Mitigating off-resonant error in the cross-resonance gate

Moein Malekakhlagh and Easwar Magesan

Mitigating off-resonant error in the cross-resonance gate

Moein Malekakhlagh[*] and Easwar Magesan[†]

IBM Quantum, IBM Thomas J. Watson Research Center,
1101 Kitchawan Rd, Yorktown Heights, NY 10598

(Dated: December 23, 2021)

Off-resonant error for a driven quantum system refers to interactions due to the input drives having non-zero spectral overlap with unwanted system transitions. For the cross-resonance gate, this includes leakage as well as off-diagonal computational interactions that lead to bit-flip error on the control qubit. In this work, we quantify off-resonant error, with more focus on the less studied off-diagonal control interactions, for a *direct* CNOT gate implementation. Our results are based on numerical simulation of the dynamics, while we demonstrate the connection to time-dependent Schrieffer-Wolff and Magnus perturbation theories. We present two methods for suppressing such error terms. First, pulse parameters need to be optimized so that off-resonant transition frequencies coincide with the local minima due to the pulse spectrum sidebands. Second, we show the advantage of a Y-DRAG pulse on the control qubit in mitigating off-resonant error. Depending on qubit-qubit detuning, the proposed methods can improve the average off-resonant error from approximately 10^{-3} closer to the 10^{-4} level for a direct CNOT calibration.

I. INTRODUCTION

Cross-resonance (CR) is a microwave-activated two-qubit gate performed by driving one of the qubits (control) at the frequency of the other (target) [1, 2]. In this architecture, superconducting qubits [3–6], typically fixed-frequency transmons [5], connect via a mediating bus resonator, resulting in a static qubit-qubit exchange interaction. The CR protocol induces various two-qubit interactions [7–13], with ZX as the dominant term, through which a CNOT gate can be calibrated. Simplicity in implementation, resilience to charge and flux noise, and scalability has made CR architecture the leading workhorse for current IBM quantum processors [12, 14, 15].

Improving CR gate fidelity necessitates both an accurate understanding of the effective interactions as well as precise microwave control. In particular, a multi-level analysis of the dynamics [7–12] is required for CR gate implementation with weakly anharmonic transmon qubits [5]. Higher qubit states can lead to both renormalization of interactions in the computational subspace [8, 11, 12] as well as leakage out of the computational subspace [10, 16]. Generally, to optimize the coherent fidelity, we need to maximize the desired ZX term and minimize unwanted computational and leakage interactions.

There are two main CNOT calibration schemes based on the CR architecture. The first is an *echo* sequence consisting of two CR pulses with flipped amplitude accompanied with single-qubit rotations [7, 11, 12, 15]. The echo removes certain unwanted terms such as the ZI, ZZ and IX, however, it induces higher-order IY and IZ error terms [11, 12]. This technique has been used to demonstrate a 280 ns echoed CR gate with an average fidelity of 99.40% [15]. The second main CNOT calibration is *direct*, with a single CR pulse [10, 15, 17]. In this calibration, we want no operation on the target qubit when the control is in state $|0_c\rangle$, hence canceling $IX + ZX$ via a separate drive on the target, and a π rotation on the target when the control is in state $|1_c\rangle$. Therefore, the IX term is not an error term anymore, and CR gate speed is determined by twice the ZX (IX) term, resulting in a faster gate. Equipped with multiple-path interference couplers [18] to suppress the static ZZ term, and using a virtual frame change [19] to cancel out Stark shifts in software, Ref. [17] demonstrated a 180 ns gate with 99.77% average gate fidelity.

Schrieffer-Wolff Perturbation Theory (SWPT) [8, 11, 20–27] is a central method in our analytical understanding of effective interactions. SWPT provides effective models by averaging high-frequency off-resonant processes systematically. Through a series of perturbative frame transformations, the interactions are partitioned into resonant (effective) and off-resonant categories. The effective interactions come from processes that connect states with equal frequency in the rotating frame of the drive, while off-resonant interactions have a net non-zero transition frequency. In contrast to the rotating-wave approximation (RWA), which simply discards off-resonant terms, SWPT takes them into account by solving for and storing the relevant frame transformations. Hence, contributions that are not resonant at a specific order may lead to resonant interactions via non-trivial higher-order mixings. The drive scheme, and the corresponding energy diagram, determines the choice of the effective frame. For CR, where the drive is resonant with the target qubit, the effective SWPT frame is block-diagonal (BD) with respect to the control qubit [8]. Under the BD approximation, analytical estimates for CR gate parameters have been derived in Refs. [8, 11] for a constant-amplitude continuous wave (CW) drive.

In this paper, we study off-resonant error due to inter-

---

actions that originate from the CR drive frequency being detuned from states of the control qubit. The corresponding dominant unwanted transitions are $|0_c\rangle \leftrightarrow |1_c\rangle$, $|0_c\rangle \leftrightarrow |2_c\rangle$ and $|1_c\rangle \leftrightarrow |2_c\rangle$ (see Fig. 1). Although off-resonant error is present for a CW drive, there is an intricate interplay with the pulse shape and in particular the pulse ramps. To model this, we employ numerical simulations based on Magnus expansion [28–31]. Furthermore, we extend the SWPT formalism for CR [8, 11] to the time-dependent case and make a connection to the Magnus method. In particular, the main role of time-independent and time-dependent perturbations are to account for *how strong* the drive amplitude and *how fast* (non-adiabatic) the pulse ramps are compared to the system transition frequencies, respectively. These two effects are independent in general, however, in the context of gate calibration, they become related based on a fixed rotation angle imposed by the intended gate.

Generally, to mitigate off-resonant error, a CR drive should have minimal spectral content at the unwanted transition frequencies. Qubit-qubit detuning and anharmonicity determine the relative configuration of off-resonant transitions in the rotating-frame of the drive [10, 11], where the error due to one or sometimes multiple transitions can be noticeable. We take the standard square Gaussian pulse, i.e. flat top with Gaussian ramps, and demonstrate further improvement. The most immediate refinement comes from optimization of the pulse rise time known as Gaussian shaping [23, 32]. The pulse ramps should be optimized so that the transition for the most dominant error type overlaps with one of the local minima of the pulse sidebands. Moreover, we show additional improvement by a Y-DRAG [23, 32–34] pulse on the control qubit. We argue that, in essence, DRAG acts as an effective filter that can be tuned to notch the error due to specific off-resonant transitions.

The remainder of this paper is organized as follows. In Sec. II, we introduce our model for a direct CNOT gate implementation [10, 17]. In Sec. III, we discuss the theory behind off-resonant error, by relating time-dependent SWPT to the Magnus method, which extends our earlier results in Refs. [8, 11]. Furthermore, using numerical simulation based on Magnus, we quantify the dependence of off-resonant error on pulse parameters. In Sec. IV, we demonstrate the advantage of Y-DRAG pulse on the control qubit in suppressing off-resonant error.

There are five appendices. In Appendix A, we discuss the derivation of a generalized time-dependent SWPT to account for the underlying pulse shapes. Using the SWPT formalism, Appendix B derives effective time-dependent Hamiltonian terms generalizing Refs. [8, 10, 11]. Appendices C and D provide the effective time evolution operator and the corresponding leading-order non-BD contributions, respectively. In Appendix E, we derive leading-order estimates for dominant off-resonant error types, and corresponding DRAG conditions for their suppression.
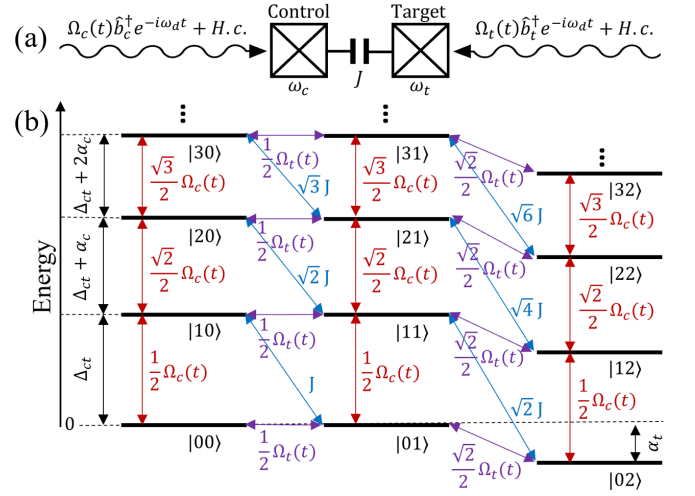

FIG. 1. (a) CR gate schematic with separate drives on each qubit. The drive amplitudes $\Omega_j(t) \equiv \Omega_{jx}(t) + i\Omega_{jy}(t)$, $j = c, t$ are complex-valued to allow for more involved control schemes like DRAG [23, 32–34]. The drive frequency is set to the target qubit frequency. (b) Energy diagram in the rotating frame of the drive under RWA (see Sec. II). The control-target detuning is denoted by $\Delta_{ct} \equiv \omega_c - \omega_t$, and control and target anharmonicities by α_c and α_t, respectively.

II. DIRECT CNOT

We consider two coupled transmon qubits [5] with individual drives on each qubit. The transmon Hamiltonian can be approximated in terms of multi-level Kerr oscillators as

$$\hat{\mathcal{H}}_q \equiv \sum_{j=c,t} \left[\omega_j \hat{b}_j^\dagger \hat{b}_j + \frac{\alpha_j}{2} \hat{b}_j^\dagger \hat{b}_j^\dagger \hat{b}_j \hat{b}_j \right] , \qquad (1)$$

with $\omega_{c,t}$ and $\alpha_{c,t}$ denoting the corresponding frequency and anharmonicity for the control and the target qubits, respectively. The transmon-transmon exchange interaction takes the approximate form

$$\hat{\mathcal{H}}_J \equiv J \left(\hat{b}_c^\dagger \hat{b}_t + \hat{b}_c \hat{b}_t^\dagger \right) , \qquad (2)$$

where J is the effective exchange coupling as a result of either direct capacitive coupling or mediated coupling through a common bus resonator [8, 11]. Furthermore, we consider separate drives on each qubit:

$$\hat{\mathcal{H}}_d(t) \equiv \frac{1}{2} \left[\Omega_c^*(t)\hat{b}_c e^{i\omega_d t} + \Omega_c(t)\hat{b}_c^\dagger e^{-i\omega_d t} \right]$$
$$+ \frac{1}{2} \left[\Omega_t^*(t)\hat{b}_t e^{i\omega_d t} + \Omega_t(t)\hat{b}_t^\dagger e^{-i\omega_d t} \right], \qquad (3)$$

with frequency ω_d, same as the target qubit frequency, and time-dependent complex-valued pulse amplitudes $\Omega_c(t) \equiv \Omega_{cx}(t) + i\Omega_{cy}(t)$ and $\Omega_t(t) \equiv \Omega_{tx}(t) + i\Omega_{ty}(t)$ (see Fig. 1). The phase of the microwave drive determines the axis, in the X–Y plane of the target qubit, to which the drive couples. We set the main CR drive to

couple to the X quadrature of the target, while the Y axis can be used for DRAG following the same conventions as single-qubit gates [23, 33, 34]. In our numerical and perturbative analysis, the system Hamiltonian is the static part $\hat{\mathcal{H}}_s \equiv \hat{\mathcal{H}}_q + \hat{\mathcal{H}}_J$, accounting for the CR eigenstate renormalization due to the exchange interaction (2), and the interaction-frame Hamiltonian is defined as $\hat{\mathcal{H}}_I(t) \equiv \exp(i\hat{\mathcal{H}}_s t)\hat{\mathcal{H}}_d(t)\exp(-i\hat{\mathcal{H}}_s t)$.

In Eqs. (1)–(3), we have adopted a Kerr model for the qubits, and applied RWA in both drive and exchange interactions. This RWA neglects terms that oscillate at approximately twice the target qubit frequency, of the order of 10 GHz, and does not change the physics of the CR gate qualitatively. In this work, we focus on the dominant error that comes from off-resonant transitions of the control qubit, with transition frequencies of the order of 100 MHz, and study the dependence on pulse parameters. A more precise model for the CR gate was discussed in Ref. [11], accounting also for eigenstate renormalization due to counter-rotating terms in the Josephson nonlinearity.

Tuning a direct CNOT gate [17] requires identity operation on the target qubit when the control is in state $|0_c\rangle$ so that

$$\int_0^{\tau_p} dt\,[\omega_{ix}(t) + \omega_{zx}(t)] = 0 , \qquad (4)$$

and a π rotation around the X axis when the control is in state $|1_c\rangle$ as

$$\int_0^{\tau_p} dt\,[\omega_{ix}(t) - \omega_{zx}(t)] = \pi . \qquad (5)$$

In particular, in our implementation, the cancellation in Eq. (4) is restricted to all times, i.e. $\omega_{ix}(t) + \omega_{zx}(t) \approx 0$, via the direct resonant drive $\Omega_t(t)$ on the target qubit. Here, $\omega_{\sigma_j \sigma_k}(t) \equiv (1/2)\mathrm{Tr}\{\hat{\mathcal{H}}_{\mathrm{CR,eff}}(t)\,(\hat{\sigma}_j \otimes \hat{\sigma}_k)\}$, $j,k \in \{i,x,y,z\}$, are effective Hamiltonian terms defined over the dressed two-qubit Pauli operators. For notation simplicity, we use $\hat{I} \equiv \hat{\sigma}_i$, $\hat{X} \equiv \hat{\sigma}_x$, $\hat{Y} \equiv \hat{\sigma}_y$, $\hat{Z} \equiv \hat{\sigma}_z$ and drop the explicit tensor product for two-qubit Pauli operators, e.g. $\hat{Z}\hat{X} \equiv \hat{\sigma}_z \otimes \hat{\sigma}_x$. Time-dependent SWPT provides reasonable estimates for the effective gate parameters (see Appendices A and B). In particular, based on Eq. (4), the main cancellation tone on the target is found as

$$\Omega_{tx}(t) = \frac{J}{\Delta_{ct}}\Omega_{cx}(t) - \frac{\alpha_c J}{2\Delta_{ct}^3(2\Delta_{ct} + \alpha_c)}\Omega_{cx}^3(t) + O\left(\Omega_{cx}^5\right), \qquad (6)$$

while for CNOT calibration [Eq. (5)] the main pulse should satisfy approximately

$$\frac{2\alpha_c J}{\Delta_{ct}(\Delta_{ct} + \alpha_c)}\Omega_{cx}[\tau_p - s_1(\tau_r)]$$
$$- \frac{(9\Delta_{ct}^3 + 15\Delta_{ct}^2\alpha_c + 11\Delta_{ct}\alpha_c^2 + 3\alpha_c^3)\alpha_c^2 J}{\Delta_{ct}^3(\Delta_{ct} + \alpha_c)^3(2\Delta_{ct} + \alpha_c)(2\Delta_{ct} + 3\alpha_c)} \qquad (7)$$
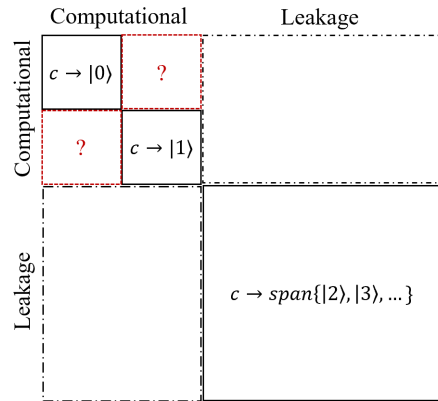$$\times \Omega_{cx}^3[\tau_p - s_3(\tau_r)] + O\left(\Omega_{cx}^5\right) = \pi .$$

FIG. 2. Schematic of the BD subspace shown in solid black, subspace corresponding to coupling between computational and leakage in dashed-dotted black, and non-BD computational subspace in red dotted lines. We are particularly interested in characterizing the error due to non-BD interactions of the form $\hat{\sigma}_x \otimes \hat{\sigma}_k$ or $\hat{\sigma}_y \otimes \hat{\sigma}_k$ for $k = i,x,y,z$.

Equations (6)–(7) provide a useful initial guess for drive parameters and facilitate more involved numerical optimization. We also find non-adiabatic corrections to the effective gate parameters in terms of the pulse derivatives, e.g. terms proportional to $J\dot{\Omega}_{cx}^2(t)\Omega_{cx}(t)$ in $\omega_{ix}(t)$ and $\omega_{zx}(t)$ (see Appendix B). Here, τ_p and τ_r denote the gate and the rise times, respectively. Moreover, $s_n(\tau_r)$ characterizes the reduced area under the curve during the ramps, compared to a square pulse, for the nth power of the pulse shape. In our simulations, we use the common square Gaussian pulse [15, 17].

Coherent error, compared to an ideal CNOT, can be traced back to the two aforementioned categories of interactions. Resonant (effective) error arises from faulty resonant rotations, terms like IX (IY) and ZX (ZY), or static and dynamic frequency shifts of the qubits, like IZ, ZI and ZZ terms. Static ZZ can in principle be substantially suppressed via tunable [35–38], opposite-anharmonicity [39, 40] and multiple-path [17, 18, 41–44] couplers, and also via auxiliary AC Stark tones (siZZle) on the qubits [45, 46]. Furthermore, Stark shifts can be removed effectively through a virtual frame change in software [17, 19]. In the following, we explore the physics of off-resonant error.

