Quantifying Nonlocality: How Outperforming Local Quantum Codes Is Expensive

Nouédyn Baspin[1] and Anirudh Krishna[2]
[1]Université de Sherbrooke, Sherbrooke, Québec, Canada J1K 2R1
[2]Stanford University, Stanford, California 94305, USA

(Received 1 December 2021; accepted 26 May 2022; published 28 July 2022)

Quantum low-density parity-check (LDPC) codes are a promising avenue to reduce the cost of constructing scalable quantum circuits. However, it is unclear how to implement these codes in practice. Seminal results of Bravyi *et al.* [Phys. Rev. Lett. **104**, 050503 (2010)] have shown that quantum LDPC codes implemented through local interactions obey restrictions on their dimension k and distance d. Here we address the complementary question of how many long-range interactions are required to implement a quantum LDPC code with parameters k and d. In particular, in 2D we show that a quantum LDPC code with distance $d \propto n^{1/2+\varepsilon}$ requires $\Omega(n^{1/2+\varepsilon})$ interactions of length $\tilde{\Omega}(n^{\varepsilon})$. Further, a code satisfying $k \propto n$ with distance $d \propto n^{\alpha}$ requires $\tilde{\Omega}(n)$ interactions of length $\tilde{\Omega}(n^{\alpha/2})$. As an application of these results, we consider a model called a stacked architecture, which has previously been considered as a potential way to implement quantum LDPC codes. In this model, although most interactions are local, a few of them are allowed to be very long. We prove that limited long-range connectivity implies quantitative bounds on the distance and code dimension.

DOI: 10.1103/PhysRevLett.129.050505

Introduction.—Finding ways to battle decoherence is among the foremost challenges on the path to implementing fault-tolerant quantum circuits. Quantum error correcting codes can address this issue, and their efficacy is guaranteed by the quantum threshold theorem [1–4]. The code we choose to use will be tailored to the advantages and disadvantages of the physical architecture on which it is implemented. For instance, we might consider how many qubits we can measure jointly, how far apart qubits involved in such measurements need to be located, or how many supplementary qubits will be needed to implement a particular algorithm fault tolerantly [5,6]. We will want the choice of code to be efficient and respect the limitations of our architecture. Consequently, there is a strong interest in understanding how physical constraints on a system can impede the efficiency of a quantum code.

Formally, a quantum error correcting code $\mathcal{C}$ on n qubits is the common $+1$ eigenspace of a set of independent commuting n-qubit Pauli operators $\{\mathsf{S}_1, \ldots, \mathsf{S}_m\}$, referred to as stabilizers,

$$\mathcal{C} = \{|\psi\rangle : \mathsf{S}_i|\psi\rangle = |\psi\rangle \quad \forall \;\; i \in \{1, \ldots, m\}\}.$$

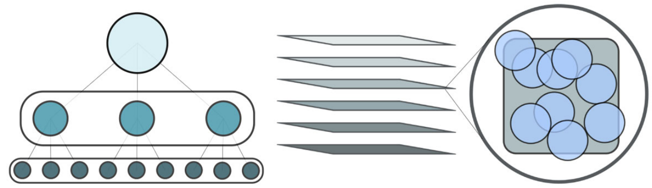
Measuring the stabilizers yields information required to detect and correct errors. Alternatively, the code space can be thought of as the ground space of a commuting Hamiltonian. For ease of implementation, we may stipulate that these measurements be *local*, i.e., that the qubits involved in a stabilizer be contained within a ball of constant radius. Let $k = \log_2 \dim \mathcal{C}$ denote the number of encoded qubits [7]; we aim to encode as many qubits as possible with a limited number of available physical qubits. Furthermore, let d denote the distance; it is a measure of the number of physical qubits that need to be corrupted to irreparably damage encoded information. Seminal works of Bravyi *et al.* [8,9] demonstrated that there are sharp trade-offs between k and d for all local codes. As a result, locality limits our ability to reduce the resource cost of implementing scalable quantum circuits. This naturally raises the following questions—Question 1: To construct an error correcting code with dimension k and distance d, how much nonlocality is needed to implement it? How do we even quantify this seemingly nebulous notion of nonlocality?

