

Constant-Overhead Fault-Tolerant Bell-Pair Distillation Using High-Rate Codes

J. Pablo Bonilla Ataides^{1,*}, Hengyun Zhou², Qian Xu^{3,4}, Gefen Baranes^{1,5},
Bikun Li⁶, Mikhail D. Lukin^{1,†} and Liang Jiang^{6,‡}

¹*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

²*QuEra Computing Incorporated, Boston, Massachusetts 02135, USA*

³*Institute for Quantum Information and Matter, Caltech, Pasadena, California 91125, USA*

⁴*Walter Burke Institute for Theoretical Physics, Caltech, Pasadena, California 91125, USA*

⁵*Department of Physics and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

⁶*Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, USA*



(Received 26 February 2025; accepted 18 August 2025; published 25 September 2025)

We present a fault-tolerant Bell-pair distillation scheme achieving constant overhead through high-rate quantum low-density parity-check (qLDPC) codes. Our approach maintains a constant distillation rate equal to the code rate while requiring no additional overhead beyond the physical qubits of the code. Full circuit-level analysis demonstrates fault-tolerance for input Bell-pair infidelities below a threshold $\sim 10\%$, readily achievable with near-term capabilities. Unlike previous proposals, our scheme keeps the output Bell pairs encoded in qLDPC codes at each node, eliminating unencoding overhead and enabling direct use in distributed quantum applications through recent advances in qLDPC computation. These results establish qLDPC-based distillation as a practical route toward resource-efficient quantum networks and distributed quantum computing.

DOI: [10.1103/s39k-r2kq](https://doi.org/10.1103/s39k-r2kq)

As quantum systems scale, establishing high-fidelity connections between nodes will be essential for distributed quantum computation, communication, and sensing [1–7] (see Fig. 1). A promising approach to achieving this is through shared Bell pairs, where one qubit of each pair is located at each node. Bell pairs provide nonlocal entanglement, enabling quantum information transfer through local operations and classical communication.

However, raw Bell pairs generated by current hardware are too noisy for practical applications. Experiments can produce Bell pairs with infidelities of $\sim 5\%$ [8–12], whereas distributed algorithms may require error rates below 10^{-10} . To bridge this gap, entanglement distillation—also known as purification—is used to convert multiple noisy Bell pairs into fewer, higher-fidelity pairs [13].

A practical Bell-pair distillation scheme should satisfy several key criteria. It must be scalable, which can be achieved by maintaining constant overhead. It should also

be fault tolerant, ensuring robustness against both network errors and local gate errors. Additionally, it should exhibit error thresholds for Bell-pair infidelity and local gate errors that are both feasible and achievable with near-term hardware. Finally, the scheme should minimize additional resource requirements, such as classical communication costs, code decoding complexity, and memory overhead from postselection in nondeterministic schemes.

Topological codes are among the leading proposals for fault-tolerant quantum computing, offering high thresholds, local low-weight checks, and compatibility across multiple architectures [14–20]. However, these codes encode only a constant number of logical qubits, resulting in a vanishing asymptotic rate. As a result, distillation schemes relying on topological codes, such as lattice surgery schemes [21–25], require a large number of physical Bell pairs to achieve the low error rates demanded by logical algorithms.

Quantum low-density parity-check (qLDPC) codes provide a promising approach for constant-rate distillation [26–30], but existing proposals lack a full fault-tolerant analysis and typically decode to physical Bell pairs, limiting robustness to local gate errors. To the best of our knowledge, no qLDPC distillation proposal has yet been analyzed at the full circuit-noise fault-tolerance level. Reference [26] suggests a fault-tolerant encoder for qLDPC codes, similar to the approach in this Letter, but does not study specific codes or conduct threshold simulations. The use of qLDPC codes for state distillation has also been

*Contact author: jbonillaataides@g.harvard.edu

†Contact author: lukin@physics.harvard.edu

‡Contact author: liang.jiang@uchicago.edu

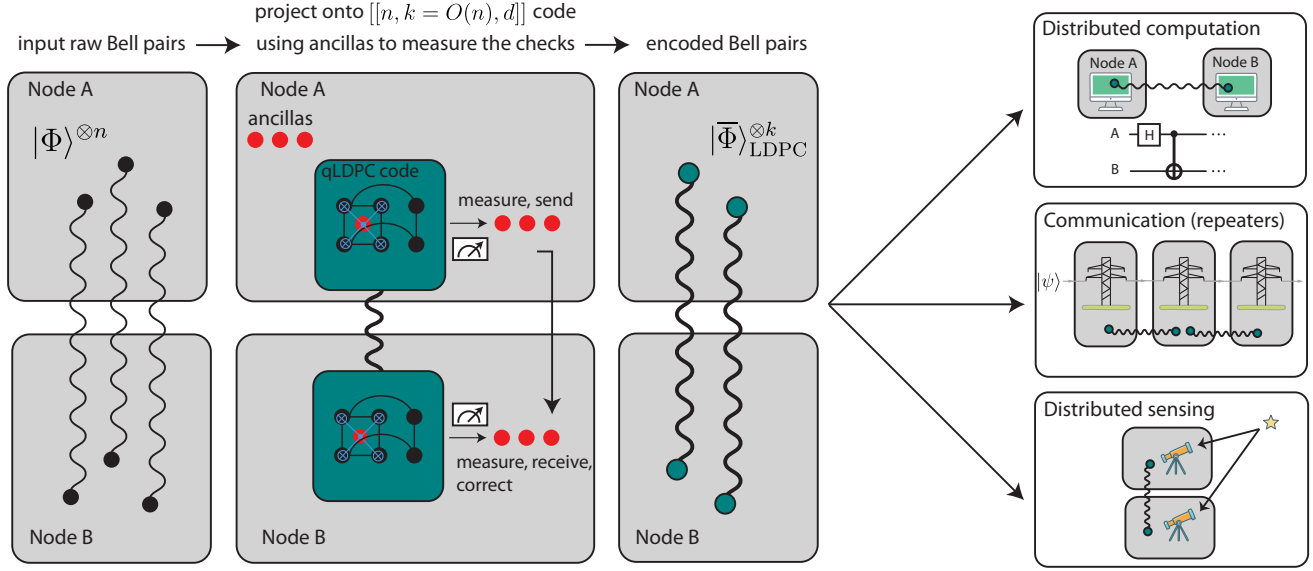


FIG. 1. Constant-overhead Bell-pair distillation using qLDPC codes. Alice and Bob share n initial raw Bell pairs. They measure the checks of a qLDPC code on each side using local ancilla qubits. Alice communicates her measurement results to Bob, who receives them and performs error correction, projecting the Bell pairs into the code state. This process results in a constant distillation rate equal to the rate of the qLDPC code, requiring only one-way communication. By performing logical operations on qLDPC codes, the encoded Bell pairs can be used for various applications, including distributed computation, communication via repeaters, and distributed sensing.

proposed in Ref. [28], though in a different context; to distill GHZ states for distributed quantum error correction (QEC). Similarly, Ref. [29] explores hyperbolic Floquet codes, implementing them nonlocally by distributing qubits across multiple nodes. In contrast, we consider a different approach, assuming local nodes large enough to implement QEC internally [19,20].