III. OFF-RESONANT ERROR

Generally, a non-zero overlap of the drive spectrum with an unwanted off-resonant system transition can cause error. For a direct CNOT calibration, the dominant off-resonant error types are bit-flip and leakage on the control qubit [10] (see Fig. 2). Unlike resonant error, which can be approximately modeled via time-independent methods accounting only for constant amplitude drive [8, 10, 11], off-resonant error exhibits sub-

Probability	Type	Time-domain	Frequency-domain
$\lvert 0_c\rangle \to \lvert 1_c\rangle$	(1) non-BD	$\approx \frac{1}{4}\left\lvert \int_0^{\tau_p} dt'\,\Omega_c(t')e^{i\Delta_{ct}t'}\right\rvert^2$	$\approx \frac{1}{4}\left\lvert \int_{-\infty}^{+\infty}\frac{d\omega'}{2\pi}\frac{\tilde{\Omega}_c(\omega')}{\omega'+\Delta_{ct}}\left[e^{i(\omega'+\Delta_{ct})\tau_p}-1\right]\right\rvert^2$
$\lvert 0_c\rangle \to \lvert 2_c\rangle$	(2) Leakage	$\approx \frac{1}{32}\left\lvert \int_0^{\tau_p} dt'\int_0^{t'} dt''\,\Omega_c(t')\Omega_c(t'')\right.$ $\left.\left[e^{i(\Delta_{ct}+\alpha_c)t'}e^{i\Delta_{ct}t''}-e^{i\Delta_{ct}t'}e^{i(\Delta_{ct}+\alpha_c)t''}\right]\right\rvert^2$	$\approx \frac{1}{32}\left\lvert \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{d\omega'}{2\pi}\frac{d\omega''}{2\pi}\frac{\alpha_c\tilde{\Omega}_c(\omega')\tilde{\Omega}_c(\omega'')}{(\omega''+\Delta_{ct}+\alpha_c)(\omega''+\Delta_{ct})}\right.$ $\left.\frac{\left[e^{i(\omega'+\omega''+2\Delta_{ct}+\alpha_c)\tau_p}-1\right]}{\omega'+\omega''+2\Delta_{ct}+\alpha_c}\right\rvert^2$
$\lvert 1_c\rangle \to \lvert 2_c\rangle$	(3) Leakage	$\approx \frac{1}{2}\left\lvert \int_0^{\tau_p} dt'\,\Omega_c(t')e^{i(\Delta_{ct}+\alpha_c)t'}\right\rvert^2$	$\approx \frac{1}{2}\left\lvert \int_{-\infty}^{+\infty}\frac{d\omega'}{2\pi}\frac{\tilde{\Omega}_c(\omega')}{\omega'+\Delta_{ct}+\alpha_c}\left[e^{i(\omega'+\Delta_{ct}+\alpha_c]\tau_p}-1\right]\right\rvert^2$

TABLE I. Summary of dominant off-resonant error types and leading-order overlap integrals in time and frequency domains. The estimates are based on a reduced model of an off-resonantly driven transmon qubit equivalent to vertical ladders in Fig. 1 (see also Appendix E). In particular, single-photon transitions are enhanced when there is large overlap with the sideband photons of the pulse, i.e. at $\omega' = -\Delta_{ct}$ and $\omega' = -(\Delta_{ct}+\alpha_c)$ for type 1 and type 3, respectively. The two-photon transition is enhanced due to simultaneous overlap with two sidebands such that $\omega'+\omega'' = -(2\Delta_{ct}+\alpha_c)$. Proximity to frequency collisions at $\Delta_{ct} = 0$, $-\alpha_c/2$, $-\alpha_c$ enhances the corresponding overlap as the transitions are excited by closer-to-center and hence stronger spectrum sidebands. For strong drive amplitude, comparable to the transition frequencies, more precise estimates can be obtained by exponentiation of the Magnus generator, compared to the perturbative expansion used in this table.

tle interplay with the pulse shapes and requires more involved time-dependent methods, such as generalized time-dependent SWPT [11, 23] and Magnus [28–30], discussed in the following.

A. Theory

The time evolution operator for the CR gate is

$$\hat{U}_I(\tau_p,0) \equiv \mathbb{T}\exp\left[-i\int_0^{\tau_p} dt'\hat{\mathcal{H}}_I(t')\right],\qquad(8)$$

where $\hat{\mathcal{H}}_I(t)$ is the drive Hamiltonian (3), expressed in the interaction frame with respect to $\hat{\mathcal{H}}_s$, $\mathbb{T}$ is the time-ordering operator and τ_p is the gate time. One standard approach for computing Eq. (8) is the Magnus method [28–30], which solves perturbatively for the generator of time evolution operator as $\hat{U}_I(\tau_p,0) \equiv \exp[-i\hat{K}(\tau_p,0)]$. Up to the second order one finds

$$\hat{K}_1(\tau_p,0) = \int_0^{\tau_p} dt'\hat{\mathcal{H}}_I(t'),\qquad(9a)$$

$$\hat{K}_2(\tau_p,0) = -\frac{i}{2}\int_0^{\tau_p} dt'\int_0^{t'} dt''[\hat{\mathcal{H}}_I(t'),\hat{\mathcal{H}}_I(t'')].\qquad(9b)$$

The Magnus method computes the two interaction categories, resonant and off-resonant, altogether. Time-dependent SWPT, however, separates the two by computing an effective resonant Hamiltonian as

$$\hat{\mathcal{H}}_{I,\text{eff}}(t) \equiv \hat{U}_{\text{SW}}^\dagger(t)\left[\hat{\mathcal{H}}_I(t)-i\partial_t\right]\hat{U}_{\text{SW}}(t),\qquad(10)$$

using the frame transformation $\hat{U}_{\text{SW}}(t) \equiv \exp[-i\hat{G}(t)]$. Similar to the Magnus expansion, we can solve for the generator $\hat{G}(t)$ and the effective Hamiltonian $\hat{\mathcal{H}}_{I,\text{eff}}(t)$ perturbatively [8, 11, 21, 23–26] (see Appendices A

and B). Therefore, SWPT is a method for implementing systematic RWA: through a perturbative frame transformation we obtain effective models with resonant interaction terms, however, the information about off-resonant processes is stored in $\hat{U}_{\text{SW}}(t)$ and hence we can reconstruct the overall time evolution operator as (see Appendices C and D)

$$\hat{U}_I(\tau_p,0) = \hat{U}_{\text{SW}}(\tau_p)\hat{U}_{I,\text{eff}}(\tau_p,0)\hat{U}_{\text{SW}}^\dagger(0).\qquad(11)$$

Equation (11) is the bridge between time-dependent Magnus and SWPT formalisms.

According to Eq. (11), the overall time evolution is in principle invariant of the frame choice. However, in practice, any perturbative treatment of $\hat{U}_{\text{SW}}(t)$ breaks the invariance. The choice for an efficient frame depends on the drive scheme and the quantities we intend to compute. For CR, to compute an effective Hamiltonian with resonant interactions, the SW frame is BD with respect to the control qubit [8, 11] so as to capture resonant X and Y target rotations (see Fig. 2). Hence, the SW frame transformation $\hat{U}_{\text{SW}}(t)$ encodes the details of off-resonant processes and non-BD interactions in particular.

Generally, analytical modeling of time-dependent error requires precise and hence very involved symbolic computer algebra as discussed in the Appendices. We base our analysis primarily on the numerical simulation, while using perturbation theory to corroborate various trends of off-resonant error. For instance, the leading-order perturbative estimate for the probability amplitude of $\lvert 0_c\rangle \to \lvert 1_c\rangle$, with time-dependent coupling $\Omega_c(t)/2$ and detuning Δ_{ct}, over time interval $[0,\tau_p]$ reads (see Ap-
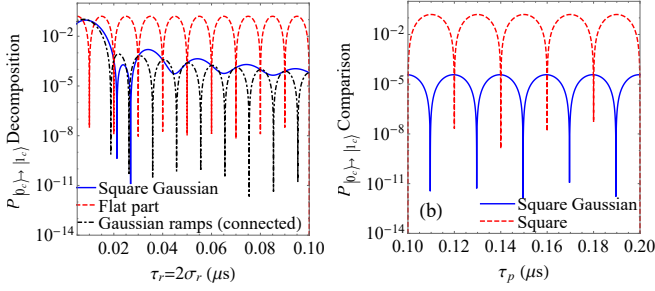
FIG. 3. (a) Decomposition of the perturbative overlap probability based on Eq. (12) into overlap with a square Gaussian pulse in Eq. (13), overlap with just the Gaussian ramps (connected), and overlap with just the flat part as a function of $\tau_r = 2\sigma_r$ for fixed $\Omega/(2\pi) = 20$ MHz, $\Delta_{ct}/(2\pi) = 50$ MHz and $\tau_p = 200$ ns. The distinct overlaps add at the level of complex probability amplitude, hence there is significant interference between separate parts of the pulse. The comparison confirms the crucial role of smooth ramps in reducing the error. (b) Comparison of overlap probability between square Gaussian, for the same parameters as in (a) and fixed $\tau_r = 2\sigma_r = 26$ ns [an optimal value in panel (a)], and a square pulse with the same τ_p as a function of τ_p. The results exhibit a periodic dependence on τ_p with period $2\pi/\Delta_{ct}$.

pendix E)

$$
\begin{aligned}
&- i \int_0^{\tau_p} dt' \frac{\Omega_c(t')}{2} e^{i\Delta_{ct} t'} \\
&= -\int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \frac{\tilde{\Omega}_c(\omega')}{2(\omega' + \Delta_{ct})} \left[e^{i(\omega' + \Delta_{ct})\tau_p} - 1 \right] \\
&= -\left\{ \sum_{n=0}^{\infty} \left[\frac{1}{\Delta_{ct}} \left(\frac{i}{\Delta_{ct}} \frac{d}{dt'} \right)^n \frac{\Omega_c(t')}{2} \right] e^{i\Delta_{ct} t'} \right\}\Bigg|_0^{\tau_p} .
\end{aligned}
$$

(12)

The first line in Eq. (12) is an example of a *time-domain* overlap integral between the pulse and transition frequency Δ_{ct}. The second line shows the *frequency-domain* representation, with pulse Fourier transform $\tilde{\Omega}_c(\omega')$, where a single sideband photon provides the energy to excite the transition. From design perspective, this suggests that mitigating the error requires *filtering* (notching) the pulse spectrum at $\omega' = -\Delta_{ct}$. Lastly, assuming that the pulse is differentiable up to arbitrary orders at the boundaries, the third line shows the spectral overlap in terms of the pulse time derivatives through *adiabatic expansion*, and suggests that DRAG [23, 33] is a natural leading-order solution for engineering spectral content (see Appendix E). Similarly, higher-order expansions in Magnus and SWPT describe probability amplitudes of multi-level off-resonant transitions as processes in which multiple sideband photons provide the total energy for the transition (see the 2nd row in Table I).

It is important to note that off-resonant error is not caused only by the pulse ramps and is present for a constant-amplitude drive. This can be seen from Eq. (12) where a constant pulse with amplitude Ω and duration τ_p results in error probability of $(\Omega/\Delta_{ct})^2 \sin^2(\Delta_{ct}\tau_p/2)$,

which has a period of $2\pi/\Delta_{ct}$ in τ_p. Smoother ramps, however, bring additional non-trivial derivative contributions that modify and in particular mitigate the error. To see this, consider the standard square Gaussian pulse defined as

$$
\Omega_{\mathrm{SG}}(t) \equiv \begin{cases} \Omega \dfrac{e^{-\frac{(t-\tau_r)^2}{2\sigma_r^2}} - e^{-\frac{\tau_r^2}{2\sigma_r^2}}}{1 - e^{-\frac{\tau_r^2}{2\sigma_r^2}}} \, , & 0 < t < \tau_r \\[4mm] \Omega \, , & \tau_r < t < \tau_p - \tau_r \\[4mm] \Omega \dfrac{e^{-\frac{[t-(\tau_p-\tau_r)]^2}{2\sigma_r^2}} - e^{-\frac{\tau_r^2}{2\sigma_r^2}}}{1 - e^{-\frac{\tau_r^2}{2\sigma_r^2}}} \, , & \tau_p - \tau_r < t < \tau_p \end{cases}
$$

(13)

with ramps comprised of a truncated Gaussian with rise time τ_r, standard deviation σ_r and a total flat time of $\tau_p - 2\tau_r$. Based on the first line of Eq. (12), contributions from different parts of the pulse, i.e. the ramps and the flat part, add with complex amplitude and can create constructive/destructive interference. This is shown in Fig. 3, where we see that the overlap probability with square Gaussian lies almost in between the individual overlap probabilities due to just the ramps or just the flat part. In particular, the overlap due to the flat part is orders of magnitude higher than the overall overlap with square Gaussian pulse, demonstrating the benefit of smooth ramps in reducing off-resonant error.

Given that CR gate operates typically in the near-detuned straddling regime ($|\Delta_{ct}| < |\alpha_c|$), the off-resonant error comprises of three dominant types involving the following transitions on the control qubit: (1) $|0_c\rangle \rightarrow |1_c\rangle$ with single-photon transition frequency Δ_{ct}, (2) $|0_c\rangle \rightarrow |2_c\rangle$ with two-photon transition frequency $2\Delta_{ct} + \alpha_c$ and (3) $|1_c\rangle \rightarrow |2_c\rangle$ with single-photon transition frequency $\Delta_{ct} + \alpha_c$ (see Table I). Depending on qubit-qubit detuning, and proximity to the underlying frequency collisions at $\Delta_{ct} = 0, -\alpha_c/2, -\alpha_c$ [10, 11], each error type can become the most dominant. For instance, type 1 (non-BD) error is dominant for qubit pairs with relatively small detuning where $|\Delta_{ct}| \ll |\Delta_{ct} + \alpha_c/2|$ and $|\Delta_{ct}| \ll |\Delta_{ct} + \alpha_c|$. In the following, we analyze these dominant off-resonant error types numerically.

B. Simulation

For the numerical integration, we employ a Magnus solver based on the following discretization of Eqs. (9a)–(9b) [31]:

$$
\hat{U}_I(t_j + h, t_j) = \exp\left[-i \sum_{n=1,2} \hat{K}_n(t_j + h, t_j) \right], \quad (14a)
$$

$$
\hat{K}_1(t_j + h, t_j) \equiv \frac{h}{2} \left[\hat{\mathcal{H}}_I(t_j + c_1 h) + \hat{\mathcal{H}}_I(t_j + c_2 h) \right], \quad (14b)
$$

$$
\hat{K}_2(t_j + h, t_j) \equiv -i\frac{\sqrt{3}h^2}{12} [\hat{\mathcal{H}}_I(t_j + c_2 h), \hat{\mathcal{H}}_I(t_j + c_1 h)], \quad (14c)
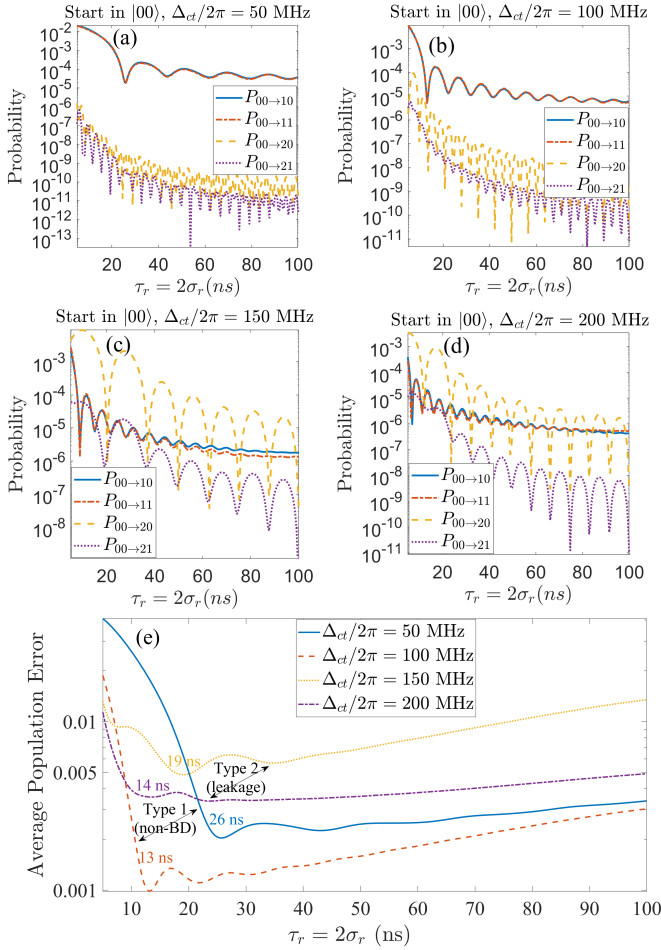$$

FIG. 4. Numerical dependence of off-resonant error on $\tau_r = 2\sigma_r$ based on the square Gaussian pulse of Eq. (13) with fixed gate time $\tau_p = 200$ ns. System parameters are set as $J/(2\pi) = 3.5$, $\omega_t/(2\pi) = 5000$, $\alpha_c/(2\pi) = \alpha_t/(2\pi) = -340$ MHz, while varying control qubit freuqency for detunings $\Delta_{ct}/(2\pi) = 50, 100, 150$ and 200 MHz. The corresponding static ZZ terms are apprximatley 147.2, 157.7, 178.8 and 220.4 KHz, respectively. The drive amplitudes Ω_{cx} and Ω_{tx} are adjusted according to the approximate conditions (6)–(7) to calibrate a CNOT gate. (a)–(d) Decomposition of error into individual transitions for initial state set to $|00\rangle$ as a point of comparison. Legends use the shorthand notation $P_{mn\to pq} \equiv |\langle pq|\hat{U}_I(\tau_p, 0)|mn\rangle|^2$. (e) Average population error in Eq. (15) calculated using the four computational states.

for an infinitesimal time evolution between t_j and $t_j + h$, where h is the time step, and $c_1 \equiv 1/2 + \sqrt{3}/6$ and $c_2 \equiv 1/2 - \sqrt{3}/6$ denote the relative position of the auxiliary midpoints with respect to t_j. The coefficients in Eq. (14a)–(14c) are chosen such that the contribution from $\hat{K}_3(t_j + h, t_j)$ is canceled, hence the numerical error is $O(h^4)$ [31]. One important advantage of a numerical solver based on the matrix exponentiation of the Magnus expansion [Eq. (14a)] is that it preserves the unitarity of the time evolution operator at any arbitrary truncation order.
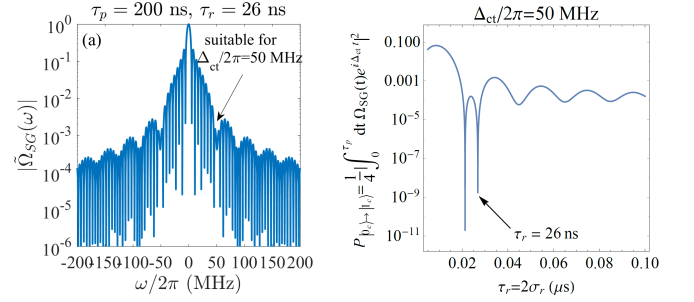

FIG. 5. (a) Normalized Fourier transform of the square Gaussian pulse (13) for $\tau_p = 200$ ns and $\tau_r = 2\sigma_r = 26$ ns. (b) Analytical reconstruction of the $|0_c\rangle \to |1_c\rangle$ error as a function of τ_r based on the first row of Table I. Compared to Fig. 3(a), the drive is adjusted at each τ_r for an approximate CNOT calibration.

