Robust Quantum Optimal Control with Trajectory Optimization

Thomas Propson,[1,2,*] Brian E. Jackson,[3] Jens Koch,[4,5] Zachary Manchester,[3] and David I. Schuster[1,2,6]

[1] *James Franck Institute, University of Chicago, Chicago, Illinois 60637, USA*
[2] *Department of Physics, University of Chicago, Chicago, Illinois 60637, USA*
[3] *Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA*
[4] *Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA*
[5] *Northwestern–Fermilab Center for Applied Physics and Superconducting Technologies, Northwestern University, Evanston, Illinois 60208, USA*
[6] *Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637, USA*

(Dated: March 24, 2021)

The ability to engineer high-fidelity gates on quantum processors in the presence of systematic errors remains the primary barrier to achieving quantum advantage. Quantum optimal control methods have proven effective in experimentally realizing high-fidelity gates, but they require exquisite calibration to be performant. We apply robust trajectory optimization techniques to suppress gate errors arising from system parameter uncertainty. We propose a derivative-based approach that maintains computational efficiency by using forward-mode differentiation. Additionally, the effect of depolarization on a gate is typically modeled by integrating the Lindblad master equation, which is computationally expensive. We employ a computationally efficient model and utilize time-optimal control to achieve high-fidelity gates in the presence of depolarization. We apply these techniques to a fluxonium qubit and suppress simulated gate errors due to parameter uncertainty below 10^{-7} for static parameter deviations on the order of 1%.

I. INTRODUCTION

Quantum optimal control (QOC) is a class of optimization algorithms for accurately and efficiently manipulating quantum systems. Early techniques were proposed for nuclear magnetic resonance experiments [1–7], and applications now include superconducting circuits [8–24], neutral atoms and ions [25–36], nitrogen-vacancy centers in diamond [37–43], and Bose-Einstein condensates [44–47]. In the context of quantum computation, optimal control is employed to achieve high-fidelity gates while adhering to experimental constraints. Experimental errors such as parameter drift, noise, and finite control resolution cause the system to deviate from the model used in optimization, hampering experimental performance [9, 14, 20, 33, 48]. Robust control improves upon standard optimal control by encoding model parameter uncertainties in optimization objectives, yielding performance guarantees over a range of parameter values [49–51]. We adapt robust control techniques from the robotics community to mitigate parameter-uncertainty errors for a superconducting fluxonium qubit.

Analytically-derived control pulses that mitigate parameter-uncertainty errors include composite pulses [52–55], pulses designed by considering dynamic and geometric phases [56, 57], and pulses obtained with the DRAG scheme [58]. As compared to analytical techniques, QOC is advantageous for designing pulses that consider all experimental constraints and performance tradeoffs [17], and for constructing operations without a known analytic solution [9, 14]. Accordingly, recent work has sought to achieve robustness in QOC frameworks using closed-loop methods [59–62] and open-loop methods [20, 42, 63–66].

In this work, we study three robust control techniques that make the system's quantum state trajectory less sensitive to uncertainties of static and time-dependent parameters:

1. A sampling method, similar to the work in Refs. [3, 20, 42, 63, 64, 67].
2. An unscented sampling method [68–70] adapted from the unscented transform [71, 72] used in state estimation.
3. A derivative method, which penalizes the sensitivity of the quantum state trajectory to uncertain parameters.

We apply these techniques to the fluxonium qubit presented in [73]. We also show that QOC can solve important problems associated with fluxonium-based qubits: exploiting the dependence of T_1 on the controls to mitigate depolarization and synchronizing the phase of qubits with distinct frequencies. To ameliorate depolarization, we perform time-optimal control and employ an efficient depolarization model for which the computational cost is independent of the Hilbert space dimension. Leveraging recent advances in trajectory optimization within the field of robotics, we solve these optimization problems using ALTRO (Augmented Lagrangian TRajectory Optimizer) [74], which can enforce constraints on the control parameters and the quantum state trajectory.

This paper is organized as follows. First, we describe ALTRO in the context of QOC in Section II. We outline realistic constraints for operating the fluxonium and define the associated QOC problem in Section III. Then, we

* tcpropson@uchicago.edu

formulate a method for suppressing depolarization in Section IV. Next, we describe three techniques for achieving robustness to static parameter uncertainties in Section V. We adapt the same techniques to mitigate $1/f$ flux noise in Section VI.

II. BACKGROUND

In this section, we review the QOC problem statement and describe the ALTRO solver [74]. QOC concerns a vector $\mathbf{a}(t)$ of time-dependent control fields that steer the evolution of a quantum state $|\psi(t)\rangle$. The evolution of the state is governed by the time-dependent Schrödinger equation (TDSE),

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = H(\mathbf{a}(t), t) |\psi(t)\rangle . \tag{1}$$

The Hamiltonian $H(\mathbf{a}(t), t)$ is determined by the quantum system and the external control fields. The QOC problem is to find the controls that minimize a functional $J[\mathbf{a}(t)]$, which we call the objective. To make the problem numerically tractable, the quantum state and controls are discretized into N time steps, $|\psi(t_k)\rangle \to |\psi_k\rangle$ and $\mathbf{a}(t_k) \to \mathbf{a}_k$ where $t_{k+1} = t_k + \Delta t$ and $k \in \{1, ..., N\}$. In the case of a single state-transfer problem, the objective is the infidelity of the time-evolved final state $|\psi_N\rangle$ and the intended target state $|\psi_T\rangle$ [to me, "f" would mean "final" rather than target...] [changed f to T], $J(\mathbf{a}) = 1 - |\langle \psi_T | \psi_N(\mathbf{a}) \rangle|^2$. Standard QOC solvers compute derivatives of the objective $\nabla J(\mathbf{a})$, which can easily be used to implement first-order optimization methods [3, 17, 75, 76].

Alternatively, the QOC problem can be formulated as a trajectory optimization problem and solved using specialized solvers developed by the robotics community [74, 77–79]. The objective $J(\mathbf{a}) = \sum_k \ell_k(\mathbf{x}_k, \mathbf{u}_k)$ is expressed in terms of the cost function at each time step ℓ_k, where $\mathbf{x}_k$ is the augmented state vector and $\mathbf{u}_k$ is the augmented control vector. We use the term *augmented* because these vectors contain all of the relevant variables in the optimization problem, not just the quantum state and the control fields, for an example see Section III. The augmented control contains all variables that the experimentalist may manipulate, and the augmented state contains all variables that depend on those in the augmented control. The variables in the augmented states depend on those in the augmented controls as defined by the differential equations governing the physical system, which are encoded in the discrete relation $\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)$. For QOC, the discrete dynamics function $\mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)$ propagates the quantum state by integrating the TDSE (1) with the given control fields using Runge-Kutta methods [80] or exponential integrators [81–84]. ["discrete dynamics function" is not physics language. Relate f to Schrodinger equation first, then give it the un- familiar name.] [I think the sentence flows better with "The discrete dynamics function $f(.,.)$ propagates" rather than "$f(.,.)$ propagates... By the way, $f(.,.)$ is called the ddf". Furthermore, I cannot produce a sentence which relates the TDSE dependence of $|\psi_k\rangle$ and $\mathbf{a}_k$ to each other and then introduces $\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{u}_k, \mathbf{x}_k)$ because then I feel it needs to be explained again that $|\psi_k\rangle$ and $\mathbf{a}_k$ are contained in $\mathbf{x}_k$ and $\mathbf{u}_k$. I think introducing the specific case of QOC later in "For QOC" is less verbose and more words will detract from understanding in this instance.] [Separate point: if x contains components other than the quantum state, then what does "obeying the physics of the system" mean?] [everything time evolves under a differential equation even if it's really simple $\mathrm{d}a/\mathrm{d}t = \dot{a}$] [definition of $\mathbf{f}$ should be separated from its numerical implementation?] [I think that the lead in "For QOC," makes it evident that this is a specific case, similar to how above it says "For example," when talking about specific components of the state and control vectors].

Additional constraints on the augmented controls and states are formulated as inequalities $\mathbf{g}_k(\mathbf{x}_k, \mathbf{u}_k) \leq \mathbf{0}$ or equalities $\mathbf{h}_k(\mathbf{x}_k, \mathbf{u}_k) = \mathbf{0}$. [How to understand the k? labels here?] [they denote the time step, which i think is consistent with everything else] The constraint functions $\mathbf{g}_k$ and $\mathbf{h}_k$ may be vector-valued to encode multiple constraints, and equalities and inequalities are understood component-wise. Each constraint's *violation* is defined as the magnitude of its deviation: $\max(g(\cdot), 0)$ or $|h(\cdot)|$, where g and h are individual components of constraint functions $\mathbf{g}_k$ and $\mathbf{h}_k$, respectively. [Is this the right place for this definition in the text. Seems unmotivated/out of place right now?] [Yes, first place it comes up in ALM paragraph is a worse place to introduce it] Stated concisely, the trajectory optimization problem is

$$\underset{\substack{\mathbf{x}_1,\ldots,\mathbf{x}_N \\ \mathbf{u}_1,\ldots,\mathbf{u}_{N-1}}}{\text{minimize}} \quad \ell_N(\mathbf{x}_N) + \sum_{k=1}^{N-1} \ell_k(\mathbf{x}_k, \mathbf{u}_k) \tag{2a}$$

$$\text{subject to} \quad \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \quad \forall\ k, \tag{2b}$$

$$\mathbf{g}_k(\mathbf{x}_k, \mathbf{u}_k) \leq \mathbf{0} \quad \forall\ k, \tag{2c}$$

$$\mathbf{h}_k(\mathbf{x}_k, \mathbf{u}_k) = \mathbf{0} \quad \forall\ k. \tag{2d}$$

[Can't we just define $\mathbf{u}_N = \mathbf{0}$ to simplify the expression in (2a)?] [I'm not certain that is the same thing as saying that $\mathbf{u}_N$ is undefined. What I think you are saying is $\ell_N(\mathbf{x}_N, \mathbf{u}_N = \mathbf{0})$, but what is really going on is that there is no dependence at all $\ell_N(\mathbf{x}_N)$]

