Universal Control in Bosonic Systems with Weak Kerr Nonlinearities

Ming Yuan,[1, *] Alireza Seif,[1] Andrew Lingenfelter,[2] David I. Schuster,[1, 2, 3] Aashish A. Clerk,[1] and Liang Jiang[1, †]

[1]*Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL 60637, USA*
[2]*Department of Physics, The University of Chicago, Chicago, IL 60637, USA*
[3]*James Franck Institute, The University of Chicago, Chicago, IL 60637, USA*
(Dated: December 27, 2023)

Resonators with weak single-photon self-Kerr nonlinearities can theoretically be used to prepare Fock states in the presence of a loss much larger than their nonlinearities. Two necessary ingredients are large displacements and a two-photon (parametric) drive. Here, we find that these systems can be controlled to achieve any desired gate operation in a finite dimensional subspace (whose dimensionality can be chosen at will). Moreover, we show that the two-photon driving requirement can be relaxed and that full controllability is achievable with only 1-photon (linear) drives. We make use of both Trotter-Suzuki decompositions and gradient-based optimization to find control pulses for a desired gate, which reduces the computational overhead by using a small blockaded subspace. We also discuss the infidelity arising from input power limitations in realistic settings, as well as from corrections to the rotating-wave approximation. Our universal control protocol opens the possibility for quantum information processing using a wide range of lossy systems with weak nonlinearities.

Introduction– Bosonic systems, such as photons in optical or microwave resonators, are a promising platform for quantum information processing. In contrast to qubits, the infinite-dimensional bosonic Hilbert space provides novel ways to encode and robustly process quantum information in a hardware-efficient manner [1]. A challenge however is the need for nonlinear operations [2]. It has been shown that, to achieve universal control in bosonic systems it is necessary and sufficient to have at least one kind of nonlinear operation in addition to linear operations, i.e., unitary evolution under a Hamiltonian linear or quadratic in bosonic raising and lowering operators [3, 4]. One approach to introduce nonlinearity is to couple the bosonic system directly to a nonlinear system such as a qubit [5–8], but for many platforms, achieving this with sufficiently strong coupling can be difficult. For example, in millimeter wave regime (around 100 GHz) there is no good superconducting qubit yet; other bosonic systems like phonons can work in a higher temperature beyond the requirement for superconducting qubits [9, 10].

Given this, it would be ideal to exploit intrinsic nonlinearities in optical or microwave resonators. An extremely common example is a self-Kerr nonlinearity. Examples include micro-ring resonators or photonic crystals with $\chi^{(3)}$ nonlinearities [11, 12] and quantum LC circuits that contain superconducting materials with high kinetic inductance [13, 14]. While in principle these nonlinearities are sufficient for universal control, in practice they are often much weaker than photon loss rates, precluding the ability to achieve high-fidelity nonlinear operations.

While nonlinearities weaker than loss might seem to be of no use for quantum nonlinear operations, recent research suggests that this might not be the case [15]. In this work, the authors propose an intriguing scheme to deterministically prepare a single-photon Fock state
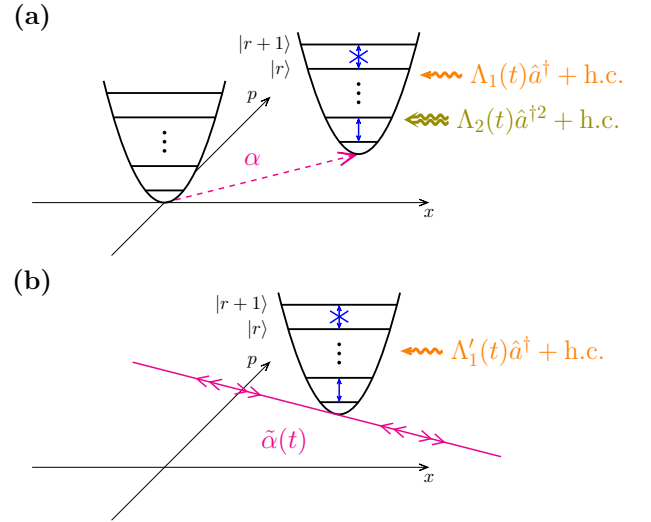

FIG. 1. Schematic diagram of the photon blockade in the displaced rotating frame with (a) both 1-photon and 2-photon drives, (b) only 1-photon drive with fast oscillation of the displaced rotating frame in phase space.

using a Kerr nonlinearity, even in cases where this nonlinearity is significantly smaller than the loss rates. This scheme relies on a novel photon blockade phenomenon that requires the displacement of a single bosonic mode together with carefully chosen 1-photon and 2-photon driving amplitudes. Notably, the speed of the operations in the blockade subspace is enhanced by the displacement amplitude, which can be large enough to counteract the effects of loss. A similar idea related to displacement-boosted gates is also presented in Ref. [7].

Here, we generalize this blockade scheme to demonstrate its applications beyond the preparation of single-photon Fock states. We show that using this scheme, one can perform any unitary operation in a blockaded sub-

space of Fock states with an arbitrarily chosen dimension. In addition to formal proof in the ideal scenario, we present a gradient-based optimization algorithm to explicitly find the control pulse sequences to implement a desired unitary operation in the blockaded subspace.

In practice, directly implementing the required 2-photon drive can be a challenging task, as this can require additional weak nonlinearities. For example, in certain platforms one could pump an auxiliary mode that interacts nonlinearly with the central mode [16]. Here we show that one can eliminate the need for an explicit 2-photon drive, by instead simply time-modulating the amplitude of the single-photon drive. This is similar in spirit to the operation of double-pumped parametric amplifiers [17].

The optimization algorithm and this modulation scheme are then integrated seamlessly to implement arbitrary operations using only a standard linear, single-photon drive. This simplification comes at the cost of requiring larger 1-photon driving amplitudes, which may result in much stronger input power or even the violation of the rotating wave approximation (RWA). We discuss these implications and possible methods to resolve the resulting imperfections. We argue that even with these limitations, our scheme performs well in experimental platforms where the strength of the single-photon self-Kerr nonlinearity is comparable to or slightly smaller than the loss rate, which was not possible using previous techniques.

System setup– We consider a setup similar to Ref. [15], consisting of a single-mode resonator with self-Kerr nonlinearity, subject to both 1-photon and 2-photon drives. Within the RWA, the Hamiltonian of the system can be written as

$$\hat{H} = \frac{\chi}{2}\hat{a}^{\dagger 2}\hat{a}^2 + \omega_c\hat{a}^{\dagger}\hat{a} + [\Lambda_1(t)e^{-i\omega_1(t)t}\hat{a}^{\dagger} \\ + \Lambda_2(t)e^{-i\omega_2(t)t}\hat{a}^{\dagger 2} + \text{h.c.}], \quad (1)$$

where χ indicates the strength of the single-photon Kerr nonlinearity, ω_c is the angular frequency of the resonator, and $\Lambda_{1(2)}$ and $\omega_{1(2)}$ are the amplitude and the frequency of the 1(2)-photon drives, respectively. Here, h.c. denotes the Hermitian conjugate of the terms in the bracket. As shown in Ref. [15], by choosing $\omega_2 = 2\omega_1$, going to a frame rotating with $\omega_1(t)$ and then displaced by $\alpha(t)$, we can obtain a blockaded Hamiltonian of the form

$$\hat{H}_{\text{dr}}[\alpha(t)] = \frac{\chi}{2}\hat{a}^{\dagger 2}\hat{a}^2 + \Delta_0\hat{a}^{\dagger}\hat{a} + [\chi\alpha(t)\hat{a}^{\dagger}(\hat{n}-r) + \text{h.c.}], \quad (2)$$

where r is an adjustable positive integer determining the block subspace's dimension, and Δ_0 is the detuning term in the displaced rotating frame. To achieve this effective Hamiltonian for a given $\alpha(t)$, Δ_0 and integer r, one needs time-dependent drive amplitudes and frequencies chosen

such that

$$\begin{cases} \Lambda_1(t) = \chi\alpha(t)[2|\alpha(t)|^2 - r] - \Delta_0\alpha(t) + i\dot{\alpha}(t), \\ \Lambda_2(t) = -\frac{\chi}{2}\alpha^2(t), \\ \omega_1(t) = \omega_c - \Delta_0 + \frac{1}{t}\int_0^t 2\chi|\alpha(t')|^2\,\text{d}t'. \end{cases} \quad (3)$$

The key term in Eq. (2) is the nonlinear drive, which has no coupling between $|r\rangle$ and $|r+1\rangle$ levels. Therefore, the dynamics generated by $\hat{H}_{\text{dr}}$ is constrained within the blockade subspace $\mathcal{H}_b$ spanned by $\{|0\rangle, |1\rangle, \ldots, |r\rangle\}$. Also, the amplitude of this nonlinear drive is set by the time-dependent displacement $\alpha(t)$, which serves as a key control parameter in what follows. Previously, it was shown that with a static α, one could use $\hat{H}_{\text{dr}}$ to generate Fock $|1\rangle$ state on a timescale much shorter than $1/\chi$ [15]. Here, we go much further: we show that in fact, by using a time-dependent $\alpha(t)$, one can generate any unitary within the blockade subspace $\mathcal{H}_b$, whose dimension can also be chosen freely.

To account for the effects of photon loss in our system, we use the master equation

$$\frac{\text{d}\hat{\rho}}{\text{d}t} = -i[\hat{H}, \hat{\rho}] + \kappa\mathcal{D}[\hat{a}]\hat{\rho}, \quad (4)$$

where $\mathcal{D}[\hat{a}]\hat{\rho} = \hat{a}\hat{\rho}\hat{a}^{\dagger} - \frac{1}{2}\{\hat{a}^{\dagger}\hat{a}, \hat{\rho}\}$ is a dissipator that models the photon loss effect and κ is the rate of this process. Note that when transforming to the same displaced rotating frame mentioned earlier, the 1-photon driving amplitude $\Lambda_1(t)$ in Eq. (3) now requires the addition of an extra term, namely $i\kappa\alpha(t)/2$, in order to obtain the same Hamiltonian $\hat{H}_{\text{dr}}[\alpha(t)]$. However, the form of the dissipator $\mathcal{D}[\hat{a}]$ and the associated loss rate κ remain unaltered in this new frame. Consequently, due to the enhancement of the nonlinear blockade drive by α in $\hat{H}_{\text{dr}}$, the operations can be performed on a significantly shorter timescale than $1/\chi$. This, in turn, allows for the mitigation of the impact of photon loss, presenting an opportunity to achieve high-fidelity gates, even when $\chi \ll \kappa$.

Demonstration of universality– We sketch the proof of universal controllability of our system governed by the Hamiltonian $\hat{H}_{\text{dr}}$ shown in Eq. (2) here and refer the reader to Ref. [18] for more details. We first focus on the dynamics within the blockade subspace $\mathcal{H}_b$. Let $\hat{\Pi}_r := \sum_{n=0}^r |n\rangle\langle n|$ denote the projector to this $N = r+1$ dimensional subspace. The projection of $\hat{H}_{\text{dr}}$ to the blockade subspace is given by

$$\hat{\Pi}_r\hat{H}_{\text{dr}}\hat{\Pi}_r = \hat{H}_{d,0} + \chi\,\text{Re}[\alpha(t)]\hat{H}_{c,R} + \chi\,\text{Im}[\alpha(t)]H_{c,I}, \quad (5)$$

where

$$\begin{cases} \hat{H}_{d,0} = \sum_{n=0}^r \left[\chi(n^2-n)/2 + \Delta_0 n\right]|n\rangle\langle n|, \\ \hat{H}_{c,R} = \sum_{n=0}^{r-1}(n-r)\sqrt{n+1}(|n+1\rangle\langle n| + \text{h.c.}), \\ \hat{H}_{c,I} = i\sum_{n=0}^{r-1}(n-r)\sqrt{n+1}(|n+1\rangle\langle n| - \text{h.c.}). \end{cases} \quad (6)$$

In what follows (using the language of quantum control theory, see e.g., [19]) $\hat{H}_{d,0}$ serves as the drift Hamiltonian, while $\hat{H}_{c,R}$ and $\hat{H}_{c,I}$ are the control Hamiltonians. The real and imaginary parts of $\alpha(t)$ are time-dependent functions that can be controlled. Following Ref. [5] we define universal control of a quantum system as the ability to realize any unitary operation $\hat{U}_{\text{tar}}$ in the U(N) group via a properly chosen control ($\alpha(t)$ in this case) and evolution time T.

A theorem in Ref. [20] suggests a sufficient condition for the Hamiltonian to make the system universally controllable. It has two requirements. First, the drift Hamiltonian $\hat{H}_{d,0}$ should be diagonal (with eigenvalues denoted as E_k for eigenstates $|k\rangle$) and contain certain type of nonlinearity, specifically, the nearest energy difference $\mu_k := E_k - E_{k+1}$ should satisfy $\mu_0 \neq 0$ and $\mu_k^2 \neq \mu_0^2$ for $k > 0$ (or similarly $\mu_{N-2} \neq 0$ and $\mu_k^2 \neq \mu_{N-2}^2$ for $k < N - 2$). Second, one of the control parts $\hat{H}_{c,j}$ should only have couplings between $|k\rangle$ and $|k+1\rangle$ for all $0 \leq k < N - 1$. If both these conditions are satisfied, then the generated dynamical Lie group will be U(N) when $\text{Tr}\left[\hat{H}_d\right] \neq 0$ and SU(N) otherwise.

We can easily verify that those requirements for U(N) group (where $N = r + 1$ in our case) generation are satisfied with our $\hat{H}_{d,0}$ and $\hat{H}_{c,R}$ as long as $r \neq -\frac{2\Delta_0}{\chi} + 1$. This allows us to fix $\Delta_0 = 0$ for $r \geq 2$ in the rest of the main text. Moreover, the two control degrees of freedom $\hat{H}_{c,R}$ and $\hat{H}_{c,I}$ provide the possibility to do any gate (up to a global phase) in an arbitrarily fast manner, as these two are also sufficient to generate full SU(N) group [18]. Finally, as r is also adjustable, we can choose any blockade dimension we want. Therefore, unitaries defined in any finite dimension are in principle achievable.

Optimal control– The generalized blockade phenomena allow one, in principle, to perform an arbitrary unitary operation in an arbitrarily chosen N-dimensional blockaded subspace in a time much faster than $1/\chi$. The question that we now address is how to design a particular control $\alpha(t)$ to realize a target unitary. In contrast to the conventional setting, where one optimizes the control in the rotating frame of the drive [21], we consider $\hat{H}_{\text{dr}}$ defined in the instantaneous displaced rotating frame and optimize $\alpha(t)$, the frame parameter. Consequently, finding an optimal $\alpha(t)$ directly determines the corresponding physical parameters $\Lambda_{1(2)}(t)$ and $\omega_1(t)$ via Eq. (3) required for implementing the desired operations in the laboratory. In other words, by hard coding the blockade condition in the evolution through Eq. (3), we simplify the optimization task to finding $\alpha(t)$ in the small blockade Hilbert space. In this way, even though the required drive amplitudes $\Lambda_{1(2)}(t)$ can be very large, we do not need to consider large photon number states in the optimization. In this section, we only optimize the controls for a fixed evolution time T in the absence of loss. We further consider the effects of loss and a realistic experi-

mental constraint on total input power (see [18]); as we show, these additional features lead to there being an optimal choice of gate time T.

Typically, optimal control algorithms such as GRAPE [22, 23] discretize the control pulse $\alpha(t)$, and maximize a figure of merit such as the fidelity by performing gradient-based optimization on the control parameters, i.e., the amplitude of the control at discrete time points. This standard approach would yield a piece-wise constant $\alpha(t)$, something that is highly problematic for our setup: discontinuous jumps in $\alpha(t)$ would require infinite driving power to implement, as the 1-photon driving amplitude has a term proportional to $i\dot{\alpha}(t)$ (c.f. Eq. (3)).

To solve this issue, in contrast to conventional methods, we expand $\alpha(t)$ on a basis of continuous functions. Since it is desirable to be in a non-displaced frame, i.e. $\alpha(0) = \alpha(T) = 0$ in the beginning ($t = 0$) and at the end ($t = T$) of the protocol, we use the following sine-basis ansatz for the control pulse

$$\alpha(t) = \sum_{k=1}^{k_{\max}} \alpha_k \sin\left(\frac{k\pi t}{T}\right). \tag{7}$$

Here, $k_{\max}$, the cutoff number for the highest harmonics that we use, is a hyperparameter that is chosen according to the complexity of the task, and α_k are complex-valued optimization variables. In practice, we choose $k_{\max}$ heuristically in the optimization procedure. If the fidelity achieved from the optimal pulses is lower than our target, we increase $k_{\max}$ for better performance. To implement a target unitary operation in N dimensions, we maximize $F_{\text{u}}(\{\alpha_k\}) = \left|\text{Tr}\left[\hat{U}_{\text{tar}}^\dagger \hat{U}(T)\right]\right|^2 / N^2$, where $\hat{U}(T)$ is obtained by solving $\frac{\text{d}}{\text{d}t}\hat{U}(t) = -i\hat{H}_{\text{dr}}[\alpha(t)]\hat{U}(t)$ for $t = T$ with the initial condition $\hat{U}(0) = \hat{I}$. When optimizing F_{u}, we implicitly ignore the irrelevant global phase. Note that the dependence of the objective on $\{\alpha_k\}$ originates from the dependence of $\hat{H}_{\text{dr}}$ on $\alpha(t)$ according to Eq. (2). Moreover, while obtaining F_{u} involves solving an ordinary differential equation, it is still differentiable and its gradient with respect to $\{\alpha_k\}$ can be calculated using the chain rule and the adjoint sensitivity method [24]. Therefore, we can use gradient-based optimization to find locally optimal $\{\alpha_k\}$.

