



This is the accepted manuscript made available via CHORUS. The article has been published as:

Learning-Based Calibration of Flux Crosstalk in Transmon Qubit Arrays

Cora N. Barrett, Amir H. Karamlou, Sarah E. Muschinske, Ilan T. Rosen, Jochen Braumüller, Rabindra Das, David K. Kim, Bethany M. Niedzielski, Meghan Schuldt, Kyle Serniak, Mollie E. Schwartz, Jonilyn L. Yoder, Terry P. Orlando, Simon Gustavsson, Jeffrey A. Grover, and William D. Oliver

Phys. Rev. Applied **20**, 024070 — Published 28 August 2023

DOI: [10.1103/PhysRevApplied.20.024070](https://doi.org/10.1103/PhysRevApplied.20.024070)

Learning-based Calibration of Flux Crosstalk in Transmon Qubit Arrays

Cora N. Barrett,^{1,2,*} Amir H. Karamlou,^{1,3,†} Sarah E. Muschinske,^{1,3} Ilan T. Rosen,¹ Jochen Braumüller,¹ Rabindra Das,⁴ David K. Kim,⁴ Bethany M. Niedzielski,⁴ Meghan Schuldt,⁴ Kyle Serniak,^{1,4} Mollie E. Schwartz,⁴ Jonilyn L. Yoder,⁴ Terry P. Orlando,^{1,3} Simon Gustavsson,¹ Jeffrey A. Grover,¹ and William D. Oliver^{1,3,4,5,‡}

¹*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

²*Department of Physics, Wellesley College, Wellesley, MA 02481, USA*

³*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

⁴*MIT Lincoln Laboratory, Lexington, MA 02421, USA*

⁵*Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

(Dated: August 11, 2023)

Superconducting quantum processors comprising flux-tunable data and coupler qubits are a promising platform for quantum computation. However, magnetic flux crosstalk between the flux-control lines and the constituent qubits impedes precision control of qubit frequencies, presenting a challenge to scaling this platform. In order to implement high-fidelity digital and analog quantum operations, one must characterize the flux crosstalk and compensate for it. In this work, we introduce a learning-based calibration protocol and demonstrate its experimental performance by calibrating an array of 16 flux-tunable transmon qubits. To demonstrate the extensibility of our protocol, we simulate the crosstalk matrix learning procedure for larger arrays of transmon qubits. We observe an empirically linear scaling in calibration time with system size while maintaining a median qubit frequency error below 300 kHz.

I. INTRODUCTION

Superconducting quantum processors comprising flux-tunable transmon qubit arrays are at the forefront of contemporary digital quantum computation [1, 2] and analog quantum simulation and emulation [3–5]. Flux-tunable qubits enable controllable, strong qubit-qubit interactions and high-fidelity two-qubit gates [6, 7] in many-qubit systems by reducing parasitic couplings and qubit frequency crowding. A central requirement for operating such quantum devices is the accurate and precise frequency control of each tunable element using local flux lines, which can be used to tune each qubit individually. While a current applied to a particular flux-line antenna is designed to address only one qubit or coupler, there are at least two mechanisms by which flux may couple to additional elements. One is through the direct, unwanted inductive coupling from the antenna to other qubits and couplers. A second is via the send and return path of the applied current, which may similarly induce unwanted magnetic flux in other qubits as it traverses the flux line and ground plane. The net effect is commonly referred to as flux crosstalk. Crosstalk can be characterized and compensated by measuring the response of each qubit to the flux generated by every flux line independently and presuming linear superposition, or by using an iterative approach [8] and optimization [9] in systems with a non-linear response.

Machine learning techniques have been applied extensively to calibrating quantum dots [10–18]. For superconducting quantum processors, machine learning approaches have been used, for example, for automated recalibration of system parameters [19], optimization of qubit frequency layouts [20], discrimination of qubit states [21–25], and calibration of single- and two-qubit gates [26, 27]. In this work, we demonstrate and analyze a learning-based approach for characterizing flux crosstalk on flux-tunable transmon devices. In comparison to previous works, our approach does not involve direct measurement of crosstalk matrix elements and requires relatively few measurements.

Flux-tunable transmons comprise two Josephson junctions forming a superconducting quantum interference device (SQUID) in parallel with a shunting capacitor [28]. In this circuit, the Josephson energy E_J is tuned by threading an external magnetic flux Φ_{ext} through the SQUID loop. The transition frequency between the ground and the first-excited state of the transmon in response to the applied magnetic flux is approximately given by [28]

$$f(\Phi_{\text{ext}}) \approx \left(f^{\text{max}} + \frac{E_C}{h} \right) \sqrt{d^2 + (1-d^2) \cos^2 \left(\pi \frac{\Phi_{\text{ext}}}{\Phi_0} \right)} - \frac{E_C}{h} \quad (1)$$

where $f^{\text{max}} = (\sqrt{8E_J E_C} - E_C)/h$ is the maximum qubit frequency, assuming $E_J \gg E_C$, and E_C is the transmon charging energy. The asymmetry parameter d of the SQUID junctions is given by $d = |(E_{J,2} - E_{J,1})/(E_{J,2} + E_{J,1})|$, where $E_{J,1}$ and $E_{J,2}$ are the Josephson energies of the two SQUID junctions. The transmon spectrum in

* These authors contributed equally to this work.

† These authors contributed equally to this work.; karamlou@mit.edu

‡ william.oliver@mit.edu

Eq. 1 provides a formula to estimate the applied magnetic flux required for tuning the qubit to a particular frequency.

In a flux-tunable transmon processor, the magnetic flux is applied by running an electric current through a flux line terminated by an antenna that is near the target qubit and inductively coupled to its SQUID loop. The current is generated at room temperature either by using an active current source or by using a voltage source outputting voltage V across a series resistance R . In either case, it is important to use a “stiff” current source with high output resistance to ameliorate the impact of temperature-dependent line resistance inside the refrigerator leading to the qubits. In this work, we use a voltage source and resistor $1\text{ k}\Omega$ to apply a current to each of our flux bias lines. The magnetic flux can be expressed as $\Phi_{\text{ext}} = V/V^{\Phi_0} + \Phi_{\text{offset}}$, where V^{Φ_0} is the voltage required to tune the qubit by one magnetic flux quantum Φ_0 , and Φ_{offset} is a flux offset due to magnetic fields produced by vortices trapped in the superconducting ground plane or other non-controllable sources of static magnetic field.

Magnetic flux crosstalk can be treated as a linear process; a vector of voltages \vec{V} applied to the flux lines is related to the magnetic flux $\vec{\Phi}_{\text{ext}}$ experienced by the qubits by the relation

$$\vec{\Phi}_{\text{ext}} = (\mathbf{V}^{\Phi_0})^{-1} \mathbf{S} \vec{V} + \vec{\Phi}_{\text{offset}}, \quad (2)$$

where \mathbf{V}^{Φ_0} is a diagonal matrix with $\mathbf{V}_{i,i}^{\Phi_0}$ corresponding to the V^{Φ_0} of qubit i . \mathbf{S} is the flux crosstalk sensitivity matrix, with $\mathbf{S}_{i,j} = \partial V_i / \partial V_j$ representing the voltage response of qubit i to a voltage signal applied to qubit j . In this representation, the diagonal elements of \mathbf{S} are 1, and characterizing \mathbf{S} along with $\vec{\Phi}_{\text{offset}}$ enables us to compensate for the crosstalk and set the qubit frequencies more precisely. Before this compensation, we observe a spread in frequency error Δf as shown in Fig. 1C.

The flux crosstalk sensitivity matrix for an array of N flux-tunable transmon qubits contains N^2 elements. We note that since the qubit array and routing layouts are not generally symmetric, the matrix \mathbf{S} is not guaranteed to be symmetric. In general, $\mathbf{S}_{i,j} \neq \mathbf{S}_{j,i}$, and so all N^2 elements need to be characterized. Typically, each element $\mathbf{S}_{i,j}$ is individually characterized by sweeping voltage V_j targeting qubit j and measuring the response of qubit i [5, 29, 30]. This approach, however, is not extensible for characterizing the flux crosstalk of large transmon qubit arrays (see Appendix B). The sample is fabricated using a 3D-integrated, flip-chip process [31], where the resonators and the flux lines are located on the interposer tier, and the qubits are located on a separate qubit tier. This learning-based calibration approach has been employed to calibrate the flux crosstalk on planar as well as flip-chip devices [4, 5, 32].

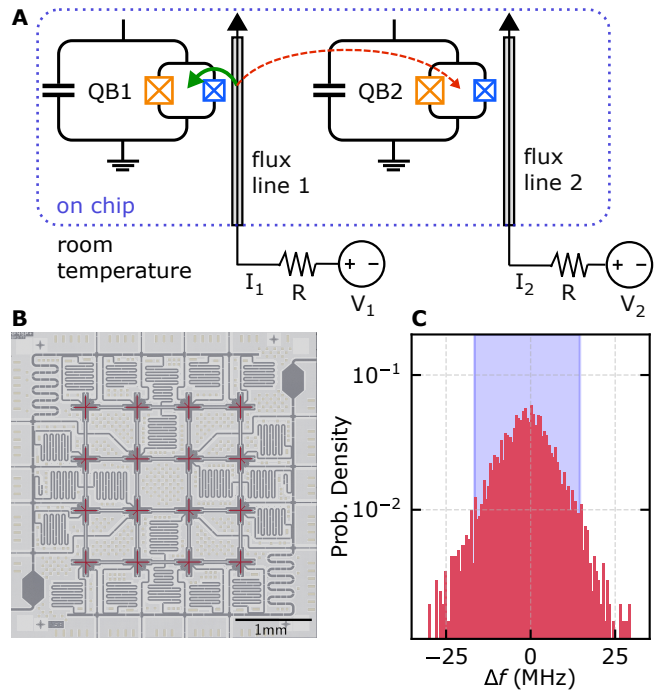


FIG. 1. **The Flux Crosstalk Model** (A) Flux crosstalk concept for superconducting transmon circuits. Current I_1 passing near the SQUID of qubit 1 supplies an intended amount of magnetic flux (green arrow). Due to crosstalk, other qubits on the device may experience an unintended flux from this current (dashed red arrow). (B) Optical image of the qubit tier of the 16-qubit transmon array fabricated using a 3D-integrated, flip-chip process (see Appendix A for details of the device). The capacitor pads of the qubits are false-colored maroon. (C) Experimental distribution of the difference between the measured qubit frequency and the target frequency $\Delta f = f_q - f_{\text{target}}$ without crosstalk correction for 200 target frequency vectors \vec{f}_{target} . For each \vec{f}_{target} , all 16 qubits are simultaneously biased to random frequencies in the region spanned by $100\text{ MHz} - 1\text{ GHz}$ below the maximum qubit frequency, and then each qubit frequency is measured. The shaded region indicates the 5th to 95th percentiles of the distribution.

II. CALIBRATION PROTOCOL

A. Learning the flux crosstalk matrix

We learn the flux crosstalk matrix \mathbf{S} by attempting to set our qubits to a target frequency layout, measuring the qubit frequencies as a result of applied external flux, and then using the frequency error to optimize \mathbf{S} . More specifically, we apply a set of voltages \vec{V} , where the i -th element of \vec{V} denotes the voltage applied to the flux line targeting qubit i , and measure the frequency of qubit i to infer the flux $\Phi_{\text{ext},i}$ via Eq. 1. Minimizing the difference between the measured flux, $\Phi_{\text{ext},i}$, and the estimated flux, which depends upon \vec{V} and \mathbf{S} , optimizes \mathbf{S} .

We generate \vec{V} by randomly selecting a frequency for

each qubit, \vec{f} , subject to a few constraints. First, we require that each qubit frequency falls in the range spanned by approximately 100 MHz – 1 GHz below its maximum qubit frequency (colloquially referred to as the “sweet spot” due to its first-order insensitivity to flux noise), as illustrated in Fig. 2A. Second, to reduce frequency shifts due to resonant interaction between qubits, we require that neighboring qubits in the array are far detuned (> 200 MHz), and the detuning between any two qubits is at least 50 MHz. The target frequency range includes regions of the transmon spectrum close to the sweet spot, which are less sensitive to changes in flux. This qubit frequency placement is necessary for sufficiently detuning the qubits. The detuning mitigates frequency shifts due to resonant interaction between qubits (see Appendix A), enabling us to accurately measure the frequencies of all qubits simultaneously.

To minimize bias in this quasi-random selection, we randomly permute the order in which target qubit frequencies within \vec{f} are chosen. This random permutation ensures that, given the detuning constraints, the qubits with higher sweet-spot frequencies in the latter half of \vec{f} are not regularly placed at the higher frequency end of the training region. Using the transmon spectrum in Eq. 1, we can calculate the magnetic flux values corresponding to the target frequencies. An initial guess for \vec{V} can be obtained from the relationship described in Eq. 2 by assuming $\mathbf{S} = \mathbf{I}$ or using an estimate of the matrix. All elements of \vec{V} will be nonzero since each qubit is biased off of its sweet spot.