Standard techniques for solving (2a)-(2d) typically fall into two categories: direct methods [85, 86] and indirect methods [87]. For indirect methods, the augmented controls are the *decision variables*, i.e., the variables the optimizer adjusts to solve the problem. The augmented states are obtained from the augmented controls using the discrete dynamics function, and they are used to evaluate derivatives of the cost functions. Then, the derivative information is employed to update the augmented controls. This approach is taken by standard QOC solvers such as GOAT [75], GRAPE [3, 17], and Krotov's method [76]. Conversely, direct methods treat

both the augmented controls and states as decision variables. In addition to minimizing the cost functions, the optimizer uses derivative information for the discrete dynamics function to satisfy the dynamics constraint (2b) to a specified tolerance. In this sense, the TDSE (1) is a constraint that may be violated for intermediate steps of the optimization, where the states need not be physical. The direct approach lends itself to a nonlinear program formulation, for which a variety of general-purpose solvers exist [88, 89]. Recent state-of-the-art solvers, such as ALTRO, combine the indirect and direct methods in a two-stage approach. First, ALTRO employs an indirect solving stage using the iterative linear-quadratic regulator (iLQR) algorithm [90] as the internal solver of an augmented Lagrangian method (ALM) [91–93]. In the second direct stage, ALTRO uses a projected Newton method [94, 95]. Next, we provide a more detailed summary of these two stages.

iLQR is an indirect method for minimizing the objective subject to the dynamics constraint, i.e., solving (2a)-(2b). [Start with the familiar or the understandable explanation. The imposing sounding names can be given afterwards. (Purely psychology.)] [resolved] First, iLQR uses an initial guess for the augmented controls to obtain the augmented states with the discrete dynamics function. iLQR then constructs quadratic models for each cost function using their zeroth-, first- and second-order derivatives in a Taylor expansion about the current augmented controls and states. These models are used in a recurrence relation between time steps to derive the locally optimal update for the augmented controls. [for some quantity (which?) at consecutive time steps] [resolved] Finally, a line search [96] is performed in the direction of this update to ensure a decrease in the objective. [again, I suspect that (2a) is not the "objective"; rather the objective is what I know as cost function, which would exclude the minimization that is part of (2a).] [resolved, i introduce objective earlier now] This procedure is repeated until convergence is reached.

While indirect solvers like iLQR are computationally efficient and maintain high accuracy for the discrete dynamics throughout the optimization, they cannot handle nonlinear equality and inequality constraints (2c)-(2d). [Seems like my edit attempts and understanding of the two-stage process are still not appropriate. On the previous page, it seemed like the projected Newton method was the final stage and a direct method at that. However, here it sounds now like it is not good enough and we need yet another method? I'm confused.] [removed discussion of projected gradient methods] For QOC, a popular approach to handle such constraints is to add the constraint functions to the objective [14, 17, 20]. [(2a) is of the form $\min_x L(x)$ do you really mean to add to the minimum? Or only to the cost function $L(x)$?] [resolved, i introduce objective earlier now] However, this strategy does not guarantee that the constraints are satisfied as the solver trades minimization of the cost functions and constraint functions against each other. ALM remedies this issue by adaptively adjusting a Lagrange multiplier estimate for each constraint function to ensure the constraints are satisfied. ALM adds terms that are linear and quadratic in the constraint functions to the objective. Then, the new objective is minimized with iLQR. If the solution obtained with iLQR does not satisfy the constraints, the prefactors for the constraint terms in the objective are increased intelligently and the procedure is repeated.

ALM converges superlinearly, but poor numerical conditioning may lead to small decreases in the constraint violations near the locally optimal solution [97]. To address this shortcoming, ALTRO projects the solution from the ALM stage onto the constraint manifold using a (direct) projected Newton method, achieving ultra-low constraint violations $\sim 10^{-8}$. For more information on the details of the ALTRO solver, see [74, 98].

As opposed to standard QOC solvers, ALTRO can satisfy constraints on both the augmented controls and states to tight tolerances. This advantage is crucial for this work, where multiple medium-priority cost functions are minimized subject to many high-priority constraints.

III. QOC FOR THE FLUXONIUM

In the following, we optimize quantum gates for the superconducting fluxonium qubit – a promising building block for quantum computers due to its high coherence times [73, 99–102] [for a physics journal, you will want as much handholding and explanation for the robotics community language as possible; by contrast, clarifying that a quantum gate is a unitary tranformation seems out of place here.] [resolved, removed unitary comment]. In this section, we use the trajectory optimization formalism (2a)-(2d) to define the optimization problem (6a)-(6i), which we extend in subsequent sections to account for experimental error channels. To high accuracy, we approximate the fluxonium Hamiltonian near the flux-frustration point as a two-level system:

$$H/h = f_q \frac{\sigma_z}{2} + a(t)\frac{\sigma_x}{2}. \tag{3}$$

Here, f_q is the qubit frequency at the flux-frustration point, $a(t)$ is the control governing the flux offset from the flux-frustration point, h is Planck's constant, and σ_z, σ_x are Pauli matrices. Although the coherent dynamics can be described with this two-level system model, our noise model, experimental constraints, and system parameters consider the full system, and they are representative of the fluxonium presented in [73].

First, we introduce the augmented control and state for this problem. To refer to the discrete moments of the flux, we introduce the notation $\int_{t_1}^{t_k} a(t)\ \mathrm{d}t \equiv \int_t a_k$, $a(t_k) \equiv a_k$, $\mathrm{d}^n a(t)/\mathrm{d}t^n|_{t=t_k} \equiv \mathrm{d}_t^n a_k$. The augmented con-

trol and state are:

$$\mathbf{u}_k = \begin{bmatrix} \mathrm{d}_t^2 a_k \end{bmatrix}, \quad \mathbf{x}_k = \begin{bmatrix} |\psi_k^0\rangle \\ |\psi_k^1\rangle \\ \int_t a_k \\ a_k \\ \mathrm{d}_t a_k \end{bmatrix}. \tag{4}$$

Here, the superscript on the quantum states $i \in \{0, 1\}$ acts as a label. (We maintain vector notation for $\mathbf{u}_k$ in anticipation of augmentation with additional components in Section IV.)] [I think this sentence is redundant given elaboration of augmented now in Section II] Since the ALTRO implementation we use does not currently support complex numbers, we represent the states in the isomorphism $\mathcal{H}(\mathbb{C}^n) \cong \mathcal{H}(\mathbb{R}^{2n})$ given in [17],

$$H|\psi\rangle \cong \begin{bmatrix} H_{\mathrm{re}} & -H_{\mathrm{im}} \\ H_{\mathrm{im}} & H_{\mathrm{re}} \end{bmatrix} \begin{bmatrix} |\psi\rangle_{\mathrm{re}} \\ |\psi\rangle_{\mathrm{im}} \end{bmatrix}. \tag{5}$$

In standard QOC frameworks, the derivatives of the control fields are obtained with finite difference methods, e.g. $\mathrm{d}_t a_k \approx (a_{k+1} - a_k)/\Delta t$ [17]. Because iLQR requires that cost functions depend only on the augmented control and state at a given time step, i.e. are of the form $\ell_k(\mathbf{x}_k, \mathbf{u}_k)$, we make $\mathrm{d}_t^2 a_k$ the decision variable and numerically integrate coupled ODEs to obtain $\mathrm{d}_t a_k$, a_k, and $\int_t a_k$ so that we may penalize them in cost functions. Similarly, the quantum states are obtained by numerically integrating the TDSE (1) with the fluxonium Hamiltonian (3) and the given flux a_k. These numerical integration rules are implemented in the discrete dynamics function for the problem, and give rise to the dynamics constraint (6b).

Next, we outline the constraints for the fluxonium gate problem. Casting this problem in terms of a multi-state transfer problem, we fix as the initial states $|\psi_1^0\rangle = |0\rangle$, $|\psi_1^1\rangle = |1\rangle$ (6c). The states at the final time step are constrained to be the target states $|\psi_N^i\rangle = |\psi_T^i\rangle \equiv U|\psi_1^i\rangle \ \forall\, i$ (6d) where $U = X/2, Y/2, Z/2$ denotes the target gate. Furthermore, we impose the normalization constraint $|\langle\psi_k^i|\psi_k^i\rangle|^2 = 1 \ \forall\, i, k$ (6e) to ensure the solver does not take advantage of discretization errors in numerical integration. We impose the zero net-flux constraint $\int_t a_N = 0$ (6f) which mitigates the inductive drift ubiquitous in flux-bias lines [73, 103, 104]. The flux is constrained by $|a_k| \leq 0.5$ GHz $\forall\, k$ (6g) to ensure the two-level approximation (3) remains valid. We also enforce the boundary condition $a_1 = a_N = 0$ (6h) [Thomas, this would actually be consistent with my suggestion to set $\mathbf{u}_N$ to zero?] [Not quite, because a_N is a component of $\mathbf{x}_N$, see (4)] so the gates may be concatenated arbitrarily. Additionally, we have the initial condition $\int_t a_1 = \mathrm{d}_t a_1 = 0$ (6i). All gates presented in this work satisfy these constraints to a maximum violation of $\sim 10^{-8}$.

The cost function at each time step is $\ell_k(\mathbf{x}_k, \mathbf{u}_k) = (\mathbf{x}_k - \mathbf{x}_T)^T Q_k (\mathbf{x}_k - \mathbf{x}_T) + \mathbf{u}_k^T R_k \mathbf{u}_k$ where Q_k and R_k are diagonal matrices we supply that assign weights to the components in the cost function. ["We supply" – sure, but not very helpful. How are they defined/constructed specifically?] [in QOC speak these are cost function weights, they are hyperparameters and set based on heuristics] The Q_k term penalizes deviations from the target augmented state $\mathbf{x}_T = \mathrm{col}(|\psi_T^0\rangle, |\psi_T^1\rangle, 0, 0, 0)$, which is consistent with the constraints we have imposed on $|\psi_N^i\rangle$, $\int_t a_N$, and a_N. Accordingly, this term penalizes the squared difference of $|\psi_k^i\rangle$ and $|\psi_T^i\rangle$ and penalizes the norm of $\int_t a_k$, a_k, and $\mathrm{d}_t a_k$. We penalize the squared difference of the final and target quantum states, rather than their infidelities, because the Hessian of the squared-difference cost function is diagonal – which makes matrix multiplications fast – and we wish to optimize $Z/2$ gates, which requires a metric that is sensitive to global phases for the initial states $|0\rangle$ and $|1\rangle$. Additionally, the R_k term penalizes the norm of $\mathrm{d}_t^2 a_k$. Penalizing the norm of $\mathrm{d}_t^2 a_k$ and $\mathrm{d}_t a_k$ makes a_k smooth, which mitigates high-frequency AWG transitions. Stated succinctly, the optimization problem takes the form:

$$\underset{\substack{\mathbf{x}_1,\ldots,\mathbf{x}_N \\ \mathbf{u}_1,\ldots,\mathbf{u}_{N-1}}}{\text{minimize}} \quad \sum_{k=1}^{N} (\mathbf{x}_k - \mathbf{x}_T)^T Q_k (\mathbf{x}_k - \mathbf{x}_T) + \sum_{k=1}^{N-1} {\mathbf{u}_k}^T R_k \mathbf{u}_k \tag{6a}$$