To illustrate the universal controllability of our scheme, we consider the problem of implementing the permutation $\hat{U}_{\text{P}}$ or the Fourier transformation $\hat{U}_{\text{FT}}$ in a 3-level blockade subspace up to a global phase spanned by $\{|m\rangle\}_{m=0}^2$, where

$$\hat{U}_{\text{P}} = |2\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 2|, \tag{8a}$$

$$\hat{U}_{\text{FT}} = \frac{1}{\sqrt{3}} \sum_{m,n} e^{i\frac{2\pi mn}{3}} |m\rangle\langle n|. \tag{8b}$$
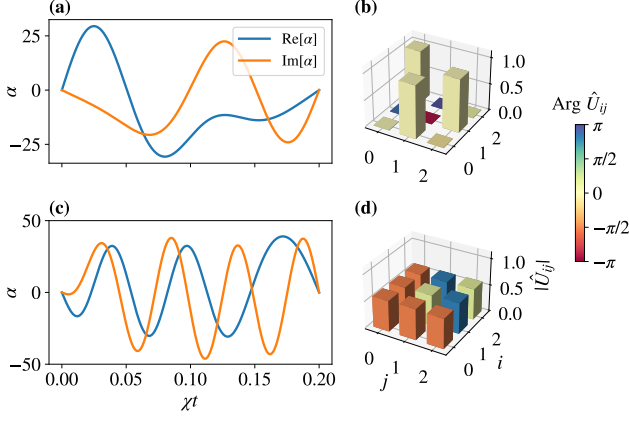
FIG. 2. Optimal control for implementing arbitrary unitary operations. (a) The optimized $\alpha(t)$ function that implements permutation (8a) in a 3-level blockade subspace with $k_{\max} = 5$ harmonics. Total evolution time $T = 0.2/\chi$. (b) The elements of the unitary $\hat{U}(T)$ generated by $\alpha(t)$ shown in (a) under the Fock basis, i.e., $\hat{U}_{ij} = \langle i | \hat{U}(T) | j \rangle$. The height and color of each bar show the absolute value and argument of the corresponding matrix elements. (c) Optimal $\alpha(t)$ for implementing Fourier transform (8b) in the same blockade subspace with $k_{\max} = 8$ harmonics and $T = 0.2/\chi$. (d) the matrix elements of $\hat{U}(T)$ generated by $\alpha(t)$ shown in (c).

We use the automatic differentiation toolbox of JAX [25], a numerical computing package, to perform the gradient-based optimization. We choose the evolution time $T = 0.2/\chi$. We also fix $\alpha(0) = \alpha(T) = 0$ and $\Delta_0 = 0$ in $\hat{H}_{\mathrm{dr}}$ and find the pulses $\alpha(t)$ that implements the two unitary operations of interest. In both cases, the algorithm finds a solution such that $\left| \mathrm{Tr}\left[\hat{U}_{\mathrm{tar}}^\dagger \hat{U}(T) \right] \right|^2 / N^2 > 1 - 10^{-4}$ (see Fig. 2).

Only 1-photon drive– So far, we have focused on the fundamental questions about controllability by explicitly constructing the 1- and 2-photon drives to achieve fast universal control for weak-Kerr systems in the presence of loss. Here, we move away from the ideal scenario and discuss issues relevant to experimental implementations. One question to address is whether one truly needs a distinct 2-photon drive. While this can be done in some platforms, e.g., superconducting qubits with flux-pumping [4], it can be challenging in other platforms. Here, we present a method that allows our control scheme to be implemented without any explicit independent 2-photon drive. As we show, this idea is intrinsically connected to squeezing by double-pumping a Kerr resonator [17].

In the absence of an independent 2-photon drive, $\Lambda_2(t) = 0$ in Eq. (1). Again, we set $\Delta_0 = 0$ and choose the 1-photon driving amplitude such that in a frame rotating with $\omega_1(t)$ and displaced by $\tilde{\alpha}(t)$, we have the same

blockade drive. In this frame, the Hamiltonian is

$$\hat{H}'_{\mathrm{dr}}[\tilde{\alpha}(t)] = \frac{\chi}{2}\hat{a}^{\dagger 2}\hat{a}^2 + [\chi\tilde{\alpha}(t)\hat{a}^\dagger(\hat{n} - r) + \mathrm{h.c.}] + \left[\frac{\chi}{2}\tilde{\alpha}^2(t)\hat{a}^{\dagger 2} + \mathrm{h.c.}\right], \tag{9}$$

where the 1-photon driving amplitude is chosen to have the form

$$\Lambda'_1(t) = \chi\tilde{\alpha}(t)[|\tilde{\alpha}(t)|^2 - r] + i\dot{\tilde{\alpha}}(t) + i\kappa\tilde{\alpha}/2. \tag{10}$$

While the nonlinear single-photon drive in Eq. (9) has the correct form, the induced two-photon drive term in the last line violates the desired blockade condition.

A key observation in our strategy to revive the blockade condition is that the desired fully blockaded Hamiltonian can be written as $\hat{H}_{\mathrm{dr}}[\alpha(t)] = \frac{1}{\sqrt{2}}\{\hat{H}'_{\mathrm{dr}}[e^{-\frac{i\pi}{4}}\alpha(t)] + \hat{H}'_{\mathrm{dr}}[e^{\frac{i\pi}{4}}\alpha(t)]\}$. Intuitively, by alternating the phase of $\tilde{\alpha}(t)$, we can average away and cancel the unwanted 2-photon drive term while retaining the desired nonlinear blockade drive. This observation combined with the Trotter formula

$$e^{-i\frac{\hat{H}_1 + \hat{H}_2}{2}\delta T} = e^{-i\hat{H}_1\delta T/4}e^{-i\hat{H}_2\delta T/2}e^{-i\hat{H}_1\delta T/4} + O[(\delta T)^3] \tag{11}$$

suggests that setting $\hat{H}_1(t) = \hat{H}'_{\mathrm{dr}}[e^{-\frac{i\pi}{4}}\alpha(t)]$ and $\hat{H}_2(t) = \hat{H}'_{\mathrm{dr}}[e^{\frac{i\pi}{4}}\alpha(t)]$ and alternating the evolution for $\delta T/2$ between the two Hamiltonians suppress the errors in violating the blockade condition to $O[(\delta T)^3]$ [26]. However, this Trotter scheme requires discretizing $\alpha(t)$ into intervals of length δT and implementing instantaneous displacements between $e^{\pm\frac{i\pi}{4}}\alpha$, which in practice introduces additional complexities.

Inspired by the discrete version of the Trotter formula, we design its continuous counterpart via Magnus expansion [18]

$$\exp\left[-i\int_0^{\delta T}\hat{H}(t)\,\mathrm{d}t\right] = \mathcal{T}\exp\left[-i\int_0^{\delta T}\hat{H}(t)\,\mathrm{d}t\right] + O[(\delta T)^3], \tag{12}$$

where $\mathcal{T}$ is the time-ordering operator and $\hat{H}(t)$ for $t \in [0, \delta T]$ is chosen to be symmetric around $\delta T/2$, i.e., $\hat{H}(t) = \hat{H}(\delta T - t)$. As a result, we construct a new function $\tilde{\alpha}(t)$ that oscillates rapidly with time. Specifically, Eq. (12) gives us a recipe for finding $\tilde{\alpha}(t)$ such that the coarse-grained evolution under $\hat{H}'_{\mathrm{dr}}[\tilde{\alpha}(t)]$ over an interval of δT is close to that under $\hat{H}_{\mathrm{dr}}[\alpha(t)]$. This translates to having the average of $\tilde{\alpha}(t)$ over a δT time interval centered on time t to satisfy

$$\overline{\tilde{\alpha}^2(t)} = 0, \quad \overline{\tilde{\alpha}(t)} = \alpha(t), \tag{13}$$

where the overline denotes the coarse-graining time average, and $\alpha(t)$ is the optimal choice of function in $\hat{H}_{\mathrm{dr}}$

defined in Eq. (2) that generates the desired target unitary operation. To satisfy these constraints, we propose using the ansatz $\tilde{\alpha}(t) = \alpha(t)f(t)$, where $f(t)$ is a periodic function with period $\delta T = \frac{T}{M}$. Here T is the total evolution time and M is the number of periods during the evolution. We also denote $\omega_r := \frac{2\pi M}{T}$ for further use. Besides, to keep the same structure as the Trotter formula in Eq. (12), we further require that

$$f(t) = f(\delta T - t). \tag{14}$$

Consequently, the constraints for $\tilde{\alpha}(t)$ in Eq. (13) translate to constraints

$$\overline{f(t)} = 1, \quad \overline{f^2(t)} = 0 \tag{15}$$

on $f(t)$ over over each period. This ensures that the overall evolution under $H'_{\mathrm{dr}}[\tilde{\alpha}(t)]$ and $\hat{H}_{\mathrm{dr}}[\alpha(t)]$ for time T closely resemble each other, that is

$$
\begin{aligned}
&\mathcal{T}\exp\left\{-i\int_0^T \hat{H}'_{\mathrm{dr}}[\tilde{\alpha}(t)]\,\mathrm{d}t\right\} \\
&\approx \mathcal{T}\exp\left\{-i\int_0^T \hat{H}_{\mathrm{dr}}[\alpha(t)]\,\mathrm{d}t\right\}.
\end{aligned} \tag{16}
$$

One possible choice of $f(t)$ is the following

$$f_{\mathrm{dp}}(t) = 1 + i\sqrt{2}\cos(\omega_r t). \tag{17}$$

An example of the shape of $\tilde{\alpha}(t)$ after implementing this modulation is shown in Fig. 3 (and schematic illustration in Fig. 1(b) as well). To obtain physical intuition on the underlying mechanism, we can split $f_{\mathrm{dp}}(t)$ into two parts. The constant part plays the role of Λ_1 in Eq. (1), and provides the desired nonlinear blockade drive in the displaced rotating frame. In contrast, the time-dependent term, $i\sqrt{2}\cos(\omega_r t)$, corresponds to the double-pumping scheme with driving frequencies $\omega_1 \pm \omega_r$, as it contributes to $\Lambda'_1(t)$ via $i\dot{\tilde{\alpha}}(t)$. This double-pumping will effectively generate the 2-photon driving Λ_2 in Eq. (1) from the Kerr interaction, as a result of 4-wave mixing [17].

The approximation in Eq. (16) is valid if the number of periods M is large enough such that (i) the error from Trotter expansion (see Eq. (11,12)) is small, and (ii) within each period of $f(t)$, the control function $\alpha(t)$ varies slowly so that the time average of $\tilde{\alpha}(t)$ deviates little from $\alpha(t)$. In the infinite M limit the two sides of Eq. (16) are identical. The induced error due to the inaccuracy of Trotter approximation is analyzed in detail in Ref. [18].

Discussion– In the previous section, we have demonstrated that in principle a distinct 2-photon drive is unnecessary for universal control and explicitly provided the modified 1-photon pulse design for targeted operations. However, our operations may suffer from coherent errors that come from inaccurate Trotter approximation.
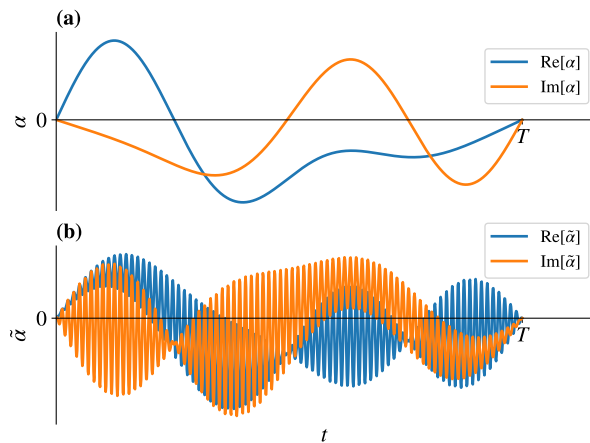

FIG. 3. Schematic comparison between the (a) original optimal pulse $\alpha(t)$ and (b) the modulated pulse $\tilde{\alpha}(t) = \alpha(t)f_{\mathrm{dp}}(t)$ using Eq. (17) needed in the absence of a 2-photon drive. In this example, we choose $M = 80$ periods for the modulation.

This, as well as the incoherent photon loss during gate execution, serve as two sources for gate infidelity.

If the 1-photon driving amplitude $|\Lambda_1|$ (or the input power P_{in}) is unlimited, both error sources can be sufficiently suppressed. We can increase the amplitude of $\alpha(t)$ so that the gate time T as well as the photon loss probability is reduced. We can also increase the displaced frame oscillating frequency ω_r to reduce Trotter errors. However, if P_{in} is limited, we have to make a trade-off to achieve optimized operation fidelity. In Ref. [18], we analyze the effect of these errors in detail for the case of preparing a single photon Fock state and show that the state preparation infidelity is given by

$$\epsilon_{\mathrm{tot}} \simeq C(\kappa_e + \kappa_i)T + \frac{C'}{P_{\mathrm{in}}^2 \kappa_e^2 \chi^{10} T^{14}}, \tag{18}$$

where $\kappa_{e(i)}$ stands for the external (internal) loss rate. Here the first term captures the effect of loss errors growing with time T, and the second term captures the Trotter errors that scale inverse polynomially with T. Therefore, there exist an optimal time and external loss rate for a given power and internal loss rate that minimizes the total error, which scales as $\epsilon_{\mathrm{tot}}^{\mathrm{opt}} \propto \kappa_i^{4/5}/(P_{\mathrm{in}}^{2/15}\chi^{2/3})$. For example, suppose we have a niobium nitride resonator [13] with $\omega_c = 2\pi \times 100$ GHz, $\chi = -1$ kHz and an improved $|\chi|/\kappa_i = 0.5$ in the future. Then, to prepare $|1\rangle$ state with 90% fidelity the required P_{in} is around 30 nW. The corresponding required $\alpha \approx 15$. These numbers suggest that our scheme may work well in the regime that self-Kerr is comparable to or slightly lower than the loss rate. However, if κ_i is increased by a factor of 10 we need an increase of P_{in} by a factor of 10^6 to keep the same $\epsilon_{\mathrm{tot}}^{\mathrm{opt}}$, which indicates that our protocol may not be power-friendly in the large κ_i regime. On the other hand, if we can implement the 2-photon drive directly which is

also power-efficient so that we only need to consider the power cost for the 1-photon drive, then the error scaling will be modified as $\epsilon \propto \kappa_i^{5/6}/(P_{\text{in}}^{1/6}\chi^{2/3})$. We leave the investigation of other power-saving protocols for universal control as further work.

We also investigate the optimality of the Trotter scheme by comparing it with a direct optimization scheme. Specifically, we assume that the two-photon drive is absent from the beginning, and directly maximize the fidelity by optimizing $\alpha(t)$ using a more expressive ansatz that utilizes neural networks with periodic activation functions [27]. Our results indicate that this heuristic approach does not yield a better solution (in terms of fidelity and required power) compared to the Trotter scheme. We refer the reader to [18] for more details.

Additionally, if our 1-photon drive is too strong, the RWA may not be valid since the frequency ω_c of the resonator is always finite. The counter-rotating (non-RWA) terms like $(\Lambda_1^*(t)e^{i\omega_1(t)t}\hat{a}^\dagger + \text{h.c.})$ should be considered in Eq. (1). In Ref. [18], we discuss the way to mitigate the dominant effects from non-RWA terms by adjusting the driving amplitude $\Lambda_1(t)$ and frequency $\omega_1(t)$. It is also worth mentioning that there will be no $(\Lambda_1^*(t)e^{i\omega_1(t)t}\hat{a}^\dagger + \text{h.c.})$ term if on the hardware we can drive both charge and flux simultaneously with properly chosen amplitudes, which provides another way to mitigate non-RWA effects. However, there could still be a fundamental limitation on the fidelity as a function of two dimensionless parameters κ_i/χ and internal quality factor $Q_i := \omega_c/\kappa_i$. For the task of preparing $|1\rangle$, we found a rough lower bound for infidelity as $\epsilon \gtrsim 3\pi(\kappa_i/\chi)^{2/3}/(16Q_i^{1/3})$ even if we can directly implement 2-photon drive or do 1-photon drives on both charge and flux quadratures [18]. We leave the further improvement of this bound as an open question.

Conclusion– We show that by making a non-trivial extension of the displacement-induced, weak-Kerr photon blockade of Ref. [15], one can achieve any unitary operation in a blockade subspace of arbitrary dimension. The speed of the operations can be enhanced by using large displacements to overcome the adverse effects of photon loss. Further, this can be implemented using only 1-photon drives, provided that the input power is not a limitation and RWA conditions are justified. Moreover, from a computational perspective, our work simplifies the task of optimal control in such systems. It reduces the computational overhead for finding control sequences that utilize large photon number states by working in a special instantaneously displaced frame. As a result, our work provides a novel and efficient quantum control protocol for weak nonlinear bosonic systems, which could be helpful for future quantum information processing tasks on suitable platforms.

———————

* yuanming@uchicago.edu
† liangjiang@uchicago.edu

[1] A. Joshi, K. Noh, and Y. Y. Gao, Quantum Science and Technology 6, 033001 (2021).

[2] P. Kok, W. J. Munro, K. Nemoto, T. C. Ralph, J. P. Dowling, and G. J. Milburn, Reviews of Modern Physics 79, 135 (2007).

[3] S. Lloyd and S. L. Braunstein, Physical Review Letters 82, 1784 (1999).

[4] A. M. Eriksson, T. Sépulcre, M. Kervinen, T. Hillmann, M. Kudra, S. Dupouy, Y. Lu, M. Khanahmadi, J. Yang, C. C. Moreno, P. Delsing, and S. Gasparinetti, arXiv:2308.15320 [quant-ph].

[5] S. Krastanov, V. V. Albert, C. Shen, C.-L. Zou, R. W. Heeres, B. Vlastakis, R. J. Schoelkopf, and L. Jiang, Physical Review A 92, 040303(R) (2015).