For the CR simulation, we keep 5 and 3 levels for the control and the target qubits, respectively. The main CR pulse, applied on the X axis of the control qubit, is taken to be the square Gaussian in Eq. (13). Drive amplitudes on the control and the target qubits should then be calibrated according to Eqs. (4)–(5). Here, we use the perturbative conditions (6)–(7) to expedite the numerical computation.

Figure 4 shows the behavior of off-resonant error as a function of pulse rise time τ_r for sample detunings $\Delta_{ct}/(2\pi) = 50, 100, 150, 200$ MHz, and $\alpha_c/(2\pi) = \alpha_t/(2\pi) = -340$ MHz. First, depending on the closeness to the three collision types in Sec. III A, we observe a crossover where either non-BD error or leakage become dominant [Figs. 4(a)–4(d)]. For instance, for $\Delta_{ct}/(2\pi) = 50$ MHz, the non-BD error is almost four orders of magnitude larger than leakage, due to proximity to a type 1 collision. Second, there are favorable local minima of the off-resonant error, as a function of τ_r, due to overlap of the underlying transition frequency with a dip in the sidebands of the pulse spectrum. In Fig. 4(e), we show average population error defined as

$$\bar{E}_{\text{pop}} \equiv 1 - \frac{1}{4}\sum_{j,k=0,1}\left|\langle j,k|\hat{U}_{\text{CNOT}}^\dagger \hat{U}_I(\tau_p, 0)|j,k\rangle\right|^2. \quad (15)$$

We use this measure intentionally to reflect the behavior of off-resonant error more clearly, instead of the overall average error [47, 48] that contains phase error due to Stark shifts and ZZ. In particular, the average population error follows similar local minima as a function of τ_r that is dominated by one, or a combination of multiple collision types. Based on Fig. 4(e), possible optimal choices of τ_r for $\Delta_{ct}/(2\pi) = 50, 100, 150$ and 200 MHz are approximately 26, 13, 19 and 14 ns, respectively. Moreover, the case of $\Delta_{ct}/(2\pi) = 100$ MHz leads to the smallest off-resonant error due to being comparably furthest from the three collisions.

To connect the simulation results to the analytical overlap integrals of Table I, note that Fourier transform

of square Gaussian consists of an overall tail, with relatively wide sidebands in frequency, whose spectral width is primarily determined by the risetime τ_r, and a series of narrower sidebands with widths determined by interplay between the overall gate time τ_p and flat time $\tau_p - 2\tau_r$ [see Fig. 5(a)]. Increasing τ_r, on the one hand, shrinks the overall spectral width and suppresses the error in general. However, to reach the same CR rotation angle, a stronger drive is needed. Up to the leading order, one- and two-photon off-resonant errors are $O(\Omega_{cx}^2)$ and $O(\Omega_{cx}^4)$, respectively. Therefore, for each parameter set, we expect distinct sweet spots for off-resonant error in terms of τ_r. We took a closer look into the case of $\Delta_{ct}/(2\pi) = 50$ MHz in Fig. 5. First, Fourier transform of Eq. (13) for $\tau_p = 200$ ns and $\tau_r = 26$ ns exhibits a sideband dip at $\omega/(2\pi) = 50$ MHz justifying why this choice is suitable for suppressing non-BD error in Fig. 4(a). Moreover, we used the leading-order analytical estimate for $|0_c\rangle \to |1_c\rangle$ in Table I to reconstruct the dependence on τ_r. We find that the position of local minima agree approximately between the numerical and leading-order analytical results in Figs. 4(a) and 5(b).

IV. DRAG

Local minima of the square Gaussian sidebands act as an intrinsic filter for off-resonant transitions as shown in Figs. 4–5. An ideal frequency composition of a CR pulse should have minimal frequency overlap with the aforementioned three collision types, and possibly higher-order transitions for stronger drive. This can in principle be achieved by applying a band-stop filter to the input pulse as

$$\tilde{\Omega}_c(\omega) = T(\omega)\tilde{\Omega}_{\mathrm{SG}}(\omega) \,, \tag{16}$$

where the transfer function $T(\omega)$ should notch every unwanted off-resonant transition (see Fig. 6). It is typically challenging to design a practical band-stop filter in hardware with sufficiently high quality factor.

Here, we explore the application of a DRAG pulse on the control qubit that results in an effective notch filter. In a leading-order Y-DRAG solution, we augment the square Gaussian, on the X axis of the control qubit, with the derivative on the Y axis as [23, 33]

$$\Omega_{cx}(t) = \Omega_{\mathrm{SG}}(t) \,, \tag{17a}$$

$$\Omega_{cy}(t) = \frac{1}{\Delta_D}\dot{\Omega}_{\mathrm{SG}}(t) \,, \tag{17b}$$

with Δ_D as the DRAG parameter. Hence, the overall pulse on the control qubit takes the form $\Omega_c(t) = \Omega_{cx}(t) + i\Omega_{cy}(t)$, and given that $\partial_t \leftrightarrow i\omega$, the effective DRAG transfer function reads

$$T_D(\omega) = 1 - \omega/\Delta_D \,. \tag{18}$$

Note that the optimal choice for Δ_D depends on the CR gate frequency allocation. In particular, due to multiple collision possibilities, tuning the DRAG coefficient to
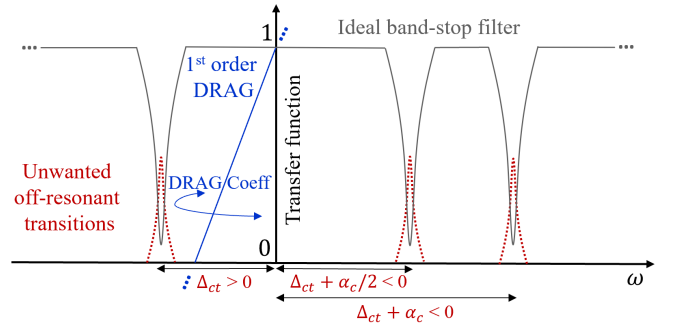

FIG. 6. Schematic transfer function of an ideal band-stop filter that mitigates off-resonant error by notching every individual collision versus a first-order DRAG solution of the form $\Omega_{cy}(t) = (1/\Delta_D)\dot{\Omega}_{cx}(t)$ and corresponding transfer function $T_D(\omega) = 1 - \omega/\Delta_D$. The frequency allocation here is for a case where non-BD error is most detrimental, i.e. $\Delta_{ct} > 0$ is the smallest transition frequency in absolute value. The idea, however, remains the same for other allocations, while the optimal DRAG parameter Δ_D varies.

mitigate a near-collision scenario can in principle cause the error due to other transitions to grow. Hence, first-order DRAG is not necessarily the optimal pulse shaping technique for CR gates, or systems with multiple collisions, as also pointed out in Ref. [34] for single-qubit gates with spectator qubits. This being said, we demonstrate noticeable improvement, especially for a qubit pair close to a type 1 (non-BD) collision.

We first discuss how to derive analytical conditions for the DRAG parameter to suppress individual collisions in Table I (see Appendix E). The derivation here is in terms of the leading-order Magnus term, hence slightly distinct from Refs. [23, 33], but reaches similar solutions. For instance, applying adiabatic expansion on the $|0_c\rangle \to |1_c\rangle$ transition probability gives

$$
\begin{aligned}
P_{|0_c\rangle \to |1_c\rangle} &\approx \frac{1}{4}\left|\int_0^{\tau_p} dt'\,\Omega_c(t')e^{i\Delta_{ct}t'}\right|^2 \\
&= \frac{1}{4}\left|\left\{\sum_{n=0}^{\infty}\left[\frac{1}{\Delta_{ct}}\left(\frac{i}{\Delta_{ct}}\frac{d}{dt'}\right)^n\Omega_c(t')\right]e^{i\Delta_{ct}t'}\right\}\Big|_0^{\tau_p}\right|^2 \,.
\end{aligned}
\tag{19}
$$

Replacing the Y-DRAG ansatz (17a)–(17b) into the 2nd line of Eq. (19) results

$$
\begin{aligned}
P_{|0_c\rangle \to |1_c\rangle} &= \frac{(1+\lambda_{01})^2}{4\Delta_{ct}^4}\left[\dot{\Omega}_{\mathrm{SG}}^2(\tau_p) + \dot{\Omega}_{\mathrm{SG}}^2(0)\right. \\
&\left. -2\cos(\Delta_{ct}\tau_p)\dot{\Omega}_{\mathrm{SG}}(\tau_p)\dot{\Omega}_{\mathrm{SG}}(0)\right] + O\left(\ddot{\Omega}_{\mathrm{SG}}^2\right) \,,
\end{aligned}
\tag{20}
$$

with normalized DRAG parameter $\lambda_{01} \equiv \Delta_{ct}/\Delta_D$. Higher-order terms are given in Appendix E, where we find that the dependence on $\ddot{\Omega}_{\mathrm{SG}}^2$ is also proportional to $(1+\lambda_{01})^2$. Therefore, setting $\lambda_{01} = -1$, i.e. $\Omega_{cy}(t) = -(1/\Delta_{ct})\dot{\Omega}_{\mathrm{SG}}(t)$, removes non-BD error up to terms of $O(\ddot{\Omega}_{\mathrm{SG}}^2)$. Similarly, the leading-order DRAG solution for suppressing type 3 error $|1_c\rangle \to |2_c\rangle$ reads
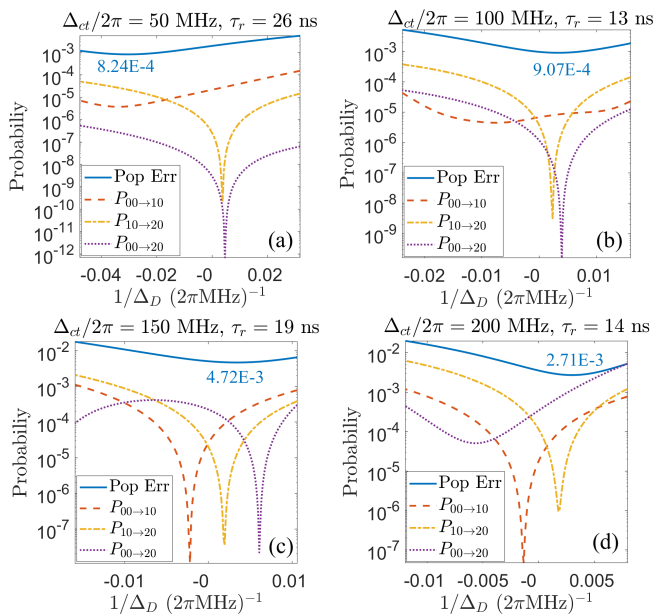
FIG. 7. Distinct off-resonant error types and average population error as a function of DRAG coefficient $1/\Delta_D$ for the same system parameters as in Fig. 4. The rise time in each case is adopted from the optimal values in Fig. 4(e) to quantify further improvement by DRAG. The numbers in blue show the average population error at the optimal DRAG parameter for each case. We note that using DRAG alone, without ramp optimization, e.g. for fixed $\tau_r = 20$ ns, results in comparable but slightly less improvement in the average population error for each case.

$\Delta_D = -(\Delta_{ct} + \alpha_c)$. Two-photon leakage is, however, more involved as the leading-order Magnus term appears as a two-time overlap integral (Table I and Appendix E). Inserting ansatz (17a)–(17b) results in a fourth-order polynomial in $1/\Delta_D$. Intuitively, we expect the optimal DRAG parameter to be approximately set according to half of the two-photon transition frequency, i.e. $\Delta_D \approx -(\Delta_{ct} + \alpha_c/2)$.

The above analytical DRAG solutions are based on an off-resonantly driven but isolated transmon qubit, while there are in principle $O(J\Omega_t)$ corrections. Moreover, DRAG correcting for one error type may increase the other (see Fig. 6). Therefore, we resort to a numerical sweep of the DRAG parameter in order to minimize the average population error. Figure 7 shows $\bar{E}_{\text{pop}}$ along with the three off-resonant error types as a function of $1/\Delta_D$, with τ_r set to the optimal value based on ramp optimization in Fig. 4(e). We see that depending on Δ_{ct}, there is a trade off between optimal Δ_D for individual error types as expected from analytics. For instance, the 50 MHz detuned pair in Fig. 7(a) is mainly limited by non-BD error without DRAG. Adding DRAG is beneficial in suppressing non-BD error, but increases the leakage to state $|2_c\rangle$. The optimal DRAG parameter is then determined by a balance between the two mechanisms. Lastly, sweeping drive amplitude (gate time) in Fig. 8
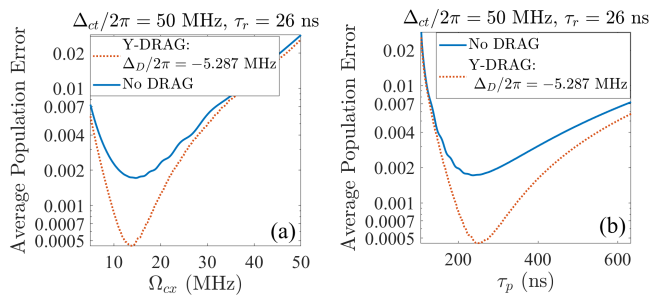

FIG. 8. Average population error as a function of (a) drive amplitude, and (b) gate time, for the $\Delta_{ct}/(2\pi) = 50$ MHz case and DRAG parameter set from the optimal value in Fig. 7(a) as $\Delta_D/(2\pi) = -5.287$ MHz. Drive parameters Ω_{cx}, Ω_{tx} and τ_p are set according to the approximate conditions (6)–(7) to calibrate a CNOT gate. The DRAG correction results in an approximate improvement of the average population error from 1.72×10^{-3} down to 4.53×10^{-4}.

shows partial improvement of the average error with a just a slightly slower gate of 246 ns. Altogether, with Gaussian shaping and Y-DRAG, the average population error for case (a) can be suppressed down to 4.53×10^{-4}. For pairs close to the type 2 collision, however, the interplay between different transitions is more involved and there is less improvement.

We note that resorting to the proposed pulse shaping techniques becomes more pertinent given the ongoing improvements in the coherence times of fixed-frequency transmon qubits. For instance, $ibmq\_peekskill$ [49] demonstrates on average $T_{1,\text{avg}} \approx 266.14$ μs and $T_{2,\text{avg}} \approx 256.61$ μs across the device, with the longest measured coherence times of $T_{1,\text{max}} \approx 453.13$ μs and $T_{2,\text{max}} \approx 663.59$ μs. Running a $\tau_p = 246$ ns direct CNOT gate results in an average incoherent error of $\bar{E}_{\text{incoh}} \approx \sum_{j=c,t}[(1/5)(\tau_p/T_{1j,\text{avg}}) + (2/5)(\tau_p/T_{2j,\text{avg}})] \approx 1.13 \times 10^{-3}$. Therefore, the improvements from ramp optimization and DRAG can be partially observed. To distinguish the full extent of the improvements demonstrated in Fig. (8), the minimum required coherence times are $T_{1j} \gtrsim 650$ μs and $T_{2j} \gtrsim 650$ μs for $j = c, t$.

V. CONCLUSION

Current calibrations of a CNOT gate using a CR architecture are premised on a BD form for effective interactions. In this work, we characterized off-resonant CR interactions, specifically non-BD contributions, as a potential source of coherent error, and illustrated its interplay with control pulse shapes. Time-dependent SW and Magnus perturbations reveal that off-resonant error occurs due to spectral overlap between the pulse and three unwanted off-resonant transitions on the control qubit, denoted as type 1–type 3. Non-BD error is enhanced for pairs in proximity of type 1, while leakage is increased

close to type 2–type 3 transitions.

