Fundamental thresholds of realistic quantum error correction circuits from classical spin models

Davide Vodola*,[1,2], Manuel Rispler*,[3], Seyong Kim[4], and Markus Müller[5,6]

[1]Dipartimento di Fisica e Astronomia "Augusto Righi" dell'Università di Bologna, I-40127 Bologna, Italy
[2]INFN, Sezione di Bologna, I-40127 Bologna, Italy
[3]QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands
[4]Department of Physics, Sejong University, 05006 Seoul, Republic of Korea
[5]Institute for Theoretical Nanoelectronics (PGI-2), Forschungszentrum Jülich, 52428 Jülich, Germany
[6]Institute for Quantum Information, RWTH Aachen University, 52056 Aachen, Germany

Mapping the decoding of quantum error correcting (QEC) codes to classical disordered statistical mechanics models allows one to determine critical error thresholds of QEC codes under phenomenological noise models. Here, we extend this mapping to admit realistic, multi-parameter noise models of faulty QEC circuits, derive the associated strongly correlated classical spin models, and illustrate this approach for a quantum repetition code with faulty stabilizer readout circuits. We use Monte-Carlo simulations to study the resulting phase diagram and benchmark our results against a minimum-weight perfect matching decoder. The presented method provides an avenue to assess fundamental thresholds of QEC circuits, independent of specific decoding strategies, and can thereby help guiding the development of near-term QEC hardware.

The interdisciplinary endeavour to develop large-scale quantum computers has revealed unexpected and strong connections across fields, in particular between quantum information theory and statistical physics. Exploring these links has been fruitful in both directions, e.g. by efficient quantum algorithms to estimate partition functions of classical spin systems [1–9], by providing effective descriptions for entanglement spreading in random circuits [10–12] or in the context of quantum error correction (QEC). For the latter, it has been shown that the decoding of leading QEC codes [13, 14] such as the topological surface [15, 16] and color codes [17, 18] can be mapped onto disordered classical statistical mechanics models. Locating the phase transitions between ordered and disordered phases in these models reveals the parameter regimes for which QEC succeeds or fails, respectively [16]. So far, however, these mappings have been largely limited to QEC codes with simple phenomenological noise models: for instance, uncorrelated [16, 19–22] or weakly correlated bit and/or phase flip errors [23], phenomenological measurement errors during syndrome readout [24, 25], qubit loss and leakage [26–30] have been considered, with recent extensions also to bosonic QEC codes [31]. Realistic modeling of experimental quantum hardware [32–41], however, requires multi-parameter noise model descriptions of failure processes of the underlying quantum circuit components.

In this work, we extend the QEC-statistical mechanics model mapping and apply it to realistic quantum circuit-noise scenarios. For Clifford measurement circuits, correlations arising from circuit noise can be efficiently related to effective phenomenological models as direct input for (possibly sub-optimal) decoders [23, 42]. Here, we use a similar approach but perform a systematic error propagation analysis and identify emerging effective noise processes in realistic quantum circuits. Instead of direct decoding, we derive the associated disordered classical spin model with correlated quenched disorder, determine its phase diagram and thereby obtain maximum threshold values agnostic to specific decoding strategies. This method allows one to assess the maximum potential

---

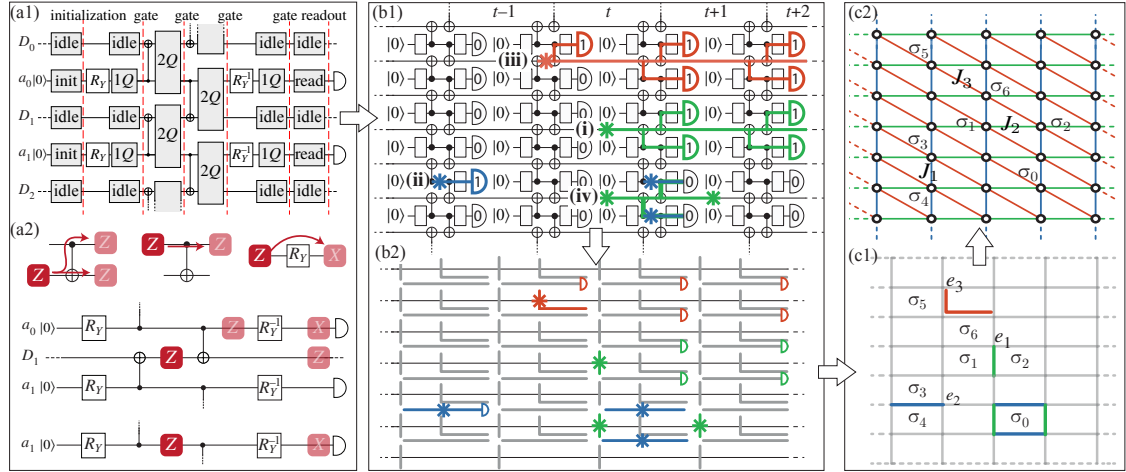*These two authors contributed equally to the present work.

