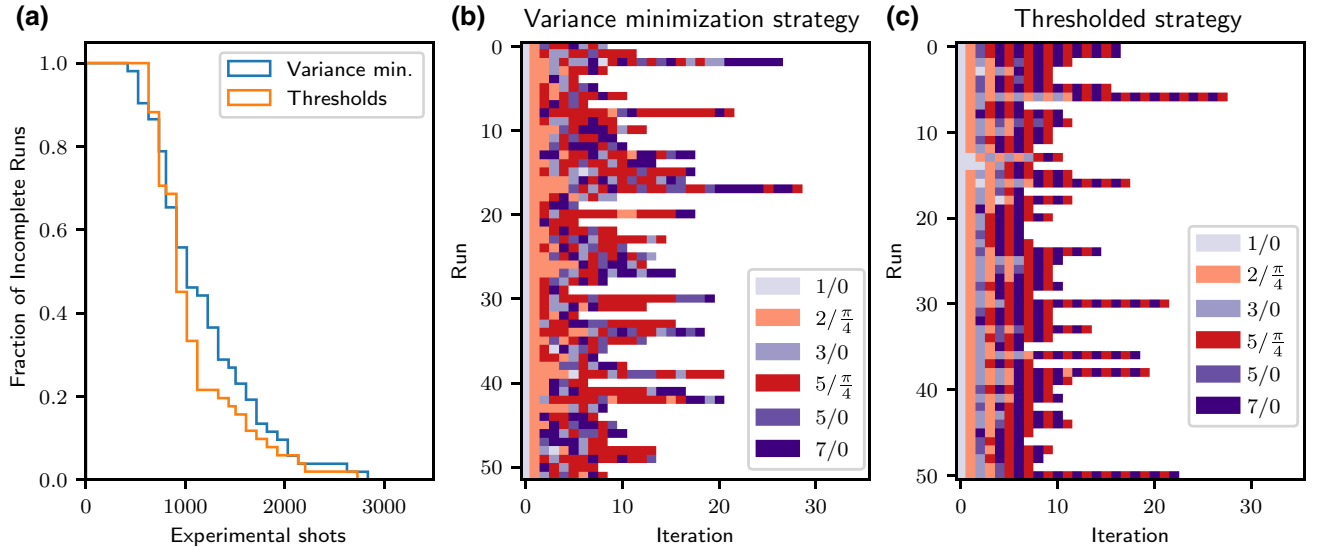
FIG. 8. The number of experimental cycles required to reach the target thresholds is shown as a fraction of all runs in (a). The thresholded strategy takes 1100 ± 500 cycles on average compared to 1200 ± 500 for the variance minimization strategy. Panels (b) and (c) show the measurement settings used during the calibration runs for the respective strategies, with the y axis corresponding to independent runs and the x axis to iterations of the algorithm during a calibration run; at each iteration 100 measurements are performed with the chosen setting. The colors indicate the measurement setting used at a given iteration of a run, characterized by the number of gates N_g (first number in the legend) and the selected phase difference $\Delta\varphi_{\text{target}}$ (second number) used for the gate sequence.

We study these effects by first obtaining a normalized distance, D_{start}, of the initial values of the parameters from their target values using the widths of the prior

$$D_{\text{start}}^2 = \sum_{\theta \in \Theta} \frac{(\theta_{\text{start}} - \theta_{\text{opt}})^2}{\sigma_\theta^2}, \tag{32}$$

where σ_θ^2 is the initial variance of the prior projected onto parameter θ, θ_{start} is the starting value, and θ_{opt} is the optimal value for that parameter.

We generate random combinations of initial gate time, center line frequency, and sideband frequency with the chosen starting distance and then run simulated calibration runs. The phase is not varied as the space is periodic, and thus limits how far the phase values can be from the optimal value. The fraction of calibration runs that converged is shown in Fig. 9, where we consider any run that ends up closer than twice the termination thresholds of the stopping criterion (Sec. IV A) in all parameters a success. Since, for each parameter, the estimated error is below the threshold, for a Gaussian distribution, at least 95% of results should be between these thresholds. Because of the statistical nature we do not expect a perfect success rate even if the particle filter were to perfectly approximate the continuous probability function.

We find that, for 2000 particles and less in the particle filter, the capture range is reduced as compared to the higher particle numbers, and performance is decreased even at a starting distance of 1 compared to the higher particle numbers. This is expected as a particle filter with a

too low number of particles will lead to an undersampling of the probability distribution, which in turn results in a poor approximation of this probability distribution. As a
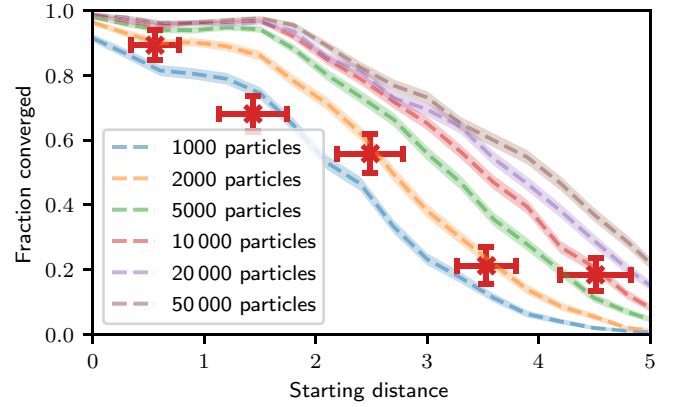