To suppress such error terms, the pulse spectrum should have minimal content at the underlying transitions. The most immediate solution lies in optimal frequency allocation, i.e. simultaneously maximizing all off-resonant transition frequencies. This requires very precise fabrication of fixed-frequency transmons, where laser annealing [50, 51] has shown promising improvement. As a more active measure, optimal control techniques can effectively reduce the collision bounds, and allow working with CR pairs in closer proximity to the unwanted transitions. To this aim, we demonstrated promising improvement using two complimentary methods of ramp optimization and Y-DRAG on the control qubit, using a square Gaussian pulse shape. These methods come with little additional cost, i.e. small enhancement of leakage, and are preferable especially since they do not increase the total gate time. More involved optimal pulse shaping is a subject of future research. Our initial experiments on IBM CR processors confirm the benefit of pulse ramp optimization and DRAG on the average error obtained from two-qubit randomized benchmarking [48, 52]. The experimental results will be presented in a subsequent work.

Appendix A: Time-dependent SWPT

Here, we review the main results of a generalized time-dependent SWPT formalism [11]. We demonstrate the connection between time-dependent and independent perturbation through adiabatic expansion, in which effective terms depend on both the input pulse shapes as well as higher-order time derivatives. Furthermore, we discuss how time-dependent SWPT and the Magnus expansion are related through a frame transformation.

Consider a driven quantum system with time-dependent Hamiltonian

$$\hat{\mathcal{H}}_s(t) = \hat{\mathcal{H}}_0 + \lambda\hat{\mathcal{H}}_{\text{int}}(t) , \qquad \text{(A1)}$$

where $\hat{\mathcal{H}}_0$ is the zeroth-order system Hamiltonian, $\hat{\mathcal{H}}_{\text{int}}(t)$ denotes the time-dependent perturbation and λ is an auxiliary expansion parameter that facilitates the bookkeeping of perturbative corrections. Moving to the interaction frame with respect to $\hat{\mathcal{H}}_0$ we find

$$\lambda\hat{\mathcal{H}}_{\text{I}}(t) \equiv e^{i\hat{\mathcal{H}}_0 t}\left[\lambda\hat{\mathcal{H}}_{\text{int}}(t)\right] e^{-i\hat{\mathcal{H}}_0 t} , \qquad \text{(A2)}$$

which simplifies the perturbation theory in what follows.

The idea behind SWPT is to average out high-frequency off-resonant processes to come up with effective resonant interactions. Formally, this is equivalent to applying a unitary SW transformation to Eq. (A2) as

$$\hat{\mathcal{H}}_{\text{I,eff}}(t) \equiv \hat{U}_{\text{SW}}^{\dagger}(t)\left[\lambda\hat{\mathcal{H}}_{\text{I}}(t) - i\partial_t\right]\hat{U}_{\text{SW}}(t) , \qquad \text{(A3)}$$

where $\hat{U}_{\text{SW}}(t) \equiv \exp[-i\hat{G}(t)]$, $\hat{G}(t)$ is the unknown SW generator and $\hat{\mathcal{H}}_{\text{I,eff}}(t)$ is the effective Hamiltonian of interest. To obtain perturbative solutions, we first perform a series expansion of $\hat{G}(t)$ and $\hat{\mathcal{H}}_{\text{I,eff}}(t)$ in terms of λ as

$$\hat{G}(t) = \sum_{\lambda=1}^{\infty}\lambda^n\hat{G}_n(t) , \qquad \text{(A4a)}$$

$$\hat{\mathcal{H}}_{\text{I,eff}}(t) = \sum_{\lambda=1}^{\infty}\lambda^n\hat{\mathcal{H}}_{\text{I,eff}}^{(n)}(t) . \qquad \text{(A4b)}$$

Collecting equal powers of λ and enforcing the frame change at any arbitrary order, we find [11]

$$O(\lambda) : \begin{cases} \hat{\mathcal{H}}_{\text{I,eff}}^{(1)} = \mathcal{B}\left(\hat{\mathcal{H}}_I\right) , \\ \dot{\hat{G}}_1 = \mathcal{N}\left(\hat{\mathcal{H}}_I\right) , \end{cases} \qquad \text{(A5a)}$$

$$O(\lambda^2) : \begin{cases} \hat{\mathcal{H}}_{\text{I,eff}}^{(2)} = \mathcal{B}\left(i[\hat{G}_1, \hat{\mathcal{H}}_I] - i/2[\hat{G}_1, \dot{\hat{G}}_1]\right) , \\ \dot{\hat{G}}_2 = \mathcal{N}\left(i[\hat{G}_1, \hat{\mathcal{H}}_I] - i/2[\hat{G}_1, \dot{\hat{G}}_1]\right) , \end{cases} \qquad \text{(A5b)}$$

$$O(\lambda^3) : \begin{cases} \hat{\mathcal{H}}_{\text{I,eff}}^{(3)} = \mathcal{B}\Big(-(i/2)[\hat{G}_1, \dot{\hat{G}}_2] - i/2[\hat{G}_2, \dot{\hat{G}}_1] \\ \qquad + 1/6[\hat{G}_1, [\hat{G}_1, \dot{\hat{G}}_1]] + i[\hat{G}_2, \hat{\mathcal{H}}_I] \\ \qquad - 1/2[\hat{G}_1, [\hat{G}_1, \hat{\mathcal{H}}_I]]\Big) , \\ \dot{\hat{G}}_3 = \mathcal{N}\Big(-i/2[\hat{G}_1, \dot{\hat{G}}_2] - i/2[\hat{G}_2, \dot{\hat{G}}_1] \\ \qquad + 1/6[\hat{G}_1, [\hat{G}_1, \dot{\hat{G}}_1]] + i[\hat{G}_2, \hat{\mathcal{H}}_I] \\ \qquad - 1/2[\hat{G}_1, [\hat{G}_1, \hat{\mathcal{H}}_I]]\Big) , \end{cases} \qquad \text{(A5c)}$$

$$O(\lambda^4): \begin{cases} \hat{\mathcal{H}}_{\mathrm{I,eff}}^{(4)} = \mathcal{B}\Big(-i/2[\hat{G}_1,\dot{\hat{G}}_3] - i/2[\hat{G}_2,\dot{\hat{G}}_2] \\ \quad -i/2[\hat{G}_3,\dot{\hat{G}}_1] + 1/6[\hat{G}_1,[\hat{G}_1,\dot{\hat{G}}_2]] \\ \quad +1/6[\hat{G}_1,[\hat{G}_2,\dot{\hat{G}}_1]] + 1/6[\hat{G}_2,[\hat{G}_1,\dot{\hat{G}}_1]] \\ \quad +i/24[\hat{G}_1,[\hat{G}_1,[\hat{G}_1,\dot{\hat{G}}_1]]] + i[\hat{G}_3,\hat{\mathcal{H}}_I] \\ \quad -1/2[\hat{G}_1,[\hat{G}_2,\hat{\mathcal{H}}_I]] - 1/2[\hat{G}_2,[\hat{G}_1,\hat{\mathcal{H}}_I]] \\ \quad -i/6[\hat{G}_1,[\hat{G}_1,[\hat{G}_1,\hat{\mathcal{H}}_I]]]\Big), \\[4pt] \dot{\hat{G}}_4 = \mathcal{N}\Big(-i/2[\hat{G}_1,\dot{\hat{G}}_3] - i/2[\hat{G}_2,\dot{\hat{G}}_2] \\ \quad -i/2[\hat{G}_3,\dot{\hat{G}}_1] + 1/6[\hat{G}_1,[\hat{G}_1,\dot{\hat{G}}_2]] \\ \quad +1/6[\hat{G}_1,[\hat{G}_2,\dot{\hat{G}}_1]] + 1/6[\hat{G}_2,[\hat{G}_1,\dot{\hat{G}}_1]] \\ \quad +i/24[\hat{G}_1,[\hat{G}_1,[\hat{G}_1,\dot{\hat{G}}_1]]] + i[\hat{G}_3,\hat{\mathcal{H}}_I] \\ \quad -1/2[\hat{G}_1,[\hat{G}_2,\hat{\mathcal{H}}_I]] - 1/2[\hat{G}_2,[\hat{G}_1,\hat{\mathcal{H}}_I]] \\ \quad -i/6[\hat{G}_1,[\hat{G}_1,[\hat{G}_1,\hat{\mathcal{H}}_I]]]\Big). \end{cases}$$

$$(A5d)$$

In Eqs. (A5a)–(A5d), $\mathcal{B}(\bullet)$ and $\mathcal{N}(\bullet)$ represent projections onto the desired frame, in which effective interactions are resonant, and the rest of the Hilbert space, respectively. For our CR model in Eqs. (1)–(3), and under a CW drive, $\mathcal{B}(\bullet)$ is equivalent to keeping contributions in the BD frame shown in Fig. 2. To generalize for a pulsed drive, we adopt the same definition of the BD frame as the CW case. The time dependence due to the pulse shape then appears as overlap integrals between the system transition frequencies and the pulse shape. The overlap integrals can also be related to the time derivatives of the pulse through an adiabatic expansion as discussed in the following. More generally, Ref. [26] argues that obtaining the effective resonant interactions for a CW drive corresponds to separating the zero-frequency part of the right-hand side of Eqs. (A5a)–(A5d) at each order in perturbation by time averaging.

Note that in solving the operator-valued ODEs (A5a)–(A5d), the initial conditions for $\hat{G}_n(t)$ appear as free parameters, where different choices here correspond to distinct frames. The goal of SWPT is to reach a desired form for the effective Hamiltonian by removing off-resonant interactions. Therefore, the homogeneous solution is not of interest and the natural choice is to solve for the *particular* solution, obtained by the *indefinite* integral of the right hand side at each order. The indefinite integral is consistent with the fact that in SWPT we are interested in effective Hamiltonian rates, while effective rotation angles can be obtained after this step using a *definite integral over the pulse duration*. To explain this point, we show the explicit expression for the effective Hamiltonian up to the second order. According to Eq. (A5a), the lowest-order generator should be set as

$$\hat{G}_1(t) = \int^t dt' \mathcal{N}\left(\hat{\mathcal{H}}_I(t')\right). \qquad (A6)$$

Replacing solution (A6) into the second-order expres-

sion (A5b) results

$$\hat{\mathcal{H}}_{\mathrm{I,eff}}^{(2)}(t) = \mathcal{B}\left(i\left[\int^t dt' \mathcal{N}\left(\hat{\mathcal{H}}_I(t')\right),\hat{\mathcal{H}}_I(t)\right]\right) \\ - \mathcal{B}\left(\frac{i}{2}\left[\int^t dt' \mathcal{N}\left(\hat{\mathcal{H}}_I(t')\right),\mathcal{N}\left(\hat{\mathcal{H}}_I(t)\right)\right]\right), \quad (A7)$$

which can be understood as possible mixings between the interaction Hamiltonian at time t and t'. Higher-order solutions account for more complex time correlations. In the limiting case where $\mathcal{B}(\hat{\mathcal{H}}_I) = 0$, and hence $\mathcal{N}(\hat{\mathcal{H}}_I) = \hat{\mathcal{H}}_I$, Eq. (A7) simplifies to $\hat{\mathcal{H}}_{\mathrm{I,eff}}^{(2)}(t) = \mathcal{B}(\frac{i}{2}[\int^t dt' \hat{\mathcal{H}}_I(t'),\hat{\mathcal{H}}_I(t)])$ as in Ref. [11].

It is importat to note that, for any driven quantum system, the dynamics is determined by the overall time evolution operator and hence is independent of our frame choice. Therefore, we need to map the effective unitary operator back to the interaction frame. This step is commonly neglected in SWPT, but is crucial in capturing the error of our effective model due to off-resonant processes. Given that the interaction and the effective frame wavefunctions are related as $|\Psi_{\mathrm{I}}(t)\rangle = \hat{U}_{\mathrm{SW}}(t)|\Psi_{\mathrm{I,eff}}(t)\rangle$ we find

$$\hat{U}_{\mathrm{I}}(t,t_0) = \hat{U}_{\mathrm{SW}}(t)\hat{U}_{\mathrm{I,eff}}(t,t_0)\hat{U}_{\mathrm{SW}}^\dagger(t_0), \qquad (A8)$$

where $\hat{U}_{\mathrm{I,eff}}(t,t_0)$ is given in terms of the effective Hamiltonian $\hat{\mathcal{H}}_{\mathrm{I,eff}}(t)$ as

$$\hat{U}_{\mathrm{I,eff}}(t,t_0) = \mathbb{T}\exp\left[-i\int_{t_0}^t dt' \hat{\mathcal{H}}_{\mathrm{I,eff}}(t')\right]. \qquad (A9)$$

Furthermore, the SW transformation in Eq. (A8) can be computed by either a matrix exponentiation of the perturbative solution for $\hat{G}$, which is norm-preserving, or by a perturbative expansion of the exponential as

$$\begin{aligned} \hat{U}_{\mathrm{SW}} = & \hat{I} - i\hat{G}_1 - i\hat{G}_2 - \frac{1}{2}\hat{G}_1^2 \\ & - i\hat{G}_3 - \frac{1}{2}\left(\hat{G}_1\hat{G}_2 + \hat{G}_2\hat{G}_1\right) + \frac{i}{6}\hat{G}_1^3 \\ & - i\hat{G}_4 - \frac{1}{2}(\hat{G}_1\hat{G}_3 + \hat{G}_3\hat{G}_1) - \hat{G}_2^2 \\ & + \frac{i}{6}\left(\hat{G}_1^2\hat{G}_2 + \hat{G}_1\hat{G}_2\hat{G}_1 + \hat{G}_2\hat{G}_1^2\right) \\ & + \frac{1}{24}\hat{G}_1^4 + O(\lambda^5). \end{aligned} \qquad (A10)$$

Time-dependent SWPT provides the means to also quantify adiabaticity, and importantly, in the limit of adiabatic response, the results agree with those found from time-independent perturbation. In particular, consider a generic correction $\int^t dt' \Omega(t')\exp[i\Delta(t-t')]$, which is a form that derives from Eq. (A7). Here, $\Omega(t)$ denotes the time-dependent drive amplitude with an intrinsic rise time τ_r, and Δ is the transition frequency for the underlying physical process. Adiabatic response is ensured

when the transition frequency is much larger than the pulse spectral width, i.e. $\Delta \gg 1/\tau_r$. However, we can quantify adiabaticity by expanding in orders of $1/(\Delta \cdot \tau_r)$, which appears naturally in terms of the time derivatives of the pulse as

$$
\int^t dt' \Omega(t') e^{i\Delta(t-t')} = \frac{\Omega(t)}{-i\Delta} - \frac{\dot{\Omega}(t)}{(-i\Delta)^2} + \cdots
$$
$$
= \sum_{n=0}^{\infty} \frac{(-1)^n}{(-i\Delta)^{n+1}} \frac{d^n \Omega(t)}{dt^n} \, . \quad \text{(A11)}
$$

The right hand side of Eq. (A11) is computed via integration by parts. Keeping the first term in the adiabatic expansion (A11) agrees with the time-independent perturbation theory, while higher-order terms capture the transient effects during pulse ramps.

Lastly, we note that the idea behind time-dependent SWPT and Magnus expansion are similar, where a perturbative expansion is made in terms of the generator (logarithm) of unitary operators. The two methods can be related to one another as

$$
\hat{U}_{\mathrm{I}}(t,t_0) \equiv e^{-i\hat{K}(t,t_0)} = e^{-i\hat{G}(t)} \hat{U}_{\mathrm{I,eff}}(t,t_0) e^{i\hat{G}(t_0)} \quad \text{(A12)}
$$

where $\hat{K}(t,t_0)$ is the Magnus generator. The main distinction is that Magnus solves for the overall time evolution operator directly, without partitioning into resonant and off-resonant sectors.

In summary, we have demonstrated the application of time-dependent SWPT for computing an effective Hamiltonian for a driven quantum system. This method is capable of accounting for the renormalization of effective Hamiltonian terms due to control pulse shapes, and also the corresponding off-resonant error.

Appendix B: Effective CR Hamiltonian

Here, we apply the time-dependent SWPT of Appendix A to the CR model in Eqs. (1)–(3) and provide expressions for the effective gate parameters. Our time-dependent results agree with and provide a natural extension of the time-independent effective CR terms given in Refs. [8, 10, 11].

For CR gates, the drive frequency is resonant with the target qubit leading to Rabi oscillations around the X or Y axis of the target. Therefore, the frame in which the effective terms are resonant is BD with respect to the control qubit (see Fig. 2). Once the effective Hamiltonian is obtained in the extended Hilbert space, we read off the CR gate parameters as

$$
\hat{\mathcal{H}}_{\mathrm{CR,eff}}(t) \equiv \sum_{m,n=i,x,y,z} \frac{1}{2} \omega_{\sigma_m \sigma_n}(t) \hat{\sigma}_m \otimes \hat{\sigma}_n \, , \quad \text{(B1a)}
$$

$$
\omega_{\sigma_m \sigma_n}(t) \equiv \frac{1}{2} \mathrm{Tr}\left((\hat{\sigma}_m \otimes \hat{\sigma}_n) \hat{\mathcal{H}}_{\mathrm{I,eff}}(t) \right) \, , \quad \text{(B1b)}
$$

where the order is control ⊗ target. In the BD frame, the effective Hamiltonian consists of IX, IY, ZX, ZY, IZ,

ZI and ZZ terms. In the following, we provide expressions in powers of the drive amplitudes $\Omega_c(t)$ and $\Omega_t(t)$.

1. Zeroth order

Up to the zeroth order, we find an effective static ZZ term as

$$
\omega_{zz}^{(0)}(t) = \left(\frac{1}{\Delta_{ct} - \alpha_t} - \frac{1}{\Delta_{ct} + \alpha_c} \right) J^2 + O\left(J^4\right) \, , \quad \text{(B2)}
$$

as a result of level repulsion between states $|11\rangle \leftrightarrow |02\rangle$ and $|11\rangle \leftrightarrow |20\rangle$.

