Laser-annealing Josephson junctions for yielding scaled-up superconducting quantum processors.

Jared B. Hertzberg, Eric J. Zhang, Sami Rosenblatt, Easwar Magesan, John A. Smolin, Jeng-Bang Yau, Vivekananda P. Adiga, Martin Sandberg, Markus Brink, Jerry M. Chow, and Jason S. Orcutt
IBM Quantum, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA
(Dated: September 16, 2020)

As superconducting quantum circuits scale to larger sizes, the problem of frequency crowding proves a formidable task. Here we present a solution for this problem in fixed-frequency qubit architectures. By systematically adjusting qubit frequencies post-fabrication, we show a nearly ten-fold improvement in the precision of setting qubit frequencies. To assess scalability, we identify the types of ‘frequency collisions’ that will impair a transmon qubit and cross-resonance gate architecture. Using statistical modeling, we compute the probability of evading all such conditions, as a function of qubit frequency precision. We find that without post-fabrication tuning, the probability of finding a workable lattice quickly approaches 0. However with the demonstrated precisions it is possible to find collision-free lattices with favorable yield. These techniques and models are currently employed in available quantum systems and will be indispensable as systems continue to scale to larger sizes.

I. INTRODUCTION

Realizing robust large-scale quantum information processors is one of the foremost challenges in quantum science. Many practical applications have been proposed for robust quantum computers, including estimating the ground state energy of chemical compounds and implementing machine learning algorithms [1–5]. Quantum advantage relative to classical computers can be realized without full fault-tolerance, but requires large quantum circuits that a classical computer cannot simulate [6]. Recent demonstrations have shown qubit circuits nearly at the threshold for demonstrating quantum advantage [7]. Much work remains in order to realize fault-tolerant quantum processors; however, scale-up of solid-state quantum circuits has shown consistent and ongoing progress [8–15]. As the qubit circuits are scaled up, they must maintain high one- and two-qubit gate fidelities, high qubit connectivity, and low cross-talk error which can be measured in a holistic sense via the quantum volume of the circuit [16]. Lattices of fixed-frequency transmon qubits represent a promising architecture for building systems of larger sizes [7]. A growing number of systems at the 20 to 50-qubit scale are now available to users through cloud access. A variety of technical challenges confront further system scaling, including improving 3-dimensional circuit integration and qubit coherence. High on the list of such challenges is the issue of ‘frequency crowding.’

Fixed frequency transmon qubits using the two-qubit cross-resonance (CR) gate form a promising architecture for scaling up quantum systems. Fixed-frequency transmons are largely insensitive to charge or flux noise, and have achieved coherence times of 100 μs and growing. The CR gate, a hardware-efficient all-microwave gate [17–20], is readily used to entangle these qubits with gate fidelities above 99%, approaching the threshold for fault-tolerant codes [21]. To achieve these fidelities, the CR gate needs not only high coherence qubits, but also a precise setting of the qubits’ frequencies. The CR gate activates a ZX interaction by driving one ‘control’ qubit with a microwave pulse at the other ‘target’ qubit’s transition frequency. The magnitude of the ZX as well as other Hamiltonian terms depends on the relative frequencies of the two qubits [22, 23]. Diminished ZX magnitude increases gate time, while other terms such as ZZ add gate errors. Neighboring qubits having the wrong detuning will exhibit a ‘frequency collision’ in which the ZX may be suppressed or other undesirable effects arise.

Maintaining high gate fidelities for all pairs in a lattice will require solving this ‘frequency crowding’ problem by precise setting of qubit frequencies to specified values, as characterized by a standard deviation $\sigma_f$. To achieve low $\sigma_f$, the tunnel-junction conductance must be controlled with high precision. Transmon frequency $f_0$ follows $h f_0 \simeq \sqrt{8 E_J E_C} - E_C$, where Josephson energy $E_J = \frac{e^2}{C}$ is many times greater than charging energy $E_C = \frac{e^2}{2 C}$ [24]. In typical transmons, a photolithographically defined capacitance $C$ has dimensions in the tens to hundreds of microns and varies little from qubit to qubit. The critical current $I_c$ is set by a tunnel barrier of area $\sim 100 \times 100 \text{ nm}$ and thickness a few nm, and is thus challenging to fabricate with precision better than a few percent [25–28]. However, tunnel barrier resistance $R_n$ is readily measurable to precision better than 0.1% and relates to $I_c$ according to the Ambegaokar-Baratoff relation $I_c = \frac{\Delta}{2 e R_n}$ (where $\Delta$ is the superconducting gap energy) [29]. We can therefore measure $R_n$ before a chip is cooled in order to assess qubit frequency imprecision. The best demonstrated precision in setting $R_n$ at time of fabrication is 2% [28]. A 2% variation in $R_n$ indicates a fractional $\sigma_f$ of 1%.