Figure 1: (a1) One cycle of the phase-flip repetition code with circuit-level noise, shown for three data qubits D_i and two ancilla qubits a_i. White boxes R_Y (R_Y^{-1}) represent single-qubit Y-rotations of $\pi/2$ ($-\pi/2$). Gray boxes (idle, init, 1Q, 2Q and read) are depolarizing channels describing faults during idling, state preparation, single-qubit, CNOT gates and measurements, respectively. (a2) (Top row) Examples of error propagation through CNOT and R_Y gates: Z errors on target qubits propagate to the control qubits; Z errors do not propagate from control to target qubits; Z errors are transformed into X errors through R_Y. (Center row) A Z error between two CNOTs on D_1 propagates also as a measurement error to the ancilla a_0. (Bottom row) A Z error between two R_Y on the ancilla results in a measurement error. (b1) Four measurement cycles showing error propagation and the affected ancilla qubits. (i) A data-qubit error (green asterisk) propagates along the green path and affects two ancilla qubits at every subsequent measurement. (ii) A measurement error (blue asterisk) affecting a single ancilla. (iii) An error between the CNOT gates of a data qubit (red asterisk) propagates along the red path and affects one ancilla at step t and two ancilla qubits in subsequent measurements. (iv) A space-time equivalence given by two qubit errors and two measurement errors, which does not trigger any ancilla. (b2) Error graph generated by the physical errors in panel (b1). Positions where qubit, measurement and correlated data phase-flip and measurement errors can happen are represented by vertical, horizontal and L-shaped edges, respectively. Colored semicircles represent ancilla qubits triggered by an error event. From this graph we derive the fundamental error events determining the couplings of the statistical mechanics Hamiltonian (panels (c1)-(c2)) and the syndrome volume. (c1) Fundamental effective errors e_1, e_2, e_3 generated by the error processes (i)-(iii) and equivalences σ_ℓ. (c2) Lattice of equivalences σ_ℓ showing the couplings J_1 (vertical blue links), J_2 (horizontal green links), J_3 (diagonal red links) of the Hamiltonian of Eq. (5).

of realistic state-of-the-art QEC architectures. This is particularly pressing for QEC codes when (near-)optimal decoding is computationally hard [43], when optimal decoders are unknown or strongly dependent on the chosen QEC circuitry and noise model details [44], e.g. in measurement circuits for high-weight stabilizers in color codes [45–47] or recently developed flag-qubit based QEC [48–51].

We illustrate the mapping of circuit-level QEC codes to statistical mechanics models for a 1D quantum repetition code with faulty circuitry described by a multi-parameter microscopic noise model accounting for a number of single- and two-qubit gate noise processes. Whereas this code does not enable correction of arbitrary errors, current experimental efforts focus on achieving repetitive QEC with this code in the regime of beneficial error suppression for increasing code sizes [52–55]. We show that in contrast to oversimplified phenomenological noise models, new effective correlated noise processes arise, which we systematically derive and quantitatively relate to error rates of the underlying microscopic noise model.

We anticipate a robust performance gap between efficient, though sub-optimal minimum-weight perfect matching (MWPM) decoding and the fundamental upper limit as established via the phase diagram of the statistical physics model. This demonstrates the viability of this complete mapping approach to identify maximal performance of QEC circuitry.