We apply \vec{V} and measure the frequency of each qubit. In our experiments, these frequency measurements are performed simultaneously. We then convert the measured frequencies into the flux experienced by each SQUID, $\vec{\Phi}_{\text{meas}}$. By repeating this procedure for M iterations, we obtain a set $\{\vec{V}_i, \vec{\Phi}_{\text{meas},i}\}_{i=1:M}$ of input voltages and the resulting measured fluxes experienced by the qubits. Using this data set, we train the elements of the k^{th} row of \mathbf{S} by minimizing the mean-squared-error cost function

$$C(\mathbf{S}_k) = \frac{1}{M} \sum_{i=1}^M \left\| (\vec{\Phi}_{\text{meas},i})_k - \left[(\mathbf{V}_{k,k}^{\Phi_0})^{-1} \mathbf{S}_k \vec{V}_i + (\vec{\Phi}_{\text{offset}})_k \right] \right\|^2. \quad (3)$$

The first term in the cost function sum corresponds to the measured flux on qubit k , and the second term corresponds to the estimated flux based on our crosstalk matrix and the applied voltages. We minimize $C(\mathbf{S}_k)$ with the L-BFGS gradient descent optimization algorithm [33] implemented in PyTorch [34] (see Appendix C for a comparison of different optimizers). For a visual schematic of the calibration protocol, see Appendix D. The minimization of the cost function will converge as the estimated fluxes approach the measured fluxes.

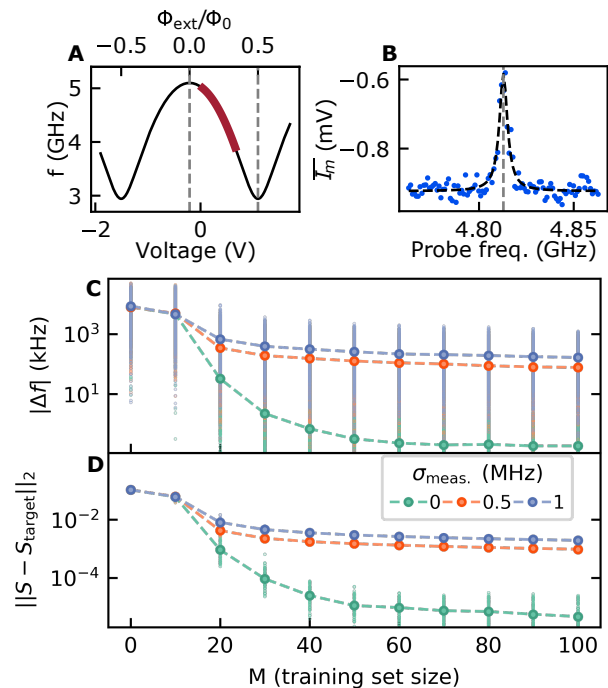


FIG. 2. **Learning protocol simulation** of measurement uncertainty dependence for a 16-qubit array. (A) Frequency spectrum of the transmon as a function of external flux applied. We bias the qubits to the maroon region of the spectrum for the flux crosstalk training. (B) We measure qubit frequencies experimentally via spectroscopy, by measuring the magnitude of the in-phase quadrature of the demodulated signal (\overline{I}_m), and extract the qubit frequency from a Lorentzian fit. (C) Simulated frequency error $|\Delta f|$ and the (D) Euclidean distance between the trained crosstalk matrix \mathbf{S} and the target matrix $\mathbf{S}_{\text{target}}$ when the matrix is learned in the presence of qubit frequency measurement uncertainty $\sigma_{\text{meas.}} \in \{0, 0.5, 1\}$ MHz. The distribution over 100 different random realizations is shown using small circles, and the median values are shown using large circles. Both quantities converge quickly in less than $M = 100$ training sets. The protocol accuracy decreases as the frequency measurement uncertainty increases.

B. Protocol simulation

We analyze the performance of our protocol for learning the crosstalk matrix on a simulated model of the 16-qubit array of transmon qubits (shown in Fig. 1B). Appendix E contains further details of the simulation model. In Fig. 2C, we report the error in achieving the desired target frequency $\Delta f = f_q - f_{\text{target}}$ when training the matrix using a training set of size M .

For the simulations, we modeled the flux crosstalk calibration protocol using realistic values for the transmon spectra parameters (\mathbf{V}^{Φ_0} , Φ_{offset} , f^{max} , E_C , and d). For the optimal crosstalk matrix $\mathbf{S}_{\text{target}}$, we used the previously-characterized crosstalk matrix for the 16-qubit array of transmons. In these simulations, each time a qubit frequency is “measured,” we add normally

distributed error with standard deviation $\sigma_{\text{meas.}}$, reflecting frequency-measurement imprecision (Fig. 2B) due to the qubit spectral linewidth arising from dephasing and power broadening, as well as frequency shifts arising from dispersive shifts and residual ZZ couplings [35]. We find that $\sigma_{\text{meas.}} \approx 0.5$ MHz most closely reflects our experimental conditions (see Appendix F for a discussion of frequency measurement error sources). Based on the trend in Fig. 2C, we expect that the plateau frequency error $|\Delta f|$ can be reduced by performing finer spectroscopy scans or measuring qubit frequencies via Ramsey interferometry, at the cost of longer overall calibration time. Another source of frequency measurement error arises due to small dispersive frequency shifts, despite the detuning scheme. We compensate for these detunings by using previously-characterized qubit-qubit couplings, but there could be some error in the calculation of the uncoupled qubit frequency.

We then create a training set of size M and train \mathbf{S} . Finally, we validate our learned crosstalk matrix \mathbf{S} with a random vector of frequencies \vec{f}_{target} independent of the training set. These random frequency vectors are chosen in the same way as the target frequency vectors used to learn the matrix, i.e. all qubits are biased to $100 \text{ MHz} - 1 \text{ GHz}$ off the sweet spot with all qubits sufficiently detuned. We attempt to set the qubits to these frequencies and record the error $|\vec{f}_q - \vec{f}_{\text{target}}|$, with no error added to the simulated measurement of \vec{f}_q .

We study the convergence of the protocol in the presence of frequency measurement errors. We observe that in the absence of measurement uncertainty ($\sigma_{\text{meas.}} = 0$ MHz), the protocol is capable of achieving a median frequency error $|\Delta f|$ on the order of 1 kHz with $M = 100$ training sets. In Fig. 2D we consider the Euclidean distance between the trained crosstalk matrix \mathbf{S} and the target crosstalk matrix $\mathbf{S}_{\text{target}}$: $\|\mathbf{S} - \mathbf{S}_{\text{target}}\|_2$. As the frequency measurement uncertainty increases, the performance of the protocol degrades, leading to a larger $|\Delta f|$ value when reaching the training plateau. We expect that with the anticipated uncertainty in measuring qubit frequencies in experiments, we should be able to achieve a median frequency error on the order of 100 kHz.

III. EXPERIMENTAL RESULTS

Next, we experimentally assess the performance of our protocol by calibrating the static flux-crosstalk matrix for a 16-qubit array of transmons. Starting from the assumption that $\mathbf{S} = \mathbf{I}$, we generate a set of 200 random voltage vectors, apply each voltage vector to the flux lines, and measure the corresponding qubit frequencies simultaneously via spectroscopy. Despite targeting > 200 MHz detuning between neighboring qubits, each qubit experiences a frequency shift due to its interaction with other qubits. We calculate the uncoupled qubit frequencies from the measured shifted frequencies using the pre-characterized qubit-qubit couplings. Using Eq. 1, we

convert these frequencies into a vector of fluxes experienced by each SQUID loop in the system corresponding to each applied voltage vector. We use different subsets of the measured set $\{\vec{V}_i, \vec{\Phi}_i^{\text{meas}}\}_{i=1:200}$ to learn the device crosstalk matrix.

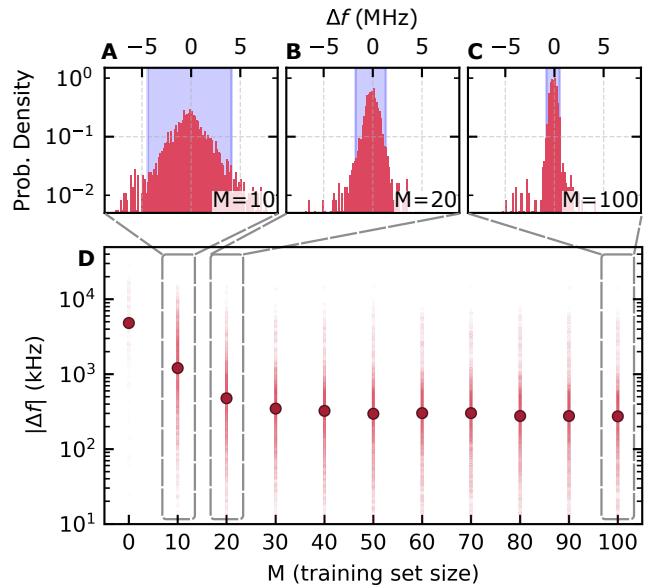


FIG. 3. **Experimental implementation** of the protocol using a 16-qubit array of transmon qubits. We see the distribution of offset in the measured qubit frequencies from the target values (Δf) for the crosstalk matrix trained with (A) $M = 10$, (B) $M = 20$, and (C) $M = 100$ number of training sets. The shaded region indicated the 5th to 95th percentile of the distribution. (D) Error in targeting the qubit frequency $|\Delta f|$ after training the crosstalk matrix with a different number of measurement sets. We observe that by increasing the training set size M , the median error (maroon points) decreases until reaching a plateau of approximately 290 kHz.

Prior to training, we selected a validation set of 10 target frequency vectors. This validation set was generated in the same way as target frequency vectors for training, with the added condition that each qubit is placed at least twice in the upper, middle, and lower regions of the frequency training region. For each value of M , we randomly selected 20 different subsets of $\{\vec{V}_i, \vec{\Phi}_i^{\text{meas}}\}$ with size M . With each subset, we learned the crosstalk matrix and then recorded the frequency error in setting the qubits to the validation frequencies.

In Figs. 3A, 3B, and 3C we show the qubit frequency deviation from the target validation set value (Δf) for training sets of size $M = 10$, 20, and 100, respectively. The shaded region in the figures indicates the 5th to 95th percentiles of the distribution. We observe that by using a larger training set, the distribution of Δf becomes narrower.

In our experiments, we notice irregularities in the transmon spectra of four of the qubits, presumably due to two-level-system (TLS) defects coupled to the qubits.

The frequency of a qubit coupled to coherent defects shifts, resulting in deviations from the transmon spectrum. In such frequency regions, we experience an error in setting the frequency of just a single qubit. Therefore, we do not include those qubits in the validation. The flux crosstalk for these four qubits can still be learned by excluding the regions of their transmon spectra impacted by TLSs. We can repeat the same learning-based protocol, biasing each of the four qubits of interest to a defect-free frequency, while applying quasi-random voltages to all other qubits and measuring the frequencies of the four qubits of interest. The resulting voltages and fluxes can be used to learn the corresponding four rows of \mathbf{S} . In Appendix E, we report the full learned \mathbf{S} for the 16-qubit array.

We demonstrate the experimental scaling of the frequency error $|\Delta f|$ as a function of M in Fig. 3D when learning the crosstalk matrix with our protocol. The frequency error generally decreases with the training set size and reaches a plateau at $M = 50$ with a median error of approximately 288 kHz.

IV. PROTOCOL SCALING

In order to demonstrate the extensibility of our approach, we simulate the learning procedure with an extension of the crosstalk model (see Appendix E) for larger transmon array sizes. We show the scaling of the frequency error for qubit arrays of size $N = 16, 64, 100$ qubits in Fig. 4A, assuming conservatively a frequency measurement uncertainty of 0.5 MHz. We observe a rapid convergence of the error in target frequency when using a training set with a size larger than the number of qubits.

In Fig. 4B, we study the error scaling by keeping the training set size constant (vertical slices of Fig. 4A). For training set sizes of $M = 200, 300, 400$, the frequency error empirically scales linearly with the number of qubits and remains below 200 kHz. The sources of error used in our model simulations are inaccuracies in characterizing the transmon spectrum parameters and uncertainty in qubit frequency measurements. Hence, we expect the frequency error to be lower in simulation compared to our experiments.

Based on the scaling observed in the simulation, in addition to the experimental demonstration, we conclude that using our learning-based crosstalk optimization protocol we would be able to accurately train the crosstalk matrix with $M \approx 2N$ spectroscopic frequency measurements per qubit. Since we can perform simultaneous frequency measurements, the total measurement time for the training set will be proportional to $2N$.