Careful design of lattices can enable error correction codes while at the same time minimizing the likelihood of ‘frequency collisions’ and therefore the required $\sigma_f$ for fabrication yield [30, 31]. Yet even the most robust designs require a fractional $\sigma_f$ of 0.25% to 0.5%, which represents a factor of 2 to 4 improvement over the
best literature results. To overcome such limits will require rework of individual qubits’ tunnel junctions after fabrication. Thermal anneal has been shown to increase tunnel resistance $R_n$, and laser heating has been demonstrated as a highly localized re-work tool [32–37]. However, the inherent variability of the anneal process itself must be overcome, and qubit frequency control utilizing such techniques at scale has never been presented in the literature.

In this paper, we introduce an adaptive post-fabrication trimming technique that we use to incrementally adjust $R_n$ on a qubit-by-qubit basis, thereby overcoming inherent variability in both initial qubit fabrication and the laser anneal. For the first time, an improvement in qubit frequency precision is demonstrated in terms of narrowed frequency distributions. Crucially, we demonstrate qubit frequency imprecision $\sigma_f$ of the same magnitude as the imprecision of predicting $f_{01}$ from $R_n$. To estimate the scalability of this technique for the fabrication of error-corrected lattices, we employ a statistical yield model based on $\sigma_f$ relative to specific collision bounds. This model predicts the severity of the frequency crowding problem for different topologies and scales of error corrected multi-qubit lattices as a function of code distance. The model demonstrates that using conventional transmon fabrication, scaled-up qubit lattices will fail to evade ‘frequency collisions’. However, our novel trimming technique achieves adequate $\sigma_f$ for scalable fabrication of distance-3 through distance-7 heavy-square and heavy-hexagon codes. In particular, this technique enables the high yield fabrication of the distance-3 and distance-5 heavy-hexagon lattices currently deployed as IBM cloud connected systems [38].

II. RESULTS

A. Frequency Precision $\sigma_f$ From Transmon Fabrication

To assess the $\sigma_f$ resulting from qubit fabrication, we developed a test vehicle containing a large number of identically-fabricated qubits (fig. 1). We cooled the chip in a dilution refrigerator and used dispersive readout through half-wave microwave resonators to measure qubit frequencies [39]. We measured the frequencies of 31 qubits to a precision better than 100 kHz using a Ramsey fringe method. The qubit frequencies had random variation $\sigma_f = 132.3$ MHz (fig. 1), or 2.3% of the median frequency. After warming the qubits to room temperature, we measured their junction resistances. In fig. 1, we show a plot of $R_n$ compared to transmon frequency, demonstrating that the observed variation in $\sigma_f$ is accounted for almost entirely by $R_n$ variation. The behavior may be fit to a power law of approximately $-\frac{1}{2}$ power, as expected from transmon theory and the Ambegaokar-Baratoff formula. For a population of transmon qubits whose frequency scatter is dominated by scatter in $R_n$, we expect the fractional standard deviation in $R_n$ to be twice that of $\sigma_f$. This is consistent with the standard deviation in junction resistances which is found to be $\sigma_R$ of 365 $\Omega$, or 4.6% of the median $R_n$. We also assess the fidelity of the frequencies to the $f$-vs-$R_n$ correlation in terms of the residual scatter after subtracting the fit line from the frequency values. This appears in the inset in fig. 1 and exhibits a standard deviation 14.5 MHz, or 0.25% of the qubit median frequency.

B. Tuning Using Selective Laser Anneal

To reduce $\sigma_f$, we developed a technique for selective laser anneal to shift tunnel resistance $R_n$ by pre-calibrated increments. (See section IV and fig. 2). We demonstrate the achievable frequency control of this technique by shifting the 31 measured qubits into a two-frequency pattern. We employed an $R_n$ vs $f$ correlation (fig. 1) to designate the target resistances. We shifted 16 junctions to one target $R_n$ and 15 to another target $R_n$. After tuning, the group of 16 junctions had median resistance 7.984 k$\Omega$ and the group of 15 had median resistance 8.798 k$\Omega$. The 31 junctions reached their targets with an overall precision of $\sigma_R = 51\Omega$, about 0.61%. In a dilution refrigerator, we re-measured the frequencies of the qubits in the two groups. Aside from two of the qubits, which we measured using CW spectroscopy (precision 2 MHz), all qubits were remeasured in the same way as in the first cooldown. The resulting frequencies appear in fig. 1. The two frequency groups are approximately normally-distributed and have medians $f_{00} = 5.430$ GHz and $f_{01} = 5.7046$ GHz. Calculating $\sigma_f = \sqrt{(f_i - f_{0,j})^2}$, where $f_{0,j}$ represents $f_{00}$ or $f_{01}$ as appropriate for a given qubit $Q_i$, we assess the overall precision $\sigma_f = 14.0$ MHz. This imprecision is nearly identical to the residual scatter from the $f(R)$ fit line (fig. 1) which guided the tuning, and the fractional precision $\sigma_f/f = 0.25\%$ is slightly better than half of the fractional precision in setting $R_n$. By these comparisons, we see that in this experiment $\sigma_f$ is limited by both the precision of setting $R_n$ and the precision of predicting $f$ from $R_n$. Drift in $R_n$ reported in the literature [32] does not appear to be a limiting factor in this study. Achieving smaller $\sigma_f$ will require improvements in setting $R_n$. As we show in section IV, the laser-anneal tuning technique is capable of precisions of 0.3% in $R_n$. On that basis the imprecision of 14.5 MHz in predicting $f$ from $R_n$ would dominate the imprecision in $\sigma_f$.