1 Noisy quantum error correcting circuits

Here we consider the phase-flip n-qubit repetition code [56, 57] from the class of stabilizer QEC codes [58]. The ± 1 eigenvalues of the $n-1$ stabilizers $S_i = X_i \otimes X_{i+1}$ (with X_i the Pauli X matrix on qubit i) form the error syndrome, with error states having at least one non-trivial syndrome. In order to diagnose potential errors, the stabilizers S_i are measured by coupling them to $n-1$ *ancilla* qubits. Phase-flip errors Z_i are detected, unless they happen on all qubits, which would be indistinguishable from a logical phase-flip $Z_L = \otimes_{i=0}^{n-1} Z_i$. Stabilizer measurements on real hardware are noisy as well and thus need to be repeated periodically. Analysing the discrete difference in time between successive measurement rounds results in another repetition code in the time domain, which yields a $(1+1)$-dimensional space-time *syndrome volume*. The code-capacity critical error threshold, below which successful QEC is feasible, is $p_c = 0.5$, when the only error source are data qubit errors. The phenomenological noise threshold, where errors are injected with a rate p per round on both ancilla and data qubits is $p_c \approx 0.11$ [59]. This is identical to the threshold of the toric code with perfect measurements that exactly maps onto the repetition code with phenomenological noise, essentially trading one spatial dimension for the time dimension introduced by repeating stabilizer measurements in time.

1.1 Microscopic circuit noise

In contrast to the phenomenological noise models, which ignore the physical reality of the circuit the QEC code is running on, in this work we analyze a multi-parameter circuit level noise model that treats every component of the circuit as faulty, as illustrated in Fig. 1(a1). Here, every operation is followed by a depolarizing noise channel (grey boxes in Fig. 1(a1)). For single-qubit operations, we inject P_i ($P_i \in X, Y, Z$) with probability $\lambda/3$, where λ can represent state preparation (p_{sp}), idling (p_{id}), single-qubit gates (p_1) and measurements (p_{m}) [44, 60]. Likewise, after a CNOT we inject a non-trivial one- and two-qubit Pauli operator $P_i \otimes P_j$ with probability $p_2/15$. Errors propagate according to the rules described in Fig. 1(a2). For the special case $p_{\mathrm{sp}} = p_{\mathrm{id}} = p_1 = p_{\mathrm{m}} = p_2 = \lambda$ the threshold was found to lie at $\lambda \approx 0.033$ for maximum likelihood decoding [61].

1.2 Effective noise processes

Any microscopic circuit-level error configuration of the multi-parameter noise model will lead to one of three effective error processes or combinations thereof, called error chains E, with associated error probabilities: (i) a single data qubit phase flip error with probability p, (ii) a measurement error with probability q, and (iii) a correlated error flipping a data-qubit and simultaneously one of the adjacent measurement outcomes with probability r (examples shown in Fig. 1(b1)). Through propagation of Pauli error generators and factorizing the effective error channel(s) we can express an arbitrary Pauli circuit noise model in terms of the three effective error rates p, q and r (see Appendix A for details):

$$p = \frac{1}{2}\left[1 - \left(1 - \frac{16p_2}{15}\right)\left(1 - \frac{4p_{\mathrm{id}}}{3}\right)^4\right],$$

$$q = \frac{1}{2}\left[1 - \left(1 - \frac{16p_2}{15}\right)\left(1 - \frac{4p_1}{3}\right)^2\left(1 - \frac{4p_{\mathrm{sp}}}{3}\right)\left(1 - \frac{4p_{\mathrm{m}}}{3}\right)\right], \qquad (1)$$

$$r = \frac{8}{15}p_2.$$

Let us emphasize that this correspondence is exact: given an arbitrary set of circuit noise parameters, the above relationship expresses this set of noise parameters in terms of p, q and r exactly to all orders. We gather the error events from the microscopic processes (Fig. 1(b1)) in the error graph of Fig. 1(b2): here vertical, horizontal and L-shaped lines indicate positions where a data-qubit (i), a measurement (ii) or a correlated error (iii) event has occurred, while colored semicircles indicate the stabilizers triggered by these, respectively. The error graph is used to construct the types of interactions appearing in the classical statistical mechanics Hamiltonian (Fig. 1(c1)) and to construct the syndrome volume as input for the MWPM decoder.

From the error graph, we can visualize the error processes by introducing the lattice in Fig. 1(c1) where vertical links represent data qubits and horizontal links are measurement steps. The three error processes (i)-(iii) define the fundamental errors events e_1, e_2, e_3 shown in Fig. 1(c1):

1. the error e_1 is generated by a data qubit error (i) or a combination of a measurement error (ii) and a correlated error (iii) with probability

$$\Pr(e_1) = p(1-q)(1-r) + (1-p)qr; \tag{2}$$

2. the error e_2 is generated by a measurement error (ii) or a combination of a data qubit error (i) and a correlated error (iii) with probability

$$\Pr(e_2) = q(1-p)(1-r) + rp(1-q); \tag{3}$$

3. the error e_3 is generated by a correlated error (iii) or a combination of a data qubit error (i) and a measurement error (ii) with probability

$$\Pr(e_3) = r(1-p)(1-q) + pq(1-r). \tag{4}$$

These fundamental errors form error chains E that can be parametrized by introducing binary variables $h_\ell, v_\ell \in \{\pm 1\}$ for each link ℓ of the lattice. h_ℓ (v_ℓ) are set to -1 if the horizontal (vertical) link ℓ belongs to E.