An alternative scheme for achieving constant-rate distillation relies on QEC codes for error detection, where errors are flagged and discarded instead of corrected. Recent work has demonstrated constant-rate distillation using code concatenation [31,32] and via scrambling [33]. However, these error-detection schemes are nondeterministic and require two-way classical communication for postselection. Additionally, this approach may require additional buffer memory overhead, otherwise, faulty pairs will bottleneck the execution by requiring one to wait until the next distillation round completes.

In this Letter, we present a scheme that satisfies all the key requirements for practical, constant-overhead entanglement distillation. We compare our method with other leading distillation schemes across multiple performance metrics in Table I. Our approach utilizes constant-rate qLDPC codes, following a stabilizer protocol [34,35], where code checks are measured to project Bell pairs onto the encoded code state (see Fig. 1).

We consider codes that are feasible for implementation in reconfigurable atom arrays [19,50–52] and possess high rates. We perform a full fault-tolerant circuit-level analysis

of our scheme and observe high thresholds achievable with near-term hardware [53]. Furthermore, recent advances in qLDPC gates enable us to leave the output Bell pairs encoded in the code. This approach eliminates the extra unencoding stage required in many current distillation schemes, allowing us to achieve distillation fidelities that are not limited by gate errors.

Ancilla-assisted encoding—Now, we describe the protocol used to distill Bell pairs using constant-rate qLDPC codes, as illustrated in Fig. 1. The goal of entanglement distillation is to take noisy input Bell pairs, $|\Phi^+\rangle^{\otimes n}$, and output higher-fidelity Bell pairs, $|\bar{\Phi}^+\rangle^{\otimes k}$, where $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ and $k \leq n$.

This process can be associated with a QEC code, where redundancy is used to achieve lower error rates [13,49]. A quantum code is labeled as $[[n, k, d]]$, where n is the number of physical qubits, k is the number of encoded logical qubits, and d is the distance of the code. A code with distance d can detect up to $d - 1$ errors and correct up to $\lfloor (d - 1)/2 \rfloor$ errors. The code rate, defined as $R = k/n$, quantifies the encoding efficiency of the code. If the QEC scheme is fault tolerant, there exists a threshold physical error rate below which increasing the code distance can exponentially suppress the error rate of the encoded logical qubits, enabling arbitrarily low error rates.

A stabilizer quantum code is specified by stabilizer generators $\mathcal{S} = \langle g_1, \dots, g_m \rangle$, where each $g_i \in \mathcal{P}^n$ is an n -qubit Pauli operator. The code space of the code is the simultaneous $+1$ eigenspace of the stabilizers of the code,

TABLE I. Summary of selected Bell-pair distillation methods. Overhead refers to the asymptotic number of input physical Bell pairs required per output logical Bell pair, assuming perfect local operations. Classical communication indicates whether the method relies on error detection (requiring two-way classical communication) or error correction (necessitating only one-way communication). Local gate error tolerance specifies the robustness of the protocol against local gate errors. Protocols that directly decode to physical Bell pairs are not robust against local gate errors. Encoding resources refer to any additional resources required by the protocol. Threshold represents the threshold against network errors of the scheme. Our scheme is simulated at the full circuit-noise level, assuming a local gate error rate of 0.1%. For further details on each method, refer to the Supplemental Material [36]. Bennett-DiVincenzo-Smolin-Wootters protocol (BDSW) and Entanglement Purification Protocol (EPP).

Method	Overhead	Classical communication	Local gate error tolerance	Encoding resources	Threshold
BDSW-2EPP [49]	$> O(1)$	Two-way	✗	Nondeterministic, requires buffer memory	50% with perfect local ops.
BDSW-1EPP [49]	$O(1)$	One-way	✗	Requires decoding random quantum codes	Not practical
Lattice surgery [22–25]	$> O(1)$	One-way	✓	$O(d)$ time ^b	~10%
Shi <i>et al.</i> [26]	$O(1)$	Two-way	✗	$O(1)$	Not studied
Pattison <i>et al.</i> [31]	$O(1)$	Two-way	✓	Nondeterministic, requires buffer memory	50% with perfect local ops. ^a
This Letter	$O(1)$	One-way	✓	$O(1)$	~10% full circuit-noise simulation

^aImperfect local operations reduce performance only slightly.

^bAlthough the primary scheme analyzed in these works involves lattice surgery at the surface code boundary, Ref. [23] suggests that the results extend to transversal gates, where the time cost can be reduced to $O(1)$.

satisfying $g|\bar{\psi}\rangle = (+1)|\bar{\psi}\rangle$ for all $g \in \mathcal{S}$. We focus on CSS codes, whose stabilizer generators are divided into X -type ($g_x \in \{X, I\}^{\otimes n}$) and Z -type ($g_z \in \{Z, I\}^{\otimes n}$).

Quantum low-density parity-check codes are those in which each check is supported on a constant number of qubits, and each qubit participates in a constant number of checks. This bounded check and qubit degree ensures that the syndrome extraction circuit remains constant depth, facilitating the fault tolerance of the QEC protocol. When qLDPC codes are used to distill Bell pairs, the distillation rate is constant and equal to the code rate. Although the checks of qLDPC codes can be highly nonlocal [54–56], various schemes and architectures have been developed to implement them efficiently [50,57–60].

To perform error correction, the stabilizers (or checks) are measured and a decoding algorithm is used to infer the error corresponding to any -1 measurement outcomes. These -1 outcomes collectively form the syndrome of the code. A single ancilla per check can be used to measure the value of each check by entangling the ancilla with the data qubits involved in the check and then measuring the ancilla. Single-ancilla syndrome extraction circuits can be made fault tolerant due to the bounded weight of the checks and, further, can be made distance preserving for some code families [50,61,62].

By measuring the stabilizers of the code on the qubits of the Bell pairs at each node, Alice and Bob can project the Bell pairs into the code space. Specifically, Alice uses local ancilla qubits to measure the checks of the qLDPC code on her side and sends the syndrome to Bob through a classical channel. Then, Bob measures the checks of the code on his

side and adds his measurements to the syndrome received from Alice. After Bob performs error correction on the joint checks, the resultant state consists of k encoded Bell pairs in the code space of the qLDPC code. A rigorous analysis of this procedure, along with a proof of the fault tolerance of the protocol, is provided in the End Matter.

