npj | quantum information

Article

Published in partnership with The University of New South Wales

https://doi.org/10.1038/s41534-024-00863-4

Experimental error suppression in Cross-Resonance gates via multi-derivative pulse shaping

Check for updates

Boxi Li [1,2] ✉, Tommaso Calarco[1,2,3] & Felix Motzoi [1,2] ✉

While quantum circuits are reaching impressive widths in the hundreds of qubits, their depths have not been able to keep pace. In particular, cloud computing gates on multi-qubit, fixed-frequency superconducting chips continue to hover around the 1% error range, contrasting with the progress seen on carefully designed two-qubit chips, where error rates have been pushed towards 0.1%. Despite the strong impetus and a plethora of research, experimental demonstration of error suppression on these multi-qubit devices remains challenging, primarily due to the wide distribution of qubit parameters and the demanding calibration process required for advanced control methods. Here, we achieve this goal, using a simple control method based on multi-derivative, multi-constraint pulse shaping, which acts simultaneously against multiple error sources. Our approach establishes a two to fourfold improvement on the default calibration scheme, demonstrated on four qubits on the IBM Quantum Platform with limited and intermittent access, enabling these large-scale fixed-frequency systems to fully take advantage of their superior coherence times. The achieved CNOT fidelities of 99.7(1)% on those publically available qubits come from both coherent control error suppression and accelerated gate time.

Superconducting qubits have experienced significant improvement in the last decade, reaching the error correction threshold[1,2] and been used to study nontrivial quantum phenomena[3,4]. Additionally, quantum devices have become more accessible outside research labs, with cloud-based platforms like the IBM Quantum platform[5] providing access to multi-qubit devices, on which near-term quantum computing applications with error mitigation have been demonstrated[6]. Although very high gate fidelity has been achieved on isolated chips[7,8], gate performance on scalable, publicly available multi-qubit devices is still bottlenecked, especially for two-qubit operations[9]. These control imperfections not only limit the fidelity and depth of quantum circuits, but also give rise to correlated errors that propagate across the qubit lattice, sabotaging quantum error correction[10], making error mitigation and benchmarking more challenging[11–14].

The Cross-Resonance (CR) gate is one of the most widely used two-qubit entangling gates for superconducting qubits, using microwave controls and avoiding noisy flux lines[15–19]. It is the default gate on most devices provided by IBM and has found applications in high-quality circuit implementation, parity measurement, and state preparation[7,20,21]. While the absence of flux control lines extends qubit coherence time, it limits qubit tunability and necessitates weak coupling between qubits. Consequently, achieving fast two-qubit gates requires a strong drive, which often leads to coherent errors due to non-adiabatic dynamics. In practice, limiting the drive amplitude and a long pulse ramping time are used to prevent undesired dynamics, including off-resonant transitions introduced by the drive[22–24] and unwanted dynamics in the effective qubits' subspace[23,25,26] (see Fig. 1a, b).

To circumvent control errors while maintaining substantial coupling strength, a combination of careful qubit parameter engineering and advanced control schemes has been employed. With these techniques, the best CR gate infidelity reported lies between 0.1% and 0.3%[7,8]. However, extending these advancements to scalable multi-qubit devices proves challenging. For instance, on the 127-qubit chip *ibm_brisbane*, the best echoed CR gate has an error of 0.35%, while the median is only around 0.8%, considerably higher than their coherence limit (median T1 ≈ 200 μs and T2 ≈ 135 μs)[5]. An important factor contributing to the challenge lies in the intentional distribution of qubit parameters over a wide range, a design choice aimed at mitigating cross-talk[27–29]. This uncertainty in the qubit parameters also stems unavoidably from the inhomogeneity in the

[1]Forschungszentrum Jülich, Institute of Quantum Control (PGI-8), D-52425 Jülich, Germany. [2]Institute for Theoretical Physics, University of Cologne, D-50937 Cologne, Germany. [3]Dipartimento di Fisica e Astronomia, Università di Bologna, 40127 Bologna, Italy. ✉e-mail: b.li@fz-juelich.de; f.motzoi@fz-juelich.de

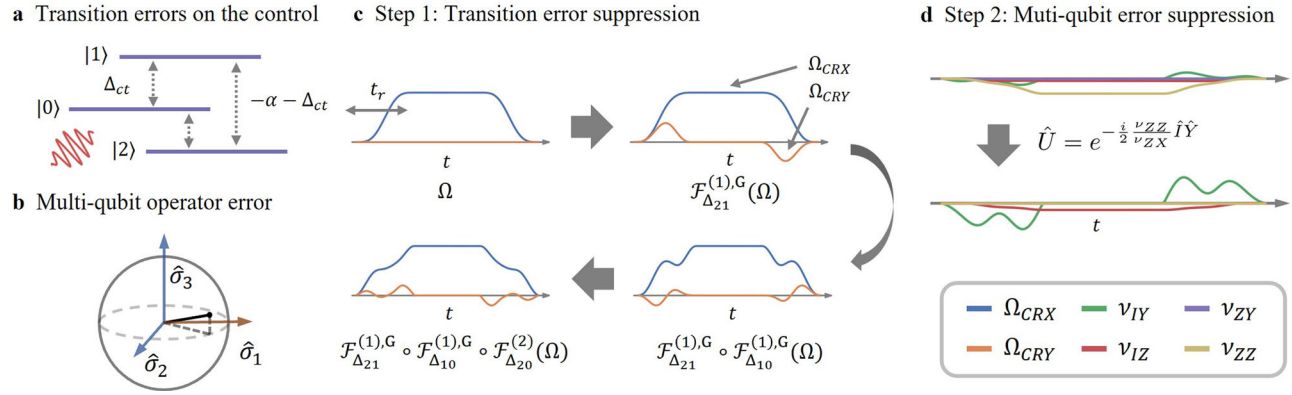

Fig. 1 | Illustration of the dominant coherent errors in the Cross-Resonance gate and the proposed pulse schemes to correct them. a Transition errors in the rotating frame for a Transmon qubit driven off-resonantly, with Δ_{ct} the detuning and α the anharmonicity. **b** Two-qubit errors of the CR gate. The three axes $\{\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3\}$ represent either $\{\hat{\sigma}_{ZX}, \hat{\sigma}_{ZY}, \hat{\sigma}_{ZZ}\}$ or $\{\hat{\sigma}_{IX}, \hat{\sigma}_{IY}, \hat{\sigma}_{IZ}\}$. The Hamiltonians *ZX* and *IX* (brown) commute and are defined as the ideal dynamics, while the others are considered errors (blue). **c** Schematic illustration of the recursive DRAG pulse that suppresses different error transitions on the control Transmon. **d** Schematic illustration of different multi-qubit errors in the effective frame during the CR operation, and the transformations of the error terms. The remaining *IY* and *IZ* errors are compensated for by an *IY*-DRAG pulse on the target qubit and the detuning of the CR drive.

fabrication process[27,30]. Therefore, designing an efficient control scheme that operates seamlessly across diverse parameter regimes is essential for achieving optimal performance across hundreds of qubits.

In this work, we devise and test a simple and scalable control scheme for CR gates that counteracts all the aforementioned control errors, following the ideal of the well-known Derivative Removal by Adiabatic Gate (DRAG) method[31–34]. For the transition error, we demonstrate that previously suggested single-derivative DRAG correction[22] is insufficient for typical parameters in multi-qubit devices, where multiple transition errors are present. We introduce a novel recursive multi-derivative DRAG pulse, considering all three possible error transitions, capable of experimentally suppressing the error to high precision without the need for any calibration, or additional free parameters. For the two-qubit rotation operator error such as the *ZZ* error in the effective two qubits' subspace, we present a different approach. While other schemes typically involve hardware modifications or additional detuned microwave drive terms[8,35,36], we show that a simple, DRAG-like correction tone applied on the target qubit, along with a detuning on the drive frequency, is sufficient to eliminate dominant entangling error terms while avoiding additional hardware engineering. An overview of the derived pulse schemes is shown in Fig. 1c, d.

In comparison to alternative pulse shaping techniques[37–39], this multi-derivative pulse Ansatz stands out for its simplicity. It provides an efficient parameterization of the control pulse as a simple expression of the qubits' frequency and anharmonicity. This simplicity is essential for scalable quantum devices as all the qubits need to be calibrated quickly and repeatedly to ensure high fidelity. With the qiskit-pulse[40] interface, we implemented our drive scheme on multi-qubit devices provided by IBM Quantum. Despite the limited calibration time due to sporadic access to busy, public machines, our experimental results validate the efficient suppression of coherent errors. We observe a two- to fourfold reduction in infidelity, achieving beyond state-of-the-art fidelities in the 99.6 -99.8% range on multiple qubit pairs publicly available on the IBM platform.

The rest of the paper is organized as follows: We start by presenting the theoretical framework of the derivative-based pulse shaping methods. Next, we derive the pulse schemes for the CR gate and experimentally validate the error suppression for both the transition errors on the control qubit and the multi-qubit errors. Finally, we demonstrate the performance and scalability of the proposed control scheme by benchmarking the custom-implemented CR gate on multi-qubit quantum hardware, accompanied by numerical simulations across a wide range of experimentally relevant regimes.

