

Boundaries of Acceptable Defectiveness: Redefining Surface Code Robustness under Heterogeneous Noise

Jacob S. Palmer
Computer Science
Northwestern University
Evanston, U.S.A.
JacobPalmer2026@u.northwestern.edu
Corresponding Author

Kaitlin N. Smith
Computer Science
Northwestern University
Evanston, U.S.A.
kns@northwestern.edu

Abstract—A variety of past research on superconducting qubits shows that these devices exhibit considerable variation and thus cannot be accurately depicted by a uniform noise model. To combat this often unrealistic picture of homogeneous noise in quantum processors during runtime, our work aims to define the boundaries of acceptable defectiveness (BADs), or the upper boundary of a qubit’s physical error, past which this defective qubit entirely degrades the logical computation and should be considered faulty and removed from the surface code mapping. Here, we present a simulation framework based on the stabilizer simulation package STIM, that allows for rapid experimentation of quantum error correction (QEC) performance under any arbitrary and unique noise model. Using this tool, QEC circuits using rotated surface codes were generated, sampled, and analyzed from distances 3 to 17, with various defective error rates and outlier defect locations (e.g., a measurement / data qubit with high error in the center of the lattice). In addition, we simulated heterogeneous noise models using the same parameters to test how increasingly deviated distributions of physical errors scale across code distances under realistic, heterogeneous noise models that are informed by current superconducting hardware. The results suggest that there are, in fact, boundaries of acceptable defectiveness in which a defective qubit, with a physical error rate ≤ 0.75 , can be left in the lattice with negligible impact on logical error rate given sufficient code distances and proper placement in the lattice. Additionally, when modeling noise as a uniform distribution, the logical error rate shows minimal impact with increasing deviation, with $\sigma \leq \mu$. As a result, we propose that defectiveness of both individual qubits and the overall uniformity of lattice fidelity should not be viewed as all or nothing, but instead as a spectrum. Our research demonstrates how heterogeneity relates to the logical error rate and, through the framework provided, facilitates the development of preliminary goals and metrics for hardware designers to meet to achieve target logical performance with imperfect or non-uniform qubit qualities.

I. INTRODUCTION

Quantum error correction (QEC) enables fault-tolerant quantum computing by utilizing multiple physical qubits to redundantly encode a single abstract logical qubit, thereby making the information more resilient to errors such as decoherence or gate noise [9]. A leading method for QEC is the surface code (SC), in which physical qubits are arranged in a 2D lattice, and operations are performed between neighboring

qubits [10]. The SC is favorable for near-term QEC demonstrations because of its high noise tolerance, and defective qubits in the 2D lattice can be managed by deforming the lattice around the defect at the cost of additional physical qubit overhead [2, 25]. Existing SC research explores (1) how logical error rate (LER) is affected by physical error rate and (2) how physical overheads are impacted by defects. However, analysis often considers physical qubit performance, specifically in regard to superconducting qubit architectures, in a rigid, homogeneous manner wherein physical qubits are either characterized as either uniform in their error rates or completely unusable. As a result, critical insights into the impact and upper bounds of the usability of qubits that range in their defectiveness are currently sparse. A better understanding of how SC tolerates heterogeneous error landscapes is necessary to improve the mapping of QEC to realistic quantum systems.

The primary goal of this work is to better understand how the SC responds when the physical error rate is not held constant across a lattice, such as with normally distributed noise or individual defects, and to provide a framework for analyzing performance in the presence of heterogeneity. Using a QEC simulation framework that supports distributed error models and is based on STIM, we investigate the impacts of defective outlier qubits and normally distributed noise on overall LER across code distances ranging from 3 to 17. Our contributions are as follows:

- For the first time, we define the concept of boundaries of acceptable defectiveness (BADs) as a way to examine and quantify the performance and thresholds for imperfect noise models. The community lacks quantitative guidance with respect to the question of “how defective is too defective” for qubit error distance from the mean. Our research shows the utility of treating the physical error rate as a distribution when selecting the best SC solution to achieve an algorithm’s target logical error. Further, our simulation results show (1) the existence of BADs in both homogeneous and heterogeneous noise models and (2)

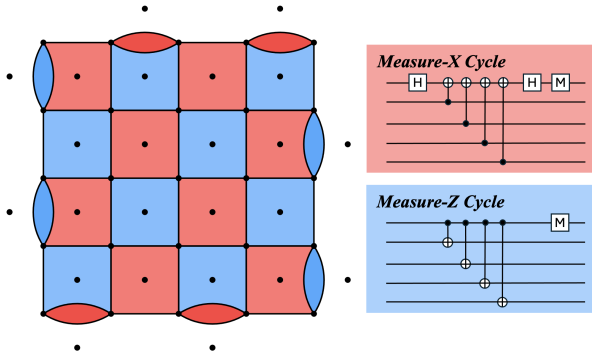


Fig. 1. An example visualization of a $d=5$ rotated surface code lattice (left) where data qubits are found on the grid boundaries, and X (Z) measurement/stabilizer qubits are found within the red (blue) squares of the grid. The sequence of operations (right) for a single Measure-X and Measure-Z surface code cycle are also illustrated.

these BADs are not static and differ greatly depending on the assumed noise model (i.e., the quality of qubits within a lattice) and targets for LER. Real quantum hardware exhibits vast variation in noise, and thus, in the pursuit of grounding our simulated models in reality, identifying the BADs allows a more accurate understanding of how performance scales given any arbitrary noise model.

- We developed a novel circuit generator framework using STIM for rapid experimentation of SC performance under any arbitrary and unique noise model. Our code, which will be open source, creates quantum circuits with granular control over each qubit’s fidelity, enabling comparisons across any configuration of noise models. It also provides a novel generation of heterogeneous noise models based on deviations from a mean physical error rate. This tool has a high potential to find utility in QEC resource estimation and design space exploration based on superconducting qubits. This code and demo notebooks used to generate the figures within this paper can be found here on GitHub.
- We model the performance of the SC under heterogeneous noise models, providing additional insights into the relationship between logical performance on imperfect quantum devices and increasingly deviant physical noise distributions.

II. BACKGROUND

A. The Surface Code

The surface code [9] (SC) is one of the most promising QEC codes for physical realization, especially with superconducting qubits, due to its implementation using a 2D nearest-neighbor qubit layout, and its high tolerance to noise [9]. Recent demonstrations on superconducting circuits have allowed logical qubits to be demonstrated and LERs to drop below threshold [22]. SCs can implement a universal set of gates, create logical entanglement with lattice surgery [17], and generate magic resource states with magic state distillation [18, 13]. In this work, we focus on the rotated SC (i.e., any further mention

of SC is assumed to be this), requiring $2d^2 - 1$ physical qubits per logical qubit. Our goal is to study the SC’s ability to store logical qubit states over time while being subject to various noise models.

An illustration of a $d=5$ SC patch can be seen on the left side of Fig. 1. In this picture, d^2 data qubits are found on the edges of the grid while $d^2 - 1$ measurement qubits are found within the red and blue squares representing Measure-X and Measure-Z stabilizers, respectively. The right side of Fig. 1 shows the sequence of operations performed during the SC cycle that allows the data qubits (bottom four lines) to interact with the measurement qubit (top line). Over many cycles, the history of detection events gives insight into the errors happening during runtime. These syndromes are then used by the decoding algorithm, minimum-weight perfect matching (MWPM) [7] in our case, to locate the error within the logical qubit so that it can be corrected. A logical error is introduced into the SC when at least $d/2$ errors occur along a non-trivial path.