III. DISCUSSION

Our post-fabrication trimming reduced $\sigma_f$ by $9.5\times$ compared to initial fabrication. To assess whether this level of precision is sufficient to reliably prepare lattices of fixed-frequency transmons capable of error-correcting codes, we must quantify the frequency-crowding problem. Transmon qubits are weakly anharmonic and have decreasing transition energies at higher levels.
Therefore, degeneracies among the $|0\rangle \rightarrow |1\rangle$, $|1\rangle \rightarrow |2\rangle$ and $|0\rangle \rightarrow |2\rangle$ transitions of nearby qubits can all contribute to ‘frequency collisions.’ We must consider the relative frequencies of both nearest-neighbors and next-nearest-neighbors in the lattice [22, 41, 42]. Fig 3 illustrates the relative positions of nearest-neighbor and next-nearest-neighbor qubits in a section of lattice, and table I lists the seven cases most likely to lead to gate errors [22]. We can think of them qualitatively as follows: Type 1 causes hybridization of states in $Q_j$ and $Q_k$, while in type 2 the CR pulse excites $Q_j$ into the non-computational $|2\rangle$ state. Type 3 excites $Q_k$ to the $|2\rangle$ state, but does not require a CR tone. In condition 4, $ZX$ is weak, which implies long gate times and increased gate error [22, 23]. In type 5, the CR gate addresses an additional neighboring qubit. In type 6, when one qubit is the target of a CR gate, its next-nearest-neighbor leaks to the $|2\rangle$ state. Type 7 causes $Q_j$ to leak to the $|2\rangle$ state during a CR gate.

Around each of the ‘frequency collisions’ described in table I, we can designate a window of undesired frequencies. This breaks the frequency space into allowed and forbidden regions. Type 4 listed in table I defines forbidden zones where $ZX$ coupling is too low. For the other six conditions, we forbid regions where the ‘frequency collision’ is the dominant source of gate error. Existing multi-qubit systems with CR gates typically exhibit two-qubit gate errors of 1 to 2 % regardless of frequency [12, 41]. Ref [22] considers an effective-Hamiltonian model for the CR gate, as a function of the relative frequency of control and target qubits. We use this model to estimate the frequency windows for nearest-neighbor collisions (table I, types 1 to 3) to cause gate errors exceeding $\sim 1 \%$. We make an assumption that similar bounds apply to next-nearest-neighbor interactions (types 5 to 7).

A useful lattice of qubits should enable high quantum volume and fault-tolerant operation while avoiding all of the ‘frequency collisions’ and forbidden regions presented in table I. Both lattice layout and the pattern of qubit frequencies are relevant. We consider three types of lattices: square, ‘heavy-square’ and ‘heavy-hexagon’ (fig. 3). Lattices comprise qubits and two-qubit connections, each qubit being linked to no more than four neighbors. In many practical implementations, these links comprise microwave-resonant buses. A square lattice facilitates ‘surface code’ fault-tolerant codes [43]. Recent literature describes hybrids of the surface code with Bacon-Shor type codes, which can be employed in ‘heavy hexagon’ and ‘heavy square’ lattices to achieve fault-tolerance, albeit with lower error thresholds than the surface code [30]. In addition to the data and ancilla qubit roles employed in the surface code, these hybrid codes assign a portion of the lattice as ‘flag’ qubits.

In the square lattice, every qubit in the bulk of the lattice lies on a degree-four vertex, while some at edges have degree two or degree three. If we populate the
square lattice with 5 distinct frequencies of qubits, $f_5 > f_4 > f_3 > f_2 > f_1$, with appropriate spacing between the frequencies, we can avoid all the forbidden regions of table I [10]. In fig. 3, we illustrate this pattern for square lattices capable of distance-5 ($d = 5$) rotated surface codes. Condition 4 of table I requires $f_{\text{control}} > f_{\text{target}}$, so the pattern also fixes the direction of CNOT gate for each pair.

In contrast to the square lattice, the ‘heavy square’ lattice includes both degree-two and degree-four vertices in the bulk. Degree-one, -two or -three vertices appear at the edges. We take advantage of this pattern to make all the degree-two vertices control qubits, using a three-frequency pattern $f_3 > f_2 > f_1$. Since every control qubit (frequency $f_3$) is linked to at most two target qubits, we need only two properly-chosen target-qubit frequencies ($f_1$ and $f_2$) to satisfy conditions 5, 6 and 7 of table I, as shown in fig. 3. A third type of lattice, the ‘heavy hexagon’, uses a similar scheme. Here the bulk of the lattice includes degree-three and degree-two vertices. Additional degree-two and degree-one vertices lie at the edges. In this lattice, all of the ‘frequency collisions’ and forbidden regions can be satisfied using only three frequencies $f_3 > f_2 > f_1$, with all control qubits residing on degree-two vertices with frequency $f_3$.