Results

Multi-derivative pulse shaping

We start explaining the general theory for the systematic, iterative error suppression with a generic two-level system

$$\hat{H} = \Delta\hat{\Pi}_j + g(t)\frac{\hat{\sigma}^+_{jk}}{2} + g^*(t)\frac{\hat{\sigma}^+_{kj}}{2} \tag{1}$$

where $\hat{\Pi}_j = |j\rangle\langle j|$ and $\hat{\sigma}^+_{jk} = |k\rangle\langle j|$, $g(t)$ denotes the coupling strength between the two levels. In the following, we omit the explicit time dependence on t for ease of notation. In general, g could take the (perturbative) form of an n-photon interaction, $\frac{\Omega^n}{\Delta_{\text{eff}}^{n-1}}$, where Δ_{eff} is an effective energy gap and Ω the drive strength. In particular, if Ω denotes the CR drive strength on the control qubit, with $n=1$, it describes the transition $|0\rangle \leftrightarrow |1\rangle$ (or $|1\rangle \leftrightarrow |2\rangle$) and with $n=2$ the two-photon transition $|0\rangle \leftrightarrow |2\rangle$.

The goal is to suppress the undesired transition introduced by the coupling g. If $g \ll \Delta$, we may perform a perturbative expansion with the antihermitian generator $\hat{S}(\tilde{g}) = \frac{\tilde{g}}{2\Delta}\hat{\sigma}^+_{jk} - \text{h.c.}$ and obtain under the transformation $\hat{H}'(g) = \hat{V}(\tilde{g})\hat{H}(g)\hat{V}^\dagger(\tilde{g}) + i\dot{\hat{V}}(\tilde{g})\hat{V}^\dagger(\tilde{g})$ with $\hat{V}(\tilde{g}) = e^{\hat{S}(\tilde{g})}$,

$$\begin{aligned}\hat{H}'(g) &= i\dot{\hat{S}}(\tilde{g}) + \hat{H}(g) + [\hat{S}(\tilde{g}), \hat{H}(g)] + \cdots \\ &= \Delta\hat{\Pi}_j + \left(g - \tilde{g} + i\frac{d}{dt}\frac{\tilde{g}}{\Delta}\right)\frac{\hat{\sigma}^+_{jk}}{2} + \text{h.c.} + \mathcal{O}(\epsilon^2),\end{aligned} \tag{2}$$

where $\epsilon \propto g/\Delta$. We deliberately distinguish between g, the actual physical coupling, and $\tilde{g}$, which is used to define the generator $\hat{S}$ that diagonalizes the Hamiltonian. As a result, for a time-dependent coupling g, to suppress the transition, we require

$$g = \tilde{g} - i\frac{d}{dt}\frac{\tilde{g}}{\Delta}. \tag{3}$$

The above equation also provides an alternative interpretation: Transition-less evolution is possible if we find a (counter-diabatic) control $g(t)$ by choosing any continuous function $\tilde{g}(t)$ and making sure that $\hat{S}(\tilde{g})$ is zero at the beginning and at the end of the time evolution[41–44]. Thus, Eq. (3) provides a substitution rule to derive a time-modulated coupling $g(t)$ with

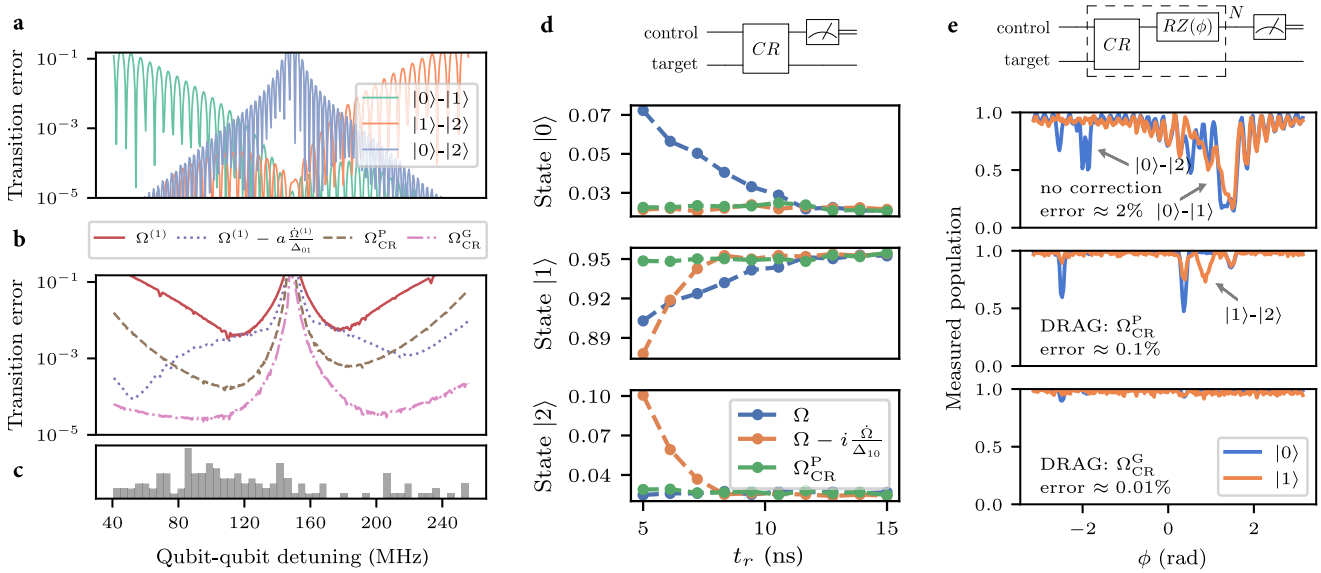

Fig. 2 | The transition errors on the control qubit with different drive schemes. **a** Simulated transition error among the 3 levels of the control Transmon introduced by the CR drive using a flat-top Gaussian pulse with $t_r = 2\sigma = 10$ ns. **b** The total transition error for different pulse schemes. We plot the envelope of the oscillation by taking the maximum over different pulse lengths with the ramping time $t_r = 10$ ns fixed. Parameters used are $\Omega_{max}/2\pi = 30$ MHz, $(\Delta_{21} - \Delta_{10})/2\pi = -300$ MHz and $\lambda = \sqrt{2}$. **c** Distribution of the qubit-qubit detunings in *ibm_brisbane*. A few outliers that are far away out of the studied range are left out. **d** The circuit and measured transition after preparing the state in $|1\rangle$ and applying a CR pulse with different rising time t_r. The data is obtained with a detuning of 120 MHz and a drive strength of about 40 MHz. **e** The amplification circuit and measured transition error. Deviations from the expected population of one indicate the presence of error. The three plots correspond to a default flat-top Gaussian, recursive DRAG pulse Ω^{P}_{CR} in Eq. (8) and Ω^{G}_{CR} in Eq. (9), with no calibration of additional parameters. The data is obtained from *ibm_nairobi* Q2 and Q1 with the drive amplitude 0.5 (≈60 MHz), $t_r = 10$ ns and $N = 30$. The qubit-qubit detuning is about 104 MHz.

the transition between the two levels suppressed. If the coupling describes an n photon interaction generated by a drive Ω, i.e., $g = \frac{\Omega^n}{\Delta_{\text{eff}}^{n-1}}$ with a constant Δ_{eff}, we obtain

$$\Omega = \mathcal{F}_{\Delta}^{(n)}(\tilde{\Omega}) := \left(\tilde{\Omega}^n - \mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\tilde{\Omega}^n}{\Delta}\right)^{\frac{1}{n}}. \tag{4}$$

Here, we choose $\tilde{g} = \frac{\tilde{\Omega}^n}{\Delta_{\text{eff}}^{n-1}}$ to keep the notation intuitive. The fractional exponent is defined for complex numbers and needs to ensure the continuity of Ω as a function of t. For $n = 1$ this gives the familiar result of the single-derivative DRAG expansion[33]. If needed, a free parameter a can be added before the derivative term to adjust the strength DRAG correction.

More generally, a two-level Hamiltonian (or subspace), in Eq. (1), is diagonalized exactly by the unitary transformation (referred to as Givens rotation)[45]

$$\hat{V} = \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) & \mathrm{e}^{-\mathrm{i}\phi}\sin\left(\frac{\theta}{2}\right) \\ -\mathrm{e}^{\mathrm{i}\phi}\sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right) \end{pmatrix}, \tag{5}$$

resulting in an exact substitution rule [c.f. Eq. (3)]:

$$g = \mathrm{e}^{\mathrm{i}\phi}\left(-(\Delta + \dot{\phi})\tan(\theta) + \mathrm{i}\dot{\theta}\right), \tag{6}$$

where θ and ϕ can in principle be chosen arbitrarily provided $\hat{V} = \mathbb{1}$ at the beginning and the end of the drive.