V. DISCUSSION

In this work, we describe an extensible approach for accurate flux crosstalk characterization and calibration

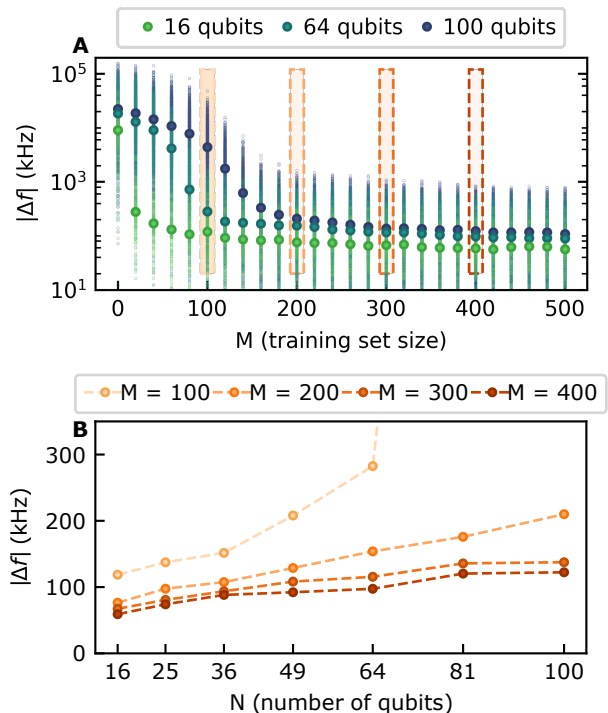


FIG. 4. **Simulation of extensibility (A)** Frequency error scaling for $N = 16, 64, 100$ qubits. We observe rapid convergence of frequency error after the training set size exceeds the number of qubits. The distribution over 10 different random realizations is shown using small circles, and the median values are shown using large circles. **(B)** Vertical slices, highlighted in **(A)** by shaded boxes, show the median frequency error as a function of the number of qubits for fixed training set size $M \in \{100, 200, 300, 400\}$. The simulation assumes each qubit frequency measurement includes a measurement uncertainty of 0.5 MHz.

based on machine learning. We experimentally verify the performance of our approach by employing the protocol to calibrate the static flux crosstalk of a 16-qubit flux-tunable transmon processor and observe convergence to a median frequency error less than 300 kHz with only $M = 50$ spectroscopy measurements of each qubit frequency, which can be done in parallel for all qubits. The protocol can also be used to calibrate the crosstalk matrix for fast flux pulses (see Appendix G).

We have demonstrated that our protocol enables us to realize a qubit frequency layout with accuracies better than < 300 kHz. In order to perform high-fidelity operations, we measure the frequency of each qubit using Ramsey interferometry after setting the qubit frequencies. In Appendix H, we discuss the impact of frequency errors on gate fidelities. The learned crosstalk matrix from our protocol also extrapolates to frequency layouts outside of the crosstalk training region defined in Fig. 2A (see Appendix I).

The accuracy of the protocol for each qubit relies on a precise characterization of the transmon spectrum fit

parameters. Furthermore, spectrum irregularities caused by TLS defects coupled to the qubit inhibit the convergence of the crosstalk optimization. To address the former, we perform simulations (see Appendix K) that show the method is robust against realistic errors in each fit parameter. To combat the impact of TLSs, we could use exclusion zones to avoid choosing frequencies in the spectrum that diverge from the transmon model.

We also study the performance of our protocol for different levels of crosstalk (Appendix N). We find that our method can effectively learn the device crosstalk when off-diagonal elements stay below roughly 10%. Flux crosstalk in current planar arrays of flux-tunable transmons is generally comfortably within this bound [4, 5, 29, 36, 37], and it is even lower in multilayer devices that can better suppress crosstalk (see Fig. 8A). This also suggests that routine recalibration of \mathbf{S} will be efficient, and we numerically find it should require fewer measurements

than the initial calibration (see Appendix L).

Advances in calibration efficiency and extensibility are essential as processor sizes increase. The learning-based flux crosstalk calibration procedure introduced in this work requires relatively few spectroscopic measurements, is robust to measurement error, and scales favorably as array size increases. Alternatively, one can use Ramsey measurements to determine the qubit frequencies, which is faster than spectroscopy for single-qubit frequency measurement. Future implementations could use simultaneous spectroscopy measurements to initially learn the matrix, and then use Ramsey measurements to fine-tune the calibration if necessary. This approach reduces the characterization time and improves the accuracy of the flux crosstalk matrix characterization—and ultimately the performance of algorithms and simulations run on superconducting qubit processors.

-
- [1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, Quantum supremacy using a programmable superconducting processor, *Nature* **574**, 505 (2019).
- [2] R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Anshmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush, D. Bacon, J. C. Bardin, J. Basso, A. Bengtsson, S. Boixo, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton, B. B. Buckley, D. A. Buell, T. Burger, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, J. Cogan, R. Collins, P. Conner, W. Courtney, A. L. Crook, B. Curtin, D. M. Debroy, A. D. T. Barba, S. Demura, A. Dunsworth, D. Eppens, C. Erickson, L. Faoro, E. Farhi, R. Fatemi, L. F. Burgos, E. Forati, A. G. Fowler, B. Foxen, W. Giang, C. Gidney, D. Gilboa, M. Giustina, A. G. Dau, J. A. Gross, S. Habegger, M. C. Hamilton, M. P. Harrigan, S. D. Harrington, O. Higgott, J. Hilton, M. Hoffmann, S. Hong, T. Huang, A. Huff, W. J. Huggins, L. B. Ioffe, S. V. Isakov, J. Iveland, E. Jeffrey, Z. Jiang, C. Jones, P. Juhas, D. Kafri, K. Kechedzhi, J. Kelly, T. Khattar, M. Khezri, M. Kieferová, S. Kim, A. Kitaev, P. V. Klimov, A. R. Klots, A. N. Korotkov, F. Kostritsa, J. M. Kreikebaum, D. Landhuis, P. Laptev, K.-M. Lau, L. Laws, J. Lee, K. Lee, B. J. Lester, A. Lill, W. Liu, A. Locharla, E. Lucero, F. D. Malone, J. Marshall, O. Martin, J. R. McClean, T. Mccourt, M. McEwen, A. Megrant, B. M. Costa, X. Mi, K. C. Miao, M. Mohseni, S. Montazeri, A. Morvan, E. Mount, W. Mruczkiewicz, O. Naaman, M. Neeley, C. Neill, A. Nersisyan, H. Neven, M. Newman, J. H. Ng, A. Nguyen, M. Nguyen, M. Y. Niu, T. E. O'Brien, A. Opremcak, J. Platt, A. Petukhov, R. Potter, L. P. Pryadko, C. Quintana, P. Roushan, N. C. Rubin, N. Saei, D. Sank, K. Sankaragomathi, K. J. Satzinger, H. F. Schurkus, C. Schuster, M. J. Shearn, A. Shorter, V. Shvarts, J. Skrzuzny, V. Smelyanskiy, W. C. Smith, G. Sterling, D. Strain, M. Szalay, A. Torres, G. Vidal, B. Villalonga, C. V. Heidweiller, T. White, C. Xing, Z. J. Yao, P. Yeh, J. Yoo, G. Young, A. Zalcman, Y. Zhang, and N. Zhu, Suppressing quantum errors by scaling a surface code logical qubit, *Nature* **614**, 676 (2023).
- [3] P. Zhang, H. Dong, Y. Gao, L. Zhao, J. Hao, J.-Y. Desautels, Q. Guo, J. Chen, J. Deng, B. Liu, W. Ren, Y. Yao, X. Zhang, S. Xu, K. Wang, F. Jin, X. Zhu, B. Zhang, H. Li, C. Song, Z. Wang, F. Liu, Z. Papić, L. Ying, H. Wang, and Y.-C. Lai, Many-body Hilbert space scarring on a superconducting processor, *Nature Physics* **19**, 120 (2023).
- [4] A. H. Karamlou, J. Braumüller, Y. Yanay, A. Di Paolo, P. M. Harrington, B. Kannan, D. Kim, M. Kjaergaard, A. Melville, S. Muschinske, B. M. Niedzielski, A. Vepsäläinen, R. Winik, J. L. Yoder, M. Schwartz, C. Tahan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Quantum transport and localization in 1d and 2d tight-binding lattices, *npj Quantum Information* **8**, 1 (2022).
- [5] J. Braumüller, A. H. Karamlou, Y. Yanay, B. Kannan, D. Kim, M. Kjaergaard, A. Melville, B. M. Niedzielski, Y. Sung, A. Vepsäläinen, R. Winik, J. L. Yoder, T. P. Orlando, S. Gustavsson, C. Tahan, and W. D. Oliver, Probing quantum information propagation with out-of-time-ordered correlators, *Nature Physics* **18**, 172 (2022).
- [6] F. Yan, P. Krantz, Y. Sung, M. Kjaergaard, D. L. Campbell, T. P. Orlando, S. Gustavsson, and W. D. Oliver,

