Quantum computing by color-code lattice surgery

Andrew J. Landahl$^{1,2,3,*}$ and Ciarán Ryan-Anderson$^{1,2,3,†}$

$^1$Advanced Device Technologies, Sandia National Laboratories, Albuquerque, NM, 87185, USA
$^2$Center for Quantum Information and Control, University of New Mexico, Albuquerque, NM, 87131, USA
$^3$Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM, 87131, USA

We demonstrate how to use lattice surgery to enact a universal set of fault-tolerant quantum operations with color codes. Along the way, we also improve existing surface-code lattice-surgery methods. Lattice-surgery methods use fewer qubits and the same time or less than associated defect-braiding methods. Furthermore, per code distance, color-code lattice surgery uses approximately half the qubits and the same time or less than surface-code lattice surgery. Color-code lattice surgery can also implement the Hadamard and phase gates in a single transversal step—much faster than surface-code lattice surgery can. Against uncorrelated circuit-level depolarizing noise, color-code lattice surgery uses fewer qubits to achieve the same degree of fault-tolerant error suppression as surface-code lattice surgery when the noise rate is low enough and the error suppression demand is high enough.

PACS numbers: 03.67.Lx

I. INTRODUCTION

Planar topological quantum error-correcting codes have emerged as promising substrates for fault-tolerant quantum computing because of their high thresholds [1], compatibility with two-dimensional (2D) local quantum processing [2], low quantum circuit overheads [3], efficient decoding algorithms [2, 4, 5], and the ability to smoothly interpolate between desired effective error rates, which concatenated codes cannot do [6]. By Anderson’s classification theorem [7], the only alternatives for planar topological stabilizer codes with nonlocal logical operators are surface codes [8] and color codes [9].

In principle, fault-tolerant quantum computing with surface codes can be achieved with transversal methods [2], defect-based methods [3, 10, 11], or lattice-surgery-based methods [12]. On 2D arrays of qubits restricted to local quantum processing and local qubit movements, transversal methods require an amount of information swapping that scales with the system size. Defect and lattice-surgery methods avoid this, improving both their runtime and their accuracy threshold [13]. Of these latter two, lattice surgery uses substantially fewer qubits to achieve a desired error rate. For example, the fewest-qubit fault-tolerant distance-three CNOT method in a topological code reported to date uses surface-code lattice surgery and only requires 53 qubits [12].

Extending transversal surface-code methods to color codes is straightforward. Fowler has also extended defect-based surface-code methods to defect-based color-code methods [14]. Notably absent are extensions of surface-code lattice-surgery methods to color-code lattice-surgery methods. Developing such methods is especially important because not only are lattice-surgery methods more qubit-efficient than defect-based methods, but also color codes are significantly more qubit-efficient than surface codes—for example, 4.8.8 color codes use about half the qubits as the qubit-optimal medial surface code [15] to achieve the same code distance [16].

Going beyond the application of a topological quantum memory [2], color-codes offer additional advantages. While transversal two-qubit operations incur penalties for swapping information around, one-qubit transversal operations do not; these advantages carry over to lattice-surgery methods. Two especially noteworthy methods are those for the encoded, or “logical,” Hadamard gate ($H$) and those for the logical phase gate ($S$) on planar color codes on the 4.8.8 lattice—both can be implemented in a single parallelized transversal step [9]. For surface codes, neither of these gates have transversal implementations on any lattice. Current surface-code solutions for these gates include elaborate multi-step code deformation procedures to implement the Hadamard gate [12, 17] and lengthy multi-gate teleportation procedures from (previously distilled) magic states to implement the phase gate [10, 18].

The only downside to color codes versus surface codes is their lower accuracy threshold, whose value has been estimated to be 0.143% against depolarizing circuit-level noise using a perfect-matching decoder [5]. Surface codes have an accuracy threshold whose value has been estimated to be in the range 0.502(1)% to 1.140(1)% [1] in the same setting. That said, surface codes have enjoyed far greater study than color codes and we expect that there are opportunities to close the gap. We will show later that, even as things stand now, at sufficiently low error rates and sufficiently low desired error rates to be achieved by encoding, color codes still use fewer qubits, despite their lower accuracy threshold.

Bolstered by the possibility of significant time and qubit reductions for fault-tolerant operations, in this article we develop methods for universal fault-tolerant quantum computation using color-code lattice surgery. We show

---

*alandahl@sandia.gov
†ciaranra@unm.edu
that our methods use fewer qubits per logical operation than surface-code lattice-surgery methods, including the smallest distance-three CNOT in a topological code—our color-code lattice-surgery methods only use 30 qubits when one allocates one syndrome qubit per face (or 22 if one uses a single mobile syndrome qubit). Along the way, we also improve the surface-code lattice-surgery methods so that the distance-three CNOT now only uses 39 qubits when one allocates one syndrome qubit per face (or 28 if one uses a single mobile syndrome qubit).

In Sec. II, we provide a brief background on triangular 4.8.8 color codes to help make our exposition better self-contained. In Sec. III, we describe fault-tolerant color-code lattice-surgery methods for performing each element in a universal set of operations. In Sec. IV, we calculate the circuit width and depth overheads required by these methods and compare them to the corresponding overheads required by surface-code lattice-surgery methods. Sec. V concludes.

II. BACKGROUND

Our color-code lattice-surgery methods are valid for any color code, but for concreteness we focus on lattice surgery of triangular color codes on the 4.8.8 lattice, namely the semiregular lattice that has a square and two octagons surrounding each vertex. These quantum stabilizer codes [19] exist for any odd code distance \( d \) and can be depicted graphically as in Fig. 1. Each vertex in this figure corresponds to one ("data") qubit in the code. Each face in the figure corresponds to two code checks, or stabilizer generators; one check acts as Pauli \( X \) on all qubits incident on the face and one check acts as Pauli \( Z \) on all qubits incident on the face. The collection of qubits and checks encode a single "logical" qubit. Representatives of the logical \( X \) and \( Z \) operators are strings of \( X \) and \( Z \) operators acting on the qubits along the bottom side of the triangle. By multiplying by a suitable collection of check operators, two other equivalent representatives are similar strings along either of the other two triangle sides.