2 Statistical mechanics model

Mapping an error model into a statistical mechanics Hamiltonian allows us to compute the probability $\Pr(\bar{E})$ of the class $\bar{E}$ formed of all the errors that differ from a reference (candidate) error chain E by a so-called *space-time equivalence* [16, 23]: A space-time equivalence $\hat{\sigma}$ (simply called equivalence in the following) is a trivial sequence of errors not detected by the stabilizers and thus producing the same space-time syndrome. For the phase flip code, equivalences are given by the action of a first phase flip on a qubit D_i at a measurement step t, followed by two measurement errors on the ancilla qubits of the stabilizers S_i adjacent to D_i and a second phase flip error on D_i before the measurement step $t + 2$ (see (iv) in Fig. 1(b1)). We associate a classical Ising spin σ to each equivalence (e.g. σ_0 in Fig. 1(c1)) such that a spin configuration with $\sigma = -1$ effectively describes the error $E' = E\hat{\sigma}$ that differs from E by the equivalence $\hat{\sigma}$. Thus, once we fix E, sampling over all possible spin configurations is equivalent to sampling over all possible errors E' in the class $\bar{E}$, and the probability of the class $\bar{E}$ is $\Pr(\bar{E}) = \sum_{\{\hat{\sigma}\}} \Pr(E\hat{\sigma})$. The limiting behavior of the complementary probabilities $\Pr(\bar{E})$ and $\Pr(\overline{EZ_L})$ with increasing system size allows us to distinguish two regimes: below threshold, either of the probabilities converges to unity and hence applying a correction from the same error class will undo the error with probability approaching unity, above threshold all probabilities remain asymptotically finite, i.e. there is no decoding choice that removes the error with high probability [21].

Given E which is described by $h_\ell \in \{\pm 1\}$ and $v_\ell \in \{\pm 1\}$ and the set of equivalences $\{\sigma\}$, the probability of the error E' can be written as $\Pr(E') = \exp[-\beta H_E(\sigma)]$ where

$$H_E(\sigma) = -\sum_{\ell \updownarrow \ell'} J_1 h_\ell \sigma_\ell \sigma_{\ell'} - \sum_{\ell \leftrightarrow \ell'} J_2 v_\ell \sigma_\ell \sigma_{\ell'} - \sum_{\ell \searrow \ell'} J_3 h_\ell v_\ell \sigma_\ell \sigma_{\ell'} \tag{5}$$

is an Ising Hamiltonian with correlated quenched disorder. The couplings J_1, J_2, J_3 and the inverse temperature β are fixed from the error model describing E by the so-called Nishimori conditions (see Eq.(14) in Appendix B). Ferromagnetic or antiferromagnetic interactions are given by the signs of v_ℓ and h_ℓ fixed by the error E. In the first sum of the r.h.s. of Eq. (5), σ_ℓ and $\sigma_{\ell'}$ are the spins horizontally adjacent to the link ℓ (e.g. spins σ_1 and σ_2 in Fig. 1(c1)); in the second sum, σ_ℓ and $\sigma_{\ell'}$ are the spins vertically adjacent to the link ℓ (e.g. spins σ_3 and σ_4 in Fig. 1(c1)); in the third sum, σ_ℓ and $\sigma_{\ell'}$ are the two spins adjacent to the two arms of the L-shaped error e_3 (e.g. spins σ_5 and σ_6 in Fig. 1(c1)). A similar Hamiltonian has been analyzed for the surface code with spatially correlated phenomenological noise [23], whereas here Hamiltonian (5) arises from multi-parameter circuit noise.
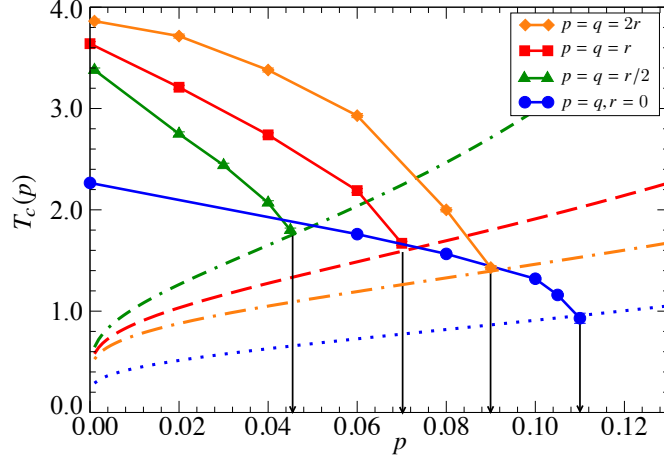
Figure 2: Phase diagram of the random bond Ising model for (I) $p = q = 2r$ (orange diamonds), (II) $p = q = r$ (red squares), (III) $p = q = r/2$ (green triangles), (IV) $p = q, r = 0$ (blue circles). The points represent the critical temperature T_c for the given disorder configuration. Solid lines are a guide to the eye, and dashed lines are the Nishimori lines for each case. The vertical black lines indicate the locations of the largest p for which T_c can be determined. Error bars are smaller than the symbol size.