- Tunable coupling scheme for implementing high-fidelity two-qubit gates, *Phys. Rev. Appl.* **10**, 054062 (2018).
- [7] Y. Sung, L. Ding, J. Braumüller, A. Vepsäläinen, B. Kannan, M. Kjaergaard, A. Greene, G. O. Samach, C. McNally, D. Kim, A. Melville, B. M. Niedzielski, M. E. Schwartz, J. L. Yoder, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Realization of High-Fidelity CZ and ZZ-Free iSWAP Gates with a Tunable Coupler, *Physical Review X* **11**, 021058 (2021).
- [8] X. Dai, D. Tennant, R. Trappen, A. Martinez, D. Melanson, M. Yurtalan, Y. Tang, S. Novikov, J. Grover, S. Disseler, J. Basham, R. Das, D. Kim, A. Melville, B. Niedzielski, S. Weber, J. Yoder, D. Lidar, and A. Lupascu, Calibration of Flux Crosstalk in Large-Scale Flux-Tunable Superconducting Quantum Circuits, *PRX Quantum* **2**, 040313 (2021).
- [9] X. Dai, R. Trappen, R. Yang, S. M. Disseler, J. I. Basham, J. Gibson, A. J. Melville, B. M. Niedzielski, R. Das, D. K. Kim, J. L. Yoder, S. J. Weber, C. F. Hirjibehedin, D. A. Lidar, and A. Lupascu, Optimizing for periodicity: a model-independent approach to flux crosstalk calibration for superconducting circuits, [arXiv:2211.01497 \[quant-ph\]](https://arxiv.org/abs/2211.01497) (2022).
- [10] J. Darulová, S. Pauka, N. Wiebe, K. Chan, G. Gardener, M. Manfra, M. Cassidy, and M. Troyer, Autonomous tuning and charge-state detection of gate-defined quantum dots, *Phys. Rev. Appl.* **13**, 054005 (2020).
- [11] N. M. van Esbroeck, D. T. Lennon, H. Moon, V. Nguyen, F. Vigneau, L. C. Camenzind, L. Yu, D. M. Zumbühl, G. A. D. Briggs, D. Sejdinovic, and N. Ares, Quantum device fine-tuning using unsupervised embedding learning, *New Journal of Physics* **22**, 095003 (2020).
- [12] D. T. Lennon, H. Moon, L. C. Camenzind, L. Yu, D. M. Zumbühl, G. A. D. Briggs, M. A. Osborne, E. A. Laird, and N. Ares, Efficiently measuring a quantum device using machine learning, *npj Quantum Information* **5**, 1 (2019).
- [13] H. Moon, D. T. Lennon, J. Kirkpatrick, N. M. van Esbroeck, L. C. Camenzind, L. Yu, F. Vigneau, D. M. Zumbühl, G. A. D. Briggs, M. A. Osborne, D. Sejdinovic, E. A. Laird, and N. Ares, Machine learning enables completely automatic tuning of a quantum device faster than human experts, *Nature Communications* **11**, 4161 (2020).
- [14] B. Severin, D. T. Lennon, L. C. Camenzind, F. Vigneau, F. Fedele, D. Jirovec, A. Ballabio, D. Chrastina, G. Isella, M. de Kruijf, M. J. Carballido, S. Svab, A. V. Kuhlmann, F. R. Braakman, S. Geyer, F. N. M. Froning, H. Moon, M. A. Osborne, D. Sejdinovic, G. Katsaros, D. M. Zumbühl, G. A. D. Briggs, and N. Ares, Cross-architecture tuning of silicon and sige-based quantum devices using machine learning, [arXiv:2107.12975 \[cond-mat, physics:quant-ph\]](https://arxiv.org/abs/2107.12975) (2021).
- [15] J. Ziegler, T. McJunkin, E. Joseph, S. S. Kalantre, B. Harpt, D. Savage, M. Lagally, M. Eriksson, J. M. Taylor, and J. P. Zvolak, Toward robust autotuning of noisy quantum dot devices, *Phys. Rev. Appl.* **17**, 024069 (2022).
- [16] J. Ziegler, F. Luthi, M. Ramsey, F. Borjans, G. Zheng, and J. P. Zvolak, Automated extraction of capacitive coupling for quantum dot systems, *Physical Review Applied* **19**, 054077 (2023).
- [17] J. P. Zvolak and J. M. Taylor, Colloquium: Advances in automation of quantum dot devices control, *Reviews of modern physics* **95**, 011006 (2023).
- [18] J. P. Zvolak, T. McJunkin, S. S. Kalantre, J. Dodson, E. MacQuarrie, D. Savage, M. Lagally, S. Coppersmith, M. A. Eriksson, and J. M. Taylor, Autotuning of double-dot devices in situ with machine learning, *Phys. Rev. Appl.* **13**, 034075 (2020).
- [19] J. Kelly, P. O'Malley, M. Neeley, H. Neven, and J. M. Martinis, Physical qubit calibration on a directed acyclic graph, [arXiv:1803.03226 \[quant-ph\]](https://arxiv.org/abs/1803.03226) (2018).
- [20] P. V. Klimov, J. Kelly, J. M. Martinis, and H. Neven, The Snake Optimizer for Learning Quantum Processor Control Parameters, [arXiv:2006.04594 \[quant-ph\]](https://arxiv.org/abs/2006.04594) (2020).
- [21] E. Magesan, J. M. Gambetta, A. D. Córcoles, and J. M. Chow, Machine learning for discriminating quantum measurement trajectories and improving readout, *Phys. Rev. Lett.* **114**, 200501 (2015).
- [22] L. A. Martinez, Y. J. Rosen, and J. L. DuBois, Improving qubit readout with hidden markov models, *Phys. Rev. A* **102**, 062426 (2020).
- [23] R. Navarathna, T. Jones, T. Moghaddam, A. Kulikov, R. Beriwal, M. Jerger, P. Pakkiam, and A. Fedorov, Neural networks for on-the-fly single-shot state classification, *Applied Physics Letters* **119**, 114003 (2021).
- [24] P. Duan, Z.-F. Chen, Q. Zhou, W.-C. Kong, H.-F. Zhang, and G.-P. Guo, Mitigating crosstalk-induced qubit readout error with shallow-neural-network discrimination, *Phys. Rev. Appl.* **16**, 024063 (2021).
- [25] B. Lienhard, A. Vepsäläinen, L. C. Govia, C. R. Hoffer, J. Y. Qiu, D. Ristè, M. Ware, D. Kim, R. Winik, A. Melville, B. Niedzielski, J. Yoder, G. J. Ribeill, T. A. Ohki, H. K. Krovi, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Deep-neural-network discrimination of multiplexed superconducting-qubit states, *Phys. Rev. Appl.* **17**, 014024 (2022).
- [26] Y. Baum, M. Amico, S. Howell, M. Hush, M. Liuzzi, P. Mundada, T. Merkh, A. R. Carvalho, and M. J. Biercuk, Experimental deep reinforcement learning for error-robust gate-set design on a superconducting quantum computer, *PRX Quantum* **2**, 040324 (2021).
- [27] N. Wittler, F. Roy, K. Pack, M. Werninghaus, A. S. Roy, D. J. Egger, S. Filipp, F. K. Wilhelm, and S. Machnes, Integrated tool set for control, calibration, and characterization of quantum devices applied to superconducting qubits, *Physical Review Applied* **15**, 034080 (2021).
- [28] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Charge-insensitive qubit design derived from the Cooper pair box, *Physical Review A* **76**, 042319 (2007).
- [29] D. M. Abrams, N. Didier, S. A. Caldwell, B. R. Johnson, and C. A. Ryan, Methods for Measuring Magnetic Flux Crosstalk between Tunable Transmons, *Physical Review Applied* **12**, 064022 (2019).
- [30] S. Krinner, N. Lacroix, A. Remm, A. Di Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann, G. J. Norris, C. K. Andersen, M. Müller, A. Blais, C. Eichler, and A. Wallraff, Realizing repeated quantum error correction in a distance-three surface code, *Nature* **605**, 669 (2022).
- [31] D. Rosenberg, D. Kim, R. Das, D. Yost, S. Gustavsson, D. Hover, P. Krantz, A. Melville, L. Racz, G. O. Samach, S. J. Weber, F. Yan, J. L. Yoder, A. J. Kerman, and W. D. Oliver, 3D integrated superconducting qubits, *npj Quantum Information* **3**, 1 (2017).
- [32] A. H. Karamlou, I. T. Rosen, S. E. Muschinske, C. N.

- Barrett, A. Di Paolo, L. Ding, P. M. Harrington, M. Hays, R. Das, D. K. Kim, B. M. Niedzielski, M. Schuld, K. Serniak, M. E. Schwartz, J. L. Yoder, S. Gustavsson, Y. Yanay, J. A. Grover, and W. D. Oliver, [Probing entanglement across the energy spectrum of a hard-core Bose-Hubbard lattice](#) (2023), [arxiv:2306.02571 \[quant-ph\]](#).
- [33] D. C. Liu and J. Nocedal, On the limited memory bfgs method for large scale optimization, [Mathematical Programming](#) **45**, 503–528 (1989).
- [34] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in [Advances in Neural Information Processing Systems 32](#) (Curran Associates, Inc., 2019) pp. 8024–8035.
- [35] D. I. Schuster, A. Wallraff, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. M. Girvin, and R. J. Schoelkopf, ac stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field, [Physical Review Letters](#) **94**, 123602 (2005).
- [36] Z. Yan, Y.-R. Zhang, M. Gong, Y. Wu, Y. Zheng, S. Li, C. Wang, F. Liang, J. Lin, Y. Xu, C. Guo, L. Sun, C.-Z. Peng, K. Xia, H. Deng, H. Rong, J. Q. You, F. Nori, H. Fan, X. Zhu, and J.-W. Pan, Strongly correlated quantum walks with a 12-qubit superconducting processor, [Science](#) **364**, 753 (2019).
- [37] X. Zhang, E. Kim, D. K. Mark, S. Choi, and O. Painter, A superconducting quantum simulator based on a photonic-bandgap metamaterial, [Science](#) **379**, 278 (2023).
- [38] M. Malekakhlagh, E. Magesan, and D. C. McKay, First-principles analysis of cross-resonance gate operation, [Physical Review A](#) **102**, 042605 (2020).
- [39] A. Patterson, J. Rahamim, T. Tsunoda, P. Spring, S. Jebari, K. Ratter, M. Mergenthaler, G. Tancredi, B. Vlastakis, M. Esposito, and P. Leek, Calibration of a Cross-Resonance Two-Qubit Gate Between Directly Coupled Transmons, [Physical Review Applied](#) **12**, 064013 (2019).

ACKNOWLEDGMENTS

The authors are grateful to Patrick M. Harrington and Francisca Vasconcelos for fruitful discussions. This work is supported in part by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum System Accelerator (QSA); in part by the Defense Advanced Research Projects Agency under the Quantum Benchmarking contract; in part by U.S. Army Research Office Grant W911NF-18-1-0411; and in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), the Department of Energy, and the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. CNB acknowledges support from the STC Center for Integrated Quantum Materials,

NSF Grant No. DMR-1231319 and the Wellesley College Samuel and Hilda Levitt Fellowship. AHK acknowledges support from the NSF Graduate Research Fellowship Program. SEM is supported by a NASA Space Technology Research Fellowship. ITR is supported by an appointment to the Intelligence Community Postdoctoral Research Fellowship Program at the Massachusetts Institute of Technology administered by Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy and the Office of the Director of National Intelligence (ODNI). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the DOE, DARPA, ODNI, IARPA, or USDR&E.

Appendix A: 3D-Integrated, Flip-Chip Device

Our experimental sample is an array of 16 flux-tunable transmon qubits fabricated using a flip-chip process [31]. Unlike a planar architecture, where all chip elements are mounted on the same surface, the flip-chip has two separate tiers which are stacked on top of each other. The qubit tier houses the qubits and the interposer tier houses all other chip elements. The benefits of the flip-chip design include decreased distances between control lines and the qubits they target and increased shielding between neighboring qubits, which significantly reduces overall crosstalk levels for DC flux control and fast flux pulses. For further details about the device see [32].

Qubit parameter	Measured Value
f^{\max}	4.887 ± 0.110 (GHz)
V^{Φ_0}	29.2 ± 2.7 (V)
d	0.35 ± 0.04
E_C/h	196.1 ± 5.2 (MHz)
$ \Phi_{\text{offset}} $	19.7 ± 5.9 ($m\Phi_0$)

TABLE I. **Qubit parameters** for the 3D-integrated flip-chip device. Mean values plus or minus one standard deviation are reported. Deviations in parameter values are due to unintentional fabrication imperfections.

The nearest neighbor coupling on this device is fixed at $J/2\pi = (5.89 \pm 0.4)$ MHz, measured at qubit frequencies of 4.5 GHz. Before performing the learning-based protocol, we characterize the qubit-qubit couplings should be characterized. With the detuning scheme we use for training (> 200 MHz detuning between nearest neighbor qubits and > 50 MHz detuning between any two qubits), we have a mean dispersive shift of around 180 kHz, which we correct for (see Appendix F). This shift is less than

the final frequency setting precision using the optimized crosstalk matrix, which is ≈ 300 kHz. In our simulations, we incorporate frequency measurement error and observe frequency error convergence for $\sigma_{\text{meas.}} = 0.5$ MHz. We, therefore, expect that this relatively small dispersive shift does not greatly impact the final frequency setting precision of our protocol.

Appendix B: Comparison to Direct Measurement Approach

One alternative to the learning-based approach to crosstalk calibration described in this paper is to directly measure each element of \mathbf{S} . A downside to this approach is that the number of elements of the crosstalk matrix scales quadratically with the number of qubits in the array.

We note that in our experimental setup, we can simultaneously measure all 16 qubits' frequencies, and we are able to leverage this simultaneous frequency measurement in the learning-based protocol. We cannot perform simultaneous frequency measurement in the direct measurement approach, because we cannot bias more than one qubit off of the sweet spot without introducing flux crosstalk effects. We might wish to sweep the voltage applied to one flux line and measure the response of all other qubits. Unfortunately, this approach to direct measurement will not work due to the broadness of the transmon spectrum (since $1/V_{i,j}^{\Phi_0} \ll 1/V_{i,i}^{\Phi_0}$ for $i \neq j$) and the relatively narrow tuning range for our voltages applied to flux lines. We will be unable to accurately fit the transmon spectrum and extract $V_{i,j}^{\Phi_0}$. To resolve elements of \mathbf{S} , we must first bias a qubit to a steep (flux-sensitive) section of the transmon spectrum and then sweep the voltage applied to another qubit's flux line. Therefore, we cannot perform a simultaneous frequency measurement to obtain all the crosstalk elements due to one flux line. There are creative methods by which we can obtain more than one crosstalk element via simultaneous measurements, but it will be impossible to obtain a full column of the crosstalk matrix at once in this fashion.

We simulate the direct measurement approach for a 16-qubit array while varying the frequency measurement uncertainty (as in Fig 2). For the direct measurement approach, we bias qubit i away from its sweet spot by applying a voltage V_i to flux line i . Then, we sweep the voltage of flux line j across its full tuning range. For our voltage source and qubit V^{Φ_0} 's, this is approximately ± 0.3 times the average V^{Φ_0} of a qubit. We fit the transmon spectrum, holding f^{max} , E_C , and d fixed. The new flux offset will be approximately:

$$\Phi_{\text{offset}} = \frac{V_i}{V_{i,i}^{\Phi_0}} + \Phi_{\text{offset},i} \quad (\text{B1})$$

Using this initial guess for the flux offset and using the slope of the measured frequencies to determine the sign

of $V_{i,j}^{\Phi_0}$ (to constrain it), we fit the curve to find $V_{i,j}^{\Phi_0}$ (we also obtain Φ_{offset} from the fit, but this information is irrelevant). Finally, the measured element of the crosstalk matrix is:

$$\mathbf{S}_{i,j} = \frac{V_{i,i}^{\Phi_0}}{V_{i,j}^{\Phi_0}} \quad (\text{B2})$$

We see that for a reasonably assumed measurement uncertainty of 0.5 MHz, the direct measurement approach reaches the same level of precision as the learning-based approach with a size $M = 30$ training set when the number of data points reaches 10 (Fig. 5A). This tells us that the direct measurement approach requires 5x as many measurements to reach the same level of precision as the learning-based approach.

In the learning-based approach, each qubit's frequency is measured 30 times. In the direct measurement approach, each qubit's frequency is measured 10 times for each flux line, for a total of $15 \cdot 10 = 150$ times. So for 16 qubits, we already achieve a $\approx 5\times$ speedup by using the learning-based protocol.

Appendix C: Gradient Descent Optimizers

One critical piece of the protocol is minimizing the mean squared error cost function in a gradient descent optimizer. Throughout this work, we use the L-BFGS optimizer in PyTorch, with a learning rate of 1.0. In order to ensure proper convergence of our gradient descent minimization, we compared the L-BFGS results to a couple of other PyTorch optimizers: SGD and Adam (see Fig. 6). On a practical level, it doesn't matter which optimizer we use, since this is not a bottleneck in our calibration protocol. However, we find that L-BFGS converges with fewer optimizer iterations in comparison to other optimizers. Additionally, other optimizers have additional parameters that need to be set. For example, SGD has a momentum parameter that may need to be changed for optimal performance.