B. Previous Work

The heterogeneous nature of superconducting qubit physical qubit error has been well-documented [20, 27, 5]. Prior work has proposed novel architectures, such as those based on chiplets [24], to promote device scaling with more uniform metrics, but it is unlikely that quantum systems comprised of imperfectly fabricated processors will ever demonstrate perfectly heterogeneous noise. This raises many challenges in terms of how to design QEC strategies that best meet the needs of algorithms while managing realistic constraints of hardware.

Previous work with the SC has made progress in combating the reality of defective qubits by leveraging “super stabilizers” and boundary deformations in the SC patch. These techniques work by routing the stabilizer measurement circuits so that they avoid faulty components [16, 6, 29]. Current deformation methodologies incur a tradeoff: deform the lattice around the defect, thus removing the defect’s negative impact on the LER *but* incur additional resource overhead in the process. This increase in overhead is typically due to a higher physical qubit requirement, but some deformation techniques also incur increased complexity during compilation or runtime through specialized steps to facilitate computation within the deformed lattice.

Most work in SC defect management assumes a rigid categorization of qubit quality in the sense that a qubit is either operating at the mean physical error rate or it is completely unusable and should be dropped out of the logical qubit mapping. However, more lenient bounds of acceptable defectiveness could exist, especially in the case where the distribution of physical qubit error rates is characterized by a standard deviation. For example, the work in [19] reports that the current variation in coherence time seen on existing superconducting qubits leads to hindered LER with the SC. This presents the question of how defects should be defined, and which qubits with higher error rates can be left in the lattice in its current state, given the QEC designer’s choice of

code distances. These sorts of estimates remain to be fully explored; as such, this work serves as a starting point for scaffolding out potential defects, their impacts, and the design choices necessary to minimize detrimental effects. To narrow our scope in this initial exploration of defining BADs, we focus on distributions associated physical qubit error (Section II).

III. HOMOGENEOUS VS. HETEROGENEOUS ASSUMPTIONS OF NOISE ON SURFACE CODE LATTICES

In general, if physical operations on the SC are performed below a sufficiently low critical noise threshold then logical error rates are exponentially suppressed with increasing code distances [10]. This scaling can be expressed by the relation:

$$\varepsilon \propto \left(\frac{p}{p_{thr}} \right)^{\frac{d+1}{2}}$$

where ε is LER, p_{thr} is threshold for physical noise tolerance, and d is the distance of our code [1]. Prior work has found $p_{thr} = 0.0057$ [10]. In this general form relation, p represents the physical error rate of all $2d^2 - 1$ physical qubits that encode our singular logical qubit in a given lattice. Thus, this relation assumes a uniform, homogeneous physical noise exhibited across the lattice. However, results from real quantum devices have shown that superconducting quantum devices do not exhibit a uniform noise over the entire processor but instead vary greatly in qubit fidelity.

IV. SIMULATION FRAMEWORK OVERVIEW

The simulation and analysis framework we devised generates and samples the performance of parameterized repetition code circuits on rotated SCs using Python and STIM [12]. The generated circuits are mapped to a 2D lattice with $n = 2d^2 - 1$ qubits scaled from a provided distance d . STIM is specifically designed for rapid design and simulation of quantum stabilizer circuits, and our code was built using the STIM Python package [11]. While other popular quantum circuit simulators exist [14, 8] STIM was selected because (1) the ability to adaptively create entire stabilizer circuits from string manipulation and (2) it offers very fast runtime allowing for sampling the logical error count from millions of shots in minutes. For example, in our distributed heterogeneous noise simulations in Section V-C, the simulations (Fig. 11 and Fig. 12) consist of nearly $4 * 10^7$ total shots over the eight distances sampled for each, with each point representing a generated circuit with a unique noise distribution with $5 * 10^4$ shots. Thus, it was important that the simulation tool be reasonably fast in order to test a wide variety of different input parameters when developing our code and analyzing our results.

The circuits are generated with a variety of input parameters to facilitate testing variations in both individual qubit noise (e.g., specific defect location and defect physical error rates) and overall lattice noise distributions (e.g., uniform or heterogeneous). Our code works in three parts (1) generating the parametrized STIM string representation of the measurement circuit, (2) parsing the simulation data and performance results provided by STIM into usable DataFrames, and (3) directly

plotting simulation data into easy-to-understand plots that allow for immediate comparison of performance across any number of code distances. Each of these circuits is built using either a uniform (homogeneous) or normally distributed / defect-injected (heterogeneous) noise distribution. For circuits built with a homogeneous noise model, all qubits within the lattice will be assigned the same physical error rate p value is given as a parameter and used as the error rate for all qubits in the lattice. For this initial foray into heterogeneous noise models with our simulator, we are assuming a simplified notion of noise. Specifically, our simulation primarily uses depolarizing channel (D) noise is a standard quantum noise model in which a qubit becomes depolarized with a given probability p [28]. In most conversations, p is referred to as the level of noise of a physical qubit, including more types than just depolarization noise [10]. However, in this simplified noise model, we rely on noise occurring as a result of Pauli-X (bit-flips) errors on state preparation and measurement, and depolarization error on gate operations to introduce noise. In reality, quantum noise can be introduced by a wide variety of sources, such as interactions with the environment, uncontrolled interactions with other qubits, imperfect operations, or leakage [23]. Additionally, noise can be introduced permanently on specific qubits as a result of manufacturing defects on the superconducting devices [3]. Implementing a more complex noise space is an area that could be greatly extended upon and is discussed further in our future work section.

A. Building Circuits as STIM Strings

For a given distance d , the generators first assign each of the $n = 2d^2 - 1$ qubits a coordinate pair (X, Y) within a grid representing the 2D lattice. The set of values for both the x-coordinate and the y-coordinate starts at 0.5, increments by 0.5, and goes up to $(d + 0.5)$.

In SCs, each qubit is either a measurement (stabilizer) qubit or a data qubit [10]. Within our generators, data qubits are always located at integer coordinates (e.g., (3, 5)) and measurement qubits are always located at non-integer coordinates ± 0.5 in x or y direction from the data qubits (e.g., (.5, .5) or (1.5, 2.5)). Within our setup, data qubit coordinates are always integer values, while the measure qubits will always be ± 0.5 from their respective data qubits.

Because STIM strings represent the stabilizer circuits in a manner similar to quantum circuit diagrams [21], each of the n qubits is referred to by a wire index from 0 to $n - 1$. Our generators map coordinate pairs to the appropriate wire index within the STIM string, allowing for the generators to accept coordinates as inputs for specific noise models and then appropriately translate between the index and coordinate references for any given qubit. Additionally, the generators will map special location keywords to coordinates on the lattice, as seen in Fig. 2. Since the distance directly determines the grid space, these special keywords facilitate placing defects at the same relative location in the lattice, simplifying cross-distance performance analysis.

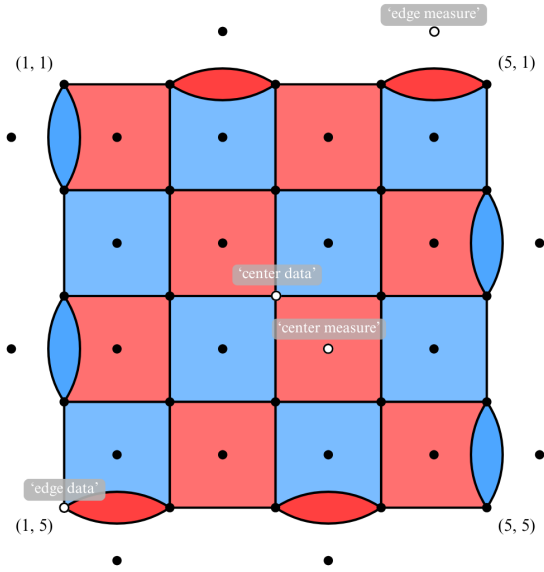


Fig. 2. For the simulations and selection of defect locations, the generators optionally allow for specific qubit locations to be designated in order to maintain similarity across various code distances. The qubit locations are marked by the special case strings, ‘center data’, ‘center measure’, ‘edge data’, ‘edge measure’. These special case strings are then converted by calculating their respective coordinate given any lattice of distance d .