Expanding our attention beyond local quantum codes is a worthwhile endeavor as certain architectures support interactions between arbitrary qubits. Prominent examples are silicon-based architectures with photon-mediated interactions that encode qubits into the spin states of silicon [10] or photonic architectures where the qubits are directly encoded in the photons and therefore not localized [11]. Other architectures include atomic arrays [12], where atoms are laid out along a single line, but long-range interactions can be used to simulate higher dimensions. Ion trap architectures that support all-to-all connectivity in a limited capacity have also been considered [13–15]. By dropping the restriction of locality, these architectures could eventually circumvent the limitations of local codes. With this motivation, we consider quantum low-density parity-check (LDPC) codes, a class that subsumes all known topological codes [2,16–18]. The study of these codes is motivated by several results showing that quantum LDPC codes can drastically reduce the number of physical qubits required to build a fault-tolerant quantum computer [19–21]. These results are theoretical and we need to better understand how to translate them for realistic implementations. In practice, we wish to understand how to implement quantum LDPC codes in a two- or three-dimensional layout. It is conceivable that implementing quantum codes where a majority of measurements are local, but some limited amount of long-range connectivity is available. This then prompts the next question concerning locality—Question 2: Can we implement good quantum LDPC codes using a setup where a majority of measurements are local?

In this Letter, we address questions 1 and 2. Through Theorem 2 we show that quantum LDPC codes require large amounts of nonlocality between qubits when the dimension k and the distance d are large. To motivate how to quantify nonlocality, we repeat an observation from [22]. It is not possible to add a limited number of long-range connections and significantly improve the performance of a local code. Any code that we consider will have to have a sufficient number of long-range interactions to work. Our quantification of nonlocality, therefore, in addition to the length of the long-range interactions, will also include the *number* of such interactions.

We highlight codes for which $k \propto n$, and $d \propto n^{\alpha}$ for $\alpha > 0$, as these codes underpin the current proposals for low-overhead quantum computation. Our results state that, to implement these codes in 2D, we require roughly n interactions of length $n^{\alpha/2}$. Therefore, implementing these codes will require an architecture able to deal with a significant amount of nonlocality. Our results are also of interest for good codes, i.e., constant-rate codes for which $\alpha = 1$ [24]. They seem to make optimal use of long-range connectivity. This is because in two dimensions the maximum distance between any two points on an $L \times L$ grid is proportional to $L \propto \sqrt{n}$, which would saturate our bound. Finally, our results suggest that it is expensive to improve the distance of a local code. For example, in 2D, Bravyi and Terhal proved that local codes cannot do better than $d \propto n^{1/2}$ [8]; we show that any code satisfying $d \propto n^{1/2+\varepsilon}$ will require a growing number of long-range interactions. Together, these results suggest that architectures limited to local interactions can only implement topological codes at best.

Next, we consider what we refer to as a *stacked* layout [25]. This model is inspired by the schematic for a concatenated code shown in Fig. 1. In the stacked model, qubits are placed on the vertices of a two-dimensional grid. The measurements required to define the code are partitioned into multiple layers as visualized in Fig. 1. Each layer of the stack represents stabilizers of a given interaction radius. The interaction radius increases as we move up the layers of the stack, while the number of stabilizers decreases. The majority of stabilizers in this model are in the lower layers. Therefore, any code implemented by a stack is mostly local. For this reason, this model has been



(a) (b)

FIG. 1. (a) A schematic for a concatenated code [28]. The qubits of the code are themselves encoded in an error correcting code and this gives rise to a hierarchical structure. (b) A two-dimensional stacked architecture. Qubits are the bottommost layer. Stabilizers, identified with their support, are assigned to different layers above and are depicted using blue circles. Stabilizers in a given layer have a radius of support depending on the layer. This interaction range increases as we move up the stack or, equivalently, the radius of the circles increases. On the other hand, the number of stabilizers in each layer decreases exponentially.

considered a potential route to implement LDPC codes. However, such an architecture cannot implement arbitrary quantum LDPC codes. In Corollary 3, we show that two-dimensional stacked layouts are limited. We show the distance is bounded by $d = \tilde{O}(n^{2/3})$ and the dimension-distance trade-off is $k^3 d^4 = \tilde{O}(n^5)$. This shows that there are strong limitations to such models; however, it does not prevent implementations of constant-rate codes with distance scaling as $\sqrt{n}$. Related work: Delfosse *et al.* provide an explicit multiplanar layout of hypergraph product codes [27]; however, within each plane the connectivity is allowed to be long-range.