2. First order

Up to the first order, we find corrections to the IX, IY, ZX and ZY terms as

$$
\omega_{ix}^{(1)}(t) = \Omega_{tx}(t) - \frac{J}{\Delta_{ct} + \alpha_c} \Omega_{cx}(t) \, , \quad \text{(B3a)}
$$

$$
\omega_{iy}^{(1)}(t) = \Omega_{ty}(t) - \frac{J}{\Delta_{ct} + \alpha_c} \Omega_{cy}(t) \, , \quad \text{(B3b)}
$$

$$
\omega_{zx}^{(1)}(t) = \left(\frac{J}{\Delta_{ct} + \alpha_c} - \frac{J}{\Delta_{ct}} \right) \Omega_{cx}(t) \, , \quad \text{(B3c)}
$$

$$
\omega_{zy}^{(1)}(t) = \left(\frac{J}{\Delta_{ct} + \alpha_c} - \frac{J}{\Delta_{ct}} \right) \Omega_{cy}(t) \, . \quad \text{(B3d)}
$$

In particular, IX and IY depend directly on the resonant target drive, while the dependence on the control drive is indirect and mediated through states $|10\rangle$ and $|20\rangle$, resulting in transition frequencies Δ_{ct} and $\Delta_{ct} + \alpha_c$ (see Fig. 1).

3. Second order

Up to the second order in drive amplitudes, there are corrections to the diagonal components ZI, IZ and ZZ, proportional to $\Omega_{cx}^2 + \Omega_{cy}^2$, $J^2(\Omega_{cx}^2 + \Omega_{cy}^2)$, $J(\Omega_{cx}\Omega_{tx} + \Omega_{cy}\Omega_{ty})$, $\Omega_{cx}\dot{\Omega}_{cy} - \Omega_{cy}\dot{\Omega}_{cx}$, $J^2(\Omega_{cx}\dot{\Omega}_{cy} - \Omega_{cy}\dot{\Omega}_{cx})$ and $J(\Omega_{tx}\dot{\Omega}_{cy} - \Omega_{ty}\dot{\Omega}_{cx})$. For simplicity, we have performed the adiabtic expansion (A11) up to the leading order in the pulse derivative. For instance, the expression for $\omega_{zi}^{(2)}(t)$ reads

$$
\begin{aligned}
\omega_{zi}^{(2)}(t) = {} & C_{zi,1}^{(2)} \left[\Omega_{cx}^2(t) + \Omega_{cy}^2(t) \right] \\
& + C_{zi,2}^{(2)} J^2 \left[\Omega_{cx}^2(t) + \Omega_{cy}^2(t) \right] \\
& + C_{zi,3}^{(2)} J \left[\Omega_{tx}(t)\Omega_{cx}(t) + \Omega_{ty}(t)\Omega_{cy}(t) \right] \\
& + C_{zi,4}^{(2)} \left[\Omega_{cx}(t)\dot{\Omega}_{cy}(t) - \Omega_{cy}(t)\dot{\Omega}_{cx}(t) \right] \\
& + C_{zi,5}^{(2)} J^2 \left[\Omega_{cx}(t)\dot{\Omega}_{cy}(t) - \Omega_{cy}(t)\dot{\Omega}_{cx}(t) \right] \\
& + C_{zi,6}^{(2)} J \left[\Omega_{tx}(t)\dot{\Omega}_{cy}(t) - \Omega_{ty}(t)\dot{\Omega}_{cx}(t) \right] \, ,
\end{aligned} \quad \text{(B4a)}
$$

where $C^{(2)}_{zi,1}$ to $C^{(2)}_{zi,6}$ are the corresponding coefficients given in the following. Similar expressions for $\omega^{(2)}_{iz}(t)$ and $\omega^{(2)}_{zz}(t)$ follow.

For the Stark shift on the control qubit, the coefficients are found as

$$C^{(2)}_{zi,1} = \frac{1}{2(\Delta_{ct} + \alpha_c)} - \frac{1}{2\Delta_{ct}} \;, \tag{B5a}$$

$$\begin{aligned}
C^{(2)}_{zi,2} = &-\frac{1}{4\alpha_c \Delta^2_{ct}} + \frac{1}{\alpha_c (\alpha_c + \Delta_{ct})^2} - \frac{3}{4\alpha_c (2\alpha_c + \Delta_{ct})^2} \\
&+ \frac{1}{\alpha^2_c \Delta_{ct}} - \frac{2}{\alpha^2_c (\alpha_c + \Delta_{ct})} - \frac{3}{\alpha^2_c (2\alpha_c + \Delta_{ct})} \\
&- \frac{4}{\alpha^2_c (\alpha_c + 2\Delta_{ct})} + \frac{12}{\alpha^2_c (3\alpha_c + 2\Delta_{ct})} \\
&+ \frac{1}{\alpha_t (\alpha_c + \Delta_{ct} - \alpha_t)^2} - \frac{1}{\alpha_c \alpha_t (\Delta_{ct} - \alpha_t)} \\
&+ \frac{1}{\alpha_c \alpha_t (\alpha_c + \Delta_{ct} - \alpha_t)} \;,
\end{aligned} \tag{B5b}$$

$$C^{(2)}_{zi,3} = \frac{\alpha_c}{2\Delta_{ct}\alpha_t (\alpha_c + \Delta_{ct})} \;, \tag{B5c}$$

$$C^{(2)}_{zi,4} = \frac{1}{2\Delta^2_{ct}} - \frac{1}{2(\Delta_{ct} + \alpha_c)^2} \;, \tag{B5d}$$

$$\begin{aligned}
C^{(2)}_{zi,5} = &\frac{6}{\alpha^3_c (\alpha_c + \Delta_{ct})} - \frac{9}{2\alpha^3_c (2\alpha_c + \Delta_{ct})} \\
&- \frac{1}{2\alpha^2_c (\alpha_c + \Delta_{ct})^2} - \frac{3}{4\alpha^2_c (2\alpha_c + \Delta_{ct})^2} \\
&+ \frac{4}{\alpha^2_c (\alpha_c + 2\Delta_{ct})^2} - \frac{12}{\alpha^2_c (3\alpha_c + 2\Delta_{ct})^2} \\
&- \frac{3}{2\alpha^3_c \Delta_{ct}} + \frac{1}{4\alpha^2_c \Delta^2_{ct}} + \frac{1}{\alpha_c \alpha^2_t (\Delta_{ct} - \alpha_t)} \\
&- \frac{1}{\alpha_c \alpha^2_t (\alpha_c + \Delta_{ct} - \alpha_t)} - \frac{1}{\alpha^2_t (\alpha_c + \Delta_{ct} - \alpha_t)^2} \;,
\end{aligned} \tag{B5e}$$

$$C^{(2)}_{zi,6} = -\frac{\alpha^2_c \Delta_{ct} + \alpha_c \Delta^2_{ct} + 2\alpha_c \Delta_{ct}\alpha_t + \alpha^2_c \alpha_t}{2\Delta^2_{ct}\alpha^2_t (\alpha_c + \Delta_{ct})^2} \;. \tag{B5f}$$

For Stark shift on the target qubit, i.e. the IZ term, the coefficients read

$$C^{(2)}_{iz,1} = 0 \;, \tag{B6a}$$

$$\begin{aligned}
C^{(2)}_{iz,2} = &\frac{1}{4\alpha_c \Delta^2_{ct}} + \frac{1}{2\alpha_c (\alpha_c + \Delta_{ct})^2} - \frac{3}{4\alpha_c (2\alpha_c + \Delta_{ct})^2} \\
&- \frac{1}{\alpha^2_c \Delta_{ct}} - \frac{4}{\alpha^2_c (\alpha_c + \Delta_{ct})} - \frac{3}{\alpha^2_c (2\alpha_c + \Delta_{ct})} \\
&+ \frac{4}{\alpha^2_c (\alpha_c + 2\Delta_{ct})} + \frac{12}{\alpha^2_c (3\alpha_c + 2\Delta_{ct})} \\
&+ \frac{1}{\alpha_t (\alpha_c + \Delta_{ct} - \alpha_t)^2} - \frac{1}{\alpha_c \alpha_t (\Delta_{ct} - \alpha_t)} \\
&+ \frac{1}{\alpha_c \alpha_t (\alpha_c + \Delta_{ct} - \alpha_t)} + \frac{1}{2\alpha_t (\Delta_{ct} - \alpha_t)^2} \;,
\end{aligned} \tag{B6b}$$

$$C^{(2)}_{iz,3} = -\frac{\Delta_{ct} + \alpha_c + \alpha_t}{2\alpha_t (\alpha_c + \Delta_{ct})^2} \;, \tag{B6c}$$

$$C^{(2)}_{iz,4} = 0 \;, \tag{B6d}$$

$$\begin{aligned}
C^{(2)}_{iz,5} = &\frac{3}{\alpha^3_c (\alpha_c + \Delta_{ct})} - \frac{9}{2\alpha^3_c (2\alpha_c + \Delta_{ct})} \\
&- \frac{1}{\alpha^2_c (\alpha_c + \Delta_{ct})^2} - \frac{3}{4\alpha^2_c (2\alpha_c + \Delta_{ct})^2} \\
&- \frac{4}{\alpha^2_c (\alpha_c + 2\Delta_{ct})^2} - \frac{12}{\alpha^2_c (3\alpha_c + 2\Delta_{ct})^2} \\
&+ \frac{3}{2\alpha^3_c \Delta_{ct}} - \frac{1}{4\alpha^2_c \Delta^2_{ct}} + \frac{1}{\alpha_c \alpha^2_t (\Delta_{ct} - \alpha_t)} \\
&- \frac{1}{\alpha_c \alpha^2_t (\alpha_c + \Delta_{ct} - \alpha_t)} - \frac{1}{2\alpha^2_t (\Delta_{ct} - \alpha_t)^2} \\
&- \frac{1}{\alpha^2_t (\alpha_c + \Delta_{ct} - \alpha_t)^2} \;,
\end{aligned} \tag{B6e}$$

$$C^{(2)}_{iz,6} = \frac{\alpha^2_c + 2\alpha_c \Delta_{ct} + \alpha_c \alpha_t + \Delta^2_{ct} + \Delta_{ct}\alpha_t + \alpha^2_t}{2\alpha^2_t (\alpha_c + \Delta_{ct})^3} \;. \tag{B6f}$$

Similarly, for the dynamic ZZ term, the coefficients

read

$$C_{zz,1}^{(2)} = 0 \,, \tag{B7a}$$

$$C_{zz,2}^{(2)} = \frac{1}{4\alpha_c \Delta_{\mathrm{ct}}^2} - \frac{1}{\alpha_c \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^2} + \frac{3}{4\alpha_c \left(2\alpha_c + \Delta_{\mathrm{ct}}\right)^2}$$
$$- \frac{1}{\alpha_c^2 \Delta_{\mathrm{ct}}} + \frac{2}{\alpha_c^2 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)} + \frac{3}{\alpha_c^2 \left(2\alpha_c + \Delta_{\mathrm{ct}}\right)}$$
$$+ \frac{4}{\alpha_c^2 \left(\alpha_c + 2\Delta_{\mathrm{ct}}\right)} - \frac{12}{\alpha_c^2 \left(3\alpha_c + 2\Delta_{\mathrm{ct}}\right)}$$
$$- \frac{1}{\alpha_t \left(\alpha_c + \Delta_{\mathrm{ct}} - \alpha_t\right)^2} + \frac{1}{\alpha_c \alpha_t \left(\Delta_{\mathrm{ct}} - \alpha_t\right)}$$
$$- \frac{1}{\alpha_c \alpha_t \left(\alpha_c + \Delta_{\mathrm{ct}} - \alpha_t\right)} \,, \tag{B7b}$$

$$C_{zz,3}^{(2)} = -\frac{\alpha_c^2 \Delta_{\mathrm{ct}} + \alpha_c \Delta_{\mathrm{ct}}^2 + 2\alpha_c \Delta_{\mathrm{ct}} \alpha_t + \alpha_c^2 \alpha_t}{2\Delta_{\mathrm{ct}}^2 \alpha_t \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^2} \,, \tag{B7c}$$

$$C_{zz,4}^{(2)} = 0 \,, \tag{B7d}$$

$$C_{zz,5}^{(2)} = -\frac{6}{\alpha_c^3 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)} + \frac{9}{2\alpha_c^3 \left(2\alpha_c + \Delta_{\mathrm{ct}}\right)}$$
$$+ \frac{1}{2\alpha_c^2 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^2} + \frac{3}{4\alpha_c^2 \left(2\alpha_c + \Delta_{\mathrm{ct}}\right)^2}$$
$$- \frac{4}{\alpha_c^2 \left(\alpha_c + 2\Delta_{\mathrm{ct}}\right)^2} + \frac{12}{\alpha_c^2 \left(3\alpha_c + 2\Delta_{\mathrm{ct}}\right)^2}$$
$$+ \frac{3}{2\alpha_c^3 \Delta_{\mathrm{ct}}} - \frac{1}{4\alpha_c^2 \Delta_{\mathrm{ct}}^2} - \frac{1}{\alpha_c \alpha_t^2 \left(\Delta_{\mathrm{ct}} - \alpha_t\right)}$$
$$+ \frac{1}{\alpha_c \alpha_t^2 \left(\alpha_c + \Delta_{\mathrm{ct}} - \alpha_t\right)} + \frac{1}{\alpha_t^2 \left(\alpha_c + \Delta_{\mathrm{ct}} - \alpha_t\right)^2} \,, \tag{B7e}$$

$$C_{zz,6}^{(2)} = -\frac{1}{2\left(\alpha_c + \Delta_{\mathrm{ct}}\right)^3} - \frac{1}{2\alpha_t \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^2}$$
$$- \frac{1}{2\alpha_t^2 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)} + \frac{1}{2\Delta_{\mathrm{ct}}^3} + \frac{1}{2\Delta_{\mathrm{ct}}^2 \alpha_t} + \frac{1}{2\Delta_{\mathrm{ct}} \alpha_t^2} \,. \tag{B7f}$$

4. Third order

Up to the third order in drive amplitudes, we find corrections to the IX, IY, ZX and ZY terms. Here, there are numerous contributions and, for brevity, we quote certain dominant corrections to e.g. the IX as

$$\omega_{ix}^{(3)}(t) = C_{ix,1}^{(3)} J \left[\Omega_{cx}^2(t) + \Omega_{cy}^2(t)\right] \Omega_{cx}(t)$$
$$+ C_{ix,2}^{(3)} J \dot{\Omega}_{cx}(t) \Omega_{cx}(t) \tag{B8}$$
$$+ C_{ix,3}^{(3)} J \dot{\Omega}_{cy}^2(t) \Omega_{cx}(t) \,.$$

Expressions for the ZX, IY, and ZY terms have a similar form.

The coefficients for the IX term are

$$C_{ix,1}^{(3)} = \frac{\alpha_c \Delta_{\mathrm{ct}}}{\left(\alpha_c + \Delta_{\mathrm{ct}}\right)^3 \left(\alpha_c + 2\Delta_{\mathrm{ct}}\right) \left(3\alpha_c + 2\Delta_{\mathrm{ct}}\right)} \,, \tag{B9a}$$

$$C_{ix,2}^{(3)} = -\frac{1}{\left(\alpha_c + \Delta_{\mathrm{ct}}\right)^5} + \frac{1}{2\alpha_c \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^4}$$
$$+ \frac{1}{12\alpha_c^2 \Delta_{\mathrm{ct}}^3} - \frac{5}{24\alpha_c^2 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^3} - \frac{1}{3\alpha_c^3 \Delta_{\mathrm{ct}}^2}$$
$$- \frac{71}{24\alpha_c^3 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^2} + \frac{2}{3\alpha_c^3 \left(\alpha_c + 2\Delta_{\mathrm{ct}}\right)^2}$$
$$- \frac{12}{\alpha_c^3 \left(3\alpha_c + 2\Delta_{\mathrm{ct}}\right)^2} + \frac{103}{8\alpha_c^4 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)}$$
$$+ \frac{9}{8\alpha_c^4 \Delta_{\mathrm{ct}}} - \frac{4}{\alpha_c^4 \left(\alpha_c + 2\Delta_{\mathrm{ct}}\right)} - \frac{24}{\alpha_c^4 \left(3\alpha_c + 2\Delta_{\mathrm{ct}}\right)} \tag{B9b}$$

$$C_{ix,3}^{(3)} = \frac{\alpha_c \left(4\alpha_c + 7\Delta_{\mathrm{ct}}\right)}{24\Delta_{\mathrm{ct}}^2 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^3 \left(\alpha_c + 2\Delta_{\mathrm{ct}}\right)^2} \,, \tag{B9c}$$

and for the ZX term are

$$C_{zx,1}^{(3)} = \frac{\alpha_c^2 \left(3\alpha_c^3 + 11\alpha_c^2 \Delta_{\mathrm{ct}} + 15\alpha_c \Delta_{\mathrm{ct}}^2 + 9\Delta_{\mathrm{ct}}^3\right)}{2\Delta_{\mathrm{ct}}^3 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^3 \left(\alpha_c + 2\Delta_{\mathrm{ct}}\right) \left(3\alpha_c + 2\Delta_{\mathrm{ct}}\right)} \,, \tag{B10a}$$

$$C_{zx,2}^{(3)} = \frac{1}{\left(\alpha_c + \Delta_{\mathrm{ct}}\right)^5} + \frac{1}{2\alpha_c \Delta_{\mathrm{ct}}^4} - \frac{1}{2\alpha_c \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^4}$$
$$- \frac{1}{2\alpha_c^2 \Delta_{\mathrm{ct}}^3} + \frac{3}{8\alpha_c^2 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^3} + \frac{1}{4\alpha_c^3 \Delta_{\mathrm{ct}}^2}$$
$$+ \frac{29}{8\alpha_c^3 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^2} - \frac{2}{\alpha_c^3 \left(\alpha_c + 2\Delta_{\mathrm{ct}}\right)^2}$$
$$+ \frac{12}{\alpha_c^3 \left(3\alpha_c + 2\Delta_{\mathrm{ct}}\right)^2} + \frac{5}{8\alpha_c^4 \Delta_{\mathrm{ct}}} - \frac{85}{8\alpha_c^4 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)}$$
$$- \frac{4}{\alpha_c^4 \left(\alpha_c + 2\Delta_{\mathrm{ct}}\right)} + \frac{24}{\alpha_c^4 \left(3\alpha_c + 2\Delta_{\mathrm{ct}}\right)} - \frac{1}{2\Delta_{\mathrm{ct}}^5} \,, \tag{B10b}$$

$$C_{zx,3}^{(3)} = -\frac{\alpha_c \left(2\alpha_c^2 + 8\alpha_c \Delta_{\mathrm{ct}} + 7\Delta_{\mathrm{ct}}^2\right)}{8\Delta_{\mathrm{ct}}^3 \left(\alpha_c + \Delta_{\mathrm{ct}}\right)^3 \left(\alpha_c + 2\Delta_{\mathrm{ct}}\right)^2} \,. \tag{B10c}$$

5. Fourth order

There are multitude of corrections up to the fourth order. Among those, here, we quote the most dominant contribution to the control qubit Stark shift as

$$\omega_{zi}^{(4)}(t) \equiv \frac{\left(3\alpha_c^5 + 11\alpha_c^4 \Delta_{ct} + 15\alpha_c^3 \Delta_{ct}^2 + 9\alpha_c^2 \Delta_{ct}^3\right)}{8\Delta_{ct}^3 \left(\alpha_c + \Delta_{ct}\right)^3 \left(\alpha_c + 2\Delta_{ct}\right) \left(3\alpha_c + 2\Delta_{ct}\right)}$$
$$\times \left[\Omega_{cx}^2(t) + \Omega_{cy}^2(t)\right]^2 \,. \tag{B11}$$

Accounting for Eq. (B11), on top of the first line of Eq. (B4a), becomes important at stronger CR drive which suppresses the Stark shift in magnitude [see also Fig. (3c) of Ref. [11]].