We use a Monte Carlo model to quantify the frequency-crowding in each lattice type. We sample the qubits at random frequencies drawn from normal distributions characterized by $\sigma_f$, and count the collisions defined in table I. (See section IV.) In fig. 4, we show the mean number of ‘frequency collisions’ predicted by the Monte Carlo model for each lattice type and frequency pattern, as a function of $\sigma_f$. As $\sigma_f \to 0$, the lattice approaches the ideal patterns of fig. 3, and has zero ‘frequency

### Table I. Seven most likely types of ‘frequency collision’

| Type | Definition | Participants | Bounds |
|------|------------|--------------|--------|
| 1    | $f_{j,01} = f_{k,01}$ | Nearest-neighbor qubits $Q_j, Q_k$ | $\pm 17$ MHz |
| 2    | $f_{j,02} = 2f_{k,01}$ | Control qubit $Q_j$, target qubit $Q_k$ | $\pm 4$ MHz |
| 3    | $f_{j,01} = f_{k,12}$ | Nearest-neighbor qubits $Q_j, Q_k$ | $\pm 30$ MHz |
| 4    | $f_{k,01} < f_{j,12}$ or $f_{j,01} < f_{k,01}$ | Control qubit $Q_j$, target qubit $Q_k$ | $-$ |
| 5    | $f_{j,01} = f_{k,01}$ | $Q_j$ is control to $Q_i$ and/or $Q_k$ & is nearest-neighbor to both. | $\pm 17$ MHz |
| 6    | $f_{j,01} = f_{k,12}$ or $f_{i,12} = f_{k,01}$ | $Q_j$ is control to $Q_i$ and/or $Q_k$ & is nearest-neighbor to both. | $\pm 25$ MHz |
| 7    | $f_{i,01} = f_{k,01}$ | $Q_j$ is control to $Q_i$ and/or $Q_k$ & is nearest-neighbor to both. | $\pm 17$ MHz |
collisions’. As $\sigma_f$ increases, the number of ‘frequency collisions’ rises steadily. As $\sigma_f \rightarrow f_{01} - f_{12}$, the different conditions appearing in table I all become likely, and a limiting number of ‘frequency collisions’ is reached. Yield follows the inverse trend, as seen in fig. 4. As $\sigma_f$ increases, the likelihood of finding a ‘collision free’ chip falls off sharply. While the step sizes between frequencies $f_1$ to $f_5$ are important, absolute frequency values are not. Setting $f_1 = 5.0$, $f_2 = 5.07$ and $f_3 = 5.14$ GHz works as well as $f_1 = 5.05$, $f_2 = 5.12$ and $f_3 = 5.19$ GHz.

The yield and mean collision number are a function of the several different collision types and bounds, so they are not readily susceptible to an analytic formulation. However, we can propose a simplified model for yield: in order for a lattice to be collision-free, every qubit in the lattice must fall within some frequency ‘window’ $\pm \Delta f$ relative to its setpoint. Presuming the qubit frequencies are normally distributed, the probability of this occurring goes as the cumulative distribution function, raised to the power $N$, where $N$ is the number of qubits:

$$\left(\int_{-\infty}^{\Delta f/\sigma_f} e^{-\frac{1}{2}x^2} dx\right)^N.$$ 

In the yield plot in fig. 4, we fit this expression to find $\Delta f$ for each lattice.

These model results allow us to predict how different lattice types and frequency patterns will respond to fabrication imprecision. As shown in fig. 4, if imprecision $\sigma_f$ is greater than 30 MHz, any $d = 5$ lattice will exhibit $> 10$ ‘frequency collisions’ of one or another of the types listed in table I, causing the affected gates to have error rates above $\sim 1\%$. However, if $\sigma_f = 10$ MHz then on average the $d = 5$ square lattice will exhibit 5 ‘frequency collisions’, while the ‘heavy square’ and ‘heavy hexagon’ lattices will exhibit 0.1 ‘frequency collision’. Considered in terms of yield, we see from fig. 4 that if $\sigma_f = 10$ MHz, then for a $d = 5$ device, a square lattice with 5-frequency pattern has a 0.8% likelihood to be ‘collision free’, whereas a ‘heavy square’ lattice with 3-frequency pattern has 90% likelihood and ‘heavy hexagon’ with 3-frequency pattern has 92% likelihood. Alternatively we can ask, how well do we have to control $\sigma_f$? If we seek a 10% yield, then fig. 4 indicates that for a $d = 5$ device, a square lattice with 5-frequency pattern requires $\sigma_f < 8$ MHz, whereas a ‘heavy square’ lattice with 3-frequency pattern requires $\sigma_f = 16$ MHz and ‘heavy hexagon’ with 3-frequency pattern requires $\sigma_f = 17$ MHz. Although the square lattice requires 10 to 20% fewer qubits than the other types at each distance $d$, it requires far better frequency precision.

The as-fabricated $\sigma_f$ seen in fig. 1 is 132.3 MHz. (See section II) The Monte Carlo modeling finds that for a heavy-hexagon lattice at $d = 3$ scale this $\sigma_f$ can enable 0.1% yield of collision-free chips. Other lattice types and larger scales will all have yield $\ll 0.1\%$. The re-tuned $\sigma_f = 14.0$ MHz demonstrated in fig. 1 will improve the yield in all types of lattice. Predictions of the Monte Carlo model for $\sigma_f = 14.0$ MHz appear in table II. At $d = 5$ scale the heavy-hexagon and heavy-square lattices and 3-frequency patterns should be collision-free nearly one-third of the time, while at $d = 7$ scale the yield is about four times smaller, still reasonable for prototype systems.