Error modeling—We perform a full fault-tolerant analysis by simulating the encoding of Bell pairs into qLDPC codes, incorporating both Bell-pair and local gate errors—extending beyond prior studies that typically assume only depolarizing noise on input pairs.

Each Bell pair undergoes a depolarizing channel with strength p_{bell} , while local two-qubit gates in the error correction circuit experience depolarizing noise with strength p . To simplify the simulation, we leverage symmetry in Bell-pair errors to shift all noise onto Bob’s side, leaving Alice’s side noiseless. This approximation allows us to simulate only Bob’s subsystem while preserving the overall logical error behavior. In the End Matter and Supplemental Material [36], we show that the increased noise on Bob’s side can be approximated as $p'_{\text{bell}} \approx 2p_{\text{bell}}$ and $p' \approx 2p$. Following convention in prior work [23,28,30,31,33], we quote p'_{bell} as the input Bell error rate. This convention aligns with experimental setups where Alice generates the Bell pairs at her node, and only the qubit sent to Bob experiences the noisy fiber channel.

Numerical simulations—We perform full fault-tolerant circuit-noise simulations using three different classes of qLDPC codes: hypergraph product (HGP) codes with a rate of ~4% [63], quasicyclic lifted product (LP) codes with a rate of ~11% [64], and a spatially coupled (SC) code with a

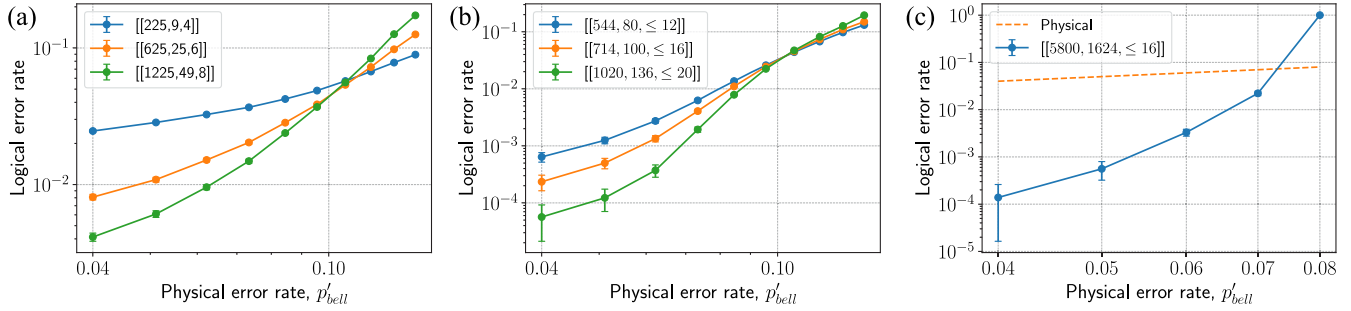


FIG. 2. Fault-tolerant simulation of qLDPC distillation. The Bell pairs experience input depolarizing noise with effective strength p'_{bell} . The circuit undergoes two-qubit gate depolarizing errors at a rate of $p' = 2 \times 0.1\%$. We simulate 14 decoding rounds with 3 cycles of syndrome extraction per round (42 cycles in total) and decode each round with BP, using BP + OSD for the final round. (a)–(c) Numerical results for the HGP, LP, and SC codes, respectively. We find a threshold $p'_{\text{bell}} \sim 10\%$ for the HGP and LP code families and a pseudothreshold of $p_{\text{bell}} \sim 7.5\%$ for the SC code. We calculate the logical error rate for the entire block per cycle, using $\bar{P} = 1 - (1 - \bar{P}_{\text{tot}})^{1/42}$, where $\bar{P}_{\text{tot}}$ is the total error after 42 cycles. Error bars are calculated assuming a binomial distribution, with N shots, giving $\sqrt{P_{\text{tot}}(1 - P_{\text{tot}})/N}$.

rate of $\sim 1/3$ [65,66]. To the best of our knowledge, a distillation rate of $1/3$ is the highest rate achieved by any deterministic and efficiently decodable distillation protocol to date [67]. Further details on code constructions can be found in the End Matter.

The fault-tolerant simulations include multiple rounds of the syndrome extraction circuit, followed by a final transversal measurement of the data qubits of the code. To extract the values of the qLDPC checks, we use a single ancilla qubit, which is entangled with the data qubits in the support of each check. A measurement of the ancilla qubit then provides the syndrome information used for decoding. Fault tolerance is maintained by scheduling the entangling gates between the ancilla and data qubits in parallel, following the coloration circuit approach [50,61]. We note that accounting for movement errors required to implement nonlocal checks on a reconfigurable atom array platform introduces negligible additional errors, as idling error is low and compact rearrangement schemes have been developed [36,50].

To decode, we turn the circuit into a spacetime code and use the belief propagation with ordered statistics decoding (BP+OSD) decoding algorithm [50,68–70]. The checks of the spacetime code correspond to products of stabilizers at subsequent time steps, and the “qubits” of the spacetime code represent fault locations in the circuit. We leverage the single-shot property of the codes, which requires only $O(1)$ cycles of QEC per decoding round to achieve fault tolerance against measurement errors [50,71]. Although, in principle, a single QEC cycle per round is sufficient for fault tolerance, we implement three cycles per round to provide additional redundancy against measurement errors. Our simulations consist of 14 rounds of three cycles each, totaling 42 cycles. We decode each round with BP only, and to project back into the code space, we decode the final round with BP + OSD. The circuits are constructed in Stim

[72], and we use the BP + OSD decoder implementation provided in the LDPC library [73].

For the error model, we assume local gate errors with strength $p = 0.1\%$, leading to an effective error strength on Bob’s side of $p' = 2 \times 0.1\%$ for the two-qubit gates in the syndrome extraction circuits. We vary the input Bell error p'_{bell} and calculate the logical error rate of the qLDPC block per cycle. The results are presented in Fig. 2. We observe a threshold of $p'_{\text{bell}} \sim 10\%$ for the HGP and LP code families and a pseudothreshold of $p_{\text{bell}} \sim 7.5\%$ for the SC code. Since we plot the block error rate, meaning that a failure is declared if any of the logical qubits of the qLDPC code fails, the error rates per logical qubit can be substantially lower than the values shown in Fig. 2.

The plateau observed in the logical error rate at low values of p'_{bell} arises from the use of a fixed circuit-level gate error. Instead, if we allow the local gate noise to vary along with the Bell error rate, this plateau disappears, enabling us to study the subthreshold scaling behavior of the codes. In the End Matter, Fig. 5 and Table II, we examine the subthreshold scaling of the codes and tabulate the achievable error rates for various instances. For example, we find, by extrapolation, that the SC code can achieve a logical error rate of $\sim 2 \times 10^{-12}$ at input physical error rate of 1% .