$$\text{subject to} \quad \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, u_k) \quad \forall\, k, \tag{6b}$$

$$|\psi_1^0\rangle = |0\rangle, \ |\psi_1^1\rangle = |1\rangle, \tag{6c}$$

$$|\psi_N^i\rangle = |\psi_T^i\rangle \quad \forall\, i, \tag{6d}$$

$$|\langle\psi_k^i|\psi_k^i\rangle|^2 = 1 \quad \forall\, i, k, \tag{6e}$$

$$\textstyle\int_t a_N = 0, \tag{6f}$$

$$|a_k| \leq 0.5 \text{ GHz} \quad \forall\, k, \tag{6g}$$

$$a_1 = a_N = 0, \tag{6h}$$

$$\textstyle\int_t a_1 = \mathrm{d}_t a_1 = 0. \tag{6i}$$

Next, we remark on our problem formulation. We put a cost function at all time steps (6a) because it benefits the iLQR solving stage [98]; putting a cost function at all time steps does not incentivize early achievement of the desired gate, as in [17], due to the target state constraint. [NOT to incentivize or to incentivize??] [NOT to incentivize] In addition to the target-state cost function, we impose a target-state constraint (6d) which requires the final state to match the target state, including its global phase, up to our chosen maximum constraint violation $\sim 10^{-8}$. If we did not impose this constraint, the optimizer would be allowed to sacrifice the closed-system gate error to achieve better performance on the other cost functions, which is undesirable. To enforce a constraint in standard QOC frameworks, the prefactor for the constraint function is manually increased between separate optimization instances until the constraint is satisfied [14, 17, 20], which becomes infeasible for more than one constraint. ALM automates these prefactor updates to find a solution that satisfies all of the given constraints. Hence, ALTRO's ability to handle multiple constraints

makes it an attractive solver for QOC problems.

In extraordinarily difficult cases of QOC [8], it may be impossible to obey the physics of the system and achieve the desired gate, i.e. the dynamics constraint (6b) and the target-state constraint (6d) may be mutually unsatisfiable. In this case, the prefactors for the constraint function terms in the ALM objective will tend to infinity – leading to numerical instability – and the optimization will not converge. To maintain a constrained approach in this situation, the maximum constraint violation for the target-state constraint can be raised to a level commensurate with the minimum acceptable gate error.

Finally, for the indirect ALM-iLQR solving stage, the states at each time step are obtained by integrating the TDSE with the discrete dynamics function, so the dynamics constraint (6b) is satisfied by construction. This is not the case for the direct projected Newton stage, where the states are free parameters that are adjusted to satisfy the TDSE. Although the final solution's deviation from the TDSE is never more than the maximum constraint violation, we explicitly integrate the TDSE when reporting gate errors to ensure accuracy. Exploring the benefit of direct optimization approaches for QOC is an interesting direction for future work.

IV. DEPOLARIZATION MITIGATION

In this section, we outline a method for optimizing the flux to mitigate depolarization. For many superconducting circuits, the depolarization time T_1 is independent of the control parameters, so the fastest possible gate incurs the least depolarization error [105]. For the fluxonium, however, T_1 is strongly dependent on the flux. We enable the optimizer to trade longer gate times for longer T_1 times, or shorter T_1 times for shorter gate times, by making the gate time a decision variable. Additionally, previous work has modeled the gate error due to depolarization by evolving density matrices under a master equation [42, 105], or evolving a large number of states in a quantum trajectory approach [106]. We avoid the increase in computational complexity required for these techniques by penalizing the integrated depolarization rate in optimization.

The integrated depolarization rate is given by,

$$D_1(t) = \int_0^t T_1^{-1}[a(t')]dt'. \tag{7}$$

For the gates we consider here, where the gate time is small compared to T_1, the integrated depolarization rate is proportional to the probability of a depolarization event. Additionally, the integrated depolarization rate is a reasonable proxy for the gate error incurred because depolarization errors are incoherent – they increase monotonically in time without interference. The integrated depolarization rate is appended to the augmented state (4) and its norm is penalized in the objective by setting the corresponding element of the target augmented state to zero. T_1 as a function of the flux is obtained by evaluating a spline fit to experimental data, see Figure 1(b). [Phys. Rev. will want this to read 1(b). Not sure how to change this the way you are doing automatic referencing.] [resolved]

Alternatively, modeling the depolarization with a master equation approach would require adding density matrices of size $n \times n$ to the augmented state, and a quantum trajectory approach would require adding many states of size n to the augmented state, where n is the dimension of the Hilbert space. By contrast, the integrated depolarization rate is a single real number; thus, the computational complexity of evaluating this depolarization model does not scale with the dimension of the Hilbert space.

[abrupt start of a new paragraph; provide transition/motivation for what's next; the entire following paragraph feels a bit out of place/inserted after the fact.] [better?] To perform time-optimal control, we make the duration between time steps a decision variable [74]. The square root of the duration $\sqrt{\Delta t_k}$ is appended to the augmented control (4) and its square $|\Delta t_k|$ is used for integration in the discrete dynamics function. Although we constrain the bounds of the duration between reasonable positive values to maintain numerical stability, the optimizer may assign negative values to the duration for intermediate optimization iterations, so this squaring approach maintains positivity.

We analyze the effect of depolarization on the $X/2$, $Y/2$, and $Z/2$ gates obtained with our numerical method and the corresponding analytic gates presented in [73]. We use the Lindblad master equation to simulate T_1 dissipation for successive gate applications, and compute the cumulative gate error after each application, see Appendix A. The gate error reported in this text is the infidelity of the evolved state and the target state averaged over 1000 pseudo-randomly generated initial states.

The flux pulses for the numerical gates are approximately periodic with amplitudes ~ 0.2GHz, see Figure 1(a). They are reminiscent of the analytically determined Floquet operations for a fluxonium described in [107] and realized in [108]. The numerical gate times are greater than the analytic gate times, but the numerical flux pulses spend more time at higher flux values, achieving higher T_1 times on average, see Figure 1(b). The single-gate errors for both the analytic and numerical gates are less than 10^{-4}, which makes them sufficient for quantum error correction–a prerequisite for fault-tolerant quantum computing [109–111]. However, the numerical gates achieve single-gate errors ~ 5 times less than those for the analytic gates, which tracks closely with their relative improvement on the integrated depolarization rate metric, see Appendix A. This advantage in single-gate errors corresponds to a significant reduction in error correction resources [112, 113]. Furthermore, for successive gate applications, the gate error due to depolarization is approximately linear in the gate count, which we expect for $t \ll T_1$, see Figure 1(c). The gate error reduction

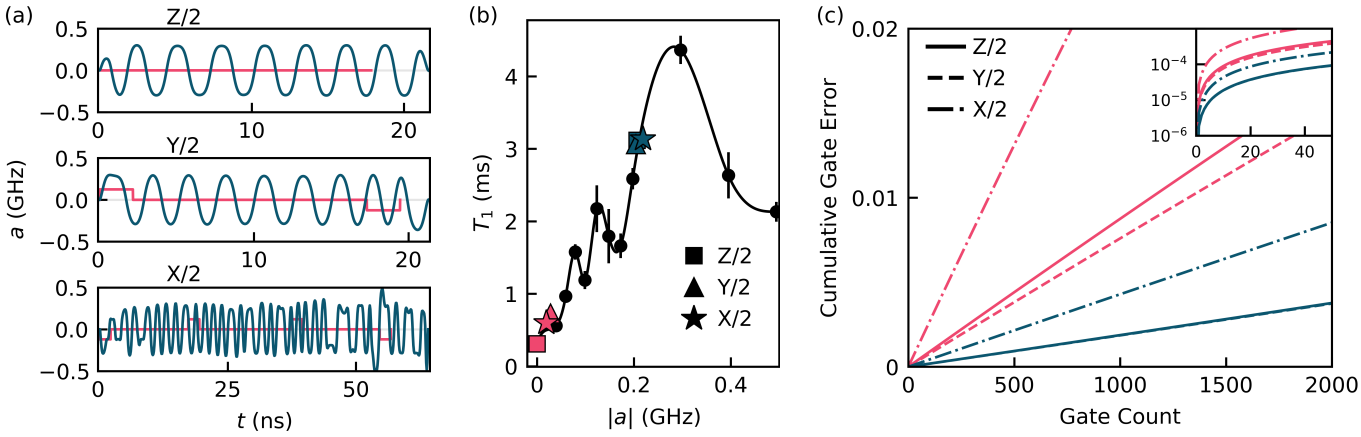

Figure 1: (a) Flux pulses for the numerical gates (dark blue) and the analytic gates (light pink). (b) T_1 interpolation function used in optimization. Circle markers indicate measured T_1 times. Non-circle markers are plotted at the time-averaged absolute flux and T_1 time for each pulse. (c) Cumulative gate errors due to depolarization as a function of the number of gates applied. Cumulative gate errors for the numerical $Z/2$ and $Y/2$ gates are indistinguishable. Inset shows log-scaled cumulative gate errors for small gate counts.

for large gate counts is important for noisy, intermediate-scale quantum (NISQ) applications. These improvements are significant given the constraints [improvements are significant FOR THE CONSTRAINTS?] [changed "for" to "given"] we have imposed on the gates, and do not represent a fundamental limit to the optimization methods we have employed.

V. ROBUSTNESS TO STATIC PARAMETER UNCERTAINTY

We have formulated the QOC problem as an *open-loop* optimization problem, i.e., we do not incorporate feedback from the experiment into the optimization. However, the precise device parameters will realistically deviate from the parameters we use in optimization, leading to poor experimental performance. We combat errors of this form using robust control techniques, making the state evolution insensitive to parameter uncertainty. As an example, we mitigate errors arising from the drift and finite measurement precision of the qubit frequency, which modifies the fluxonium Hamiltonian (3) by $f_q \to f_q + \delta f_q$. We consider three robust control techniques to accomplish this task: a sampling method, an unscented sampling method, and a derivative method.