[6] S. Hacohen-Gourgy, L. S. Martin, E. Flurin, V. V. Ramasesh, K. B. Whaley, and I. Siddiqi, Nature 538, 491–494 (2016).

[7] A. Eickbusch, V. Sivak, A. Z. Ding, S. S. Elder, S. R. Jha, J. Venkatraman, B. Royer, S. M. Girvin, R. J. Schoelkopf, and M. H. Devoret, Nature Physics 18, 1464–1469 (2022).

[8] A. A. Diringer, E. Blumenthal, A. Grinberg, L. Jiang, and S. Hacohen-Gourgy, arXiv:2301.09831 [quant-ph].

[9] J. Chan, T. P. M. Alegre, A. H. Safavi-Naeini, J. T. Hill, A. Krause, S. Gröblacher, M. Aspelmeyer, and O. Painter, Nature 478, 89 (2011).

[10] R. N. Patel, T. P. McKenna, Z. Wang, J. D. Witmer, W. Jiang, R. Van Laer, C. J. Sarabalis, and A. H. Safavi-Naeini, Physical Review Letters 127, 133602 (2021).

[11] Z. Vernon and J. E. Sipe, Physical Review A 91, 053802 (2015).

[12] H. Choi, M. Heuck, and D. Englund, Physical Review Letters 118, 223605 (2017).

[13] A. Anferov, A. Suleymanzade, A. Oriani, J. Simon, and D. I. Schuster, Physical Review Applied 13, 024056 (2020).

[14] M. Xu, R. Cheng, Y. Wu, G. Liu, and H. X. Tang, PRX Quantum 4, 010322 (2023).

[15] A. Lingenfelter, D. Roberts, and A. A. Clerk, Science Advances 7, eabj1916 (2021).

[16] A. W. Bruch, X. Liu, J. B. Surya, C.-L. Zou, and H. X. Tang, Optica 6, 1361 (2019).

[17] A. Kamal, A. Marblestone, and M. Devoret, Physical Review B 79, 184301 (2009).

[18] See Supplemental Material for the detailed discussion on the universality proof, error analysis for our scheme, the estimation of the feasibility on several experimental platforms to justify the scheme, as well as other numerical methods to perform optimal control for our problem.

[19] U. Boscain, M. Sigalotti, and D. Sugny, PRX Quantum **2**, 030203 (2021).

[20] S. G. Schirmer, H. Fu, and A. I. Solomon, Physical Review A **63**, 063410 (2001).

[21] S. Ashhab, F. Yoshihara, T. Fuse, N. Yamamoto, A. Lupascu, and K. Semba, Physical Review A **105**, 042614 (2022).

[22] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Journal of Magnetic Resonance **172**, 296 (2005).

[23] S. Machnes, U. Sander, S. J. Glaser, P. de Fouquières, A. Gruslys, S. Schirmer, and T. Schulte-Herbrüggen, Physical Review A **84**, 022305 (2011).

[24] L. Pontryagin, V. Boltânskij, K. M. R. Collection, K. Trirogoff, L. Neustadt, R. Gamkrelidze, and E. Mîŝenko, *The Mathematical Theory of Optimal Processes* (Interscience Publishers, 1962).

[25] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, "JAX: composable transformations of Python+NumPy programs," (2018).

[26] We should also note that suppressing errors beyond the third order using the discrete Trotter scheme is not possible due to the non-existence of positive decomposition as shown in Ref. [28]. In other words, higher order suppression require changing the sign of $\hat{H}'_{\mathrm{dr}}$, and that is not possible since the sign of the nonlinearity χ is fixed in a device.

[27] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, Advances in Neural Information Processing Systems **33**, 7462 (2020).

[28] M. Suzuki, Journal of Mathematical Physics **32**, 400 (1991).

[29] M. Suzuki, Journal of Mathematical Physics **26**, 601 (1985).

[30] U. B. Hoff, B. M. Nielsen, and U. L. Andersen, Optics Express **23**, 12013 (2015).

[31] S. Mittal, G. Moille, K. Srinivasan, Y. K. Chembo, and M. Hafezi, Nature Physics **17**, 1169 (2021).

[32] Z. Vernon, N. Quesada, M. Liscidini, B. Morrison, M. Menotti, K. Tan, and J. Sipe, Physical Review Applied **12**, 064024 (2019).

Supplemental Material for "Universal Control in Bosonic Systems with Weak Kerr Nonlinearities"

Ming Yuan,[1, *] Alireza Seif,[1] Andrew Lingenfelter,[2] David I. Schuster,[1, 2, 3] Aashish A. Clerk,[1] and Liang Jiang[1, †]

[1]*Pritzker School of Molecular Engineering, The University of Chicago, Chicago, IL 60637, USA*
[2]*Department of Physics, The University of Chicago, Chicago, IL 60637, USA*
[3]*James Franck Institute, The University of Chicago, Chicago, IL 60637, USA*
(Dated: December 27, 2023)

S1. DETAILED UNIVERSALITY PROOF

In this part, we show that the Hamiltonian $\hat{H}_{dr}$ (see Eq. (2)) can generate arbitrary unitary operations in U(N) within the $N = r + 1$ dimensional blockade subspace $\mathcal{H}_b$. Since r is an adjustable parameter, our protocol allows for "universal control" in bosonic systems with Kerr nonlinearities in the sense that any unitary operator with any chosen dimension is realizable in the system.

In quantum control theory [S1], a Hamiltonian defined in an N dimensional Hilbert space spanned by $\{|k\rangle\}_{k=0}^{N-1}$ contains a drift part $\hat{H}_d$ and several control parts $\hat{H}_{c,j}$. It can be expressed as

$$\hat{H}(t) = \hat{H}_d + \sum_j v_j(t)\hat{H}_{c,j}. \tag{S1}$$

Here, we consider the drift and one of the control parts with the following form

$$\hat{H}_d = \sum_{k=0}^{N-1} E_k |k\rangle\langle k| ,$$
$$\hat{H}_{c,1} = \sum_{k=0}^{N-2} d_k(|k+1\rangle\langle k| + |k\rangle\langle k+1|). \tag{S2}$$

Here $d_k \in \mathbb{R}$ and $d_k \neq 0$. The "universal control" [S2] is named as the ability to realize any unitary operation $\hat{U}_{tar}$ in U(N) with properly chosen $v_j(t)$ and evolution time T, such that

$$\hat{U}_{tar} = \hat{U}(T) = \mathcal{T} \exp\left[-i \int_0^T \hat{H}(t')\, dt'\right]. \tag{S3}$$

A Theorem in Ref. [S3] suggests a sufficient condition for the choice of E_k and d_k to make the system universally controllable. We first repeat the theorem here and check that our Hamiltonian $\hat{H}_{dr}$ restricted in blockade subspace $(\hat{\Pi}_r \hat{H}_{dr} \hat{\Pi}_r$, see Eq. (5)) satisfies those criteria.

Theorem 1 (Ref. [S3]) *Denote* $\mu_k = E_k - E_{k+1}$. *If* $\mu_0 \neq 0$ *and* $\mu_k^2 \neq \mu_0^2$ *for* $k > 0$ *(or similarly if* $\mu_{N-2} \neq 0$

and $\mu_k^2 \neq \mu_{N-2}^2$ for $k < N - 2$), then the dynamical Lie group of the system $\hat{H}(t)$ defined in Eq. (S1, S2) is at least SU(N). Further, if $\text{Tr}\left[\hat{H}_d\right] \neq 0$, the dynamical Lie group is U(N).

We notice that $\hat{\Pi}_r \hat{H}_{dr} \hat{\Pi}_r$ contains the drift part $\hat{H}_{d,0}$ and two control parts $\hat{H}_{c,R}$ and $\hat{H}_{c,I}$. It is easy to verify that $\hat{H}_{c,R}$ meets the requirement for the control part. Then we should check the nonlinear condition for $\hat{H}_{d,0}$. In this case, we have $E_k = \chi(k^2 - k)/2 + \Delta_0 k$, which gives the nearest energy difference $\mu_k = -\chi k - \Delta_0$. To match the condition in Theorem 1, we need to make sure $\mu_0^2 \neq \mu_{N-2}^2$, which leads to $r \neq -\frac{2\Delta_0}{\chi} + 1$. Further, since μ_k is monotonic in k, we cannot find both k_1, k_2 $(0 < k_1, k_2 < N-2)$ such that $\mu_0^2 = \mu_{k_1}^2$ and $\mu_{N-2}^2 = \mu_{k_2}^2$. This concludes the proof of the universal controllability of our system, even if we can fix $\text{Im}[\alpha(t)] = 0$ all the time.

Moreover, the ability to control $\text{Im}[\alpha(t)]$ provides us with an additional degree of freedom to control the system, which in principle makes it possible to perform operations arbitrarily fast. To demonstrate this, we first show that the nested commutators between $i\hat{H}_{c,R}$ and $i\hat{H}_{c,I}$ are sufficient to form a complete basis of the Lie algebra associated with SU(N). The reason is that $[i\hat{H}_{c,R}, i\hat{H}_{c,I}] = \sum_{n=0}^{r} -2[3n^2 - (4r+1)n + r^2]i |n\rangle\langle n|$ is diagonal in Fock basis, has zero trace, and fulfills the anharmonicity requirement as that for the drift term $\hat{H}_d$ in Theorem 1. Further, as proved in Theorem 1, $[i\hat{H}_{c,R}, i\hat{H}_{c,I}]$ and $i\hat{H}_{c,R}$ are sufficient to generate a set of complete basis of Lie algebra $\mathfrak{su}(N)$. In practice, the overall phase for unitary operations has no physical meaning, which makes SU(N) group sufficient for universality. Second, since $\chi \text{Re}[\alpha(t)]\hat{H}_{c,R} + \chi \text{Im}[\alpha(t)]\hat{H}_{c,I}$ is sufficient to achieve the desired unitary, we can simply increase the amplitude of $\alpha(t)$ to do things arbitrarily fast. In the short time limit, the drift term $\hat{H}_{d,0}$ does not contribute to the dynamics and therefore does not impose a speed constraint.

S2. BASIC ERROR ANALYSIS

In this section, we analyze the errors in implementing a desired operation using our protocol where only 1-photon drive is allowed. Notice that, even if all the driving parameters are perfectly implemented, there are

———
[*] yuanming@uchicago.edu
[†] liangjiang@uchicago.edu

errors stemming from photon loss and inaccuracies of the Trotter design in Eq. (16). Reducing the total time of the protocol T helps to mitigate the photon loss errors, but at the same time increases the required input power, which could be limited in practice. This limited power also introduces coherent errors in the Trotter approximation of the unitary operator of interest. Therefore, there is a trade-off between these two sources of error, which we discuss in detail in the following.

For simplicity, we only consider the error in state preparation tasks characterized by infidelity

$$\epsilon = 1 - \langle \psi_{\text{tar}} | \hat{\rho}(T) | \psi_{\text{tar}} \rangle, \qquad (S4)$$

where $|\psi_{\text{tar}}\rangle$ is a pure state that we are interested in preparing and $\hat{\rho}(T)$ is the state of the system at the end of the evolution. Here, $\hat{\rho}(T)$ is the state obtained after evolving for time T under the Lindblad equation

$$\frac{d\hat{\rho}}{dt} = -i[\hat{H}'_{\text{dr}}, \hat{\rho}] + \kappa \mathcal{D}[\hat{a}]\hat{\rho}, \qquad (S5)$$

where $\hat{H}'_{\text{dr}}$ is introduced in Eq. (9).

To study the scaling of error ϵ with relevant physical parameters in the problem, we focus our attention on the case of single-photon state preparation when the blockade subsystem has only 2 dimension and $\alpha(t)$ is constant, as originally proposed in Ref. [S4]. Therefore, in this case we set $|\psi_{\text{tar}}\rangle = |1\rangle$, and $\hat{\rho}(0) = |0\rangle\langle 0|$. To prepare the target state $|1\rangle$, we choose $\Delta_0 = 0$ and $\alpha(t) = \alpha$, a constant that is assumed to be a real number for simplicity. Therefore, the effective Hamiltonian in the blockade subspace is

$$\hat{H}_{\text{dr}}^{\text{qb}} = -\chi \alpha \hat{\sigma}_x. \qquad (S6)$$

Here $\hat{\sigma}_x = |0\rangle\langle 1| + |1\rangle\langle 0|$. Moreover, in this subspace the photon loss dissipator $\mathcal{D}[\hat{a}]$ in Eq. (S5) is now simply modified to $\mathcal{D}[\hat{\sigma}_-]$, where $\hat{\sigma}_- = |0\rangle\langle 1|$. Note that in the ideal case, when there is no loss or Trotter error, we can perfectly prepare $|1\rangle$ by evolving $|0\rangle$ under $\hat{H}_{\text{dr}}^{\text{qb}}$ in Eq. (S6) for time $T = \frac{\pi}{2\chi\alpha}$. For simplicity here we ignore the part of the dynamics where $\alpha(t)$ increases from 0 to α in the beginning and decreases to 0 in the end [S4], while only discussing errors with a constant $\alpha(t)$. In fact, using a time-dependent $\alpha(t)$ will introduce additional imperfection for Trotter approximation. However, as discussed in Sec. S3, given the assumption that $\alpha(t)$ varies slowly over time, these additional errors are higher-order effects compared with the major Trotter error scaling we focus on later.

To find the scaling properties of ϵ, we treat the photon loss (ϵ_{loss}) and Trotter (ϵ_{tt}) errors independently.

First, for the photon loss error, we assume that the Hamiltonian $\hat{H}_{\text{dr}}^{\text{qb}}$ in Eq. (S6) is perfectly implemented so that the quantum state is confined to the blockade subspace all the time. If $\chi\alpha \gg \kappa$, we can treat the dissipative term $\kappa\mathcal{D}[\sigma_-]$ perturbatively. So, to the lowest order, the error from the photon loss process satisfies $\epsilon_{\text{loss}} = c_1 \kappa T$, where the coefficient c_1 in general depends on the function $\alpha(t)$. In our case, we find that $c_1 = \frac{3}{8}$ (see Sec. S9).

Next, to estimate the error from the Trotter approximation, we ignore the photon loss process and assume that the system undergoes unitary evolution with Hamiltonian $\hat{H}'_{\text{dr}}$ shown in Eq. (9). For simplicity, we use the discrete Trotter formula to illustrate analysis, but the results apply to the proposed continuous version as well (see Sec. S8). Let $\hat{U}_{\text{tar}} = e^{-i\frac{\hat{H}_1 + \hat{H}_2}{2}\delta T}$ and $\hat{U} = e^{-i\hat{H}_1\delta T/4}e^{-i\hat{H}_2\delta T/2}e^{-i\hat{H}_1\delta T/4}$. Using the discrete version of Trotter formula [S5], we have

$$\hat{U}_{\text{tar}}^{-1/2}\hat{U}\hat{U}_{\text{tar}}^{-1/2} = \exp\{(\delta T)^3 \hat{A}_3 + O[(\delta T)^5]\}, \qquad (S7)$$

where $\hat{A}_3 = i[[\hat{H}_1, \hat{H}_2], \hat{H}_1 + 2\hat{H}_2]/192$ is anti-Hermitian. Further, recall the form of $\hat{H}'_{\text{dr}}[\tilde{\alpha}]$. Since we work in the regime that $|\alpha| \gg \sqrt{N}$ (N is the dimension of the blockade subspace), the dominant contribution to $\hat{A}_3$ comes from the $(\frac{\chi}{2}\alpha_{1,2}^2 \hat{a}^{\dagger 2} + \text{h.c.})$ terms in the Hamiltonian $\hat{H}_1$ and $\hat{H}_2$, and therefore the matrix elements of $\hat{A}_3$ that we focus on scale as $O[(\chi\alpha^2)^3]$. Finally, as mentioned before, we need to sequentially apply $M = T/\delta T$ repetitions of the Trotter operation $\hat{U}$ to approximately achieve our desired unitary $\hat{U}_{\text{tar}}^M$, where $|\psi_{\text{tar}}\rangle\langle\psi_{\text{tar}}| = \hat{U}_{\text{tar}}^M \hat{\rho}(0) \hat{U}_{\text{tar}}^{\dagger M}$. Therefore, we find

$$\begin{aligned} \epsilon_{\text{tt}} &= 1 - \langle\psi_{\text{tar}}| \hat{U}^M \hat{\rho}(0) \hat{U}^{\dagger M} |\psi_{\text{tar}}\rangle \\ &= O[(M(\chi\alpha^2\delta T)^3)^2]. \end{aligned} \qquad (S8)$$

Recalling the relation that $T \propto 1/(\chi\alpha)$ and $M = T/\delta T$, we finally obtain $\epsilon_{\text{tt}} = c_2/[M^4(\chi T)^6]$, where c_2 is again the coefficient that depends on specific form of our pulse design and can be estimated numerically (see Sec. S9). Note that in our numerical work $f(t)$ function is chosen as

$$f(t) = \frac{i\pi}{2} e^{-i2\pi|s(t) - 1/2|}, \qquad (S9)$$

where $s(t) = t/\delta T - \lfloor t/\delta T \rfloor$. This choice of $f(t)$ also satisfies the requirements shown in Eq. (14, 15), and also $|f(t)|$ is constant over time. In this situation, the reference displaced frame (characterized by $\tilde{\alpha}(t)$) oscillates back and forth along a semi-circle. In practice, it can be implemented with 1-photon drives using two interleaved tones with frequencies $\omega_1 \pm \omega_r$.