FIG. 9. Capture range of the algorithm. The width of the initial Gaussian prior has been held constant at $\sigma_\Omega = 0.2 \ \Omega_{\text{opt}}$ Rabi frequency uncertainty, $\sigma_\delta = 1 \times 2\pi \, \text{kHz}$ sideband uncertainty, $\sigma_{\omega_{\text{cl}}} = 1 \times 2\pi \, \text{kHz}$ center line uncertainty, and $\sigma_{\Delta\varphi} = 0.33\pi$ phase uncertainty. The starting distance has been calculated using Eq. (32). The fraction captured represents the number of runs that terminate within 2 times the target thresholds of the optimal parameter values. For each distance, 2000 tuneups are run. The shaded area corresponds to one standard deviation. For 10 000 particles, we additionally measure the capture range experimentally, where the results are represented with red squares. The error bars correspond to the statistical deviation of the data points.
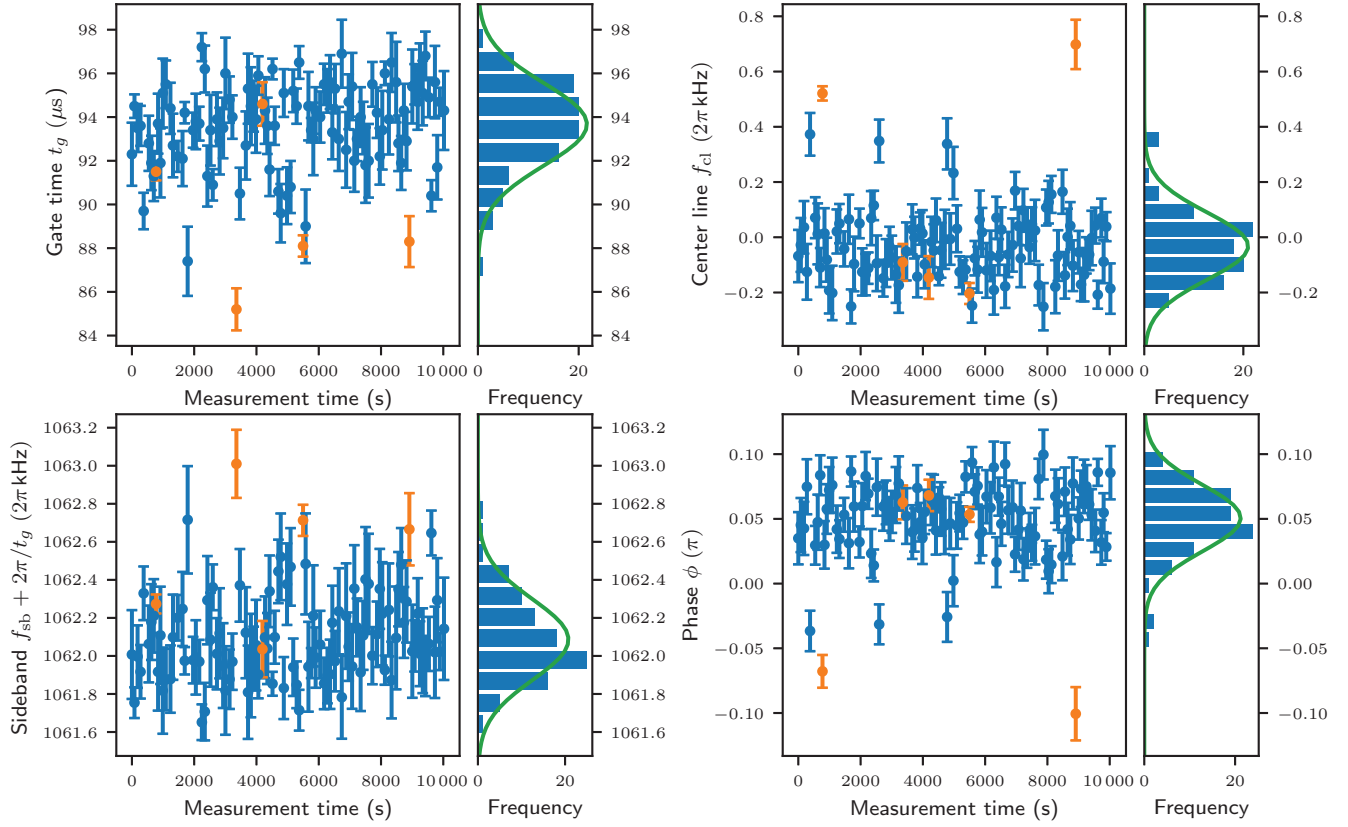
FIG. 10. Final parameter estimates and uncertainties for 103 iterations of tuning up the Mølmer-Sørensen gate versus time elapsed. The orange points are runs rejected by the confirmation criterion described in Sec. IV E, corresponding to 5(2)% of all runs. Removing the rejected points reduces the standard deviation of the distributions of final parameter estimates between 13%–20%. Green curves show a Gaussian distribution fitted to a histogram the distribution of final parameter values.

consequence, discretization artifacts will appear, such as the particle filtering not being able to track well enough the peaks of the distribution.