3 Numerical results

3.1 Monte Carlo simulations

Mapping the QEC code to Eq. (5) allows us to access the critical thresholds of the code from thermodynamic properties of $H_E(\sigma)$ [16]. The error threshold separating the correctable from non-correctable parameter regime of the code corresponds to the phase transition between ordered and disordered phases of $H_E(\sigma)$. We determine this transition from Monte Carlo (MC) simulations and a scaling analysis of the two-body spin correlation length for various disorder strengths (see Appendix C for details). To this end, we transform the lattice in Fig. 1(c1) into the triangular one of Fig. 1(c2) where classical spins (white circles) reside on vertices and couplings J_1, J_2, J_3 are represented by vertical (blue), horizontal (green) and diagonal (red) links, respectively.

Figure 2 shows the resulting phase diagram for (I) $p = q = 2r$, (II) $p = q = r$, (III) $p = q = r/2$, and for the uncorrelated case (IV) $p = q$ and $r = 0$ (this reduces to the random-bond Ising model on a square lattice). We find that as the correlated error strength r increases, the thresholds decrease to $p_c(\text{I}) = 0.0925(25)$, $p_c(\text{II}) = 0.0725(25)$, $p_c(\text{III}) = 0.0475(25)$ from the uncorrelated threshold $p_c(\text{IV}) = 0.110(5)$. Near the Nishimori lines, longer MC runs are required to locate p_c and T_c accurately. The error in p_c is determined within the present MC statistics (see Appendix C).

3.2 Minimum weight matching decoder

The thresholds extracted via MWPM provide lower bounds to the thresholds obtained through the statistical mechanics mapping since they correspond to a decoding decision that is not necessarily optimal: MWPM looks for the likeliest errors given a syndrome, which in the statistical mechanics mapping corresponds to minimizing the energy as opposed to the free energy [16]. MWPM therefore disregards the degeneracy of errors, because the likeliest error need not necessarily be in the most likely class of equivalent errors. A discrepancy between MWPM and optimal (maximum-likelihood) decoding is known to exist already for phenomenological noise [62]. In order to make a fair comparison, we optimize the weight metric going into the MWPM algorithm, which uses the information available about the syndrome graph. A detailed derivation of the error weight metric used for decoding and information on the general MWPM strategy are provided in Appendix D.
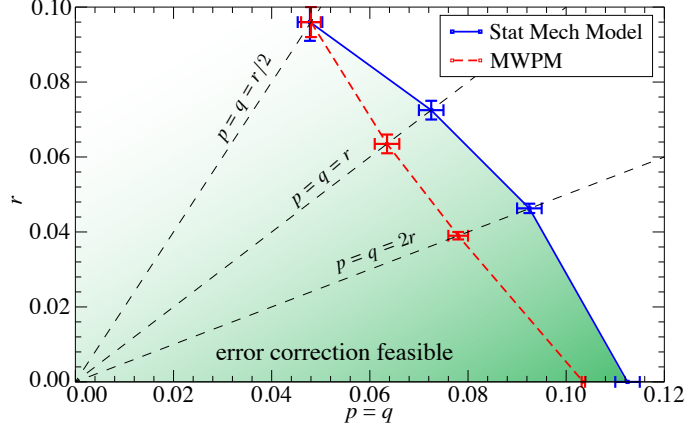
Figure 3: Threshold probabilities from the statistical mechanics model (blue points) and the MWPM decoder (red points). Lines are a guide to the eye. The region where the MWPM decoder finds correctability lies within the ordered region (green) of the statistical mechanics model. The latter region represents the fundamental regime, independent of specific decoding strategies, where correction via the repetition code is possible in principle.