As seen from the Monte Carlo analysis, the laser-anneal rework method can scale to the $> 100$ qubit size, enabling a well-chosen lattice and frequency pattern to implement $d = 7$ error-correction codes free of frequency-crowding. To examine needs for the next generation of chips up to the 1000-qubit level, we can coarsely estimate requirements by extrapolating the fixed window model for the

FIG. 3. **Lattice and frequency-pattern examples.** Lattices are capable of $d = 5$ codes. Patterns of qubit frequencies avoid all conditions in table I. Statistical model is applied to these examples and to equivalent lattices at $d = 3$ and $d = 7$. (See Supplementary Figures 1-9.) Square lattice includes 49 qubits in 5-frequency pattern. Heavy-square lattice includes 73 qubits in 3-frequency pattern. In a portion of the heavy-hexagon lattice, we indicate qubits’ intended gate roles: control (black circles) or target (white circles); as well as code roles: data (D), ancilla (A) or flag (F) [30]. Inset shows relative positions of qubits for collision definitions of table I. $Q_i$ is coupled to nearest-neighbors $Q_j$ and to $Q_k$. Qubits $Q_i$ and $Q_k$ are next-nearest-neighbors.
heavy-hexagon lattice as shown in fig. 5. While the $\sigma_f = 14.0$ MHz demonstrated here enables practical yield up to the 100 to 200 qubit scale, it is clear that roughly a factor of two further improvement is needed to scale towards 1000 qubits. Since this precision is also better than the resistance-to-frequency prediction precision shown in this work, development of further refinements in tuning and frequency prediction approaches will be necessary as the scale of fixed-frequency transmon circuits surpass the 100 qubit milestone.

![FIG. 4. Frequency-crowding trends as a function of $\sigma_f$. Results of Monte Carlo simulation. (See section IV.) (a) Average number of collisions. and (b) Fraction of cases having zero ‘frequency collisions’. Simulation was applied to the lattices and frequency patterns shown in fig. 3, capable of distance-5 codes, as well as to $d = 3$ and $d = 7$ scale lattices of square, heavy-square or heavy-hexagon type. (See Supplementary Figures 1-9 for lattice layouts and table II for numbers of qubits.) Color-coded dotted lines in part (b) are fits of each lattice yield to expression $\int_{-\infty}^{(\Delta f/\sigma_f)} e^{-\frac{x^2}{2\sigma_f^2}} dx \right)^N$, where $N$ is the number of qubits and $\pm \Delta f$ defines an allowable ‘window’ around frequency set-points. (See table II.) Using this expression, two solid red lines predict yield for the heavy-hexagon lattice type at 300 and 1000 qubits, using $\Delta f = 27.99$ and 26.32 MHz, respectively. See fig. 5 for estimation of $\Delta f$ as a function of qubit number.

![FIG. 5. Yield scaling for heavy hexagon lattice. Fixed window size model. (left axis) Model yields for $\sigma_f$ ranging from the 14 MHz demonstrated in this work to 6 MHz that would optimize future scaling yield beyond the 1000Q level. (right axis) Fixed window size fits to the $d=3$, $d=5$ and $d=7$ heavy hexagon lattice, as well as a fit of these values to expression $A + B \cdot \log(N_{QB})$, extrapolated from 20 to 1000 qubits. This trend illustrates varying frequency crowding constraints as a function of lattice size.

IV. METHODS

A. Chip Fabrication

A chip of the kind used to determine $\sigma_f$ and to test our laser-anneal rework process appears in fig. 1. All microwave elements comprise Nb films $\sim 200$ nm thick on a silicon substrate. Each qubit is coupled to a readout resonator but is not directly coupled to any nearby qubits. All transmon capacitors are identical. Junctions are fabricated using identical electron-beam lithographic patterns and deposited simultaneously using double-angle deposition and oxidation [44]. The individual qubit design is similar to that used in Ref [21] with aharmonicity $f_{12} - f_{01} \approx -330$ MHz. Junctions have linear dimension $\sim 100$ nm and are designed for $I_c$ of $\sim 30$ nA. During packaging, we accidentally damaged three of the 36 qubits and found these to be non-functional when cooled in a dilution refrigerator. We left two of the remaining 33 qubits un-tuned as experimental controls, so that our tuning demonstration includes 31 qubits.