There are two primary generator types, which assume what type of noise distribution the lattice displays:

1) *Homogeneous Distributions*: The homogeneous generator takes as input a physical error rate p that will be set as the standard physical error rate for all qubits in the lattice.

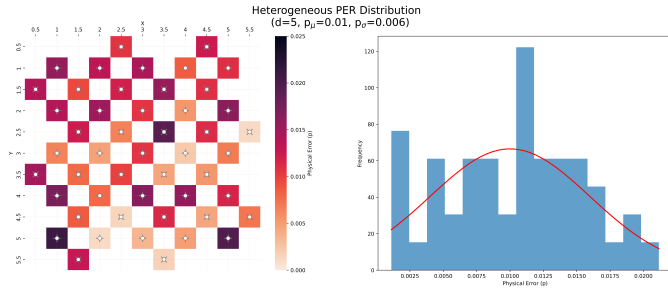


Fig. 3. A heatmap and distribution representing a randomly distributed heterogeneous noise distribution on a $d=5$ code. The distribution on the left is shown to demonstrate that, although the truncated normal distribution results in slight differences, the samples still form a distribution around p_μ and p_σ . For our experiments in Case 3 (Section V-C) and Case 4 (Section V-D), we utilize an α parameter and the actual sampled distribution mean to ensure that any shifts from the input mean are reflected (i.e., when $p_\sigma > p_\mu$ at sufficiently small p_μ values). In this figure, with a $p_\sigma = 0.006$ and $p_\mu = 0.01$, the α parameter would be $\alpha = 0.6$, given that $p_\sigma = \alpha * p_\mu$. Within all heatmaps shown, data qubits are denoted by + and measurement qubits are denoted with X.

2) *Heterogeneous Distributions*: As discussed earlier, a primary motivation for this research is to provide methods for exploring and analyzing simulation models that more accurately resemble real superconducting devices [19, 1]. To achieve this, the heterogeneous generator creates a circuit

that can be used to model the performance of a non-perfect lattice with individual qubits’ physical error rate spanning any arbitrary distribution of noise values, rather than a static p .

In this first iteration of our code, the heterogeneous generator defaults to a truncated Gaussian distribution. For each qubit in a lattice, it’s physical error rate is randomly sampled from a truncated normal distribution created based on two input parameters: a mean physical error rate, p_μ , and either standard deviation, p_σ , or a scalar, α , that determines the p_σ as such:

$$p_{\sigma} = \alpha * p_{\mu}$$

Each time a circuit is generated with a given input p_μ and p_σ , a new random distribution will be sampled from, creating unique distributions of heterogeneous noise between circuit generations (even with identical parameters), as seen in Fig. 3 for a $d=5$ code. This unique distribution for each generated circuit is, in part, a modest reflection of how the noise exhibited by superconducting qubits can greatly fluctuate during or between runtime, but introduces some variability in our simulation’s performance results that leads to some trends being non-monotonic, something that is discussed in Section V later on.

It is important to note that noise distribution on current superconducting devices does not necessarily follow a Gaussian model [26]. Thus, as a preliminary method for analyzing the impact of increasingly distributed noise in a lattice, analyzing how the SC reacts under a Gaussian distribution is an ideal starting point, and our parameterized modeling allows for other distributions to be explored.

B. Defective Qubits

Both homogeneous and heterogeneous circuit generators accept a number of parameters, including distance, rounds, maximum shots. More importantly, the generator methods allow individual defect(s) to be arbitrarily created at any part in the lattice by providing a qubits coordinate (x, y) pair within the lattice (as seen in Fig. 2) and a physical error rate for each defect, p_{def} . In practice, the generators could be used to manually set the physical error rate for each qubit in the lattice simultaneously, providing complete granular control over the simulated noise.

C. Operational Noise

In the generators, noise is introduced right before circuit elements (e.g., gates and measurements) as a noise operator with the general format for 1q operations:

<1q noise type> <q1 index> <pval_q1>

and for 2q operations:

<2q noise type> <q1 index> <q2 index> <p_2q>

Each qubit is assigned an individual error rate, p_i , where i is the index of the qubit. p_i determines the error rate for all operation involving this qubit. These operations include all multi-qubit operations (“DEPOLARIZE2”), single-qubit operations (“DEPOLARIZE1”), and measurements (“X-ERROR”)

on the qubit. The p_i is scaled based on single or two-qubit operations. Generally, multi-qubit operations are more prone to errors than operations involving only a single qubit. Thus, it should be reflected that CX operations involving the defective qubits would exhibit a higher error rate than their single-qubit operations. Thus, for two-qubit depolarization noise associated with CX operations, the error rate is set as a mean of the p_i corresponding to the two qubits involved in the operation. Since two-qubit operations are more prone to error than single-qubit operations [1, 19], two-qubit noise has an additional scaling factor, s , on the $q_i + q_j$ mean, in the form:

$$p_{CX} = \left(\frac{p_{q1} + p_{q2}}{2} \right) * s$$

This scaling factor is set as $s=1.2$, representing a 20% increase over single qubit operations. For single-qubit depolarization and Pauli-X, the value for each operation on a qubit q_i is set directly to its corresponding p_i value.

V. CASE STUDIES AND ANALYSIS

To demonstrate the performance disparity between heterogeneous and homogeneous noise assumptions, we examined how LER scales across code distances 3-17 when assuming a homogeneous vs. heterogeneous noise model. In each case, the logical error rate per round, ε , is the metric used to evaluate performance. Every measurement circuit generated for these experiments contains 3 rounds, and the logical error rate is considered as the average number of errors that occur per round in the measurement circuit:

$$\varepsilon_{\text{round}} = \left(\frac{\text{errors}}{\text{shots} * \text{rounds}} \right)$$

To facilitate comparison and benchmarking across the four cases, we establish a target logical error rate, $\varepsilon_{\text{target}}$, of 0.005. The $\varepsilon_{\text{target}}$ to use as a baseline depends primarily on the fidelity required to support a specific algorithm or application on a particular quantum device. However, for our purposes, we assume that a $\varepsilon_{\text{target}} = 0.005$ is considered to be sufficient fidelity for these memory experiments. For each case, a heatmap was generated to visually represent an approximate noise distribution on a $d=5$ rotated SC lattice.

In our experiments, four primary cases were selected under increasingly complex models of noise: (1) a completely uniform homogeneous lattice, (2) a uniform homogeneous lattice with a singular defect in the center data qubit, (3) a heterogeneous lattice, and (4) a heterogeneous lattice with a singular defect in the center data qubit. In each case, the simulations increase in complexity of the noise models, allowing differences in performance scaling with distance to become evident.

In these experiments, we once again return to the idea of the “boundary of acceptable defectiveness”, or BADs. For each code distance, the boundary of acceptable defectiveness (BAD) defines the maximum physical noise tolerance after which the LER surpasses the $\varepsilon_{\text{target}}$. Within the figures for each of our case experiments, the BAD for each code distance is

precisely at the intersection point between the target $\varepsilon_{\text{target}}$ and the distance’s performance curve. Once the physical noise level has surpassed the BAD, we can consider the lattice to be effectively too noisy and the computation unusable. Importantly, these boundaries of acceptable defectiveness for a given distance are not held constant across different models of noise. Thus, using these boundaries provides a meaningful way to quantitatively compare performance across various noise models.