3.3 Discussion

Figure 3 shows the thresholds from MC simulations (blue points, solid line) and from MWPM decoding (red points, dotted line). The region where MWPM successfully decodes lies within error bars inside the fundamental region of correctability determined with the statistical mechanics mapping. This demonstrates a finite interval between MWPM and a higher fundamental threshold that could be approached or achieved by improving the decoding strategy. To connect the phase diagram to experimental situations, let us discuss the relative strength of r and its meaning. First, p and q both are typically dominated by similar error processes that for many qubit platforms are of comparable error rates, hence we set $p = q$ for simplicity, and additionally, we dial up the strength of r. The known case $r = 0$ reduces to the phenomenological case without correlated errors, which would correspond to perfect two-qubit gates ($p_2 = 0$). Increasing r to $r = p/2 = q/2$ corresponds to $p_2 = 5p_1 + \mathcal{O}(p_1^2)$, which is roughly compatible with the two-qubit gate being an order of magnitude worse than single-qubit operations (error rate or infidelity) as observed in many experimental realizations [36, 40, 55, 63]. Considering $r = p = q$ corresponds to p_2 being dominant and all other error sources negligible ($p_{\mathrm{sp}} = p_{\mathrm{id}} = p_1 = p_{\mathrm{m}} = 0$). The case $r = 2p = 2q$ goes beyond the depolarizing noise model that underlies Eq. (1): not only is the two-qubit gate the only error source, it furthermore specifically produces errors leading to the r type instead of p and q, which would be described by an asymmetric depolarizing noise channel biased towards r. Finally, pure r errors ($p = q = 0$) correspond to a situation where syndrome information is perfectly trustable if interpreted correctly, akin to a repetition code with perfect measurements (with a threshold approaching 0.5 asymptotically). We thus see that the most experimentally relevant region $r \neq p \approx q$ is where we also find a clear separation between MWPM and the fundamental threshold. The separation seems to become narrower for larger contributions of r, which is compatible with the expectation in the extreme case of only r errors with an asymptotic threshold of 0.5. From a practical operational viewpoint, the estimated thresholds for ideal decoding and MWPM for $p = q = 2r$ correspond to single-qubit errors $p_1 \approx 0.02$ (ideal) vs. 0.017 (MWPM) and the two-qubit error rate $p_2 \approx 0.087$ (ideal) vs. 0.074 (MWPM). For $p = q = r$ the threshold is $p_2 \approx 0.14$ (ideal) vs. 0.12 (MWPM).

Interestingly, in a recent experimental realization of the phase-flip code [55] circuit errors are reported to be well described by Pauli errors. Casting the experimentally obtained error rates into Eq. (1), we find that they correspond to effective error rates $p = 0.032$, $q = 0.0285$ and $r = 0.0035$, which is in excellent agreement with the observation that this experiment is thoroughly in the error-suppression regime.

4 Outlook

The techniques presented in this work can be extended to more complex complete QEC codes, including surface and color codes, and concatenated codes. We anticipate that as the code circuitry becomes more complex, it will be necessary to include more types of effective noise processes, which in turn will give rise to new statistical mechanics models with even richer interaction and disorder properties. It will also be interesting to extend the mapping technique to non-Clifford dynamics, and to temporally or spatially correlated circuit-level noise, to study the QEC potential of an even broader class of realistic quantum processors.

Acknowledgments

References

[1] D. A. Lidar and O. Biham, "Simulating Ising spin glasses on a quantum computer," Phys. Rev. E **56**, 3661 (1997).

[2] R. D. Somma, C. D. Batista, and G. Ortiz, "Quantum approach to classical statistical mechanics," Phys. Rev. Lett. **99**, 030603 (2007).

[3] M. Van den Nest, W. Dür, and H. J. Briegel, "Classical spin models and the quantum-stabilizer formalism," Phys. Rev. Lett. **98**, 117207 (2007).

[4] J. Geraci and D. A. Lidar, "On the exact evaluation of certain instances of the Potts partition function by quantum computers," Commun. Math. Phys. **279**, 735 (2008).