A. Sampling Method

[Consider structuring the following into subsections to clearly separate the three methods.] [better?] The sampling method incentivizes the optimizer to ensure that multiple copies of a state, each evolving with a distinct value of the uncertain parameter, achieve the same target state. Variants of this technique have been proposed in the context of QOC [3, 20, 42, 63, 67] and applied experimentally [64]. For each initial state, we add two sample states $|\psi^{\pm}\rangle$ to the augmented state (4). The discrete dynamics function is modified so the sample states evolve under the fluxonium Hamiltonian (3) with $f_q \to f_q \pm \sigma_{f_q}$ for a fixed standard deviation σ_{f_q} of the qubit frequency, acting as a hyperparameter. We penalize the infidelities of the sample states with respect to the target state by adding a cost function to the objective of the form $\sum_{k,\pm} b_k(1 - |\langle\psi_k^{\pm}|\psi_T\rangle|^2)$ where b_k is a constant we supply. For this method, the standard orthonormal basis states are an insufficient choice for the initial states. As an example, a $Z/2$ gate achieved by idling at the flux frustration point ($a_k = 0\ \forall\ k$) will be robust to qubit frequency detunings for the initial states $|0\rangle$ or $|1\rangle$ because the infidelity metric is insensitive to global phases, but this gate will not be robust for any other initial states. Therefore, we choose the four initial states $\{|0\rangle, |1\rangle, (|0\rangle + i|1\rangle)/\sqrt{2}, (|0\rangle - |1\rangle)/\sqrt{2}\}$ [114], whose outer products span the operators on the Hilbert space, and we refer to them as the operator basis. [I cannot find the term "operator basis" in the rest of the paper. If not used, then don't introduce terminology.] [Operator basis is used in the appendix D]

B. Unscented Sampling Method

Whereas the sampling method penalizes the deviations of the sample states from the target state, the unscented sampling method penalizes the deviations of the sample states from the nominal state [68–70]. Accordingly, the cost function we add to the objective takes the form $\sum_{k,j} c_k(\psi_k^j - \psi_k)^T(\psi_k^j - \psi_k)$, where c_k is a constant we

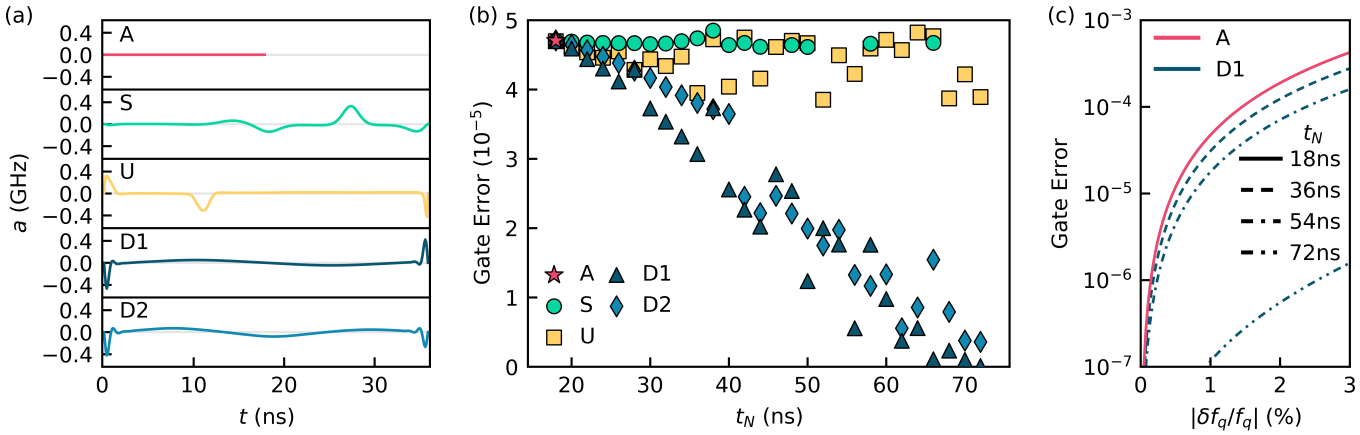

Figure 2: (a) Flux pulses for $Z/2$ gates robust to qubit frequency detunings constructed with the analytic (A), sampling (S), unscented sampling (U), and the 1st- and 2nd-order derivative methods (D1, D2). The flux pulses shown for the sampling, unscented sampling, and derivative methods are optimized for twice the gate time of the analytic gate. (b) Single-gate error at a one-percent qubit frequency detuning as a function of the gate time. Missing data points represent gates with a gate error greater than $5 \cdot 10^{-5}$. (c) Single-gate error as a function of the qubit frequency detuning. The gate errors for the analytic and 1st-order derivative methods are shown for gate times which are multiples of $1/4f_q \sim 18$ns. The gate errors for the two methods are indistinguishable at the gate time 18ns.

supply, ψ_k is the evolved initial state (nominal state), and ψ_k^j is a sample state that evolves under a modified Hamiltonian similar to that in the sampling method. We omit bra-ket notation here to emphasize that the states are real vectors and are given by the right-hand-side of the complex-to-real isomorphism (5). The sample states are chosen to encode a unimodal distribution over the $2n$ elements of the nominal state, modeling the uncertainty in the state as a result of the uncertainty in the parameter. We use the unscented transform [71, 72] to accurately propagate the mean and covariance of this distribution between time steps, or equivalently, through the transformation of the TDSE (1). Unlike the sampling method, the cost function for the unscented sampling method is sensitive to global phases. Accordingly, we do not observe a performance increase when using more than one initial state. A detailed procedure for the unscented transformation is given in Appendix B.

C. Derivative Method

The derivative method penalizes the sensitivity of the state to the uncertain parameter, which is encoded in the l^{th}-order state derivative $|\partial_{f_q}^l \psi\rangle \equiv \partial_{f_q}^l |\psi\rangle$. In the m^{th}-order derivative method, we append all state derivatives of order $1, \ldots, m$ to the augmented state vector (4) for each initial state. We obtain the state derivatives at each time step by performing forward-mode differentiation on the TDSE (1). For example, the dynamics for the 1st-order derivative method are:

$$i\hbar \frac{d}{dt} |\psi\rangle = H |\psi\rangle , \tag{8}$$

$$i\hbar \frac{d}{dt} |\partial_{f_q}\psi\rangle = H |\partial_{f_q}\psi\rangle + (\partial_{f_q} H) |\psi\rangle . \tag{9}$$

We integrate the coupled ODEs with exponential integrators, see Appendix C. [integrators IN A function? sounds like mix-up between the function vs. how it is implemented numerically] [removed ref. to ddf] While the state $|\psi\rangle$ has unit norm, the state derivatives $|\partial_{f_q}^l \psi\rangle$ need not, as is evident from the non-unitary dynamics (9). We penalize the norms of the state derivatives in the objective by setting the corresponding elements of the target augmented state to zero. Intuitively, this corresponds to penalizing the sensitivity of each state element to the uncertain parameter. As was the case for the unscented sampling method, we do not observe a performance increase when using more than one initial state for the derivative method. We present the runtimes of our implementations of the three robust control methods in Appendix D. [a "technique" is abstract/general; a technique does not have a concrete runtime; do you mean computational complexity?], [better?]

D. Comparison

We examine the gate errors due to a static qubit frequency detuning for the $Z/2$ gates obtained with the robust control techniques and the analytic $Z/2$ gate. To compute the gate error, an initial state is evolved under the fluxonium Hamiltonian (3) two separate times with

the transformations $f_q \to f_q \pm \delta f_q$ at the stated qubit frequency detuning δf_q. The reported gate error is the infidelity of the evolved state and the target state averaged over the two transformations for each of 1000 pseudorandomly generated initial states. We set $\sigma_{f_q}/f_q = 1\%$ for the sampling and unscented sampling methods.

The analytic gate corresponds to idling at the flux frustration point $a_k = 0 \; \forall \; k$, see Figure 2(a). Its gate time $1/4f_q \sim 18$ns is the shortest possible for a $Z/2$ gate on the device. The gate's erroneous rotation angle $2\pi\delta f_q/4f_q$ is linear in the qubit frequency detuning, resulting in a gate error that is quadratic in the detuning. At a one-percent detuning $|\delta f_q/f_q| = 1\%$, the gate error is $\sim 4.7 \cdot 10^{-5}$, which is sufficient for quantum error correction.

For the sampling method, the gate error at a one-percent qubit frequency detuning does not decrease substantially over the range of gate times, and begins to increase above $5 \cdot 10^{-5}$ for gate times greater than ~ 50ns, see Figure 2(b). Optimization results for the sampling method reveal that it is typically able to achieve a high fidelity for one sample $|\psi^{\pm}\rangle$, but not the other $|\psi^{\mp}\rangle$, indicating that it is difficult for the optimizer to make progress on both objectives. For the unscented sampling method, the gate error at a one-percent detuning does not decrease substantially over the gate times, but it does reach a minimum of $\sim 3.9 \cdot 10^{-5}$ near fractions of the Larmor period $2/4f_q \sim 36$ns, $3/4f_q \sim 54$ns, $4/4f_q \sim 72$ns.

The two derivative methods converge on qualitatively similar flux pulses that idle near the flux frustration point and use fast triangle movements at the boundaries, similar to the flux pulse produced by the unscented sampling method. For both derivative methods, the gate error at a one-percent qubit frequency detuning decreases superlinearly in the gate time. For the 1^{st}-order method, the gate error at a one-percent detuning reaches 10^{-7} at the Larmor period $1/f_q \sim 72$ns, see Figure 2(c). This result mimics the ability of composite pulses to mitigate parameter uncertainty errors to arbitrary order with sufficiently many pulses [55]. It is difficult to choose an appropriate composite pulse for the problem studied here due to our Hamiltonian and experimental constraints. We propose comparisons between composite pulses and numerical techniques for future work.

Furthermore, the ability to perform Z-type gates in any given time is critical for synchronizing phases in multi-qubit experiments, where the qubits have distinct frequencies. Notably, the analytic gate studied here cannot be extended to gate times other than $1/4f_q$. We can find gates using the numerical methods at all gate times above 18ns, see Figure 2(b). These numerical methods offer an effective scheme for synchronizing multi-qubit experiments.

VI. ROBUSTNESS TO TIME-DEPENDENT PARAMETER UNCERTAINTY

An additional source of experimental error arises from time-dependent parameter uncertainty. For many flux-biased and inductively-coupled superconducting circuit elements, magnetic flux noise is the dominant source of coherent errors [115–118]. Flux noise modifies the fluxonium Hamiltonian (3) by $a(t) \to a(t) + \delta a(t)$ where $\delta a(t)$ is the flux noise. The spectral density of flux noise is observed to follow a $1/f$ distribution [73, 115–120], so the noise is dominated by low-frequency components. The analytic gate considered here takes advantage of the low-frequency characteristic and treats the noise as quasi-static, performing a generalization of the spin-echo technique to compensate for erroneous drift [121, 122].

We modify the robust control techniques presented in the previous section to combat $1/f$ flux noise. The unscented sampling method is modified so that the sample states are subject to $1/f$ flux noise. The noise is generated by filtering white noise sampled from a standard normal distribution with a finite impulse response filter [123]. The noise is then scaled by the flux noise amplitude of our device $A_\Phi = 5.21\mu\Phi_0 \implies \sigma_a = 2.5 \cdot 10^{-5}$GHz. In principle, we could modify the sampling method similarly; however, we choose to subject the sample states to static noise $a(t) \to a(t) \pm \sigma_a$ for comparison. The derivative methods require no algorithmic modification from the static case, but the TDSE is now differentiated with respect to $a(t)$ instead of f_q as in (9).