Since errors from the loss and Trotter approximation are independent, to the lowest order they can be added directly to find the total infidelity as

$$\epsilon_{\text{tot}} = \epsilon_{\text{loss}} + \epsilon_{\text{tt}} = c_1 \kappa T + \frac{c_2}{M^4(\chi T)^6}. \qquad (S10)$$

If the driving amplitude is unbounded, we can use an arbitrarily short time T with arbitrary fast oscillating $\tilde{\alpha}(t)$ ($M \to \infty$) such that both ϵ_{loss} and ϵ_{tt} are suppressed to zero. However, in practice there are possible limitations that prevent us from doing that. Here we consider a key limitation, namely the input power P_{in} that can be applied to the system and evaluate the optimal ϵ_{tot} that we can achieve under this constraint.
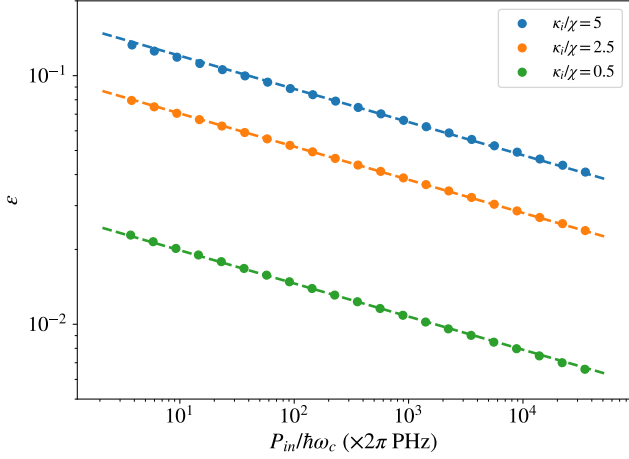
FIG. S1. Numerical results of the infidelity of Fock state $|1\rangle$ preparation with given input power and corresponding optimal choice of gate time. The choice of Kerr parameter is $\chi = 2\pi \times 3$ kHz. The dynamics is calculated with the assumption that RWA is valid. Different colors indicate different κ_i/χ ratio. The dashed lines are references for $\epsilon \propto P_{\text{in}}^{-2/15}$ scaling.

First, we assume that $\chi|\alpha|^2 \ll \omega_c$, which can be checked self-consistently since otherwise RWA will not apply, as discussed in Sec. S5. This assumption allows us approximately write down the input power as

$$P_{\text{in}} \simeq \frac{|\Lambda_1|^2}{\kappa_e} \hbar\omega_c, \tag{S11}$$

where κ_e is the external loss rate induced by coupling to the driving port. The total loss κ is composed of $\kappa = \kappa_e + \kappa_i$, where κ_i is the internal loss rate of the resonator. We notice from Eq. (10) that, $\Lambda_1 \sim O(\chi\tilde{\alpha}^3) + i\dot{\tilde{\alpha}}$. We then need to estimate the contribution from $\dot{\tilde{\alpha}}$. In our design we have $\dot{\tilde{\alpha}} \sim O(\alpha M/T)$. Besides, since we are only considering lowest order errors from the Trotter expansion, the Trotter error should be small. Therefore, we should have $\chi\alpha^2 \delta T \ll 1$, or equivalently $\chi\alpha^2 \ll \omega_r$. This implies that $M \gg \chi\alpha^2 T$, and as a result $i\dot{\tilde{\alpha}}$ is the dominating term in Λ_1. So, we can write P_{in} as

$$P_{\text{in}} \sim O(\frac{M^2\alpha^2}{\kappa_e T^2}) := \frac{c_3 M^2}{\kappa_e \chi^2 T^4}, \tag{S12}$$

where $\hbar\omega_c$ and the proportionality constant originating from $f(t)$ are both absorbed in c_3. By eliminating M in Eq. (S10) using Eq. (S12) we obtain

$$\epsilon_{\text{tot}} = c_1(\kappa_e + \kappa_i)T + \frac{c_2 c_3^2}{P_{\text{in}}^2 \kappa_e^2 \chi^{10} T^{14}}. \tag{S13}$$

As a result, we find that the scaling of T^{opt}, namely the optimal time that minimizes ϵ_{tot}, and the corresponding optimal error $\epsilon_{\text{tot}}^{\text{opt}}$ is given by

$$T^{\text{opt}} = \left(\frac{14 c_2 c_3^2}{c_1}\right)^{1/15} \cdot \frac{1}{P_{\text{in}}^{2/15} \chi^{2/3} \kappa_e^{2/15} (\kappa_e + \kappa_i)^{1/15}},$$

$$\epsilon_{\text{tot}}^{\text{opt}} = 15 \left(\frac{c_1^{14} c_2 c_3^2}{14^{14}}\right)^{1/15} \cdot \left(\frac{1}{P_{\text{in}}^{2/15} \chi^{2/3}} \frac{(\kappa_e + \kappa_i)^{14/15}}{\kappa_e^{2/15}}\right). \tag{S14}$$

It is also natural to assume that κ_i is a fixed property of the device while κ_e is adjustable to further optimize $\epsilon_{\text{tot}}^{\text{opt}}$. We can find the optimal ratio of $\kappa_e/\kappa_i = \frac{1}{6}$ that minimizes $\epsilon_{\text{tot}}^{\text{opt}}$ while keeping χ, κ_i and input power P_{in} fixed and using the optimal T^{opt}. This gives

$$\epsilon_{\text{tot}}^{\text{opt}} = \frac{15(c_1^{14} c_2 c_3^2)^{1/15}}{2^{14/15} \cdot 6^{4/5}} \cdot \left(\frac{\kappa_i^{4/5}}{P_{\text{in}}^{2/15} \chi^{2/3}}\right). \tag{S15}$$

We also numerically investigate the infidelity ϵ as prescribed by Eq. (S4) as a function of input power. To find the optimal time for the protocol in our numerical simulations we first need to determine the constants in Eq. (S13). More details can be found in Sec. S9. We first obtain c_2 by varying M and T in the simulations and extracting the corresponding proportionality constant in Eq. (S10). We then obtain c_3 by using the explicit form of $f(t)$ in Eq. (S9), which results in $c_3 \simeq \frac{\pi^6}{4}\hbar\omega_c$. Therefore, together with $c_1 = \frac{3}{8}$ from perturbative analysis, we can then find T^{opt} for a given power, which subsequently determines the required M using Eq. (S12). In this way, both the optimal protocol time T^{opt} and $\tilde{\alpha}(t)$ are fully determined. Here we also use the optimal choice of $\kappa_e = \frac{1}{6}\kappa_i$.

In FIG. S1, we observe that the numerically obtained infidelities agree well with the estimated $\epsilon \propto P_{\text{in}}^{-2/15}$ scaling in a wide range of P_{in} values.

Finally, in the numerical simulation we have assumed the validity of RWA so that the dynamics is irrelevant with ω_c after going to the rotating frame. We notice that if the required α or M (equivalently ω_r) is so large such that Λ_1 is comparable with the frequency of the resonator ω_c, then those off-resonant terms which has been ignored under RWA in the beginning may lead to non-negligible effects on the dynamics. In Sec. S5, we use scaling analysis to briefly discuss those effects and strategies to partially compensate them by adjusting $\Lambda_1(t)$ and $\omega_1(t)$, or driving both charge and flux quadratures together rather than only one of them on hardware.

S3. TROTTER ERRORS WITH TIME-DEPENDENT $\alpha(t)$

In this section, we will also discuss the error scaling properties of our protocol, but specifically focus on the Trotter error in the case that $\alpha(t)$ designed by optimal control algorithm is a slowly varied function that depends

on time t. Here, by "slowly varied", we mean the time derivative of $\alpha(t)$ scales as $\dot{\alpha}(t) \sim O(\alpha/T)$, and in general $d^n\alpha(t)/dt^n \sim O(\alpha/T^n)$. To be more precise, we will calculate the difference between both sides of Eq. (16) in detail, where on the left-hand side is the evolution we can achieve with only 1-photon drive, and on the right-hand side is the target unitary operation that we want to achieve. Notice that, unlike the constant α case in the main text that $T \propto 1/(\chi\alpha)$, here we do not have such property rigorously. But, it is still reasonable to assume that T roughly scales as $O(1/\chi\alpha)$ since the dominant part in $\hat{H}_{\mathrm{dr}}$ is still $[\chi\alpha(t)\hat{a}^\dagger(\hat{n}-r)+\mathrm{h.c.}]$ if α is large. We will keep this assumption in the following derivation.

First, let us calculate the difference of unitary operation within each time slice $t \in [k\delta T, (k+1)\delta T]$ where k is an integer and $0 \leq k < M$. We denote $\hat{U}_k = \mathcal{T}\exp\{-i\int_{k\delta T}^{(k+1)\delta T} \hat{H}'_{\mathrm{dr}}[\tilde{\alpha}(t)]\,dt\}$ and $\hat{U}_{\mathrm{tar},k} = \mathcal{T}\exp\{-i\int_{k\delta T}^{(k+1)\delta T} \hat{H}_{\mathrm{dr}}[\alpha(t)]\,dt\}$. Recall that $\tilde{\alpha}(t) = \alpha(t)f(t)$, where $f(t)$ is a periodic function with period δT, scales as $O(1)$, and satisfies Eq. (14, 15). Without loss of generality, we can just focus on $k=0$ case and calculate $\hat{U}_0\hat{U}^\dagger_{\mathrm{tar},0}$, since $\hat{U}_k\hat{U}^\dagger_{\mathrm{tar},k}$ with any k has a similar structure as the situation with $k=0$. By analogy with the Magnus expansion, we have

$$
\hat{U}_0\hat{U}^\dagger_{\mathrm{tar},0} = \exp\left(-i\int_0^{\delta T} dt\,[\hat{H}'_{\mathrm{dr}}(t) - \hat{H}_{\mathrm{dr}}(t)] - \frac{1}{2}\int_0^{\delta T} dt_2\int_0^{t_2} dt_1\{[\hat{H}'_{\mathrm{dr}}(t_2), \hat{H}'_{\mathrm{dr}}(t_1)] - [\hat{H}_{\mathrm{dr}}(t_2), \hat{H}_{\mathrm{dr}}(t_1)]\}\right.
$$
$$
\left. + \frac{1}{2}\int_0^{\delta T} dt_2\int_0^{\delta T} dt_1[\hat{H}'_{\mathrm{dr}}(t_2), \hat{H}_{\mathrm{dr}}(t_1)] + \hat{R}_3\right). \tag{S16}
$$

It is easy to see that $\hat{R}_3$ is anti-Hermitian and $\hat{R}_3 \sim O[(\chi\alpha^2\delta T)^3]$. As we discussed in the main text, if $\alpha(t)$ is a constant within $t \in [0, \delta T]$, the first and the second order terms written explicitly in Eq. (S16) are zero. But here we want to talk about the more general case that $\alpha(t)$ is time dependent. Even in this situation, we will show that the contribution of the error from the first two orders are small compared with the third order term.

We can write $\alpha(t)$ as

$$
\alpha(t) = \alpha(t_0) + \dot{\alpha}(t_0)(t - t_0) + \frac{1}{2}\ddot{\alpha}(t_0)(t - t_0)^2 + O[(t - t_0)^3], \tag{S17}
$$

where $t_0 = \frac{\delta T}{2}$. Since we assumed that $\alpha(t)$ is a slowly varied function, we have $\dot{\alpha} \sim O(\alpha/T)$ and $\ddot{\alpha} \sim O(\alpha/T^2)$. As a result, based on the structure of $f(t)$, for the first order term we have

$$
\int_0^{\delta T} dt\,\chi[\tilde{\alpha}(t) - \alpha(t)] \sim O\left[\frac{\chi\alpha(\delta T)^3}{T^2}\right],
$$
$$
\int_0^{\delta T} dt\,\chi\tilde{\alpha}^2(t) \sim O\left[\frac{\chi\alpha^2(\delta T)^3}{T^2}\right]. \tag{S18}
$$

Since we have assumed that $T \sim O(1/\chi\alpha)$, both of the two terms are smaller compared with $O[(\chi\alpha^2\delta T)^3]$ when α is large.

A similar analysis can be performed for the second-order terms. The dominant part lies in $[\hat{H}'_{\mathrm{dr}}(t_2), \hat{H}'_{\mathrm{dr}}(t_1)]$

since it contains $O[(\chi\alpha^2)^2]$ coefficients. We find that

$$
\int_0^{\delta T} dt_2\int_0^{t_2} dt_1\,\chi^2[\tilde{\alpha}^2(t_2)\tilde{\alpha}^{*2}(t_1) - \tilde{\alpha}^{*2}(t_2)\tilde{\alpha}^2(t_1)]
$$
$$
\sim O\left(\frac{\chi^2\alpha^4(\delta T)^3}{T}\right), \tag{S19}
$$

which is also a higher order term compared with $O[(\chi\alpha^2\delta T)^3]$. So, the dominant scaling of Trotter error does not change even if $\alpha(t)$ is a slowly varied function introduced in Eq. (7).

S4. ERRORS FROM INACCURATE CONTROL

In this part, we will investigate the extra infidelity induced by inaccurate control of $\Lambda_1(t)$. As shown in Sec. S2, $|\Lambda_1| \sim O(\alpha M/T)$ can be strong enough so that a deviation for a small portion of it may cause a huge impact. Here we consider the problem that how the infidelity changes if the actual driving pulse we use is $(1+\eta)\Lambda_1(t)$ where η is a small dimensionless number.

We first briefly derive the contribution of the extra $\eta\Lambda_1(t)$ to the dynamics. Notice that, if we go to the same displaced rotating frame as mentioned in the main text, now the Hamiltonian should be written as

$$
\hat{H}_\eta = \hat{H}'_{\mathrm{dr}}[\tilde{\alpha}(t)] + [\eta\Lambda_1(t)\hat{a}^\dagger + \mathrm{h.c.}], \tag{S20}
$$

where $\hat{H}'_{\mathrm{dr}}[\tilde{\alpha}(t)]$ has been defined in Eq. (9) and $\Lambda_1(t) = \chi\tilde{\alpha}(t)[|\tilde{\alpha}(t)|^2 - r] + i\kappa\tilde{\alpha}(t)/2 + i\dot{\tilde{\alpha}}(t)$ (as in Eq. (10)). We can again perform a frame transformation such that $\hat{a} \to$
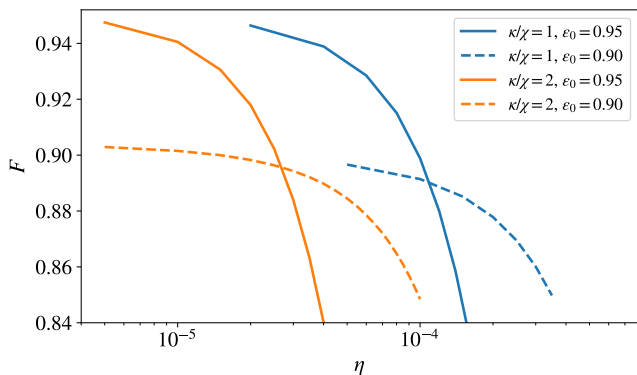
FIG. S2. Fidelity changes due to the amplitude deviation with the ideal pulses for the Fock $|1\rangle$ state preparation task, as explained in Sec. S4.

$\hat{a} + \beta(t)$ where $\beta(t)$ satisfies

$$\dot{\beta}(t) + \frac{\kappa}{2}\beta(t) = -i\eta\Lambda_1(t). \tag{S21}$$

Further, due to the specific form of $\Lambda_1(t)$, we can write $\beta(t) = \beta_1(t) + \beta_2(t)$ where $\beta_1(t) = \eta\tilde{\alpha}(t)$ and $\beta_2(t) = -i\eta \int_0^t e^{-\kappa(t-t')/2}\chi\tilde{\alpha}(t')[|\tilde{\alpha}(t')|^2 - r]\,dt'$. It is worth mentioning that, although $i\dot{\tilde{\alpha}}$ dominates in Λ_1, its contribution to β is much smaller compared with the contribution from the $O(\chi\alpha^3)$ term, provided that $\alpha \gg 1$.

Now the Hamiltonian in the new frame should be

$$\begin{aligned}
\hat{H}'_\eta = {} & \frac{\chi}{2}(\hat{a}^\dagger + \beta^*)^2(\hat{a} + \beta)^2 \\
& + \{\chi\tilde{\alpha}(\hat{a}^\dagger + \beta^*)[(\hat{a}^\dagger + \beta^*)(\hat{a} + \beta) - r] + \text{h.c.}\} \\
& + [\frac{\chi}{2}\tilde{\alpha}^2(\hat{a}^\dagger + \beta^*)^2 + \text{h.c.}].
\end{aligned} \tag{S22}$$

Clearly $\hat{H}'_\eta$ deviates from the ideal $\hat{H}'_{\text{dr}}[\tilde{\alpha}(t)]$ as β increases from zero. Besides, in general $\beta(T) \neq 0$, which suggests that we need an extra displacement operation $\hat{D}[\beta(T)]$ to let the states go back to the non-displaced frame. But in practice we do not know what the $\beta(T)$ is and will not actively apply this operation, which may lead to extra errors.

Finally we try to numerically consider a specific example, which is the Fock $|1\rangle$ state preparation with constant α as mentioned in Sec. S2. We assume that, T and M are chosen optimally such that for a given target infidelity ϵ_0 we have $c_1\kappa T = \frac{14}{15}\epsilon_0$ and $\frac{c_2}{M^4(\chi T)^6} = \frac{1}{15}\epsilon_0$. However, during the constant α evolution time the actual 1-photon driving amplitude we implement is $(1 + \eta)\Lambda_1(t)$. From FIG. S2 we can see how the fidelity $F = 1 - \epsilon$ decreases as the increase of the inaccuracy η. We notice that for a smaller ϵ_0 expected, we need a larger α. However, larger α will result in larger deviation β. The fidelity will be more sensitive to η with larger α and therefore it decays faster as η grows when ϵ_0 is small.