A. Case 1: Homogeneous Noise

For the first scenario, Case 1 covers the assumption of uniform noise in which all physical qubits within a lattice unilaterally express the same p noise. A heatmap for this uniform homogeneous noise model can be seen in Fig. 4. The performance results for Case 1 are shown in Fig. 5 for $d=[3, 5, 7, 9, 11, 13, 15, 17]$. In past literature, this most closely resem-

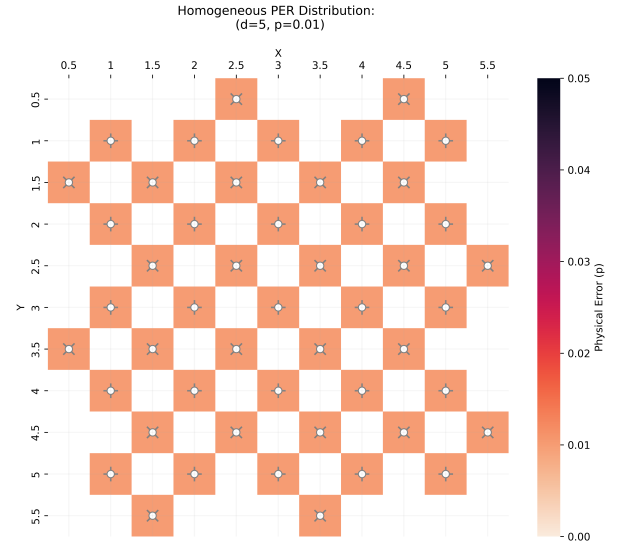


Fig. 4. (Case Study 1) Heatmap for a homogeneous noise model on a $d=5$ code with no defects or deviations. In this heatmap, data qubits are denoted by + and measurement qubits are denoted with X.

bles the ideal and common notion of how noise is represented in SC. As shown in Fig. 5, all distances (with the exception of $d=3$) cross the $\varepsilon_{\text{target}}$. For the highest distance sampled, $d=17$, the BAD is at $p \approx 0.006$. Any $d=17$ lattice with a $p \leq 0.006$ would, in this theoretical model where $\varepsilon_{\text{target}} = 0.005$, be sufficient for computation. However, if our goal is to minimize the code distance and thus our physical qubit overhead, we would select the minimum viable code distance for which our device’s expressed p is below the BAD for that distance. In this uniform homogeneous model, the $d=5$ code has a BAD at approximately $p \approx 0.0029$. Thus, if deciding an optimal code distance for a physical device capable of performing computation with physical noise, $p \approx 0.0029$, we should select the $d=5$ code. While achieving $p = 0.0029$ on a real superconducting device might be out of reach with current hardware [15], analyzing the BADs for a given code distance

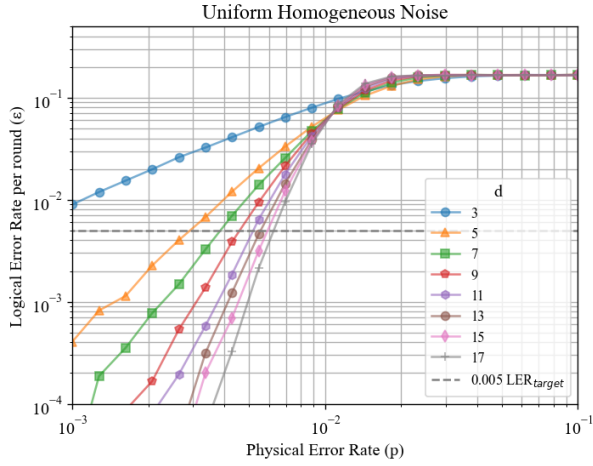


Fig. 5. (Case Study 1) Performance results for a completely uniform, homogeneous noise model across distances $d=[3, 5, 7, 9, 11, 13, 15, 17]$. Case 1 represents the most ideal noise model assumptions when considering SCs in that all qubits within the lattice have an identical quality of p and no qubit deviates from this value. In Cases 2 and 3, the blue line in each performance result figure is the baseline for each distance when no deviation is expected. As such, each line in this figure is functionally equivalent to the blue baseline in the Case 2 and Case 3 performance results for the corresponding distance. Because errors are probabilistically applied, small differences in these baselines can occur across simulations.

with a realistic physical noise model allows for additional optimization considerations. In Case 1, performance requirements are met at all but the smallest distance, $d=3$. While achieving uniform noise across all physical qubits results in the highest overall fidelity and the most viable code distances, noise data from superconducting devices exhibit significant variations in noise within a given chip. Thus, this uniform model in Case 1, at least for superconducting devices, is naive and an unrealistic basis for accurately modeling SC robustness [15, 19].

B. Case 2: Homogeneous Noise with Outlier Defects

To test individual qubit impact within a homogeneous noise model, two separate circuits were generated with a defect located at the center data qubit and near-center measure qubits. In this second scenario, we examine the impact of a singular defect on the SC's performance under an otherwise uniform noise model. A heatmap for the homogeneous noise model with an outlier defect center data qubit can be seen in Fig. 6. The performance results for Case 2 are shown in Fig. 7 for $d=[3, 5, 7, 9]$ and Fig. 8 for $d=[11, 13, 15, 17]$.

In this experiment, all lattices across the sampled distances have a defect located at the data qubit in the center of the lattice. We simulated the performance with the defect-center data qubit p_{def} value ranging from 0.05 to 0.75. At the upper end of the p_{def} values sampled, the individual qubit is essentially completely unreliable, with roughly 75% of operations on this qubit resulting in an erroneous computation. The results of this experiment (Fig. 7, Fig. 8, and Fig. 9) show an incredibly strong resilience to the individual defect. Similar to that of the unilateral homogeneous model in Case 1 (Fig. 5), all but the smallest code distance $d=3$ (Fig. 7)

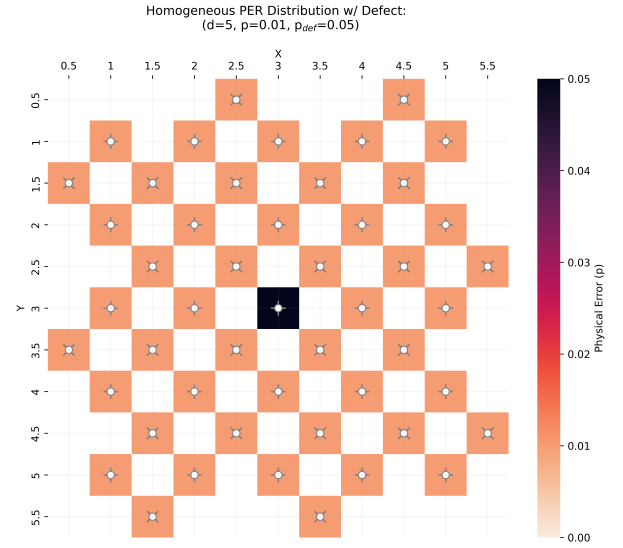


Fig. 6. (Case Study 2) Heatmap for a homogeneous noise model on a $d=5$ code with a defect on the center data qubit. The p_{def} shown in this heatmap is 0.05 to show this distribution, but significantly higher values p_{def} are also used in the Case 2 experiments as well. In this heatmap, data qubits are denoted by + and measurement qubits are denoted with X.