Outlook—We presented and analyzed a one-way constant-rate distillation scheme based on qLDPC error correction that achieves high fault-tolerant thresholds, rates as high as $1/3$, and requires no additional overhead beyond the qubits of the qLDPC code. Leaving the Bell pairs encoded at each node eliminates the unencoding step required in many existing schemes and ensures protection against local errors for future computations.

However, performing local operations on these encoded Bell pairs requires the ability to manipulate qLDPC-encoded qubits. In this direction, numerous resource-

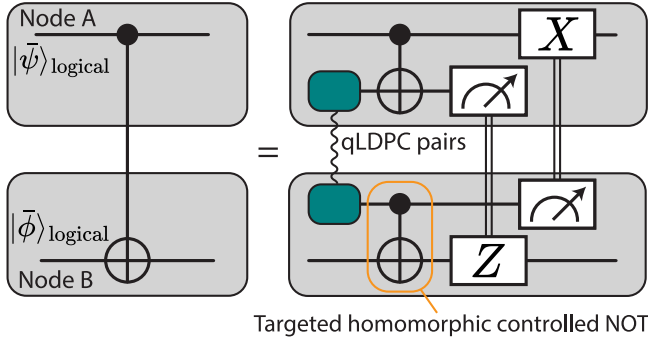


FIG. 3. Distributed CNOT using qLDPC-distilled Bell pairs. The distributed CNOT gate can be implemented using local targeted CNOT operations between logical qubits and one of the k encoded Bell pairs of the qLDPC code. These targeted CNOT gates can be performed using lattice surgery techniques or homomorphic gadgets. In product codes such as HGP codes, transversal approaches using homomorphic gadgets provide a constant-overhead implementation compatible with reconfigurable neutral atom architectures.

efficient schemes have recently been proposed to enable selective computation with qLDPC codes [74–81].

As an example, consider the implementation of a distributed controlled NOT (CNOT) gate, as illustrated in Fig. 3. This gate can be realized using the high-fidelity Bell pairs distilled with the qLDPC code and local logical CNOT gates. The local logical CNOTs must target a single Bell pair among the k encoded pairs in the qLDPC code.

This targeted CNOT operation can be implemented in various ways. Lattice surgery techniques [50,75,76,78,82–84] provide a viable approach, while certain product codes, such as HGP codes, enable a more efficient implementation via homomorphic gadgets [74]. The latter maintains constant overhead and is naturally compatible with the movement capabilities of reconfigurable atom arrays [50,74]. In the proposed implementation, the error rate of the distributed CNOT gate is primarily determined by the error rate of the qLDPC-encoded Bell pairs, as this is the only stage where nonlocal (and noisy) physical elements are involved. Thus, the results of Fig. 2 are also applicable to the distributed CNOT gadget shown in Fig. 3. We note that, while the distillation protocol itself requires only one-way communication for error correction, distributed entangling operations, such as the one illustrated in Fig. 3, may practically necessitate two-way communication, for example, to implement feedforward Pauli corrections.

A challenge that future distillation protocols may face is the rate of Bell-pair generation. Since the codes required for error correction often demand numerous Bell pairs for encoding, one might naively wait for all Bell pairs to be available before starting the encoding process. This idling time can introduce additional errors on the Bell pairs. Convolutional codes, such as the SC code, offer a potential solution by allowing

encoding to proceed in smaller segments, each of which is itself a code. This structure enables on-the-fly encoding, which may help mitigate these idling errors.

Focusing on distributed computation, we note that our protocol is compatible with the recent paradigm of algorithmic fault tolerance (AFT) for fast quantum computation [85,86]. By leveraging AFT and the transversal operations available in qLDPC codes [74], we can achieve a factor of $O(d)$ reduction in time cost per transversal logical operation compared to schemes based on lattice surgery [23,24], which is relevant for the execution of resource-efficient distributed algorithms.

With near-term Bell pair generation rates and infidelities expected to reach values of 10^5 s^{-1} and $p_{\text{bell}} = 0.1\%$, respectively [53], qLDPC-based Bell-pair distillation presents a promising approach for high-fidelity quantum interconnects.

Exploring hybrid strategies that first reduce physical Bell-pair error rates via standard distillation before qLDPC encoding, as well as schemes that interpolate flexibly between error detection and error correction using decoder confidence metrics, are compelling directions for further research that may enhance practical performance.

Finally, we emphasize that our distillation protocol is broadly applicable, being compatible with any quantum computing architecture capable of implementing the long-range connectivity required by high-rate qLDPC codes [57,87–91].

Acknowledgments—We acknowledge helpful discussions with Andi Gu, Senrui Chen, Christopher Pattison, Siyi Yang, Dolev Bluvstein, Nishad Maskara, Marcello Laurel, Rohan Mehta, Patrick Rall, Josiah Sinclair, Nazli Ugur Koyluoglu, and Varun Menon. We acknowledge support from the ARO (Grant No. W911NF-23-1-0077), ARO MURI (Grants No. W911NF-21-1-0325 and No. W911NF-20-1-0082), AFOSR MURI (Grants No. FA9550-19-1-0399, No. FA9550-21-1-0209, and No. FA9550-23-1-0338), DARPA (Grants No. HR0011-24-9-0359 and No. HR0011-24-9-0361), NSF (Grants No. OMA-1936118, No. ERC-1941583, No. OMA-2137642, No. OSI-2326767, No. CCF-2312755, No. PHY-2012023 and No. CCF-2313084), NTT Research, Packard Foundation (Grant No. 2020-71479), the Marshall and Arlene Bennett Family Research Program, IARPA and the ARO, under the Entangled Logical Qubits Program (W911NF-23-2-0219), the Center for Ultracold Atoms (a NSF Physics Frontiers Center, PHY-1734011), and the NSF Center for Quantum networks (Grant No. EEC-1941583). This material is based upon work supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers and Advanced Scientific Computing Research (ASCR)

Program under Contract No. DE-AC02-06CH11357 as part of the InterQnet quantum networking project. Q. X. is funded in part by the Walter Burke Institute for Theoretical Physics at Caltech. G. B. acknowledges support from the MIT Patrons of Physics Fellows Society.

Data availability—The data that support the findings of this Letter are not publicly available. The data are available from the authors upon reasonable request.