We analyze the gate errors due to $1/f$ flux noise for the $X/2$ gates constructed with the robust control techniques and the analytic $X/2$ gate. To compute the gate error, we evolve an initial state under the fluxonium Hamiltonian (3) where the optimized flux is modified $a(t) \to a(t) + \delta a(t)$. We generate the flux noise as we described for the unscented sampling method. The reported gate error is the infidelity averaged over 1000 pseudorandomly generated initial states, each of which is subject to a distinct pseudorandomly generated flux noise instance. To observe the effect of interfering coherent errors, we simulate successive applications of the gate constructed by each method; we compute the cumulative gate error after each application, see Figure 3. Both the analytic and numerical gates yield single-gate errors sufficient for quantum error correction. Despite converging on qualitatively different solutions, the numerical gates perform similarly in the concatenated gate application comparison. Their gate errors after 200 gate applications $\sim 11\mu$s are two orders of magnitude less than the gate error produced by the analytic gate. $1/f$ flux noise is a significant source of coherent errors in NISQ applications, and these numerical techniques offer effective avenues to mitigate it.

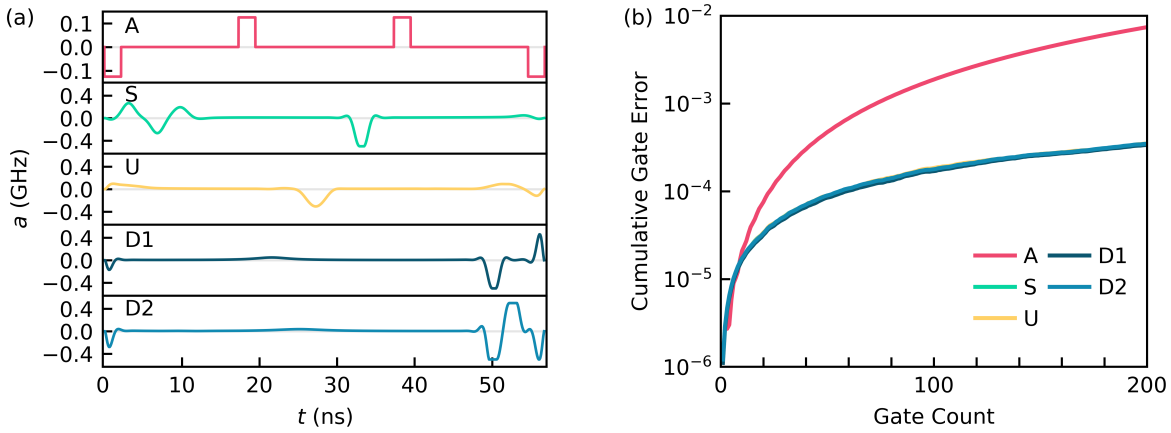

Figure 3: (a) Flux pulses for $X/2$ gates robust to flux noise constructed with the analytic (A), sampling (S), unscented sampling (U), and the 1st- and 2nd-order derivative methods (D1, D2). (b) Cumulative gate error due to $1/f$ flux noise for successive gate applications. The cumulative gate errors for the sampling, unscented sampling, and the derivative methods are indistinguishable.

VII. CONCLUSION

We have applied state-of-the-art trajectory optimization techniques to mitigate decoherence and achieve robustness to parameter uncertainty errors on a quantum system. We have proposed a scheme for suppressing depolarization with time-optimal control and the integrated depolarization rate model. The computational complexity of evaluating this model is independent of the dimension of the Hilbert space, enabling inexpensive optimization on high-dimensional quantum systems. [I would argue that a *model* has no cost... Computing something does.] [better?] We have also proposed the derivative method for robust control which achieves super-linear gate error reductions in the gate time for the static parameter uncertainty problem we studied. We have shown that the derivative, sampling, and unscented sampling methods can mitigate $1/f$ flux noise errors–which dominate coherent errors for flux controlled qubits. These robust control techniques can be applied to any Hamiltonian, allowing experimentalists in all domains to engineer robust operations on their quantum systems. These methods can be used to achieve the low gate errors required for fault-tolerant quantum computing applications. Our implementations of the techniques described in this work are available at `https://github.com/SchusterLab/rbqoc`.

ACKNOWLEDGMENTS

We thank Helin Zhang for experimental assistance and Taylor Howell, Tanay Roy, Colm Ryan, and Daniel Weiss for useful discussions. This work is funded in part by EPiQC, an NSF Expedition in Computing, under grant CCF-1730449. This work was supported by the Army Research Office under Grant No. W911NF1910016. This work was made possible by many open source software projects, including but not limited to: Altro.jl [74], DifferentialEquations.jl [124], Distributions.jl [125], ForwardDiff.jl [126], Matplotlib [127], NumPy [128], and Zygote.jl [129].

Appendix A: Depolarization

We comment on the depolarization metrics and then give our procedure for integrating the Lindblad master equation. The integrated depolarization rate and the gate error due to depolarization [The error is due to depolarization METRICS?] [resolved, removed the word metrics] are compared in Table I for the numerical experiment described in Section IV. The ratio of the value obtained on the metric with the analytic technique to the value obtained with the numerical technique is similar across the two metrics. what is relative performance? replaced relative performance with more precise phrasing

Gate	$D_{1\mathrm{A}}$ (10^{-5})	$D_{1\mathrm{N}}$ (10^{-5})	$D_{1\mathrm{A}}/D_{1\mathrm{N}}$	$\mathrm{GE_A}$ (10^{-5})	$\mathrm{GE_N}$ (10^{-5})	$\mathrm{GE_A}/\mathrm{GE_N}$
Z/2	5.745	1.149	5.000	0.888	0.185	4.791
Y/2	5.253	1.157	4.540	0.770	0.186	4.132
X/2	16.251	2.660	6.109	2.674	0.432	6.200

Table I: Single-gate integrated depolarization rate (D_1) and single-gate error due to depolarization (GE). Values are reported for the analytic (A) and numerical (N) gates.

We employ the Lindblad master equation to compute the gate error due to depolarization. This equation takes

the form:

$$\frac{d}{dt}\rho = -\frac{i}{\hbar}[H,\rho] + \sum_i \gamma_i (L_i \rho L_i^\dagger - \frac{1}{2}\{L_i^\dagger L_i, \rho\}), \quad \text{(A1)}$$

[I suggest to remove the rest of the sentence completely; should not state any of this for physics journal] [sentence on commutator definitions removed] For depolarization, $\gamma_\pm = T_\pm^{-1}$, $L_\pm = \sigma^\pm \equiv (\sigma_x \pm i\sigma_y)/2$. The depolarization times $T_+ = T_- = 2T_1$ are obtained at each time step from the spline shown in Figure 1(b). We obtain the T_1 values in this spline by driving the qubit at the desired flux bias and monitoring the resultant decay. For more details on these measurements, consult [73]. Because T_1 depends on the flux $a(t)$, so do the decay rates $\gamma_\pm$. Integrating the master equation with time-dependent decay rates provides a heuristic for how gates might perform in the experiment. This procedure may not be strictly correct when decay rates change significantly on the time scale of the relaxation time, which is the regime we are operating in. Standard derivations of the Lindblad master equation do not account for time-dependent decay rates [130]. A more thorough treatment of this regime in future work would unlock new insights for quantum computing platforms where decoherence is strongly dependent on the control parameters.

In order to use exponential integrators, we employ the vectorization/Choi-Jamiolkowski isomorphism [131],

$$\frac{d}{dt}\text{vec}(\rho) = \hat{\mathcal{L}}\,\text{vec}(\rho), \quad \text{(A2)}$$

$$\hat{\mathcal{L}} = -i(\mathbb{1}\otimes H - H^T \otimes \mathbb{1}) + \sum_i \gamma_i (L_i^* \otimes L_i - \frac{1}{2}(\mathbb{1}\otimes L_i^\dagger L_i - L_i^T L_i^* \otimes \mathbb{1})), \quad \text{(A3)}$$

where $\rho = \sum_{i,j} \alpha_{ij} |i\rangle\langle j|$ and $\text{vec}(\rho) = \sum_{i,j} \alpha_{ij} |i\rangle \otimes |j\rangle$. Because $a(t)$ is constant between time steps $[t_k, t_{k+1})$, due to our numerical discretization, [for real, because of the AWG? even that would get filtered, right? or as an approximation because of discretization?] [added comment bout discretization], so are $H(a(t))$ and $\gamma_\pm(a(t))$. Therefore, the exact solution is $\text{vec}(\rho_{k+1}) = \exp(\Delta t_k \hat{\mathcal{L}}_k)\text{vec}(\rho_k)$. The vector isomorphism transforms $(n\times n)\times(n\times n)$ matrix-matrix multiplications to $(n^2\times n^2)\times n^2$ matrix-vector multiplications. For small n, we find that it is faster to use an exponential integrator on the vectorized equation than to perform Runge-Kutta on the unvectorized equation. [what is zero-order hold?] [removed ZOH comment] The latter requires decreasing the interval Δt_k to maintain accuracy, resulting in more time steps.

Appendix B: Unscented Sampling Method

In this section, we outline the full unscented sampling procedure. We consider a state $\psi \in \mathbb{R}^{2n}$, an uncertain set of parameters $\lambda \in \mathbb{R}^d$, and discrete dynamics $\psi_{k+1} = f(\psi_k, \lambda_k)$. We abandon bra-ket notation to emphasize that the state is a real vector, and it is given by the right-hand-side of the complex-to-real isomorphism (5). The nominal initial state is given by $\bar{\psi}_1$ with an associated covariance matrix $P_1 \in \mathbb{S}^{2n}_{++}$ [why do we need a double plus?] [the convention is + = positive-semidefinite, ++ = positive-definite] which describes the uncertainty in the initial state. We use the notation $\mathbb{S}^m_{++}$ to denote the set of real, symmetric, and positive-definite $m\times m$ matrices. By the positive-definite requirement, P_1 must be non-zero even if the state-preparation error is negligible. The uncertain parameter has zero-mean and its distribution is given by the covariance matrix $L_k \in \mathbb{S}^d_{++}$ at time step k. The zero-mean assumption is convenient for deriving the update procedure. A non-zero mean can be encoded in the discrete dynamics function $f(\psi_k, \lambda_k)$.