This exact diagonalization becomes useful in scenarios involving strong drive amplitudes or small detunings. In those cases, Eq. (6) provides a compact expression for the DRAG pulse beyond the perturbation limit. To be consistent with the perturbative solution, we set $\theta = \arctan(-|\tilde{g}|/\Delta)$ and define ϕ as the complex phase of the coupling, i.e., $\tilde{g} = \mathrm{e}^{\mathrm{i}\phi}|\tilde{g}|$. We note that, in general, Δ could also depend on g and Eq. (6) becomes an implicit equation for g instead of a closed-form expression. To obtain an expression for the drive strength Ω, one needs to invert the dependence of $g = f(\Omega)$. For instance, for a linear dependence, $g = \kappa\Omega$ (and $\tilde{g} = \kappa\tilde{\Omega}$), with κ a constant factor, we get

$$\Omega = \mathcal{F}_{\Delta}^{(1),\mathrm{G}}(\tilde{\Omega}) := \frac{\Delta + \dot{\phi}_{\tilde{\Omega}}}{\Delta}\tilde{\Omega} + \frac{\mathrm{i}\mathrm{e}^{\mathrm{i}\phi_{\tilde{\Omega}}}}{\kappa}\frac{\mathrm{d}}{\mathrm{d}t}\arctan\left(-\frac{|\kappa\tilde{\Omega}|}{\Delta}\right) \tag{7}$$

with $\mathrm{e}^{\mathrm{i}\phi_{\tilde{\Omega}}} = \tilde{\Omega}/|\tilde{\Omega}|$. Eqs. (4) and (7) will be the building blocks throughout the remaining of this article as we extend our analysis to multilevel systems.

Application to control-qubit errors

The CR interaction is typically activated by driving the control qubit with the frequency of the target[23,25,26], leading to a rotation in the target qubit depending on the state of the control, equivalent to a CNOT gate up to single-qubit operations. Ideally, the state of the control qubit should remain unaltered at the end of the gate. However, despite the detuning, the drive may still excite the control qubit, especially when operating in the straddling regime for fast entanglement[23,26], where the qubit-qubit detuning is smaller than their anharmonicities. Depending on the parameter regimes, it manifests both as single-photon transitions between $|0\rangle \leftrightarrow |1\rangle, |1\rangle \leftrightarrow |2\rangle$ as well as the two-photon transition between $|0\rangle \leftrightarrow |2\rangle$[22].

To counter these transition errors, the single-derivative DRAG pulse has been employed which introduces a term proportional to the first derivative of the drive pulse, i.e., $\Omega - \mathrm{i}a\frac{\dot{\Omega}}{\Delta}$, with a constant factor a to be optimized[22]. This heuristic is proven useful when the qubit-qubit detuning is very small, ranging from 50 MHz to 70 MHz[7,24], because in this range there is only one dominant transition. In contrast, for the scalable, multi-qubit fixed-frequency architecture such as the IBM Quantum plaform, the detuning between neighboring qubits is distributed over a much broader range (Fig. 2c). An efficient drive scheme must be able to suppress the error in all the operating parameter regimes. As shown later in this section, in some prevalent parameter regimes, the single-derivative DRAG pulse provides only minimal improvement. Even with a numerically optimized scale factor, a compromise arises among different transitions[22].

Following the multi-derivative pulse described above, we propose the following pulse shape derived by recursively applying the DRAG correction targeting at the three dominant transitions

$$\Omega_{\mathrm{CR}}^{\mathrm{P}} = \mathcal{F}_{\Delta_{21}}^{(1)} \circ \mathcal{F}_{\Delta_{10}}^{(1)} \circ \mathcal{F}_{\Delta_{20}}^{(2)}(\Omega) \tag{8}$$

with the perturbative substitution Eq. (4) or

$$\Omega_{\mathrm{CR}}^{\mathrm{G}} = \mathcal{F}_{\Delta_{21}}^{(1),\mathrm{G}} \circ \mathcal{F}_{\Delta_{10}}^{(1),\mathrm{G}} \circ \mathcal{F}_{\Delta_{20}}^{(2)}(\Omega) \tag{9}$$

using the exact expression Eq. (7) for the two single-photon transitions. The energy difference between state $|j\rangle$ and $|k\rangle$ in the rotating frame is denoted by Δ_{jk}. The symbol $\circ$ denotes the composition of different substitutions $\mathcal{F}$, applied sequentially from right to left on the pulse shape. Recursively chaining the DRAG correction as above suppresses all three dominant errors. While the two single-photon corrections $\mathcal{F}^{(1)}$ are interchangeable, the substitution for the two-photon transitions needs to be applied first, as detailed in the Methods section. The explicit formula for the perturbative DRAG pulse [Eq. (8)] is given by the following recursive expressions:

$$\Omega_{\mathrm{CR}}^{\mathrm{P}} = \Omega_1 - \mathrm{i}\frac{\dot{\Omega}_1}{\Delta_{10}} \tag{10}$$

$$\Omega_1 = \Omega_2 - \mathrm{i}\frac{\dot{\Omega}_2}{\Delta_{21}} \tag{11}$$

$$\Omega_2 = \sqrt{\Omega_3^2 - \mathrm{i}\frac{2\Omega_3\dot{\Omega}_3}{\Delta_{20}}}. \tag{12}$$

Here, Ω_3 needs to be chosen such that the obtained pulse is continuous and starts and ends in zero. Without the last equation for the two-photon transition, the derived pulse aligns with the multi-derivative DRAG solution proposed for multiple linear couplings in[33]. Notably, if one of the DRAG correction strengths is fine-tuned by an additional parameter, it will not affect the other correction significantly because of the recursive design.

Typically, a CR pulse consists of a rising, a holding and a lowering period, during which the pulse is turned on from zero to the maximum, held for a while and then turned off. We choose the rising portion of the pulse to be

$$\Omega^{(m)}(t) = \Omega_{\max}\mathcal{I}_0 \int_0^t \mathrm{d}t' \sin^m\left(\frac{\pi t'}{2t_{\mathrm{r}}}\right), \quad 0 \le t \le t_{\mathrm{r}} \tag{13}$$

with the normalization $\mathcal{I}_0$ fixed via $\Omega^{(m)}(t_{\mathrm{r}}) = \Omega_{\max}$. This definition ensures that the pulse is m times differentiable and the derivatives are zero at $t = 0$ and $t = t_{\mathrm{r}}$, which guarantees the validity of the frame transformation $\hat{V}$ introduced above. Other pulse shapes can also be used as long as this property is satisfied. After the holding time, the lowering phase takes the time-reversed shape. An example of the CR pulse is shown in Fig. 1c. For $m = 1$ and with zero holding time, the pulse is the same as the Hann window, very close to the flat-top Gaussian pulse commonly adopted. It is important to note that, as m increases, more high-frequency components become incorporated into the pulse shape, leading to higher non-adiabatic transition error if not compensated for. Therefore, it is advisable to keep m as small as possible. For our study, we use $m = 3$ as the initial shape for the recursive DRAG pulse.

To verify the performance of the error suppression, we numerically simulate the dynamics of the three-level Hamiltonian of the control Transmon (see Methods) and the result is shown in Fig. 2a, b. First, we examine in Fig. 2a the contribution of the three transition errors for an uncorrected pulse, across the typical experimentally relevant qubit-qubit detuning values. The error is defined as the probability of unwanted population transfer among different states. The plot indicates that all three transitions must be considered for sufficient error suppression. Moreover, we observe that partial suppression of the errors (using only one or two derivatives) may increase the unsuppressed ones, as demonstrated in detail in Supplementary Note 1, making them non-negligible even if they were initially small.

Next, we compare the total transition error introduced by different pulse schemes in Fig. 2b. To better illustrate the difference between the pulse schemes, we take the sum of the three transition errors and the maximum over pulses with various holding lengths. In this way, the oscillation caused by the pulse timing is removed and only the upper envelope remains. As a baseline, we plot the error for pulse shape $\Omega^{(1)}$, which is similar to the flat-top Gaussian pulse used in qiskit-pulse[40]. The recursive DRAG pulse shapes we derived, $\Omega_{\mathrm{CR}}^{\mathrm{P}}$ and $\Omega_{\mathrm{CR}}^{\mathrm{G}}$, suppress the error by several orders of magnitude, without any numerical optimization, as long as the drive is not resonant with the two-photon transition. In comparison, the single-derivative DRAG pulse, used in previous works[7,22,24], performs well only when the error is dominated by one single-photon transition (very large or very small detuning). Outside of these regimes, its performance is restricted due to the compromise between different transitions, even if the DRAG coefficient a is calibrated to minimize the total error.

This observation is further supported by experimental results shown in Fig. 2d, e. In Fig. 2d, the state was initialized in state $|1\rangle$, and a CR pulse of 200 ns with varying rising times t_{r} was applied. As the rising time decreases, the error transition grows quadratically. Without any correction, the error is dominated by the transition between $|0\rangle \leftrightarrow |1\rangle$. Applying a DRAG pulse designed to suppress this transition, i.e., $\Omega - \mathrm{i}\dot{\Omega}/\Delta_{10}$, effectively suppresses this error but introduces a new transition error between $|1\rangle \leftrightarrow |2\rangle$. Calibrating the DRAG coefficient only compromises between these errors. In contrast, with the recursive pulse shape defined in Eq. (8), all errors are suppressed below the state preparation and measurement error.

Typically, achieving high-fidelity quantum operations requires the transition errors to be suppressed to the order of 10^{-4}. Resolving this population error often needs a large number of sampling points. Therefore, we employ the error amplification circuits outlined in ref. 24, which add virtual Z gates $RZ(\phi)$ between the repetitions with different phases ϕ (Fig. 2). Different transition errors will be selected by different choices of ϕ, as detailed in Supplementary Note 3. To perform the measurement, we calibrate an X gate between states $|1\rangle$, $|2\rangle$ and build a measurement discriminator for qutrits[46].

In Fig. 2e, the measured population of the state $|0\rangle(|1\rangle)$ is plotted after the initial preparation in $|0\rangle(|1\rangle)$ and 30 repetition of the CR pulse. A population close to one implies negligible errors, while any deviation indicates a transition to other states. For instance, overlapping blue and orange curves indicate the $|0\rangle \leftrightarrow |1\rangle$ transition, while a drop solely in the blue curve suggests the $|0\rangle \leftrightarrow |2\rangle$ transition. It is evident that for this short rising time ($t_{\mathrm{r}} = 10$ ns), there exists a significant transition error between state $|0\rangle$ and $|1\rangle$, but also a non-negligible contribution from other transitions between $|0\rangle$ and $|2\rangle$. After applying the perturbative DRAG pulse, a substantial reduction in the error is observed, with some remaining small transitions. Using the recursive DRAG pulse derived by Givens rotation proves highly effective, suppressing all transition errors below the threshold. In both cases, no calibration of DRAG coefficients is required, and the analytical formulas are completely predictive. In general, free parameters can be added to each substitution before the derivative terms to fine-tune the strength of DRAG corrections for each individual transition error.

It is crucial to highlight that previous applications of a single-derivative DRAG[7,22,24] to CR gates primarily focus on the case of very small qubit-qubit detuning (ranging from 50 MHz to 70 MHz), the errors of which is dominated only by the $|0\rangle \leftrightarrow |1\rangle$ transition. However, qubit pairs on IBM Quantum Platform have detuning distributed in a much larger range from 40 to 260 MHz (Fig. 2c), where other transitions become nonnegligible (Fig. 2a). In contrast, the recursive DRAG solution showcased in this study exhibits remarkable universal performance even in the presence of multiple types of errors, without any calibration necessary. In Supplementary Note 5, we show similar error suppression on qubit pairs with qubit-qubit deutning of 143 MHz and 189 MHz, together with an example where the single-derivative DRAG fails to suppress the error even with a full-sweep calibration of the DRAG coefficient.

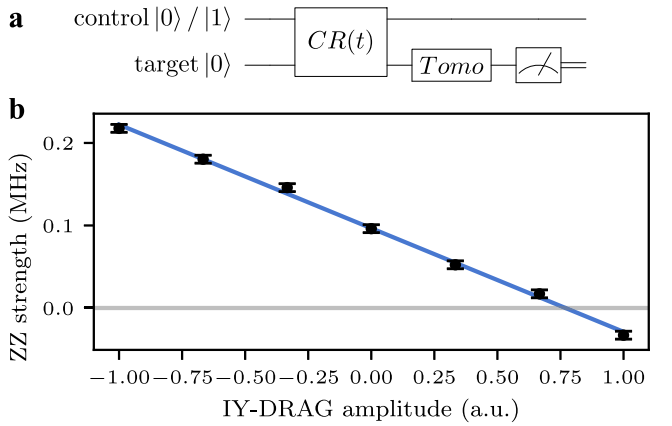

Fig. 3 | Measurement of the ZZ error and calibration of the IY-DRAG schemes. **a** Circuit used for the CR Hamiltonian tomography[19]. **b** Experimental data for calibrating the *IY*-DRAG amplitude to minimize the *ZZ* coupling, obtained from *ibm_perth*. The error bar denotes the standard deviation of the measured *ZZ* strength.

Application to multi-qubit operator errors

A second major part of the error in the CR operation comes from the remaining dynamical operators in the two-qubit subspace that do not commute with the ideal dynamics *ZX*. Assuming the transition errors on the control qubit are all suppressed, the effective Hamiltonian in the two-qubit subspace is given by

$$H = \frac{\nu_{ZX}}{2}\hat{Z}\hat{X} + \frac{\nu_{ZY}}{2}\hat{Z}\hat{Y} + \frac{\nu_{ZZ}}{2}\hat{Z}\hat{Z} + \frac{\nu_{IX}}{2}\hat{I}\hat{X} + \frac{\nu_{IY}}{2}\hat{I}\hat{Y} + \frac{\nu_{IZ}}{2}\hat{I}\hat{Z}. \tag{14}$$

The coefficient $\nu(t)$ for each term can be derived by perturbative expansion, with the explicit expressions given in appendix C of ref. 25. Experimentally, they can be measured by Hamiltonian tomography[19]. An overview of the multi-qubit errors and the frame transformations used below to remove them is shown in Fig. 1d.

When implementing a CR gate, the *ZY* term is removed by calibrating the phase of the CR drive and the single-qubit rotations, *IX* and *IY*, compensated for by a target drive[19]. To achieve high-fidelity operations, we iteratively fine-tune the drive pulse until the error terms *ZY*, *IY* and *IX* are all below 0.015 MHz (see Supplementary Note 6). After this standard calibration procedure, one can describe the dynamics with the following effective Hamiltonian

$$\hat{H} = \frac{\nu_{ZX}}{2}\hat{Z}\hat{X} + \frac{\nu_{ZZ}}{2}\hat{Z}\hat{Z} + \frac{\nu_{IZ}}{2}\hat{I}\hat{Z}, \tag{15}$$

where the first term is the desired Hamiltonian dynamic while the other two are multi-qubit errors to be suppressed.

We now show that an *IY*-DRAG correction and a detuning are sufficient to suppress the remaining errors. Note that the two Hamiltonian terms *ZX* and *ZZ* are connected by a rotation along the *IY* axis. Hence, we define the transformation

$$\hat{V}_{ZZ} = \hat{I} \otimes \exp\left(-\mathrm{i}\beta(t)\hat{Y}/2\right) \tag{16}$$

with $\beta = \arctan(\frac{\nu_{ZZ}}{\nu_{ZX}}) \approx \frac{\nu_{ZZ}}{\nu_{ZX}}$. In the perturbaitve expansion, the coefficients are given as $\nu_{ZZ} = \frac{J^2|\Omega_{CR}|^2}{2\Delta_{eff}(\alpha_1+\Delta)2}$ and $\nu_{ZX} + \mathrm{i}\nu_{ZY} = -\frac{J\Omega_{CR}\alpha_1}{\Delta(\alpha_1+\Delta)}$, where Δ_{eff} is a constant depending on the Transmons' frequency and anharmonicity[25]. This transformation results in an enhanced *ZX* strength $\sqrt{\nu_{ZX}^2 + \nu_{ZZ}^2}$ and an additional single-qubit term $\dot{\beta}\hat{I}\hat{Y}/2$ to be compensated by a $\hat{Y}$ drive on the target qubit. It is not difficult to verify that this corresponds to a DRAG-like correction which is non-zero only during the pulse ramping time.

In general, to completely remove the error, one needs to match the shape of the *IY*-DRAG pulse exactly with $\dot{\beta}/2$. In typical CR gates, the holding period is much longer than the ramping time. Therefore, we can neglect the coherent error introduced by the time-dependent part $0 < t < t_r$ and focus only on the holding period $t_r < t < t_f - t_r$. This simplified approach allows us to neglect the shape of the *IY*-DRAG pulse and only calibrate the area (amplitude) such that the *ZZ* error is removed during the holding period. We choose the *IY*-DRAG shape as the first derivative of the target drive, i.e., $c_{IY}\dot{\Omega}_{IX}$. Given that the *ZZ* error is typically small (<0.1 MHz), the *IY*-DRAG correction is also very weak. Thus, the correct coefficient c_{IY} can be obtained by measuring the *ZZ* coupling strength for a few different c_{IY} and conducting a linear fit, as illustrated in Fig. 3. In practice, we find that three sampling points are sufficient for the accurate calibration of the *IY*-DRAG amplitude. It is worth noting that in the calibration, the removed *ZZ* error consists of both the dynamic ones introduced by the drive and the static *ZZ* terms caused by residual coupling[47].

Compared to the previous approach applied by IBM, the target rotatory pulse[48], our proposed method require only three sampling points and a linear fit, employing the same tomography circuit as used in the standard calibration[19], which renders it more practical for implementation on the IBM platform with limited calibration time. In contrast, the calibration of the target rotatory pulse amplitude requires a sweep across various amplitudes and finding a minimum of total measured errors. Furthermore, our method does not require the echoed CNOT structure and thus can be used to construct a direct CNOT gate. The two methods can also be combined, introducing new degrees of freedom to suppress more residual errors at the same time, which are left for future study.

Finally, the only untreated error, the *IZ* term, is compensated for by detuning the CR drive. In general, the exact cancellation of the *IZ* error requires time-dependent detuning, i.e., a chirped pulse or phase ramping. Here, as the *IZ* term is usually small, it is sufficient to compensate for it with a constant detuning. This is implemented by adding an additional phase term to the pulse shape $\exp(-\mathrm{i}\nu_{IZ}t/2)$, where ν_{IZ} denotes the measured *IZ* coefficient from the tomography, similar to the phase error in single qubit gates[49].

Benchmarking the improved CR gate

The investigations outlined above underscore the performance of our proposed methods in addressing both the transition errors on the control Transmon induced by rapid driving and the multi-qubit operator errors arising from (static and dynamic) residual coupling. The improved precision in control not only reduces the coherent error but also facilitates the exploration of higher drive amplitudes and faster tuning speeds, which usually introduces more coherent error if left uncompensated[23,26]. As a result, the attained reduction in gate time allows us to exceed the impact of decoherence and achieve higher fidelities. For instance, by reducing t_r from 28 ns (default qiskit-pulse parameter) to 10 ns, one gains about 35 ns for an echoed CR gate. Moreover, because of the simplified calibration procedure, a complete removal of multi-qubit errors is achieved only when the drive is at its maximum. Therefore, the reduction in t_r contributes not only to shorter gate time but also to improved accuracy in the CR drive.

In consideration of these factors, we select pairs of qubits from the available Transmon qubits on the IBM Quantum Platform that exhibit sufficiently large coupling and relatively long coherence time (>30 μs). Due to our limited access, a comprehensive search for the speed limit at different drive strengths to determine the optimal gate time is not feasible. Therefore, we opt for an empirical approach in choosing the CR drive amplitude based on the qubit-qubit detuning and the effective coupling strength. In addition to the commonly used echoed CNOT gates[19], we also calibrate direct CNOT

gates, eliminating the two single-qubit gates and further reducing the gate time. Examples of the applied pulse shapes are shown in Fig. 4a, b, which include the recursive-DRAG pulse correction on the CR drive (ramping up of the CR pulse), the *IY*-DRAG correction (green pulse), and the frequency detuning (asymmetry between the ramping up and off). The calibration procedure is explained in detail in Supplementary Note 6.

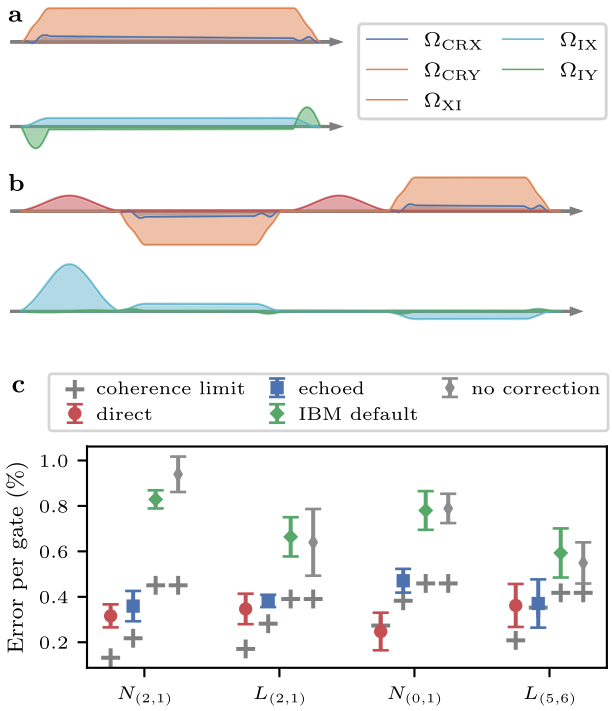

Fig. 4 | Calibrated pulses for the CNOT gate and randomized benchmarking. Examples of the custom pulse shapes used are shown for the direct (**a**) and echoed (**b**) CNOT gate on *ibm_lagos* qubit Q2 and Q1. The amplitudes for different drives are rescaled for visualization purposes. **c** Error per gate of the self-calibrated CNOT and the default CNOT gate, estimated by interleaved randomized benchmarking. The coherence limit is calculated including relaxation (T1) and dephasing (T2) noise provided by IBM Quantum. Data obtained from *ibm_lagos* (*L*) and *ibm_nairobi* (*N*). The error bar represents the standard deviation of five randomized benchmarking experiments.

To accurately characterize the fidelity, we measure the infidelity of the self-calibrated CR gates through interleaved randomized benchmarking. For each pulse configuration, we repeat the experiment five times (each takes about 5 to 10 minutes including the classical communication time) and compute the mean and standard deviation of the measured gate error. Therefore, the presented gate error should be interpreted as the average error over the following hours after the calibration, including possible detrimental drift in the prior system parameters over the ensuing time period. As shown in Fig. 4, we obtain a significant reduction of the error on several pairs of the qubits compared to the default CNOT gate. Over the four pairs of qubits studied, we obtain an average gate fidelity of 99.7(1)%. We further compare it to the self-calibrated CNOT gate with no corrections applied, using the default Gaussian shape, drive amplitude and t_r. This comparison verifies that the observed improvement is not solely attributable to our more recent calibration. Detailed information on the used qubits, measured effective coupling strength and drive parameters is presented in Supplementary Note 4.

Next, we characterize the possible improvement over a wider range of typical and prospective parameter regimes. In particular, we perform thorough numerical simulations to demonstrate what is possible beyond current bottlenecks given by present-day coherence times and limited calibration access on high-demand systems.

To show the applicability of the derived pulse on a large-scale quantum device, it is important to evaluate its performance across various parameter regimes representative of a real quantum system. The Hamiltonian model is chosen to have similar coupling strength and *ZZ* error rate to those qubits on the IBM Quantum Platform (see Methods). Since we focus on the CR operation, we exclude the error that can be removed by single-qubit corrections. We perform a sweep for different Ω_{max} and t_r while also varying the qubit-qubit detuning $\Delta/2\pi = \{70, 110, 200\}$ MHz in the straddling regime.

In Fig. 5, we compare the infidelity between the proposed pulse and the flat-top Gaussian pulse. It shows a drastic reduction in the coherent error in all regimes via our approach, with orders of magnitude suppression similarly seen for the three detunings. The recursive DRAG correction reduces the ramping time while keeping a low transition error rate. Meanwhile, the *IY*-DRAG correction cancels the *ZZ* error and allows for stronger drive amplitude. The observed optimal selection of the pulse ramping time between 10 to 15 ns in the simulations results from a compromise between the static *ZZ* error in IBM qubit parameters and the transition error. In our simulation, we considered IBM hardware with fixed coupler frequencies, resulting in a static *ZZ* error that cannot be fully corrected, especially during the ramping period. Shorter ramping times lead to reduced accumulation of this static *ZZ* error, at the expense of increased transition error.

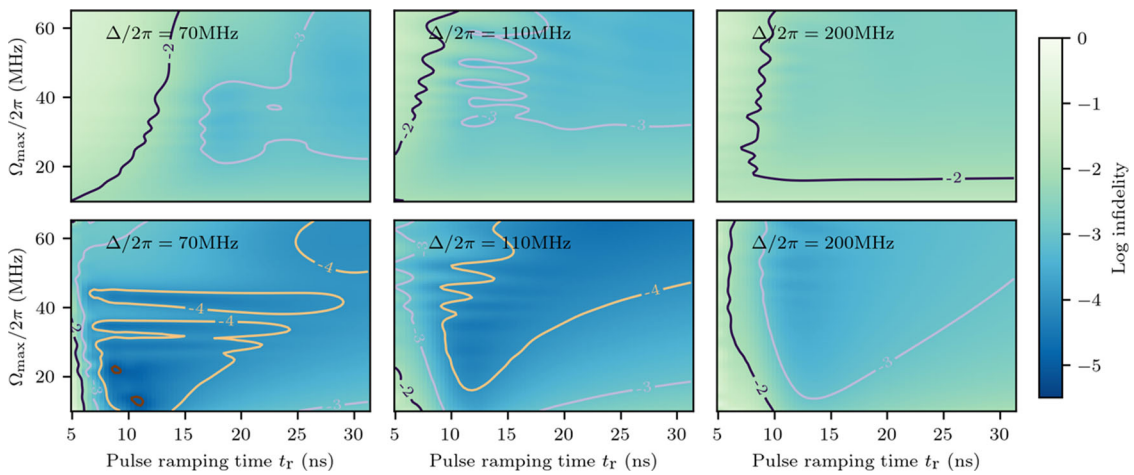

Fig. 5 | Comparison between the drive pulses. Infidelities of simulated CR gates implemented by the flat-top Gaussian pulse (top) and the proposed multi-derivative pulse (bottom), including the derived recursive DRAG correction (Ω_{CR}^{G}) and the *IY*-DRAG correction. All the pulses are closed-form and computed deterministically without numerical calibration. The simulation is repeated for 3 different values of the qubit-qubit detuning Δ.

For commensurate qubit lifetimes in the range of milliseconds, as already demonstrated in refs. 50,51, errors as low as 10^{-4} are within reach using our proposed pulse, while the standard pulse would be limited to an order of magnitude larger error. Along with the shorter gate times coming from larger amplitudes and shorter ramping times, the large coherent error suppression further amplifies any expected gains coming from improvements in qubit fabrication. This is already seen, for example, in Fig. 4, where not only is the coherent error on the IBM Quantum devices suppressed, but there is also a reduction in the total gate duration (with reduced coherence limit). This is especially important looking forward, as advantages in coherence times for fixed-frequency architectures vs. tunable-qubit architectures tilts the advantage towards the former with appropriate pulse shaping. Note also that even if the parasitic ZZ error is engineered to be very small[35,36,45,52], as coherence times improve, the standard pulses must choose a long ramping time to match the incoherent error, while our pulse shaping approach can continue to use very short times, fully taking advantage of such improvements.

Importantly, these pulses are constructed following the analytical expression without additional optimization or fitting parameters. This means that compared to all but the simplest approaches available, including ref. 22, these high-performance pulses are much faster and more straightforward to calibrate. Additionally, we observe that the transition error is barely affected by the drift of the drive strength and is also relatively robust against frequency drift (see Supplementary Note 2).

Discussion

We introduced an analytical multi-derivative pulse shape tailored for driving the CR interaction in superconducting qubits, adept at eliminating undesired transitions on the control qubit and unwanted multi-qubit dynamics. Our approach extends the DRAG formalism to a recursive structure capable of suppressing multiple error transitions simultaneously. Additionally, we developed a novel technique to eliminate multi-operator errors by dynamically transforming the errors into the desired entangling form. This resulted in several orders of magnitude suppression in the coherent error when simulating across the range of typical c-QED regimes, without extensive requirement on calibration. The simplicity and universality of the proposed pulse shape make it well-suited for implementation on the IBM Quantum Platform as an efficient high-quality calibration across hundreds of qubits. We demonstrate this on several qubits, showing a significant suppression in the state-of-the-art error using our customized pulse shape. The results are reproducible over a wide range of qubit frequency spacings and with prescriptive pulse shapes across the spectrum.

These analytical approaches are general and also applicable to other entangling gates in c-QED and various quantum technologies[53–55]. The control error addressed in this work extends beyond CR gates and is relevant to other off-resonant drive schemes, such as microwave-activated gates[8,56–58], as well as the use of microwave drives for suppressing quantum cross-talk and leakage[29,33,59,60]. The coherent error suppression demonstrated here also has implications for fixed-frequency architectures, allowing them to take advantage of longer coherence times compared to tunable architectures. Moreover, errors involving a spectator qubit[26,29,48,59] can be addressed by incorporating the ancillary level into the modeling and introducing new derivative-based corrections accordingly.

Apart from the pursuit of improving multi-qubit gates, it is noteworthy that the suppression of coherent errors also indirectly enhances the fabrication process' yield. For instance, the transition errors addressed in this work were also identified as frequency collisions in refs. 27,28. The proposed drive scheme effectively increases the threshold for frequency collisions, thereby contributing to an increased fabrication yield. Similar error models for frequency collision also apply to the tunable-coupler architecture[61], extending the potential application domain.

Methods

Derivation of the recursive DRAG pulse

We use the following three-level Hamiltonian to model the control Transmon

$$\hat{H} = \frac{\epsilon\Omega_{\mathrm{CR}}}{2}(\sigma_{01}^{+} + \lambda\sigma_{12}^{+}) + \text{h.c.} + \Delta_{10}\hat{\Pi}_1 + (\Delta_{10} + \Delta_{21})\hat{\Pi}_2, \tag{17}$$

where λ is the relative coupling strength of the second transition and ϵ is used to denote the perturbation order. For detuning $\Delta_{10} = 0$, the pulse is on resonance and implements a single-qubit gate. When the drive is resonant with the frequency of the target qubit, a CR operation is activated. In the rotating frame with respect to the driving frequency, we have Δ_{10} equal to the qubit-qubit detuning and $\Delta_{21} = \Delta_{10} + \alpha_c$, with α_c the anharmonicity. To the leading order perturbation, the coupling strength is proportional to Ω_{CR}[25]. An ideal CR pulse generates rotations on the target qubit depending on the control qubit state while leaving the latter intact. This approximation holds well as long as the dressing of the qubit is perturbative. Therefore, we aim at finding a pulse Ω_{CR} with non-zero real integral but introducing no population transfer among any of the three levels of the control qubit. This model, Eq. (17), includes both the leakage error and population flipping on the control qubit[23].

In the following, we show the derivation of the substitution rule of Ω in Eq. (8) via Schrieffer Wolff perturbation. We omit the perturbative corrections to the diagonal part of the Hamiltonian as they have no effect on the leading-order perturbative coupling strength. The derivation includes three steps, each targeting one coupling. The perturbative transformation generated by an anti-hermitian matrix $\hat{S}$ is defined as

$$\hat{H}' = \mathrm{i}\dot{\hat{S}} + \hat{H} + \left[\hat{S}, \hat{H}\right] + \frac{1}{2}\left[\hat{S}, \left[\hat{S}, \hat{H}\right]\right] + \cdots. \tag{18}$$

First, we apply the perturbative diagonalization targeting the $|0\rangle \leftrightarrow |1\rangle$ transition

$$\hat{S}_1 = \frac{\epsilon}{2}\left(\frac{\Omega_1}{\Delta_{10}}\hat{\sigma}_{01}^{+} + \frac{\lambda\Omega_1}{\Delta_{10}}\hat{\sigma}_{12}^{+}\right) - \text{h.c.} \tag{19}$$

The first component in $\hat{S}_1$ is chosen to remove the $|0\rangle \leftrightarrow |1\rangle$ coupling perturbatively. According to the derivation in the main text, we define a substitution for $\Omega_{\mathrm{CR}}^{\mathrm{P}}$

$$\Omega_{\mathrm{CR}}^{\mathrm{P}} = \Omega_1 - \mathrm{i}\frac{\dot{\Omega}_1}{\Delta_{10}}. \tag{20}$$

The second term in Eq. (19) is chosen such that $\mathrm{i}\dot{\hat{S}}_1$ is proportional to the Y control Hamiltonian. This ensures that in the derived effective Hamiltonian, no $\dot{\Omega}_1$ appears in the $|1\rangle \leftrightarrow |2\rangle$ coupling, because it is absorbed in Ω_{CR}. Note that it does not diagonalize the $|1\rangle \leftrightarrow |2\rangle$ coupling, which would need $\frac{\lambda\Omega_1}{\Delta_{21}}\hat{\sigma}_{12}^{+}$ instead. As a result, we obtain

$$\begin{aligned}\hat{H}_1 = &\left(1 - \tfrac{\Delta_{21}}{\Delta_{10}}\right)\left(\tfrac{1}{2}\lambda\Omega_1\epsilon\hat{\sigma}_{12}^{+} - \tfrac{\lambda\Omega_1^2\epsilon^2}{8\Delta_{10}}\hat{\sigma}_{02}^{+}\right) + \text{h.c.}\\ &+ \text{diag} + \mathcal{O}(\epsilon^3).\end{aligned} \tag{21}$$

In the second step, we perform another perturbative diagonalization that removes the $|1\rangle \leftrightarrow |2\rangle$ transition:

$$S_2 = \left(1 - \frac{\Delta_{21}}{\Delta_{10}}\right)\frac{\lambda\Omega_2\epsilon}{2\Delta_{21}}\hat{\sigma}_{12}^{+} - \text{h.c.} \tag{22}$$

and substitute

$$\Omega_1 = \Omega_2 - \mathrm{i}\frac{\dot{\Omega}_2}{\Delta_{21}}. \tag{23}$$

This gives the effective Hamiltonian

$$\begin{aligned}\hat{H}_2 &= \left(\tfrac{\Delta_{21}}{\Delta_{10}} - 1\right)\left(\Omega_2 - \mathrm{i}\tfrac{\dot{\Omega}_2}{\Delta_{21}}\right)^2 \tfrac{\lambda\epsilon^2}{8\Delta_{10}}\hat{\sigma}_{02}^{+} + \text{h.c.} \\ &\quad + \text{diag} + \mathcal{O}(\epsilon^3)\end{aligned} \tag{24}$$

where both single-photon transitions are removed to the leading order. Notice that the DRAG pulse shape is independent of the relative drive amplitude λ in this first-order approximation.

It may seem strange that the remaining coupling for the $|0\rangle \leftrightarrow |2\rangle$ transition is not symmetric with respect to the order of the transformations of $|0\rangle \leftrightarrow |1\rangle$ and $|1\rangle \leftrightarrow |2\rangle$, although the two substitutions commute. In fact, we can perform a transformation $\hat{S}_3$

$$\hat{S}_3 = -\left(\frac{1}{\Delta_{21}} - \frac{1}{\Delta_{10}}\right)\frac{(\lambda\Omega_2^2\epsilon^2)}{8\Delta_{10}}\hat{\sigma}_{02}^{+} - \text{h.c.} \tag{25}$$

which only removes the $\Omega\dot{\Omega}$ term and gives

$$\begin{aligned}H_3 &= \tfrac{1}{8}\lambda\epsilon^2\left(\tfrac{1}{\Delta_{21}} - \tfrac{1}{\Delta_{10}}\right)\left(\tfrac{\dot{\Omega}_2^2}{\Delta_{21}\Delta_{10}} + \Omega_2^2\right)\hat{\sigma}_{02}^{+} + \text{h.c.} \\ &\quad + \text{diag} + \mathcal{O}(\epsilon^3).\end{aligned} \tag{26}$$

Lastly, we perform the third step to suppress the remaining $|0\rangle \leftrightarrow |2\rangle$ coupling. To fully remove this transition one needs to solve the differential equation

$$\left(\frac{\dot{\Omega}_2^2}{\Delta_{21}\Delta_{10}} + \Omega_2^2\right) = \left(\frac{\dot{\Omega}_3^2}{\Delta_{21}\Delta_{10}} + \Omega_3^2\right) - \mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\left(\frac{\dot{\Omega}_3^2}{\Delta_{21}\Delta_{10}} + \Omega_3^2\right)}{\Delta_{20}}, \tag{27}$$

which is difficult because of the non-linearity. Moreover, it may result in a pulse that does not fulfill the boundary condition, unless Ω_3 is carefully chosen to ensure that. In practice, numerical solutions may be employed to solve the equation, though it will pose challenges for fast calibration. For simplicity, we here assume that the pulse ramping is quasi-adiabatic i.e. $\Omega_2^2 \gg \frac{\dot{\Omega}_2^2}{\Delta_{10}\Delta_{21}}$. For the parameters studied in this work, with $\dot{\Omega} \approx \frac{\Omega}{t_r}$, this threshold lies around $t_r \approx 6$ ns. In this case, we can ignore the term proportional to $\dot{\Omega}_2^2$.

We then define the last transformation that diagonalizes the $|0\rangle \leftrightarrow |2\rangle$ transition

$$\hat{S}_4 = \frac{1}{8}\lambda\epsilon^2\left(\frac{1}{\Delta_{21}} - \frac{1}{\Delta_{10}}\right)\frac{\Omega_2^2}{\Delta_{20}}\hat{\sigma}_{02}^{+} - \text{h.c.} \tag{28}$$

and substitute

$$\Omega_2 = \sqrt{\Omega_3^2 - \mathrm{i}\frac{2\Omega_3\dot{\Omega}_3}{\Delta_{20}}}. \tag{29}$$

Here, $\Omega_3 = \Omega^{(3)}$, defined in Eq. (13), which is a continuously three-times differentiable function and ensures that the final expression starts and ends at zero. As a result, we suppress all three transitions up to $\mathcal{O}(\epsilon^3) + \mathcal{O}(\dot{\Omega}^2/\Delta^4)$.

Combining the three expressions, we obtain the explicit formula for the perturbative recursive DRAG pulse presented in the main text. As simple as the perturbative DRAG expression is, it may not sufficiently suppress the error if the qubits frequencies are very close to the one of $|0\rangle \leftrightarrow |1\rangle$ or $|1\rangle \leftrightarrow |2\rangle$ and the perturbative approximation no longer holds, as shown in Fig. 2b. To address this limitation, we replace the substitutions for the single-photon transitions with the exact diagonalization based on Givens rotations defined in Eq. (7). It is important to note that the substitution $\mathcal{F}^{(1),G}$ is only exact concerning the two-level subsystem; corrections to the energy gaps and other couplings are still disregarded. Nevertheless, it still significantly improves the performance compared to the perturbative expressions.

Numerical simulation of the CR gate

In the simulation, we use an effective Duffing model[62] truncated at 4 levels

$$\hat{H}_0 = \omega_a\hat{a}^\dagger\hat{a} + \sum_{j=1,2}\omega_j\hat{b}_j^\dagger\hat{b}_j + \frac{\alpha_j}{2}\hat{b}_j^\dagger\hat{b}_j^\dagger\hat{b}_j\hat{b}_j + g_j(\hat{b}_j\hat{a}^\dagger + \hat{b}_j^\dagger\hat{a}) \tag{30}$$

where $\hat{b}_j$ and $\hat{a}$ are the annihilation operators for qubit j and the resonator, respectively, and g_j is the coupling strength. The microwave drive on qubit j is written as

$$\begin{aligned}\hat{H}_c &= \mathrm{Re}(\Omega_{CR})\cos(\omega_d t)(\hat{b}_j^\dagger + \hat{b}_j) \\ &\quad + \mathrm{i}\,\mathrm{Im}(\Omega_{CR})\cos(\omega_d t)(\hat{b}_j^\dagger - \hat{b}_j)\end{aligned} \tag{31}$$

where ω_d is the driving frequency, initially chosen as the frequency of the target qubit. For simplicity, we use the same drive frequency for both the control and the target qubit.

We choose the anharmonicity $\alpha = -300$ MHz and $g_j = 80$ MHz. The detuning of the coupler from the control qubit, i.e., $\omega_1 - \omega_r$, is about -1.4 GHz and adjusted such that the effective qubit-qubit coupling strength is about 3 MHz, with ZZ crosstalk around 0.06 MHz, similar to the Transmons on the IBM platform (see Supplementary Note 4). Based on the model above, we derive the CR pulse following the analytical expressions derived in this paper. The effective coupling strength of ZX and ZZ are computed using the Non-Perturbative Analytical Diagonalization (NPAD) method[45], from which the gate time, i.e., the holding duration of the pulse, is calculated.

When computing the fidelity in the simulation, we ignore the contribution of the (commuting) single-qubit corrections ZI and IX, because they can be easily calibrated in the experiment. Given an ideal unitary $\hat{\mathcal{U}}_I$ for a two-qubit gate, the average gate fidelity is defined as[63]

$$F[\hat{\mathcal{U}}_Q] = \frac{\mathrm{Tr}\left[\hat{\mathcal{U}}_Q\hat{\mathcal{U}}_Q^\dagger\right]}{d(d+1)} + \frac{\left|\mathrm{Tr}\left[\hat{\mathcal{U}}_Q\hat{\mathcal{U}}_I^\dagger\right]\right|^2}{d(d+1)} \tag{32}$$

where $\hat{\mathcal{U}}_Q$ is the full unitary truncated to the two-qubit subspace and $d = 4$. Because we ignore the possible single-qubit correction ZI and IX, we compute the maximal fidelity optimized over the possible single-qubit rotation angles

$$\tilde{F} = \max_{\{\theta_1,\theta_2\}} F\left[\mathrm{e}^{-\mathrm{i}(\theta_1\hat{I}\hat{X}+\theta_2\hat{Z}\hat{I})}\hat{\mathcal{U}}_Q\mathrm{e}^{\mathrm{i}(\theta_1\hat{I}\hat{X}+\theta_2\hat{Z}\hat{I})}\right]. \tag{33}$$

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability

The code for calibrating the Cross-Resonance gate can be accessed via this link https://github.com/BoxiLi/qiskit-CR-calibration.

Received: 20 January 2024; Accepted: 21 June 2024;
Published online: 02 July 2024

References

1. Zhao, Y. et al. Realization of an Error-Correcting Surface Code with Superconducting Qubits. *Phys. Rev. Lett.* **129**, 030501 (2022).
2. Acharya, R. et al. Suppressing quantum errors by scaling a surface code logical qubit. *Nature* **614**, 676–681 (2023).
3. Satzinger, K. J. et al. Realizing topologically ordered states on a quantum processor. *Science* **374**, 1237–1241 (2021).
4. Andersen, T. I. et al. Non-Abelian braiding of graph vertices in a superconducting processor. *Nature* **618**, 264–269 (2023).
5. IBM Quantum, https://quantum.ibm.com/ (2024).
6. Kim, Y. et al. Evidence for the utility of quantum computing before fault tolerance. *Nature* **618**, 500–505 (2023).
7. Kandala, A. et al. Demonstration of a High-Fidelity CNOT Gate for Fixed-Frequency Transmons with Engineered ZZ Suppression. *Phys. Rev. Lett.* **127**, 130501 (2021).
8. Wei, K. X. et al. Hamiltonian Engineering with Multicolor Drives for Fast Entangling Gates and Quantum Crosstalk Cancellation. *Phys. Rev. Lett.* **129**, 060501 (2022).
9. Willsch, D., Nocon, M., Jin, F., De Raedt, H. & Michielsen, K. Gate-error analysis in simulations of quantum computers with transmon qubits. *Phys. Rev. A* **96**, 062302 (2017).
10. Ghosh, J., Fowler, A. G., Martinis, J. M. & Geller, M. R. Understanding the effects of leakage in superconducting quantum-error-detection circuits. *Phys. Rev. A* **88**, 062329 (2013).
11. McEwen, M. et al. Removing leakage-induced correlated errors in superconducting quantum error correction. *Nat. Commun.* **12**, 1761 (2021).
12. Miao, K. C. et al. Overcoming leakage in quantum error correction. *Nat. Phys.* **19**, 1780–1786 (2023).
13. Varbanov, B. M. et al. Leakage detection for a transmon-based surface code. *npj Quantum Inf.* **6**, 102 (2020).
14. Bultink, C. C. et al. Protecting quantum entanglement from leakage and qubit errors via repetitive parity measurements. *Sci. Adv.* **6**, eaay3050 (2020).
15. Paraoanu, G. S. Microwave-induced coupling of superconducting qubits. *Phys. Rev. B* **74**, 140504(R) (2006).
16. de Groot, P. C. et al. Selective darkening of degenerate transitions demonstrated with two superconducting quantum bits. *Nat. Phys.* **6**, 763–766 (2010).
17. Rigetti, C. & Devoret, M. Fully microwave-tunable universal gates in superconducting qubits with linear couplings and fixed transition frequencies. *Phys. Rev. B* **81**, 134507 (2010).
18. Groszkowski, P., Fowler, A. G., Motzoi, F. & Wilhelm, F. K. Tunable coupling between three qubits as a building block for a superconducting quantum computer. *Phys. Rev. B* **84**, 144516 (2011).
19. Sheldon, S., Magesan, E., Chow, J. M. & Gambetta, J. M. Procedure for systematically tuning up cross-talk in the cross-resonance gate. *Phys. Rev. A* **93**, 060302(R) (2016).
20. Takita, M. et al. Demonstration of Weight-Four Parity Measurements in the Surface Code Architecture. *Phys. Rev. Lett.* **117**, 210505 (2016).
21. Takita, M., Cross, A. W., Córcoles, A. D., Chow, J. M. & Gambetta, J. M. Experimental Demonstration of Fault-Tolerant State Preparation with Superconducting Qubits. *Phys. Rev. Lett.* **119**, 180501 (2017).
22. Malekakhlagh, M. & Magesan, E. Mitigating off-resonant error in the cross-resonance gate. *Phys. Rev. A* **105**, 012602 (2022).
23. Tripathi, V., Khezri, M. & Korotkov, A. N. Operation and intrinsic error budget of a two-qubit cross-resonance gate. *Phys. Rev. A* **100**, 012301 (2019).
24. Wei, K. X., Pritchett, E., Zajac, D. M., McKay, D. C. & Merkel, S. Characterizing non-Markovian off-resonant errors in quantum gates. *Phys. Rev. Appl.* **21**, 024018 (2024).
25. Magesan, E. & Gambetta, J. M. Effective Hamiltonian models of the cross-resonance gate. *Phys. Rev. A* **101**, 052308 (2020).
26. Malekakhlagh, M., Magesan, E. & McKay, D. C. First-principles analysis of cross-resonance gate operation. *Phys. Rev. A* **102**, 042605 (2020).
27. Hertzberg, J. B. et al. Laser-annealing Josephson junctions for yielding scaled-up superconducting quantum processors. *npj Quantum Inf.* **7**, 129 (2021).
28. Zhang, E. J. et al. High-performance superconducting quantum processors via laser annealing of transmon qubits. *Sci. Adv.* **8**, eabi6690 (2022).
29. Zhao, P. et al. Quantum Crosstalk Analysis for Simultaneous Gate Operations on Superconducting Qubits. *PRX Quantum* **3**, 020301 (2022).
30. Tolpygo, S. K. et al. Fabrication process and properties of fully-planarized deep-submicron nb/al- AlO_x/Nb josephson junctions for vlsi circuits. *IEEE Trans. Appl. Supercond.* **25**, 1–12 (2015).
31. Motzoi, F., Gambetta, J. M., Rebentrost, P. & Wilhelm, F. K. Simple Pulses for Elimination of Leakage in Weakly Nonlinear Qubits. *Phys. Rev. Lett.* **103**, 110501 (2009).
32. Gambetta, J. M., Motzoi, F., Merkel, S. T. & Wilhelm, F. K. Analytic control methods for high-fidelity unitary operations in a weakly nonlinear oscillator. *Phys. Rev. A* **83**, 012308 (2011).
33. Motzoi, F. & Wilhelm, F. K. Improving frequency selection of driven pulses using derivative-based transition suppression. *Phys. Rev. A* **88**, 062318 (2013).
34. Theis, L. S., Motzoi, F., Machnes, S. & Wilhelm, F. K. Counteracting systems of diabaticities using DRAG controls: The status after 10 years. *EPL* **123**, 60001 (2018).
35. Mundada, P., Zhang, G., Hazard, T. & Houck, A. Suppression of Qubit Crosstalk in a Tunable Coupling Superconducting Circuit. *Phys. Rev. Appl.* **12**, 054023 (2019).
36. Xu, X. & Ansari, M. Parasitic-Free Gate: An Error-Protected Cross-Resonance Switch in Weakly Tunable Architectures. *Phys. Rev. Appl.* **19**, 024057 (2023).
37. Kirchhoff, S. et al. Optimized cross-resonance gate for coupled transmon systems. *Phys. Rev. A* **97**, 042348 (2018).
38. Dalgaard, M., Motzoi, F., Sørensen, J. J. & Sherson, J. Global optimization of quantum dynamics with AlphaZero deep exploration. *npj Quantum Inf.* **6**, 6 (2020).
39. Baum, Y. et al. Experimental Deep Reinforcement Learning for Error-Robust Gate-Set Design on a Superconducting Quantum Computer. *PRX Quantum* **2**, 040324 (2021).
40. Alexander, T. et al. Qiskit Pulse: Programming Quantum Computers Through the Cloud with Pulses. *Quantum Sci. Technol.* **5**, 044006 (2020).
41. Unanyan, R., Yatsenko, L., Bergmann, K. & Shore, B. Laser-induced adiabatic atomic reorientation with control of diabatic losses. *Opt. Commun.* **139**, 48–54 (1997).
42. Demirplak, M. & Rice, S. A. Adiabatic Population Transfer with Control Fields. *J. Phys. Chem. A* **107**, 9937–9945 (2003).
43. Chen, X., Lizuain, I., Ruschhaupt, A., Guéry-Odelin, D. & Muga, J. G. Shortcut to Adiabatic Passage in Two- and Three-Level Atoms. *Phys. Rev. Lett.* **105**, 123003 (2010).
44. Guéry-Odelin, D. et al. Shortcuts to adiabaticity: Concepts, methods, and applications. *Rev. Mod. Phys.* **91**, 045001 (2019).
45. Li, B., Calarco, T. & Motzoi, F. Nonperturbative Analytical Diagonalization of Hamiltonians with Application to Circuit QED. *PRX Quantum* **3**, 030313 (2022).
46. Morvan, A. et al. Qutrit Randomized Benchmarking. *Phys. Rev. Lett.* **126**, 210504 (2021).
47. Xu, X. & Ansari, M. H. ZZ freedom in two qubit gates. *Phys. Rev. Appl.* **15**, 064074 (2021).
48. Sundaresan, N. et al. Reducing Unitary and Spectator Errors in Cross Resonance with Optimized Rotary Echoes. *PRX Quantum* **1**, 020318 (2020).

49. Chen, Z. et al. Measuring and Suppressing Quantum State Leakage in a Superconducting Qubit. *Phys. Rev. Lett.* **116**, 020501 (2016).
50. Somoroff, A. et al. Millisecond Coherence in a Superconducting Qubit. *Phys. Rev. Lett.* **130**, 267001 (2023).
51. Wang, C. et al. Towards practical quantum computers: Transmon qubit with a lifetime approaching 0.5 milliseconds. *npj Quantum Inf.* **8**, 3 (2022).
52. Goerz, M. H., Motzoi, F., Whaley, K. B. & Koch, C. P. Charting the circuit QED design landscape using optimal control theory. *npj Quantum Inf.* **3**, 37 (2017).
53. Nesterov, K. N., Wang, C., Manucharyan, V. E. & Vavilov, M. G. CNOT Gates for Fluxonium Qubits via Selective Darkening of Transitions. *Phys. Rev. Appl.* **18**, 034063 (2022).
54. Dogan, E. et al. Two-Fluxonium Cross-Resonance Gate. *Phys. Rev. Appl.* **20**, 024011 (2023).
55. Ficheux, Q. et al. Fast Logic with Slow Qubits: Microwave-Activated Controlled-Z Gate on Low-Frequency Fluxoniums. *Phys. Rev. X* **11**, 021026 (2021).
56. Mitchell, B. K. et al. Hardware-Efficient Microwave-Activated Tunable Coupling between Superconducting Qubits. *Phys. Rev. Lett.* **127**, 200502 (2021).
57. Kim, Y. et al. High-fidelity three-qubit iToffoli gate for fixed-frequency superconducting qubits. *Nat. Phys.* **18**, 783–788 (2022).
58. Goss, N. et al. High-fidelity qutrit entangling gates for superconducting circuits. *Nat. Commun.* **13**, 7481 (2022).
59. Cai, T. Q. et al. Impact of Spectators on a Two-Qubit Gate in a Tunable Coupling Superconducting Circuit. *Phys. Rev. Lett.* **127**, 060505 (2021).
60. Marques, J. F. et al. All-Microwave Leakage Reduction Units for Quantum Error Correction with Superconducting Transmon Qubits. *Phys. Rev. Lett.* **130**, 250602 (2023).
61. Osman, A. et al. Mitigation of frequency collisions in superconducting quantum processors. *Phys. Rev. Res.* **5**, 043001 (2023).
62. Khani, B., Gambetta, J. M., Motzoi, F. & Wilhelm, F. K. Optimal generation of Fock states in a weakly nonlinear oscillator. *Phys. Scr.* **2009**, 014021 (2009).
63. Pedersen, L. H., Møller, N. M. & Mølmer, K. Fidelity of quantum operations. *Phys. Lett. A* **367**, 47–51 (2007).

Acknowledgements

The authors would like to thank Adrian Lupascu and Francisco Cárdenas López for the insightful discussions. This work was funded by the Federal Ministry of Education and Research (BMBF) within the framework programme “Quantum technologies - from basic research to market” (Project QSolid, Grant No. 13N16149), by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy - Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1 - 390534769, by HORIZON-CL4-2022-QUANTUM-01-SGA Project under Grant 101113946 OpenSuperQPlus100 and the European Union’s Horizon Programme (HORIZON-CL4-2021-DIGITALEMERGING-02-10) Grant Agreement 101080085 QCFD. We acknowledge the use of IBM Quantum Platforms for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

Author contributions

B.L., T.C. and F.M. contributed to the conceptualization. Detailed analysis was done by B.L. with input from F.M. The experimental investigation was performed by B.L. B.L. wrote the manuscript with input from F.M. and T.C., T.C. and F.M. supervised the project.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41534-024-00863-4.

Correspondence and requests for materials should be addressed to Boxi Li or Felix Motzoi.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2024