We choose 10 000 particles as our default value to work with since in simulation the gains for using more than 5000 particles decrease significantly.

Experimentally, we find a reduction of 10%–25% of the success rate compared to the simulations using the same sample size for starting distances < 2 due to effects not included in the model affecting the algorithm. For larger distances, the success rate is comparable between experiment and simulation.

While we are unable to pin down the exact cause of failures, we hypothesize that these events may be due to transients in the control fields.

These effects may lead to a difference between our estimated likelihoods and the true ones, which then suppress the posterior probability distribution around the true values. This can then shift the parameter estimates away from the true values, but also lead to underestimating the variances compared to the true error. The algorithm might then use unsuitable measurement settings, or prematurely terminate if the variances are small enough to fulfill the stopping

criterion. These effects can then lead to final parameter estimates of the algorithm far from the ideal values.

E. Confirmation measurements

To use the calibration algorithm even with less than perfect success rate, we insert confirmation measurements after completing the procedure to detect and then reject outliers in the final gate parameters predicted by the algorithm. For rejected calibration runs, the algorithm can then be repeated.

To detect outliers, we employ check sequences. We choose a sequence of eight consecutive $U_{MS}^0(\pi/2)$ gates and a sequence of six gates with $\Delta\varphi_{target} = \pm\pi/4$ that both ideally return all populations to the ground state. The sequences are chosen because they are sensitive to Rabi frequency and phase difference miscalibrations, respectively. We choose an acceptance threshold of at least 85 out of 100 measurements being in the target state. This threshold is chosen as a compromise between acceptance rate and the infidelity of the accepted runs, as more stringent thresholds start decreasing acceptance rate without significant improvements of either maximum or
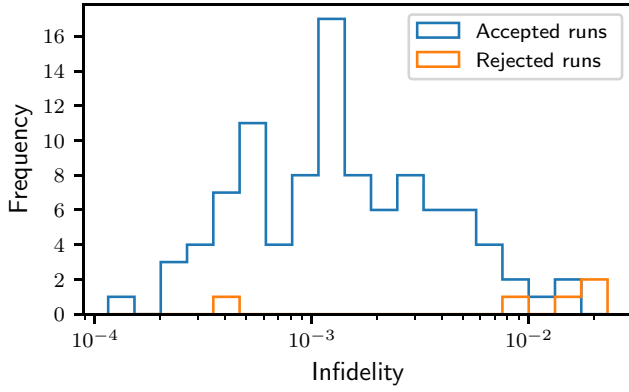
FIG. 11. Histogram of gate infidelities due to imperfect calibration. Randomized benchmarking is used to compute a per-gate infidelity from the difference of the final parameter estimates of independent calibration runs compared to their mean value (Fig. 10). We calculate a median infidelity of $1.3(1) \times 10^{-3}$ caused by imperfect calibration and a 95 percentile of 7.4×10^{-3}.

median infidelity of accepted runs. We run repeated calibration runs including confirmation measurements and record the final parameter estimates produced by the calibration runs. Of the runs, 95(2)% pass the confirmation test (Fig. 10). The distribution of accepted runs has standard deviations of $\sigma_{t_g} = 1.9(1)$ μs, $\sigma_{f_{cl}} = 0.12(1) \times 2\pi$ kHz, $\sigma_{f_{sb}} = 0.21(2) \times 2\pi$ kHz, $\sigma_\phi = 0.025(3)$ π. Translating these deviations to the gate parameters Θ, we find that they are in agreement or below the thresholds set in Sec. IV A.

We evaluate the estimated infidelity due to imperfect calibration of the accepted runs of the algorithm by simulating randomized cycle benchmarking using the difference between the final parameter estimates and the mean of all accepted calibration runs as miscalibration, and calculate the expected infidelity due to calibration error. We achieve a median calibration infidelity of $1.3(1) \times 10^{-3}$ (Fig. 11). This value is lower than what was calculated for independent errors in Sec. IV A, which we attribute to the correlations between the parameters partially compensating each other. We can compare this to the fidelity of Bell states [72] generated by repeated application of the MS gate on the same system, and find that the infidelity per gate is observed to be $4(1) \times 10^{-3}$. While this state fidelity is not suited to characterize the effects of coherent errors, it does pose a limit on the achievable fidelity due to incoherent errors present in the system. We expect laser phase noise and dephasing of the motional mode as the leading sources of decoherence. From this infidelity we conclude that our calibration routine can produce parameter estimates such that our entangling gates are not limited by calibration errors, but limited by the performance of the machine.

V. CONCLUSION AND OUTLOOK

In this work we proposed and implemented a procedure to use Bayesian parameter estimation to optimize the control parameters of an entangling two-qubit Mølmer-Sørensen gate. We described how the gate can be simulated as a function of a basic set of parameters, and how a particle filtering algorithm can be used to estimate them from measurements. Two strategies for selecting the measurement settings were compared, and we showed that a simple heuristic strategy can match the performance of a more complex variance minimization method.