-
- [1] C. Monroe, R. Raussendorf, A. Ruthven, K. R. Brown, P. Maunz, L.-M. Duan, and J. Kim, *Phys. Rev. A* **89**, 022317 (2014).
 - [2] S. Muralidharan, L. Li, J. Kim, N. Lütkenhaus, M. D. Lukin, and L. Jiang, *Sci. Rep.* **6**, 20463 (2016).
 - [3] W. Ge, K. Jacobs, Z. Eldredge, A. V. Gorshkov, and M. Foss-Feig, *Phys. Rev. Lett.* **121**, 043604 (2018).
 - [4] T. J. Proctor, P. A. Knott, and J. A. Dunningham, *Phys. Rev. Lett.* **120**, 080501 (2018).
 - [5] Q. Zhuang, Z. Zhang, and J. H. Shapiro, *Phys. Rev. A* **97**, 032329 (2018).
 - [6] Z. Eldredge, M. Foss-Feig, J. A. Gross, S. L. Rolston, and A. V. Gorshkov, *Phys. Rev. A* **97**, 042337 (2018).
 - [7] Z. Yang, A. Ghubaish, R. Jain, H. Shapourian, and A. Shabani, *AVS Quantum Sci.* **6**, 013801 (2024).
 - [8] L. J. Stephenson, D. P. Nadlinger, B. C. Nichol, S. An, P. Drmota, T. G. Ballance, K. Thirumalai, J. F. Goodwin, D. M. Lucas, and C. J. Ballance, *Phys. Rev. Lett.* **124**, 110501 (2020).
 - [9] M. Mirhosseini, A. Sipahigil, M. Kalaei, and O. Painter, *Nature (London)* **588**, 599 (2020).
 - [10] B. Jing, X.-J. Wang, Y. Yu, P.-F. Sun, Y. Jiang, S.-J. Yang, W.-H. Jiang, X.-Y. Luo, J. Zhang, X. Jiang *et al.*, *Nat. Photonics* **13**, 210 (2019).
 - [11] C. M. Knaut, A. Suleymanzade, Y.-C. Wei, D. R. Assumpcao, P.-J. Stas, Y. Q. Huan, B. Machielse, E. N. Knall, M. Sutula, G. Baranes *et al.*, *Nature (London)* **629**, 573 (2024).
 - [12] S. Meesala, D. Lake, S. Wood, P. Chiappina, C. Zhong, A. D. Beyer, M. D. Shaw, L. Jiang, and O. Painter, *Phys. Rev. X* **14**, 031055 (2024).
 - [13] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, *Phys. Rev. Lett.* **76**, 722 (1996).
 - [14] A. Y. Kitaev, *Ann. Phys. (N.Y.)* **303**, 2 (2003).
 - [15] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, *Phys. Rev. A* **86**, 032324 (2012).
 - [16] J. P. Bonilla Ataides, D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, *Nat. Commun.* **12**, 1 (2021).
 - [17] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas, M. Newman, M. Li, K. R. Brown, M. Cetina *et al.*, *Nature (London)* **598**, 281 (2021).
 - [18] Google Quantum AI, *Nature (London)* **614**, 676 (2023).
 - [19] D. Bluvstein *et al.*, *Nature (London)* **626**, 58 (2023).
 - [20] Google Quantum AI, *Nature (London)* **638**, 920 (2025).
 - [21] D. Litinski, *Quantum* **3**, 128 (2019).
 - [22] A. G. Fowler, D. S. Wang, C. D. Hill, T. D. Ladd, R. Van Meter, and L. C. L. Hollenberg, *Phys. Rev. Lett.* **104**, 180503 (2010).
 - [23] J. Ramette, J. Sinclair, N. P. Breuckmann, and V. Vuletić, [arXiv:2302.01296](https://arxiv.org/abs/2302.01296).
 - [24] J. Sinclair, J. Ramette, B. Grinkemeyer, D. Bluvstein, M. Lukin, and V. Vuletić, [arXiv:2408.08955](https://arxiv.org/abs/2408.08955).
 - [25] H. Leone, S. Srikara, P. P. Rohde, and S. Devitt, [arXiv:2209.00151](https://arxiv.org/abs/2209.00151).
 - [26] Y. Shi, A. Patil, and S. Guha, [arXiv:2408.06299](https://arxiv.org/abs/2408.06299).
 - [27] N. P. Breuckmann and J. N. Eberhardt, *PRX Quantum* **2**, 040101 (2021).
 - [28] N. Rengaswamy, N. Raveendran, A. Raina, and B. Vasić, *Quantum* **8**, 1233 (2024).
 - [29] E. Sutcliffe, B. Jonnadula, C. Le Gall, A. E. Moylett, and C. M. Westoby, [arXiv:2501.14029](https://arxiv.org/abs/2501.14029).
 - [30] Y. Shi, A. Patil, and S. Guha, preceding Letter, *Phys. Rev. Lett.* **135**, 130803 (2025).
 - [31] C. A. Pattison, G. Baranes, J. Ataides, M. D. Lukin, and H. Zhou, [arXiv:2408.15936](https://arxiv.org/abs/2408.15936).
 - [32] H. Yamasaki and M. Koashi, *Nat. Phys.* **20**, 247 (2024).
 - [33] A. Gu, L. Leone, K. Goodenough, and S. Khatri, [arXiv:2502.09483](https://arxiv.org/abs/2502.09483).
 - [34] T. Brun, I. Devetak, and M.-H. Hsieh, *Science* **314**, 436 (2006).
 - [35] M. M. Wilde, H. Krovi, and T. A. Brun, *Proceedings of the 2010 IEEE International Symposium on Information Theory* (2010), p. 2657.
 - [36] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/s39k-r2kq> for more details, which includes Refs. [37–48].
 - [37] M. Murao, M. B. Plenio, S. Popescu, V. Vedral, and P. L. Knight, *Phys. Rev. A* **57**, R4075 (1998).
 - [38] K. Vollbrecht and F. Verstraete, *Phys. Rev. A* **71**, 062325 (2005).
 - [39] N. Isailovic, Y. Patel, M. Whitney, and J. Kubiawicz, [arXiv:quant-ph/0604048](https://arxiv.org/abs/quant-ph/0604048).
 - [40] E. Hostens, J. Dehaene, and B. De Moor, *Phys. Rev. A* **73**, 062337 (2006).
 - [41] A. W. Leung and P. W. Shor, [arXiv:quant-ph/0702155](https://arxiv.org/abs/quant-ph/0702155).
 - [42] P. Hayden, M. Horodecki, A. Winter, and J. Yard, *Open Syst. Inf. Dyn.* **15**, 7 (2008).
 - [43] A. W. Leung, *Phys. Rev. A* **77**, 012322 (2008).
 - [44] S. Krastanov, V. V. Albert, and L. Jiang, *Quantum* **3**, 123 (2019).
 - [45] C. Gidney, [arXiv:2311.10971](https://arxiv.org/abs/2311.10971).
 - [46] K. Goodenough, A. Sajjad, E. Kaur, S. Guha, and D. Towsley, [arXiv:2406.02427](https://arxiv.org/abs/2406.02427).
 - [47] V. Siddhu, D. Abdelhadi, T. Jochym-O'Connor, and J. Smolin, [arXiv:2405.06231](https://arxiv.org/abs/2405.06231).
 - [48] C. Wang, J. Harrington, and J. Preskill, *Ann. Phys. (N.Y.)* **303**, 31 (2003).
 - [49] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, *Phys. Rev. A* **54**, 3824 (1996).
 - [50] Q. Xu, J. P. Bonilla Ataides, C. A. Pattison, N. Raveendran, D. Bluvstein, J. Wurtz, B. Vasić, M. D. Lukin, L. Jiang, and H. Zhou, *Nat. Phys.* **20**, 1084 (2024).
 - [51] D. Bluvstein, H. Levine, G. Semeghini, T. T. Wang, S. Ebadi, M. Kalinowski, A. Keesling, N. Maskara, H. Pichler, M. Greiner, V. Vuletić, and M. D. Lukin, *Nature (London)* **604**, 451 (2022).
 - [52] S. J. Evered, D. Bluvstein, M. Kalinowski, S. Ebadi, T. Manovitz, H. Zhou, S. H. Li, A. A. Geim, T. T. Wang,