B. Tuning Using Selective Laser Anneal

We have built an integrated junction rework system that can measure and modify the junction resistance. Fig. 2 shows a schematic of our laser annealing system, which we call “Laser Annealing of Stochastically Impaired Qubits” (LASIQ). The laser output is generated by a diode-pumped solid-state laser, frequency doubled to 532 nm. Active power control of the anneal beam is performed using a piezo-rotary mounted waveplate and polarizing beam splitter (PBS), which is adaptively adjusted based on a pick-off beam measured on a downstream silicon photodiode. A precision-timed shutter is
| Lattice & frequency pattern | Code distance | Qubits | $\sigma_f = 132.3$ MHz | $\sigma_f = 14$ MHz |
|-----------------------------|--------------|--------|----------------------|---------------------|
|                             |              |        | Mean number of collisions | Yield of collision-free devices | Mean number of collisions | Yield of collision-free devices | $\Delta f$ Window (MHz) |
| Square, 5-frequency         | $d = 3$      | 17     | 9                     | $\ll 0.1\%$          | 3                  | 6%                      | 13.96               |
|                             | $d = 5$      | 49     | 35                   | $\ll 0.1\%$          | 10                 | $< 0.1\%$              | 13.23               |
|                             | $d = 7$      | 97     | 78                   | $\ll 0.1\%$          | 23                 | $\ll 0.1\%$           | 12.12               |
| Heavy square, 3-frequency   | $d = 3$      | 25     | 10                   | $\ll 0.1\%$          | 0.4                | 67%                     | 30.89               |
|                             | $d = 5$      | 73     | 33                   | $\ll 0.1\%$          | 1.5                | 27%                     | 29.49               |
|                             | $d = 7$      | 145    | 70                   | $\ll 0.1\%$          | 3.5                | 6%                      | 29.06               |
| Heavy hexagon, 3-frequency  | $d = 3$      | 23     | 8                    | 0.1%                 | 0.4                | 70%                     | 31.61               |
|                             | $d = 5$      | 65     | 25                   | $\ll 0.1\%$          | 1.2                | 33%                     | 29.91               |
|                             | $d = 7$      | 127    | 51                   | $\ll 0.1\%$          | 2.7                | 8%                      | 29.29               |

TABLE II. Monte-Carlo model predictions of mean number of collisions and yield of collision-free chips. Full results of Monte-Carlo model of frequency-crowding appear in fig. 4. Here we show values for as-fabricated precision $\sigma_f = 132.3$ MHz, and for demonstrated frequency precision of the laser-anneal method, $\sigma_f = 14$ MHz. (fig. 1) in different lattices and frequency patterns (fig. 3 and Supplementary Figures 1-9). The $\Delta f$ values correspond to the fit lines in fig. 4: in order for the device to be collision-free, every qubit in the lattice must fall within ± this ‘window’, relative to its set-point.

used to control the anneal duration, and beam alignment is performed using a mechanical mirror mount which directs the beam via pattern recognition to the transmon junction center. The beam is shaped as needed to avoid illuminating the junction directly [40].

By careful control of laser power and pulse duration, we use this system to adjust $R_n$. This process overcomes the imprecision due to transmon fabrication, with a residual imprecision $\sigma_f$ due to the rework process. To develop the process, we prepared a set of more than 150 junctions identically to qubit junctions, and measured their response to a range of laser powers and exposure times. We recorded $R_n$ shifts up to 15% relative to initial $R_n$, for anneal durations varying by an order of magnitude and laser powers varying by 20%. Response to laser power in particular was highly nonlinear. Based on these empirical calibrations of $R_n$ shift to power and exposure, we established a qubit tuning process: We first measure the transmon junction’s $R_n$ using four-point probing of the transmon capacitor pads at 25 °C. Using a $f(R_n)$ prediction based on a previously determined correlation curve (fig. 1), we assign the junction a target resistance corresponding to the target frequency in a multiqubit chip lattice. Because the anneal can shift $R_n$ in only one direction, the target must be higher than the initial $R_n$. We anneal the qubit junction using laser power and duration chosen from our calibration set, then re-measure its $R_n$. A junction requiring large shifts in $R_n$ may require repeated anneals to reach its target, as shown in fig. 2. The control algorithm increases the resistance until the measured value is within 0.3% of the target value. In a separate trial of tuning precision, more than 300 junctions were tuned to target $R_n$s ranging from 0.4% to 14.5% above their initial values, and landed successfully within this 0.3% margin. We expect 0.3% imprecision in $R_n$ to introduce 0.15% imprecision in transmon frequency.

C. Monte Carlo Frequency-Crowding Model

Using a Monte Carlo model, we can estimate the incidence of ‘frequency collisions’ in a lattice as a function of $\sigma_f$. We assume that imperfect frequency-setting will distribute qubit frequencies normally around their design frequencies with standard deviation $\sigma_f$. For lattices of the type shown in fig. 3, we designate 3 to 5 frequencies $f_1$, $f_2$, $f_3$, $f_4$, $f_5$ spaced at regular intervals in the pattern shown. We set $f_1 = 5$ GHz, similar to real-world transmons [38, 45]. We sample the qubit frequencies randomly around these values and count the collisions throughout the lattice, as listed in table I. This process is illustrated in fig. 6. We repeat the frequency-assignment and counting to build statistics for a given lattice and frequency pattern. We then repeat the model for a range of $\sigma_f$ values from 0 to 150 MHz. We repeat the entire process over a range of frequency spacings, to find the spacing that minimizes ‘frequency collisions’ at each value of $\sigma_f$. As a function of $\sigma_f$ we can then extract 1) the mean number of total collisions in the lattice, and 2) the fraction of repetitions which result in zero collisions (‘yield’). Our simulations used 1000 repetitions except to find yield below 1% in $d = 5$ lattices and below 0.2% in $d = 3$ lattices, which used 4000 repetitions, and in $d = 7$ lattices to find mean collisions for $\sigma_f < 16$ MHz or yield above 50% (100 repetitions) or to find mean collisions for $\sigma_f > 16$ MHz (40 repetitions).
FIG. 6. Frequency collision statistical model. Left: Square lattice, $d = 3$ with 5-frequency pattern of fig. 3: $f_1 = 5.00$, $f_2 = 5.07$, $f_3 = 5.14$, $f_4 = 5.21$, $f_5 = 5.28$ GHz. To model the lattice statistically, treat the frequencies $f_i$ to $f_5$ as means of distributions. Right: Mean frequencies and normal distributions characterized by $\sigma_f$. For each position in the lattice, sample from the local distribution. Choose random frequencies in this fashion, count collisions as described in table I and repeat to gather statistical sample. Process is repeated for differing spacings between mean frequencies $f_i$ to $f_5$, and for different distribution widths $\sigma_f$.

[1] Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017).
[2] Havlíček, V. et al. Supervised learning with quantum-enhanced feature spaces. *Nature* **567**, 209–212 (2019).
[3] Hempel, C. et al. Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator. *Phys. Rev. X* **8**, 031022 (2018).
[4] Colless, J. et al. Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm. *Phys. Rev. X* **8**, 011021 (2018).
[5] Nam, Y. et al. Ground-state energy estimation of the water molecule on a trapped-ion quantum computer. *npj Quantum Inf.* **6**, 1–6 (2020).
[6] Kandala, A. et al. Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491–495 (2019).
[7] Córcoles, A. D. et al. Challenges and Opportunities of Near-Term Quantum Computing Systems. *Proc. IEEE* 1–15 (2019).
[8] Chow, J. M. et al. Implementing a strand of a scalable fault-tolerant quantum computing fabric. *Nat. Comm.* **5**, 4015 (2014).
[9] Córcoles, A. et al. Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. *Nat. Comm.* **6**, 6979 (2015).
[10] Gambetta, J. M., Chow, J. M. & Steffen, M. Building logical qubits in a superconducting quantum computing system. *npj Quantum Inf.* **3**, 2 (2017).
[11] Takita, M. et al. Demonstration of weight-four parity measurements in the surface code architecture. *Phys. Rev. Lett.* **117**, 210505 (2016).
[12] Takita, M., Cross, A. W., Córcoles, A. D., Chow, J. M. & Gambetta, J. M. Experimental demonstration of fault-tolerant state preparation with superconducting qubits. *Phys. Rev. Lett.* **119**, 180501 (2017).
[13] Kelly, J. et al. State preservation by repetitive error detection in a superconducting quantum circuit. *Nature* **519**, 66–69 (2015).
[14] Risté, D. et al. Detecting bit-flip errors in a logical qubit using stabilizer measurements. *Nat. Comm.* **6**, 6983 (2015).
[15] Ofek, N. et al. Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature* **536**, 441–445 (2016).
[16] Cross, A. W., Bishop, L. S., Sheldon, S., Nation, P. D. & Gambetta, J. M. Validating quantum computers using randomized model circuits. *Phys. Rev. A* **100**, 032328 (2019).
[17] Chow, J. M. et al. Simple All-Microwave Entangling Gate for Fixed-Frequency Superconducting Qubits. *Phys. Rev. Lett.* **107**, 080502 (2011).
[18] Chow, J. M. et al. Universal Quantum Gate Set Approaching Fault-Tolerant Thresholds with Superconducting Qubits. *Phys. Rev. Lett.* **109**, 060501 (2012).
[19] de Groot, P. C. et al. Selective darkening of degenerate transitions for implementing quantum-controlled-NOT gates. *New J. Phys.* **14**, 073038 (2012).
[20] Rigetti, C. & Devoret, M. Fully microwave-tunable universal gates in superconducting qubits with linear couplings and fixed transition frequencies. *Phys. Rev. B* **81**, 134507 (2010).
[21] Sheldon, S., Magesan, E., Chow, J. M. & Gambetta, J. M. Procedure for systematically tuning up crosstalk in the cross-resonance gate. *Phys. Rev. A* **93**, 060302 (2016).
[22] Magesan, E. & Gambetta, J. M. Effective Hamiltonian models of the cross-resonance gate. *Phys. Rev. A* **101**, 052308 (2020).
[23] Ware, M. et al. Cross-resonance interactions between superconducting qubits with variable detuning (2019).
Preprint at http://arxiv.org/abs/1905.11480v1.

[24] Koch, J. et al. Charge-insensitive qubit design derived from the Cooper pair box. Phys. Rev. A 76, 042319 (2007).

[25] Potts, A., Parker, G., Baumberg, J. & de Groot, P. CMOS compatible fabrication methods for submicron Josephson junction qubits. IEEE Proc.: Sci. Meas. Technol. 148, 225–228 (2001).

[26] Wu, X. et al. Overlap junctions for high coherence superconducting qubits. Appl. Phys. Lett. 111, 032602 (2017).

[27] Costache, M. V., Bridoux, G., Neumann, I. & Valenzuela, S. O. Lateral metallic devices made by a multiangle shadow evaporation technique. J. Vac. Sci. Technol., B 30, 04E105 (2012).

[28] Kreikebaum, J. M., O’Brien, K. P., Morvan, A. & Siddiqi, I. Improving wafer-scale Josephson junction resistance variation in superconducting quantum coherent circuits. Supercond. Sci. Technol. 33, 06LT02 (2020).

[29] Ambegaokar, V. & Baratoff, A. Tunneling Between Superconductors. Phys. Rev. Lett. 10, 486–489 (1963).

[30] Chamberland, C., Zhu, G., Yoder, T. J., Hertzberg, J. B. & Cross, A. W. Topological and Subsystem Codes on Low-Degree Graphs with Flag Qubits. Phys. Rev. X 10, 011022 (2020).

[31] Koppinen, P. J., Vist, L. M. & Maasilta, I. J. Complete stabilization and improvement of the characteristics of tunnel junctions by thermal annealing. Appl. Phys. Lett. 90, 053503 (2007).

[32] Granata, C. et al. Trimming of critical current in niobium Josephson devices by laser annealing. J. Phys.: Conference Series 97, 012110 (2008).

[33] Muthusubramanian, N. et al. Local trimming of transmon qubit frequency by laser annealing of Josephson junctions (2019). American Physical Society March Meeting, B29.015.

[34] Oliva, A. & Monaco, R. Annealing properties of high quality Nb/Al-AlO_x/Nb tunnel junctions. IEEE Trans. Appl. Supercond. 4, 25–32 (1994).

[35] Lehnhert, T., Billon, D., Grassl, C. & Gundlach, K. H. Thermal annealing properties of NbAl/AlOxNb tunnel junctions. J. Appl. Phys. 72, 3165–3168 (1992).

[36] Orcutt, J. S. & Rosenblatt, S. United States Patent: 10418540 - Adjustment of qubit frequency through annealing (2019).

[37] Jurcevic, P. et al. Demonstration of quantum volume 64 on a superconducting quantum computing system (2020). Preprint at https://arxiv.org/abs/2008.08571.

[38] Blais, A., Huang, R.-S., Wallraff, A., Girvin, S. M. & Schoelkopf, R. J. Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation. Phys. Rev. A 69, 062320 (2004).

[39] McKay, D. C., Sheldon, S., Smolin, J. A., Chow, J. M. & Gambetta, J. M. Three-Qubit Randomized Benchmarking. Phys. Rev. Lett. 122, 200502 (2019).

[40] Malekakhlagh, M., Magesan, E. & McKay, D. C. First-principles analysis of cross-resonance gate operation (2020). Preprint at http://arxiv.org/abs/2005.00133.

[41] Sheldon, S. et al. Characterizing errors on qubit operations via iterative randomized benchmarking. Phys. Rev. A 93, 012301 (2016).

ACKNOWLEDGEMENTS

We acknowledge funding from the Intelligence Advanced Research Projects Activity (IARPA) under contract W911NF-16-1-0114, for the multi-qubit test vehicle and frequency-vs-resistance correlation studies. We thank N. Bronn, M. Carroll, C. Chamberland, A. Cross, J. Gambetta, J. Ku, M. Malekakhlagh, D. McKay, B. Plourde, E. Pritchett, A. Rosenbluth, M. Takita, J. Timmerwilke and G. Zhu for helpful discussions. We thank E. Porter for coding assistance, Y. Martin and R. Haiht for assistance in constructing the laser optics, and R. Patel for photomicroscopy.
SUPPLEMENTARY INFORMATION for
Laser annealing Josephson junctions for yielding scaled-up superconducting quantum processors)

Jared B. Hertzberg, Eric J. Zhang, Sami Rosenblatt, Easwar Magesan, John A. Smolin, Jeng-Bang Yau, Vivekananda P. Adiga, Martin Sandberg, Markus Brink, Jerry M. Chow, and Jason S. Orcutt

IBM Quantum, IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA

(Dated: September 3, 2020)
Supplementary Figure 1. Distance-3 square lattice, 17 qubits, 5-frequency pattern

Supplementary Figure 2. Distance-3 heavy-square lattice, 25 qubits, 3-frequency pattern

Supplementary Figure 3. Distance-3 heavy-hexagon lattice, 23 qubits, 3-frequency pattern
Supplementary Figure 4. Distance-5 square lattice, 49 qubits, 5-frequency pattern

Supplementary Figure 5. Distance-5 heavy-square lattice, 73 qubits, 3-frequency pattern
Supplementary Figure 6. Distance-5 heavy-hexagon lattice, 65 qubits, 3-frequency pattern
Supplementary Figure 7. Distance-7 square lattice, 97 qubits, 5-frequency pattern
Supplementary Figure 8. Distance-7 heavy-square lattice, 145 qubits, 3-frequency pattern
Supplementary Figure 9. Distance-7 heavy-hexagon lattice, 127 qubits, 3-frequency pattern