We demonstrated that the algorithm can be used to calibrate a gate to a median residual infidelity due to parameter miscalibration of $1.3(1) \times 10^{-3}$ (Fig. 11) in 1200 ± 500 experimental shots, achieving a significant speedup over manually iterating over 1D parameter scans (see the Appendix). These results demonstrate the suitability of the described approach to reliably and consistently calibrate a set of control parameters to achieve a gate fidelity that is commensurate with the maximum fidelity set by incoherent processes in the quantum system. Our approach considerably eases the operational burden and can produce valuable time savings for the operating experimentalist. Furthermore, periodic recalibration can be included within long experimental data-taking sequences, protecting against slow parameter drifts, and thus guaranteeing high-quality output over considerably longer time.

The two-qubit gate optimization routine we present here would benefit the quantum CCD ion-trapped quantum computing architecture [76], where electric fields can be used to trap ions in multiple independent, spatially separate regions. By a combination of shuttling, splitting, and rotations, entangling operations can be performed between any two ions in the N-qubit register by isolating them in the appropriate "interaction" region [77,78]. Once the gate is calibrated in the interaction region, owing to all ions having identical atomic properties, the gate remains calibrated for the whole qubit register. The current implementation is also suited for experiments that generate 2-out-of-N ion entanglement by selectively addressing ions in a 1D chain through their radial modes of motion [79]. In the presence of additional sources of miscalibration such as crosstalk and position-dependent Stark shift, these schemes can easily be extended such that gates between each ion pair will have their own simulated interpolator.

While the discussion here has focused on trapped-ion implementations, we emphasize that this multidimensional Bayesian approach to gate calibration is not limited to just trapped ions and can be applied to any quantum system where a classical simulation of the entangling gate yields an accurate representation of the physical implementation.

A limitation is that the classical information stored in the interpolator of the gate action will grow exponentially with the number of control parameters. We can thus ask

what is the outlook for extending the scheme. The MS gate naturally scales to N-qubit entangling gates using the exact same set of control parameters. In this case the current scheme would work as long as the classical simulation remains tractable and can effectively describe other experimental imperfections such as magnetic field gradients or unequal Rabi frequencies on the individual ions. While this step might be time intensive for multiple qubits, it only needs to be performed once. When considering 2-out-of-N entanglement generation on radial modes, modifications would be required in situations where many motional modes begin to participate in the MS interaction and complex amplitude or frequency modulation waveforms are required to disentangle the spins from these other modes [80,81]. Extensions to include such optimal control techniques into the calibration protocol provide an avenue for further research.

Another extension of the current scheme is the integration of more sophisticated control parameter tracking. Instead of periodically running the full calibration routine starting from an unbiased prior and $N_g = 1$, one would retain the Bayesian prior from the previous calibration and seek to update it in as few measurements as possible to achieve further speedup. For the same amount of experimental shots, we could shorten the time between successive recalibrations and would thus make the system even more robust against parameter drift, as has been previously shown for single-qubit systems [32,33,82].

While the calibration is robust because runs that fail our convergence criterion are rejected, we continue to investigate the source of these occasional failures. We hypothesize that such events can be caused by large nonstatistical deviations of the gate Hamiltonian due to transients in control fields. One method to characterize this would be to try an alternative automated calibration method, such as a machine learning approach, and compare the average achieved infidelity. This would allow us to characterize how sensitive competing methods are to such instabilities. Furthermore, we would obtain another point of comparison to benchmark the convergence speed against. In a similar vein, we could potentially enhance the success rate by further improving our classical simulation of the gate model by substituting the unitary predictions with a master equation description that includes measured noise sources such as heating and dephasing in our experimental setup. An interesting application of this open system extension would be to assess the suitability of our approach to calibrating experimental control systems that exhibit significant deviations from the ideal, unitary gate action.

ACKNOWLEDGMENTS

APPENDIX: COMPARISON TO DIFFERENT CALIBRATION METHODS

One of the desiderata of an optimization routine is having a fast convergence rate to the desired parameter accuracy. In other words, we would like to minimize the number of times the experiment has to be queried since each measurement incurs by far the largest operational time overhead. Since in this publication we do not claim to have the optimal strategy, can we still claim that our Bayesian optimization is fast relative to what might be considered a traditional or manual optimization? In practice, each experimentalist might have their own slightly idiosyncratic way of optimizing the gate parameters and will also have different amounts of prior knowledge about the parameter values. There is no set prescription that is followed by a majority consensus in the ion trap literature. However, Ref. [25] provides an algorithmic approach that iteratively performs one-dimensional parameter scans to optimize the gate performance. We find this method closely mirrors what we would perform in our lab in the absence of the Bayesian optimization routine and hence we use it as a benchmark. In summary, the algorithm can be reduced to the following.

1. Roughly find the sideband detuning for desired gate time.
2. Scan the center line detuning at 2 times the gate time and maximize the population in p_{ee}.