S5. DYNAMICS BEYOND ROTATING WAVE APPROXIMATION

A. Overview

In the main text, we did not talk about any property related to the resonator frequency ω_c, since it simply disappears after we go to the rotating frame and perform the rotating wave approximation (RWA). In practice, however, the validity of RWA will set a limitation on our current protocol. If we can do either charge or flux drive but not both, then the driving amplitude Λ_1 cannot be too strong, which further limits the maximum $|\alpha|$ and the oscillation frequency ω_r for $\tilde{\alpha}(t)$ that is used to suppress the infidelity. In this section, we will discuss the effects of those off-resonant terms in the aspect of scaling, and we also provide some strategies to partially compensate for them.

Before we start to talk about the specific problem, we will briefly explain the analysis strategy we are going to use. Rather than going to the displaced rotating frame we mentioned in the main text, which is treated as the 0-th order reference frame here, we try to find another reference frame with higher order corrections on the original one and compare the difference between Hamiltonian there and the one ($\hat{H}'_{\text{dr}}[\tilde{\alpha}(t)]$) we desire to engineer. To make sure the dynamics induced by the difference between the two Hamiltonian are negligible, we need higher-order corrections on the 1-photon driving amplitude $\Lambda_1(t)$ and frequency $\omega_1(t)$, together with several assumptions on the relationship among some relevant parameters (see Eq. (S30) for example).

As we shall see, there are two kinds of terms in the difference between desired and actual Hamiltonian. One kind includes terms that have a fast-oscillating coefficient, which is usually proportional to $e^{ik\omega_c t}$ or $e^{ik\omega_r t}$ with nonzero integer k; while another kind contains those terms without fast-oscillating factors, which are called as "slowly varied" terms. Besides, we also call terms with coefficients $e^{ik\omega_r t}$ "slowly varied" when they are compared with $e^{ik\omega_c t}$ (or we call "slowly varied" compared with ω_c for short).

For both cases, the dynamics induced by those terms is negligible if the amplitudes (absolute value) of their coefficients are too small such that the integration of the amplitudes over gate time T is far less than 1. For example, terms with slowly varied coefficient $c(t)$ or the fast oscillating $c(t)e^{i\omega t}$ can be both ignored if $\int_0^T |c(t)|\,dt \ll 1$. However, for terms with fast oscillating coefficients that do not belong to the former situation, we need to consider some effective Hamiltonian corresponding to those terms. For example, the dynamics generated by a Hamiltonian like $\hat{H} = \hat{A}e^{i\omega t} + \hat{A}^\dagger e^{-i\omega t}$ can be approximated by that generated via an effective Hamiltonian $\hat{H}_{\text{eff}} = [\hat{A}, \hat{A}^\dagger]/\omega$, provided that ω is much larger than some specific matrix norm of $\hat{A}$. This can be simply demonstrated via Magnus expansion. In this way, we will focus on the scaling of the coefficients for those slowly varied effective counterparts

and see if the time integration of those is far less than 1 and therefore negligible as well. Similarly, remember that in the parameter regime we want to work in, the coherent error induced by $(\frac{\chi}{2}\tilde{\alpha}^2\hat{a}^{\dagger 2} + \text{h.c.})$ is negligible. In general, for terms that oscillate in the form of $f^2(t)$ (where $f(t)$ is a periodic function with frequency ω_r and satisfies Eq. (14, 15)), their contribution to the dynamics can be ignored even when their amplitudes scale as $O(\chi\alpha^2)$. We will also make use of this fact later. These discussions provide support for the following derivations.

Our scaling analysis also relies on some assumptions for the ideal blockade Hamiltonian. We assume the dimension of the blockade subspace is small, such that in the scaling analysis the matrix elements of $\hat{a}$ and $\hat{a}^{\dagger}$ are of the order $O(1)$ when dynamics is restricted into the blockade subspace. Also, we assume that the control function $\alpha(t)$ is slowly varied in time, which (as mentioned in Sec. S3) implies that the time derivative of $\alpha(t)$ scales as $\dot{\alpha}(t) \sim O(\alpha/T)$, and similarly $\mathrm{d}^n\alpha(t)/\mathrm{d}t^n \sim O(\alpha/T^n)$.

B. Detailed scaling analysis on the dynamics

In general, the Hamiltonian in the lab frame without RWA should be written as

$$
\begin{aligned}
\hat{H} = &\frac{\chi}{12}(\hat{a}^{\dagger} + \hat{a})^4 + (\omega_c - \chi)\hat{a}^{\dagger}\hat{a} \\
&+ [\Lambda_1(t)e^{-i\omega_1(t)t} + \Lambda_1^*(t)e^{i\omega_1(t)t}](\hat{a}^{\dagger} + \hat{a}).
\end{aligned} \tag{S23}
$$

From then on we denote $\phi_d(t) := \omega_1(t)t - \omega_c t$ and as shown in the main text we have $\dot{\phi}_d(t) \sim O(\chi\alpha^2)$. If we still go to the displaced rotating frame shown in the main text, the Hamiltonian will become

$$
\begin{aligned}
\hat{H}_{\text{NR}} = &\frac{\chi}{2}(\hat{a}^{\dagger} + \tilde{\alpha}^*)^2(\hat{a} + \tilde{\alpha})^2 - \dot{\phi}_d(t)(\hat{a}^{\dagger} + \tilde{\alpha}^*)(\hat{a} + \tilde{\alpha}) + [(\Lambda_1(t) - i\dot{\tilde{\alpha}} - i\kappa\tilde{\alpha}/2)\hat{a}^{\dagger} + \text{h.c.}] \\
&+ (\Lambda_1^*(t)e^{2i\phi_d(t)}e^{2i\omega_c t}\hat{a}^{\dagger} + \text{h.c.}) \\
&+ \left\{ \left[\frac{\chi}{3}(\hat{a}^{\dagger} + \tilde{\alpha}^*)^3(\hat{a} + \tilde{\alpha}) + \frac{\chi}{2}(\hat{a}^{\dagger} + \tilde{\alpha}^*)^2 \right] e^{2i\phi_d(t)}e^{2i\omega_c t} + \frac{\chi}{12}(\hat{a}^{\dagger} + \tilde{\alpha}^*)^4 e^{4i\phi_d(t)}e^{4i\omega_c t} + \text{h.c.} \right\}.
\end{aligned} \tag{S24}
$$

It can be easily seen that we can recover the Hamiltonian under RWA by throwing away all the fast oscillating terms related to $e^{i\omega_c t}$. Therefore, as proposed earlier we can choose $\Lambda_1^{[0]}$ and $\phi_d^{[0]}$ to achieve the desired Hamiltonian under RWA by

$$
\begin{aligned}
\Lambda_1^{[0]} &= \dot{\phi}_d^{[0]}\tilde{\alpha}^{[0]} - \chi\tilde{\alpha}^{[0]}(|\tilde{\alpha}^{[0]}|^2 + r) + i\dot{\tilde{\alpha}}^{[0]} + i\kappa\tilde{\alpha}^{[0]}/2, \\
\dot{\phi}_d^{[0]} &= 2\chi|\tilde{\alpha}^{[0]}|^2.
\end{aligned} \tag{S25}
$$

where $\tilde{\alpha}^{[0]}$ is the $\tilde{\alpha}$ function we designed in the main text.

Now we want to focus on those fast oscillating terms. One way is to go to a slightly different displaced frame to absorb some large components. Note that $\Lambda_1 \sim i\dot{\tilde{\alpha}} \sim \alpha\omega_r \gg \chi\alpha^3$, which means $(\Lambda_1^*(t)e^{2i\phi_d(t)}e^{2i\omega_c t}\hat{a}^{\dagger} + \text{h.c.})$ has the largest amplitude and we want to deal with it first. Basically, we want to choose $\tilde{\alpha} = \tilde{\alpha}^{[0]} + \tilde{\alpha}^{[1]}$ such that the amplitude of the linear $\hat{a}^{\dagger}$ term can be mostly absorbed in $\tilde{\alpha}^{[1]}$. We first introduce $\tilde{\alpha}^{[1']}$ tentatively such that

$$
\begin{aligned}
i\dot{\tilde{\alpha}}^{[1']} + i\kappa\tilde{\alpha}^{[1']}/2 = &[\Lambda_1^{*[0]} + \chi\tilde{\alpha}^{*[0]}(|\tilde{\alpha}^{[0]}|^2 + 1)]e^{2i\phi_d^{[0]}(t)}e^{2i\omega_c t} \\
&+ \frac{\chi}{3}(\tilde{\alpha}^{*[0]})^3 e^{4i\phi_d^{[0]}(t)}e^{4i\omega_c t} + \frac{\chi}{3}(\tilde{\alpha}^{[0]})^3 e^{-2i\phi_d^{[0]}(t)}e^{-2i\omega_c t}.
\end{aligned} \tag{S26}
$$

For simplicity, we denote the right-hand-side (RHS) of the equation above as $g(t)$.

The solution for $\tilde{\alpha}^{[1']}$ can be written down explicitly as

$$
\tilde{\alpha}^{[1']}(t) = -i \int_0^t e^{-\kappa(t-t')/2} g(t') \, \mathrm{d}t'. \tag{S27}
$$

Via the technique of integration by parts, we have the following identity

$$
\begin{aligned}
\int_0^t h(t')e^{i\omega t'} \, \mathrm{d}t' = &\sum_{n=0}^{k} \left(\frac{i}{\omega} \right)^{n+1} [h^{(n)}(0) - h^{(n)}(t)e^{i\omega t}] \\
&+ \left(\frac{i}{\omega} \right)^{k+1} \int_0^t h^{(k+1)}(t')e^{i\omega t'} \, \mathrm{d}t',
\end{aligned} \tag{S28}
$$

where $h^{(n)}(t) = \frac{\mathrm{d}^n}{\mathrm{d}t^n} h(t)$. In Eq. (S28) we have implicitly assumed that $h^{(n)}(t)$ is a continuous function for $n \leq k$. Notice that if for any m, $\frac{1}{\omega}|\frac{\mathrm{d}}{\mathrm{d}t} \ln h^{(m)}(t)| = \frac{1}{\omega} \frac{|h^{(m+1)}(t)|}{|h^{(m)}(t)|} \ll 1$, then the summation in the first line of Eq. (S28) can be convergent quickly and the residual terms can be ignored. In our situation here when calculating $\tilde{\alpha}^{[1']}(t)$, we need to assume $\omega_r/\omega_c \ll 1$ since $\tilde{\alpha}^{[0]}(t)$ and $\Lambda_1^{[0]}(t)$ contains frequency components related to ω_r. In fact, as mentioned in former sections we always assume that $\kappa \ll \chi\alpha \ll \dot{\phi}_d \sim \chi\alpha^2 \ll \omega_r$ to guarantee high-fidelity operations, which means the assumption of $\omega_r/\omega_c \ll 1$ implies $\kappa/\omega_c \ll 1$ and $\chi\alpha^2/\omega_c \ll 1$. Later on, we simply ignore contributions from the factor $e^{-\kappa(t-t')/2}$ since we want to work in the regime that $\kappa t \leq \kappa T \ll 1$ to achieve high fidelity. Also, it varies slowly over time, so it will not affect any scaling analysis when taking time derivatives of this.

With the trick mentioned above, we can write $\tilde{\alpha}^{[1']}$ in the following form:

$$\tilde{\alpha}^{[1']} = \tilde{\alpha}^{[1',0]} + \tilde{\alpha}^{[1',1]}e^{2i\omega_c t} + \tilde{\alpha}^{[1',2]}e^{4i\omega_c t} + \tilde{\alpha}^{[1',-1]}e^{-2i\omega_c t}. \tag{S29}$$

We first analyze $\tilde{\alpha}^{[1',1]}$ since $\Lambda_1^{*[0]}$ dominates in $g(t)$. For the lowest order term ($n = 0$) in Eq. (S28), notice that $\Lambda_1^{[0]}$ contains $i\dot{\tilde{\alpha}}^{[0]}$ term where $\tilde{\alpha}^{[0]}(t) = \alpha(t)f(t)$, we have the leading contribution in $\tilde{\alpha}^{[1',1]}$ as $\tilde{\alpha}^{[1',1]} \simeq \frac{i\alpha^* \dot{f}^*}{2\omega_c}e^{2i\phi_d^{[0]}} \sim O(\frac{\alpha\omega_r}{\omega_c})$. It also contains $O(\frac{\chi\alpha^3}{\omega_c})$ terms that come from those $O(\chi\alpha^3)$ parts in $g(t)$ and those implicitly in $\Lambda_1^{[0]}$. Also, since we assumed $\alpha(t)$ varies slowly that $\dot{\alpha}(t) \sim O(\alpha/T) \sim O(\chi\alpha^2)$, it gives a correction to the estimation of $\tilde{\alpha}^{[1',1]}$ with $O(\frac{\dot{\alpha}f}{\omega_c}) \sim O(\frac{\chi\alpha^2}{\omega_c})$. Besides, the second order term ($n = 1$) in Eq. (S28) can give a contribution of $O(\frac{\alpha\ddot{f}}{\omega_c^2}) \sim O(\frac{\alpha\omega_r^2}{\omega_c^2})$. Using similar argument, we can show that $\tilde{\alpha}^{[1',2]} \sim O(\frac{\chi\alpha^3}{\omega_c})$ and $\tilde{\alpha}^{[1',-1]} \sim O(\frac{\chi\alpha^3}{\omega_c})$. The analysis for $\tilde{\alpha}^{[1',0]}$ is a little different. As mentioned in the main text, in practice we start with $\alpha(0) = 0$ while $\dot{\alpha} \sim O(\alpha/T) \sim O(\chi\alpha^2)$. Therefore, we have $\Lambda_1^{[0]}(0) \sim O(\chi\alpha^2)$ so that $\tilde{\alpha}^{[1',0]} \sim O(\frac{\chi\alpha^2}{\omega_c})$.

Other terms from $g(t)$ will only cause higher order corrections.

Later we want to show how we can partially compensate for the effects of those fast-oscillating terms and even with this technique we still need further assumptions that

$$\frac{\chi\alpha^3}{\omega_c} \ll 1, \quad \frac{\alpha\omega_r^2}{\omega_c^2} \ll 1, \tag{S30}$$

so that the residual effects could be small.

We focus on the change of the Hamiltonian under the new displaced frame with $\tilde{\alpha} = \tilde{\alpha}^{[0]} + \tilde{\alpha}^{[1']}$. We want first to check those extra terms (in comparison with $\hat{H}'_{\mathrm{dr}}[\tilde{\alpha}(t)]$) which are induced by $\tilde{\alpha}^{[1']}$ and do not contain fast oscillating factor (factor $e^{ik\omega_c t}$ where $|k| \in \mathbb{N}_+$). We try to show the dominant effect from each term in Eq. (S24), and evaluate their contributions under assumptions Eq. (S30).

For the corrections with cubic $\hat{a}$ or $\hat{a}^\dagger$ terms (including $\hat{a}^{\dagger 3}$, $\hat{a}^{\dagger 2}\hat{a}$ and their Hermitian conjugate), their coefficients will be of $O(\chi\tilde{\alpha}^{[1']})$, whose amplitudes scale at most $O(\chi\alpha\frac{\omega_r}{\omega_c})$. Even without the fast-oscillating factor, the integration of their amplitudes over time will give us $\chi\alpha T\frac{\omega_r}{\omega_c} \sim \frac{\omega_r}{\omega_c} \ll 1$ (where we have used $T \sim \frac{1}{\chi\alpha}$), which means the contribution from those correction terms to the dynamics are small.

For the correction induced by slowly varied (in comparison with ω_c) $\hat{a}^\dagger\hat{a}$ terms, it can come from either $2\chi|\tilde{\alpha}|^2\hat{a}^\dagger\hat{a}$ or $(\chi\tilde{\alpha}^{*2}e^{2i\phi_d^{[0]}}e^{2i\omega_c t} + \text{c.c.})\hat{a}^\dagger\hat{a}$, where c.c. stands for "complex conjugate". The former will lead extra slowly varied $\hat{a}^\dagger\hat{a}$ terms with coefficients like $(2\chi\tilde{\alpha}^{[0]}\tilde{\alpha}^{*[1',0]} + \text{c.c.}) \sim O(\frac{1}{T}\frac{\chi\alpha^2}{\omega_c})$ and $2\chi|\tilde{\alpha}^{[1',1]}|^2 \sim O(\frac{1}{T}\frac{\alpha\omega_r^2}{\omega_c^2})$, while the coefficient of the dominant correction from the later will be $(2\chi\tilde{\alpha}^{*[0]}\tilde{\alpha}^{*[1',1]}e^{2i\phi_d^{[0]}} + \text{c.c.}) \sim O(\chi\alpha^2\frac{\omega_r}{\omega_c})$. We introduce $\phi_d^{[1]}(t)$ function and make its time derivative $\dot{\phi}_d^{[1]}(t)$ to be equal to the summation of all the slowly varied coefficients of the correction terms, as

$$\dot{\phi}_d^{[1]}(t) = 2\chi\left\{\sum_{k=-1}^2 |\tilde{\alpha}^{[1',k]}|^2 + (\tilde{\alpha}^{[0]}\tilde{\alpha}^{*[1',0]} + \text{c.c.}) + \{[(\tilde{\alpha}^{*[0]} + \tilde{\alpha}^{*[1',0]})\tilde{\alpha}^{*[1',1]} + \tilde{\alpha}^{[1',2]}\tilde{\alpha}^{*[1',-1]}]e^{2i\phi_d^{[0]}} + \text{c.c.}\right\}. \tag{S31}$$

Apparently we still have $\dot{\phi}_d^{[1]} \sim O(\chi\alpha^2\frac{\omega_r}{\omega_c})$. This term may not be ignored and later we will show that we can compensate for that by changing the frequency $\omega_1(t)$ of the linear drive.