The initial $4n+2d$ sample states and initial $4n+2d$ uncertain parameters are sampled from the initial distributions,

$$\begin{bmatrix}\psi_1^j \\ \lambda_1^j\end{bmatrix} = \begin{bmatrix}\bar{\psi}_1 \\ 0\end{bmatrix} \pm \beta \sqrt{\begin{bmatrix}P_1 & 0 \\ 0 & L_1\end{bmatrix}}^j. \quad \text{(B1)}$$

β is a hyperparameter that controls the spacing of the covariance contour. The $(\pm)$ is understood to take $(+)$ for $j \in \{1,\ldots,2n+d\}$ and $(-)$ for $j \in \{2n+d+1,\ldots,4n+2d\}$. We use the Cholesky factorization to compute the square root of the joint covariance matrix, though other methods such as the principal square root may be employed. The superscript on the matrix square root indicates the j^{th} column (mod $2n+d$) of the lower triangular Cholesky factor. Then, the sample states are normalized,

$$\psi_1^j \to \frac{\psi_1^j}{\sqrt{{\psi_1^j}^T \psi_1^j}}. \quad \text{(B2)}$$

The sample states are propagated to the next time step,

$$\psi_2^j = f(\psi_1^j, \lambda_1^j). \quad \text{(B3)}$$

The mean and covariance of the sample states are computed,

$$\bar{\psi}_2 = \frac{1}{4n+2d}\sum_{j=1}^{4n+2d} \psi_2^j, \quad \text{(B4)}$$

$$P_2 = \frac{1}{2\beta^2}\sum_{j=1}^{4n+2d} (\psi_2^j - \bar{\psi}_2)(\psi_2^j - \bar{\psi}_2)^T. \quad \text{(B5)}$$

The sample states are then resampled and propagated

to the next time step using (B1), (B2), and (B3). Our choice of sample states (sigma points) follows equation 11 of [71]. Prescriptions that require fewer sigma points exist [132].

Appendix C: Derivative Method

Here, we outline how to efficiently integrate the dynamics for the derivative method using exponential integrators. General exponential integrators break the dynamics into a linear term and a non-linear term. For example, the dynamics for the first state derivative are $\frac{d}{dt}|\partial_\lambda\psi\rangle = -\frac{i}{\hbar}H|\partial_\lambda\psi\rangle - \frac{i}{\hbar}(\partial_\lambda H)|\psi\rangle$. The linear term is $L = -\frac{i}{\hbar}H$ and the non-linear term is $N = -\frac{i}{\hbar}(\partial_\lambda H)|\psi\rangle$. With zero-order hold on the controls the exact solution is:

$$|\partial_\lambda\psi_{k+1}\rangle = \exp(\Delta t_k L_k)|\partial_\lambda\psi_k\rangle + \int_0^{\Delta t_k}\exp\Big((\Delta t_k - t')L_k\Big)N(t_k+t')dt'. \tag{C1}$$

General exponential integrators proceed by breaking the integral in (C1) into a discrete sum, similar to the procedure for Runge-Kutta schemes. We use a simple approximation known as the Lawson-Euler method [82],

$$|\partial_\lambda\psi_{k+1}\rangle \approx \exp(\Delta t_k L_k)|\partial_\lambda\psi_k\rangle + \exp(\Delta t_k L_k)N_k\Delta t_k. \tag{C2}$$

This method provides a good tradeoff between accuracy and efficiency, requiring one unique matrix exponential computation per stage. Integration accuracy is not of the utmost importance because the state derivatives guide the optimization, and do not correspond to experimental parameters which must be realized with high accuracy.

Appendix D: Computational Performance

We provide runtimes for our optimizations and comment on the problem size of the robustness methods. The runtimes for the base optimization in Section III, the depolarization optimization in Section IV, and the robust optimizations in Section V are presented in Table II for a $Z/2$ gate at gate times which are multiples of $1/4f_q \sim 18$ns. We performed optimizations on a single core of an AMD Ryzen Threadripper 3970X 32-Core Processor. Future work will parallelize the robustness methods using GPUs [17], which will enable fast optimizations on high-dimensional Hilbert spaces.

	Average Runtime (s)		
t_N (ns)	18	36	72
Base	0.155 ± 0.008	7.0 ± 0.4	15.9 ± 0.8
Depol.	1.69 ± 0.08	-	-
S	1.77 ± 0.09	48 ± 2	280 ± 10
U	75 ± 4	340 ± 20	400 ± 20
D1	6.1 ± 0.3	27 ± 1	65 ± 3
D2	15.7 ± 0.8	17.3 ± 0.9	54 ± 3

Table II: Average runtimes for $Z/2$ optimizations using the base, depolarization, sampling (S), unscented sampling (U), and the 1st- and 2nd-order derivative methods (D1, D2).

Now we note the size of the augmented state vector for the robustness methods. For the sampling method, the size of the augmented state vector is $O(dn^3)$, where d is the number of uncertain parameters and n is the dimension of the Hilbert space. There are n^2 initial states in the operator basis, $2d$ sample states per initial state, and each state has $2n$ real numbers. For the unscented sampling method, the size of the augmented state vector is $O(dn^3+n^4)$. There are n^2 initial states in the operator basis, $2(2n+d)$ sample states per initial state, and each state has $2n$ real numbers. For the m^{th}-order derivative method, the size of the augmented state vector is $O(dmn^3)$. There are n^2 initial states in the operator basis, dm state derivatives per initial state, and each state has $2n$ real numbers.

[1] L. M. K. Vandersypen and I. L. Chuang, Nmr techniques for quantum control and computation, Rev. Mod. Phys. **76**, 1037 (2005).

[2] C. T. Kehlet, A. C. Sivertsen, M. Bjerring, T. O. Reiss, N. Khaneja, S. J. Glaser, and N. C. Nielsen, Improving solid-state nmr dipolar recoupling by optimal control, Journal of the American Chemical Society **126**, 10202 (2004).

[3] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Optimal control of coupled spin dynamics: design of nmr pulse sequences by gradient ascent algorithms, Journal of magnetic resonance **172**, 296 (2005).

[4] I. I. Maximov, Z. Tošner, and N. C. Nielsen, Optimal control design of nmr and dynamic nuclear polarization experiments using monotonically convergent algorithms, The Journal of Chemical Physics **128**, 184505 (2008).

[5] N. C. Nielsen, C. Kehlet, S. J. Glaser, and N. Khaneja, Optimal control methods in nmr spectroscopy, in *eMagRes* (American Cancer Society, 2010).

[6] T. E. Skinner, T. O. Reiss, B. Luy, N. Khaneja, and S. J.

Glaser, Application of optimal control theory to the design of broadband excitation pulses for high-resolution nmr, Journal of Magnetic Resonance **163**, 8 (2003).

[7] Z. Tošner, T. Vosegaard, C. Kehlet, N. Khaneja, S. J. Glaser, and N. C. Nielsen, Optimal control in nmr spectroscopy: Numerical implementation in simpson, Journal of Magnetic Resonance **197**, 120 (2009).

[8] M. Abdelhafez, B. Baker, A. Gyenis, P. Mundada, A. A. Houck, D. Schuster, and J. Koch, Universal gates for protected superconducting qubits using optimal control, Phys. Rev. A **101**, 022321 (2020).

[9] S. Chakram, K. He, A. V. Dixit, A. E. Oriani, R. K. Naik, N. Leung, H. Kwon, W.-L. Ma, L. Jiang, and D. I. Schuster, Multimode photon blockade (2020), arXiv:2010.15292 [quant-ph].

[10] D. J. Egger and F. K. Wilhelm, Optimized controlled-z gates for two superconducting qubits coupled through a resonator, Superconductor Science and Technology **27**, 014001 (2013).

[11] R. Fisher, F. Helmer, S. J. Glaser, F. Marquardt, and T. Schulte-Herbrüggen, Optimal control of circuit quantum electrodynamics in one and two dimensions, Phys. Rev. B **81**, 085328 (2010).

[12] P. Gokhale, Y. Ding, T. Propson, C. Winkler, N. Leung, Y. Shi, D. I. Schuster, H. Hoffmann, and F. T. Chong, Partial compilation of variational algorithms for noisy intermediate-scale quantum machines, in *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture* (2019) pp. 266–278.

[13] S.-Y. Huang and H.-S. Goan, Optimal control for fast and high-fidelity quantum gates in coupled superconducting flux qubits, Phys. Rev. A **90**, 012318 (2014).

[14] R. W. Heeres, P. Reinhold, N. Ofek, L. Frunzio, L. Jiang, M. H. Devoret, and R. J. Schoelkopf, Implementing a universal gate set on a logical qubit encoded in an oscillator, Nature communications **8**, 1 (2017).

[15] J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, and J. M. Martinis, Optimal quantum control using randomized benchmarking, Phys. Rev. Lett. **112**, 240504 (2014).

[16] Z. Leng, P. Mundada, S. Ghadimi, and A. Houck, Robust and efficient algorithms for high-dimensional black-box quantum optimization (2019), arXiv:1910.03591 [quant-ph].

[17] N. Leung, M. Abdelhafez, J. Koch, and D. Schuster, Speedup for quantum optimal control from automatic differentiation based on graphics processing units, Phys. Rev. A **95**, 042318 (2017).

[18] S. Li, T. Chen, and Z.-Y. Xue, Fast holonomic quantum computation on superconducting circuits with optimal control, Advanced Quantum Technologies **3**, 2000001 (2020).

[19] P. J. Liebermann and F. K. Wilhelm, Optimal qubit control using single-flux quantum pulses, Phys. Rev. Applied **6**, 024022 (2016).

[20] P. Reinhold, *Controlling Error-Correctable Bosonic Qubits*, Ph.D. thesis, Yale University (2019).

[21] P. Rebentrost and F. K. Wilhelm, Optimal control of a leaking qubit, Phys. Rev. B **79**, 060507 (2009).

[22] P. Rebentrost, I. Serban, T. Schulte-Herbrüggen, and F. K. Wilhelm, Optimal control of a qubit coupled to a non-markovian environment, Phys. Rev. Lett. **102**, 090401 (2009).

[23] R. J. Spiteri, M. Schmidt, J. Ghosh, E. Zahedinejad, and B. C. Sanders, Quantum control for high-fidelity multi-qubit gates, New Journal of Physics **20**, 113009 (2018).

[24] A. Spörl, T. Schulte-Herbrüggen, S. J. Glaser, V. Bergholm, M. J. Storcz, J. Ferber, and F. K. Wilhelm, Optimal control of coupled josephson qubits, Phys. Rev. A **75**, 012302 (2007).

[25] I. Brouzos, A. I. Streltsov, A. Negretti, R. S. Said, T. Caneva, S. Montangero, and T. Calarco, Quantum speed limit and optimal control of many-boson dynamics, Phys. Rev. A **92**, 062110 (2015).