[5] G. De las Cuevas, W. Dür, H. J. Briegel, and M. A. Martin-Delgado, "Unifying all classical spin models in a lattice gauge theory," Phys. Rev. Lett. **102**, 230502 (2009).

[6] G. De las Cuevas, W. Dür, M. V. den Nest, and M. A. Martin-Delgado, "Quantum algorithms for classical lattice models," New J. Phys. **13**, 093021 (2011).

[7] Y. Xu, G. De las Cuevas, W. Dür, H. J. Briegel, and M. A. Martin-Delgado, "The U(1) lattice gauge theory universally connects all classical models with continuous variables, including background gravity," J. Stat. Mech-Theory E **2011**, P02013 (2011).

[8] G. De las Cuevas and T. S. Cubitt, "Simple universal models capture all classical spin physics," Science **351**, 1180 (2016).

[9] M. H. Zarei and A. Montakhab, "Dual correspondence between classical spin models and quantum Calderbank-Shor-Steane states," Phys. Rev. A **98**, 012337 (2018).

[10] Y. Li and M. P. A. Fisher, "Statistical mechanics of quantum error correcting codes," Phys. Rev. B **103**, 104306 (2021).

[11] A. Nahum, S. Vijay, and J. Haah, "Operator spreading in random unitary circuits," Phys. Rev. X **8**, 021014 (2018).

[12] T. Zhou and A. Nahum, "Emergent statistical mechanics of entanglement in random unitary circuits," Phys. Rev. B **99**, 174205 (2019).

[13] B. M. Terhal, "Quantum error correction for quantum memories," Rev. Mod. Phys. **87**, 307 (2015).

[14] D. A. Lidar and T. A. Brun, eds., *Quantum Error Correction* (Cambridge University Press, Cambridge, England, 2013).

[15] A. Kitaev, "Fault-tolerant quantum computation by anyons," Ann. Phys. **303**, 2 (2003).

[16] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, "Topological quantum memory," J. Math. Phys. **43**, 4452 (2002).

[17] H. Bombin and M. A. Martin-Delgado, "Topological quantum distillation," Phys. Rev. Lett. **97**, 180501 (2006).

[18] H. Bombin and M. A. Martin-Delgado, "Topological computation without braiding," Phys. Rev. Lett. **98**, 160502 (2007).

[19] H. G. Katzgraber, H. Bombin, and M. A. Martin-Delgado, "Error threshold for color codes and random three-body Ising models," Phys. Rev. Lett. **103**, 090501 (2009).

[20] H. G. Katzgraber, H. Bombin, R. S. Andrist, and M. A. Martin-Delgado, "Topological color codes on union jack lattices: a stable implementation of the whole clifford group," Phys. Rev. A **81**, 012319 (2010).

[21] H. Bombin, R. S. Andrist, M. Ohzeki, H. G. Katzgraber, and M. A. Martin-Delgado, "Strong resilience of topological codes to depolarization," Phys. Rev. X **2**, 021004 (2012).

[22] R. S. Andrist, H. G. Katzgraber, H. Bombin, and M. A. Martin-Delgado, "Error tolerance of topological codes with independent bit-flip and measurement errors," Phys. Rev. A **94**, 012318 (2016).

[23] C. T. Chubb and S. T. Flammia, "Statistical mechanical models for quantum codes with correlated noise," Ann. Inst. Henri Poincaré Comb. Phys. Interact. **8**, 269 (2021).

[24] T. Ohno, G. Arakawa, I. Ichinose, and T. Matsui, "Phase structure of the random-plaquette Z_2 gauge model: accuracy threshold for a toric quantum memory," Nucl. Phys. B **697**, 462 (2004).

[25] R. S. Andrist, H. G. Katzgraber, H. Bombin, and M. A. Martin-Delgado, "Tricolored lattice gauge theory with randomness: fault tolerance in topological color codes," New J. Phys. **13**, 83006 (2011).

[26] T. C. Ralph, A. J. F. Hayes, and A. Gilchrist, "Loss-tolerant optical qubits," Phys. Rev. Lett. **95**, 100501 (2005).

[27] T. M. Stace, S. D. Barrett, and A. C. Doherty, "Thresholds for Topological Codes in the Presence of Loss," Phys. Rev. Lett. **102**, 200501 (2009).