We find that L-BFGS consistently converges in the least iterations compared to other optimizers, and additionally requires no tuning of parameters, making it a good choice for practical use in the laboratory.

Appendix D: Schematic of Learning-based Protocol

The goal of flux crosstalk calibration is to control qubit frequencies precisely. The learning-based approach described in this paper is an intuitive approach to this problem since it actively addresses the goal of the calibration via the calibration process. The intuitiveness of the learning-based approach stands in contrast to a direct measurement approach, which seeks to resolve individual elements of the crosstalk matrix \mathbf{S} , but does not directly

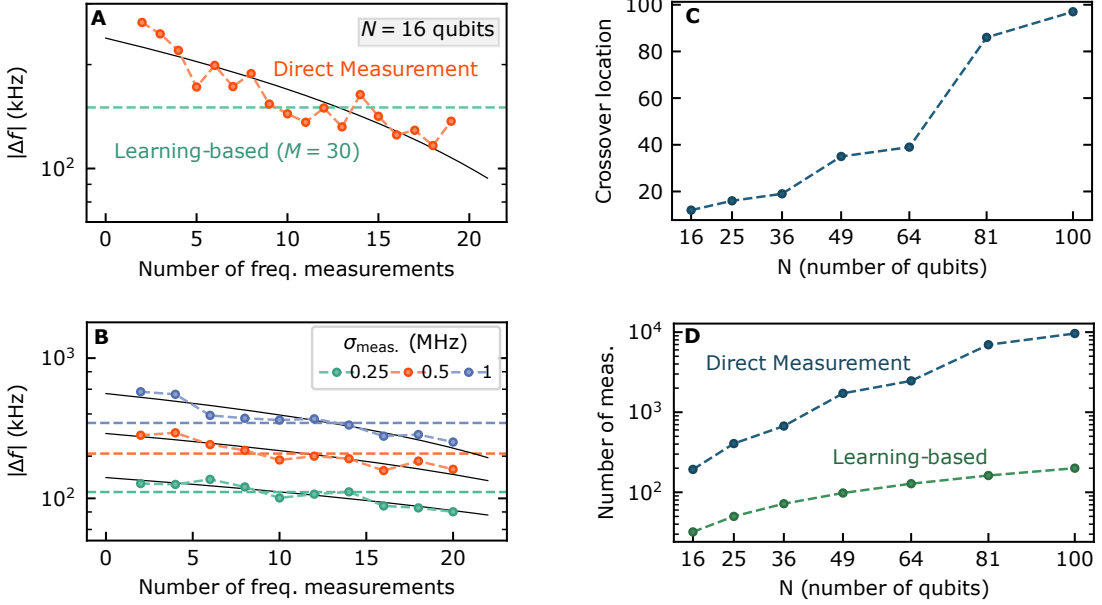


FIG. 5. **Scaling of direct measurement approach** Simulation of direct measurement convergence for a frequency measurement uncertainty of 0.5 MHz. We sweep the number of frequency measurements between $V_j = -0.3V^{\Phi_0}$ and $V_j = +0.3V^{\Phi_0}$. **(A)** For $N = 16$ qubits, the performance of the direct measurement approach exceeds the performance of the learning-based approach with $M = 30 \approx 2 \times N$ training sets when the number of frequency measurements is ≈ 10 . The median frequency errors for the direct measurement approach are fitted linearly with a black line, which crosses over the learning-based median frequency error at ≈ 12 data points. The medians are computed over 10 simulation repetitions. In **(B)**, we examine the performance of the direct measurement approach for $N = 16$ qubits under varying levels of frequency measurement uncertainty $\sigma_{\text{meas.}}$. We find that the crossover location (again, for the median frequency error using the learning-based approach with $M = 30 \approx 2 \times N$ training sets) increases as $\sigma_{\text{meas.}}$ increases. The medians are computed over 20 simulation repetitions. In **(C)**, we examine the scaling of this crossover location for up to $N = 100$ qubits, assuming $\sigma_{\text{meas.}} = 0.5$ MHz. The number of frequency measurements required to match the learning-based protocol's performance grows empirically linearly with the number of qubits. **(D)** The number of frequency measurements required for each qubit, given by the crossover location times $N - 1$, diverges from the learning-based $M = 2 \times N$ measurements per qubit.

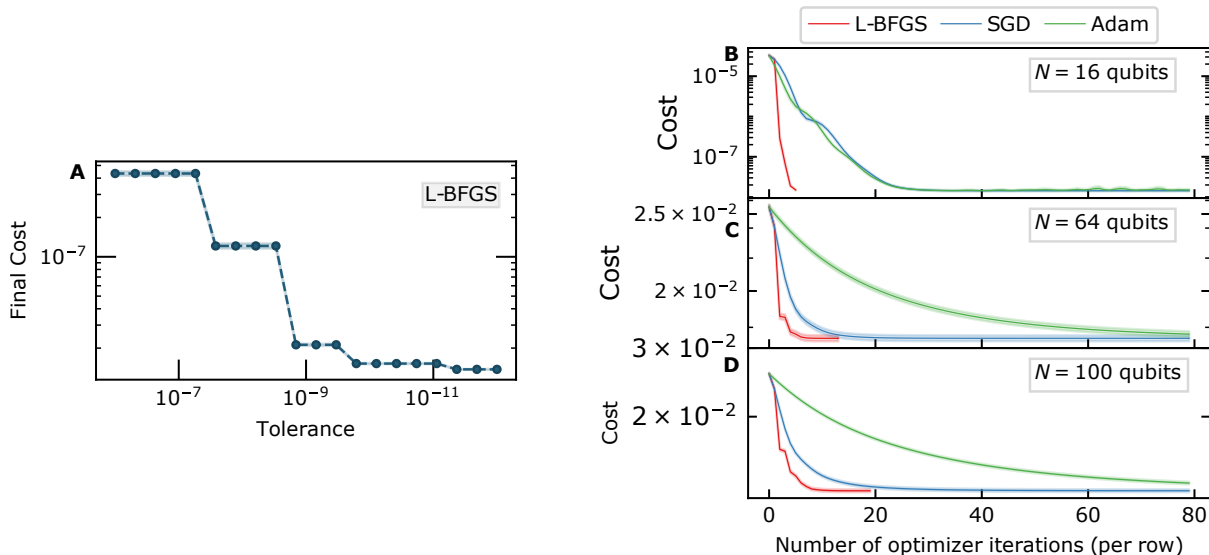


FIG. 6. **Comparison of gradient descent optimizers** We minimize a mean squared error cost function to optimize \mathbf{S} row-by-row. In (A), we consider the final minimized cost versus the function change tolerance of the L-BFGS optimizer for $N = 16$ qubits. The mean costs were computed over 100 simulation repetitions, and the x-axis is inverted for clarity. The shaded region is the 95% confidence interval for the mean. In (B,C,D), we compare the minimization of the cost function of the L-BFGS, SGD, and Adam optimizers for $N \in \{16, 64, 100\}$ qubits. L-BFGS terminates on tolerance, whereas SGD and Adam terminate after a given number of iterations. The L-BFGS tolerance is set to 10^{-10} with a learning rate of 1.0. The SGD learning rate is 1.0 with a momentum of 0.7. The Adam learning rate is 0.002 with betas of (0.7, 0.999). The SGD and Adam parameters were selected for the best performance for $N = 16$ qubits. The mean costs were computed over 10 simulation repetitions, and the shaded region is the 95% confidence interval for the mean.

evaluate the ability of the calibrated \mathbf{S} to set qubit frequencies precisely. For a quantitative comparison of these two approaches, see Appendix B.

The general concept behind the learning-based approach to flux crosstalk calibration is to use an initial estimate for \mathbf{S} to target frequencies \vec{f}_{target} and then measure the frequencies of the qubits. The difference between the measured and targeted frequencies gives us insight into how to adjust \mathbf{S} to minimize frequency error. Specifically, we utilize the linear flux crosstalk relation (Eq. 2) and the transmon spectrum (Eq. 1) to obtain a cost function with a well-defined gradient which can be minimized in a gradient descent optimizer. In Fig. 7, we provide a flowchart visualization of the calibration process.

We note that the cost function we are minimizing is convex due to the constrained frequency range for target frequencies. Although the transmon spectrum is periodic, we restrict our target frequencies to one portion of the spectrum, allowing for a 100 MHz buffer from the upper sweet spot (and an even larger buffer from the lower sweet spot). With this restriction, assuming reasonable crosstalk levels of $< 10\%$, we eliminate the possibility of converting the measured frequency to an incorrect experienced flux.

For the purposes of this paper, which are to introduce and demonstrate the efficacy and extensibility of the

learning-based approach, we did not optimize end-to-end calibration time. One way the efficiency of this calibration protocol could be improved is to learn the crosstalk matrix iteratively. We could easily modify this protocol to learn \mathbf{S} in batches. Instead of applying all voltages and measuring all frequencies, we can instead perform a size m training set with $m < M$, and then use the resultant trained \mathbf{S} as our new estimate for \mathbf{S} for the next training set. The advantage to this batched approach is that each iterative \mathbf{S} will become closer and closer to the optimal \mathbf{S} . Therefore, the actual qubit frequencies will be closer to the target frequencies in the training sets, which means we need to scan over a smaller range of frequencies to locate our qubits. Since frequency measurements are the most time-consuming part of this protocol, minimizing the time per measurement will speed up the protocol.

Appendix E: Crosstalk model

For all simulations of the calibration process for a 16-qubit array, we used the measured flux crosstalk matrix for the 16-qubit device discussed in this paper. The crosstalk matrix for this device is shown in Fig. 8A. The measured crosstalk matrix we used was from a previous cooldown, but the overall crosstalk matrix changes rela-

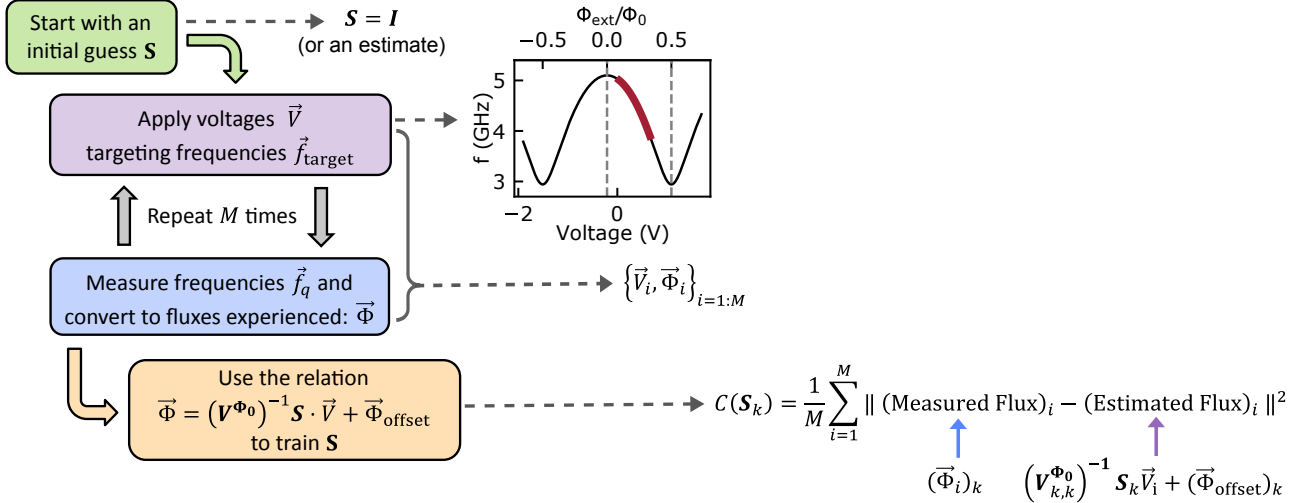


FIG. 7. **Flowchart for learning-based protocol** We begin with an initial S , which could be the identity or an estimate. We select target frequencies from a range of ≈ 100 MHz to ≈ 1 GHz off the sweet spot and apply voltages \vec{V} targeting these frequencies. We measure the qubits' frequencies and convert them to fluxes experienced. After repeating this process M times, we have a size- M training set of pairs of applied voltages and experienced fluxes. We use the linear flux crosstalk relation to train each row of S (S_k) by minimizing the mean-squared-error cost function $C(S_k)$ of the measured flux and the estimated flux in a gradient descent optimizer. The resultant S minimizes the difference between measured and experienced fluxes.

tively little from one cooldown to the next. For the simulation of protocol scaling, we needed a model for crosstalk to generate realistic crosstalk matrices for larger array sizes.