- N. Maskara, H. Levine, G. Semeghini, M. Greiner, V. Vuletić, and M. D. Lukin, *Nature (London)* **622**, 268 (2023).
- [53] Y. Li and J. D. Thompson, *PRX Quantum* **5**, 020363 (2024).
- [54] N. Baspin and A. Krishna, *Quantum* **6**, 711 (2022).
- [55] N. Baspin and A. Krishna, *Phys. Rev. Lett.* **129**, 050505 (2022).
- [56] N. Baspin, V. Guruswami, A. Krishna, and R. Li, *Quantum Sci. Technol.* **10**, 015021 (2024).
- [57] S. Bravyi, A. W. Cross, J. M. Gambetta, D. Maslov, P. Rall, and T. J. Yoder, *Nature (London)* **627**, 778 (2024).
- [58] S. Stein, S. Xu, A. W. Cross, T. J. Yoder, A. Javadi-Abhari, C. Liu, K. Liu, Z. Zhou, C. Guinn, Y. Ding, Y. Ding, and A. Li, *arXiv:2411.03202*.
- [59] J. Viszlai, W. Yang, S. F. Lin, J. Liu, N. Nottingham, J. M. Baker, and F. T. Chong, *arXiv:2311.16980*.
- [60] C. Poole, T. M. Graham, M. A. Perlin, M. Otten, and M. Saffman, *Phys. Rev. A* **111**, 022433 (2025).
- [61] M. A. Tremblay, N. Delfosse, and M. E. Beverland, *Phys. Rev. Lett.* **129**, 050504 (2022).
- [62] A. G. Manes and J. Claes, *arXiv:2308.15520*.
- [63] J. P. Tillich and G. Zemor, *IEEE Trans. Inf. Theory* **60**, 1193 (2014).
- [64] N. Raveendran, N. Rengaswamy, F. Rozpędek, A. Raina, L. Jiang, and B. Vasić, *Quantum* **6**, 767 (2022).
- [65] S. Yang and R. Calderbank, *arXiv:2305.00137*.
- [66] M. Hagiwara, K. Kasai, H. Imai, and K. Sakaniwa, in *Proceedings of the 2011 IEEE International Symposium on Information Theory Proceedings* (IEEE, New York, 2011), pp. 638–642.
- [67] The quoted rates for the LP and SC codes correspond to finite-size instances. For further details, see the End Matter and Supplemental Material [36].
- [68] P. Panteleev and G. Kalachev, *Quantum* **5**, 585 (2019).
- [69] J. Roffe, D. R. White, S. Burton, and E. Campbell, *Phys. Rev. Res.* **2**, 043423 (2020).
- [70] O. Higgott and N. P. Breuckmann, *PRX Quantum* **4**, 020332 (2023).
- [71] A. O. Quintavalle, M. Vasmer, J. Roffe, and E. T. Campbell, *PRX Quantum* **2**, 020340 (2021).
- [72] C. Gidney, *Quantum* **5**, 497 (2021).
- [73] J. Roffe, LDPC: Python tools for low density parity check codes (2022).
- [74] Q. Xu, H. Zhou, G. Zheng, D. Bluvstein, J. Ataides, M. D. Lukin, and L. Jiang, *arXiv:2407.18490*.
- [75] A. Cross, Z. He, P. Rall, and T. Yoder, *arXiv:2407.18393*.
- [76] D. J. Williamson and T. J. Yoder, *arXiv:2410.02213*.
- [77] N. P. Breuckmann and S. Burton, *Quantum* **8**, 1372 (2024).
- [78] E. Swaroop, T. Jochym-O'Connor, and T. J. Yoder, *arXiv:2410.03628*.
- [79] N. P. Breuckmann, M. Davydova, J. N. Eberhardt, and N. Tantivasadakarn, *arXiv:2410.16250*.
- [80] Z. He, V. Vaikuntanathan, A. Wills, and R. Y. Zhang, *arXiv:2502.01864*.
- [81] A. J. Malcolm, A. N. Glaudell, P. Fuentes, D. Chandra, A. Schotte, C. DeLisle, R. Haenel, A. Ebrahimi, J. Roffe, A. O. Quintavalle, S. J. Beale, N. R. Lee-Hone, and S. Simmons, *arXiv:2502.07150*.
- [82] L. Z. Cohen, I. H. Kim, S. D. Bartlett, and B. J. Brown, *Sci. Adv.* **8**, eabn1717 (2022).
- [83] B. Ide, M. G. Gowda, P. J. Nadkarni, and G. Dauphinais, *arXiv:2410.02753*.
- [84] A. Cross, Z. He, P. Rall, and T. Yoder, *arXiv:2407.18393*.
- [85] M. Cain, C. Zhao, H. Zhou, N. Meister, J. P. Bonilla Ataides, A. Jaffe, D. Bluvstein, and M. D. Lukin, *arXiv:2403.03272*.
- [86] H. Zhou, C. Zhao, M. Cain, D. Bluvstein, C. Duckering, H.-Y. Hu, S.-T. Wang, A. Kubica, and M. D. Lukin, *arXiv:2406.17653*.
- [87] M. Ye and N. Delfosse, *arXiv:2503.22071*.
- [88] N. Berthussen, D. Devulapalli, E. Schoute, A. M. Childs, M. J. Gullans, A. V. Gorshkov, and D. Gottesman, *PRX Quantum* **6**, 010306 (2025).
- [89] C. A. Pattison, A. Krishna, and J. Preskill, *Quantum* **9**, 1728 (2025).
- [90] T. J. Yoder, E. Schoute, P. Rall, E. Pritchett, J. M. Gambetta, A. W. Cross, M. Carroll, and M. E. Beverland, *arXiv:2506.03094*.
- [91] K. Wang *et al.*, *arXiv:2505.09684*.
- [92] A. Grospellier, L. Grouès, A. Krishna, and A. Leverrier, *Quantum* **5**, 432 (2021).
- [93] N. P. Breuckmann and J. N. Eberhardt, *IEEE Trans. Inf. Theory* **67**, 6653 (2020).