TABLE I. Number of points used in the scans for the different cases for a 1D parameter scan tuneup considered.

	Case 1	Case 2	Case 3	Case 4
Points t_g	4	8	16	16
Points δ	4	8	8	10
Points ω_{cl}	4	6	8	12
Points $\Delta\varphi$	4	6	8	10
Total measurements	2000	3500	4800	5900

3. Scan the sideband detuning at the presumed gate time and balance the populations p_{ee} and p_{gg}, so that $p_{ee} \approx p_{gg}$.
4. Scan the gate time and minimize the populations of $p_{eg} + p_{ge}$.
5. Repeat steps 3 and 4.
6. Final repeat of step 3.
7. Repeat step 2.
8. Scan the phase after two gates.
9. Repeat steps 7 and 8.

To estimate the parameters from the 1D scans, we employ standard least square fitting. For the center line scan, a Gaussian fit function is used on the p_{ee} population, and for the sideband detuning scan, the difference $p_{ee} - p_{gg}$ is fitted with a linear slope to balance the populations. For the gate time scan, the dip in the unwanted populations $p_{eg} + p_{ge}$ is fitted with an offset Gaussian. Lastly, in the phase scan the difference $p_{ee} - p_{gg}$ is fitted with a fixed-frequency cosine function.

For each scan, both the number of points and the range of the scan are free parameters that can be altered. To find reasonable combinations of scan width and points, we numerically simulate scans with only a single misset parameter for combinations of ranges and points, and fit the simulated data. We then choose the range with the smallest distance between the 13th and 87th percentiles.

We simulate 100 executions of the algorithm described previously for different numbers of points per scan (see Table I), for a gate with ideal values of $t_g = 100$ μs, $\omega_{cl} = 0$ kHz, $\Delta\varphi = 0$. Each data point consists of 50 measurements. Results with extreme parameter values are discarded ($t_g < 0$ μs, $t_g > 200$ μs, $|2\pi\omega_{cl}| > 10$ kHz, $2\pi\delta > 20$ kHz, $2\pi\delta < 0$ kHz), which occurs 21(2)% of the time in all cases considered here. For the remainder, we determine the fidelity of the gate given the parameter estimates in the same manner the Bayesian parameter estimates were evaluated (see Fig. 11), except that the true value set in the simulation can be substituted for the mean. The results for the different settings can be seen in Fig. 12.

It should be noted that the robustness of this algorithm could be increased by introducing sanity checks to discard poor fits and adaptively choosing if a parameter should be
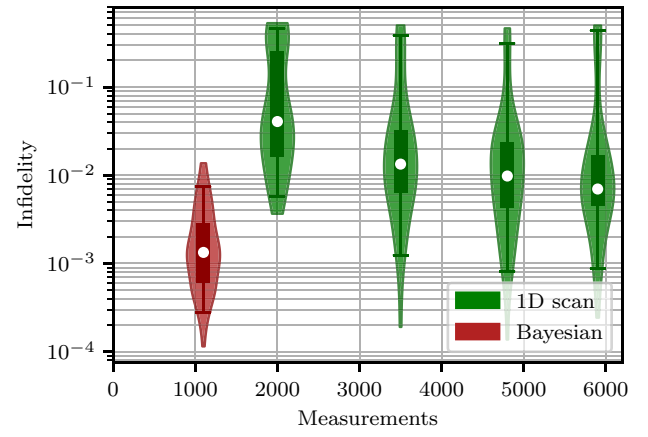

FIG. 12. Simulation of manual tuneups using 1D scans for four different measurement budgets compared to the data obtained using the Bayesian method. The distribution of infidelities is shown as a violine plot, with median, quartiles, and 5–95 percentiles indicated. For all four settings under consideration, the method has a 21(2)% rate of failure. For all settings, the median simulated parameter infidelities are larger than the infidelities obtained from the Bayesian method.

remeasured. However, doing so will increase the number of measurements further.

We find that, for all settings considered here, the experimental results obtained using the Bayesian algorithm outperform the simulations of iterative parameter scans both in terms of median infidelity achieved and measurements required even in the ideal case of a noise-free system. Possible reasons for this include that the fit functions are only approximations to the true shape. Only part of the measurement data is used for the fit. Furthermore, corrections are always applied to a single variable per scan, leading to correlated effects needing to be corrected by a successive approximation.

In conclusion, the Bayesian algorithm outperforms a 1D manual parameter scan even granting for the most optimistic outcome, and in typical operation it is several times more efficient in the number of required experimental shots.

[1] M. Saffman, Quantum computing with atomic qubits and Rydberg interactions: Progress and challenges, J. Phys. B: At. Mol. Opt. Phys. **49**, 202001 (2016).
[2] G. Wendin, Quantum information processing with superconducting circuits: A review, Rep. Prog. Phys. **80**, 106001 (2017).
[3] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).
[4] H.-L. Huang, D. Wu, D. Fan, and X. Zhu, Superconducting quantum computing: A review, Sci. China Inf. Sci. **63**, 180501 (2020).

[5] S. Slussarenko and G. J. Pryde, Photonic quantum information processing: A concise review, Appl. Phys. Rev. **6**, 041303 (2019).