Background and intuition.—An $[\![n, k, d]\!]$ quantum code $\mathcal{C}$ is a 2^k-dimensional subspace of the complex Euclidean space $\mathbb{C}^{2^n}$ associated with n qubits. The code space is specified as the joint $+1$ eigenspace of a set of commuting Pauli operators $\mathcal{S} \subset \{\mathsf{I}, \mathsf{X}, \mathsf{Y}, \mathsf{Z}\}^{\otimes n}$ called the stabilizer group. The distance d is the minimum number of qubits that are acted on nontrivially by a Pauli operator to map one element of $\mathcal{C}$ to another. Suppose the group is generated by some elements $\{\mathsf{S}_i\}_{i=1}^{n-k}$. The code is said to be a LDPC code if each generator only acts on a constant number of qubits, and each qubit is only involved in a constant number of generators.

We represent a quantum code $\mathcal{C}$ on n qubits using a "connectivity graph" $G = G(\mathcal{C}) = (V, E)$. Here V refers to the set of vertices of the graph and $E \subseteq V \times V$ the set of edges. Each vertex $v \in V$ of G corresponds to a qubit of $\mathcal{C}$ and two vertices share an edge $e \in E$ if both qubits participate in the same stabilizer generator S_i. The connectivity graph of a LDPC code is *sparse*, i.e., only a constant number of edges emanate from each vertex. In [22], we showed that there is an intimate relationship between the properties of a quantum code and the corresponding connectivity graph. We build on these results to show that the properties of quantum LDPC codes with desired parameters are severely restricted. For an in-depth discussion of this lemma, including the proof, we point the interested reader to [22]. For brevity, we use the following notation in our inequalities (see Ref. [29] for details): consider two functions $f, g: X \to Y$ with real domain and image, i.e., $X, Y \subseteq \mathbb{R}$. If there exists an $x_0 \in \mathbb{R}$ such that for all $x \geq x_0$, (a) there exists a constant c such that $f(x) \geq cg(x)$, we say that $f(x) = O(g(x))$; (b) if there exist constants c_-, c_+ such that $c_- g(x) \leq f(x) \leq c_+ g(x)$, we say that $f(x) = T(g(x))$; and (c) if there exists a constant c such that $f(x) \leq cg(x)$, we say $f(x) = O(g(x))$. These are modified with a tilde when the bounds hold only up to polylogarithmic factors. For example, $f = \tilde{\Omega}(g)$ implies that $f(x) = O[g(x) log^c(x)]$ for some constant c. We use this shorthand because we are interested in the scaling of resources, and this notation allows us to highlight the most important features of this scaling.

Main result: Embedding codes in D dimensions.—In this section, we consider how to embed quantum LDPC codes in $\mathbb{R}^D$. This section is inspired by results from metric geometry that consider the distortion of expander graphs embedded in $\mathbb{R}^D$. Here we show that a class of graphs called ε expanders are difficult to embed. As a consequence, we show that constant-rate quantum codes require a growing number of long-range interactions between qubits.

Definition.—For a graph $G = (V, E)$, a map $\eta: V \to \mathbb{R}^D$ is called an embedding. η satisfies the following condition for all pairs of distinct vertices $u, v \in V$, $|\eta(u) - \eta(v)| \geq 1$. We use $|.|: \mathbb{R}^D \to \mathbb{R}$ to denote the standard Euclidean metric.