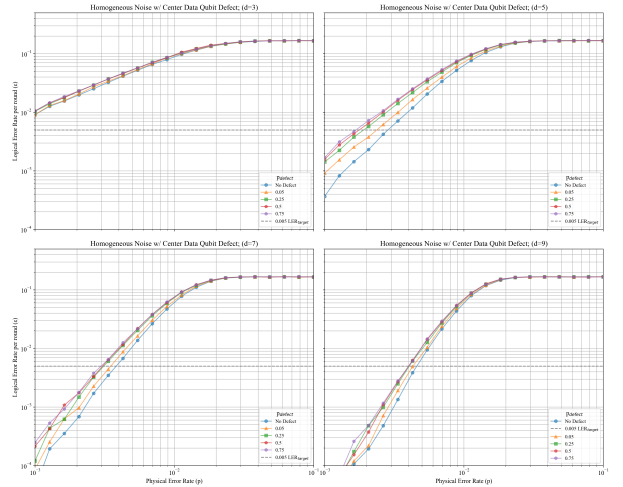


Fig. 7. (Case Study 2) Performance of homogeneous noise model with outlier center data defect over $d=3$ through $d=9$.

were able to meet the ϵ_{target} . In the worst case with $p_{def}=0.75$, the BAD for the $d=17$ code (Fig. 8) is $p \approx 0.0058$. In comparing this to the $d=17$ performance in Case 1 (Fig. 5), the maximally defective qubit only reduced our BAD by 3%. For the minimum viable code distance in Case 2, $d=5$ (Fig. 7), the BAD is located at $p \approx 0.0017$.

Specifically, for $d=5$ in Case 2 with the BAD for $d=5$ ($p \approx 0.0029$) in Case 1 (Fig. 5), the presence of the defect with $p_{def}=0.75$ resulted in a larger reduction to our acceptable boundary by about 50%. As with physical noise tolerance, the results here suggest that under a homogeneous noise model, an increase in code distance also directly results in a significant increase in the resilience to individual defects with high levels

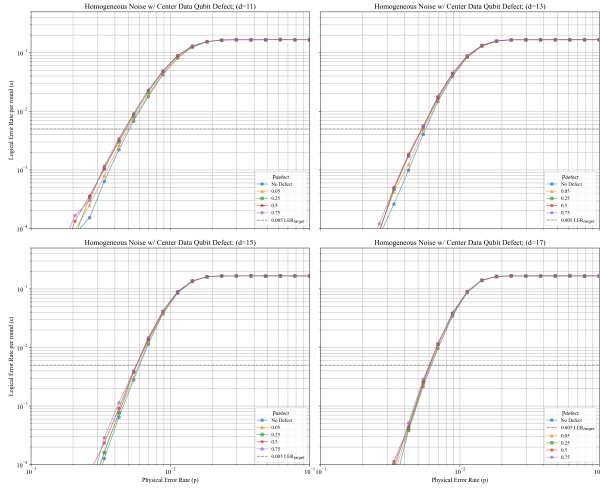


Fig. 8. (Case Study 2) Performance of homogeneous noise model with outlier center data defect $d=11$ through $d=17$.

of noise, which can be clearly seen in Fig. 9.

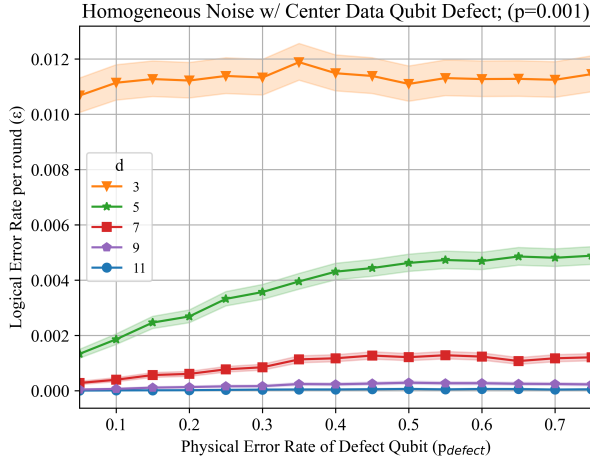


Fig. 9. (Case Study 2) Logical error rate vs. defective physical qubit error rate for code distances $d=[3, 5, 7, 9, 11]$. To further examine how performance is affected across code distances, this figure shows a similar homogeneous noise model to that in Case 2, except that the noise for all other qubits is kept constant at $p = 0.001$. As we increase our code distance, the lattice becomes increasingly resilient to the noise introduced by the defective data qubit, as expected. At distances of 9 or above, the performance is essentially unaffected by the presence of the defective qubit. Thus, this further suggests that the impact of highly defective qubits becomes negligible at a sufficiently large code distance.

C. Case 3: Heterogeneous Noise

In the third scenario, we expand our noise model to a heterogeneous normal distribution of physical errors and examine the impact on the logical performance. A heatmap demonstrating a potential configuration for a $d=5$ heterogeneous noise model can be seen in Fig. 10. The performance results for Case 3 are shown in Fig. 11 for $d=[3, 5, 7, 9]$ and Fig. 12 for $d=[11, 13, 15, 17]$. For this experiment, we simulate heterogeneity by assigning each qubit a p_r randomly sampled from a truncated

normal distribution with an input mean parameter, p_μ , and an input standard deviation, p_σ . To examine the impact of this Gaussian model of heterogeneous noise on logical performance, we simulated 4 α values ranging from 0.01 to 0.9, corresponding to normal distributions with a deviation of 1% to 90% of the p_μ . For each α value examined, we simulated their performance with 20 different p_μ values evenly distributed between 10^{-3} and 10^{-1} on a logarithmic scale. To prevent negative probabilities from being sampled, a truncated distribution is used. When a distribution is truncated, the input mean used to create the distribution, p_μ , can differ from the *actual* mean, $p_{\mu-out}$, of the values sampled from the distribution. Thus, to accurately reflect any shifts between input and actual mean, the $p_{\mu-out}$ is used as the x-axis in our performance plots. For clarity, if p_{q_i} is the randomly sampled physical noise for qubit q_i and $n = 2d^2 - 1$, then the average mean of the sampled distribution is as follows:

$$p_{\mu-out} = \frac{\sum_{i=0}^{n-1} p_{q_i}}{n}$$

Similar to Case 1 (Fig. 5) and 2 (Fig. 7 and Fig. 8), the performance of the heterogeneous noise is benchmarked against a uniform homogeneous distribution, which is equivalent to when the $\alpha = p_\sigma = 0.0$. The non-monotonic behavior of the plotted data in Fig. 11 and Fig. 12 is a result of this randomly sampled distribution, in which some circuits have a higher distribution of physical noise placed on more detrimental locations, such as higher noise on measurement qubits or edge qubits of which have a greater impact on LER [9]. However, the noise and coherence properties of individual superconducting qubits can change frequently [15]. Thus, purely monotonic performance curves when utilizing superconducting qubits is

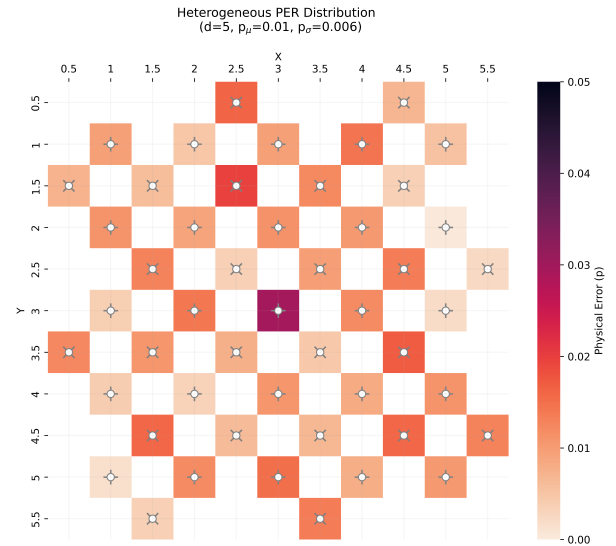


Fig. 10. (Case Study 3) Heatmap for a heterogeneous noise model on a $d=5$ code. Data qubits are denoted by + and measurement qubits are denoted with X.

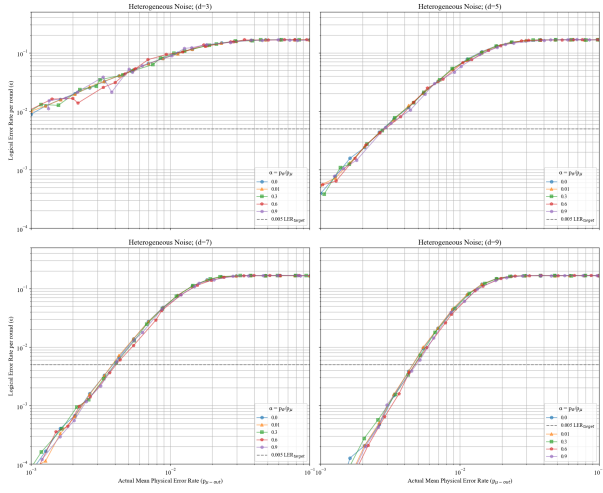


Fig. 11. (Case Study 3) Performance of heterogeneous noise model for distances 3 through 9. The blue line ($p_\sigma = 0$) represents a unilateral homogeneous noise model.

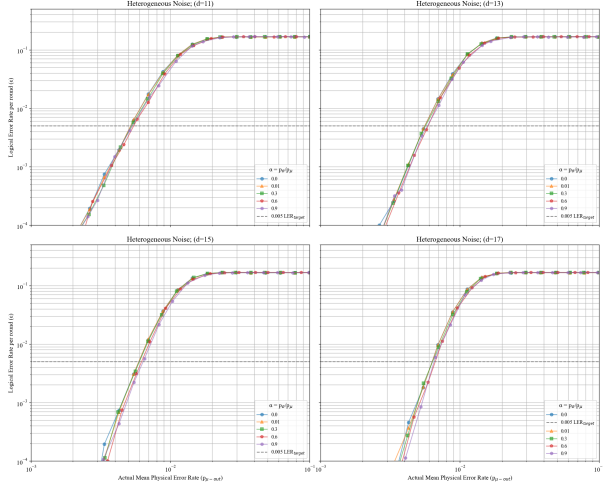


Fig. 12. (Case Study 3) Performance of heterogeneous noise model for distances 11 through 17. The blue line ($\alpha = 0.0$) represents a unilateral homogeneous noise model covered in Case 1.

not a guarantee.

This is most notable with $d=3$ in Fig. 11 in which the ε_{round} appears to fluctuate more drastically as α increases. In a $d=3$, the proportion of 2-weight stabilizers on the edge of a lattice to 4-weight stabilizers in the bulk of the lattice is higher. With a higher proportion of low-weight stabilizers in $d=3$, the increased deviation leads to greater fluctuations, as shown for $\alpha = 0.9$ and $\alpha = 0.6$ in Fig. 11.

D. Case 4: Heterogeneous Noise with Outlier Defects

In the fourth scenario, we examine performance changes under a heterogeneous noise model in the presence of a center data-defect qubit. A heatmap demonstrating a potential configuration for a $d=5$ heterogeneous noise model with a center defect can be seen in Fig. 13. The performance results

for Case 3 are shown in Fig. 14 for $d=[3, 5, 7, 9]$ and Fig. 15 for $d=[11, 13, 15, 17]$.

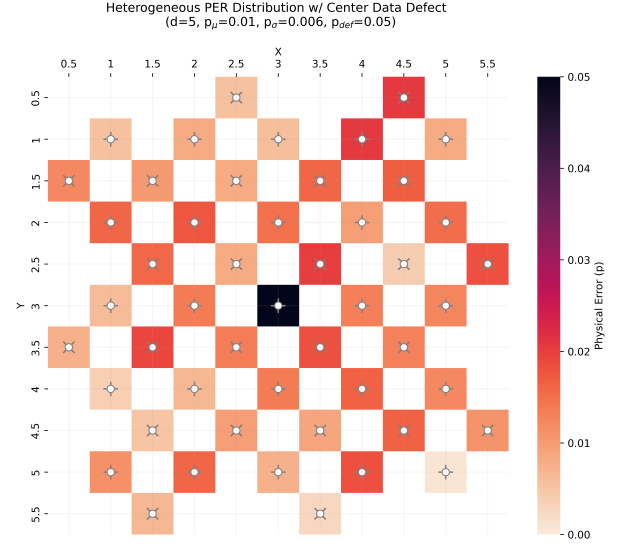


Fig. 13. (Case Study 4) Heatmap for a heterogeneous noise model on a $d=5$ code with an outlier center data qubit. In this heatmap, data qubits are denoted by + and measurement qubits are denoted with X.

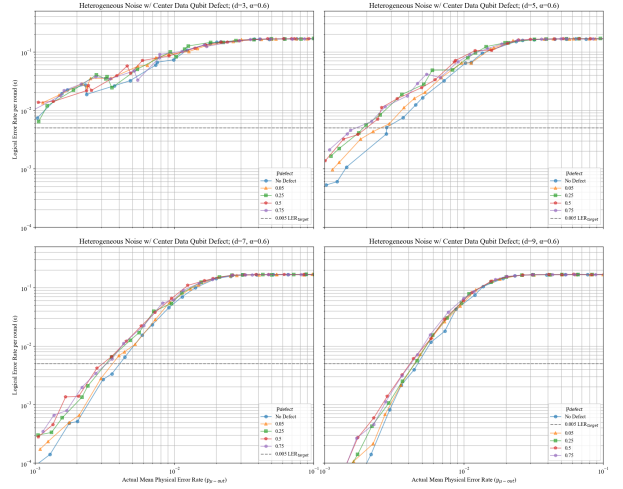


Fig. 14. (Case Study 4) Performance of heterogeneous noise model with center data defect for distances 3 through 9. In Case 4, we simulated the performance of heterogeneous noise resembling a Gaussian distribution with a singular defect at the center data qubit with a p_{def} in the range of 0.05 to 0.75. In all simulations in Case 4, the α parameter, which determines the deviation as a percentage of the input mean, is held constant at $\alpha = 0.6$. The blue line represents a baseline heterogeneous noise model with no defects present, which is the same as the $\alpha = 0.6$ simulations in Case 3.

The location of the defect (center data qubit) and p_{def} range (from 0.05 to 0.75) is identical to the setup in Case 2, with the exception that the noise of all other qubits is normally distributed rather than constant. As with the setup of the experiments in Case 3, we use the actual mean of the physical noise that was randomly sampled, $p_{\mu-out}$, as the horizontal axis. Additionally, like in Case 2, we want to assess how

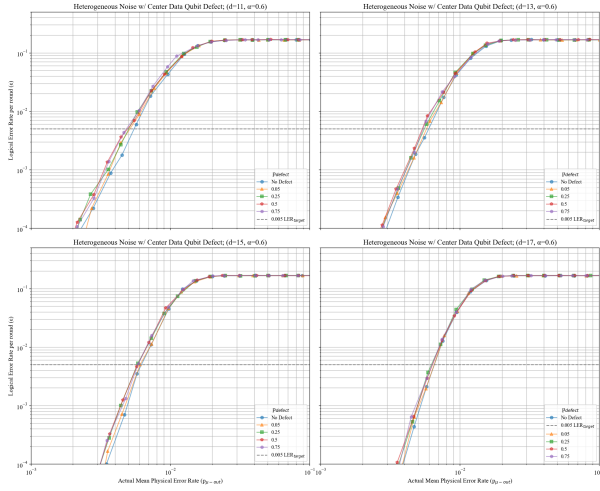


Fig. 15. (Case Study 4) Performance of heterogeneous noise model with center data defect for distances 11 through 17. The blue line represents a baseline heterogeneous noise model with no defects present.

increasing the noise of all other qubits in the presence of a defect impacts the LER. Thus, our $p_{\mu-out}$ does not include the p_{def} value as it is not sampled from the distribution itself. For clarity, in Case 4, the $p_{\mu-out}$ is as follows:

$$p_{\mu-out} = \frac{\sum_{i=0}^{n-1} p_{q_i} - p_{def}}{n - 1}$$

For this experiment, we selected a singular α used in Case 3, $\alpha = .6$, and used the same p_{def} values that were selected in Case 2. Interestingly, the performance trends shown in Case 4 (Fig. 14 and Fig. 15) follow a pattern similar to that observed in Case 2 (Fig. 7 and Fig. 8) in which higher distances show a strong resilience to increasing magnitudes of p_{def} .