![FIG. 1: Triangular 4.8.8 color codes of distances 3, 5, and 7. The number of data qubits for distance \( d \) is \((d^2 - 1)/2 + d\). The number of faces (which is half the number of checks) is \((d^2 + 2d - 3)/4\).](image)

Syndrome qubits are associated with the faces in the graph; how many syndrome qubits are associated with each face is a nuanced function of the syndrome extraction protocol one uses. At a minimum, one can use a single syndrome qubit over and over again, but it would have to be moved either physically or by SWAP gates in such a way that it interacted with every data qubit on the interior six times, every data qubit on the edge four times, and every data qubit on a corner twice, because that is the number of checks each of these types of data qubits are involved in. A faster syndrome extraction is possible by allocating one syndrome qubit per face so that each syndrome qubit is used to measure both the \( X \) and the \( Z \) check on each face. By allocating two qubits per face, syndrome extraction can run faster still, with the \( X \) and \( Z \) check measurements scheduled in an interleaved fashion [16].

Adding more syndrome qubits can lead to better performance, such as a higher accuracy threshold or less error propagation; we examine these tradeoffs in greater detail in Sec. IV. One way to increase the number of syndrome qubits is to allocate five syndrome qubits per each octagonal face and two per each square face, extracting the syndrome into two-qubit and verified four-qubit cat states [5, 14]. By doubling this number of syndrome qubits, two cat states per face can be prepared in parallel and used in the interleaved schedule for \( X \) and \( Z \) check measurements. Going even further, one can allocate one syndrome qubit for every data qubit to enact even more robust Shor-style [20] or Steane-style [21] syndrome extraction. This number of qubits can be doubled further to enact Knill-style syndrome extraction with the same robustness but a faster extraction circuit [22]. We are not aware of any schemes that use even more syndrome qubits to any advantage, so the number of syndrome qubits can range anywhere from one to twice the number of data qubits. In this article, we will generally restrict attention to schemes which use either one syndrome qubit per face or one syndrome qubit per check (two per face), as we believe these offer the closest comparison to the most widely-studied surface-code syndrome layout scheme, namely the one with one syndrome qubit per check (one per face) [2].

Color codes are frequently considered in one of three broad classes of error models [16]. In code-capacity models, data qubits are subject to error but syndrome qubits are not. In phenomenological models, both data and syndrome qubits are subject to error. In circuit-level models, data qubits, syndrome qubits, and the individual quantum gates that act upon them are subject to error. This latter class is the most realistic and is the one we focus on in this article. However, because the available operations at the circuit level are very hardware-dependent, we abstract away the specifics of the hardware-level gate basis wherever possible.

Even when the physical circuit gate basis is known, it can be the case that the error model on that gate basis is not well known. In the absence of an experimentally-informed circuit-level error model, a frequently used surrogate is the independent identically distributed (iid) depolarizing noise model, as it is kind of a “worst case” noise model for iid stochastic errors. In the iid depo-
larizing noise model, noise acts independently and identically on the outputs of each quantum circuit element, including the identity gate. Depolarizing noise causes an error to occur with probability $p$, and it selects the error equiprobably among the possible non-identity Pauli operators on the outputs. For single-qubit measurement operations, it also flips the classical bit output with probability $p$ (because a measurement error is a disagreement between the recorded measurement outcome and the actual state). While this noise model is not without its flaws even for iid stochastic errors (see, for example, Refs. [1, 23, 24]), it is widely used.

The syndrome extracted from a color code can be decoded in myriad ways. For the best performance, one could use the optimal decoder. Although optimal decoding of stabilizer codes is #P-hard in general [25], it is possible that an efficient optimal decoder (or one that approximates it arbitrarily well) for color codes will be found. For example, the optimal-decoder-approximating PEPS decoder for surface-codes might be extended to color codes [26]. Alternatively, one could use a slightly weaker integer-program-based decoder that identifies the most likely error given the syndrome [16]. Weaker still but faster yet, one could matching-based decoder, such as a minimum-weight perfect matching decoder [5, 27], a renormalization-group matching decoder [28–30], a local greedy matching decoder [31–34], or a “global attractive-force” local cellular automaton matching decoder [35, 36]. It is also possible to exploit the local equivalence between a color code and a finite number of copies of the surface code to arrive at a decoding solution from multiple surface-code decoders [37, 38]. Developing new color-code decoders is an active research front, where the trade space between decoding complexity and decoding performance is being explored.

### III. UNIVERSAL GATE SET

In this section, we describe how to fault-tolerantly perform a universal set of operations by lattice surgery on 4.8.8 triangular color codes. We use the notation from Ref. [39] to denote gates, states, measurements, and quantum circuits. The universal set we effect in encoded form by lattice surgery is as follows:

$$\{ I, |0\rangle, |+\rangle, M_Z, M_X, S, H, T|+\rangle, \text{CNOT}\}.$$  

(1)

In the absence of hardware-informed circuit-level details, we imagine that the same set of operations is available on the physical qubits as well, with the CNOT gates restricted to nearest-neighbor data-ancilla qubit pairs.

With this gate basis, Pauli operators never need to be applied or even synthesized from the other gates. By the Gottesman-Knill theorem [40], Pauli operators can be propagated through all stabilizer operations (Clifford gates plus Pauli preparations and measurements) efficiently classically and used solely to reinterpret measurement results. Since this gate basis consists solely of stabilizer operations and the $T|+\rangle$ preparation, and because Pauli operators never need to be propagated through preparations, no Pauli operators are ever needed. Importantly, this means that if a decoding algorithm calls for Pauli operators to be applied as a corrective action, the data need not be touched by the Pauli operators and the classical “Pauli frame” can be updated instead. That said, to avoid polynomial-time classical computation, it might be useful to implement the Pauli-frame updates from time to time. For example, if errors are not corrected but only tracked, then the observed syndrome bit rate will climb until it reaches a steady state close to 50%, at which point decoding may take longer than if the tracked Pauli errors had been actually reversed.