One possible way to model crosstalk is to consider the strength of crosstalk versus the Euclidean distance between the site of the flux line and the influenced qubit. For the 16-qubit device, the magnitude of crosstalk vs Euclidean distance is shown in Fig. 8B.

One compelling reason to model crosstalk in this way is that once current flows through a local flux line, it must flow through the superconducting ground plane to ground. The paths that these currents take through the chip are unknown and cannot be modeled. We do know, however, that a current dispensed at the end of a flux line must flow to the edge of the chip. If we consider a circle of radius r centered on the end of the flux line, we can assume there is an equal probability of the current flowing through any point on this circle (namely $s/2\pi r$ probability, for an arc of length s). We can therefore assume that the probability of a current flowing past a point a distance r from the end of a flux line is proportional to $1/r$.

We fit the curve in Fig 8B with a shifted $1/r$ decay: $\frac{1}{ax+1} + c$. This equation is a $1/r$ decay that passes through $(0, 100\%)$ when $c = 0$. We obtain fit values of $a = 178.2$ and $c = 0.264$. The model qualitatively fits the data, with the nearest neighbor crosstalk slightly elevated compared to longer-range crosstalk.

To generate crosstalk matrices for larger arrays, we used this exponential model. Since we do not have experimental measurements for Euclidean distances beyond

the furthest distance on a 16-qubit array, we used the standard deviation of the magnitude of all off-diagonal crosstalk elements for the 16-qubit array: $\sigma = 0.342$. For a given crosstalk element $S_{i,j}$ with $i \neq j$, we compute the crosstalk level ℓ corresponding to the Euclidean distance between qubit i and qubit j and then pick the magnitude of $S_{i,j}$ from a normal distribution $\mathcal{N}(\ell, \sigma)$. Then, we randomly multiply the element by ± 1 .

Appendix F: Frequency Measurement Error Sources

One of the limiting factors in the frequency precision of our protocol is the measurement error inherent in the training sets. In each training set measurement, we measure the frequency of each qubit and convert that to a flux experienced. The accuracy of this frequency measurement therefore directly influences the frequency error plateau for the protocol.

We have identified three primary sources of error in our frequency measurements: spectroscopic measurement inaccuracies, uncompensated dispersive shifts, and residual ZZ couplings. We report the estimated error from each of these sources in Table II. In this section, we discuss these estimates and their implications.

We measure our qubit frequencies via spectroscopy. The spectroscopy scans have a frequency step of 0.5 MHz to rapidly scan $f_{\text{target}} \pm 30$ MHz for each qubit. We fit the peak to a Lorentzian lineshape. To obtain an estimate for the error of the fitted spectroscopy peak, we consider the distribution of fit uncertainties for the size $M = 200$ initial training set used in our experiment. After remov-

ing outliers where no peak was found, the mean error in the fitted peak is 118 kHz. We report the estimated error from spectroscopic inaccuracies as 120 kHz.

The qubits experience dispersive shifts due to interactions with nearest neighbor qubits via direct capacitive coupling and with other qubits in the array via stray capacitances. We characterize the qubit-qubit couplings before calibrating flux crosstalk, and we use these coupling strengths and the measured frequency of each qubit to determine the dispersive shift experienced by each qubit and compensate for it. From the initial size $M = 200$ training set in our experiment, the mean value of the calculated dispersive shift is approximately 180 kHz.

Given a vector of measured frequencies $\vec{f}_{\text{meas.}}$ and the qubit-qubit coupling matrix \mathbf{J} , we determine the uncoupled qubit frequencies \vec{f} in the following way. The elements of \vec{f} are the frequencies of the qubits in the case

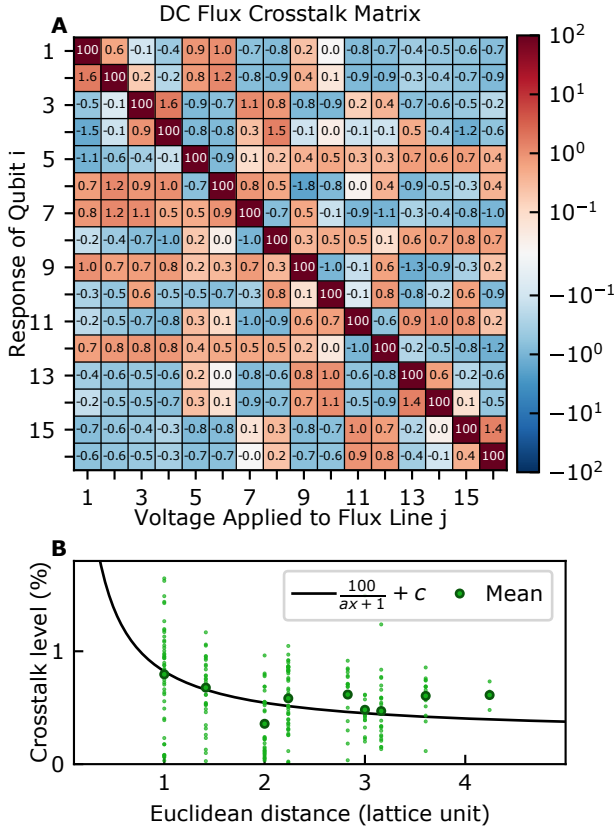


FIG. 8. **The DC crosstalk matrix (A)** The measured flux crosstalk sensitivity matrix for the 16-qubit flip-chip device in our lab, rescaled by a factor of 100 such that each element is a percentage. The average off-diagonal crosstalk level is $< 1\%$. We use the distribution of crosstalk sensitivity $S_{i,j}$ versus Euclidean distance between qubit i and qubit j to generate realistic crosstalk matrices for simulations of protocol scaling. **(B)** The off-diagonal crosstalk versus Euclidean distance between flux line and qubit can be fitted with a $1/r$ decay defined by $\ell = \frac{100}{ax+1} + c$, with $a = 178.2$ and $c = 0.264$.

Error Source	Estimated Error
Spectroscopic Inaccuracy	120 kHz
Uncompensated Dispersive Shift	120 kHz
Residual ZZ Coupling	230 kHz
Total Estimate	280 kHz

TABLE II. **Error budget** for frequency measurement based on the known primary sources of uncertainty. Estimates for average error are reported.

of no qubit-qubit coupling: $\mathbf{J} = 0$. Therefore, we diagonalize the matrix $\vec{f} + \mathbf{J}$, where \vec{f} is a diagonal matrix whose entries are f . The entries of this diagonalized matrix, which are the eigenvalues of $\vec{f} + \mathbf{J}$, will be $\vec{f}_{\text{meas.}}$ (although perhaps in a different order).

To find \vec{f} , we first compute the eigenvalues $\vec{\lambda}$ of the matrix $\vec{f} + \mathbf{J}$ and sort them in ascending order. We determine \vec{f} by using a Nelder-Mead optimizer to minimize the cost function:

$$C(\vec{f}) = \left\| \vec{\lambda}(\vec{f}) - \vec{f}_{\text{meas.}} \right\|^2 \quad (\text{F1})$$

where both $\vec{\lambda}(\vec{f})$ and $\vec{f}_{\text{meas.}}$ are sorted in ascending order.

This dispersive shift compensation scheme relies on the accuracy of the characterized \mathbf{J} . In practice, we do not characterize the full matrix \mathbf{J} , but instead, only measure nearest and next-nearest neighbor couplings. Stray capacitances, however, can exist between any two qubits in an array. Using the measured frequencies from the size $M = 200$ initial training set, we compare the calculated \vec{f} when using \mathbf{J} versus using a modified \mathbf{J}' with stray capacitances. We set the beyond-next-nearest-neighbor elements of \mathbf{J}' to the absolute value of random numbers normally distributed around zero, with a standard deviation of $\sigma \in J \times \{0.01, 0.03, 0.05\}$, where $J = 2\pi \times 5.98$ MHz is the mean nearest-neighbor coupling for our device. The corresponding mean frequency error, assuming \mathbf{J}' is the true coupling matrix, is $\Delta f = 49, 122, 164$ kHz for $\sigma = 0.01J, 0.03J, 0.05J$. We therefore estimate that the error due to dispersive shift compensation is 120 kHz. This error is around 67% of the average dispersive shift compensation of 180 kHz, which means that the overall benefits of our dispersive shift compensation are minimal, given our experimental conditions.

The qubits also have residual ZZ coupling, which leads to shifts in the qubit frequencies depending on the states of their neighbors. Assuming the qubits have only three energy levels, the frequency shift due to residual ZZ coupling between two qubits (1 and 2) is given by:

$$\Delta\omega_1 = \frac{\alpha_1 + \alpha_2}{(\Delta_{12} + \alpha_1)(\Delta_{12} - \alpha_2)} J^2 \quad (\text{F2})$$

where α_i is the anharmonicity of qubit i , J is the coupling strength between the qubits, and $\Delta_{12} = \omega_1 - \omega_2$ is the frequency detuning between the qubits [38, 39].

From the measured frequencies of the size $M = 200$ training set used in our experiment, we can find the mean shift due to nearest-neighbor ZZ coupling. We assume that each anharmonicity is the mean of the anharmonicities: $\alpha_i/2\pi = -218.4$ MHz. We also assume that each nearest neighbor coupling is the mean of the nearest neighbor couplings, $J/2\pi = 5.89$ MHz, for our device, and the coupling between qubits that are not directly capacitively coupled is 0. We find that the standard deviation of the frequency shifts due to ZZ coupling is approximately 231 kHz. Since next-nearest neighbor qubits also interact somewhat strongly and non-nearest neighbor qubits also interact via stray capacitances, we believe this estimate of error (230 kHz) to be conservative.

In Table II, we report the estimated error for each of these sources, as well as a total estimated error. We assume these sources of error are uncorrelated and compute the total estimated error by adding in quadrature. We estimate a total average error of 280 kHz, which is less than the frequency measurement uncertainty of $\sigma_{\text{meas.}} = 0.5$ MHz we use in our protocol simulations. Other potential sources of frequency error are more challenging to characterize, such as small frequency shifts due to weak coupling with coherent defects or ZZ coupling due to higher transmon energy levels. From the conservative error budget discussed above and other potential sources of error, we believe that $\sigma_{\text{meas.}} = 0.5$ MHz is a reasonable upper bound for the frequency measurement uncertainty on our device.

Appendix G: Fast flux crosstalk calibration

We can control our qubits via the local flux lines by applying DC voltages or baseband fast flux pulses. These fast flux pulses are square pulses with a cosine ramp and no modulations. The learning-based protocol described in this paper can also be used for fast flux crosstalk calibration. In Fig 9, we report the fast flux crosstalk matrix for our 16-qubit flip-chip device. We find that the off-diagonal crosstalk levels are much smaller for fast flux pulses, with most off-diagonal elements having a magnitude on the order of 0.01%. The average off-diagonal DC flux crosstalk level (Fig. 8A) is around 1%, so the fast flux crosstalk levels are around two orders of magnitude smaller than DC flux crosstalk levels.

In order to learn the fast flux crosstalk for each qubit, we tune that qubit to a target frequency using a 100 ns fast flux pulse and measure the qubit frequency via spectroscopy. We then use the same pulse amplitude for the target qubit, apply flux pulses with random amplitudes through the other flux lines, and measure the change in the target qubit’s frequency. The voltages applied and the changes in the frequency of the target qubit from the training set that we use for learning the fast flux crosstalk

matrix are shown in Fig. 9.

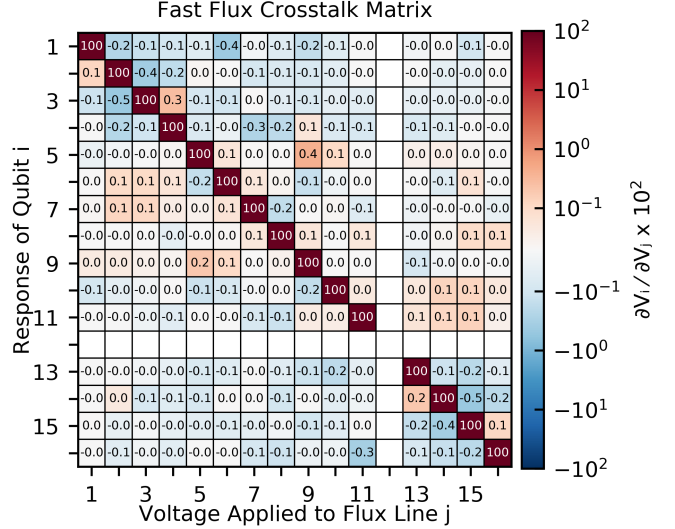


FIG. 9. **The fast-flux crosstalk matrix** The measured fast flux crosstalk matrix for the 16-qubit flip-chip device, rescaled by a factor of 100 such that each element is a percentage. The average off-diagonal crosstalk level is approximately two orders of magnitude smaller for fast flux crosstalk than for DC flux crosstalk. Elements labeled ± 0.0 are crosstalk percentages of magnitude $< 0.05\%$. The fast flux microwave line for qubit 12 was broken, so there is no information for this qubit.