[26] G. De Chiara, T. Calarco, M. Anderlini, S. Montangero, P. J. Lee, B. L. Brown, W. D. Phillips, and J. V. Porto, Optimal control of atom transport for quantum gates in optical lattices, Phys. Rev. A **77**, 052333 (2008).

[27] M. Grace, C. Brif, H. Rabitz, I. A. Walmsley, R. L. Kosut, and D. A. Lidar, Optimal control of quantum gates and suppression of decoherence in a system of interacting two-level particles, Journal of Physics B: Atomic, Molecular and Optical Physics **40**, S103 (2007).

[28] M. H. Goerz, T. Calarco, and C. P. Koch, The quantum speed limit of optimal controlled phasegates for trapped neutral atoms, Journal of Physics B: Atomic, Molecular and Optical Physics **44**, 154011 (2011).

[29] J. Guo, X. Feng, P. Yang, Z. Yu, L. Q. Chen, C.-H. Yuan, and W. Zhang, High-performance raman quantum memory with optimal control in room temperature atoms, Nature Communications **10**, 148 (2019).

[30] J. H. M. Jensen, J. J. Sørensen, K. Mølmer, and J. F. Sherson, Time-optimal control of collisional $\sqrt{\text{swap}}$ gates in ultracold atomic systems, Phys. Rev. A **100**, 052314 (2019).

[31] A. Larrouy, S. Patsch, R. Richaud, J.-M. Raimond, M. Brune, C. P. Koch, and S. Gleyzes, Fast navigation in a large hilbert space using quantum optimal control, Phys. Rev. X **10**, 021058 (2020).

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

[33] A. Omran, H. Levine, A. Keesling, G. Semeghini, T. T. Wang, S. Ebadi, H. Bernien, A. S. Zibrov, H. Pichler, S. Choi, J. Cui, M. Rossignolo, P. Rembold, S. Montangero, T. Calarco, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, Generation and manipulation of schrödinger cat states in rydberg atom arrays, Science **365**, 570 (2019).

[34] S. Rosi, A. Bernard, N. Fabbri, L. Fallani, C. Fort, M. Inguscio, T. Calarco, and S. Montangero, Fast closed-loop optimal control of ultracold atoms in an optical lattice, Phys. Rev. A **88**, 021601 (2013).

[35] P. Treutlein, T. W. Hänsch, J. Reichel, A. Negretti, M. A. Cirone, and T. Calarco, Microwave potentials and optimal control for robust quantum gates on an atom chip, Phys. Rev. A **74**, 022312 (2006).

[36] S. van Frank, M. Bonneau, J. Schmiedmayer, S. Hild, C. Gross, M. Cheneau, I. Bloch, T. Pichler, A. Negretti, T. Calarco, *et al.*, Optimal control of complex atomic quantum systems, Scientific reports **6**, 34187 (2016).

[37] Y. Chou, S.-Y. Huang, and H.-S. Goan, Optimal control of fast and high-fidelity quantum gates with elec-

tron and nuclear spins of a nitrogen-vacancy center in diamond, Phys. Rev. A **91**, 052315 (2015).
[38] F. Dolde, V. Bergholm, Y. Wang, I. Jakobi, B. Naydenov, S. Pezzagna, J. Meijer, F. Jelezko, P. Neumann, T. Schulte-Herbrüggen, *et al.*, High-fidelity spin entanglement using optimal control, Nature communications **5**, 1 (2014).
[39] J. Geng, Y. Wu, X. Wang, K. Xu, F. Shi, Y. Xie, X. Rong, and J. Du, Experimental time-optimal universal control of spin qubits in solids, Phys. Rev. Lett. **117**, 170501 (2016).
[40] T. Nöbauer, A. Angerer, B. Bartels, M. Trupke, S. Rotter, J. Schmiedmayer, F. Mintert, and J. Majer, Smooth optimal quantum control for robust solid-state spin magnetometry, Phys. Rev. Lett. **115**, 190801 (2015).
[41] F. Poggiali, P. Cappellaro, and N. Fabbri, Optimal control for one-qubit quantum sensing, Phys. Rev. X **8**, 021059 (2018).
[42] P. Rembold, N. Oshnik, M. M. Müller, S. Montangero, T. Calarco, and E. Neu, Introduction to quantum optimal control for quantum sensing with nitrogen-vacancy centers in diamond, AVS Quantum Science **2**, 024701 (2020).
[43] J. Tian, T. Du, Y. Liu, H. Liu, F. Jin, R. S. Said, and J. Cai, Optimal quantum optical control of spin in diamond, Phys. Rev. A **100**, 012110 (2019).
[44] S. Amri, R. Corgier, D. Sugny, E. M. Rasel, N. Gaaloul, and E. Charron, Optimal control of the transport of bose-einstein condensates with atom chips, Scientific reports **9**, 1 (2019).
[45] P. Doria, T. Calarco, and S. Montangero, Optimal control technique for many-body quantum dynamics, Phys. Rev. Lett. **106**, 190501 (2011).
[46] J. J. Sørensen, J. Jensen, T. Heinzel, and J. F. Sherson, Qengine: A c++ library for quantum optimal control of ultracold atoms, Computer Physics Communications **243**, 135 (2019).
[47] J. J. W. H. Sørensen, M. O. Aranburu, T. Heinzel, and J. F. Sherson, Quantum optimal control in a chopped basis: Applications in control of bose-einstein condensates, Phys. Rev. A **98**, 022119 (2018).
[48] P. V. Klimov, J. Kelly, J. M. Martinis, and H. Neven, The snake optimizer for learning quantum processor control parameters (2020), arXiv:2006.04594 [quant-ph].
[49] K. Zhou, *Essentials of Robust Control*, 1st ed. (Pearson, 1997).
[50] J. Morimoto and C. Atkeson, Minimax differential dynamic programming: An application to robust biped walking, Advances in neural information processing systems **15**, 1563 (2002).
[51] Z. Manchester and S. Kuindersma, Robust direct trajectory optimization using approximate invariant funnels, Autonomous Robots 10.1007/s10514-018-9779-5 (2018).
[52] H. K. Cummins and J. A. Jones, Use of composite rotations to correct systematic errors in NMR quantum computation, New Journal of Physics **2**, 6 (2000).
[53] H. K. Cummins, G. Llewellyn, and J. A. Jones, Tackling systematic errors in quantum logic gates with composite rotations, Phys. Rev. A **67**, 042308 (2003).
[54] Ä. Kupce and R. Freeman, Stretched adiabatic pulses for broadband spin inversion, Journal of Magnetic Resonance, Series A **117**, 246 (1995).
[55] J. T. Merrill and K. R. Brown, Progress in compensating pulse sequences for quantum computation, Quantum Information and Computation for Chemistry , 241 (2014).
[56] Z. Han, Y. Dong, B. Liu, X. Yang, S. Song, L. Qiu, D. Li, J. Chu, W. Zheng, J. Xu, *et al.*, Experimental realization of universal time-optimal non-abelian geometric gates (2020), arXiv:2004.10364 [quant-ph].
[57] J. Xu, S. Li, T. Chen, and Z.-Y. Xue, Nonadiabatic geometric quantum computation with optimal control on superconducting circuits (2020), arXiv:2004.10199 [quant-ph].
[58] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Simple pulses for elimination of leakage in weakly nonlinear qubits, Phys. Rev. Lett. **103**, 110501 (2009).
[59] D. J. Egger and F. K. Wilhelm, Adaptive hybrid optimal quantum control for imprecisely characterized systems, Phys. Rev. Lett. **112**, 240503 (2014).
[60] G. Feng, F. H. Cho, H. Katiyar, J. Li, D. Lu, J. Baugh, and R. Laflamme, Gradient-based closed-loop quantum optimal control in a solid-state two-qubit system, Phys. Rev. A **98**, 052341 (2018).
[61] J. Li, X. Yang, X. Peng, and C.-P. Sun, Hybrid quantum-classical approach to quantum optimal control, Phys. Rev. Lett. **118**, 150503 (2017).
[62] N. Wittler, F. Roy, K. Pack, M. Werninghaus, A. S. Roy, D. J. Egger, S. Filipp, F. K. Wilhelm, and S. Machnes, An integrated tool-set for control, calibration and characterization of quantum devices applied to superconducting qubits (2020), arXiv:2009.09866 [quant-ph].
[63] J. Allen, *Robust Optimal Control of the Cross-Resonance Gate in Superconducting Qubits*, Ph.D. thesis, University of Surrey (2019).
[64] A. R. Carvalho, H. Ball, M. J. Biercuk, M. R. Hush, and F. Thomsen, Error-robust quantum logic optimization using a cloud quantum computer interface (2020), arXiv:2010.08057 [quant-ph].
[65] R. L. Kosut, M. D. Grace, and C. Brif, Robust control of quantum gates via sequential convex programming, Phys. Rev. A **88**, 052326 (2013).
[66] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, Universal quantum control through deep reinforcement learning, npj Quantum Information **5**, 33 (2019).
[67] H. Ball, M. Biercuk, A. Carvalho, J. Chen, M. R. Hush, L. A. de Castro, L. Li, P. J. Liebermann, H. Slatyer, C. Edmunds, V. Frey, C. Hempel, and A. Milne, Software tools for quantum control: Improving quantum computer performance through noise and error suppression, Quantum Science and Technology 10.1088/2058-9565/abdca6 (2021).
[68] T. A. Howell, C. Fu, and Z. Manchester, Direct policy optimization using deterministic sampling and collocation (2020), arXiv:2010.08506 [cs.RO].
[69] A. Lee, Y. Duan, S. Patil, J. Schulman, Z. McCarthy, J. van den Berg, K. Goldberg, and P. Abbeel, Sigma hulls for gaussian belief space planning for imprecise articulated robots amid obstacles, in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2013) pp. 5660–5667.
[70] S. Thangavel, R. Paulen, and S. Engell, Robust multistage nonlinear model predictive control using sigma points, Processes **8**, 851 (2020).
[71] S. J. Julier and J. K. Uhlmann, Unscented filtering and nonlinear estimation, Proceedings of the IEEE **92**, 401