In summary, our time-dependent formalism captures the previously known expressions for the gate parameters, e.g. Eq. (B5a) for ZI, Eq. (B9a) for IX and Eq. (B10a) for ZX, while also quantifying non-adiabatic response in terms of the derivatives of underlying pulse shapes.

Appendix C: Effective CR time evolution operator

Here, we calculate the effective time evolution operator for the CR gate as

$$\hat{U}_{\text{CR,eff}}(t,0) = \mathbb{T} \exp\left[-i \int_0^t dt' \hat{\mathcal{H}}_{\text{CR,eff}}(t')\right], \quad \text{(C1)}$$

where $\hat{\mathcal{H}}_{\text{CR,eff}}(t)$ is the effective BD Hamiltonian in Appendix B. In general, time-dependent corrections in the effective terms due to pulse ramps do not commute, hence explicit time-ordering is needed. For simplicity, however, we keep the leading order in the adiabatic expansion which corresponds to the constant mid-part of the pulses. The resulting approximate expressions are helpful for designing specific gate calibrations and reverse engineering the required drive scheme [11, 12].

In the control=$|0\rangle$ subspace, the $|00\rangle\langle00|$, $|00\rangle\langle01|$, $|01\rangle\langle00|$ and $|01\rangle\langle01|$ components of $\hat{U}_{\text{CR,eff}}(t,0)$ are found respectively as

$$\frac{e^{-\frac{1}{2}i\omega_{zi}t}\left[\omega_+ \cos\left(\frac{\omega_+ t}{2}\right) - i(\omega_{iz} + \omega_{zz})\sin\left(\frac{\omega_+ t}{2}\right)\right]}{\omega_+}, \quad \text{(C2a)}$$

$$-\frac{e^{-\frac{1}{2}i\omega_{zi}t}\left[i(\omega_{ix} + \omega_{zx}) + (\omega_{iy} + \omega_{zy})\right]\sin\left(\frac{\omega_+ t}{2}\right)}{\omega_+}, \quad \text{(C2b)}$$

$$-\frac{e^{-\frac{1}{2}i\omega_{zi}t}\left[i(\omega_{ix} + \omega_{zx}) - (\omega_{iy} + \omega_{zy})\right]\sin\left(\frac{\omega_+ t}{2}\right)}{\omega_+}, \quad \text{(C2c)}$$

$$\frac{e^{-\frac{1}{2}i\omega_{zi}t}\left[\omega_+ \cos\left(\frac{\omega_+ t}{2}\right) + i(\omega_{iz} + \omega_{zz})\sin\left(\frac{\omega_+ t}{2}\right)\right]}{\omega_+}, \quad \text{(C2d)}$$

where $\omega_\pm$ are collective CR frequencies [11, 12] defined as

$$\omega_\pm \equiv \left[(\omega_{ix} \pm \omega_{zx})^2 + (\omega_{iy} \pm \omega_{zy})^2 + (\omega_{iz} \pm \omega_{zz})^2\right]^{1/2}. \quad \text{(C3)}$$

Similarly, in the control=$|1\rangle$ subspace, the $|10\rangle\langle10|$, $|10\rangle\langle11|$, $|11\rangle\langle10|$ and $|11\rangle\langle11|$ components of $\hat{U}_{\text{CR,eff}}(t,0)$ read

$$\frac{e^{\frac{1}{2}i\omega_{zi}t}\left[\omega_- \cos\left(\frac{\omega_- t}{2}\right) - i(\omega_{iz} - \omega_{zz})\sin\left(\frac{\omega_- t}{2}\right)\right]}{\omega_-}, \quad \text{(C4a)}$$

$$-\frac{e^{\frac{1}{2}i\omega_{zi}t}\left[i(\omega_{ix} - \omega_{zx}) + (\omega_{iy} - \omega_{zy})\right]\sin\left(\frac{\omega_- t}{2}\right)}{\omega_-}, \quad \text{(C4b)}$$

$$-\frac{e^{\frac{1}{2}i\omega_{zi}t}\left[i(\omega_{ix} - \omega_{zx}) - (\omega_{iy} - \omega_{zy})\right]\sin\left(\frac{\omega_- t}{2}\right)}{\omega_-}, \quad \text{(C4c)}$$

$$\frac{e^{\frac{1}{2}i\omega_{zi}t}\left[\omega_- \cos\left(\frac{\omega_- t}{2}\right) + i(\omega_{iz} - \omega_{zz})\sin\left(\frac{\omega_- t}{2}\right)\right]}{\omega_-}. \quad \text{(C4d)}$$

According to Eq. (C2a)–(C4d), to calibrate a direct CNOT gate based on cross-resonance, we set $\omega_{ix}(t) + \omega_{zx}(t) = 0$ and tune $\int_0^{\tau_p} dt'[\omega_{ix}(t') - \omega_{zx}(t')] = \pi$ as a π pulse. These conditions along with perturbative estimates for gate parameters in Appendix B lead to the approximate calibration conditions (6) and (7) of the main text.

Alternatively, we can express the effective unitary in the two-qubit Pauli basis as $u_{\text{eff},\sigma_m\sigma_n}(t,0) \equiv (1/4)\,\text{Tr}\left\{(\hat{\sigma}_m \otimes \hat{\sigma}_n)\hat{U}_{\text{CR,eff}}(t,0)\right\}$. The Pauli decomposition reads

$$u_{\text{eff},ii}(t,0) = \frac{1}{2}\left[e^{\frac{1}{2}i\omega_{zi}t}\cos\left(\frac{\omega_- t}{2}\right) + e^{-\frac{1}{2}i\omega_{zi}t}\cos\left(\frac{\omega_+ t}{2}\right)\right], \quad \text{(C5a)}$$

$$u_{\text{eff},ix}(t,0) = -i\left[\frac{\omega_+(\omega_{ix} - \omega_{zx})e^{\frac{1}{2}i\omega_{zi}t}\sin\left(\frac{\omega_- t}{2}\right)}{2\omega_+\omega_-} + \frac{\omega_-(\omega_{ix} + \omega_{zx})e^{-\frac{1}{2}i\omega_{zi}t}\sin\left(\frac{\omega_+ t}{2}\right)}{2\omega_+\omega_-}\right], \quad \text{(C5b)}$$

$$u_{\text{eff},iy}(t,0) = -i\left[\frac{\omega_+(\omega_{iy} - \omega_{zy})e^{\frac{1}{2}i\omega_{zi}t}\sin\left(\frac{\omega_- t}{2}\right)}{2\omega_+\omega_-} + \frac{\omega_-(\omega_{iy} + \omega_{zy})e^{-\frac{1}{2}i\omega_{zi}t}\sin\left(\frac{\omega_+ t}{2}\right)}{2\omega_+\omega_-}\right], \quad \text{(C5c)}$$

$$u_{\text{eff},iz}(t,0) = -i\left[\frac{\omega_+(\omega_{iz}-\omega_{zz})e^{\frac{1}{2}i\omega_{zi}t}\sin\left(\frac{\omega_-t}{2}\right)}{2\omega_+\omega_-}\right.$$
$$\left.+\frac{\omega_-(\omega_{iz}+\omega_{zz})e^{-\frac{1}{2}i\omega_{zi}t}\sin\left(\frac{\omega_+t}{2}\right)}{2\omega_+\omega_-}\right],$$
(C5d)

$$u_{\text{eff},zi}(t,0) = \frac{1}{2}\left[e^{-\frac{1}{2}i\omega_{zi}t}\cos\left(\frac{\omega_+t}{2}\right)\right.$$
$$\left.-e^{\frac{1}{2}i\omega_{zi}t}\cos\left(\frac{\omega_-t}{2}\right)\right],$$
(C5e)

$$u_{\text{eff},zx}(t,0) = i\left[\frac{\omega_+(\omega_{ix}-\omega_{zx})e^{\frac{1}{2}i\omega_{zi}t}\sin\left(\frac{\omega_-t}{2}\right)}{2\omega_+\omega_-}\right.$$
$$\left.-\frac{\omega_-(\omega_{ix}+\omega_{zx})e^{-\frac{1}{2}i\omega_{zi}t}\sin\left(\frac{\omega_+t}{2}\right)}{2\omega_+\omega_-}\right],$$
(C5f)

$$u_{\text{eff},zy}(t,0) = i\left[\frac{\omega_+(\omega_{iy}-\omega_{zy})e^{\frac{1}{2}i\omega_{zi}t}\sin\left(\frac{\omega_-t}{2}\right)}{2\omega_+\omega_-}\right.$$
$$\left.-\frac{\omega_-(\omega_{iy}+\omega_{zy})e^{-\frac{1}{2}i\omega_{zi}t}\sin\left(\frac{\omega_+t}{2}\right)}{2\omega_+\omega_-}\right],$$
(C5g)

$$u_{\text{eff},zz}(t,0) = i\left[\frac{\omega_+(\omega_{iz}-\omega_{zz})e^{\frac{1}{2}i\omega_{zi}t}\sin\left(\frac{\omega_-t}{2}\right)}{2\omega_+\omega_-}\right.$$
$$\left.-\frac{\omega_-(\omega_{iz}+\omega_{zz})e^{-\frac{1}{2}i\omega_{zi}t}\sin\left(\frac{\omega_+t}{2}\right)}{2\omega_+\omega_-}\right].$$
(C5h)

In summary, based on Eqs. (C2a)–(C5h), the time evolution operator in the BD frame consists of beatings between three CR frequencies: ω_+, ω_- and ω_{zi}.

Appendix D: Non-BD terms in overall time evolution operator

The BD decomposition of the effective time evolution operator, found in Appendix C, contains only the effective interactions. As discussed in Appendix A, we can also quantify the off-resonant contributions by mapping the effective unitary operator back to the initial interaction frame as

$$\hat{U}_I(t,0) = \hat{U}_{\text{SW}}(t)\hat{U}_{I,\text{eff}}(t,0)\hat{U}_{\text{SW}}^\dagger(0)\,. \qquad (D1)$$

We then read off $\hat{U}_{\text{CR}}(t,0)$ as the projection of $\hat{U}_I(t,0)$ onto the computational subspace. Upon the frame transformation (D1), both the effective BD and non-BD subspaces of $\hat{U}_I(t,0)$ are renormalized, while the corrections to the BD subspace are higher-order. Here, we provide the lowest-order adiabatic expressions for the non-BD elements of $\hat{U}_{\text{CR}}(t,0)$.

We begin with the lowest-order expression for the XI element of $\hat{U}_{\text{CR}}(t,0)$ as

$$u_{xi}(t,0) = \frac{\Omega_c^*(0)+e^{-i\Delta_{ct}t}\Omega_c^*(t)}{4\Delta_{ct}}e^{\frac{1}{2}i\omega_{zi}t}\cos\left(\frac{\omega_-t}{2}\right)$$
$$-\frac{\Omega_c(0)+e^{i\Delta_{ct}t}\Omega_c(t)}{4\Delta_{ct}}e^{-\frac{1}{2}i\omega_{zi}t}\cos\left(\frac{\omega_+t}{2}\right)$$
$$+i\frac{\dot{\Omega}_c^*(0)-e^{-i\Delta_{ct}t}\dot{\Omega}_c^*(t)}{4\Delta_{ct}^2}e^{\frac{1}{2}i\omega_{zi}t}\cos\left(\frac{\omega_-t}{2}\right)$$
$$+i\frac{\dot{\Omega}_c(0)-e^{i\Delta_{ct}t}\dot{\Omega}_c(t)}{4\Delta_{ct}^2}e^{-\frac{1}{2}i\omega_{zi}t}\cos\left(\frac{\omega_+t}{2}\right),$$
(D2)

truncated up to the first-order derivative $\dot{\Omega}_c(t)$. Compared to the BD part of the time evolution, there exist faster oscillations in terms of qubit-qubit detuning Δ_{ct} on top of effective slower oscillations characterized by ω_{zi} and $\omega_\pm$. Given that the drive amplitude is set to zero at $t=0,\tau_p$, we find that the XI error at $t=\tau_p$ is determined by the last two-terms as

$$u_{xi}(\tau_p,0) = i\frac{\dot{\Omega}_c^*(0)-e^{-i\Delta_{ct}\tau_p}\dot{\Omega}_c^*(\tau_p)}{4\Delta_{ct}^2}e^{\frac{1}{2}i\omega_{zi}\tau_p}\cos\left(\frac{\omega_-\tau_p}{2}\right)$$
$$+i\frac{\dot{\Omega}_c(0)-e^{i\Delta_{ct}\tau_p}\dot{\Omega}_c(\tau_p)}{4\Delta_{ct}^2}e^{-\frac{1}{2}i\omega_{zi}\tau_p}\cos\left(\frac{\omega_+\tau_p}{2}\right).$$
(D3)

Based on Eq. (D3), the lowest-order XI error is determined by $\dot{\Omega}_c(t)|_{t=0,\tau_p}/\Delta_{ct}^2$ and can be mitigated by (i) larger qubit-qubit detuning and (ii) smoother ramps.