In our fault-tolerant constructions, all but the $T|+\rangle$ preparation become exponentially more tolerant to faults as the code distance increases. To increase the fidelity of $T|+\rangle$ preparations, any of a number of magic-state distillation protocols can be used [41–44]. These protocols use high-fidelity operations from the rest of the set to “distill” multiple $T|+\rangle$ preparations into fewer $T|+\rangle$ preparations of higher fidelity.

#### A. The identity gate $I$

To fault-tolerantly implement the encoded identity gate on a triangular color code, we simply perform fault-tolerant quantum error correction by measuring the syndrome for $d$ rounds and run a classical decoding algorithm on the data, such as one of the decoders described in Refs. [5, 16, 27, 28, 30], to infer a corrective action.

#### B. Preparation of $|0\rangle$ and $|+\rangle$ states

To fault-tolerantly prepare an encoded $|0\rangle$ state (the $+1$ eigenstate of the encoded $Z$ operator), we first prepare each data qubit in a triangular color code in the state $|0\rangle$ (the $+1$ eigenstates of the physical $Z$ operators). We then perform fault-tolerant quantum error correction by measuring the syndrome $d$ times and running it through a decoder. The process of measuring all of the code checks transforms the set of single-qubit $Z$ checks into a set consisting of (a) the $Z$ checks of the color code and (b) the encoded $Z$ operator for the color code.

The process for fault-tolerantly preparing an encoded $|+\rangle$ state (the $+1$ eigenstate of the encoded $X$ operator) is identical, except that the individual qubits are initially prepared in $+1$ $X$ eigenstates instead of $+1$ $Z$ eigenstates.

#### C. Measurement $M_Z$ and $M_X$

To fault-tolerantly measure the encoded $Z$ operator, $M_Z$, on a logical qubit, we measure each of the data qubits in the logical qubit in the $Z$ basis in a single round and perform classical error correction on the result. This
measurement is “destructive” in that it takes the logical qubit out of the code space. A non-destructive measurement can be implemented by augmenting this destructive measurement with an encoded CNOT gate using Steane’s ancilla-coupled measurement method [21].

Fault-tolerantly measuring the encoded $X$ operator, $M_X$, is similar: we measure each of the data qubits in the logical qubit in the $X$ basis in a single round and perform classical error correction on the result. It is also a destructive measurement, with a non-destructive version achievable using Steane’s method.

D. Phase and Hadamard gates ($S$ and $H$)

Because the 2D color codes are strong CSS codes (meaning that not only do the checks factor into $X$-type and $Z$-type classes but also they have identical support), the transversal Hadamard gate will swap the two types of checks. For triangular color codes (but not, e.g., for color codes on compact surfaces [15]), the logical $X$ and $Z$ operators can be made to be coincident so that the transversal Hadamard gate exchanges these as well. The net result is that the transversal Hadamard gate is a fault-tolerant logical Hadamard gate for triangular color codes.

As shown by Bombin in Ref. [45], the $S$ gate is transversal for 2D color codes as well, with a suitable choice of which physical qubits to apply $S$ to and which to apply $S^\dagger$ to. The 2D triangular color codes on the 4.8.8 lattice have perhaps the simplest allocation choice: use the transversal $S$ operator if the code distance is congruent to 1 mod 4 and the transversal $S^\dagger$ operator if the code distance is congruent to 3 mod 4.

E. The CNOT gate

To fault-tolerantly implement the encoded CNOT gate, we use a sequence of lattice surgery operations. These operations are intended to mimic either the circuit in Fig. 2 or the circuit in Fig. 3, both of which are equivalent to a CNOT gate; these circuits were leveraged heavily in Ref. [46] to combat biased noise.

The Pauli corrections in these circuits can be omitted in our approach because of our choice of gate basis; we simply use them to re-interpret future measurement results as needed. The only operations depicted in these circuits that we have not provided methods for yet are the $M_{XX}$ and $M_{ZZ}$ measurements; with them, we can construct the encoded CNOT operation.
Using these methods for $M_{XX}$ and $M_{ZZ}$ measurements, we describe step-by-step how to implement a fault-tolerant CNOT gate by lattice surgery using a simulation of the circuit in Fig. 2; the simulation of the circuit in Fig. 3 is similar. While our construction works for arbitrary code distances, we depict an example of each step for $d = 5$, with the layout of control, ancilla, and target regions as depicted in Fig. 6; other choices of orientation are possible.

1. Prepare the data qubits in the ancilla region in $|0\rangle$ states ($Z = +1$ eigenstates), as depicted in Fig. 7.

2. Measure the checks in the ancilla region for $d$ rounds and correct errors fault-tolerantly, as depicted in Fig. 8.

3. Measure the checks that fuse the target and ancilla logical qubits in an $M_{XX}$ measurement for $d$ rounds and correct errors fault-tolerantly, as depicted in Fig. 9.

4. Stop measuring the $M_{XX}$-fusing checks and measure the checks for the target and ancilla logical qubits separately, splitting them apart again, for $d$ rounds and correct errors fault-tolerantly, as depicted in Fig. 10.

5. Measure the checks that fuse the control and ancilla logical qubits in an $M_{ZZ}$ measurement for $d$ rounds and correct errors fault-tolerantly, as depicted in Fig. 11.

6. Stop measuring the $M_{ZZ}$-fusing checks and measure the checks for the control and ancilla logical qubits separately, splitting them apart again, for $d$ rounds and correct errors fault-tolerantly, as depicted in Fig. 12.