(2004).
[72] J. K. Uhlmann, *Dynamic map building and localization: New theoretical foundations*, Ph.D. thesis, University of Oxford Oxford (1995).
[73] H. Zhang, S. Chakram, T. Roy, N. Earnest, Y. Lu, Z. Huang, D. K. Weiss, J. Koch, and D. I. Schuster, Universal fast-flux control of a coherent, low-frequency qubit, Phys. Rev. X **11**, 011010 (2021).
[74] T. A. Howell, B. E. Jackson, and Z. Manchester, Altro: A fast solver for constrained trajectory optimization, in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2019) pp. 7674–7679.
[75] S. Machnes, E. Assémat, D. Tannor, and F. K. Wilhelm, Tunable, flexible, and efficient optimization of control pulses for practical qubits, Phys. Rev. Lett. **120**, 150401 (2018).
[76] M. H. Goerz, D. Basilewitsch, F. Gago-Encinas, M. G. Krauss, K. P. Horn, D. M. Reich, and C. P. Koch, Krotov: A python implementation of krotov's method for quantum optimal control, SciPost physics **7**, 10.21468/SciPostPhys.7.6.080 (2019).
[77] J. Schulman, J. Ho, A. X. Lee, I. Awwal, H. Bradlow, and P. Abbeel, Finding locally optimal, collision-free trajectories with sequential convex optimization., in *Robotics: science and systems*, Vol. 9 (Citeseer, 2013) pp. 1–10.
[78] R. Tedrake and the Drake Development Team, Drake: A planning, control, and analysis toolbox for nonlinear dynamical systems (2016).
[79] A. Hereid and A. D. Ames, Frost: Fast robot optimization and simulation toolkit, in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE/RSJ, Vancouver, BC, Canada, 2017).
[80] L. Jørgensen, D. L. Cardozo, and E. Thibierge, *Numerical Resolution Of The Schrödinger Equation*, Tech. Rep. (École Normale Supérieure de Lyon, 2011).
[81] N. Auer, L. Einkemmer, P. Kandolf, and A. Ostermann, Magnus integrators on multicore cpus and gpus, Computer Physics Communications **228**, 115 (2018).
[82] H. Berland and B. Skaflestad, Solving the nonlinear Schrödinger equation using exponential integrators, Modeling, Identification and Control **27**, 201 (2006).
[83] L. Einkemmer, M. Tokman, and J. Loffeld, On the performance of exponential integrators for problems in magnetohydrodynamics, Journal of Computational Physics **330**, 550 (2017).
[84] R. Shillito, J. A. Gross, A. D. Paolo, Élie Genois, and A. Blais, Fast and differentiable simulation of driven quantum systems (2020), arXiv:2012.09282 [quant-ph].
[85] C. R. Hargraves and S. W. Paris, Direct trajectory optimization using nonlinear programming and collocation, J. Guidance **10**, 338 (1987).
[86] M. Kelly, An introduction to trajectory optimization: How to do your own direct collocation, SIAM Review **59**, 849 (2017).
[87] J. T. Betts, Survey of numerical methods for trajectory optimization, Journal of guidance, control, and dynamics **21**, 193 (1998).
[88] P. E. Gill, W. Murray, and M. A. Saunders, Snopt: An sqp algorithm for large-scale constrained optimization, SIAM review **47**, 99 (2005).
[89] A. Wächter and L. T. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, Mathematical programming **106**, 25 (2006).
[90] W. Li and E. Todorov, Iterative Linear Quadratic Regulator Design for Nonlinear Biological Movement Systems, in *Proceedings of the 1st International Conference on Informatics in Control, Automation and Robotics* (Setubal, Portugal, 2004).
[91] G. Lantoine and R. P. Russell, A hybrid differential dynamic programming algorithm for constrained optimal control problems. part 1: Theory, Journal of Optimization Theory and Applications **154**, 382 (2012).
[92] B. Plancher, Z. Manchester, and S. Kuindersma, Constrained unscented dynamic programming, in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2017) pp. 5674–5680.
[93] J. Nocedal and S. Wright, *Numerical optimization* (Springer Science & Business Media, 2006).
[94] D. P. Bertsekas, Projected newton methods for optimization problems with simple constraints, SIAM Journal on control and Optimization **20**, 221 (1982).
[95] C. V. Rao, S. J. Wright, and J. B. Rawlings, Application of interior-point methods to model predictive control, Journal of Optimization Theory and Applications **99**, 723 (1998).
[96] L. Zhang, W. Zhou, and D. Li, Global convergence of a modified fletcher–reeves conjugate gradient method with armijo-type line search, Numerische Mathematik **104**, 561 (2006).
[97] D. Bertsekas, Constrained optimization and lagrange multiplier methods (second edn.) athena scientific: Belmont (1996).
[98] B. E. Jackson, T. Punnoose, D. Neamati, K. Tracy, R. Jitosho, and Z. Manchester, Altro-c: A fast solver for conic model-predictive control, in *International Conference on Robotics and Automation ICRA* (2021) in Review.
[99] N. Earnest, S. Chakram, Y. Lu, N. Irons, R. K. Naik, N. Leung, L. Ocola, D. A. Czaplewski, B. Baker, J. Lawrence, J. Koch, and D. I. Schuster, Realization of a Λ system with metastable states of a capacitively shunted fluxonium, Phys. Rev. Lett. **120**, 150504 (2018).
[100] Y.-H. Lin, L. B. Nguyen, N. Grabon, J. San Miguel, N. Pankratova, and V. E. Manucharyan, Demonstration of protection of a superconducting qubit from energy decay, Phys. Rev. Lett. **120**, 150503 (2018).
[101] V. E. Manucharyan, J. Koch, L. I. Glazman, and M. H. Devoret, Fluxonium: Single cooper-pair circuit free of charge offsets, Science **326**, 113 (2009).
[102] L. B. Nguyen, Y.-H. Lin, A. Somoroff, R. Mencia, N. Grabon, and V. E. Manucharyan, High-coherence fluxonium qubit, Phys. Rev. X **9**, 041041 (2019).
[103] 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).
[104] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer's guide to superconducting qubits, Applied Physics Re-

views **6**, 021318 (2019).

[105] T. Schulte-HerbrÃŒggen, A. SpÃ¶rl, N. Khaneja, and S. J. Glaser, Optimal control for generating quantum gates in open dissipative systems, Journal of Physics B: Atomic, Molecular and Optical Physics **44**, 154013 (2011).

[106] M. Abdelhafez, D. I. Schuster, and J. Koch, Gradient-based optimal control of open quantum systems using quantum trajectories and automatic differentiation, Phys. Rev. A **99**, 052327 (2019).

[107] Z. Huang, P. S. Mundada, A. Gyenis, D. I. Schuster, A. A. Houck, and J. Koch, Engineering dynamical sweet spots to protect qubits from $1/f$ noise (2020), arXiv:2004.12458 [quant-ph].

[108] P. S. Mundada, A. Gyenis, Z. Huang, J. Koch, and A. A. Houck, Floquet-engineered enhancement of coherence times in a driven fluxonium qubit (2020), arXiv:2007.13756 [quant-ph].

[109] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, SIAM Journal on Computing **38**, 1207 (2008).

[110] E. Knill, Quantum computing with realistically noisy devices, Nature **434**, 39–44 (2005).

[111] D. Gottesman, Stabilizer codes and quantum error correction (1997), arXiv:quant-ph/9705052 [quant-ph].

[112] A. Paetznick, *Resource optimization for fault-tolerant quantum computing*, Ph.D. thesis, University of Waterloo (2014).

[113] M. Suchara, A. Faruque, C.-Y. Lai, G. Paz, F. T. Chong, and J. Kubiatowicz, Comparing the overhead of topological and concatenated quantum error correction (2013), arXiv:1312.2316 [quant-ph].

[114] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Randomized benchmarking and process tomography for gate errors in a solid-state qubit, Phys. Rev. Lett. **102**, 090502 (2009).

[115] R. C. Bialczak, R. McDermott, M. Ansmann, M. Hofheinz, N. Katz, E. Lucero, M. Neeley, A. D. O'Connell, H. Wang, A. N. Cleland, and J. M. Martinis, $1/f$ flux noise in josephson phase qubits, Phys. Rev. Lett. **99**, 187006 (2007).

[116] K. Kakuyanagi, T. Meno, S. Saito, H. Nakano, K. Semba, H. Takayanagi, F. Deppe, and A. Shnirman, Dephasing of a superconducting flux qubit, Phys. Rev. Lett. **98**, 047004 (2007).

[117] P. Kumar, S. Sendelbach, M. A. Beck, J. W. Freeland, Z. Wang, H. Wang, C. C. Yu, R. Q. Wu, D. P. Pappas, and R. McDermott, Origin and reduction of $1/f$ magnetic flux noise in superconducting devices, Phys. Rev. Applied **6**, 041001 (2016).

[118] F. Yoshihara, K. Harrabi, A. O. Niskanen, Y. Nakamura, and J. S. Tsai, Decoherence of flux qubits due to $1/f$ flux noise, Phys. Rev. Lett. **97**, 167001 (2006).

[119] R. H. Koch, D. P. DiVincenzo, and J. Clarke, Model for $1/f$ flux noise in squids and qubits, Phys. Rev. Lett. **98**, 267003 (2007).

[120] F. Yoshihara, Y. Nakamura, and J. S. Tsai, Correlated flux noise and decoherence in two inductively coupled flux qubits, Phys. Rev. B **81**, 132502 (2010).

[121] E. L. Hahn and D. E. Maxwell, Spin echo measurements of nuclear spin coupling in molecules, Phys. Rev. **88**, 1070 (1952).

[122] S. Meiboom and D. Gill, Modified spin-echo method for measuring nuclear relaxation times, Review of scientific instruments **29**, 688 (1958).

[123] J. O. Smith, *Spectral Audio Signal Processing* (2020) online book, 2011 edition.

[124] C. Rackauckas and Q. Nie, Differentialequations.jl–a performant and feature-rich ecosystem for solving differential equations in julia, Journal of Open Research Software **5** (2017).

[125] M. Besançon, D. Anthoff, A. Arslan, S. Byrne, D. Lin, T. Papamarkou, and J. Pearson, Distributions.jl: Definition and modeling of probability distributions in the juliastats ecosystem (2019), arXiv:1907.08611 [stat.CO].

[126] J. Revels, M. Lubin, and T. Papamarkou, Forward-mode automatic differentiation in Julia, arXiv:1607.07892 [cs.MS] (2016).

[127] J. D. Hunter, Matplotlib: A 2d graphics environment, Computing in Science & Engineering **9**, 90 (2007).

[128] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del R'ıo, M. Wiebe, P. Peterson, P. G'erard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, Array programming with NumPy, Nature **585**, 357 (2020).

[129] M. Innes, Don't unroll adjoint: Differentiating ssa-form programs (2018), arXiv:1810.07951 [cs.PL].

[130] D. Manzano, A short introduction to the lindblad master equation, AIP Advances **10**, 025106 (2020).

[131] G. T. Landi, Lecture notes on quantum information and quantum noise (2018).

[132] S. J. Julier and J. K. Uhlmann, Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations, in *Proceedings of the 2002 American Control Conference (IEEE Cat. No. CH37301)*, Vol. 2 (IEEE, 2002) pp. 887–892.