In the following sections, we will frequently refer to the length of an edge. We mean that any embedding η naturally endows an edge (u, v) with a length. Equivalently, the length of an edge (u, v) is $|\eta(u) - \eta(v)|$. The condition on the embedding guarantees that two qubits are not squeezed arbitrarily close together.

Theorem 2 (Main).—Let $\mathscr{C} = \{\mathcal{C}_n\}$ be a family of $[\![n, k, d]\!]$ quantum LDPC codes. Further suppose $\mathscr{C}$ is associated with the nontrivial connectivity graphs $\mathcal{G} = \{G_n = (V_n, E_n)\}_n$. For any θ embedding $\eta: V_n \to \mathbb{R}^D$, there exists some β, n_0 such that for code sizes $n > n_0$, and any $\alpha \in (0, 1)$, the following propositions hold: η induces (1) $\Omega(d)$ edges of length $\tilde{\Omega}\{(d)/(n^{(D-1)/D})\}$, (2) $\tilde{\Omega}(\sqrt{(k/n)}d)$ edges of length $\tilde{\Omega}(\sqrt{(k/n)}d^{1/D})$, and (3) $\tilde{\Omega}(\sqrt{[(1-\alpha)k]/(n)}^{1/\log_n(d)}\alpha k)$ edges of length $\tilde{\Omega}(\sqrt{[(1-\alpha)k]/(n)}d^{1/D})$ if $kd^{2/D} \geq \beta n \log(n)^2/(1-\alpha)$.

To understand its implications, we proceed to a short discussion of the theorem and the intuition for the proof. The proof is presented in Sec. II of the Supplemental Material [30]; the proof uses references [31–34]. As shown in [22], a quantum LDPC code with good parameters k and d requires a connectivity graph with a lot of connectivity. We can measure the distance between two vertices on the graph using the graph metric, which is simply the minimum number of edges to traverse between the two vertices. In a tightly connected graph, the minimum distance between vertices is small. For example, in what are known as expander graphs, there is high degree of connectivity. On an expander graph of size n, the maximum distance between two points is $O[\log(n)]$. On the other hand, this distance can be quite large for a poorly connected graph such as the grid graph. For example, for the grid graph in two dimensions, the maximum distance between two points can be proportional to $\sqrt{n}$. In general, any embedding η from the connectivity graph will try to respect the graph metric. This is to minimize distorting the graph and make edges longer than necessary. However, there is only a limited extent to which it can do so, as we have constrained the density of the embedding η. Recall that η cannot place two qubits in D dimensions closer than unit distance apart. It is forced to distort the graph metric for a well-connected graph when embedding in two dimensions. This, in turn, forces some edges of the graph to be very long.

Discussion.—As a reminder, an edge of length l implies that there exist a stabilizer measurement involving at least two qubits that are embedded at a distance at least l from each other. We say that such stabilizer has range at least l. If an embedding induces m edges of length l, then, since the codes we consider are LDPC, there exist at least $\Theta(m)$ stabilizers of range at least l.

(1) We focus on the case $D = 2$. The first observation is that a code of distance $\Omega(n^{1/2+\varepsilon})$ will induce $\Omega(n^{1/2+\varepsilon})$ edges of length $\tilde{\Omega}(n^{\varepsilon})$ from claim 1. This underlines how hard it is to break free of the natural restrictions space imposes on the distance: the case $\varepsilon = 0$ can be obtained readily using topological codes and only nearest-neighbor interactions, but $\varepsilon > 0$ will require a significant amount of nonlocality. In particular, implementing a linear distance code will induce $\Omega(n)$ edges of length $\tilde{\Omega}(n^{1/2})$. In that particular case, the length of the edges are tight up to logarithmic factors, since any code can be implemented on a $\sqrt{n} \times \sqrt{n}$ square lattice such that all qubits are at a distance at most $O(n^{1/2})$ from each other. In D dimensions, this result can also be seen as complementing the Bravyi-Terhal claim [8]—if we desire that the code be local, then the longest edges of its connectivity graph have length $O(1)$, the distance must obey $d = \tilde{O}(n^{(D-1)/D})$.