[6] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, Trapped-ion quantum computing: Progress and challenges, Appl. Phys. Rev. **6**, 021314 (2019).

[7] C. J. Ballance, T. P. Harty, N. M. Linke, M. A. Sepiol, and D. M. Lucas, High-Fidelity Quantum Logic Gates Using Trapped-Ion Hyperfine Qubits, Phys. Rev. Lett. **117**, 060504 (2016).

[8] A. C. Hughes, V. M. Schäfer, K. Thirumalai, D. P. Nadlinger, S. R. Woodrow, D. M. Lucas, and C. J. Ballance, Benchmarking a High-Fidelity Mixed-Species Entangling Gate, Phys. Rev. Lett. **125**, 080504 (2020).

[9] Y. Wang, S. Crain, C. Fang, B. Zhang, S. Huang, Q. Liang, P. H. Leung, K. R. Brown, and J. Kim, High-Fidelity Two-Qubit Gates Using a Microelectromechanical-System-Based Beam Steering System for Individual Qubit Addressing, Phys. Rev. Lett. **125**, 150505 (2020).

[10] P. H. Leung, K. A. Landsman, C. Figgatt, N. M. Linke, C. Monroe, and K. R. Brown, Robust 2-Qubit Gates in a Linear Ion Crystal Using a Frequency-Modulated Driving Force, Phys. Rev. Lett. **120**, 020501 (2018).

[11] J. P. Gaebler, T. R. Tan, Y. Lin, Y. Wan, R. Bowler, A. C. Keith, S. Glancy, K. Coakley, E. Knill, D. Leibfried, and D. J. Wineland, High-Fidelity Universal Gate Set for ^{9}Be$^+$ Ion Qubits, Phys. Rev. Lett. **117**, 1 (2016).

[12] S. S. Hong, A. T. Papageorge, P. Sivarajah, G. Crossman, N. Didier, A. M. Polloreno, E. A. Sete, S. W. Turkowski, M. P. Da Silva, and B. R. Johnson, Demonstration of a parametrically activated entangling gate protected from flux noise, Phys. Rev. A **101**, 1 (2020).

[13] R. Barends et al., Superconducting quantum circuits at the surface code threshold for fault tolerance, Nature **508**, 500 (2014).

[14] M. A. Rol, F. Battistel, F. K. Malinowski, C. C. Bultink, B. M. Tarasinski, R. Vollmer, N. Haider, N. Muthusubramanian, A. Bruno, B. M. Terhal, and L. DiCarlo, Fast, High-Fidelity Conditional-Phase Gate Exploiting Leakage Interference in Weakly Anharmonic Superconducting Qubits, Phys. Rev. Lett. **123**, 120502 (2019).

[15] W. Huang, C. H. Yang, K. W. Chan, T. Tanttu, B. Hensen, R. C. C. Leon, M. A. Fogarty, J. C. C. Hwang, F. E. Hudson, K. M. Itoh, A. Morello, A. Laucht, and A. S. Dzurak, Fidelity benchmarks for two-qubit gates in silicon, Nature **569**, 532 (2019).

[16] T. M. Graham, M. Kwon, B. Grinkemeyer, Z. Marra, X. Jiang, M. T. Lichtman, Y. Sun, M. Ebert, and M. Saffman, Rydberg-Mediated Entanglement in a Two-Dimensional Neutral Atom Qubit Array, Phys. Rev. Lett. **123**, 230501 (2019).

[17] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[18] A. Montanaro, Quantum algorithms: An overview, Npj Quantum Inf. **2**, 15023 (2016).

[19] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum (NISQ) algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[20] A. Patterson, J. Rahamim, T. Tsunoda, P. Spring, S. Jebari, K. Ratter, M. Mergenthaler, G. Tancredi, B. Vlastakis, M. Esposito, and P. Leek, Calibration of a Cross-Resonance Two-Qubit Gate Between Directly Coupled Transmons, Phys. Rev. Appl. **12**, 064013 (2019).

[21] J. Kelly et al., Scalable in situ qubit calibration during repetitive error detection, Phys. Rev. A **94**, 032321 (2016).

[22] F. t. Arute, Quantum supremacy using a programmable superconducting processor (supplementary material), Nature **574**, 505 (2019).

[23] P. V. Klimov, J. Kelly, J. M. Martinis, and H. Neven, The snake optimizer for learning quantum processor control parameters, arXiv:2006.04594 [quant-ph] (2020).

[24] Y. Li, L. Pezzè, M. Gessner, Z. Ren, W. Li, and A. Smerzi, Frequentist and Bayesian quantum phase estimation, Entropy **20**, 25 (2018).

[25] N. Akerman, N. Navon, S. Kotler, Y. Glickman, and R. Ozeri, Universal gate-set for trapped-ion qubits using a narrow linewidth diode laser, New J. Phys. **17**, 113060 (2015).

[26] S. G. Schirmer and F. C. Langbein, Ubiquitous problem of learning system parameters for dissipative two-level quantum systems: Fourier analysis versus Bayesian estimation, Phys. Rev. A - At. Mol. Opt. Phys. **91**, 1 (2015).