[28] T. M. Stace and S. D. Barrett, "Error correction and degeneracy in surface codes suffering loss," Phys. Rev. A **81**, 022317 (2010).

[29] D. Vodola, D. Amaro, M. A. Martin-Delgado, and M. Müller, "Twins percolation for qubit losses in topological color codes," Phys. Rev. Lett. **121**, 060501 (2018).

[30] D. Amaro, J. Bennett, D. Vodola, and M. Müller, "Analytical percolation theory for topological color codes under qubit loss," Phys. Rev. A **101**, 032317 (2020).

[31] C. Vuillot, H. Asasi, Y. Wang, L. P. Pryadko, and B. M. Terhal, "Quantum error correction with the toric Gottesman-Kitaev-Preskill code," Phys. Rev. A **99**, 032344 (2019).

[32] A. D. Córcoles, E. Magesan, S. J. Srinivasan, A. W. Cross, M. Steffen, J. M. Gambetta, and J. M. Chow, "Demonstration of a quantum error detection code using a square lattice of four superconducting qubits," Nat. Commun. **6**, 6979 (2015).

[33] M. Saffman, "Quantum computing with atomic qubits and Rydberg interactions: progress and challenges," J. Phys. B: At., Mol. Opt. Phys. **49**, 202001 (2016).

[34] V. Negnevitsky, M. Marinelli, K. K. Mehta, H. Y. Lo, C. Flühmann, and J. P. Home, "Repeated multi-qubit readout and feedback with a mixed-species trapped-ion register," Nature **563**, 527 (2018).

[35] C. K. Andersen, A. Remm, S. Lazar, S. Krinner, J. Heinsoo, J.-C. Besse, M. Gabureac, A. Wallraff, and C. Eichler, "Entanglement stabilization using ancilla-based parity detection and real-time feedback in superconducting circuits," npj Quantum Inf. **5**, 69 (2019).

[36] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, "Trapped-ion quantum computing: Progress and challenges," Applied Physics Reviews **6** (2019).

[37] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, "Superconducting qubits: Current state of play," Annual Review of Condensed Matter Physics **11**, 369 (2020).

[38] C. K. Andersen, A. Remm, S. Lazar, S. Krinner, N. Lacroix, G. J. Norris, M. Gabureac, C. Eichler, and A. Wallraff, "Repeated quantum error detection in a surface code," Nat. Phys. **16**, 875 (2020).

[39] S. Pezzagna and J. Meijer, "Quantum computer based on color centers in diamond," Applied Physics Reviews **8**, 011308 (2021).

[40] A. Chatterjee, P. Stevenson, S. De Franceschi, A. Morello, N. P. de Leon, and F. Kuemmeth, "Semiconductor qubits in practice," Nat. Rev. Phys. **3**, 157 (2021).

[41] L. Egan *et al.*, "Fault-tolerant control of an error-corrected qubit," Nature **598**, 281 (2021).

[42] L. P. Pryadko, "On maximum-likelihood decoding with circuit-level errors," Quantum **4**, 304 (2020), 1909.06732v4 .

[43] P. Iyer and D. Poulin, "Hardness of Decoding Quantum Stabilizer Codes," IEEE Trans. Inf. Theory **61**, 5209 (2015).

[44] A. M. Stephens, "Fault-tolerant thresholds for quantum error correction with the surface code," Phys. Rev. A **89**, 022321 (2014).

[45] D. Wang, A. Fowler, C. Hill, and L. Hollenberg, "Graphical algorithms and threshold error rates for the 2d color code," Quant. Inf. Comp **10**, 780 (2010).

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

[47] M. E. Beverland, A. Kubica, and K. M. Svore, "Cost of universality: A comparative study of the overhead of state distillation and code switching with color codes," PRX Quantum **2**, 020341 (2021).

[48] R. Chao and B. Reichardt, "Fault-tolerant quantum computation with few qubits," npj Quantum Inf. **4** (2017).

[49] C. Chamberland and M. E. Beverland, "Flag fault-tolerant error correction with arbitrary distance codes," Quantum **2**, 53 (2018).

[50] C. Chamberland, A. Kubica, T. J. Yoder, and G. Zhu, "Triangular color codes on trivalent graphs with flag qubits," New J. Phys. **22**, 023019 (2020).

[51] C. Chamberland and K. Noh, "Very low overhead fault-tolerant magic state preparation using redundant ancilla encoding and flag qubits," npj Quantum Inf. **6** (2020).

[52] P. