Supplementary Information for “Independent, extensible control of same-frequency superconducting qubits by selective broadcasting”

S. Asaad∗,1, 2 C. Dickel∗,1, 2 N. K. Langford,1, 2 S. Poletto,1, 2 A. Bruno,1, 2 M. A. Rol,1, 2 D. Deurloo,1, 3 and L. DiCarlo1, 2

1 QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands
2 Kavli Institute of Nanoscience, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands
3 Netherlands Organisation for Applied Scientific Research (TNO), P.O. Box 155, 2600 AD Delft, Netherlands

(Dated: April 6, 2016)

PACS numbers:

This supplement provides experimental details and additional data supporting the claims in the main text. The device and experimental setup, including images of the device and a full wiring diagram, are described in Section I. The microwave performance of the vector switch matrix (VSM) and its use for independent control of two same-frequency qubits are demonstrated in Sec. II A. The techniques employed for tuning qubit pulses are discussed in Sec. III. Section IV contains the results of randomized benchmarking (RB) experiments in the two-qubit global broadcasting context. Our technique for assessing the effects of leakage is detailed in Sec. V. Numerical simulations of the effects of cross-excitations on RB are shown in Sec. S8. In Section VI B, we introduce a method for generating selective-broadcasting pulse sequences that are robust to cross-excitation effects. The decompositions of the 24 single-qubit Cliffords into a minimal set of pulses and the 5-primitive pulses are given in Sec. VII. Finally, the algorithm used to compile optimal pulse sequences for implementing independent single-qubit Clifford gates on multiple qubits is explained in Sec. VIII.

I. QUANTUM CHIP AND EXPERIMENTAL SETUP

A. Chip design and fabrication

Our quantum chip consists of three transmons (top: QT, middle (ancilla): QM, and bottom: QB) with dedicated voltage drive lines (DT, DM and DB, respectively), flux-bias lines, and readout resonators. All readout resonators are capacitively coupled to one common feedline which crosses various on-chip components using airbridge crossovers. Qubits QT and QM are coupled by one bus resonator, and QB and QM by another (fundamental frequencies 4.9 and 5.0 GHz, respectively). All resonators are open-ended on the coupling side, and short-circuit at the other.

*These authors contributed equally to this work.

FIG. S1: Scanning electron micrograph (with added false colour) of a device twin to the one used in the experiment and fabricated in the same batch. Notably, the feedline crosses the three readout resonators.

The chip fabrication method is similar to that in ref. 1, but with some important differences which we now explain. Rather than sapphire, we use a high-resistivity intrinsic silicon substrate prepared by HF dip and HMDS surface passivation before sputtering a 300-nm-thick film of NbTiN, as introduced in ref. 2. This change aims to improve the substrate-metal interface and thereby intrinsic quality factors for both resonators and qubits. After sputtering, the patterns are etched into the superconducting layer using reactive-ion etching with an SF6/O2 plasma. In contrast with the Al transmon capacitor plates commonly used, in this experiment we make them also from NbTiN, with an aim to improve the substrate-metal interface and avoid large AlOx surfaces which may house unwanted two-level systems. Only the Josephson junctions are made by the standard technique of Al-AlOx-Al double-angle evaporation. A further HF dip just prior to evaporation also helps to contact the junctions directly to the NbTiN capacitor plates. In ref. 1, air bridges were already used to cross the feed-
line over flux-bias lines on chip. Here, we extend this technique to allow the feedline to cross three readout resonators.

A key requirement for this experiment was the ability to match qubit frequencies without sacrificing coherence. Flux-bias lines allow easy compensation for mismatch, but at the cost of reduced coherence in the qubit detuned from its maximal frequency (coherence sweet spot\(^3\)). We aimed for identical maximum frequencies of \(Q_T\) and \(Q_B\), as determined by capacitor and junction geometries. The capacitors were easily matched in fabrication. We then selected the chip with the closest matching room-temperature resistance values for the relevant qubit junctions.

**B. Experimental Setup**

The chip is anchored to a copper cold-finger connected to the mixing chamber of a Leiden Cryogenics CS81 \(^3\)He/\(^4\)He dilution refrigerator with 7 mK base temperature. A copper can seals the sample space, with an inner surface that is coated with a mixture of Stycast 2850 and silicon carbide granules (15 to 1000 nm diameter) used for infrared absorption\(^4\). The copper can is in turn magnetically shielded by an aluminum enclosure and two outer Cryophy enclosures (1 mm thick)\(^2\).

A complete wiring schematic showing all cryogenic and room-temperature components is shown in Fig. S2. The four analog channels of the Tektronix AWG5014C create the in-phase and in-quadrature pulses for \(Q_T\) and \(Q_B\) by single-sideband modulation of a common carrier. Because single-sideband modulation requires two AWG channels to modulate an IQ mixer, independent derivative-removal-via-adiabatic-gate (DRAG) tuning with the VSM therefore requires four AWG channels, irrespective of the number of qubits. The VSM can be scaled up to many output channels, and direct hardware savings can be realized as soon as three or more same-frequency qubits are driven by a single set of AWG inputs. These pulses are input at ports 1 and 2 of the VSM. The VSM combines these pulses with individually tuned insertion loss and phase to each of two outputs (labelled T and B). Input-output combinations can be switched on nanosecond timescales using the gate inputs of the VSM, provided by digital markers of the AWG5014C. A second AWG5014C with the appropriate carrier frequency is used to excite transmons \(Q_T\) and \(Q_B\) to the second-excited state (Sec. V), to pulse measurement tones, and to trigger the AlazarTech ATS9870 acquisition card.

**FIG. S2:** Detailed schematic of the experimental setup and optical image of the chip (installed in a Leiden Cryogenics CS81 dilution refrigerator). Note that while our VSM has two pairs of analog inputs [(1,2) and (3,4)] and four gate inputs (activating links from each pair to each output), the pair (3,4) and its associated gate inputs are not used throughout this experiment.
C. Device frequencies

| Qubit | $f_{\text{max}}$ (GHz) | $f_{\text{bias}}$ (GHz) | $f_{\text{res}}$ (GHz) |
|-------|-----------------|-----------------|-----------------|
| $Q_T$ | 6.277           | 6.220           | 6.700           |
| $Q_M$ | 6.551           | 6.551           | 6.733           |
| $Q_B$ | 6.220           | 6.220           | 6.800           |

TABLE S1: Table of sweet-spot frequencies $f_{\text{max}}$ and bias-point frequencies $f_{\text{bias}}$ of the three qubits, as well as the fundamental frequencies $f_{\text{res}}$ of their dedicated readout resonators at the bias point. The qubits are tuned into the bias point by a combination of spectroscopy and standard Ramsey experiments.

D. Qubit coherence times

![Graph showing qubit coherence times](image)

FIG. S3: Measurements of relaxation ($T_1$, top), Ramsey dephasing ($T_2^*$, middle) and echo dephasing ($T_2^E$, bottom) times for the three qubits at the bias point. When measuring $Q_T$ or $Q_B$, the other qubit is detuned by −50 MHz to suppress cross-coupling effects. Ramsey fringes for $Q_T$ (middle, left panel) fit better to a Gaussian (shown) than an exponential decay, reflecting the susceptibility of $Q_T$ to low-frequency flux noise away from its sweet-spot. $P_1$ denotes excited-state population.

II. VECTOR SWITCH MATRIX

A. Measured isolation

To characterize the isolation of the VSM in the range 4 to 8 GHz, we have measured the insertion loss between all input (1 and 2) and output ports (T and B) with static settings at the two gate inputs (Fig. S4). Ideally, each gate activates (on state) and deactivates (off state) the link of both inputs 1 and 2 to one output, independent of the other gate. As shown in Fig. S4, the typical relative isolation with the relevant gate in the off state is $\sim 50$ dB.

![Graph showing insertion loss between inputs and outputs of the VSM](image)

FIG. S4: Insertion loss between inputs and outputs of the VSM for four static combinations of the gate inputs. The insertion loss is measured relative to the level with both gate inputs activated ($|a_{\text{off}}|$). The black curves indicate the noise background in our scalar network analyzer measurement. The dashed vertical lines indicate the common frequency (6.220 GHz) of $Q_T$ and $Q_B$ at the bias point.

B. Individual qubit tune-up

The VSM enables independent control of the on/off state, insertion loss and phase for every input-output combination. We exploit this feature to perform DRAG-compensated pulses on $Q_T$ and $Q_B$, that are individually tailored for each qubit. Different types of gate errors, such as non-ideal in-phase and in-quadrature amplitudes, can be distinguished using an AllXY sequence$^5$, consisting of 21 combinations of two pulses drawn from the set $\{I, X, Y, X\pi/2, Y\pi/2\}$ (Table S2). Figure S5 shows AllXY sequence results for $Q_T$ and $Q_B$ as the amplitude of each quadrature on $Q_T$ is varied independently. While the AllXY signature of $Q_T$ reveals changing levels of amplitude and phase errors, there is no noticeable change in the AllXY signature of $Q_B$. This demonstrates the use of the VSM for individual tune-up of pulses for same-frequency qubits.
TABLE S2: The 21 two-pulse combinations comprising the AllXY pulse sequence$^8$.

III. PULSE-CALIBRATION ROUTINES

We tune up qubit pulses by alternating the calibration of in-phase and in-quadrature pulse amplitudes until a simultaneous optimum is found. The two calibration routines are discussed below.

A. Accurate in-phase pulse amplitude calibration

The in-phase quadrature amplitude is calibrated by first applying a $\pi/2$ pulse to the qubit, followed by a train of $\pi$ pulses. The pulse sequence is $(X_\pi)^N X_{\pi/2}|g\rangle$, where $|g\rangle$ is the qubit ground state and $N \in [0, 49]$. In the absence of gate errors and decoherence, the driven qubit would end on the equator of its Bloch sphere for all $N$. However, over- or under-driving produces a positive or negative initial slope on $P_1$ versus $N$, respectively (Fig. S6). We choose the in-quadrature amplitude that minimizes the absolute slope.

B. DRAG-parameter calibration

To minimize phase errors resulting from the presence of the second- and higher-excited states, we optimize the scaling of the in-quadrature pulse. As in conventional DRAG$^{6,7}$, we choose as the envelope of the in-quadrature pulse the derivative of the Gaussian envelope on the in-phase pulse. The DRAG scaling parameter is calibrated using the method detailed in ref. 5. Specifically, we measure the difference in excited-state population produced by the $X_\pi X_{\pi/2}$ and $X_\pi Y_{\pi/2}$ pulse combinations (AllXY ID 10 and 11). Ideally, for both, the final qubit state would lie on the equator. However, any phase error shifts the final excited-state population in opposite directions in these cases. We choose the DRAG scaling parameter minimizing this shift.

IV. GLOBAL BROADCASTING

Aside from single-qubit control and selective broadcasting, the VSM also allows global broadcasting of pulses to all qubits simultaneously by keeping the markers for both qubits on (Fig. S7). While this does provide simultaneous control of $Q_T$ and $Q_B$, marker control is needed to achieve independent control. Using RB, we measure the performance of both qubits when broadcasting pulses to both qubits, and compare the results with those obtained from single-qubit control (Fig. S7). The global broadcasting RB measurements were alternated with the single-qubit RB measurements, and aside from marker settings all other settings were identical. Comparison of the results in Fig. S7 show that the qubit gate performance does not depend on whether a single qubit is controlled, or both are controlled simultaneously through global broadcasting.
VI. CROSS-DRIVING EFFECTS

The isolated single-qubit control experiments in the main text [Fig. 2(g,h)] show that significant spurious excitations can build up in the idling qubit over the course of the long gate sequences tested in RB (particularly in the case of idling QB while driving QT). It may therefore initially be somewhat surprising that virtually the same individual qubit-control performance is achieved in both selective-broadcasting (Fig. 4, main text) and global-broadcasting (Fig. S7) scenarios. As discussed in the main text, the observed cross-excitation is unlikely to result from cross-coupling, primarily because a symmetrical quantum coupling should not result in strongly asymmetric effects on the different qubits. We show in Sec. VI A that the results are, however, consistent with the effects of cross-driving by numerically simulating RB with cross-driving under experimentally realistic conditions (using independently measured qubit and cross-driving parameters). Simulations are performed using QuTiP9,10. In Sec. VI B, we show that the residual cross-driving effects can be largely compensated for by an appropriate choice of Clifford decompositions.

A. Simulating cross-driving

We model our system as two uncoupled qubits, QT and QB, subject to T1 relaxation (with corresponding relaxation times) and cross-driving. We approximate the system dynamics using instantaneous unitary pulse operators from the standard Pauli set \( \{X, Y, X_{\pm \pi/2}, Y_{\pm \pi/2}\} \) with 20 ns delays of T1-only qubit relaxation between pulses implemented using a master equation. When applying a pulse to one qubit, cross-driving of the other qubit is implemented by applying a pulse with the same rotation axis, but with the original rotation angle multiplied by the relevant cross-driving ratio. We note that also trying to model the effects of qubit dephasing using a simple master equation does not produce RB data consistent with the experimental observations (e.g., Figs 2 and 4 of the main text). This reflects the non-uniform phase noise spectrum which affects the transmon qubit. The long RB pulse sequences consisting of \( \pi \) and \( \pi/2 \) assume that leakage is irreversible. We therefore model leakage using the following difference equation for \( \langle P_2 \rangle \):

\[
\langle P_2[m+1] \rangle - \langle P_2[m] \rangle \simeq t_p \langle N_p \rangle - \frac{t_p \langle N_p \rangle}{T_{2\rightarrow1}} \langle P_2[m] \rangle,
\]

where \( T_{2\rightarrow1} \) is the second-to-first-excited-state relaxation time. Assuming no initial population in the second-excited state, the solution is Eq. (2), which shows good agreement with measured data. We extract \( \kappa \) by fitting Eq. (2) to \( \langle P_2 \rangle \) data. \( T_{2\rightarrow1} \) is obtained from the best-fit decay constant (not directly measured) and \( \kappa \) from the best-fit prefactor.

V. LEAKAGE TO SECOND EXCITED STATE

Leakage is fundamentally different from unitary qubit errors. To quantify leakage, we monitor the populations \( P_i \) of the three lowest energy states \( \{0, 1, 2\} \) during RB and calculate the average values \( \langle P_i \rangle \) over all seeds. To do this, we calibrate the average signal levels \( V_i \) for the transmons in level \( i \), and perform each RB measurement twice, the second time with an added final \( \pi \) pulse on the 0–1 transition. This final \( \pi \) pulse swaps \( P_0 \) and \( P_1 \), leaving \( P_2 \) unaffected. Under the assumption that higher levels are unpopulated \( (P_0 + P_1 + P_2 = 1) \),

\[
\begin{bmatrix}
V_0 - V_2 & V_1 - V_2 \\
V_1 - V_2 & V_0 - V_2
\end{bmatrix}
\begin{bmatrix}
P_0 \\ P_1
\end{bmatrix}
= \begin{bmatrix}
S - V_2 \\ S' - V_2
\end{bmatrix},
\]

where \( S \) (\( S' \)) is the measured signal level without (with) final \( \pi \) pulse. The populations are extracted by matrix inversion.

Measuring \( \langle P_2 \rangle \) as a function of the number of Clifford gates allows us to estimate an average leakage per Clifford, \( \kappa \). Because the populations are ensemble averages over different random seeds, we assume that leakage of the average qubit-space populations to \( \langle P_2 \rangle \) is incoherent, and, provided \( \langle P_2 \rangle \) remains small (\( \kappa \) small), we also

FIG. S7: Global broadcasting of DRAG pulses to same-frequency qubits. (a) Illustration of global broadcasting. Two simultaneous pulses, one with Gaussian envelope at input 1, and another with derivative-of-Gaussian envelope at input 2, are simultaneously directed to QT and QB (both markers always on). The insertion loss and phase shift of each pulse is separately optimized for each output to produce precision DRAG pulses for each qubit. (b,c) Comparison of single-qubit driving versus driving both qubits (broadcasting) by RB of Clifford gates composed from \( \pi/2 \) and \( \pi \) pulses.\(^{\text{x}} \) Average population of QT and QB in the ground, first- and second-excited state (\( \langle P_0 \rangle \), \( \langle P_1 \rangle \) and \( \langle P_2 \rangle \), resp.) as a function of the number of Clifford gates applied. Curves are the best fits of single exponentials with offsets to the populations. The single-qubit Clifford-gate fidelity for each qubit is extracted from the decay of the corresponding ground-state population when using global broadcasting.
FIG. S8: Simulation of cross-driving effects during RB. The results shown are averaged over ten runs, using the single-qubit minimal-set decomposition (red), and using the selective-broadcasting asymmetric and symmetric 5-primitives schemes (green and blue, respectively). Cross-driving effects are largely suppressed in the 5-primitives schemes by choosing the five pulse primitives such that constituent pulses largely cancel out. The symmetric 5-primitives scheme further reduces cross-driving effects by alternating between the five pulse primitives and the inverse pulses.

Randomized benchmarking is implemented by generating independent Clifford sequences for each qubit. We decompose Clifford gates using either the minimal set decomposition or one of the selective-broadcasting schemes. Figure S8 shows simulated results of cross-driving for the isolated single-qubit control scenario reported in Fig. 2 of the main text. In this section, we are only concerned with the red curves, which correspond to implementing single-qubit RB with the standard set of pulse decompositions\(^8\). These simulations can be compared directly with the curves in Fig. 2(g,h). While the maximum excitation population observed in the simulations is larger than the value observed in the experiments, the simulations for both qubits show the same qualitative behaviour as the measured data. The quantitative difference may be explained by the fact that the direct measurements of cross-driving were made at a different time from the main measurement run and we observed few-percent fluctuations in cross-driving levels over time.

As discussed in the main text, while the plots of cross-excitation during RB are useful diagnostics of the presence of a spurious cross-driving effect, they may give a misleading impression when presented in parallel with RB results. Although the decay curves look superficially similar, they should not be interpreted in the same way. By contrast, the technique of interleaved RB (IRB), which was introduced to enable rigorous quantification of the performance of individual gates, allows us to calculate a meaningful error per Clifford for the idling operation\(^11\). In IRB, the usual random sequence of Cliffords is alternated with identical repetitions of an individual gate. By comparing the interleaved decay rate with the decay rate for a standard RB measurement, it is possible to calculate a robust error per gate for the individual gate in question. In this context, the target gate is the nominal identity operation on one qubit which results from a random Clifford being applied to the other qubit. The IRB pulse sequence is therefore identical to the sequence implemented in the sequential selective-broadcasting scheme (see Fig. 4 of the main text). In the main text, we use the formulas in ref. 11 to calculate the idling error per Clifford, but for these simulations, the performance of sequential selective-broadcasting already provides a simple way to assess the performance of cross-excitation during idling. Figure S9 shows that idling performance as quantified by RB is limited mainly by \(T_1\) relaxation. Finally, when identical gate sequences are being applied to both qubits, cross-driving will result in a small amount of over-driving on each qubit (overdriving ratio \(r_o\)), which would also look like an error in pulse rotation angle. Figure S10 shows that the Clifford error is insensitive to both cross-driving and over-driving to first order.
B. Making pulse sequences robust to cross-driving

We have already shown that cross-excitation does not have a dominant effect on single-qubit control in both global and selective broadcasting. We show here that any residual effect can be largely eliminated also while a qubit is idling by choosing robust pulse sequences for decomposing the Clifford gates.

If a qubit is idle, every pulse that is applied to the driven qubit rotates the idle qubit by an amount depending on the cross-driving ratio. The random application of successive pulses to the driven qubit can therefore be viewed as a random walk for the idle qubit. As we will discuss in more detail in Sec. VIII, there are many ways to compose a given Clifford gate from a small set of standard rotations. By choosing the constituent pulses in such a way that their combined application largely cancels out, cross-driving effects can be greatly reduced. In the standard set of Clifford decompositions, the decompositions involve a majority of pulses rotating in the positive direction, biasing the random walk and producing a pronounced net cross-driving effect. This effect can be countered by choosing decompositions which minimize the bias. We have implemented this in the 5-primitives scheme, by choosing the first three pulses, \(\{X_{\pi/2}, Y_{\pi/2}, X_{\pi/2}\}\), to be positive rotations, and the last two, \(\{X_{-\pi}, Y_{-\pi}\}\), to move in the negative direction. Even though the pulse subset that is applied depends on the Clifford chosen, the pulses still largely cancel out after applying many Cliffords. Furthermore, as the single-qubit Clifford operations form a group, the inverse of all Cliffords also form the Clifford group. The complete inverse of the five pulse primitives, \(\{X_{\pi}, Y_{\pi}, X_{-\pi/2}, Y_{-\pi/2}, X_{-\pi/2}\}\), can therefore also generate each of the 24 Cliffords using an appropriate subset of the pulses. By alternating between the normal five-pulse primitives and the inverted five pulses, cross-driving effects can be further reduced. (In fact, we note that this exactly eliminates all cross-driving that occurs via leakage in the VSM, because all pulses are always present at that distribution stage.) We refer to this as the symmetric 5-primitives technique and this is the technique we implement in the main experiments described in Fig. 4 of the main text. Our simulations in Fig. S8 show that the asymmetric 5-primitives technique already dramatically reduces the effect of cross-driving, and in the case of isolated single-qubit control, cross-driving is effectively eliminated completely using the symmetric 5-primitives scheme. This is also confirmed by measurements of cross-driving for the three selective-broadcasting schemes (Fig. S11), where we implement the symmetric 5-primitives technique. While we have only demonstrated this technique for the 5-primitives scheme, it could also be relatively straightforwardly applied to the sequential scheme, but the far better scaling of the 5-primitives scheme make it more interesting for scaling up to larger system sizes.

VII. CLIFFORD PULSE DECOMPOSITION

The decompositions of the 24 single-qubit Clifford gates into a minimal set of \(\pi/2\) and \(\pi\) pulses and into the 5-primitives scheme are shown in Table S3. We note that the specific choice of pulse primitives in the 5-primitives scheme is not unique, but at least five are required (four pulses allow a maximum of 16 unique gate decompositions, compared with the 24 single-qubit Cliffords).

```
| Clifford ID | Minimal set decomposition | 5-primitives decomposition |
|-------------|---------------------------|---------------------------|
|             | First                     | Second                    | Third                     | First | Second | Third |
| 1            | \(X\)                     | \(0\)                     | \(0\)                     | \(0\) | \(0\)  | \(0\)  |
| 2            | \(Y_{\pi/2}\) X_{\pi/2}  | \(0\)                     | \(1\)                     | \(1\) | \(0\)  | \(0\)  |
| 3            | \(X_{-\pi/2}\) Y_{-\pi/2} | \(1\)                     | \(1\)                     | \(0\) | \(1\)  | \(0\)  |
| 4            | \(X_{\pi}\)               | \(0\)                     | \(0\)                     | \(0\) | \(1\)  | \(1\)  |
| 5            | \(Y_{-\pi/2}\) X_{-\pi/2} | \(0\)                     | \(1\)                     | \(1\) | \(0\)  | \(0\)  |
| 6            | \(X_{\pi/2}\) Y_{-\pi/2}  | \(1\)                     | \(1\)                     | \(0\) | \(0\)  | \(0\)  |
| 7            | \(Y_{-\pi/2}\) X_{-\pi/2} | \(0\)                     | \(1\)                     | \(1\) | \(1\)  | \(1\)  |
| 8            | \(X_{\pi/2}\) Y_{-\pi/2}  | \(1\)                     | \(1\)                     | \(0\) | \(0\)  | \(0\)  |
| 9            | \(X_{\pi/2}\) Y_{-\pi/2}  | \(1\)                     | \(1\)                     | \(0\) | \(0\)  | \(0\)  |
| 10           | \(X_{\pi/2}\) Y_{-\pi/2}  | \(1\)                     | \(1\)                     | \(0\) | \(0\)  | \(0\)  |
| 11           | \(Y_{\pi/2}\) X_{-\pi/2}  | \(0\)                     | \(1\)                     | \(1\) | \(0\)  | \(0\)  |
| 12           | \(X_{-\pi/2}\) Y_{-\pi/2} | \(1\)                     | \(1\)                     | \(0\) | \(0\)  | \(0\)  |
| 13           | \(Y_{-\pi/2}\) X_{\pi/2}  | \(0\)                     | \(1\)                     | \(1\) | \(0\)  | \(0\)  |
| 14           | \(X_{-\pi/2}\) Y_{-\pi/2} | \(0\)                     | \(1\)                     | \(1\) | \(0\)  | \(0\)  |
| 15           | \(X_{\pi/2}\) Y_{-\pi/2}  | \(1\)                     | \(1\)                     | \(0\) | \(0\)  | \(0\)  |
| 16           | \(Y_{-\pi/2}\) X_{\pi/2}  | \(0\)                     | \(1\)                     | \(1\) | \(0\)  | \(0\)  |
| 17           | \(X_{\pi/2}\) Y_{-\pi/2}  | \(0\)                     | \(1\)                     | \(1\) | \(0\)  | \(0\)  |
| 18           | \(X_{\pi/2}\) Y_{-\pi/2}  | \(1\)                     | \(1\)                     | \(0\) | \(0\)  | \(0\)  |
| 19           | \(Y_{-\pi/2}\) X_{\pi/2}  | \(0\)                     | \(1\)                     | \(1\) | \(0\)  | \(0\)  |
| 20           | \(X_{\pi/2}\) Y_{-\pi/2}  | \(1\)                     | \(1\)                     | \(0\) | \(0\)  | \(0\)  |
| 21           | \(Y_{-\pi/2}\) X_{\pi/2}  | \(0\)                     | \(1\)                     | \(1\) | \(0\)  | \(0\)  |
| 22           | \(X_{\pi/2}\) Y_{\pi/2}   | \(1\)                     | \(1\)                     | \(0\) | \(0\)  | \(0\)  |
| 23           | \(Y_{\pi/2}\) X_{\pi/2}   | \(1\)                     | \(1\)                     | \(0\) | \(0\)  | \(0\)  |
| 24           | \(X_{\pi/2}\) Y_{\pi/2}   | \(1\)                     | \(1\)                     | \(0\) | \(0\)  | \(0\)  |
```

TABLE S3: Two decompositions of the 24 single-qubit Clifford gates. The first, taken from ref. 8, minimizes the number of \(\pi/2\) and \(\pi\) pulses around the \(\pm x\) and \(\pm y\) axes. The second is our decomposition into 5 primitives. Pulses are applied from left to right.
VIII. COMPILED SELECTIVE BROADCASTING ALGORITHM

A. Finding the optimal pulse sequence

When using a selective broadcasting architecture to send pulse sequences to multiple qubits, pulses can be directed to any subset of the qubits, but distinct pulses may not be applied simultaneously. In compiled selective broadcasting, the total number of pulses required to implement single-qubit gates on all qubits of a multi-qubit system is minimized by searching all possible combinations of single-qubit Clifford decompositions and grouping together like pulses where possible. In this section, we introduce an algorithm for determining the shortest compiled pulse sequence implementing independent single-qubit Clifford gates on \( n \) qubits.

On average, there are approximately 38 distinct decompositions for each single-qubit Clifford gate, given the basis set of \( X \) and \( Y \) pulses: \( \{ I, X, Y, X \pm iY \} \), resulting in approximately \( 38^n \) different decompositions for a given \( n \)-qubit combination of Cliffords. Here, we only consider sequences of up to four pulses, because the 5-primitives decomposition identified in Table S3 already provides a recipe for decomposing an arbitrary \( n \)-qubit Clifford combination into five pulses. We do not include trivial decompositions where sequential pulses cancel out.

Given a particular choice of \( n \) Cliffords \( (C_1, \ldots, C_n) \), where \( \alpha_i \) is the Clifford ID for qubit \( i \), we write a specific decomposition as \( (P_1^{\alpha_1}, \ldots, P_m^{\alpha_m}) \), where \( P_i^{\alpha} \) is the \( i \)-th of \( m \) pulses which implement \( C_{\alpha_i} \). While this already fixes the order in which pulses must be applied to individual qubits, we still have the freedom to choose in which order the distinct pulses are applied to different qubits. For each possible decomposition, we use the following recursive algorithm to search and minimize over all possible pulse orderings.

We first define an empty broadcasting sequence \( P_{\text{seq}} \) to store the compiled multi-qubit pulse sequence. In order to convert from parallel single-qubit pulse sequences to the single broadcasting sequence \( P_{\text{seq}} \), we define a vector of indices \( \beta = (\beta_1, \ldots, \beta_n) \) to store the current position in each single-qubit sequence. Initially, \( \beta = (1, \ldots, 1) \). At each instant, \( P_{\beta} = (P_{\beta_1}^{\alpha_1}, \ldots, P_{\beta_n}^{\alpha_n}) \) contains the next pulses to be applied to each qubit. When \( \beta_i = m_i + 1 \), the pulse sequence for that qubit is completed, and so there is no \( P_{\beta_i} \) to be added to \( P_{\beta} \). The recursive part of the algorithm then proceeds as follows:

1. Define \( \beta_1 \) to be the set of distinct pulses in \( P_{\beta_1} \).

   **If:** \( P_{\beta_1} \) is empty, store the number of pulses in \( P_{\text{seq}} \) and abort this recursion branch (\( P_{\text{seq}} \) is a completed pulse sequence such that all Cliffords are applied to the corresponding qubits).

   **Else:** Continue.

2. For each pulse \( P \) in \( P_{\beta_1} \), perform the following steps:

   (i) Append \( P \) to \( P_{\text{seq}} \).

   (ii) Copy indices \( \beta \) to \( \beta_{\text{new}} = (\beta_{\text{new}}^1, \ldots, \beta_{\text{new}}^n) \).

   (iii) For all indices \( i \) for which \( \beta_{\text{new}}^i = P \), increase the index \( \beta_{\text{new}}^i = \beta_i + 1 \).

   (iv) Recursively loop to step 1 using the new indices \( \beta_{\text{new}} \).

After considering all possible pulse sequences and looping over all possible decompositions, choose the sequence with the minimum number of pulses \( N_p \) in \( P_{\text{seq}} \).

This algorithm determines the minimum number of pulses \( N_p \) required to implement a given \( n \)-qubit combination of single-qubit Clifford gates. However, this becomes prohibitively resource intensive as the number of qubits increases. It is therefore important to assess how the performance of the optimal compiled decomposition compares with the 5-primitives decomposition, which can be applied to any number of qubits without any extra overhead in resources (in neither calculation time nor sequence length). To do this, we use the average number of pulses \( N_p \) required per \( n \)-qubit combination of Cliffords (per Clifford).

In the case of compiled selective broadcasting, finding \( \langle N_p \rangle \) requires minimizing the sequence length for all \( 24^n \) possible Clifford combinations. This problem again scales exponentially with \( n \). For example, for \( n = 5 \) qubits, this requires \( 24^5 \cdot 38^5 \approx 6.3 \cdot 10^{14} \) repetitions of the complete recursive search described above. Nevertheless, by employing a number of optimizations described in the next section, we have exactly calculated \( \langle N_p \rangle \) for \( 1 \leq n \leq 5 \) qubits in under 2 hours. Using random sampling (finding the shortest pulse sequence for a random sample of Clifford combinations), we also approximated \( \langle N_p \rangle \) for \( 5 \leq n \leq 10 \). Exact and approximate results for \( n = 5 \) agree. As shown in Table S4, the improvement offered by compiled selective broadcasting over the 5-primitives method is already less than one pulse per Clifford and continues to decrease rapidly. Considering how badly the resource overhead scales for increasing numbers of qubits for finding a compiled sequence, it is questionable whether compiling offers any significant benefit over using the prescriptive 5-primitives approach when scaling up to larger system sizes.
TABLE S4: The average number of pulses \( \langle N_p \rangle \) required to perform one Clifford gate on each of \( n \) qubits in compiled selective broadcasting. Exact values were obtained for \( 1 \leq n \leq 5 \) in under two hours, using the outlined optimizations. Approximate values were obtained for \( 5 \leq n \leq 10 \) using random sampling. These results are plotted in Fig. 4 of main text.

| \( n \) | Exact calculation \( \langle N_p \rangle \) | Random sampling \( \langle N_p \rangle \) |
|---|---|---|
| 1 | 1.875 | 5.000 (2) |
| 2 | 2.925 | 6.380 (12) |
| 3 | 3.521 | 7.570 (15) |
| 4 | 3.874 | 8.721 (10) |
| 5 | 4.137 | 9.781 (8) |
| 6 | 4.253 | 10. 4.857 (24) |

B. Optimizations for the Clifford compilation algorithm

In the first optimization, we place an upper bound \( N^{ub} \) on the pulse sequence length. The upper bound \( N^{ub} \) is given by the minimum number of pulses found so far that can compile a given Clifford combination. At each stage, the algorithm checks if the sum of the pulses in \( P_{seq} \) and all distinct pulses left is equal to or greater than \( N^{ub} \). If this is the case, a shorter combination of pulses using \( P_{seq} \) is not possible, so we stop considering this sequence and proceed to the next one. Initially \( N^{ub} = 5 \), as the 5-primitives method proves that there is always a decomposition of an arbitrary number of Cliffords into 5 pulses. Note that, as the limit \( N^{ub} \) moves down, the frequency at which the algorithm stops considering sequences increases.

The second optimization relies on decompositions with fewer pulses being more likely to result in an optimal Clifford compilation. The decompositions of every Clifford are therefore arranged in ascending number of pulses. The first decompositions compared are then those with the minimum number of pulses; these have the highest probability of finding an optimal Clifford compilation. Even if an optimal Clifford compilation is not found, it is more likely that \( N^{ub} \) will be low. This optimization is especially effective in combination with the first optimization.

The third optimization places a lower bound \( N^{lb} \) on the number of pulses. For a given Clifford combination \((C_{\alpha_1}, \ldots, C_{\alpha_n})\), \( N^{lb} \) is found by looking at the minimum number of pulses \( N_p \) previously found for all \( n-1 \) Clifford subsets. Since \( N_p \) for the \( n \) Cliffords can never be less than \( N_p \) for any of the \( n-1 \) Clifford subsets, the maximum length of the \( n-1 \) Clifford subsets therefore places a lower bound \( N^{lb} \) on \( N_p \) for the \( n \) Cliffords. This means that if a pulse sequence is found whose length is equal to \( N^{lb} \), it is an optimal Clifford compilation, and all further search is aborted. This is in contrast to the first optimization, where only the particular sequence of pulses is aborted upon reaching \( N^{lb} \). Furthermore, as \( n \) increases, it becomes increasingly likely that the lower bound is 5. In this case, \( N^{lb} = N^{ub} \), and so the 5-primitives method is an optimal Clifford compilation. This optimization results in the largest gain in computation time, by several orders of magnitude.

In the fourth and most complicated optimization, all decompositions composed of three pulses or less are separated from those composed of four pulses. First, all combinations of Clifford decompositions composed of three pulses or less are compared. This reduces the average number of decompositions per Clifford, from 38 to 7, resulting in an exponentially reduced number of total decomposition combinations. It is, however, not always the case that the optimal Clifford compilation is found using only up to three pulses per decomposition; sometimes optimal Clifford compilations require that one of the decompositions is composed of four pulses. However, after comparing decompositions of three pulses or less, these four-pulse decompositions only need to be considered when \( N^{lb} \leq 4 \) and \( N^{ub} = 5 \). If there is a sequence containing a four-pulse decomposition that outperforms any found using up to three-pulse decompositions and the 5-primitives method, the sequence must consist of four pulses. Only one Clifford then has a four-pulse decomposition, while all other Cliffords are subsets of these four pulses. We therefore loop, for every Clifford, over each of the four-pulse decompositions, and test whether every other Cliffords can be decomposed into a subset of these four pulses. This changes the comparison of four-pulse decompositions from scaling exponentially with \( n \) to scaling linearly.

The fifth and final optimization is only of use when all different Clifford combinations need to be considered to determine \( \langle N_p \rangle \). It stems from the observation that an optimal compilation for a certain Clifford combination \((C_{\alpha_1}, \ldots, C_{\alpha_n})\) is the same as for any permutation of these Cliffords. We therefore only determine an optimal Clifford compilation when \( \beta_1 \leq \cdots \leq \beta_n \). This reduces the number of calculations exponentially (81 times fewer computations for \( n = 5 \)).

[1] Ristè, D., Poletto, S., Huang, M. Z., Bruno, A., Vesterinen, V., Saira, O. P. et al., Detecting bit-flip errors in a logical qubit using stabilizer measurements. *Nat. Commun.* 6 (2015).
[2] Bruno, A., de Lange, G., Asaad, S., van der Enden, K. L., Langford, N. K. & DiCarlo, L., Reducing intrinsic loss in superconducting resonators by surface treatment and deep etching of silicon substrates. *Appl. Phys. Lett.* 106, 182601 (2015).
[3] Schreier, J. A., Houck, A. A., Koch, J., Schuster,
D. I., Johnson, B. R., Chow, J. M. et al., Suppressing charge noise decoherence in superconducting charge qubits. Phys. Rev. B 77, 180502 (2008).

[4] Barends, R., Wenner, J., Lenander, M., Chen, Y., Bialczak, R. C., Kelly, J. et al., Minimizing quasiparticle generation from stray infrared light in superconducting quantum circuits. Appl. Phys. Lett. 99, 113507 (2011).

[5] Reed, M., Entanglement and quantum error correction with superconducting qubits. PhD Dissertation, Yale University (2013).

[6] 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).

[7] Chow, J. M., DiCarlo, L., Gambetta, J. M., Motzoi, F., Frunzio, L., Girvin, S. M. et al., Optimized driving of superconducting artificial atoms for improved single-qubit gates. Phys. Rev. A 82, 040305 (2010).

[8] Epstein, J. M., Cross, A. W., Magesan, E. & Gambetta, J. M., Investigating the limits of randomized benchmarking protocols. Phys. Rev. A 89, 062321 (2014).

[9] Johansson, J., Nation, P. & Nori, F., Qutip: An open-source python framework for the dynamics of open quantum systems. Computer Physics Communications 183, 1760 – 1772 (2012).

[10] Johansson, J., Nation, P. & Nori, F., Qutip 2: A python framework for the dynamics of open quantum systems. Computer Physics Communications 184, 1234 – 1240 (2013).

[11] Magesan, E., Gambetta, J. M., Johnson, B. R., Ryan, C. A., Chow, J. M., Merkel, S. T. et al., Efficient measurement of quantum gate error by interleaved randomized benchmarking. Phys. Rev. Lett. 109, 080505 (2012).