7. Measure the data qubits in the ancilla region in the $X$ basis destructively and perform classical error correction on the result, as depicted in Fig. 13.
FIG. 8: (Step 2.) The checks in the ancilla region (A) are measured for \(d\) rounds and errors are corrected fault-tolerantly. (color online)

FIG. 9: (Step 3.) The checks that fuse the target and ancilla logical qubits in an \(M_{XX}\) measurement are measured for \(d\) rounds and errors are corrected fault-tolerantly. (color online)

FIG. 10: (Step 4.) The \(M_{XX}\)-fusing checks stop being measured. Instead, the target and ancilla logical qubits checks are measured for \(d\) rounds and errors are corrected fault-tolerantly. (color online)

FIG. 11: (Step 5.) The checks that fuse the control and ancilla logical qubits in an \(M_{ZZ}\) measurement are measured for \(d\) rounds and errors are corrected fault-tolerantly. (color online)

FIG. 12: (Step 6.) The \(M_{ZZ}\)-fusing checks stop being measured. Instead, the control and ancilla logical qubits checks are measured for \(d\) rounds and errors are corrected fault-tolerantly. (color online)

FIG. 13: (Step 7.) The qubits in the ancilla region are measured in the \(X\) basis, implementing a destructive \(M_{X}\) measurement. The result is error-corrected classically. The control and target logical qubit checks are measured for \(d\) rounds and errors are corrected fault-tolerantly. (color online)
As described, this method takes one round of data-qubit preparation, 5d rounds of syndrome extraction, and one round of data-qubit measurement. However, this time can be sped up considerably.

As a starter, a preparation operation on a logical qubit and a fusing operation between that logical qubit and another logical qubit can be combined into a single step—instead of thinking of the operations as “prepare-then-fuse,” one can think of them as a single “grow one of the logical qubits” operation. Step 2 can therefore be eliminated and, without loss of generality, we can omit step 1 and use the state it prepares as the initial state of the method. This reduces the number of rounds of parallelized measurements to $4d + 1$.

Next, a splitting operation between two logical qubits that “heals” the interface between them can happen simultaneously with a fusing operation acting on a different side of one of the logical qubits and a side of a third logical qubit. Running these operations simultaneously does not hamper the fault-tolerance of the method—the code distances do not drop by this kind of parallelization. This observation allows us to eliminate step 4, reducing the number of rounds of parallelized measurements to $3d + 1$. It also means that the target logical qubit is free to use one of its other sides after just $d$ rounds of measurements.

Finally, a splitting operation between two logical qubits can happen simultaneously with a destructive measurement operation that follows on one of them; again, the operations do not interfere with one another. Because the destructive measurement operation only takes one round of parallelized measurements, the time savings is not very great—the number of rounds is reduced to $3d$ with this observation.

F. Preparation of $T|+\rangle$ states

To fault-tolerantly prepare an encoded $T|+\rangle$ state, we use the process of code injection. Figures 14–16 depict the injection process for distances $d = 3, 5,$ and $7$. The coloring in these figures is chosen so that the blue side of the final triangular code is always on the left for ease of discussion. The top two rows of qubits in these figures represent two isolated Bell pairs for $d = 3$ and $d = 7$, even though they look like they are connected to the rest of the surface via a square and a digon.

In the first step, we prepare a single qubit in the state $T|+\rangle$ and we prepare an adjacent region in an auxiliary state that consists of a distance $d - 1$ color-code stabilizer state, along with two additional Bell pairs if $d \equiv 3$ mod 4. For $d > 3$, we prepare the two Bell pairs to $O(p^2)$ error by post-selection, with a mean waiting time of $(1 - p)^{-4} \approx 1 + 4p$ rounds of measurement. In parallel, we measure the rest of the checks three times and use a classical decoding algorithm to suppress errors in the distance $d - 1$ code state to $O(p^2)$. We handle the case of $d = 3$ separately; the auxiliary state is just three Bell states in this case, so we prepare it by post-selection to $O(p^2)$ error with a mean waiting time of $(1 - p)^{-6} \approx 1 + 6p$ rounds of measurement.

To inject the state, in the second step we measure the new blue $X$ and $Z$ checks along the left side, accepting whatever syndrome values we obtain as being “correct.” This causes the Pauli $X$ and $Z$ operators on the single qubit being injected to extend to distance-$d$ logical Pauli $X$ and $Z$ operators along that edge of the triangle. In parallel, we cease measuring the green checks along the
left side, including the digon operator if one is present. However, in parallel we do measure all of the other checks for the code.

For \( d > 3 \), the checks that persist are capable of detecting up to two errors on any pair of data qubits, excluding the state to be injected. Any single or two-qubit error on the interior data qubits will be detected because the code distance is sufficiently high. Any single-qubit error on data qubits along the left boundary will be incident on a red check or the bottom-left green check, so it will be detected as well. If a two-qubit error afflicts two data qubits on different red checks on the left side or a red check and the bottom-left green check, they will also be detected. If a two-qubit error afflicts two data qubits on a single red check, at least one other persistent check will detect it, by inspection. Since the persistent checks can detect up to two errors, one can use a classical decoding algorithm on three rounds of extracted syndrome to correct any single error, suppressing errors to \( O(p^2) \). The case of \( d = 3 \) can be handled as a special case with, e.g., postselection on the entire injection process.

The total number of rounds of syndrome extraction in the state-injection process is six: three to prepare the ancillary state and three to decode the full distance-\( d \) code. The error in the process is \( O(p) \), where the multiplicative constant is solely a function of the circuit elements in the check measurement circuit that act on the state to be injected. Importantly, this constant does not grow with the distance of the code. To reduce this error further once it is encoded, an encoded magic-state distillation protocol may be used.

IV. RESOURCE ANALYSIS

A. Overhead per code distance

Table I summarizes the space and time resource overheads used by our color-code lattice-surgery methods for the scenario in which one syndrome qubit is allocated per check (two per face).

| Gate | \( T[+], I[0], I[+] \) | \( M_Z, M_X \) | \( H, S \) | \( CNOT \) |
|------|----------------|---------------|---------------|---------------|
| Depth | \( 6 + d \) | \( d \) | \( \frac{3d}{2} \) | \( \frac{3d}{2} \) |
| Qubits | \( d^2 + 2d - 2 \) | \( 3d^2 + 6d - 6 \) | \( 3d^2 + 6d - 6 \) | \( 3d^2 + 6d - 6 \) |
| Error | \( O(p) \) | \( O(p^2) \) | \( O(p^2) \) | \( O(p^2) \) |

TABLE I: Resources used by fault-tolerant 4.8.8 triangular color-code lattice surgery on distance-\( d \) codes when two syndrome bits per face are allocated. Depth is measured in number of measurement rounds. Qubit counts include both data and syndrome qubits. Error is reported in big-\( O \) notation because syndrome-extraction-circuit implementation details can change the constants.

While surface-code lattice-surgery was first explored in by Dennis et al. in the context of state injection [2], the first exploration of a universal set of logical gates on surface codes using lattice-surgery methods was performed by Horsman et al. [12]. Inspired by our color-code lattice surgery methods, we improved the methods presented in Ref. [12] so that they now use fewer qubits for the \( CNOT \), \( H \), and \( S \) gates, using the layout depicted in Fig. 17. We also developed a new six-step surface-code state-injection method similar to our color-code state-injection method; the surface-code layout is depicted in Fig. 18. Table II lists the resources used by these improved surface-surgery methods on the “rotated” or “medial” surface code, with an allocation of one syndrome per check (one per face).

Surface-code lattice surgery (1 syndrome qubit/check)  

| Gate | \( T[+], I[0], I[+] \) | \( M_Z, M_X \) | \( H, S \) | \( CNOT \) |
|------|----------------|---------------|---------------|---------------|
| Depth | \( 6 \) | \( d \) | \( 6d + 1 \) | \( 3d + 1 \) |
| Qubits | \( 2d^2 - 2d + 1 \) | \( 6d^2 - 6d + 3 \) | \( 6d^2 - 6d + 3 \) | \( 6d^2 - 6d + 3 \) |
| Error | \( O(p^2) \) | \( O(p^2) \) | \( O(p^2) \) | \( O(p^2) \) |

TABLE II: Resources used by fault-tolerant medial surface-code lattice surgery on distance-\( d \) codes when one syndrome bit per face is allocated. Depth is measured in number of measurement rounds. Qubit counts include both data and syndrome qubits. Error is reported in big-\( O \) notation because syndrome-extraction-circuit implementation details can change the constants. The logical \( S \) gate is implemented by catalytic teleportation from the \( HS[+]+ \) state, which requires two logical \( CNOT \) gates and a logical Hadamard gate [18]. The logical \( H \) gate is performed by lattice surgery as in Ref. [12], but qubits are shifted \( d \) sites horizontally and \( d \) sites vertically in the method to ensure that the size of the logical operators do not drop below \( d \), making the operation fault-tolerant.

From these tables, we see that while color codes use approximately half as many qubits as surface codes to achieve the same order of error suppression. Color-code lattice surgery also performs encoded gates in essentially the same time or faster than they are performed via surface-code lattice surgery. Even when both models are optimized for qubits by exploiting a single roving syndrome qubit, the color-code \( CNOT \) uses \( (3d^2 + 6d − 1)/2 \) qubits whereas the surface-code \( CNOT \) uses \( 3d^2 + 1 \) qubits—again about half as many.

B. Overhead per desired level of error suppression

Because the accuracy threshold against circuit-level depolarizing noise is smaller for color codes than for surface codes, a color code will need a larger code distance than a surface would need to achieve the same level of error suppression (i.e., to achieve the same logical failure probability \( p_{\text{fail}} \)). This erodes the factor-of-two qubit savings
that color codes provide at the same code distance, and could possibly eliminate the savings entirely.

To compute the qubit overhead $\Omega$ to achieve a given $p_{\text{fail}}$ for a logical operation, one inverts the relationship $p_{\text{fail}}(d)$ and plugs the solution $d(p_{\text{fail}})$ into the appropriate expression for the number of qubits per operation, e.g., from the “Qubits” entry in Table I or Table II. The analytic expression best-suited for $p_{\text{fail}}(d)$ depends on the relative magnitudes of $d$ and the depolarizing probability $p$ [2, 11, 32, 47–49]; for example, Watson and Barrett have shown that the scaling of $p_{\text{fail}}$ with $d$ is qualitatively different in the regime $d < 1/4p$ and $d > 1/4p$ for code-capacity and phenomenological error models [49]. Since overhead comparisons are most relevant for non-asymptotic $d$ and for $p$ below the relevant pseudothreshold (i.e., the $p$ at a fixed code distance below which $p_{\text{fail}} < p$), and because we are most interested in the scaling for circuit-level error models, we use the expression for fixed $d$ and low $p$ for these models that Fowler found fit well to surface-codes in Ref. [48], namely

$$p_{\text{fail}} = A(d) \left( \frac{p}{p_{\text{th}}} \right)^{d/2}.$$  \hfill (2)

It is an interesting question as to whether color codes can exhibit the same scaling at this in the low-$p$ regime. Stephens has noted that his color-code matching decoder in Ref. [5] does not attain the full algebraic code distance, suggesting that the exponent in Eq. (2) using his decoder will be $\alpha d$, where $\alpha < 1/2$. In contrast, the integer-program (IP) decoder in Ref. [16] should attain the full code distance at the cost of running more slowly. If only one syndrome qubit per face or one per check is used with the IP decoder, though, errors may spread badly, cutting in to the effective code distance. Using Shor-, Steane-, or Knill-style syndrome extraction should eliminate this problem at the cost of many extra syndrome qubits. It may suffice to use the verified four-cat and two-cat states per octagonal and square faces respectively as used in Refs. [5, 14] with the IP decoder to achieve this scaling, but currently that is an open question. Although the IP decoder appears to be inefficient at high error rates, at low error rates it can be expected to run quickly. Moreover, the recent linear-time PEPS decoder for surface codes [26] gives hope that a truly efficient color-code decoder that achieves the scaling of Eq. (2) will be found. For the purposes of comparison, and with this optimism in mind, we will assume that the scaling law in Eq. (2) holds for both surface and color codes. However, we urge caution in reading too much into the results derived from this assumption.

Using Eq. (2), the color-code distance $d_c$ that gives the same error-suppression power as a surface code with distance $d_s$ is

$$d_c = d_s \left( \frac{\log p_{\text{th}}/p}{\log p_{\text{th}}/p_{\text{th}}^{(c)}} \right) + 2 \left( \frac{\log A_s(d)/A_c(d)}{\log p_{\text{th}}/p_{\text{th}}^{(c)}} \right).$$  \hfill (3)

Fowler’s numerical simulations suggest that $A_s(d)$ is approximately a constant function of $d$ for $d$ up to 10 [48]; there is no reason to expect that $A_s(d)$ is not also a comparably-sized constant function of $d$ in the same range, or indeed that $A_c(d)$ and $A_s(d)$ should scale substantially differently for any $d$. The numerator in the second term of Eq. (2) should therefore be quite small because of the logarithm. Moreover, the denominator gets larger as $p$ is reduced below the color-code (pseudo)threshold, making the overall term even smaller. For these reasons, we will neglect the second term in Eq. (2) in our subsequent analysis.
Using the expressions in Tables I and II for the color-code and surface-code qubit overheads, which we denote by $\Omega_c(d)$ and $\Omega_s(d)$, and the relationship in Eq. (3), we plot the ratio $\Omega_c(d_c(d_s)) / \Omega_s(d_s)$ versus $p$ for several values of $d_s$. This ratio is sensitive to the estimates for $p_{th}^{(c)}$ and $p_{th}^{(s)}$, so we present two plots at the extremes of the estimates. Figure 19 is the plot using the highest estimate for the color-code accuracy threshold $(0.143\%)$ and the lowest estimate for the color-code accuracy threshold $(0.502\%)$. Figure 20 is the plot using the lowest estimate for the color-code accuracy threshold $(0.082\%)$ and the highest estimate for the surface-code accuracy threshold $(1.140\%)$.

From these plots, we see that for distances greater than 11, as long as $p$ is below a value bracketed approximately somewhere between $10^{-5}$ to $10^{-7}$, color codes use fewer qubits to achieve the same level of error suppression.

![Figure 19: Ratio of color-code to surface-code qubit overhead $\Omega_c / \Omega_s$ versus circuit-level depolarizing probability $p$ when both codes are tuned via Eq. (3) to achieve the same logical qubit failure probability. Plots assume a color-code accuracy threshold of 0.143% and a surface-code accuracy threshold of 0.502%. (color online)](image)

This conclusion could be sharpened by direct numerical simulations, which we believe would be an interesting future research project. Rather than assuming a phenomenological scaling law as in Eq. (2) for the failure probability and using it to infer the overhead, one could perform direct numerical estimation of the overhead as a function of $d$ and $p$ and compare the results for color codes and surface codes. In addition to removing the need to fit an assumed scaling law, this approach would also remove the need to estimate accuracy thresholds, or even pseudothresholds, because it gets directly at the question at hand.

![Figure 20: Ratio of color-code to surface-code qubit overhead $\Omega_c / \Omega_s$ versus circuit-level depolarizing probability $p$ when both codes are tuned via Eq. (3) to achieve the same logical qubit failure probability. Plots assume a color-code accuracy threshold of 0.082% and a surface-code accuracy threshold of 1.140%. (color online)](image)

### C. Overhead for small logical CNOT gates

Because of the interest expressed in Ref. [12] in designing the fewest-qubit implementation of a CNOT gate with a topological stabilizer code, we thought it would be valuable to list the qubit overheads required by the methods described here for small distances. As mentioned in Sec. II, the number of syndrome qubits used by an implementation of a topological stabilizer code is design dependent: a single roving syndrome qubit would suffice, but one could use a number of syndrome qubits up to twice the number of data qubits to some advantage.

In Table III, we list the qubit overhead required for the low end of the syndrome-allocation spectrum for the following methods: (a) color-code transversal methods, (b) our color-code lattice-surgery methods, (c) surface-code transversal methods, (d) our surface-code lattice-surgery methods, and (e) the surface-code lattice-surgery methods described in Ref. [12].

As noted in our introduction, transversal methods are not well-suited to local quantum processing on two-dimensional arrays of qubits restricted to local movements; we list the overheads here despite this because at small distances, one might be able to exploit non-local processing and/or nonlocal qubit movement. For example, a recent demonstration in a trapped-ion quantum computer of a single-round of error correction on a distance-three color code exploited the fact that all seven $^{40}$Ca$^+$ ions involved were trapped in a single Paul trap [50]. (The minimal extra “roving” syndrome qubit was not used in the experiment because the protocol was not fault-tolerant—instead of repeating syndrome measurements into one or more auxiliary qubits,
the data-qubit ions were measured destructively once via resonance-fluorescence.)

For all methods in Table III, we consider the allocations of (i) a single roving syndrome qubit, (ii) one syndrome qubit per face, and (iii) one syndrome qubit per check, which is the same as one per face for surface codes but is two per face for color codes. For transversal methods, we also consider an in-between variant with (iv) one syndrome qubit per two faces, because one might want to share the syndrome qubits transversally between the two logical qubits. (The case of sharing two syndrome qubits per face between the two logical qubits has the same overhead count as having both logical qubits use one syndrome qubit per face.)

| Color transversal: 1 total  | 15  | 35 | 63 | 99 | 143 |
|----------------------------|-----|----|----|----|-----|
| Color transversal: faces/2 | 17  | 42 | 77 | 122| 177 |
| Color transversal: faces   | 20  | 50 | 92 | 146| 212 |
| Color transversal: 2×faces | 26  | 66 | 122| 194| 282 |
| Color surgery: 1 total     | 22  | 52 | 94 | 148| 214 |
| Color surgery: faces       | 30  | 75 | 138| 219| 318 |
| Color surgery: 2×faces     | 39  | 99 | 183| 291| 423 |
| Surface transversal: 1 total| 19  | 51 | 99 | 163| 243 |
| Surface transversal: faces/2| 22  | 66 | 134| 226| 342 |
| Surface transversal: faces | 26  | 82 | 170| 290| 442 |
| Surface surgery: 1 total   | 28  | 76 | 148| 244| 364 |
| Surface surgery: faces     | 39  | 123| 255| 435| 666 |
| Surface surgery [12]: 1 total| 34  | 86 | 162| 262| 386 |
| Surface surgery [12]: faces| 53  | 149| 293| 485| 725 |

TABLE III: Number of qubits needed to implement a logical CNOT gate for several color-code and surface-code methods for small values of the code distance $d$, assuming that the number of syndrome qubits used is as indicated.

V. CONCLUSIONS

Our color-code lattice-surgery methods open new possibilities for achieving fault-tolerant quantum computation using fewer resources. Per code distance, they are manifestly superior to surface-code lattice-surgery methods, using approximately half the qubits and the same time or less to perform logical quantum operations. Although we did not discuss it, they also use fewer qubits and the same time or less than defect-based “spacetime braiding” methods for both surface-codes [10] and color-codes [14]. Transversal methods do use fewer qubits per code distance than color-code lattice surgery to perform logical operations [2], but transversal methods cannot be implemented in systems utilizing local quantum processing on two-dimensional arrays of qubits restricted to local movements.

Because color codes are estimated to have a lower accuracy threshold than surface codes against uncorrelated circuit-level depolarizing noise [1, 5, 16], the superiority of color codes only becomes manifest at sufficiently low depolarizing error probabilities and sufficiently large code distances. Subject to an assumed scaling law given by Eq. (2) for both surface codes and color codes, the depolarizing probability cutoff is approximately somewhere in the range $p = 10^{-5}$ to $p = 10^{-7}$ with a corresponding distance cutoff of $d = 11$. Color-code decoder research is only in its infancy, and we believe that the regime of superiority can be expanded with further study. For example, the recent linear-time PEPS decoder by Bravyi et al. [26] might be extended to color codes, allowing one to approximate the optimal decoder quite well with only linear-time processing. The close relationship between color codes and surface codes at the topological-phase level [37] means that the decoding complexity, if not the performance, can always be made comparable for the two classes of codes [30, 37, 38].

It would seem then, color codes are equal to or superior to surface codes, at least insofar as space and time overhead considerations are concerned, for systems that are sufficiently mature, meaning that they have sufficiently low error rates and sufficiently many qubits available. When technology brings us to this point, we believe the transition from (two-colorable) surface-codes to (three-colorable) color codes will resemble the transition of television broadcasts from black-and-white to color: perhaps a little bumpy at first, but inevitable. Until then, the mandate for color-code research is to bring that horizon closer to the present.

Acknowledgments

The authors were supported in part by the Laboratory Directed Research and Development program at Sandia National Laboratories. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-AC04-94AL85000. The authors would like to thank Eric Bahr, Chris Cesare, Anand Ganti, Setso Metodi, and Uzoma Onunkwo for helpful discussions.

[1] A. M. Stephens, Fault-tolerant thresholds for quantum error correction with the surface code, Phys. Rev. A 89, 022321 (2014), doi:10.1103/PhysRevA.89.022321, arXiv:1311.5003.
[2] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, J. Math. Phys. 43, 4452 (2002), doi:10.1063/1.1499754, arXiv:quant-ph/0110143.
[3] R. Raussendorf and J. Harrington, Fault-tolerant quantum computation with high threshold in two dimensions, Phys. Rev. Lett. 98, 190504.
(2007), doi:10.1103/PhysRevLett.98.190504, arXiv:quant-ph/0610082.

[4] G. Duclos-Cianci and D. Poulin, Fast decoders for topological quantum codes, Phys. Rev. Lett. 104, 050504 (2010), doi:10.1103/PhysRevLett.104.050504, arXiv:0911.0581.

[5] A. M. Stephens, Efficient fault-tolerant decoding of topological color codes (2014), arXiv:1402.3037.

[6] A. W. Cross, D. P. DiVincenzo, and B. M. Terhal, Comparative code study for quantum fault tolerance, Quant. Inf. Comp. 9, 541 (2009), arXiv:0711.1556, URL http://www.rintonpress.com/xxqic9/qic-9-78/0541-0572.pdf.

[7] J. T. Anderson, Homological stabilizer codes (2011), arXiv:1107.2502.

[8] A. Y. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. 303, 2 (2003), doi:10.1016/S0003-4916(02)00018-0, arXiv:quant-ph/9707021.

[9] H. Bombin and M. A. Martin-Delgado, Topological quantum distillation, Phys. Rev. Lett. 97, 180501 (2006), doi:10.1103/PhysRevLett.97.180501, arXiv:quant-ph/0605138.

[10] R. Raussendorf, J. Harrington, and K. Goyal, A fault-tolerant one-way quantum computer, Ann. Phys. 321, 2242 (2006), doi:10.1016/j.aop.2006.01.012, arXiv:quant-ph/0510135.

[11] R. Raussendorf, J. Harrington, and K. Goyal, Topological fault-tolerance in cluster state quantum computation, New J. Phys. 9, 199 (2007), doi:10.1088/1367-2630/9/6/199, arXiv:quant-ph/0703143.

[12] C. Horsman, A. G. Fowler, S. Devitt, and R. van Meter, Surface code quantum computing by lattice surgery, New J. Phys. 14, 123011 (2012), doi:10.1088/1367-2630/14/12/123011, arXiv:1111.4022.

[13] F. M. Spedalieri and V. P. Roychowdhury, Latency in local, two-dimensional, fault-tolerant quantum computing, Quant. Inf. Comp. 9, 666 (2009), arXiv:0805.4213, URL http://www.rintonpress.com/xxqic9/qic-9-78/0666-0682.pdf.

[14] A. G. Fowler, Two-dimensional color-code quantum computation, Phys. Rev. A 83, 042310 (2011), doi:10.1103/PhysRevA.83.042310, arXiv:0806.4827v3.

[15] H. Bombin and M. A. Martin-Delgado, Optimal resources for topological two-dimensional stabilizer codes: Comparative study, Phys. Rev. A 76, 012305 (2007), doi:10.1103/PhysRevA.76.012305, arXiv:quant-ph/0703272.

[16] A. J. Landahl, J. T. Anderson, and P. R. Rice, Fault-tolerant quantum computing with color codes (2011), arXiv:1108.5738.

[17] A. G. Fowler, Low-overhead surface code logical Hadamard, Quant. Inf. Comp. 12, 970 (2012), arXiv:1202.2639, URL http://www.rintonpress.com/xxqic12/qic-12-1112/0970-0982.pdf.

[18] P. Aliferis, Level reduction and the quantum threshold theorem, Ph.D. thesis, Caltech (2007), arXiv:quant-ph/0703230.

[19] D. Gottesman, Stabilizer codes and quantum error correction, Ph.D. thesis, Caltech (1997), arXiv:quant-ph/9705052.

[20] P. W. Shor, Fault-tolerant quantum computation, in Proceedings of the 37th Annual Symposium on Foundations of Computer Science, edited by R. S. Sipple, IEEE (IEEE Press, Los Alamitos, CA, 14–16 Oct. 1996, Burlington, VT, USA, 1996), pp. 56–65, ISBN 0-8186-7594-2, doi:10.1103/80097539795293172, arXiv:quant-ph/9605011.

[21] A. M. Steane, Space, time, parallelism and noise requirements for reliable quantum computing, Fortschr. Phys. 46, 443 (1998), arXiv:quant-ph/9708021, URL http://onlinelibrary.wiley.com/doi/10.1002/%28SICI%292152-3978(20060629464%5C3A433::AID-PRA43%3E3.0.CO;2-8/abstract.

[22] E. Knill, Quantum computing with realistically noisy devices, Nature 434, 39 (2005), doi:10.1038/nature03350, arXiv:quant-ph/0410199.

[23] E. Magesan, D. Poulin, Hardness of decoding quantum stabilizer codes (2013), arXiv:1310.3235.

[24] S. Bravyi, M. Suchara, and A. Vargo, Efficient algorithms for maximum likelihood decoding in the surface code (2014), arXiv:1405.4883.

[25] D. S. Wang, A. G. Fowler, C. D. Hill, and L. C. L. Hollenberg, Graphical algorithms and threshold error rates for the 2d color code, Quant. Inf. Comp. 10, 780 (2010), arXiv:0907.1708, URL http://www.rintonpress.com/xxqic10/qic-10-910/0780-0802.pdf.

[26] P. Sarvepalli and R. Raussendorf, Efficient decoding of topological color codes (2011), arXiv:1111.0831.

[27] G. Duclos-Cianci, H. Bombin, and D. Poulin, Fast decoding algorithm for subspace and subsystem color codes and local equivalence of topological phases (2011), personal communication.

[28] G. Duclos-Cianci and D. Poulin, Fault-tolerant renormalization group decoder for Abelian topological codes, Quant. Inf. Comp. 14, 0721 (2014), arXiv:1304.6100, URL http://www.rintonpress.com/xxqic14/qic-14-910/0721-0740.pdf.

[29] E. Dennis, Purifying quantum states: quantum and classical algorithms, Ph.D. thesis, University of California at Santa Barbara (2003), arXiv:quant-ph/0503169.

[30] S. Bravyi and A. Vargo, Simulation of rare events in quantum error correction, Phys. Rev. A 88, 062308 (2013), doi:10.1103/PhysRevA.88.062308, arXiv:1308.6270.

[31] J. R. Wootton, A simple decoder for topological codes (2013), arXiv:1310.2393.

[32] H. Anwar, B. J. Brown, E. T. Campbell, and D. E. Browne, Efficient decoders for quotient topological codes, New J. Phys. 16, 063038 (2014), doi:10.1088/1367-2630/16/6/063038, arXiv:1311.4895.

[33] J. W. Harrington, Analysis of quantum error-correcting codes: symplectic lattice codes and toric codes, Ph.D. thesis, Caltech (2004).

[34] M. Herold, E. T. Campbell, J. Eisert, and M. J. Kast-
ryano, *Cellular-automaton decoders for topological quantum memories* (2014), arXiv:1406.2338.

[37] H. Bombin, G. Duclos-Cianci, and D. Poulin, *Universal topological phase of 2D stabilizer codes*, New J. Phys. 14, 073048 (2012), arXiv:1103.4606.

[38] N. Delfosse, *Decoding color codes by projection onto surface codes*, Phys. Rev. A 89, 012317 (2014), doi:10.1103/PhysRevA.89.012317, arXiv:1308.6207.

[39] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000), ISBN 0-521-63235-8 (Hardback), 0-521-63503-9 (Paperback).

[40] D. Gottesman, *The Heisenberg representation of quantum computers*, in *Group22: Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics*, edited by S. P. Corney, R. Delbourgo, and P. D. Jarvis (International Press, Cambridge, MA, 13–17 Jul. 1998, Hobart, Australia, 1999), pp. 32–43, arXiv:quant-ph/9807006.

[41] S. Bravyi and A. Kitaev, *Universal quantum computation with ideal Clifford gates and noisy ancillas*, Phys. Rev. A 71, 022316 (2005), doi:10.1103/PhysRevA.71.022316, arXiv:quant-ph/0403025.

[42] A. M. Meier, B. Eastin, and E. Knill, *Magic-state distillation with the four-qubit code* (2012), arXiv:1204.4221.

[43] S. Bravyi and J. Haah, *Magic state distillation with low overhead* (2012), arXiv:1209.2426.

[44] A. J. Landahl and C. Cesare, *Complex instruction set computing architecture for performing accurate quantum $z$ rotations with less magic* (2013), arXiv:1302.3240.

[45] H. Bombin, *Gauge color codes* (2013), arXiv:1311.0879.

[46] P. Aliferis and J. Preskill, *Fault-tolerant quantum computation against biased noise*, Phys. Rev. A 78, 052331 (2008), doi:10.1103/PhysRevA.78.052331, arXiv:0710.1301.

[47] C. Wang, J. Harrington, and J. Preskill, *Confinement-Higgs transition in a disordered gauge theory and the accuracy threshold for quantum memory*, Ann. Phys. 303, 31 (2003), doi:10.1016/S0003-4916(02)00009-2, arXiv:quant-ph/0207088.

[48] A. G. Fowler, *Analytic asymptotic performance of topological codes*, Phys. Rev. A 87, 040310(R) (2013), doi:10.1103/PhysRevA.87.040301, arXiv:1208.1334.

[49] F. H. Watson and S. D. Barrett, *Estimating overhead for topological quantum codes* (2013), arXiv:1312.5213.

[50] D. Nigg, M. Mueller, E. A. Martinez, P. Schindler, M. Hennrich, T. Monz, M. A. Martin-Deîgado, and R. Blatt, *Experimental quantum computations on a topologically encoded qubit* (2014), arXiv:1403.5426.