[27] F. Martínez-García, D. Vodola, and M. Möller, Adaptive Bayesian phase estimation for quantum error correcting codes, New J. Phys. **21**, 123027 (2019).

[28] B. Teklu, S. Olivares, and M. G. A. Paris, Bayesian estimation of one-parameter qubit gates, J. Phys. B: At. Mol. Opt. Phys. **42**, 035502 (2009).

[29] Y. Yang, B. Gong, and W. Cui, Real-time quantum state estimation in circuit QED via the Bayesian approach, Phys. Rev. A **97**, 1 (2018).

[30] A. H. Kiilerich and K. Mølmer, Quantum Zeno effect in parameter estimation, Phys. Rev. A - At. Mol. Opt. Phys. **92**, 1 (2015).

[31] N. Widmer, Master's thesis, 2018.

[32] T. Proctor, M. Revelle, E. Nielsen, K. Rudinger, D. Lobser, P. Maunz, R. Blume-Kohout, and K. Young, Detecting and tracking drift in quantum information processors, Nat. Commun. **11**, 5396 (2020).

[33] J. F. Ralph, K. Jacobs, and C. D. Hill, Frequency tracking and parameter estimation for robust quantum state estimation, Phys. Rev. A - At. Mol. Opt. Phys. **84**, 1 (2011).

[34] A. Sørensen and K. Mølmer, Quantum Computation with Ions in Thermal Motion, Phys. Rev. Lett. **82**, 1971 (1999).

[35] A. Sørensen and K. Mølmer, Entanglement and quantum computation with ions in thermal motion, Phys. Rev. A **62**, 022311 (2000).

[36] G. Kirchmair, J. Benhelm, F. Zähringer, R. Gerritsma, C. F. Roos, and R. Blatt, Deterministic entanglement of ions in thermal states of motion, New J. Phys. **11**, 023002 (2009).

[37] C. F. Roos, Ion trap quantum gates with amplitude-modulated laser beams, New J. Phys. **10**, 013002 (2008).

[38] M. Riebe, K. Kim, P. Schindler, T. Monz, P. O. Schmidt, T. K. Körber, W. Hänsel, H. Häffner, C. F. Roos, and R. Blatt, Process Tomography of Ion Trap Quantum Gates, Phys. Rev. Lett. **97**, 220407 (2006).

[39] I. L. Chuang and M. A. Nielsen, Prescription for experimental determination of the dynamics of a quantum black box, J. Mod. Opt. **44**, 2455 (1997).

[40] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, Self-consistent quantum process tomography, Phys. Rev. A **87**, 1 (2013).

[41] 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).

[42] S. Mavadia, C. L. Edmunds, C. Hempel, H. Ball, F. Roy, T. M. Stace, and M. J. Biercuk, Experimental quantum verification in the presence of temporally correlated noise, Npj Quantum Inf. **4**, 1 (2017).

[43] A. Erhard, J. J. Wallman, L. Postler, M. Meth, R. Stricker, E. A. Martinez, P. Schindler, T. Monz, J. Emerson, and R. Blatt, Characterizing large-scale quantum computers via cycle benchmarking, Nat. Commun. **10**, 5347 (2019).

[44] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography, Nat. Commun. **8**, 14485 (2017).

[45] I. A. Pogorelov, G. I. Struchalin, S. S. Straupe, I. V. Radchenko, K. S. Kravtsov, and S. P. Kulik, Experimental adaptive process tomography, Phys. Rev. A **95**, 012302 (2017).

[46] C. Granade, C. Ferrie, and S. T. Flammia, Practical adaptive quantum tomography, New J. Phys. **19**, 113017 (2017).

[47] M. A. Rol, C. C. Bultink, T. E. O'Brien, S. R. de Jong, L. S. Theis, X. Fu, F. Luthi, R. F. L. Vermeulen, J. C. de Sterke, A. Bruno, D. Deurloo, R. N. Schouten, F. K. Wilhelm, and L. DiCarlo, Restless Tuneup of High-Fidelity Qubit Gates, Phys. Rev. Appl. **7**, 041001 (2017).

[48] E. Greplova, C. K. Andersen, and K. Mølmer, Quantum parameter estimation with a neural network, arXiv, 1 (2017).

[49] M. Brandl, M. Van Mourik, L. Postler, A. Nolf, K. Lakhmanskiy, R. Paiva, S. Möller, N. Daniilidis, H. Häffner, and V. Kaushal *et al.*, Cryogenic setup for trapped ion quantum computing, Rev. Sci. Instrum. **87**, 113103 (2016).

[50] D. Leibfried, R. Blatt, C. Monroe, and D. Wineland, Quantum dynamics of single trapped ions, Rev. Mod. Phys. **75**, 281 (2003).

[51] D. F. V. James, Quantum dynamics of cold trapped ions with application to quantum computation, Appl. Phys. B: Lasers Opt. **66**, 20 (1997).

[52] P. Schindler, D. Nigg, T. Monz, J. T. Barreiro, E. Martinez, S. X. Wang, S. Quint, M. F. Brandl, V. Nebendahl, and C. F. Roos *et al.*, A quantum information processor with trapped ions, New J. Phys. **15**, 123012 (2013).