(2) Similarly, our results yield nontrivial bounds on codes with constant rate. First, consider the case with $k \propto n$ and $d \propto 1$. Such a code can be achieved using $\Theta(n)$ disjoint patches of a 2D topological code, and this implementation requires zero nonlocal interactions. However, claim 3 shows that escaping from this constant distance is challenging. For example, achieving $d \propto n^{\alpha}$ requires $\tilde{\Omega}(n)$ interactions of length $\tilde{\Omega}(n^{\alpha/2})$: quite a dramatic change.

(3) The Panteleev-Kalachev codes [24] seem to make optimal use of nonlocality, as they almost saturate claim 3. For example, we could implement $n^{1-\alpha}$ disjoint blocks of good codes, each with size n^{α}. Then we have $k \propto n$,

$d \propto n^{\alpha}$, and at most $O(n)$ edges of length $n^{\alpha/2}$, which minimizes the bound as discussed in the previous point. This suggests that good quantum codes will likely be essential in decreasing the experimental cost of quantum error correction.

(4) There is a gap between the Bravyi *et al.* result and our results with respect to the conditional statement in claim 3. Recall that they stated that if quantum LDPC codes are local, then $kd^{2/(D-1)} = O(n)$. However, we require for claim 3 that $kd^{2/D}$ is roughly greater than $n\log(n)^2$. What are the classes of codes that lie in the gap? Claim 3 itself cannot be sharpened to yield nontrivial bounds on codes satisfying $kd^{2/(D-1)} = \Omega(n)$. Suppose we naively substitute the conditional $\sqrt{k/nd}^{1/D}$ by $\sqrt{k/nd}^{1/(D-1)}$. Then in two dimensions, for any distance larger than $n^{1/2+\varepsilon}$ and constant rate, we would find some edges larger than $n^{1/2+\varepsilon}$. However, this is impossible: we can always place the qubits in a $\sqrt{n}\times\sqrt{n}$ square with edges of length $O(\sqrt{n})$. This seems to imply that, if that substitution worked, there exists no constant-rate quantum LDPC code with a distance larger than $\sqrt{n}$. However, we know this to be false because of the recent result by Panteleev and Kalachev [24].

Application of main theorem to the stacked model.—We return now to the stacked architecture and provide strong evidence that the properties of any code implemented this way will be limited. We begin by describing the model in more detail. Suppose we wish to design an error correcting code using a stacked layout in two dimensions. Consider the following proposal where qubits are laid out on a square grid of size $n = 2^{l_m}\times 2^{l_m}$ as shown in Fig. 1. In total, there are l_m layers in this stack, where the generators at level l act within a ball of radius $r_l = 2^l/\sqrt{2}$. At the very top, we have a highly nonlocal stabilizer associated with a ball of radius $r_{l_m} = 2^{l_m}/\sqrt{2}$. To be clear, while the stabilizer in the topmost layer has a radius of r_{l_m}, it still only jointly measures some constant number of qubits, and each qubit is involved in a constant number of generators. The radius merely constrains where these qubits are allowed to be located. In the next layer we have four stabilizers, but these stabilizers are each only supported within a ball of radius $r_{l-1} = 2^{l-1}/\sqrt{2}$. This proceeds until we hit the very last layer—there are 4^{l_m-l} such generators in layer l—until we hit layer 0, which consists of stabilizers supported entirely within a ball of constant radius. It follows that the majority of the stabilizers are in the last layer, or in other words, the majority of stabilizers are local with $r = O(1)$ locality. A natural question then is whether the nonlocal checks are numerous enough to allow for good codes.

A corollary of our results is that the average length of the interactions in the implementation of a code limits code properties. For example, a family of codes with linear distance requires $\Omega(n)$ edges of length $\tilde{\Omega}(n^{1/2})$. If this system is sparse, then the average length is $\tilde{\Omega}(n^{1/2})$. Conversely, if the average length of the interactions is not $\tilde{\Omega}(n^{1/2})$, then the system cannot implement a family of linear distance codes.

Extending this idea, we can use a direct edge-counting argument together with Theorem 2 to bound the distance and obtain a trade-off between k and d.

Corollary.—The stacked model satisfies $d = n^{2/3}\log(n)^{2/3}$, and $k^3d^4 = O[n^5\log(n)^{10}]$.

The proof is presented in Sec. III of the Supplemental Material [30]. The distance bound immediately implies that this limited amount of nonlocality only yields a limited amount of leeway. The distance of a two-dimensional local code, with this limited nonlocality, is constrained like that of a three-dimensional local code. We do not know if this bound can be saturated, but it does not readily forbid the implementation of constant-rate codes, with $d\propto\sqrt{n}$. The Panteleev-Kalachev codes [22] achieve code dimension and distance that scale as $\Theta(n)$; these codes clearly violate the above bounds. However, it is still not clear whether the codes that do not violate these bounds *can* be implemented via a stacked architecture; our techniques do not rule out this possibility.

Conclusions.—We considered how much nonlocality is needed to implement quantum LDPC codes. In our results, this question is addressed by lower bounding the number of long-range connections between qubits and their length. In particular, in 2D we show that a quantum LDPC code with distance $d\propto n^{1/2+\varepsilon}$ requires $\Omega(n^{1/2+\varepsilon})$ interactions of length $\tilde{\Omega}(n^{\varepsilon})$. We also focus on constant-rate quantum LDPC codes, as the cost of encoding a logical qubit in such a code remains fixed. For such a code to exhibit a distance $d\propto n^{\alpha}$, we find that one requires $\tilde{\Omega}(n)$ interactions of length $\tilde{\Omega}(n^{\alpha/2})$. We then considered a stacked architecture to implement quantum LDPC codes. In this model, although most stabilizers are local, a few are capable of long-range connections. We showed that the distance of this architecture is bounded. Furthermore, it too witnesses a sharp trade-off between k and d. We hope these tools can be used to understand the difficulty of implementing efficient codes, as well as the limitations of particular architectures.

We would like to thank Guillaume Duclos-Cianci for facilitating this collaboration. We thank Anthony Leverrier for his comments on a draft of this Letter. A. K. is supported by the Bloch postdoctoral fellowship at Stanford University, NSF Grants No. CCF-1844628 and No. CCF-1763299. A. K. thanks Emily Davis, Dripto Debroy, and Sam Roberts for pointing him to references on implementations of long-range connectivity.

[1] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error, In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing* (ACM, 1997), pp. 176–188.

[2] A. Y. Kitaev, Quantum computations: Algorithms and error correction, Russ. Math. Surv. **52**, 1191 (1997).

[3] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation: Error models and thresholds, Proc. R. Soc. A **454**, 365 (1998).

[4] P. Aliferis, D. Gottesman, and J. Preskill, Quantum accuracy threshold for concatenated distance-3 codes, Quantum Inf. Comput. **6**, 97 (2006).

[5] J. O'Gorman and E. T. Campbell, Quantum computation with realistic magic-state factories, Phys. Rev. A **95**, 032338 (2017).

[6] Y. R. Sanders, D. W. Berry, P. C. Costa, L. W. Tessler, N. Wiebe, C. Gidney, H. Neven, and R. Babbush, Compilation of fault-tolerant quantum heuristics for combinatorial optimization, PRX Quantum **1**, 020312 (2020).

[7] We often refer to k as the number of *logical* qubits, or as the *dimension* of the code.

[8] S. Bravyi and B. Terhal, A no-go theorem for a two-dimensional self-correcting quantum memory based on stabilizer codes, New J. Phys. **11**, 043029 (2009).

[9] S. Bravyi, D. Poulin, and B. Terhal, Tradeoffs for Reliable Quantum Information Storage in 2D Systems, Phys. Rev. Lett. **104**, 050503 (2010).

[10] L. Bergeron, C. Chartrand, A. Kurkjian, K. Morse, H. Riemann, N. Abrosimov, P. Becker, H.-J. Pohl, M. Thewalt, and S. Simmons, Silicon-integrated telecommunications photon-spin interface, PRX Quantum **1**, 020301 (2020).

[11] H. Bombin, I. H. Kim, D. Litinski, N. Nickerson, M. Pant, F. Pastawski, S. Roberts, and T. Rudolph, Interleaving: Modular architectures for fault-tolerant photonic quantum computing, arXiv:2103.08612.

[12] A. Periwal, E. S. Cooper, P. Kunkel, J. F. Wienand, E. J. Davis, and M. Schleier-Smith, Programmable interactions and emergent geometry in an atomic array, Nature (London) **600**, 630 (2021).

[13] C. Monroe, R. Raussendorf, A. Ruthven, K. Brown, P. Maunz, L.-M. Duan, and J. Kim, Large-scale modular quantum-computer architecture with atomic memory and photonic interconnects, Phys. Rev. A **89**, 022317 (2014).

[14] N. M. Linke, D. Maslov, M. Roetteler, S. Debnath, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, Experimental comparison of two quantum computing architectures, Proc. Natl. Acad. Sci. U.S.A. **114**, 3305 (2017).

[15] P. Murali, D. M. Debroy, K. R. Brown, and M. Martonosi, Architecting noisy intermediate-scale trapped ion quantum computers, In *2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA)* (IEEE, New York, 2020), pp. 529–542.

[16] H. Bombin and M. A. Martin-Delgado, Topological Quantum Distillation, Phys. Rev. Lett. **97**, 180501 (2006).

[17] S. Bravyi and A. Y. Kitaev, Quantum codes on a lattice with boundary, arXiv:quant-ph/9811052.

[18] A. Kubica and M. E. Beverland, Universal transversal gates with color codes: A simplified approach, Phys. Rev. A **91**, 032330 (2015).

[19] D. Gottesman, Fault-tolerant quantum computation with constant overhead, Quantum Inf. Comput. **14**, 1339 (2014).

[20] A. A. Kovalev and L. P. Pryadko, Fault tolerance of quantum low-density parity check codes with sublinear distance scaling, Phys. Rev. A **87**, 020304(R) (2013).

[21] O. Fawzi, A. Grospellier, and A. Leverrier, Constant overhead quantum fault-tolerance with quantum expander codes, In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, New York, 2018), pp. 743–754.

[22] P. Panteleev and G. Kalachev, Quantum LDPC codes with almost linear minimum distance, arXiv:2012.04068.

[23] N. Baspin and A. Krishna, Connectivity constrains quantum codes, Quantum **6**, 711 (2022).

[24] P. Panteleev and G. Kalachev, Asymptotically good quantum and locally testable classical LDPC codes, arXiv:2111.03654.

[25] We recently learned that this model was first proposed by Fernando Pastawski. We heard about it through David Poulin; something similar was also mentioned by Daniel Gottesman [26].

[26] D. Gottesman, Fault tolerance with LDPC codes, Proceedings available at https://www.youtube.com/watch?v=PD4h6ZIV2gY.

[27] M. A. Tremblay, N. Delfosse, and M. E. Beverland, preceding Letter, Constant-Overhead Quantum Error Correction with Thin Planar Connectivity, Phys. Rev. Lett. **129**, 050504 (2022).

[28] E. Knill and R. Laflamme, Concatenated quantum codes, arXiv:quant-ph/9608012.

[29] Wikipedia contributors, Big o notation, https://en.wikipedia.org/wiki/Big_O_notation.

[30] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.129.050505 for Theorem 2 and Corollary 3 in the Main Text, including relevant background material from reference [23].

[31] J. Matoušek, Lecture notes on metric embeddings, Technical Report, ETH Zürich, 2013.

[32] L. Hogben, *Handbook of Linear Algebra, Second Edition*, Discrete Mathematics and Its Applications (Chapman and Hall/CRC, Boca Raton, FL, 2013).

[33] J. H. S. Noga Alon, *The Probabilistic Method*, 4th ed., Wiley Series in Discrete Mathematics and Optimization (John Wiley & Sons, New York, 2016).

[34] J. Böttcher, K. P. Pruessmann, A. Taraz, and A. Würfl, Bandwidth, expansion, treewidth, separators and universality for bounded-degree graphs, Eur. J. Combinatorics **31**, 1217 (2010).