Similar expressions can be obtained for other non-BD components. For instance, the XX, XY, and XZ components of $\hat{U}_{\text{CR}}(\tau_p,0)$ read

$$u_{xx}(\tau_p,0) = \frac{(\omega_{ix}-\omega_{zx})\left[\dot{\Omega}_c^*(0)-e^{-i\Delta_{ct}\tau}\dot{\Omega}_c^*(\tau_p)\right]}{4\Delta_{ct}^2\omega_-}$$
$$\times e^{\frac{1}{2}i\omega_{zi}\tau}\sin\left(\frac{\omega_-\tau_p}{2}\right)$$
$$+\frac{(\omega_{ix}+\omega_{zx})\left[\dot{\Omega}_c(0)-e^{i\Delta_{ct}\tau_p}\dot{\Omega}_c(\tau_p)\right]}{4\Delta_{ct}^2\omega_+}$$
$$\times e^{-\frac{1}{2}i\omega_{zi}\tau_p}\sin\left(\frac{\omega_+\tau_p}{2}\right),$$
(D4)

$$u_{xy}(\tau_p, 0) = \frac{(\omega_{iy} - \omega_{zy})\left[\dot{\Omega}_c^*(0) - e^{-i\Delta_{ct}\tau_p}\dot{\Omega}_c^*(\tau_p)\right]}{4\Delta_{ct}^2\omega_-}$$
$$\times e^{\frac{1}{2}i\omega_{zi}\tau_p}\sin\left(\frac{\omega_-\tau_p}{2}\right)$$
$$+ \frac{(\omega_{iy} + \omega_{zy})\left[\dot{\Omega}_c(0) - e^{i\Delta_{ct}\tau_p}\dot{\Omega}_c(\tau_p)\right]}{4\Delta_{ct}^2\omega_+}$$
$$\times e^{-\frac{1}{2}i\omega_{zi}\tau_p}\sin\left(\frac{\omega_+\tau_p}{2}\right) , \tag{D5}$$

$$u_{xz}(\tau_p, 0) = \frac{(\omega_{iz} - \omega_{zz})\left[\dot{\Omega}_c^*(0) - e^{-i\Delta_{ct}\tau_p}\dot{\Omega}_c^*(\tau_p)\right]}{4\Delta_{ct}^2\omega_-}$$
$$\times e^{\frac{1}{2}i\omega_{zi}\tau_p}\sin\left(\frac{\omega_-\tau_p}{2}\right)$$
$$+ \frac{(\omega_{iz} + \omega_{zz})\left[\dot{\Omega}_c(0) - e^{i\Delta_{ct}\tau_p}\dot{\Omega}_c(\tau_p)\right]}{4\Delta_{ct}^2\omega_+}$$
$$\times e^{-\frac{1}{2}i\omega_{zi}\tau}\sin\left(\frac{\omega_+\tau_p}{2}\right) . \tag{D6}$$

Furthermore, the YI, YX, YY and YZ components of $\hat{U}_{\mathrm{CR}}(\tau_p, 0)$ read

$$u_{yi}(\tau_p, 0) = \frac{\dot{\Omega}_c^*(0) + e^{-i\Delta_{ct}\tau_p}\dot{\Omega}_c^*(\tau_p)}{4\Delta_{ct}^2}e^{\frac{1}{2}i\omega_{zi}\tau_p}\cos\left(\frac{\omega_-\tau_p}{2}\right)$$
$$- \frac{\dot{\Omega}_c(0) + e^{i\Delta_{ct}\tau_p}\dot{\Omega}_c(\tau_p)}{4\Delta_{ct}^2}e^{-\frac{1}{2}i\omega_{zi}\tau_p}\cos\left(\frac{\omega_+\tau_p}{2}\right) , \tag{D7}$$

$$u_{yx}(\tau_p, 0) = -i\frac{(\omega_{ix} - \omega_{zx})\left[\dot{\Omega}_c^*(0) + e^{-i\Delta_{ct}\tau_p}\dot{\Omega}_c^*(\tau_p)\right]}{4\Delta_{ct}^2\omega_-}$$
$$\times e^{\frac{1}{2}i\omega_{zi}\tau_p}\sin\left(\frac{\omega_-\tau_p}{2}\right)$$
$$+ i\frac{(\omega_{ix} + \omega_{zx})\left[\dot{\Omega}_c(0) + e^{i\Delta_{ct}\tau_p}\dot{\Omega}_c(\tau_p)\right]}{4\Delta_{ct}^2\omega_+}$$
$$\times e^{-\frac{1}{2}i\omega_{zi}\tau_p}\sin\left(\frac{\omega_+\tau_p}{2}\right) , \tag{D8}$$

$$u_{yy}(\tau_p, 0) = -i\frac{(\omega_{iy} - \omega_{zy})\left[\dot{\Omega}_c^*(0) + e^{-i\Delta_{ct}\tau}\dot{\Omega}_c^*(\tau_p)\right]}{4\Delta_{ct}^2\omega_-}$$
$$\times e^{\frac{1}{2}i\omega_{zi}\tau}\sin\left(\frac{\omega_-\tau_p}{2}\right)$$
$$+ i\frac{(\omega_{iy} + \omega_{zy})\left[\dot{\Omega}_c(0) + e^{i\Delta_{ct}\tau}\dot{\Omega}_c(\tau_p)\right]}{4\Delta_{ct}^2\omega_+}$$
$$\times e^{-\frac{1}{2}i\omega_{zi}\tau_p}\sin\left(\frac{\omega_+\tau_p}{2}\right) , \tag{D9}$$

$$u_{yz}(\tau_p, 0) = -i\frac{(\omega_{iz} - \omega_{zz})\left[\dot{\Omega}_c^*(0) + e^{-i\Delta_{ct}\tau_p}\dot{\Omega}_c^*(\tau_p)\right]}{4\Delta_{ct}^2\omega_-}$$
$$\times e^{\frac{1}{2}i\omega_{zi}\tau_p}\sin\left(\frac{\omega_-\tau_p}{2}\right)$$
$$+ i\frac{(\omega_{iz} + \omega_{zz})\left[\dot{\Omega}_c(0) + e^{i\Delta_{ct}\tau}\dot{\Omega}_c(\tau_p)\right]}{4\Delta_{ct}^2\omega_+}$$
$$\times e^{-\frac{1}{2}i\omega_{zi}\tau_p}\sin\left(\frac{\omega_+\tau_p}{2}\right) . \tag{D10}$$

Equations (D3)–(D10) are the main results of this appendix. They provide the leading-order correction, i.e. up to the first-order derivative of the pulse, to the non-BD subspace of the time evolution operator. Importantly, on top of the effective frequencies that appear in the BD subspace, characterized by $\omega_\pm$ and ω_{zi}, there exists higher-frequency oscillation in the non-BD subspace that is set by the qubit-qubit detuning Δ_{ct}.

Appendix E: Off-resonantly driven transmon

In this appendix, we consider an off-resonantly driven transmon qubit as a simpler model that still captures the main physics of off-resonant error on the control qubit. In comparison to the energy diagram in Fig. 1, this corresponds to one of the vertical ladders consisting of the control qubit states. First, in Sec. E 1, we derive the leading-order overlap integrals presented in Table I. Second, in Sec. E 2, we derive leading-order analytical DRAG solutions that minimize each specific error type.

1. Derivation of overlap integrals

In the rotating frame of the drive, the system and drive Hamiltonian for an off-resonantly driven transmon can be approximated using a Kerr model as

$$\hat{\mathcal{H}}_q \equiv \Delta_{cd}\hat{b}_c^\dagger\hat{b}_c + \frac{\alpha_c}{2}\hat{b}_c^\dagger\hat{b}_c^\dagger\hat{b}_c\hat{b}_c , \tag{E1}$$

$$\hat{\mathcal{H}}_d(t) \equiv \frac{1}{2}\left[\Omega_c^*(t)\hat{b}_c + \Omega_c(t)\hat{b}_c^\dagger\right] , \tag{E2}$$

where $\Delta_{cd} \equiv \omega_c - \omega_d$ is the qubit-drive detuning. The interaction-frame Hamiltonian is then found as

$$\hat{\mathcal{H}}_I(t) \equiv e^{i\hat{\mathcal{H}}_q t}\hat{\mathcal{H}}_d(t)e^{-i\hat{\mathcal{H}}_q t}$$
$$= \sum_{n_c=1}^{d}\left[\frac{\sqrt{n_c}}{2}\Omega_c(t)e^{i\Delta_{nd}t}|n_c\rangle\langle n_c - 1| + \mathrm{H.c.}\right] , \tag{E3}$$

with level-dependent detuning $\Delta_{nd} \equiv \Delta_{cd} + (n_c - 1)\alpha_c$ and level cut-off d. Assuming the detuning lies in the straddling regime, and in the drive range relevant to CR, the leakage to levels beyond the second excited state is

typically negligible. Hence, a three-level model is sufficient for quantifying leading-order transition probabilities $|0_c\rangle \rightarrow |1_c\rangle$, $|0_c\rangle \rightarrow |2_c\rangle$ and $|1_c\rangle \rightarrow |2_c\rangle$.

To solve for the time evolution operator, and the resulting overlap integrals, we follow the Magnus method up to the second order, while the same result can also be obtained via time-dependent SWPT. A perturbative expansion of the time evolution operator $\hat{U}_I(\tau_p, 0) \equiv \exp[-i\hat{K}(\tau_p, 0)]$ in terms of the generator $\hat{K}(\tau_p, 0)$ yields

$$
\begin{aligned}
\hat{U}_I(\tau_p, 0) = &\hat{I} - i\hat{K}_1(\tau_p, 0) \\
&- i\hat{K}_2(\tau_p, 0) - \frac{1}{2}\hat{K}_1^2(\tau_p, 0) + O(\hat{\mathcal{H}}_I^3) ,
\end{aligned}
\tag{E4a}
$$

where $\hat{K}_1(\tau_p, 0)$ and $\hat{K}_2(\tau_p, 0)$ read [30]

$$
\hat{K}_1(\tau_p, 0) = \int_0^{\tau_p} dt' \hat{\mathcal{H}}_I(t') ,
\tag{E4b}
$$

$$
\hat{K}_2(\tau_p, 0) = -\frac{i}{2}\int_0^{\tau_p} dt' \int_0^{t'} dt'' [\hat{\mathcal{H}}_I(t'), \hat{\mathcal{H}}_I(t'')] .
\tag{E4c}
$$

In Eq. (E4a), first-order (single-photon) transitions occur via $-i\hat{K}_1(\tau_p, 0)$, while 2nd order (two-photon) transitions occur via both $-i\hat{K}_2(\tau_p, 0)$ and $-(1/2)\hat{K}_1^2(\tau_p, 0)$. The two contributions up to the 2nd order, however, describe different physical processes: $-(1/2)\hat{K}_1^2(\tau_p, 0)$ quantifies a two-photon spectral overlap as a product of individual single-photon overlaps, while $-i\hat{K}_2(\tau_p, 0)$ quantifies a two-time (two-frequency) spectral overlap as shown in the following.

We begin by single-photon transitions. The transition probability of $|0_c\rangle \rightarrow |1_c\rangle$ is found as

$$
\begin{aligned}
P_{|0_c\rangle \rightarrow |1_c\rangle} &= \left| \langle 1_c| \hat{U}_I(\tau_p, 0) |0_c\rangle \right|^2 \\
&= \left| -i \langle 1_c| \hat{K}_1(\tau_p, 0) |0_c\rangle + O(\hat{\mathcal{H}}_I^3) \right|^2 \\
&= \left| -i \int_0^{\tau_p} dt' \langle 1_c| \hat{\mathcal{H}}_I(t') |0_c\rangle + O(\hat{\mathcal{H}}_I^3) \right|^2
\end{aligned}
\tag{E5}
$$

Replacing $\langle 1_c| \hat{\mathcal{H}}_I(t') |0_c\rangle = (1/2)\Omega_c(t')e^{i\Delta_{cd}t'}$ from Eq. (E3) gives

$$
\begin{aligned}
P_{|0_c\rangle \rightarrow |1_c\rangle} &\approx \frac{1}{4}\left| \int_0^{\tau_p} dt' \Omega_c(t')e^{i\Delta_{cd}t'} \right|^2 \\
&= \frac{1}{4}\left| \int_{-\infty}^{+\infty} \frac{d\omega'}{2\pi} \frac{\tilde{\Omega}_c(\omega')}{\omega' + \Delta_{cd}} \left[e^{i(\omega'+\Delta_{cd})\tau_p} - 1 \right] \right|^2 .
\end{aligned}
\tag{E6}
$$

Equation (E6) is the leading-order measure for non-BD error in the CR gate. It shows that the transition is run by a sideband photon as a result of spectral overlap between the pulse and qubit-drive detuning. Similarly,

the transition probability of $|1_c\rangle \rightarrow |2_c\rangle$ reads

$$
\begin{aligned}
P_{|1_c\rangle \rightarrow |2_c\rangle} &\approx \frac{1}{2}\left| \int_0^{\tau_p} dt' \Omega_c(t')e^{i(\Delta_{cd}+\alpha_c)t'} \right|^2 \\
&= \frac{1}{2}\left| \int_{-\infty}^{+\infty} \frac{d\omega'}{2\pi} \frac{\tilde{\Omega}_c(\omega')}{\omega' + \Delta_{cd} + \alpha_c} \left[e^{i(\omega'+\Delta_{cd}+\alpha_c)\tau_p} - 1 \right] \right|^2 ,
\end{aligned}
\tag{E7}
$$

where the distinct prefactor and transition frequency come from $\langle 2_c| \hat{\mathcal{H}}_I(t') |1_c\rangle = (\sqrt{2}/2)\Omega_c(t')e^{i(\Delta_{cd}+\alpha_c)t'}$.

Two-photon transition probability $|0_c\rangle \rightarrow |2_c\rangle$ is obtained as

$$
\begin{aligned}
P_{|0_c\rangle \rightarrow |2_c\rangle} &= \left| \langle 2_c| \hat{U}_I(\tau_p, 0) |0_c\rangle \right|^2 \\
&= \left| -\frac{1}{2} \langle 2_c| \hat{K}_1^2(\tau_p, 0) |0_c\rangle \right. \\
&\quad \left. - i \langle 2_c| \hat{K}_2(\tau_p, 0) |0_c\rangle + O(\hat{\mathcal{H}}_I^3) \right|^2 ,
\end{aligned}
\tag{E8}
$$

where contributions from $-\frac{1}{2}\hat{K}_1^2(\tau_p, 0)$ and $-i\hat{K}_2(\tau_p, 0)$ read

$$
\begin{aligned}
-\frac{1}{2} \langle 2_c| \hat{K}_1^2(\tau_p, 0) |0_c\rangle = &-\frac{\sqrt{2}}{8}\left[\int_0^{\tau_p} dt' \Omega_c(t')e^{i\Delta_{cd}t'} \right] \\
&\times \left[\int_0^{\tau_p} dt' \Omega_c(t')e^{i(\Delta_{cd}+\alpha_c)t'} \right] ,
\end{aligned}
\tag{E9}
$$

$$
\begin{aligned}
-i \langle 2_c| \hat{K}_2(\tau_p, 0) |0_c\rangle = &-\frac{\sqrt{2}}{8}\int_0^{\tau_p} dt' \int_0^{t'} dt'' \Omega_c(t')\Omega_c(t'') \\
&\left[e^{i(\Delta_{ct}+\alpha_c)t'}e^{i\Delta_{ct}t''} - e^{i\Delta_{ct}t'}e^{i(\Delta_{cd}+\alpha_c)t''} \right] .
\end{aligned}
\tag{E10}
$$

Equation (E9) is a product of simultaneous overlaps with Δ_{cd} and $\Delta_{cd} + \alpha_c$. Depending on frequency allocation, either one or both overlaps are small and hence their product cannot grow too large. However, Eq. (E10) contains higher-order correlations leading to spectral overlap with the $|0_c\rangle \rightarrow |2_c\rangle$ transition frequency $2\Delta_{cd} + \alpha_c$:

$$
\begin{aligned}
P_{|0_c\rangle \rightarrow |2_c\rangle} \approx \frac{1}{32}\left| \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{d\omega'}{2\pi}\frac{d\omega''}{2\pi}\alpha_c\tilde{\Omega}_c(\omega')\tilde{\Omega}_c(\omega'') \right. \\
\left. \frac{\left[e^{i(\omega'+\omega''+2\Delta_{ct}+\alpha_c)\tau_p} - 1 \right]}{(\omega'' + \Delta_{ct} + \alpha_c)(\omega'' + \Delta_{ct})(\omega' + \omega'' + 2\Delta_{ct} + \alpha_c)} \right|^2 .
\end{aligned}
\tag{E11}
$$

Equation (E11) is an approximate form for the frequency-domain representation of Eq. (E10). Here, where we kept only the dominant two-photon contribution in which the transition is excited by two sideband photons with total energy $\omega' + \omega'' = -(2\Delta_{cd} + \alpha_c)$. The overlap grows when the two-photon gap is small, i.e. close to the collision at $\Delta_{cd} = -\alpha_c/2$.

2. Derivation of DRAG conditions

We next derive leading-order DRAG solutions to suppress single-photon transitions discussed in Eqs. (E6) and (E7). The derivation here is slightly distinct compared to Refs. [23, 33] and done in two consecutive steps: we first perform perturbative expansion in the drive amplitude $\Omega_c(t)$ and then an adiabatic expansion in $\frac{1}{\Delta}\frac{d}{dt}$ allowing for more flexibility. Nevertheless, we recover the solution in Ref. [33] as a special case.

To this aim, we consider the following leading-order Y-DRAG ansatz

$$\Omega_c(t) = \Omega_{\rm SG}(t) + \frac{i}{\Delta_D}\dot{\Omega}_{\rm SG}(t) , \qquad \text{(E12)}$$

where the DRAG coefficient Δ_D needs to be determined such that a specific off-resonant overlap is suppressed.