VI. DISCUSSION AND FUTURE WORK

Our experiments aimed to highlight how heterogeneity affects performance on the SC, and the results indicate that the presumed model of noise greatly impacts the purported performance of the SC. We proposed in this work that, in order to accurately model SC on superconducting qubits, accepting some level of heterogeneity is necessary due to the inherent non-uniform noise exhibited by superconducting qubits. The aim of our work is thus to provide a preliminary accounting of how the performance of heterogeneous distributions directly affects the LER. As shown in our experiments, the presence of a single defective qubit with an error rate as high as 75% significantly affects the LER only at small code distances and becomes increasingly negligible as the code distance increases.

In regards to the impacts of individual defects, the performance data for both a singular defect in the presence of heterogeneous and homogeneous noise models show that the defective qubit's impact is only non-negligible at distances at or closely above the minimum viable code distance (the distance at which the target LER is reachable).

In the performance experiments for Case 2 (Fig. 7 and Fig. 8), the minimum viable code distance occurs at $d=5$. At $d=3$, the target LER is not reachable even when there are no defects present. Although the performance at $d=3$ is insufficient, the performance across the increasing p_{def} noise is nearly identical, indicating that this defect has little impact on the LER. At $d=5$, the trend lines suddenly deviate, and the impact of higher p_{def} noise on performance is noticeable. At $d=5$ through $d=9$, this performance deviation between p_{def} noise begins to converge. At $d=11$ and above, the impact of an increasingly defective qubit is nearly negligible again. What we can extrapolate from this is that there exists a sort of “Goldilocks zone” in which the detrimental impact of increasing noise on the defective qubit is evident. Within this “Goldilocks zone,” choosing to drop out or mitigate the defective qubit with large p_{def} values could result in noticeable performance gains. At $d=11$ and above, however, the impact of this defective qubit, even at $p=0.75$, is negligible and thus implementing a mitigation technique incurs additional cost, such as an increase in physical overhead [29, 24] or circuit depth [6], but results in no performance gains. Ideally, when making design decisions, techniques that incur additional costs should only be used when there is a noticeable performance gain. In this way, our research suggests that treating defect mitigation should be viewed as an optimization problem in which we aim to only drop out qubits in which the resulting performance gain is non-negligible in light of the cost incurred by the mitigation method.

Regarding the impact of heterogeneous noise that resembles a Gaussian distribution, our results show that the LER is not greatly affected by increasing deviation. In comparing the performance curves in Cases 1 with static p and Cases 3 with normally distributed noise, we see nearly identical trends, indicating the increased deviation had a minimal impact on the LER. In Case 4, the same defect configuration (center data qubit) is tested under the heterogeneous model with a fixed $\alpha = 0.6$. In the performance experiments conducted (Fig. 14 and Fig. 15), the defect noise reflects a significant outlier from the remainder of the comparatively small normally distributed noise. These results closely resemble the trends discussed in Case 2. When comparing the insights of Case 1 vs. Case 3 and Case 2 vs. Case 4, our results indicate that while evenly distributed distributions may not impact LER, outlier defects much higher than the average physical noise do. The simulations used 3 rounds across all distances, which was done to enable rapid experimenting over many configurations and distances while we were exploring these preliminary models of noise. We performed the same experiments for each case study with rounds=distance. While higher p_{def} values became slightly more noticeable impact on LER at higher distances, the trends discussed for all cases involving both defects and impact of Gaussian noise were the same.

It appears that if noise is normally distributed, the slightly higher fidelity qubits essentially cancel out their slightly noisier counterparts. However, this balancing requires that for every bad qubit in the lattice, there is an equally good qubit,

which is an unrealistic assumption. Further, the presence of a significant defect cannot be balanced and thus has a noticeable impact on the LER. After all, it's much easier to fabricate a noisy qubit than it is to fabricate a perfect qubit. Ultimately, a heterogeneous noise model based on a Gaussian distribution does not appear to capture the complexity or impact of real heterogeneous noise, nor does it accurately reflect what we observe on real devices, further arguing that real quantum noise on superconducting devices cannot be represented as Gaussian.

Overall, the results of our case studies concur with prior work examining the impact of individual defects and normally distributed heterogeneous [4], and expand these insights across a wide range of distances, outlier defects, and distribution deviations. In moving forward, we believe that examining the impacts of heterogeneity on SC is vital as we continue to develop both physical devices and QEC techniques in order to reach fault-tolerant quantum computing. Looking ahead, we believe there are many areas in which the work here could be expanded. A particular area of improvement would be to expand the simplified noise model that we use to create our circuits, such as introducing additional scaling factors to more accurately represent the different operational error rates [19]. Secondly, the experiments focused primarily on testing individual defects on a singular data qubit, but expanding these experiments to encompass how additional configurations of defect characteristics affect performance is an area for future work. Additionally, identifying how these defects accumulate would provide further insight into when deformation techniques are absolutely necessary. The experimental results presented in our research demonstrate preliminary evidence showing the existence of BADs as well as the disparity between viewing the SC under a homogeneous model vs. a heterogeneous model. If the outcome of the simulation is to optimize performance for computation on real devices, viewing the SC in only a homogeneous regime results in vastly different design outcomes, such as picking optimal code distances to meet a target logical error rate.

VII. CONCLUSION

In this work, we explore the impact of heterogeneous noise on the SC's performance. This work is among the first to study the physical error rate of device qubits as a distribution and the first to provide a framework and preliminary toolset for rapidly evaluating how configurations of non-identical noise affect the LER in rotated SCs. Using the BADs framework, we explore the impact of both outlier qubits and normally distributed physical errors on SC LER. As a result of our work, we discovered that one bad qubit does not necessarily spoil the bunch. Furthermore, we demonstrate that if heterogeneous noise follows a Gaussian distribution, the SC LER scales only with the mean of the SC lattice, not its deviation. This ultimately indicates that modeling heterogeneous noise as a Gaussian distribution is functionally equivalent to treating it as homogeneous, further emphasizing the need to develop more complex models that capture the effects of superconducting

qubit deviation. To conclude, this work motivates the inclusion of BADs in QEC analysis based on the SC, and we hope that the software framework developed here will open the door to design space exploration, enabling shorter timelines to fault-tolerant quantum computing.

VIII. DECLARATIONS

A. Data Availability

The data and figures generated for this paper were created from our code and can be found on our GitHub repository at: <https://github.com/JacobSPalmer/quantum-ddq-toolkit/tree/icrc-25>. Additionally, all data used to generate the specific figures shown in the manuscript can be found here: <https://github.com/JacobSPalmer/quantum-ddq-toolkit/tree/icrc-25/supplementary-materials>.