Appendix H: Single-qubit Gate Fidelity

In this paper, we report an experimental median frequency error on our 16-qubit device of ≈ 300 kHz after learning the crosstalk matrix with a size $M = 50$ training set. We now address the effect of this error on single-qubit gate fidelity.

We would like to be able to arrange the qubits in frequency layouts with high fidelity so that we can perform quantum simulations and/or computations. If qubit frequencies are off from their target frequencies by over 10 MHz (as is the case on our device before learning the crosstalk), then qubits may interact more or less strongly than intended. An important thing to note is that after setting qubits in a frequency layout, we would perform Ramsey measurements to further refine the qubit frequencies (i.e. ensure we are driving qubits on resonance). Also, we could repeat the crosstalk learning protocol for a specific frequency layout to further reduce frequency errors.

In the event that we do not tune up our drive frequencies, we could consider the impact of the frequency error on our single-qubit gate fidelities. For an X -gate, a drive detuning tilts the effective driving field away from the X -axis in the $X - Z$ plane. In Fig. 10, we can see the unitary fidelity for $\Delta f \in \{0.1, 0.3, 0.5\}$ MHz as we apply successive 2π rotations around the X -axis, with a

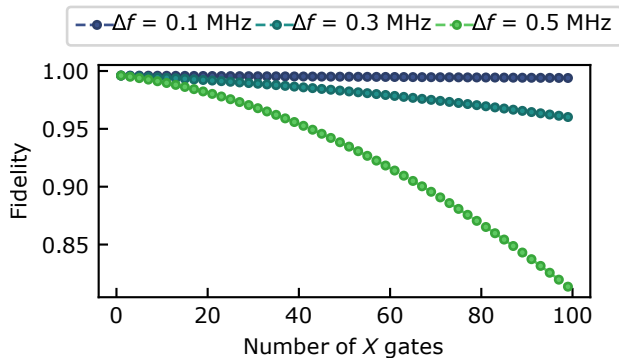


FIG. 10. **Simulations of single-qubit gate fidelity** in the presence of frequency error. The qubit is prepared in the excited state and then 2π -pulses are successively applied around the X -axis, in the presence of frequency detuning Δf . We assume a drive strength that induces a Rabi frequency of $\omega_R = 2\pi \cdot 10$ MHz.

pulse amplitude corresponding to a Rabi frequency of $\omega_R = 2\pi \cdot 10$ MHz. For $\Delta f < 0.3$ MHz, we maintain over 95% fidelity with 100 successive 2π - X gates. We therefore consider 300 kHz to be an acceptable median frequency error, especially since we can further improve upon this frequency error by performing Ramsey measurements or learning the crosstalk for a specific frequency layout.

Appendix I: Bias Region Extrapolation

We use the frequency region spanned by 100 MHz to 1 GHz off the sweet spot to learn our crosstalk matrix. We now explore whether the crosstalk matrix learned in this region extrapolates to other bias regions.

In Fig. 11, we simulate the protocol’s performance when evaluated in different bias regions. We still learn the crosstalk matrix using the standard training region (100 MHz to 1 GHz off the sweet spot), and then report the frequency error when attempting to set qubit frequencies to the 5 different bias regions highlighted in Fig. 11A. For each bias region, we choose frequency vectors that place each qubit in that bias region (i.e. all voltages are nonzero). We note that in these simulations, we disregard the detuning conditions normally required in our experiments.

In Fig. 11B, we see that the performance of the protocol is best for the bias region closest to the sweet spot, which makes sense since the applied voltages are the smallest. The performance of the protocol decreases the further the bias region is from the upper sweet spot, with the worst performance occurring in regions 4 and 5, with ≈ 800 kHz and ≈ 850 kHz median frequency error, respectively, for training set size $M = 50$.

In Appendix H, we found that a frequency error of 500 kHz impacts gate fidelities, particularly for more

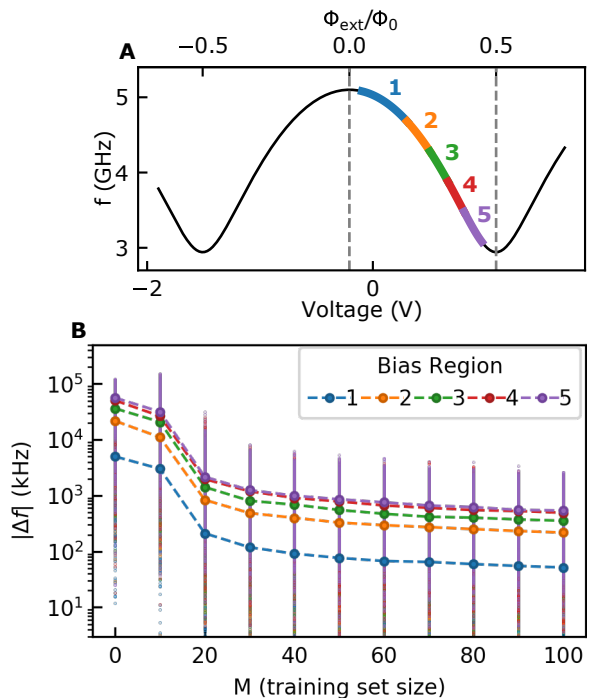


FIG. 11. **Simulation of frequency error in other bias regions.** The crosstalk is learned by placing the qubit frequencies randomly in the standard training region spanned by 100 MHz to 1 GHz off the sweet spot, and then the error in setting qubit frequencies is determined for 5 different bias regions, highlighted in (A). We target random frequency vectors in each bias region to evaluate the learned crosstalk matrix. For these simulations, we disregard the detuning conditions normally used in our experiments. In (B), we show the convergence of frequency error in each bias region as the training set size increases. The frequency precision is best for Region 1, closest to the sweet spot, and decreases the further the bias region is from the upper sweet spot. The medians are computed over 100 simulation repetitions, and all data points are shown on the plot.

than 20 successive gates. Therefore, we expect that using the learned crosstalk matrix from the standard training region for frequency layouts with qubits primarily placed in bias regions 4 and 5 will result in worsened computational performance.

We note, however, that the frequency error in these regions follows the same convergence pattern as in other regions while starting from a higher initial error. In this sense, the learned crosstalk matrix extrapolates to different bias regions. Furthermore, we expect that frequency layouts with different qubits placed in different bias regions (as described in Fig. 11A) will have a median frequency error somewhere in between the minimum and the maximum errors in Fig. 11B. To improve frequency setting precision in other bias regions, we would recommend repeating crosstalk training for the specific frequency layouts or bias regions desired. Additionally, we can perform Ramsey measurements to further refine

qubit frequencies.

Appendix J: Single-Qubit Characterization

Before we can conduct the learning-based protocol described in this paper, we must perform single-qubit characterizations. One of the essential assumptions of the protocol is that we can convert accurately between flux experienced by a qubit's SQUID loop and its frequency. This conversion is described by the transmon spectrum (Eq. 1). In the case of a single qubit, the protocol relies on the ability to convert between voltage V applied to its flux line and flux Φ_{ext} experienced by the qubit's SQUID loop. This conversion is described by $\Phi_{\text{ext}} = V/V^{\Phi_0} + \Phi_{\text{offset}}$, where V^{Φ_0} is the voltage required to tune the qubit by one magnetic flux quantum Φ_0 , and Φ_{offset} is a flux offset.

We determine V^{Φ_0} , Φ_{offset} , f^{max} , d , and E_C by performing qubit spectroscopy. We sweep the voltage V across the full extent of our tuning range (which in practice turns out to be approximately $\pm 0.3V^{\Phi_0}$) and measure the qubit frequency. We fit this curve with Eq. 1, substituting $\Phi_{\text{ext}} = V/V^{\Phi_0} + \Phi_{\text{offset}}$, where V^{Φ_0} . Naturally, the more data points we take for the spectroscopy fit, the more precise our fit parameters will become. We investigate the precision of our fit parameters under the presence of frequency measurement uncertainty $\sigma_{\text{meas.}} \in \{0.1, 0.5, 1.0\}$ MHz.

In Section K, we investigate how errors in fit parameters impact the convergence of the learning-based protocol. From our analysis in this section (Fig. 12), we can see that we can achieve the required precision in fit parameters with < 15 frequency measurements per qubit.

Appendix K: Error Analysis

We investigated, via simulation, the effect of error in the system on the efficacy of the protocol. One key assumption of the protocol is that we can convert between a qubit's frequency f and flux experienced by its SQUID loops Φ using the transmon spectrum, Eq. 1. If the parameters f^{max} , d , or E_C have significant errors, the protocol will break down because we will be unable to accurately determine the fluxes experienced by the SQUIDS to train \mathbf{S} (i.e. we won't have an accurate training set $\{\vec{V}_i, \vec{\Phi}_i^{\text{meas}}\}_{i=1:M}$). We note that Eq 1 is an estimation, which is another potential source of error in our experimental results. We assume, however, that the transmon spectrum is an exact equation for all simulations of protocol performance.

We simulated the calibration protocol for the 16-qubit array while varying the standard deviation of error for these parameters, shown in Fig. 13. We find that the protocol begins to fail when the standard deviation of f_{max} exceeds 0.5 MHz, the standard deviation of d exceeds 1%, or the standard deviation of E_C exceeds 10%.

Another key assumption of the protocol is that we can form an accurate estimate for the flux experienced by the SQUID loops, given an optimized crosstalk matrix and the voltages applied. This is the linear crosstalk relation:

$$\vec{\Phi}_{\text{ext}} = (\mathbf{V}^{\Phi_0})^{-1} \mathbf{S} \vec{V} + \vec{\Phi}_{\text{offset}} \quad (\text{K1})$$

This assumption breaks down when there are errors in the measured \mathbf{V}^{Φ_0} and $\vec{\Phi}_{\text{offset}}$. Errors in \mathbf{V}^{Φ_0} and $\vec{\Phi}_{\text{offset}}$ will render the argument of the cost function inaccurate:

$$C(\mathbf{S}_k) = \frac{1}{M} \sum_{i=1}^M \left\| (\vec{\Phi}_i)_k - \left[(\mathbf{V}_{k,k}^{\Phi_0})^{-1} \mathbf{S}_k \vec{V}_i + (\vec{\Phi}_{\text{offset}})_k \right] \right\|^2 \quad (\text{K2})$$

preventing us from converging to the correct minimum.

We simulated the calibration protocol for the 16-qubit array while varying the standard deviation of error for these parameters, shown in Fig. 14. We find that the protocol fails when the standard deviation of \mathbf{V}^{Φ_0} exceeds 0.1% or the standard deviation of $\vec{\Phi}_{\text{offset}}$ exceeds 1%.

Appendix L: Efficient Recalibration

An advantage to the learning-based approach to calibration is that it can be used to efficiently recalibrate the crosstalk matrix. With a direct measurement approach, there is no simple way to recalibrate the crosstalk without repeating the measurements over again.

We follow the same procedure, except we use a good initial guess for \mathbf{S} instead of the identity. This will mean that the qubit frequencies will already be close to their targets with no additional training, enabling us to perform a finer spectroscopy scan with less measurement uncertainty. In Fig 15, we can see that the frequency error quickly converges to the median error from Fig 2 (grey line, assuming $M = 30$ and measurement uncertainty = 0.5 MHz). This means that we can use the same protocol to recalibrate our crosstalk matrix efficiently.

Appendix M: Training V^{Φ_0} and Φ_{offset}

Another advantage of this learning-based protocol is that it can be used to efficiently retrain \mathbf{V}^{Φ_0} and $\vec{\Phi}_{\text{offset}}$. For our superconducting devices, the DC flux crosstalk matrix remains essentially constant throughout an entire cooldown and also changes minimally from cooldown to cooldown. The qubits' V^{Φ_0} and Φ_{offset} , however, can drift throughout a cooldown. In particular, the flux offsets can drift substantially.

Once we have learned \mathbf{S} , we can use the same learning-based protocol to retrain \mathbf{V}^{Φ_0} and $\vec{\Phi}_{\text{offset}}$. Instead of optimizing \mathbf{S} in the gradient descent optimizer, we can instead optimize one or both of \mathbf{V}^{Φ_0} and $\vec{\Phi}_{\text{offset}}$ by minimizing $C(\mathbf{V}^{\Phi_0}, \vec{\Phi}_{\text{offset}})$.