Then we discuss the corrections for slowly varied (in comparison with ω_c) $\hat{a}^{\dagger 2}$ terms which come from $\frac{\chi}{2}(\tilde{\alpha} + \tilde{\alpha}^*e^{2i\phi_d^{[0]}}e^{2i\omega_c t})^2\hat{a}^{\dagger 2}$. Those corrections are

$$\frac{\chi}{2}\{(2\tilde{\alpha}^{[0]} + \tilde{\alpha}^{[1',0]} + \tilde{\alpha}^{*[1',1]}e^{2i\phi_d^{[0]}})(\tilde{\alpha}^{[1',0]} + \tilde{\alpha}^{*[1',1]}e^{2i\phi_d^{[0]}})$$
$$+ 2[\tilde{\alpha}^{*[1',1]} + (\tilde{\alpha}^{*[0]} + \tilde{\alpha}^{*[1',0]})e^{2i\phi_d^{[0]}}](\tilde{\alpha}^{*[1',-1]} + \tilde{\alpha}^{*[1',2]}e^{2i\phi_d^{[0]}})\}\hat{a}^{\dagger 2}. \tag{S32}$$

The dominant coefficient of the correction will be $\chi\tilde{\alpha}^{[0]}\tilde{\alpha}^{*[1',1]}e^{2i\phi_d^{[0]}} \sim O(\frac{\chi\alpha^2\omega_r}{\omega_c})$. Although the amplitude of this term might be large, we notice that $\tilde{\alpha}^{[1',1]} \simeq \frac{i\alpha^*\dot{f}^*}{2\omega_c}e^{2i\phi_d^{[0]}}$ with corrections of the order $O(\frac{\chi\alpha^2}{\omega_c})$ and $O(\frac{\alpha\omega_r^2}{\omega_c^2})$. As a result, $\chi\tilde{\alpha}^{[0]}\tilde{\alpha}^{*[1',1]}e^{2i\phi_d^{[0]}} \simeq -i\frac{\chi\alpha^2 f\dot{f}}{2\omega_c}$. We notice that $\frac{f\dot{f}}{\omega_c}$ is a periodic function with period $\delta T = \frac{2\pi}{\omega_r}$. By averaging within each period, we have $\overline{f\dot{f}} = \int_0^{\delta T} f\dot{f}\,dt = 0$. So, $f\dot{f}$ is a linear superposition of fast oscillating $e^{ik\omega_r t}$ functions with nonzero integer k. We can now use the effective Hamiltonian argument we brought up previously. The commutator between this and the 0-th order $[\chi(\tilde{\alpha}^{[0]})^2\hat{a}^{\dagger 2}+\text{h.c.}]$ will give an effective term with amplitude scaled as $\frac{1}{\omega_r}(\frac{\chi\alpha^2\omega_r}{\omega_c})(\chi\alpha^2) \sim \frac{\chi\alpha^3}{\omega_c}\frac{1}{T}$,

which is negligible after integration over time T under our assumptions Eq. (S30). However, this trick does not apply to higher order correction terms of $\tilde{\alpha}^{*[1',1]}$ like $-\frac{\alpha\ddot{f}}{4\omega_c^2} \sim O(\frac{\alpha\omega_r^2}{\omega_c^2})$, since it may lead to slowly varied corrections (compared with ω_r) in $\chi\tilde{\alpha}^{[0]}\tilde{\alpha}^{*[1',1]}e^{2i\phi_d^{[0]}}$. But with the second assumption in Eq. (S30), after integration over time we still have $\chi\alpha\frac{\alpha\omega_r^2}{\omega_c^2}T \ll 1$, which means the contribution from the higher order correction of $\tilde{\alpha}^{*[1',1]}$ can still be ignored. We can similarly argue that other terms in Eq. (S32) can be ignored since the amplitudes of their coefficients are also too small.

Now we start the procedure of partial compensation. We adjust the 1-photon driving frequency such that $\phi_d = \phi_d^{[0]} + \phi_d^{[1]}$. Therefore, we choose a new correction $\tilde{\alpha}^{[1]}$ for the displaced frame which satisfies

$$i\dot{\tilde{\alpha}}^{[1]} + i\kappa\tilde{\alpha}^{[1]}/2 = [\Lambda_1^{*[0]} + \chi\tilde{\alpha}^{*[0]}(|\tilde{\alpha}^{[0]}|^2 + 1)]e^{2i[\phi_d^{[0]}(t)+\phi_d^{[1]}(t)]}e^{2i\omega_c t}$$
$$+ \frac{\chi}{3}(\tilde{\alpha}^{*[0]})^3 e^{4i[\phi_d^{[0]}(t)+\phi_d^{[1]}(t)]}e^{4i\omega_c t} + \frac{\chi}{3}(\tilde{\alpha}^{[0]})^3 e^{-2i[\phi_d^{[0]}(t)+\phi_d^{[1]}(t)]}e^{-2i\omega_c t}. \tag{S33}$$

The only difference between here and Eq. (S26) is that we use $e^{2ki[\phi_d^{[0]}(t)+\phi_d^{[1]}(t)]}$ instead of $e^{2ki\phi_d^{[0]}(t)}$. We denote

$$\tilde{\alpha}^{[1]} = \tilde{\alpha}^{[1,0]} + \tilde{\alpha}^{[1,1]}e^{2i\omega_c t} + \tilde{\alpha}^{[1,2]}e^{4i\omega_c t} + \tilde{\alpha}^{[1,-1]}e^{-2i\omega_c t}. \tag{S34}$$

Then, from Eq. (S28) we can find that

$$\tilde{\alpha}^{[1,1]} = \tilde{\alpha}^{[1',1]}e^{2i\phi_d^{[1]}(t)} + O(\frac{\Lambda_1}{\omega_c}\frac{\dot{\phi}_d^{[1]}}{\omega_c})$$
$$= \tilde{\alpha}^{[1',1]}e^{2i\phi_d^{[1]}(t)} + O(\frac{\chi\alpha^3\omega_r^2}{\omega_c^3}). \tag{S35}$$

Similarly, we can also find that the difference between $\tilde{\alpha}^{[1,k]}$ and $\tilde{\alpha}^{[1',k]}e^{2ik\phi_d^{[1]}(t)}$ (for $k = -1, 0, 2$) are

even smaller compared with $O(\frac{\chi\alpha^3\omega_r^2}{\omega_c^3})$ in the scaling aspect. The difference between $\tilde{\alpha}^{[1]}$ and $\tilde{\alpha}^{[1']}$ can lead to extra non-fast-oscillating (compared with ω_c) terms. For example, the extra coefficient induced in this way for $\hat{a}^\dagger\hat{a}$ will be at most the same scaling as $[2\chi\tilde{\alpha}^{*[0]}(\tilde{\alpha}^{*[1,1]}e^{2i(\phi_d^{[0]}+\phi_d^{[1]})} - \tilde{\alpha}^{*[1',1]}e^{2i\phi_d^{[0]}}) + \text{c.c.}] \sim O(\frac{\chi\alpha^3\omega_r^2}{T\omega_c^3})$, which will be far less that 1 and therefore can be ignored after integration over time. We can use the same way to argue the extra effect from non-fast-oscillating correction of $\hat{a}^{\dagger 2}$ term can be ignored, due to the similarity in the structure of $\tilde{\alpha}^{[1]}$ and $\tilde{\alpha}^{[1']}$.

The coefficient of slowly varied correction (in comparison with ω_c) for $\hat{a}^\dagger$ term will be at most $O(\chi\alpha^2\tilde{\alpha}^{[1,1]}) \sim O(\chi\alpha^3\frac{\omega_r}{\omega_c})$. To deal with it, we introduce a correction for $\Lambda_1 = \Lambda_1^{[0]} + \Lambda_1^{[1]}$, where

$$
\begin{aligned}
\Lambda_1^{[1]} = &-\chi[(|\tilde{\alpha}^{[0]} + \tilde{\alpha}^{[1,0]}|^2 + |\tilde{\alpha}^{[1,1]}|^2 + |\tilde{\alpha}^{[1,2]}|^2 + |\tilde{\alpha}^{[1,-1]}|^2)(\tilde{\alpha}^{[0]} + \tilde{\alpha}^{[1,0]} + \tilde{\alpha}^{*[1,1]}e^{2i\phi_d}) - |\tilde{\alpha}^{[0]}|^2\tilde{\alpha}^{[0]}] \\
&-\chi[\tilde{\alpha}^{[1,1]}(\tilde{\alpha}^{*[0]} + \tilde{\alpha}^{*[1,0]}) + (\tilde{\alpha}^{[0]} + \tilde{\alpha}^{[1,0]})\tilde{\alpha}^{*[1,-1]} + \tilde{\alpha}^{[1,2]}\tilde{\alpha}^{*[1,1]}](\tilde{\alpha}^{[1,-1]} + \tilde{\alpha}^{*[1,2]}e^{2i\phi_d}) \\
&-\chi[\tilde{\alpha}^{*[1,1]}(\tilde{\alpha}^{[0]} + \tilde{\alpha}^{[1,0]}) + (\tilde{\alpha}^{*[0]} + \tilde{\alpha}^{*[1,0]})\tilde{\alpha}^{[1,-1]} + \tilde{\alpha}^{*[1,2]}\tilde{\alpha}^{[1,1]}][\tilde{\alpha}^{[1,1]} + (\tilde{\alpha}^{*[0]} + \tilde{\alpha}^{*[1,0]})e^{2i\phi_d}] \\
&-\chi[(\tilde{\alpha}^{[0]} + \tilde{\alpha}^{[1,0]})\tilde{\alpha}^{*[1,2]} + \tilde{\alpha}^{[1,-1]}\tilde{\alpha}^{*[1,1]}](\tilde{\alpha}^{[1,2]} + \tilde{\alpha}^{*[1,-1]}e^{2i\phi_d}) + \chi\tilde{\alpha}^{*[1,1]}e^{2i\phi_d} \\
&+[\dot{\phi}_d^{[0]}\tilde{\alpha}^{[1,0]} + \dot{\phi}_d^{[1]}(\tilde{\alpha}^{[0]} + \tilde{\alpha}^{[1,0]})] \\
&-\chi[\tilde{\alpha}^{*[1,2]}(\tilde{\alpha}^{*[0]} + \tilde{\alpha}^{*[1,0]})^2 + 2\tilde{\alpha}^{*[1,2]}\tilde{\alpha}^{*[1,1]}\tilde{\alpha}^{*[1,-1]} + (\tilde{\alpha}^{*[1,1]})^2(\tilde{\alpha}^{*[0]} + \tilde{\alpha}^{*[1,0]})]e^{4i\phi_d} \\
&-\chi[2\tilde{\alpha}^{[1,-1]}(\tilde{\alpha}^{[0]} + \tilde{\alpha}^{[1,0]})\tilde{\alpha}^{[1,2]} + \tilde{\alpha}^{[1,-1]}(\tilde{\alpha}^{[1,1]})^2 + (\tilde{\alpha}^{[0]} + \tilde{\alpha}^{[1,0]})^2\tilde{\alpha}^{[1,1]}]e^{-2i\phi_d}.
\end{aligned}
\tag{S36}
$$

Here $\Lambda_1^{[1]} \sim O(\chi\alpha^3\frac{\omega_r}{\omega_c})$ is chosen to fully absorb those non-fast-oscillating corrections. It is still worth to mention that after this change there will be extra $\Lambda_1^{*[1]}e^{2i\phi_d(t)}e^{2i\omega_c t}$ and some other fast-oscillating terms with amplitude scaled at most $O(\chi\alpha^3\frac{\omega_r}{\omega_c})$ in the new Hamiltonian, since all others with larger amplitudes have been absorbed due to Eq. (S33).

Further, we can introduce a correction $\tilde{\alpha}^{[2]} \sim O(\chi\alpha^3\frac{\omega_r}{\omega_c^2}) \ll 1$ and repeat all the steps we did (including

$\phi_d^{[2]}$ and $\Lambda_1^{[2]}$ compensation) so that the contribution from $\tilde{\alpha}^{[2]}$ are too small and we can ignore this. After k rounds of all those corrections, terms $\hat{a}^\dagger$ with coefficients oscillating in the speed of ω_c will be at most $O[\alpha\omega_r(\frac{\chi\alpha^2}{\omega_c})^k]$, which will be small compared with $\frac{1}{T}$ for $k \geq 3$ under the assumptions in Eq. (S30).

Finally we talk about the effects from those fast-oscillating terms (under ω_c) of $\hat{a}^{\dagger 2}$, $\hat{a}^\dagger\hat{a}$ and $\hat{a}^2$ in Eq. (S24), whose coefficients are at most $O(\chi\alpha^2)$. We denote the summation of them as $\hat{H}_{\mathrm{NR},2}(t)$, which can be written as

$$
\begin{aligned}
\hat{H}_{\mathrm{NR},2}(t) = &\left\{\left[\frac{\chi}{2}[(2|\tilde{\alpha}^{[0]}|^2 + 1)e^{2i\phi_d}e^{2i\omega_c t} + (\tilde{\alpha}^{*[0]})^2 e^{4i\phi_d}e^{4i\omega_c t}] + O(\frac{\chi\alpha^2\omega_r}{\omega_c}e^{2ik\omega_c t})\right]\hat{a}^{\dagger 2} + \mathrm{h.c.}\right\} \\
&+\left[[\chi(\tilde{\alpha}^{*[0]})^2 e^{2i\phi_d}e^{2i\omega_c t} + \mathrm{c.c.}] + O(\frac{\chi\alpha^2\omega_r}{\omega_c}e^{2ik\omega_c t})\right]\hat{a}^\dagger\hat{a}.
\end{aligned}
\tag{S37}
$$

We denote $\hat{U}_{\mathrm{NR},2}(t) := \mathcal{T}\exp\left(-i\int_0^t \hat{H}_{\mathrm{NR},2}(t')\,\mathrm{d}t'\right)$, then for any Hamiltonian $\hat{H}$ we can perform a frame transformation to get a new one $\hat{H}' = \hat{U}_{\mathrm{NR},2}^\dagger\hat{H}\hat{U}_{\mathrm{NR},2} + i\dot{\hat{U}}_{\mathrm{NR},2}^\dagger\hat{U}_{\mathrm{NR},2} = \hat{U}_{\mathrm{NR},2}^\dagger(\hat{H} - \hat{H}_{\mathrm{NR},2})\hat{U}_{\mathrm{NR},2}$. We can choose

$\hat{H}$ as the Hamiltonian after going to the displaced rotating frame with all rounds of corrections. To achieve $\hat{H}'$, we can replace $\hat{a}$ with $\hat{a}(t) := \hat{U}_{\mathrm{NR},2}^\dagger(t)\hat{a}\hat{U}_{\mathrm{NR},2}(t)$ and $\hat{a}^\dagger$ with $\hat{a}^\dagger(t) := \hat{U}_{\mathrm{NR},2}^\dagger(t)\hat{a}^\dagger\hat{U}_{\mathrm{NR},2}(t)$ in $(\hat{H} - \hat{H}_{\mathrm{NR},2})$ (also if we consider loss we need to change $\mathcal{D}[\hat{a}]$ into $\mathcal{D}[\hat{a}(t)]$). From the definition of $\hat{a}(t)$ and $\hat{a}^\dagger(t)$, we have

$$
\begin{aligned}
\frac{\mathrm{d}\hat{a}(t)}{\mathrm{d}t} = &(-i)[\chi(2|\tilde{\alpha}^{[0]}|^2 + 1)e^{2i\phi_d}e^{2i\omega_c t} + \chi(\tilde{\alpha}^{*[0]})^2 e^{4i\phi_d}e^{4i\omega_c t} + O(\frac{\chi\alpha^2\omega_r}{\omega_c}e^{2ik\omega_c t})]\hat{a}^\dagger(t) \\
&-i\left[[\chi(\tilde{\alpha}^{*[0]})^2 e^{2i\phi_d}e^{2i\omega_c t} + \mathrm{c.c.}] + O(\frac{\chi\alpha^2\omega_r}{\omega_c}e^{2ik\omega_c t})\right]\hat{a}(t),
\end{aligned}
\tag{S38}
$$

and also $\dot{\hat{a}}^\dagger(t) = [\dot{\hat{a}}(t)]^\dagger$ is also equal to a linear superposition of $\hat{a}^\dagger(t)$ and $\hat{a}(t)$. From scaling analysis we know that $\frac{(\chi\alpha^2)^2}{\omega_c}T \sim \frac{\chi\alpha^3}{\omega_c} \ll 1$, which allows us to use Dyson

series expansion to solve Eq. (S38) and only keep the lower order outcomes. Together with Eq. (S28), we can achieve the solution for $\hat{a}(t)$ and $\hat{a}^\dagger(t)$ approximately with dominant terms in the following:

$$\hat{a}(t) \simeq \{1 - \frac{\chi}{2\omega_c}[(\tilde{\alpha}^{*[0]})^2 e^{2i\phi_d}e^{2i\omega_c t} - \text{c.c.}] + \frac{i\chi^2}{\omega_c}\int_0^t [2(|\tilde{\alpha}^{[0]}(t')|^2 + \frac{1}{2})^2 + \frac{|\tilde{\alpha}^{[0]}(t')|^4}{4} + O(\frac{\alpha^4\omega_r}{\omega_c})]\,dt'\}\hat{a} +$$
$$\{ -[\frac{\chi(2|\tilde{\alpha}^{[0]}|^2 + 1)}{2\omega_c}e^{2i\phi_d}e^{2i\omega_c t} + \frac{\chi(\tilde{\alpha}^{*[0]})^2}{4\omega_c}e^{4i\phi_d}e^{4i\omega_c t}] \tag{S39}$$
$$+ \frac{i\chi^2}{\omega_c}\int_0^t \{[\tilde{\alpha}^{[0]}(t')]^2(2|\tilde{\alpha}^{[0]}(t')|^2 + 1) + O(\frac{\alpha^4\omega_r}{\omega_c})\}\,dt'\}\hat{a}^\dagger.$$

We can see that the difference between $\hat{a}(t)$ and $\hat{a}$ is linear in $\hat{a}$ and $\hat{a}^\dagger$ whose coefficients are at most $O(\frac{\chi^2\alpha^4 T}{\omega_c})$, which are far less than 1. Let us focus on the terms without $e^{ik\omega_c t}$ factors first. Notice that the leading-order terms come from the time integration of some polynomials of $\tilde{\alpha}^{[0]}$ and $\tilde{\alpha}^{*[0]}$ with at most fourth order, which is denoted as $p(\tilde{\alpha}^{[0]}, \tilde{\alpha}^{*[0]})$ later in general. Also $\tilde{\alpha}^{[0]}(t) = \alpha(t)f(t)$ where $f(t)$ is a periodical function that can be written as a Fourier series with basis $e^{ik\omega_r t}$ ($k \in \mathbb{Z}$). Due to the time integration in $\frac{\chi^2}{\omega_c}\int_0^t p(\tilde{\alpha}^{[0]}, \tilde{\alpha}^{*[0]})\,dt'$, only the coefficient of $e^{ik\omega_r t}$ with $k = 0$ scales as $O(\frac{\chi^2\alpha^4 T}{\omega_c}) \sim O(\frac{\chi\alpha^3}{\omega_c})$, while terms that oscillate as $e^{ik\omega_r t}$ (with nonzero k) only have amplitudes that scale as at most $O(\frac{\chi^2\alpha^4}{\omega_r\omega_c})$. Therefore, for the corrections that come from $\frac{\chi}{2}(\tilde{\alpha}^{[0]})^2([\hat{a}^\dagger(t)]^2 - \hat{a}^{\dagger 2})$, they contain terms with amplitude $O(\chi\alpha^2 \cdot \frac{\chi\alpha^3}{\omega_c}) \ll O(\chi\alpha^2)$ but oscillate in the form of $f^2(t)$, whose effect is ignorable as we argued at the beginning of the section. They also contain slowly varied part with coefficients at most $O(\chi\alpha^2 \cdot \frac{\chi^2\alpha^4}{\omega_c\omega_r}) \sim O(\frac{\chi\alpha^2}{\omega_r} \cdot \frac{\chi\alpha^3}{\omega_c} \cdot \frac{1}{T})$, which are again ignorable after integration over time T. We can similarly demonstrate that the corrections coming from $\chi\tilde{\alpha}^{[0]}\tilde{\alpha}^{*[1,1]}e^{2i\phi_d}[\hat{a}^\dagger(t)]^2$ and other high-order terms are negligible.