End Matter

qLDPC encoder—To demonstrate that the ancilla-assisted procedure described in the main text correctly encodes onto the qLDPC code, let the stabilizers of the qLDPC code be given by $\mathcal{S}_{\text{qLDPC}} = \langle g_x \sqcup g_z \rangle$ where $g_x = \{g_{x,1}, \dots, g_{x,m}\}$ and $g_z = \{g_{z,1}, \dots, g_{z,m}\}$. By performing the encoding procedure, the stabilizers of the new code become

$$\begin{aligned} \mathcal{S}' &= \langle g'_x \sqcup g'_z \rangle, \\ g'_x &= \{g_{x,1}^A \otimes g_{x,1}^B, \dots, g_{x,m}^A \otimes g_{x,m}^B\}, \\ g'_z &= \{g_{z,1}^A \otimes g_{z,1}^B, \dots, g_{z,m}^A \otimes g_{z,m}^B\}. \end{aligned}$$

This holds because, for example, when Alice sends Bob the parity of her first X stabilizer, $g_{x,1}^A$, Bob combines it with the parity of his first X stabilizer, $g_{x,1}^B$, to obtain the value of $g_{x,1}^A \otimes g_{x,1}^B$. After performing error correction on his side, he projects the joint code state into the $+1$ eigenspace of $g_{x,1}^A \otimes g_{x,1}^B$. The same logic applies to the remaining stabilizers. As a result, the stabilizers of the joint code become $\mathcal{S}' = \{s \otimes s \mid s \in \mathcal{S}_{\text{qLDPC}}\}$, which is isomorphic to the stabilizers of the qLDPC code: $\mathcal{S}_{\text{qLDPC}} \cong \mathcal{S}'$ via $s \mapsto s \otimes s$.

The logical operators of the code are elements in the centralizer of the stabilizer group. Since the structure of the

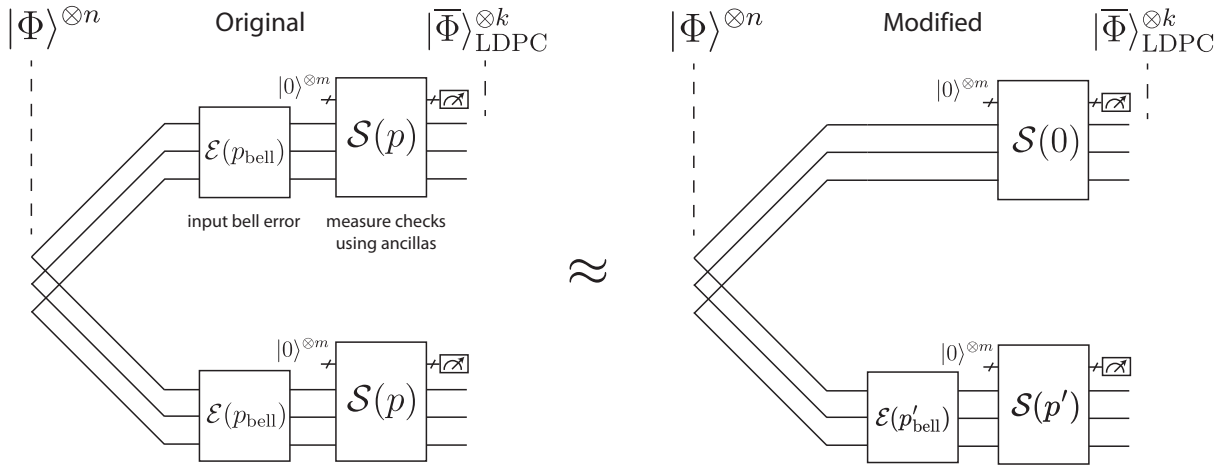


FIG. 4. Local and network noise in the encoding protocol. Alice and Bob initially share n noisy Bell pairs, modeled by a depolarizing channel with error parameter p_{bell} . Syndrome extraction is performed using local ancilla qubits, with a two-qubit gate error rate p at each node. To approximate the full noisy circuit (left), where both sides experience noise, we use a simplified model (right) in which Alice's side is noiseless and Bob's side has increased noise, with parameters $p'_{\text{bell}} \approx 2p_{\text{bell}}$ and $p' \approx 2p$.

stabilizers of the qLDPC code is preserved after the joint projection, the logical operators are also preserved. Explicitly, suppose the logical operators of the qLDPC code are given by $\mathcal{L}_x = \{\bar{X}_1, \dots, \bar{X}_k\}$ and $\mathcal{L}_z = \{\bar{Z}_1, \dots, \bar{Z}_k\}$, where $[\bar{X}_i, \bar{Z}_j] = 0$ for $i \neq j$ and $\{\bar{X}_i, \bar{Z}_i\} = 0$. Because of the isomorphism, the logical operators of the output state are given by

$$\mathcal{L}'_x = \{\bar{X} \otimes \bar{X} | \bar{X} \in \mathcal{L}_x\} \quad \text{and} \quad \mathcal{L}'_z = \{\bar{Z} \otimes \bar{Z} | \bar{Z} \in \mathcal{L}_z\},$$

which are the logical stabilizers of the k output Bell pairs. We highlight that the encoding protocol is fault-tolerant because the logical Bell stabilizers consist of physical Bell stabilizers that commute with the checks of the joint code, allowing for deterministic fault-tolerant error correction [35].

Error modeling—The error channels are illustrated in Fig. 4 (left). Each Bell pair undergoes independent depolarizing noise at each node, modeled as

$$\mathcal{E}(p_{\text{bell}}) = (1 - p_{\text{bell}})\rho + \frac{p_{\text{bell}}}{3}(X\rho X + Y\rho Y + Z\rho Z). \quad (\text{B1})$$

The QEC circuits experience two-qubit gate depolarizing errors with strength p , given by

$$\mathcal{S}(p) = (1 - p)\rho + \frac{p}{15} \sum_{P \in \{IX, IY, \dots, ZZ\}} P\rho P. \quad (\text{B2})$$

To simplify simulations, we shift errors from Alice's side to Bob's, effectively increasing the noise on Bob's side. This allows us to simulate only Bob's end, where the effective Bell error rate and circuit error rate are denoted p'_{bell} and p' (Fig. 4, right). In Supplemental Material [36], we use the symmetry of Bell pairs under Pauli operators to

show that $p'_{\text{bell}} \approx 2p_{\text{bell}}$. As discussed in the main text, we quote p'_{bell} as the input Bell error rate.