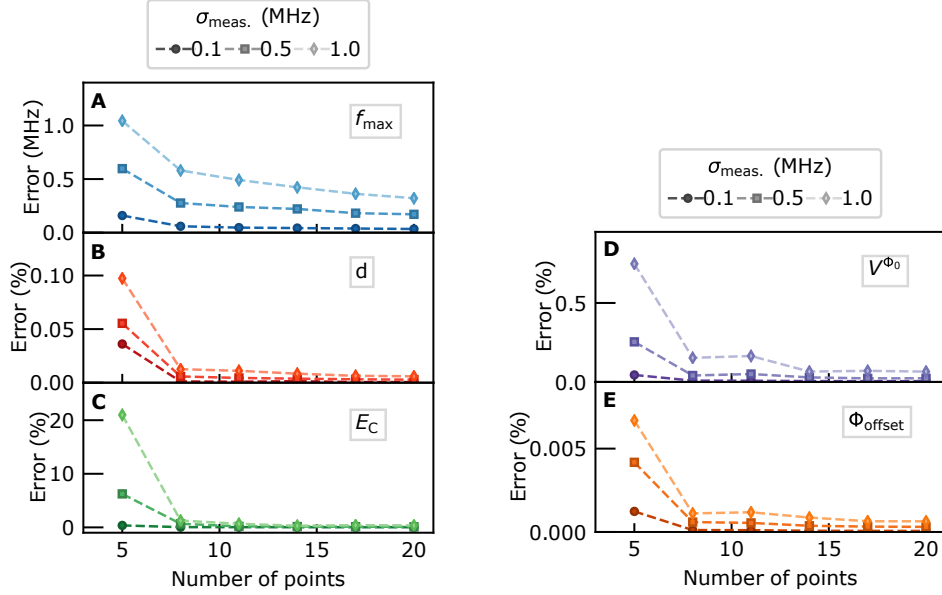


FIG. 12. **Sensitivity analysis of transmon spectrum fits** The mean error in (A) f^{\max} (MHz), and the mean percent error in (B) d , (C) E_C , (D) V^{ϕ_0} , and (E) Φ_{offset} , versus the number of data points in the spectroscopy sweep for varying levels of frequency measurement uncertainty $\sigma_{\text{meas.}}$. The error converges by around 15 measurements. The means are computed over 50 simulation repetitions.

Appendix N: Crosstalk level scaling

The calibration approach outlined in this paper is applicable to other systems with low to moderate levels of crosstalk. We simulated the convergence of the protocol for a 16-qubit array while varying the magnitude of the off-diagonal crosstalk elements of the target crosstalk matrix $\mathbf{S}_{\text{target}}$. For a given off-diagonal crosstalk level ℓ , we randomly assign each off-diagonal element of $\mathbf{S}_{\text{target}}$ to be $\pm\ell$. In each simulation repetition, this arrangement of plus and minus changes, but the magnitude of

each off-diagonal element remains fixed.

We find that the protocol breaks down when the off-diagonal crosstalk levels exceed 10%. For a training set of size $M = 100$, the median crosstalk matrix error Fig. 16D begins to rise after off-diagonal crosstalk levels reach 6%. While crosstalk matrix error is a potentially useful metric for protocol simulations, it is not measurable experimentally, as the target crosstalk matrix is unknown. The experimentally measurable quantity, median frequency error Fig. 16C, begins to rise later after off-diagonal crosstalk levels reach 10%. This protocol will function well for systems with crosstalk levels $\leq 10\%$.

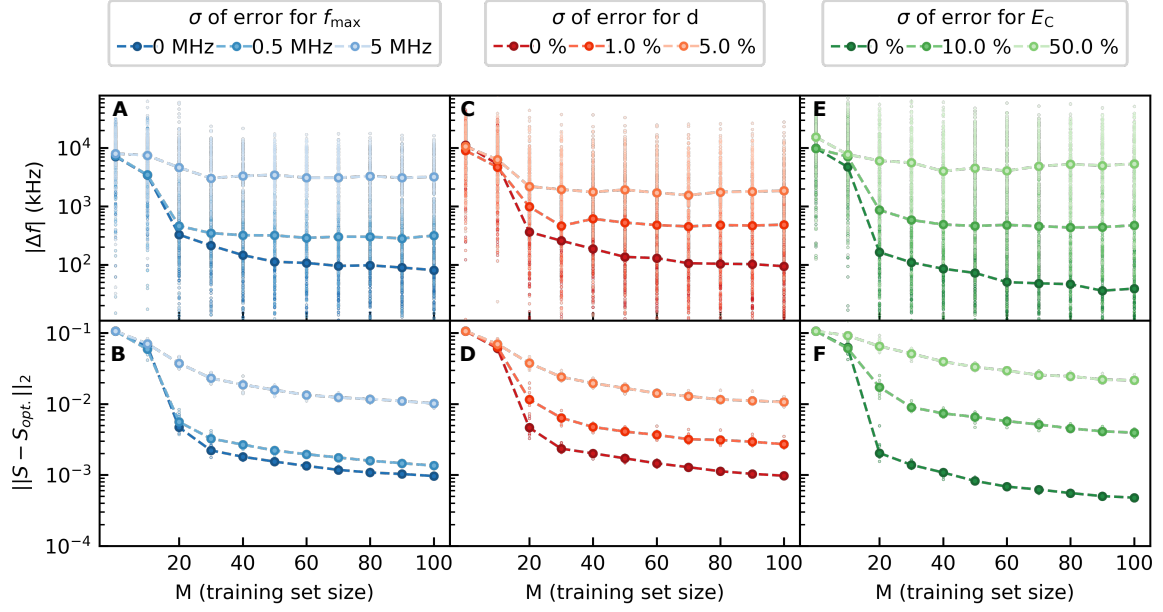


FIG. 13. **Effect of error in transmon spectrum parameters on crosstalk precision** Simulation of protocol convergence for varying levels of error in f^{\max} (A,B), d (C,D), and E_c (E,F). The frequency error is shown in (A,C,E) and the Euclidean norm between the trained crosstalk matrix $S_{\text{opt.}}$ and the actual crosstalk matrix S is shown in (B,D,F). The distribution over 10 different random realizations is shown using small circles, and the median values are shown using large circles. The convergence of the protocol is robust to small errors in these parameters but breaks down in the large error limit. The medians are computed over 10 simulation repetitions, and all data points are shown on the plot. A frequency measurement uncertainty of 0.5 MHz (see Fig. 2) was used for these simulations.

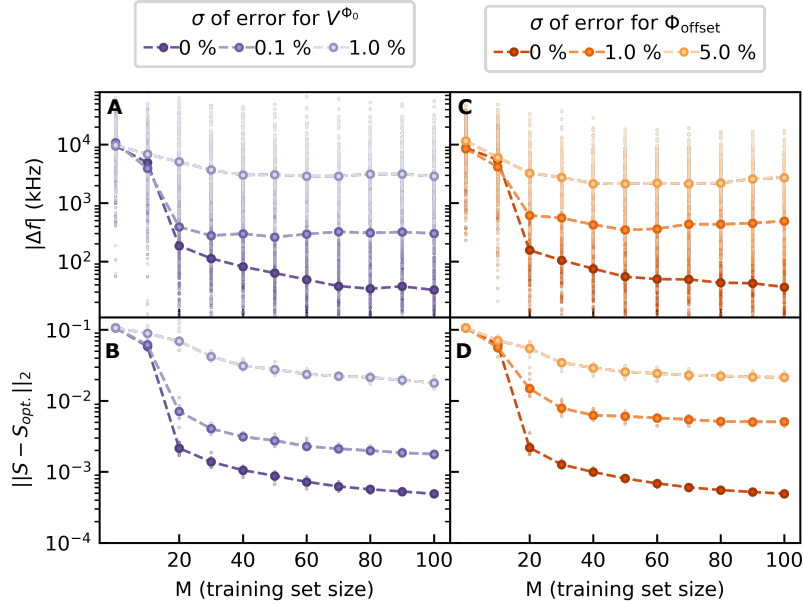


FIG. 14. **V^{Φ_0} and Φ_{offset} error analysis** Simulation of protocol convergence for varying levels of error in V^{Φ_0} (A,B) and Φ_{offset} (C,D). The frequency error is shown in (A,C) and the Euclidean norm between the trained crosstalk matrix $S_{\text{opt.}}$ and the actual crosstalk matrix S is shown in (B,D). The distribution over 10 different random realizations is shown using small circles, and the median values are shown using large circles. The convergence of the protocol is particularly sensitive to errors in V^{Φ_0} , and fairly sensitive to errors in Φ_{offset} . A frequency measurement uncertainty of 0.5 MHz (see Fig. 2) was used for these simulations.

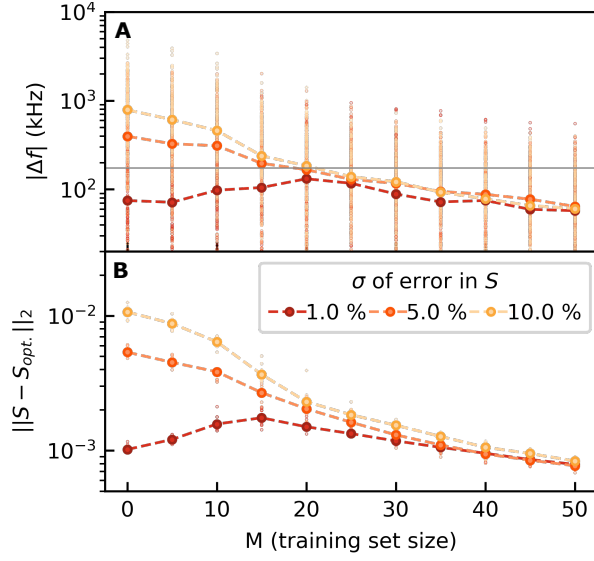


FIG. 15. **Crosstalk matrix recalibration** Simulation of protocol convergence for varying levels of error in the off-diagonal estimates of the initial crosstalk matrix used at the start of training. Both the (A) median frequency error and the (B) median Euclidean norm between the trained crosstalk matrix $S_{\text{opt.}}$ and the actual crosstalk matrix S converge by $M = 20$ to the originally trained median from Fig 2 (grey line, assuming $M = 30$ and measurement uncertainty = 0.5 MHz). The medians are computed over 10 simulation repetitions, and all data points are shown on the plot. A frequency measurement uncertainty of 0.25 MHz was used for these simulations. Since we already know where the qubits are, we can assume a finer frequency scan (lower power leading to less power broadening) and therefore less measurement uncertainty.

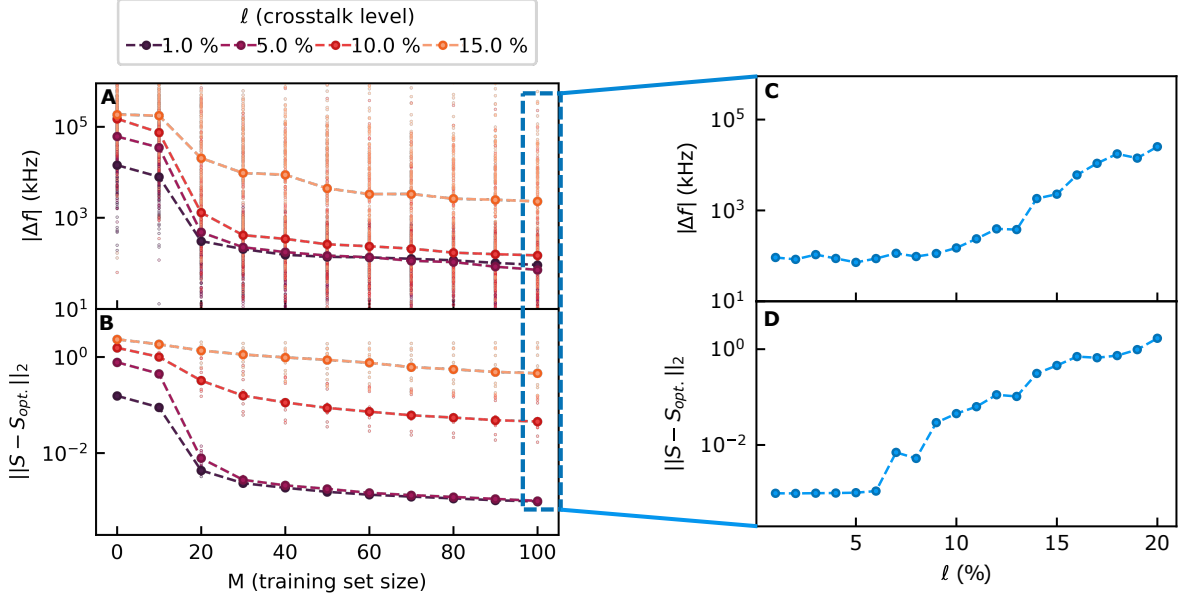


FIG. 16. **Off-diagonal crosstalk level** Simulation of protocol convergence for varying levels of off-diagonal crosstalk in the target crosstalk matrix S_{target} . Both the median frequency error (A) and the median Euclidean norm between the trained crosstalk matrix S_{opt} and the actual crosstalk matrix S (B) begin to fail to converge when off-diagonal crosstalk levels exceed 10%. For a training set of size $M = 100$, the median frequency error (C) and median crosstalk matrix error (D) grow exponentially for large off-diagonal crosstalk levels. The medians are computed over 10 simulation repetitions, and all data points are shown on the plot. A frequency measurement uncertainty of 0.5 MHz (see Fig. 2) was used for these simulations.