Besides, after going to the reference frame with $\hat{U}_{\text{NR},2}$, the fast-oscillating (under ω_c) $\hat{a}^{\dagger 2}$ or $\hat{a}^\dagger\hat{a}$ terms left have amplitudes at most $O(\frac{\chi^2\alpha^4}{\omega_c}) \sim \frac{\chi\alpha^3}{\omega_c} \cdot \frac{1}{T})$, which are sufficiently small. Finally, since we work in a small-dimensional blockade subspace with only several excitations as assumed, as well as $\frac{\chi^2\alpha^4 T}{\omega_c} \ll 1$, the corrections due to the frame transformation $\hat{U}_{\text{NR},2}^\dagger(T)$ on the final states are negligible. This concludes all the evidence that under assumptions in Eq. (S30), the dynamics affected by terms beyond RWA in Eq. (S24) could be ignorable, in the aspect of scaling analysis.

C. Both charge and flux drives

Things will be different if we can drive both charge and flux simultaneously. In this case, our 1-photon driving

term can be implemented as

$$\text{Re}[\Lambda_1(t)e^{-i\omega_1(t)t}](\hat{a} + \hat{a}^\dagger)$$
$$+ \text{Re}[i\Lambda_1(t)e^{-i\omega_1(t)t}] \cdot i(\hat{a} - \hat{a}^\dagger) \tag{S40}$$
$$= \Lambda_1(t)e^{-i\omega_1(t)t}\hat{a}^\dagger + \Lambda_1^*(t)e^{i\omega_1(t)t}\hat{a}.$$

In this way, our leading non-RWA term in Eq. (S23) is gone. As a result, the improved hardware controllability can also help to mitigate the non-RWA effects.

D. Rough lower bound on operation infidelity

We want to point out that the requirement for $\frac{\chi\alpha^3}{\omega_c} \ll 1$ is actually necessary since without the assumption the $\hat{H}_{\text{NR},2}$ term will affect the dynamics a lot. Even if we can directly implement 2-photon drive, or we can do both charge and flux drives, the requirement is still there due to the difference between the original Kerr term $\chi(\hat{a} + \hat{a}^\dagger)^4/12$ and the simplified one $\chi\hat{a}^{\dagger 2}\hat{a}^2/2$ after RWA. We can use the $\frac{\chi\alpha^3}{\omega_c} \ll 1$ condition to derive a lower bound on the infidelity of Fock $|1\rangle$ state preparation using constant α. Since $\alpha < (\omega_c/\chi)^{1/3}$, we have

$$T > \frac{\pi}{2(\omega_c\chi^2)^{1/3}}. \tag{S41}$$

So, even in the case that power is not a constraint, and we are allowed to design κ_e as small as possible, we still have

$$\epsilon > c_1\kappa_i T > \frac{3\pi(\kappa_i/\chi)}{16(\omega_c/\chi)^{1/3}} = \frac{3\pi(\kappa_i/\chi)^{2/3}}{16Q_i^{1/3}}, \tag{S42}$$

where we have already put $c_1 = \frac{3}{8}$ in it, and $Q_i = \omega_c/\kappa_i$ is the internal quality factor. We denote $\epsilon_{\text{min}} := \frac{3\pi(\kappa_i/\chi)^{2/3}}{16Q_i^{1/3}}$, which serves as a lower bound of the infidelity that can be achieved.

S6. ROUGH LOWER BOUND FOR INPUT POWER NEEDED

Here we consider some extreme cases to provide a rough lower bound of the input power needed for our protocol. We assume that we could directly implement 2-photon drives so that the infidelity only comes from

photon loss that $\epsilon = c_1(\kappa_e + \kappa_i)T$. Besides, we further assume that the 2-photon drives can be implemented in some power-efficient manners such that we only need to consider the power consumed by 1-photon drives. However, from Eq. (3) we still have $\Lambda_1 \simeq 2\chi\alpha^3$, and therefore the input power we need scales as $P_{\text{in}} \propto \alpha^6$. To prepare $|1\rangle$ state with constant α, the power-dependent infidelity satisfies the following equation

$$\epsilon = c_1 \frac{\pi}{2} \frac{\kappa_e + \kappa_i}{\kappa_e^{1/6}} \left(\frac{4\hbar\omega_c}{P_{\text{in}}\chi^4} \right)^{1/6}. \qquad (S43)$$

We can choose $\frac{\kappa_e}{\kappa_i} = \frac{1}{5}$ to further minimize ϵ when κ_i is fixed, and therefore achieve the error scaling $\epsilon \propto \kappa_i^{5/6}/(P_{\text{in}}^{1/6}\chi^{2/3})$. Equivalently, the rough lower bound of the input power P_{in} for a given infidelity ϵ satisfies

$$P_{\text{in}} = \frac{4(3\pi c_1)^6}{5^5}\hbar\omega_c \frac{\kappa_i^5}{\chi^4\epsilon^6}. \qquad (S44)$$

S7. FEASIBILITY ON ACTUAL PLATFORMS

In this section, we try to consider parameters from different kinds of experimental platforms and check if it is feasible to use our protocol to achieve some high-fidelity operations on those devices. For simplicity we still consider the Fock $|1\rangle$ state preparation task. We will check the requirements for RWA and the input power needed in order to achieve a certain fidelity derived under RWA. It is worth mentioning that, although our RWA requirements are derived under assumptions that $\alpha(t)$ is slowly varied with $\alpha(0) = \alpha(T) = 0$ and $f(t)$ itself is analytic, in this section we still focus on the method with constant α and $f(t)$ defined in Eq. (S9) to calculate relevant parameters. More physical $\alpha(t)$ and $f(t)$ functions will lead to a change of those c_i coefficients instead of the scaling properties, so we can still use those ideal choices to get a taste of the feasibility of our protocol on different platforms.

A. Optical ring resonators

In this part, we talk about the possibility of implementing our protocol on optical ring resonators with $\chi^{(3)}$ nonlinearity. We first estimate the self-Kerr χ value with experimentally achievable platforms, and then check the RWA requirements by calculating ϵ_{min}, which is a rough lower bound for infidelity achieved by perturbative analysis together with parts of the RWA requirements (See Sec. S5 D).

The self-Kerr parameter χ in optical ring resonators with $\chi^{(3)}$ nonlinearity can be calculated by [S7]:

$$\chi = -\frac{\hbar\omega_c^2 cn_2}{2n^2 A_{\text{eff}} L}, \qquad (S45)$$

where ω_c is the frequency of the mode, n and n_2 are the refractive index and nonlinear refractive index correspondingly, and A_{eff} and L are effective area and length of the waveguide. In the system proposed in Ref. [S8], the authors there assumed $\frac{\omega_c cn_2}{n^2 A_{\text{eff}} L} = 3.7 \times 10^{20}$ W^{-1}s^{-2} without pointing out which kind of materials they use nor those geometrical parameters of the resonators. Besides, the authors hope to achieve an internal loss rate $\kappa_i = 2\pi \times 25$ MHz, and they mentioned that this assumption is reasonable since the corresponding intrinsic quality factor $Q_i \sim 8 \times 10^6$ is achievable with the silicon-nitride platform. This discussion indicates that they have chosen $\omega_c \approx 2\pi \times 200$ THz. With those parameters, the corresponding $\chi = -2\pi \times 3.90$ Hz. However, when putting these numbers into Eq. (S42), we can find out that $\epsilon_{\text{min}} = 102 \gg 1$, which indicates that our perturbative analysis should fall and RWA will be significantly violated if we want to use our current protocol to prepare a Fock $|1\rangle$ state.

A similar result is achieved using parameters from another literature [S9]. In their simulation, the parameters are chosen as $\gamma_{\text{NL}} := \frac{n_2\omega_c}{cA_{\text{eff}}} = 1$ W^{-1}s^{-2}, together with $\omega_c = 2\pi \times 193$ THz, $n = 1.7$, and $L = 400$ μm. With these numbers, we can find that $\chi = -2\pi \times 0.79$ Hz. The intrinsic Q-factor is chosen as 2×10^6, which corresponds to $\kappa_i = 2\pi \times 96.5$ MHz. This gives $\epsilon_{\text{min}} = 1151$, which again indicates that our protocol does not work.

B. Photonic crystals

Photonic crystals may perform better to increase Kerr nonlinearities with smart design by suppression of effective mode volume. Our discussion in this part builds on the work from Ref. [S6], where the authors discussed the χ and κ values that can be possibly achieved using their novel design of nanocavities with ultra-small mode volumes. It is claimed to be promising for achieving a "single-photon Kerr nonlinearity" ($\chi \sim \kappa$) regime. In that work, χ is determined by

$$\chi = -\frac{3\chi^{(3)}\hbar\omega_c^2}{2\epsilon_0 n^4} \frac{V_M}{V_{\text{eff}}^2}, \qquad (S46)$$

where V_{eff} is the effective mode volume, and V_M is another parameter related to cavity design with the unit of volume.

The authors there provided two types of cavity design: tip design and bridge design. With the tip cavity design there, in simulations people could achieve $\frac{QV_M}{V_{\text{eff}}^2} \approx 2 \times 10^7 \lambda^{-3}$ and quality factor $Q \approx 10^6$. The λ here is the wavelength of the mode. Besides, in that work, those authors assumed that the cavity radiation loss is much larger than material loss so that the same quality factor applies to cavities with the same design but different materials. Here we simply follow this assumption to derive κ_i.

For organic materials like J aggregate (PIC), it has $|\chi^{(3)}|/n^4 \sim 1.1 \times 10^{-15}$ m^2/V^2 at $\lambda = 575$ nm, which

TABLE I. A summary of power needed to achieve 90% fidelity of Fock state $|1\rangle$ with our protocol in photonic crystals designed in [S6], as well as the check of the self-consistency requirements for RWA.

Cavity design	Nonlinear materials	$\omega_c/(2\pi)$	$|\chi|/(2\pi)$	$\kappa_i/(2\pi)$	$\epsilon_{\min}$	$P_{\text{in},0.9}$	$(\omega_r/\omega_c)_{0.9}$	$(\alpha\omega_r^2/\omega_c^2)_{0.9}$	$(\chi\alpha^3/\omega_c)_{0.9}$
Tip	PIC	521 THz	3.5 THz	0.52 GHz	$\chi \gg \kappa_i$				
	ITO	255 THz	1.9 GHz	0.26 GHz	1.5×10^{-3}	1.3×10^{-5} W	5.8×10^{-5}	3.3×10^{-9}	7.1×10^{-6}
Bridge	GaAs	283 THz	2.7 MHz	0.19 GHz	0.088	4.3×10^8 W	0.46	112	1.3
	Ge	94.6 THz	9.9 kHz	63 MHz	1.76	$\epsilon_{\min} > 1$			

gives $\chi = -2\pi \times 3.5$ THz and $\kappa_i = 2\pi \times 0.52$ GHz. In this case χ is several orders of magnitude larger than κ_i, so there will be no need to use our protocol to achieve universal control.

Inorganic materials like indium tin oxide (ITO) has $|\chi^{(3)}|/n^4 \sim 2.12 \times 10^{-17} \text{m}^2/\text{V}^2$ at $\lambda = 1175$ nm. It gives $\chi = -2\pi \times 1.9$ GHz with $\kappa_i = 2\pi \times 0.26$ GHz. The parameters still lie in the strong Kerr-nonlinearity regime as $|\chi|/\kappa_i > 1$, and we only need input power $P_{\text{in}} = 1.3 \times 10^{-5}$ W to achieve 90% fidelity of a single photon state with our protocol.

Similarly, for their bridge cavity design they have $\frac{QV_M}{V_{\text{eff}}^2} \approx 5.5 \times 10^7 \lambda^{-3}$ and quality factor $Q \approx 1.5 \times 10^6$, which in general requires a small $|\chi^{(3)}|/n^4$ to achieve the same Kerr nonlinearity. But, as they mentioned, for usual semiconductor materials like gallium arsenide (GaAs) with $|\chi^{(3)}|/n^4 \sim 0.97 \times 10^{-20} \text{m}^2/\text{V}^2$ at $\lambda = 1.06$ μm, we can get $\chi = -2\pi \times 2.7$ MHz and $\kappa_i = 2\pi \times 0.19$ GHz. In this case, if we want to achieve 90% fidelity at the optimal point we derived in Sec. S2 under RWA, in fact we can see that $\alpha\omega_r^2/\omega_c^2 = 112 \gg 1$, which means the RWA is actually violated. If people can get a better device with κ_i 10 times smaller (so that $\kappa_i = 2\pi \times 19$ MHz and $\kappa_i/\chi = 7.0$) while other parameters do not change, it is promising to pass the RWA requirement

with $\omega_r/\omega_c = 1.5 \times 10^{-3}$, $\chi\alpha^3/\omega_c = 1.3 \times 10^{-3}$ and $\alpha\omega_r^2/\omega_c^2 = 1.1 \times 10^{-4}$ if we want to achieve 90% fidelity. The power needed will be $P = 4.3 \times 10^2$ W.

For bridge cavity design using Germanium (Ge) as nonlinear materials with $|\chi^{(3)}|/n^4 \sim 0.86 \times 10^{-20} \text{m}^2/\text{V}^2$ at $\lambda = 3.17$ μm, we have $\chi = -2\pi \times 9.9$ kHz and $\kappa_i = 2\pi \times 63$ MHz. However, we find that $\epsilon_{\min} = 1.76 > 1$, which rules out the possibility of using our protocol due to the violation of RWA.

S8. DERIVATION OF CONTINUOUS VERSION FORMULAE

In this section, we will use the Magnus expansion method to derive some "continuous version" formulae including Eq. (12) and (S7).