Similarly, we relate the two-qubit gate error rates of the QEC circuit, p and p' , by considering physical error events that lead to logical errors on the encoded Bell pairs. In Supplemental Material [36], we show that $p' \approx 2p$ is a valid approximation for single-round QEC circuits. We also numerically verify that, for a single round of QEC, joint decoding yields approximately the same threshold as decoding only Bob's side using the effective error model. For a large number of QEC rounds, $p' \approx 2p$ becomes a conservative, as argued in SM [36]. Nonetheless, we use $p' = 2p$ for the simulations in the main text, where the local gate error is fixed at 0.1%, resulting in an effective simulated gate error of $p' = 0.2\%$.

Codes—The HGP codes are constructed as the hypergraph product of two classical LDPC codes based on (3,4) regular Tanner graphs [50,61,92]. The LP codes extend this construction by incorporating a lifting operation, which reduces qubit overhead while preserving high code distances [64,68,93]. Finally, the SC codes build upon the LP structure by introducing spatial coupling of the lifted component matrices—a form of parity-check matrix concatenation—which further enhances qubit efficiency, achieving an encoding rate of $\sim 1/3$ [65,66]. All codes are optimized to (1) reduce small loops in the Tanner graph, which improves belief propagation decoding, (2) enhance expansion properties, facilitating single-shot decoding, and (3) maximize distance, improving error correction performance. Further details on code constructions are provided in Supplemental Material [36].

Resource requirements—To investigate the subthreshold scaling behavior of the codes, we modify

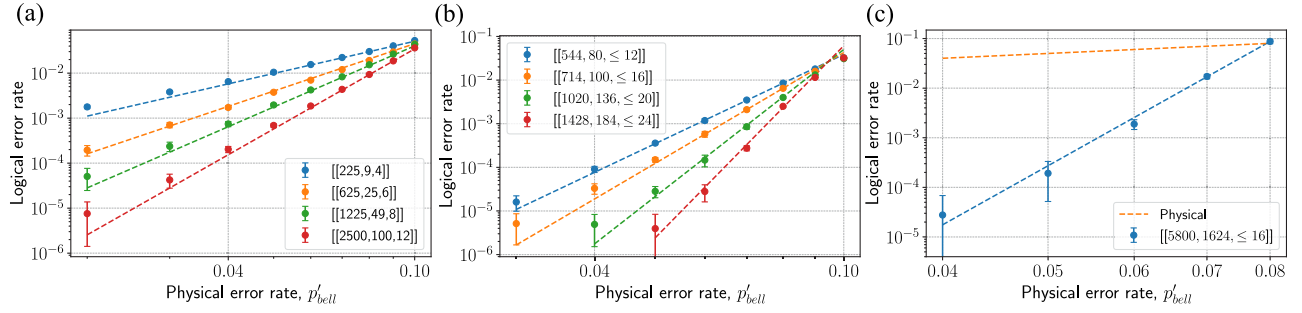


FIG. 5. Subthreshold data and fitted ansatz. Bell pairs have input depolarizing noise with strength $p'_{\text{bell}} = p$, and two-qubit gate depolarizing errors at $p'_{2q} = p/50$. We simulate 14 decoding rounds, each with three syndrome extraction cycles (42 total). Each round is decoded using BP, with the final round using BP + OSD. (a) Numerical results for the HGP codes. (b) Numerical results for the LP codes. (c) Numerical results for the SC code. Dotted lines fit the subthreshold data to the ansatz $\bar{P} = A(p/p_{\text{th}})^{Bn^C}$, where n is the code size, p_{th} is the threshold error rate, and A, B, C are fitting parameters. The logical error rate per cycle is $\bar{P} = 1 - (1 - \bar{P}_{\text{tot}})^{1/42}$, where $\bar{P}_{\text{tot}}$ is the total error after 42 cycles. Error bars assume a binomial distribution with N shots, computed as $\sqrt{\bar{P}_{\text{tot}}(1 - \bar{P}_{\text{tot}})/N}$.

the error model so that both the Bell-pair error and circuit error scale with the noise strength. We set $p_{\text{bell}} = p, p_{2q} = (p/50)$ so that the observed thresholds in the main text remain approximately preserved. Then, we vary p and fit the data to a subthreshold scaling ansatz. The ansatz we use is $\bar{P} = A(p/p_{\text{th}})^{Bn^C}$ [50], where A, B, C are fitting parameters, n is the code size, and p_{th} is the

threshold physical error rate (also fitted). The numerical data collected and the fitted ansatz are shown in Fig. 5. The fitted logical error rates are given by $\bar{P}_{\text{HGP}} = 0.07(p/0.11)^{0.31n^{0.38}}$, $\bar{P}_{\text{LP}} = 0.026(p/0.094)^{0.05n^{0.79}}$, $\bar{P}_{\text{SC}} = 0.26(p/0.087)^{0.26n^{0.44}}$. We document achievable logical Bell error rates as a function of input Bell error rates for select codes in Table II.

TABLE II. Error rates of encoded Bell pairs as a function of input Bell error rates. Values marked with * represent extrapolated error rates obtained by fitting to subthreshold data. Unmarked values are obtained from exact numerical simulations. The overhead (O) indicates the number of physical Bell pairs required per logical Bell pair.

Code (overhead)	Input Bell error									
	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
$[[2500, 100]]$ ($O = 25$)	$6.0 \times 10^{-9*}$	7.6×10^{-6}	4.2×10^{-5}	2.0×10^{-4}	6.9×10^{-4}	1.9×10^{-3}	4.4×10^{-3}	9.3×10^{-3}	1.9×10^{-2}	3.6×10^{-2}
$[[1428, 184]]$ ($O = 7.8$)	$2.0 \times 10^{-17*}$	$9.5 \times 10^{-13*}$	$5.1 \times 10^{-10*}$	$4.5 \times 10^{-8*}$	4.0×10^{-6}	2.8×10^{-5}	2.8×10^{-4}	2.5×10^{-3}	1.1×10^{-2}	3.2×10^{-2}
$[[5800, 1624]]$ ($O = 3.6$)	$2.3 \times 10^{-12*}$	$7.9 \times 10^{-9*}$	$9.4 \times 10^{-7*}$	2.8×10^{-5}	1.9×10^{-4}	1.9×10^{-3}	1.7×10^{-2}	8.7×10^{-2}		