B. Author Contributions

J. S. P. led code development and experiment implementations. K.N.S. advised project direction, and both authors contributed to the analysis and writing of the manuscript.

C. Funding

No funding was received in relation to this study. (Funding: Not Applicable)

D. Competing Interests

The authors of this paper declare no competing interests.

REFERENCES

- [1] Rajeev Acharya et al. "Quantum error correction below the surface code threshold". en. In: *Nature* 638.8052 (Feb. 2025). Publisher: Nature Publishing Group, pp. 920–926. ISSN: 1476-4687. DOI: 10.1038/s41586-024-08449-y. URL: <https://www.nature.com/articles/s41586-024-08449-y>.
- [2] James M Auger et al. "Fault-tolerance thresholds for the surface code with fabrication errors". In: *Physical Review A* 96.4 (2017), p. 042316.
- [3] James M. Auger et al. "Fault-tolerance thresholds for the surface code with fabrication errors". In: *Physical Review A* 96.4 (Oct. 2017). arXiv:1706.04912 [quant-ph], p. 042316. ISSN: 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.96.042316. URL: <http://arxiv.org/abs/1706.04912>.
- [4] Malcolm S. Carroll, James R. Wootton, and Andrew W. Cross. *Subsystem Surface and Compass Code Sensitivities to Non-Identical Infidelity Distributions on Heavy-Hex Lattice*. Version 1. Feb. 14, 2024. DOI: 10.48550/arXiv.2402.08203. arXiv: 2402.08203 [quant-ph]. URL: <http://arxiv.org/abs/2402.08203>. Pre-published.
- [5] Samudra Dasgupta and Travis S Humble. "Reliable devices yield stable quantum computations". In: *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*. Vol. 2. IEEE. 2023, pp. 223–226.

- [6] Dripto M. Debroy et al. *LUCI in the Surface Code with Dropouts*. arXiv:2410.14891 [quant-ph]. Oct. 2024. DOI: 10.48550/arXiv.2410.14891. URL: <http://arxiv.org/abs/2410.14891>.
- [7] Eric Dennis et al. “Topological quantum memory”. In: *Journal of Mathematical Physics* 43.9 (2002), pp. 4452–4505.
- [8] Cirq Developers. *Cirq*. Zenodo. Aug. 2025. DOI: 10.5281/ZENODO.16867504.
- [9] Austin G Fowler et al. “Surface codes: Towards practical large-scale quantum computation”. In: *Physical Review A—Atomic, Molecular, and Optical Physics* 86.3 (2012), p. 032324.
- [10] Austin G. Fowler et al. “Surface codes: Towards practical large-scale quantum computation”. In: *Physical Review A* 86.3 (Sept. 2012). arXiv:1208.0928 [quant-ph], p. 032324. ISSN: 1050-2947, 1094-1622. DOI: 10.1103/PhysRevA.86.032324. URL: <http://arxiv.org/abs/1208.0928>.
- [11] Craig Gidney. *Stim*. Version v1.15.0. May 2025. DOI: 10.5281/zenodo.15354872. URL: <https://doi.org/10.5281/zenodo.15354872>.
- [12] Craig Gidney. “Stim: a fast stabilizer circuit simulator”. In: *Quantum* 5 (July 2021). arXiv:2103.02202 [quant-ph], p. 497. ISSN: 2521-327X. DOI: 10.22331/q-2021-07-06-497. URL: <http://arxiv.org/abs/2103.02202>.
- [13] Craig Gidney and Austin G. Fowler. “Efficient Magic State Factories with a Catalyzed CCZ to T Transformation”. In: *Quantum* 3 (Apr. 2019), p. 135. DOI: 10.22331/q-2019-04-30-135.
- [14] Ali Javadi-Abhari et al. *Quantum Computing with Qiskit*. June 2024. DOI: 10.48550/arXiv.2405.08810. arXiv: 2405.08810 [quant-ph].
- [15] Youngseok Kim et al. “Error Mitigation with Stabilized Noise in Superconducting Quantum Processors”. In: *Nature Communications* 16.1 (Sept. 2025), p. 8439. ISSN: 2041-1723. DOI: 10.1038/s41467-025-62820-9.
- [16] Sophia Fuhui Lin et al. “Codesign of quantum error-correcting codes and modular chiplets in the presence of defects”. In: *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*. arXiv:2305.00138 [quant-ph]. Apr. 2024, pp. 216–231. DOI: 10.1145/3620665.3640362. URL: <http://arxiv.org/abs/2305.00138>.
- [17] Daniel Litinski. “A game of surface codes: Large-scale quantum computing with lattice surgery”. In: *Quantum* 3 (2019), p. 128.
- [18] Daniel Litinski. “Magic state distillation: Not as costly as you think”. In: *Quantum* 3 (2019), p. 205.
- [19] Masoud Mohseni et al. *How to Build a Quantum Supercomputer: Scaling from Hundreds to Millions of Qubits*. arXiv:2411.10406 [quant-ph]. Jan. 2025. DOI: 10.48550/arXiv.2411.10406. URL: <http://arxiv.org/abs/2411.10406>.
- [20] Prakash Murali et al. “Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers”. In: *Proceedings of the twenty-fourth international conference on architectural support for programming languages and operating systems*. 2019, pp. 1015–1029.
- [21] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. <https://www.cambridge.org/highereducation/books/quantum-computation-and-quantum-information/01E10196D0A682A6AEFFEA52D53BE9AE>. Dec. 2010. DOI: 10.1017/CBO9780511976667.
- [22] “Quantum error correction below the surface code threshold”. In: *Nature* 638.8052 (2025), pp. 920–926.
- [23] Salonik Resch and Ulya R. Karpuzcu. “Benchmarking Quantum Computers and the Impact of Quantum Noise”. In: *ACM Comput. Surv.* 54.7 (July 2021), 142:1–142:35. ISSN: 0360-0300. DOI: 10.1145/3464420. URL: <https://dl.acm.org/doi/10.1145/3464420>.
- [24] Kaitlin N Smith et al. “Scaling superconducting quantum computers with chiplet architectures”. In: *2022 55th IEEE/ACM International Symposium on Microarchitecture (MICRO)*. IEEE. 2022, pp. 1092–1109.
- [25] Armands Strikis, Simon C Benjamin, and Benjamin J Brown. “Quantum computing is scalable on a planar array of qubits with fabrication defects”. In: *Physical Review Applied* 19.6 (2023), p. 064081.
- [26] Youngkyu Sung et al. “Non-Gaussian Noise Spectroscopy with a Superconducting Qubit Sensor”. In: *Nature Communications* 10.1 (Sept. 2019), p. 3715. ISSN: 2041-1723. DOI: 10.1038/s41467-019-11699-4.
- [27] Swamit S Tannu and Moinuddin K Qureshi. “Not all qubits are created equal: A case for variability-aware policies for NISQ-era quantum computers”. In: *Proceedings of the twenty-fourth international conference on architectural support for programming languages and operating systems*. 2019, pp. 987–999.
- [28] Yu Tomita and Krysta M. Svore. “Low-Distance Surface Codes under Realistic Quantum Noise”. In: *Physical Review A* 90.6 (Dec. 2014), p. 062320. ISSN: 1050-2947, 1094-1622. DOI: 10.1103/PhysRevA.90.062320. arXiv: 1404.3747 [quant-ph].
- [29] Keyi Yin et al. *Surf-Deformer: Mitigating Dynamic Defects on Surface Code via Adaptive Deformation*. arXiv:2405.06941 [quant-ph]. Sept. 2024. DOI: 10.48550/arXiv.2405.06941. URL: <http://arxiv.org/abs/2405.06941>.