We first use Magnus expansion to compute $\hat{U} = \mathcal{T}\exp\left[-i\int_0^{\delta T} \hat{H}(t)\,dt\right]$ up to fourth order explicitly with $\hat{H}(t) = \hat{H}(\delta T - t)$. Notice that

$$\hat{U} = \exp\left\{\sum_{m=1}^4 (-i)^m \hat{\Omega}_m + O[(\delta T)^5]\right\}, \tag{S47}$$

where

$$\hat{\Omega}_1 = \int_0^{\delta T} \hat{H}_1\, dt_1, \quad \hat{\Omega}_2 = \frac{1}{2}\int d^2t\, [\hat{H}_1, \hat{H}_2], \quad \hat{\Omega}_3 = \frac{1}{6}\int d^3t\, ([\hat{H}_1, [\hat{H}_2, \hat{H}_3]] + [\hat{H}_3, [\hat{H}_2, \hat{H}_1]]),$$

$$\hat{\Omega}_4 = \frac{1}{12}\int d^4t\, ([[[\hat{H}_1, \hat{H}_2], \hat{H}_3], \hat{H}_4] + [\hat{H}_1, [[\hat{H}_2, \hat{H}_3], \hat{H}_4]] + [\hat{H}_1, [\hat{H}_2, [\hat{H}_3, \hat{H}_4]]] + [\hat{H}_2, [\hat{H}_3, [\hat{H}_4, \hat{H}_1]]]). \tag{S48}$$

Here we use abbreviations for $\hat{H}_k := \hat{H}(t_k)$ and

$$\int d^k t := \int_{0 \leq t_k \leq t_{k-1} \leq \cdots \leq t_1 \leq \delta T} \cdots \int dt_1 \ldots dt_k. \tag{S49}$$

By making use of the property $\hat{H}(t) = \hat{H}(\delta T - t)$, we have the following:

$$\hat{\Omega}_2 = \frac{1}{2}\int d^2t\, [\hat{H}_2, \hat{H}_1] = -\hat{\Omega}_2 = 0 \tag{S50}$$

$$\int d^4t\, [[[\hat{H}_1, \hat{H}_2], \hat{H}_3], \hat{H}_4] = \int d^4t\, [[[\hat{H}_4, \hat{H}_3], \hat{H}_2], \hat{H}_1] = -\int d^4t\, [\hat{H}_1, [\hat{H}_2, [\hat{H}_3, \hat{H}_4]]] \tag{S51}$$

Then, from Jacobi identity we have

$$
\begin{aligned}
\int \mathrm{d}^4 t \ [\hat{H}_1, [[\hat{H}_2, \hat{H}_3], \hat{H}_4]] &= \int \mathrm{d}^4 t \ -[\hat{H}_4, [\hat{H}_1, [\hat{H}_2, \hat{H}_3]]] - [[\hat{H}_2, \hat{H}_3], [\hat{H}_4, \hat{H}_1]] \\
&= \int \mathrm{d}^4 t \ -[\hat{H}_1, [\hat{H}_4, [\hat{H}_3, \hat{H}_2]]] - [[\hat{H}_2, \hat{H}_3], [\hat{H}_4, \hat{H}_1]] \\
&= \int \mathrm{d}^4 t \ -[\hat{H}_1, [[\hat{H}_2, \hat{H}_3], \hat{H}_4]] - [[\hat{H}_2, \hat{H}_3], [\hat{H}_4, \hat{H}_1]] \\
&= -\frac{1}{2} \int \mathrm{d}^4 t \ [[\hat{H}_2, \hat{H}_3], [\hat{H}_4, \hat{H}_1]].
\end{aligned} \tag{S52}
$$

Similarly, we have

$$
\int \mathrm{d}^4 t \ [\hat{H}_2, [\hat{H}_3, [\hat{H}_4, \hat{H}_1]]] = -\frac{1}{2} \int \mathrm{d}^4 t \ [[\hat{H}_4, \hat{H}_1], [\hat{H}_2, \hat{H}_3]], \tag{S53}
$$

which will lead to $\hat{\Omega}_4 = 0$.

In summary, we have

$$
\begin{aligned}
\hat{U} &= \exp\left\{-i\hat{\Omega}_1 + i\hat{\Omega}_3 + O[(\delta T)^5]\right\}, \\
\hat{U}_{\mathrm{tar}} &= \exp\left[-i \int_0^{\delta T} \hat{H}(t)\, \mathrm{d}t\right] = \exp\left[-i\hat{\Omega}_1\right].
\end{aligned} \tag{S54}
$$

Since $\hat{\Omega}_3$ is in the order of $O[(\delta T)^3]$, it is easy to see that

$$
\hat{U} - \hat{U}_{\mathrm{tar}} = O[(\delta T)^3], \tag{S55}
$$

which gives the proof of Eq. (12).

Similarly, with the Baker–Campbell–Hausdorff (BCH) formula, we can find that

$$
\hat{U}_{\mathrm{tar}}^{-1/2} \hat{U} \hat{U}_{\mathrm{tar}}^{-1/2} = \exp\left\{i\hat{\Omega}_3 + O[(\delta T)^5]\right\}, \tag{S56}
$$

which gives the proof of the "continuous version" correspondence of Eq. (S7).

S9. ESTIMATION OF UNDETERMINED FACTORS IN EQ. (S13)

In this section, we will show how we estimate the coefficients c_1, c_2 and c_3 in Eq. (S13) when focusing on a specific task described in Sec. S2, which is to prepare Fock state $|1\rangle$ with constant α and oscillation function $f(t)$ defined in Eq. (S9).

A. Estimation of c_1

In the estimation of c_1, we can assume that the photon blockade is perfectly achieved with ideal Hamiltonian $\hat{H}_{\mathrm{dr}}^{\mathrm{qb}}$ in Eq. (S6), so we only need to take care of the infidelity induced by photon loss. We can start with the

following Lindblad equation with operators restricted in the two-level blockade subspace:

$$
\frac{\mathrm{d}\hat{\rho}}{\mathrm{d}t} = -i[-\chi\alpha\hat{\sigma}_x, \hat{\rho}] + \kappa\mathcal{D}[\hat{\sigma}_-]\hat{\rho}. \tag{S57}
$$

The state is initialized as $\hat{\rho}(0) = |0\rangle\langle 0|$. To prepare the $|1\rangle$ state, we require the evolution to last for $T = \frac{\pi}{2\chi\alpha}$. Since we work in the regime that $\kappa \ll \chi\alpha$, we will treat the loss as a perturbation. We first go to the interaction picture, and then denote $\hat{U}_0(t) = e^{i\chi\alpha t\hat{\sigma}_x}$ as well as $\hat{\rho}_I(t) = \hat{U}_0^\dagger(t)\hat{\rho}(t)\hat{U}_0(t)$. Therefore, the evolution of $\hat{\rho}_I(t)$ satisfies

$$
\frac{\mathrm{d}\hat{\rho}_I(t)}{\mathrm{d}t} = \kappa\mathcal{D}[\hat{U}_0^\dagger(t)\hat{\sigma}_-\hat{U}_0(t)]\hat{\rho}_I(t). \tag{S58}
$$

The final state $\hat{\rho}_I(T)$ up to first order in κ is

$$
\hat{\rho}_I(T) = \hat{\rho}_I(0) + \kappa \int_0^T \mathcal{D}[\hat{U}_0^\dagger(t)\hat{\sigma}_-\hat{U}_0(t)]\hat{\rho}_I(0)\, \mathrm{d}t. \tag{S59}
$$

Recall our definition for infidelity ϵ in Eq. (S4), we have

$$
\begin{aligned}
\epsilon &= 1 - \langle 1| \hat{\rho}(T) |1\rangle = 1 - \langle 0| \hat{\rho}_I(T) |0\rangle \\
&= -\kappa \langle 0| \left[\int_0^T \mathcal{D}[\hat{U}_0^\dagger(t)\hat{\sigma}_-\hat{U}_0(t)]\hat{\rho}_I(0)\, \mathrm{d}t\right] |0\rangle \\
&= \frac{3}{8}\kappa T,
\end{aligned} \tag{S60}
$$

where $T = \frac{\pi}{2\chi\alpha}$. This result directly indicates that $c_1 = \frac{3}{8}$ in the protocol we considered.

B. Estimation of c_2

Notice that c_2 is introduced as a factor in the expression of Trotter error $\epsilon_{\mathrm{tt}} = c_2/[M^4(\chi T)^6]$, so we can numerically estimate it by calculating the infidelity of the
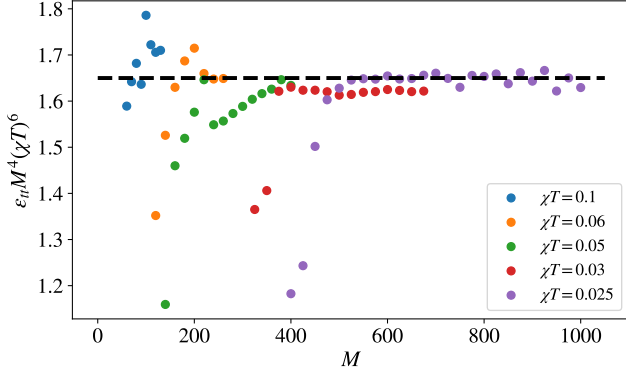
FIG. S3. Determine $c_2 = \epsilon_{\mathrm{tt}} M^4 (\chi T)^6$ numerically by varying M and T. The black dashed line is the reference for $c_2 = 1.65$.

state preparation under Hamiltonian $\hat{H}'_{\mathrm{dr}}[\tilde{\alpha}(t)]$ evolution while varying M and T. We show the corresponding plot in FIG. S3 where the data points on it are chosen to satisfy $\epsilon_{\mathrm{tt}} < 0.2$, and find out that in the small χT (to reduce infidelity coming from photon loss) but large M (to reduce ϵ_{tt}) regime we have $c_2 \approx 1.65$. Since we mainly care about the scaling behavior in our protocol instead of exact numbers, we will just use $c_2 \approx 1.65$ as a rough estimation when it is needed.

C. Estimation of c_3

The derivation of c_3 is straightforward, since we just need to put the exact form of $\tilde{\alpha}(t) = \alpha f(t)$ into Eq. (9) and then calculating the power. As we mentioned in Sec. S2 that $i\dot{\tilde{\alpha}}$ is the dominant term in Λ_1, and due to our specific design of $f(t)$ in Eq. (S9), we have

$$|\Lambda_1| \simeq |\dot{\tilde{\alpha}}(t)| = |\alpha \dot{f}(t)| = \frac{\pi^2 M \alpha}{T} = \frac{\pi^3 M}{2\chi T^2}, \qquad (S61)$$

in which we have used $\alpha = \frac{\pi}{2\chi T}$. Therefore, we can finally achieve

$$P_{\mathrm{in}} = \frac{|\Lambda_1|^2}{\kappa_e} \hbar \omega_c = \frac{\pi^6}{4} \hbar \omega_c \frac{M^2}{\kappa_e \chi^2 T^4}, \qquad (S62)$$

which directly gives us that $c_3 = \frac{\pi^6}{4} \hbar \omega_c$.

S10. GENERATION OF CONTROL SEQUENCE WITH NEURAL NETWORK

In this part, we will discuss about the way to use neural network to generate control pulses in order to solve the imperfect blockade situation due to the assumption that we do not have the direct 2-photon drive in experiment.

We still first go to the rotating frame and then the displaced frame, without making constraint between $\alpha(t)$ and $\Lambda_1(t)$. Besides, we can choose $\omega_1(t)$ such that there

is no detuning term ($\hat{a}^\dagger \hat{a}$) in the new Hamiltonian. Therefore, we will end up with the following Hamiltonian:

$$
\begin{aligned}
\hat{H}_{\mathrm{dr}}^{\mathrm{new}} = {} & \frac{\chi}{2} \hat{a}^{\dagger 2} \hat{a}^2 + \{\chi \alpha^{\mathrm{new}}(t) \hat{a}^\dagger \hat{n} + \frac{\chi}{2} [\alpha^{\mathrm{new}}(t)]^2 \hat{a}^{\dagger 2} \\
& + \Lambda_1^{\mathrm{new}}(t) \hat{a}^\dagger + \mathrm{h.c.}\}.
\end{aligned}
\tag{S63}
$$

In this case, both $\alpha^{\mathrm{new}}(t)$ and $\Lambda_1^{\mathrm{new}}(t)$ are controllable. We note that, although r does not come into $\hat{H}_{\mathrm{dr}}^{\mathrm{new}}$ explicitly, we still try to achieve the effective photon blockade during evolution by optimizing the fidelity of the unitary operations or state preparation tasks within the subspace spanned by $\{|0\rangle, \ldots, |r\rangle\}$ that we want to block, while adding a penalty function to penalize the average population leakage out of the blockade subspace during evolution. Specifically, we can choose $g_{\mathrm{u}} = \frac{1}{(r+1)T} \int_0^T \mathrm{d}t \, \mathrm{Tr}\left[\hat{U}^\dagger(t)(\hat{I} - \hat{\Pi}_r)\hat{U}(t)\hat{\Pi}_r\right]$ as the penalty function in the unitary operation task, or $g_{\mathrm{st}} = \frac{1}{T} \int_0^T \mathrm{d}t \, \langle\psi(t)| (\hat{I} - \hat{\Pi}_r) |\psi(t)\rangle$ for the state preparation task, where as in the main text $\hat{\Pi}_r = \sum_{n=0}^r |n\rangle\langle n|$ is the projection operator onto the blockade subspace. Here we can understand the dynamics as that, the states that initially lie in the blockade subspace will always evolves within it with small leakage, by keeping track of the evolution with a proper time-dependent displace frame $\alpha^{\mathrm{new}}(t)$. This also provides another advantage that we do not have to take a large dimension cutoff which always associates with numerical computation in bosonic systems, since we just have to focus on the dynamics confined in a small subspace with slight leakage.

To encode the variables $\alpha^{\mathrm{new}}(t)$ and $\Lambda_1^{\mathrm{new}}(t)$ in the optimization problem, we use sinusoidal representation neural network with three layers that takes t as an input and returns the real and imaginary part of the corresponding control parameter by $\Phi^{(t)} = \mathbf{W}_2 \sin\left(\mathbf{W}_1 \sin(\mathbf{W}_0 t + \mathbf{b}_0) + \mathbf{b}_1\right) + \mathbf{b}_2$, where $\mathbf{W}_0 \in \mathbb{R}^{20 \times 1}$, $\mathbf{W}_1 \in \mathbb{R}^{20 \times 20}$, and $\mathbf{W}_2 \in \mathbb{R}^{2 \times 20}$, and $\mathbf{b}_i$ are vectors with a dimension consistent with their corresponding $\mathbf{W}_i$ [S10]. We then maximize $F_{\mathrm{task}}(\boldsymbol{\alpha}) - g(\boldsymbol{\alpha})$ by finding an appropriate set of $\mathbf{W}_i$ and $\mathbf{b}_i$ that specifies the controls, where $F_{\mathrm{task}}(\boldsymbol{\alpha}) = \left|\mathrm{Tr}\left[\hat{U}_{\mathrm{tar}}^\dagger \hat{U}(T)\right]\right|^2 / (r+1)^2$ for unitary operations and $F_{\mathrm{task}}(\boldsymbol{\alpha}) = \left|\langle\psi_{\mathrm{tar}}|\psi(T)\rangle\right|^2$ for state preparation tasks. To do this we first fix the total evolution time χT and perform gradient based optimization until a convergence criteria is satisfied. After finding $\alpha^{\mathrm{new}}(t)$ and $\Lambda_1^{\mathrm{new}}(t)$ at the end of the optimization, we compute the infidelity by taking the photon loss $(\kappa_i + \kappa_e)\mathcal{D}[\hat{a}]$ into account and also the input power needed. We fix κ_i and vary κ_e to find infidelity as a function of the required power.

Note that as opposed to the Trotter based method with the sine-basis ansatz of Eq. (7), we do not impose $\alpha(0) = \alpha(T) = 0$ for the neural network ansatz to allow for more flexibility. These potential non-zero values in the beginning and end of the protocol can be realized by
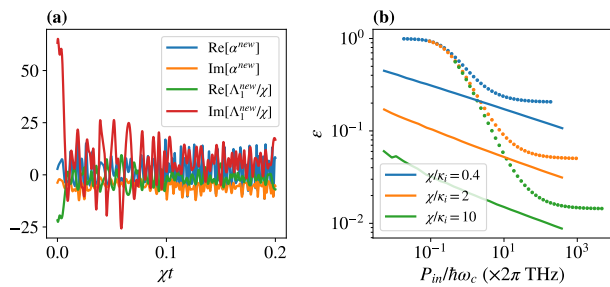
FIG. S4. (a) The outcome of pulse sequences $\alpha^{\text{new}}(t)$ and $\Lambda_1^{\text{new}}(t)$ from the neural network with fixed $\chi T = 0.2$. (b) The comparison of maximum power needed and infidelities of state preparation between pulses from neural network optimization (shown with dots) with fixed total time T and Trotter-based design (shown with solid lines) with optimized T and M.

fast displacements. It is also possible to impose a constraint on these values by adding another penalty term to the cost function.

To compare this method with the Trotter scheme, we consider the single photon Fock $|1\rangle$ state preparation task. Unlike the protocol examined in Sec. S2, where the total time T can be optimized analytically based on χ, κ_i, and the input power allowed, here we fix T without any optimization attempt, and vary κ_i to compare the infidelity as a function of the power for the pulse sequence generated by the neural network with those obtained in Sec. S2 based on analytical intuition. We show an example of the outcome in FIG. S4.

We believe that the inferior performance of the neural network is due to the complexity of finding highly oscillatory functions (as also seen in the Trotter based scheme) required for relaxing the $\Lambda_2(t)$ requirement. Solving ordinary differential equations with highly oscillatory functions is computationally costly and increases the optimization time. Therefore, even though, in principle, a highly expressive neural network is capable of representing the solutions found using the Trotter scheme, it is difficult to find those (or better) solutions directly using our optimization techniques. It is interesting to devise more efficient optimization methods using prior information from the success of the Trotter based scheme. We leave these questions for future work.

[S1] U. Boscain, M. Sigalotti, and D. Sugny, PRX Quantum **2**, 030203 (2021).

[S2] S. Krastanov, V. V. Albert, C. Shen, C.-L. Zou, R. W. Heeres, B. Vlastakis, R. J. Schoelkopf, and L. Jiang, Physical Review A **92**, 040303(R) (2015).

[S3] S. G. Schirmer, H. Fu, and A. I. Solomon, Physical Review A **63**, 063410 (2001).

[S4] A. Lingenfelter, D. Roberts, and A. A. Clerk, Science Advances **7**, eabj1916 (2021).

[S5] M. Suzuki, Journal of Mathematical Physics **26**, 601 (1985).

[S6] H. Choi, M. Heuck, and D. Englund, Physical Review Letters **118**, 223605 (2017).

[S7] U. B. Hoff, B. M. Nielsen, and U. L. Andersen, Optics Express **23**, 12013 (2015).

[S8] S. Mittal, G. Moille, K. Srinivasan, Y. K. Chembo, and M. Hafezi, Nature Physics **17**, 1169 (2021).

[S9] Z. Vernon, N. Quesada, M. Liscidini, B. Morrison, M. Menotti, K. Tan, and J. Sipe, Physical Review Applied **12**, 064024 (2019).

[S10] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, Advances in Neural Information Processing Systems **33**, 7462 (2020).