[53] V. Nebendahl, H. Häffner, and C. Roos, Optimal control of entangling operations for trapped-ion quantum computing, Phys. Rev. A **79**, 012312 (2009).

[54] H. Häffner, S. Gulde, M. Riebe, G. Lancaster, C. Becher, J. Eschner, F. Schmidt-Kaler, and R. Blatt, Precision Measurement and Compensation of Optical Stark Shifts for an Ion-Trap Quantum Processor, Phys. Rev. Lett. **90**, 143602 (2003).

[55] P. Schindler, Master's thesis, University of Innsbruck, 2008.

[56] 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).

[57] P. Lee, *Bayesian Statistics: An Introduction* (Wiley, Chichester, West Sussex Hoboken, N.J, 2012).

[58] Y. Ho, and R. Lee, A Bayesian approach to problems in stochastic estimation and control, IEEE Trans. Automat. Contr. **9**, 382 (1964).

[59] D. Q. Mayne, A solution of the smoothing problem for linear dynamic systems, Automatica **4**, 73 (1966).

[60] J. E. Handschin and D. Q. Mayne, Monte Carlo techniques to estimate the conditional expectation in multi-stage nonlinear filtering, Int. J. Control **9**, 547 (1969).

[61] J. S. Liu and R. Chen, Sequential Monte Carlo methods for dynamic systems, J. Am. Stat. Assoc. **93**, 1032 (1998).

[62] The nomenclature of a filter stems from filtering estimates out of a stream of noisy measurement data.

[63] CPU: Intel i5-4670S@3.10GHz.

[64] P. A. Walker, Quadcubic interpolation: A four-dimensional spline method, arXiv preprint arXiv:1904.09869 (2019).

[65] CPU: Intel i7-6700K@4.00GHz.

[66] N. J. Gordon, D. J. Salmond, and A. F. Smith, in *IEE proceedings F (radar and signal processing)*, (IET, 1993), Vol. 140, p. 107.

[67] D. B. Rubin, Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm, J. Am. Stat. Assoc. **82**, 543 (1987).

[68] J. Liu and M. West, in *Sequential Monte Carlo Methods in Practice* (Springer, 2001), p. 197.

[69] F. Huszár and N. M. Houlsby, Adaptive Bayesian quantum tomography, Phys. Rev. A - At. Mol. Opt. Phys. **85**, 052120 (2012).

[70] K. S. Kravtsov, S. S. Straupe, I. V. Radchenko, N. M. Houlsby, F. Huszár, and S. P. Kulik, Experimental adaptive Bayesian tomography, Phys. Rev. A - At. Mol. Opt. Phys. **87**, 062122 (2013).

[71] N. Wiebe and C. Granade, Efficient Bayesian Phase Estimation, Phys. Rev. Lett. **117**, 010503 (2016).

[72] R. Jozsa, Fidelity for mixed quantum states, J. Mod. Opt. **41**, 2315 (1994).

[73] 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).

[74] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, J. Opt. B: Quantum Semiclassical Opt. **7**, S347 (2005).

[75] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and Robust Randomized Benchmarking of Quantum Processes, Phys. Rev. Lett. **106**, 8 (2011).

[76] D. Kielpinski, C. Monroe, and D. J. Wineland, Architecture for a large-scale ion-trap quantum computer, Nature **417,** 709 (2002).

[77] J. M. Pino, J. M. Dreiling, C. Figgatt, J. P. Gaebler, S. A. Moses, M. S. Allman, C. H. Baldwin, M. Foss-Feig, D. Hayes, K. Mayer, C. Ryan-Anderson, and B. Neyenhuis, Demonstration of the trapped-ion quantum CCD computer architecture, Nature **592,** 209 (2021).

[78] V. Kaushal, B. Lekitsch, A. Stahl, J. Hilder, D. Pijn, C. Schmiegelow, A. Bermudez, M. Müller, F. Schmidt-Kaler, and U. Poschinger, Shuttling-based trapped-ion quantum information processing, AVS Quantum Sci. **2,** 014101 (2020).

[79] I. Pogorelov, T. Feldker, C. D. Marciniak, L. Postler, G. Jacob, O. Krieglsteiner, V. Podlesnic, M. Meth, V. Negnevitsky, M. Stadler, B. Höfer, C. Wächter, K. Lakhmanskiy, R. Blatt, P. Schindler, and T. Monz, Compact Ion-Trap Quantum Computing Demonstrator, PRX Quantum **2,** 020343 (2021).

[80] Y. Lu, S. Zhang, K. Zhang, W. Chen, Y. Shen, J. Zhang, J.-N. Zhang, and K. Kim, Global entangling gates on arbitrary ion qubits, Nature **572,** 363 (2019).

[81] K. Wright *et al.*, Benchmarking an 11-qubit quantum computer, Nat. Commun. **10,** 1 (2019).

[82] S. Kimmel, G. H. Low, and T. J. Yoder, Robust calibration of a universal single-qubit gate set via robust phase estimation, Phys. Rev. A **92,** 062315 (2015).