Here, the main pulse is taken as square Gaussian, but the DRAG condition is in principle independent of this choice. Performing an adiabtic expansion on $|0_c\rangle \to |1_c\rangle$ probability in Eq. (E6) results in

$$
\begin{aligned}
P_{|0_c\rangle \to |1_c\rangle} &\approx \frac{1}{4}\left|\int_0^{\tau_p} dt'\Omega_c(t')e^{i\Delta_{cd}t'}\right|^2 \\
&= \frac{1}{4}\left|\left\{\sum_{n=0}^{\infty}\left[\frac{1}{\Delta_{cd}}\left(\frac{i}{\Delta_{cd}}\frac{d}{dt'}\right)^n\Omega_c(t')\right]e^{i\Delta_{cd}t'}\right\}\Big|_0^{\tau_p}\right|^2 .
\end{aligned}
$$
(E13)

We then substitute the DRAG ansatz (E12) into Eq. (E13) and read off contributions at $t = 0, \tau_p$ as a measure for off-resonant error. In terms of the normalized DRAG parameter $\lambda_{01} \equiv \Delta_{cd}/\Delta_D$, and given that $\Omega_{\rm SG}(0) = \Omega_{\rm SG}(\tau_p) = 0$, we find

$$
\begin{aligned}
P_{|0_c\rangle \to |1_c\rangle} ={}& \frac{(1+\lambda_{01})^2}{4\Delta_{cd}^4}\left[\dot{\Omega}_{\rm SG}^2(\tau_p) + \dot{\Omega}_{\rm SG}^2(0) - 2\cos(\Delta_{cd}\tau_p)\dot{\Omega}_{\rm SG}(\tau_p)\dot{\Omega}_{\rm SG}(0)\right] \\
&+ \frac{(1+\lambda_{01})^2}{4\Delta_{cd}^6}\left[\ddot{\Omega}_{\rm SG}^2(\tau_p) + \ddot{\Omega}_{\rm SG}^2(0) - 2\cos(\Delta_{cd}\tau_p)\ddot{\Omega}_{\rm SG}(\tau_p)\ddot{\Omega}_{\rm SG}(0)\right] \\
&+ \frac{\lambda_{01}^2}{4\Delta_{cd}^8}\left[\dddot{\Omega}_{\rm SG}^2(\tau_p) + \dddot{\Omega}_{\rm SG}^2(0) - 2\cos(\Delta_{cd}\tau_p)\dddot{\Omega}_{\rm SG}(\tau_p)\dddot{\Omega}_{\rm SG}(0)\right] \\
&- \frac{\lambda_{01}(1+\lambda_{01})}{2\Delta_{cd}^6}\left[\dddot{\Omega}_{\rm SG}(\tau_p)\dot{\Omega}_{\rm SG}(\tau_p) + \dddot{\Omega}_{\rm SG}(0)\dot{\Omega}_{\rm SG}(0)\right. \\
&\left. - \cos(\Delta_{cd}\tau_p)\dddot{\Omega}_{\rm SG}(\tau_p)\dot{\Omega}_{\rm SG}(0) - \cos(\Delta_{cd}\tau_p)\dddot{\Omega}_{\rm SG}(0)\dot{\Omega}_{\rm SG}(\tau_p)\right] + O\left(\frac{d^4}{dt^4}\Omega_{\rm SG}(t)\Big|_{t=0,\tau_p}\right) .
\end{aligned}
$$
(E14)

Based on Eq. (E14), up to the leading order, the optimal DRAG parameter λ_{01} is determined as the roots of a 2nd order polynomal, whose coefficients are generally determined by the pulse spectrum (derivatives), gate time τ_p and the transition frequency Δ_{cd}. A special solution, however, is found as $\lambda_{01} = -1$, i.e. $\Delta_D = -\Delta_{ct}$, which sets the 1st, 2nd and the 4th term in Eq. (E14) to zero resulting in residual error in terms of $\ddot{\Omega}_{\rm SG}(t)$ as

$$\frac{\left[\dddot{\Omega}_{\rm SG}^2(\tau_p) + \dddot{\Omega}_{\rm SG}^2(0) - 2\cos(\Delta_{cd}\tau_p)\dddot{\Omega}_{\rm SG}(\tau_p)\dddot{\Omega}_{\rm SG}(0)\right]}{4\Delta_{cd}^8} .$$
(E15)

This DRAG choice corresponds to X and Y control pulses $\Omega_{cx}(t) = \Omega_{\rm SG}(t)$ and $\Omega_{cy}(t) = -(1/\Delta_{cd})\dot{\Omega}_{\rm SG}(t)$.

Adiabatic expansion of the $|1_c\rangle \to |2_c\rangle$ transition prob-

ability has a similar form:

$$
\begin{aligned}
P_{|1_c\rangle \to |2_c\rangle} &\approx \frac{1}{2}\left|\int_0^{\tau_p} dt'\Omega_c(t')e^{i(\Delta_{cd}+\alpha_c)t'}\right|^2 \\
&= \frac{1}{2}\left|\left\{\sum_{n=0}^{\infty}\left[\frac{1}{\Delta_{cd}+\alpha_c}\left(\frac{i}{\Delta_{cd}+\alpha_c}\frac{d}{dt'}\right)^n\Omega_c(t')\right]\right.\right. \\
&\qquad\qquad \left.\left. \times e^{i(\Delta_{cd}+\alpha_c)t'}\right\}\Big|_0^{\tau_p}\right|^2 ,
\end{aligned}
$$
(E16)

where the transition frequency is replaced by $\Delta_{cd}+\alpha_c$. In terms of $\lambda_{12} \equiv (\Delta_{cd}+\alpha_c)/\Delta_D$, the derivative expansion takes the same form as in Eq. (E14) and a possible DRAG solution is $\Omega_{cx}(t) = \Omega_{\rm SG}(t)$ and $\Omega_{cy}(t) = -1/(\Delta_{cd}+\alpha_c)\dot{\Omega}_{\rm SG}(t)$. In the resonant scenario where $\Delta_{cd} = 0$, this solution recovers that of Ref. [33] for single-qubit gates.

We note that the same method of adiabatic expansion can also be applied to the $|0_c\rangle \to |2_c\rangle$ probability in Eqs. (E8)–(E10). However, the computation is much more involved, and the leading-order DRAG condition is determined as the roots of a fourth-order polynomial in

$1/\Delta_D$. Therefore, we resort to numerical optimization as

discussed in Sec. IV and Fig. 7.

[1] GS Paraoanu. "microwave-induced coupling of superconducting qubits'. *Physical Review B*, 74(14):140504, 2006.

[2] Chad Rigetti and Michel Devoret. "fully microwave-tunable universal gates in superconducting qubits with linear couplings and fixed transition frequencies". *Physical Review B*, 81(13):134507, 2010.

[3] Yu Nakamura, Yu A Pashkin, and JS Tsai. "coherent control of macroscopic quantum states in a single-cooper-pair box". *Nature*, 398(6730):786–788, 1999.

[4] Andreas Wallraff, David I Schuster, Alexandre Blais, Luigi Frunzio, R-S Huang, Johannes Majer, Sameer Kumar, Steven M Girvin, and Robert J Schoelkopf. "strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics". *Nature*, 431(7005):162–167, 2004.

[5] Jens Koch, Terri M Yu, Jay M Gambetta, Andrew A Houck, David I Schuster, Joseph Majer, Alexandre Blais, Michel H Devoret, Steven M Girvin, and Robert J Schoelkopf. "charge-insensitive qubit design derived from the Cooper pair box". *Phys. Rev. A*, 76(4):042319, October 2007.

[6] John Clarke and Frank K. Wilhelm. Superconducting quantum bits. *Nature*, 453(7198):1031–1042, June 2008.

[7] Sarah Sheldon, Easwar Magesan, Jerry M Chow, and Jay M Gambetta. "procedure for systematically tuning up cross-talk in the cross-resonance gate". *Physical Review A*, 93(6):060302, 2016.

[8] Easwar Magesan and Jay M. Gambetta. "effective hamiltonian models of the cross-resonance gate". *Phys. Rev. A*, 101:052308, May 2020.

[9] Susanna Kirchhoff, Torsten Keßler, Per J Liebermann, Elie Assémat, Shai Machnes, Felix Motzoi, and Frank K Wilhelm. "optimized cross-resonance gate for coupled transmon systems". *Physical Review A*, 97(4):042348, 2018.

[10] Vinay Tripathi, Mostafa Khezri, and Alexander N. Korotkov. "operation and intrinsic error budget of a two-qubit cross-resonance gate". *Phys. Rev. A*, 100:012301, Jul 2019.

[11] Moein Malekakhlagh, Easwar Magesan, and David C. McKay. "first-principles analysis of cross-resonance gate operation". *Phys. Rev. A*, 102:042605, Oct 2020.

[12] Neereja Sundaresan, Isaac Lauer, Emily Pritchett, Easwar Magesan, Petar Jurcevic, and Jay M Gambetta. "reducing unitary and spectator errors in cross resonance with optimized rotary echoes". *PRX Quantum*, 1(2):020318, 2020.

[13] Kentaro Heya and Naoki Kanazawa. "cross cross resonance gate". *arXiv preprint arXiv:2103.00024*, 2021.

[14] Andrew W Cross, Lev S Bishop, Sarah Sheldon, Paul D Nation, and Jay M Gambetta. "validating quantum computers using randomized model circuits". *Physical Review A*, 100(3):032328, 2019.

[15] Petar Jurcevic, Ali Javadi-Abhari, Lev S Bishop, Isaac Lauer, Daniela F Bogorin, Markus Brink, Lauren Capelluto, Oktay Günlük, Toshinari Itoko, Naoki Kanazawa, et al. "demonstration of quantum volume 64 on a superconducting quantum computing system". *Quantum Science and Technology*, 6(2):025020, 2021.

[16] Christopher J. Wood and Jay M. Gambetta. "quantification and characterization of leakage errors". *Phys. Rev. A*, 97:032306, Mar 2018.

[17] Abhinav Kandala, Ken X. Wei, Srikanth Srinivasan, Easwar Magesan, Santino Carnevale, George A. Keefe, David Klaus, Oliver Dial, and David C. McKay. "demonstration of a high-fidelity cnot for fixed-frequency transmons with engineered zz suppression". *arXiv preprint quant-ph/2011.07050*, 2020.

[18] Pranav Mundada, Gengyan Zhang, Thomas Hazard, and Andrew Houck. "suppression of qubit crosstalk in a tunable coupling superconducting circuit". *Physical Review Applied*, 12(5):054023, 2019.

[19] David C McKay, Christopher J Wood, Sarah Sheldon, Jerry M Chow, and Jay M Gambetta. "efficient z gates for quantum computing". *Phys. Rev. A*, 96:022330, Aug 2017.

[20] John R Schrieffer and Peter A Wolff. "relation between the anderson and kondo hamiltonians". *Physical Review*, 149(2):491, 1966.

[21] Maxime Boissonneault, Jay M Gambetta, and Alexandre Blais. "dispersive regime of circuit qed: Photon-dependent qubit dephasing and relaxation rates". *Phys. Rev. A*, 79:013819, Jan 2009.

[22] Sergey Bravyi, David P DiVincenzo, and Daniel Loss. "schrieffer–wolff transformation for quantum many-body systems". *Annals of physics*, 326(10):2793–2826, 2011.

[23] Jay M Gambetta, F Motzoi, ST Merkel, and Frank K Wilhelm. "analytic control methods for high-fidelity unitary operations in a weakly nonlinear oscillator". *Physical Review A*, 83(1):012308, 2011.

[24] Moein Malekakhlagh, Alexandru Petrescu, and Hakan E. Türeci. "lifetime renormalization of weakly anharmonic superconducting qubits. I. role of number nonconserving terms". *Phys. Rev. B*, 101:134509, Apr 2020.

[25] Alexandru Petrescu, Moein Malekakhlagh, and Hakan E. Türeci. "lifetime renormalization of driven weakly anharmonic superconducting qubits. II. the readout problem". *Phys. Rev. B*, 101:134510, Apr 2020.

[26] Alexandru Petrescu, Camille Le Calonnec, Catherine Leroux, Agustin Di Paolo, Pranav Mundada, Sara Sussman, Andrei Vrajitoarea, Andrew A Houck, and Alexandre Blais. "accurate methods for the analysis of strong-drive effects in parametric gates". *arXiv preprint arXiv:2107.02343*, 2021.

[27] Moein Malekakhlagh, William Shanks, and Hanhee Paik. "optimization of the resonator-induced phase gate for superconducting qubits". *arXiv preprint arXiv:2110.01724*, 2021.

[28] Wilhelm Magnus. "on the exponential solution of differential equations for a linear operator". *Communications on pure and applied mathematics*, 7(4):649–673, 1954.

[29] Sergio Blanes, Fernando Casas, Jose-Angel Oteo, and José Ros. "the magnus expansion and some of its applications". *Physics reports*, 470(5-6):151–238, 2009.

[30] Sergio Blanes, Fernando Casas, JA Oteo, and J Ros. "a pedagogical approach to the magnus expansion". *Euro-*

pean journal of physics, 31(4):907, 2010.

[31] Ernst Hairer, Marlis Hochbruck, Arieh Iserles, and Christian Lubich. "geometric numerical integration". *Oberwolfach Reports*, 3(1):805–882, 2006.

[32] Jerry M Chow, L DiCarlo, Jay M Gambetta, F Motzoi, L Frunzio, Steven M Girvin, and Robert J Schoelkopf. "optimized driving of superconducting artificial atoms for improved single-qubit gates". *Physical Review A*, 82(4):040305, 2010.

[33] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm. "simple pulses for elimination of leakage in weakly nonlinear qubits". *Phys. Rev. Lett.*, 103:110501, Sep 2009.

[34] R Schutjens, F Abu Dagga, DJ Egger, and FK Wilhelm. "single-qubit gates in frequency-crowded transmon systems". *Physical Review A*, 88(5):052330, 2013.

[35] S. H. W. van der Ploeg, A. Izmalkov, Alec Maassen van den Brink, U. Hübner, M. Grajcar, E. Il'ichev, H.-G. Meyer, and A. M. Zagoskin. "controllable coupling of superconducting flux qubits". *Phys. Rev. Lett.*, 98:057004, Feb 2007.

[36] Yu Chen, C Neill, P Roushan, N Leung, M Fang, R Barends, J Kelly, B Campbell, Z Chen, B Chiaro, et al. "qubit architecture with high coherence and fast tunable coupling". *Physical review letters*, 113(22):220502, 2014.

[37] David C. McKay, Ravi Naik, Philip Reinhold, Lev S. Bishop, and David I. Schuster. "high-contrast qubit interactions using multimode cavity qed". *Phys. Rev. Lett.*, 114:080501, Feb 2015.

[38] J Stehlik, DM Zajac, DL Underwood, T Phung, J Blair, S Carnevale, D Klaus, GA Keefe, A Carniol, M Kumph, et al. "tunable coupling architecture for fixed-frequency transmons". *arXiv preprint arXiv:2101.07746*, 2021.

[39] Jaseung Ku, Xuexin Xu, Markus Brink, David C. McKay, Jared B. Hertzberg, Mohammad H. Ansari, and B. L. T. Plourde. "suppression of unwanted zz interactions in a hybrid two-qubit system". *Phys. Rev. Lett.*, 125:200504, Nov 2020.

[40] Peng Zhao, Peng Xu, Dong Lan, Ji Chu, Xinsheng Tan, Haifeng Yu, and Yang Yu. "high-contrast zz interaction using superconducting qubits with opposite-sign anharmonicity". *arXiv preprint arXiv:2002.07560*, 2020.

[41] Fei Yan, Philip Krantz, Youngkyu Sung, Morten Kjaergaard, Daniel L Campbell, Terry P Orlando, Simon Gustavsson, and William D Oliver. "tunable coupling scheme for implementing high-fidelity two-qubit gates". *Physical Review Applied*, 10(5):054062, 2018.

[42] Youngkyu Sung, Leon Ding, Jochen Braumüller, Antti Vepsäläinen, Bharath Kannan, Morten Kjaergaard, Ami Greene, Gabriel O Samach, Chris McNally, David Kim, et al. "realization of high-fidelity cz and zz-free

[43] Michele C Collodo, Johannes Herrmann, Nathan Lacroix, Christian Kraglund Andersen, Ants Remm, Stefania Lazar, Jean-Claude Besse, Theo Walter, Andreas Wallraff, and Christopher Eichler. "implementation of conditional phase gates based on tunable zz interactions". *Physical Review Letters*, 125(24):240502, 2020.

[44] Yuan Xu, Ji Chu, Jiahao Yuan, Jiawei Qiu, Yuxuan Zhou, Libo Zhang, Xinsheng Tan, Yang Yu, Song Liu, Jian Li, et al. "high-fidelity, high-scalability two-qubit gate scheme for superconducting qubits. *Physical Review Letters*, 125(24):240503, 2020.

[45] KX Wei, E Magesan, I Lauer, S Srinivasan, DF Bogorin, S Carnevale, GA Keefe, Y Kim, D Klaus, W Landers, et al. "quantum crosstalk cancellation for fast entangling gates and improved multi-qubit performance". *arXiv preprint arXiv:2106.00675*, 2021.

[46] Bradley K Mitchell, Ravi K Naik, Alexis Morvan, Akel Hashim, John Mark Kreikebaum, Brian Marinelli, Wim Lavrijsen, Kasra Nowrouzi, David I Santiago, and Irfan Siddiqi. "hardware-efficient microwave-activated tunable coupling between superconducting qubits". *arXiv preprint arXiv:2105.05384*, 2021.

[47] Line Hjortshøj Pedersen, Niels Martin Møller, and Klaus Mølmer. "fidelity of quantum operations". *Physics Letters A*, 367(1-2):47–51, 2007.

[48] Easwar Magesan, J. M. Gambetta, and Joseph Emerson. "scalable and robust randomized benchmarking of quantum processes". *Phys. Rev. Lett.*, 106:180504, May 2011.

[49] *ibmq_peekskill* is a 27-qubit processor based on Falcon r8 design (see `https://quantum-computing.ibm.com/services?services=systems&system=ibm_peekskill`).

[50] Jared B Hertzberg, Eric J Zhang, Sami Rosenblatt, Easwar Magesan, John A Smolin, Jeng-Bang Yau, Vivekananda P Adiga, Martin Sandberg, Markus Brink, Jerry M Chow, et al. "laser-annealing josephson junctions for yielding scaled-up superconducting quantum processors". *arXiv preprint arXiv:2009.00781*, 2020.

[51] Eric J Zhang, Srikanth Srinivasan, Neereja Sundaresan, Daniela F Bogorin, Yves Martin, Jared B Hertzberg, John Timmerwilke, Emily J Pritchett, Jeng-Bang Yau, Cindy Wang, et al. "high-fidelity superconducting quantum processors via laser-annealing of transmon qubits". *arXiv preprint arXiv:2012.08475*, 2020.

[52] Easwar Magesan, Jay M Gambetta, Blake R Johnson, Colm A Ryan, Jerry M Chow, Seth T Merkel, Marcus P Da Silva, George A Keefe, Mary B Rothwell, Thomas A Ohki, et al. "efficient measurement of quantum gate error by interleaved randomized benchmarking". *Physical review letters*, 109(8):080505, 2012.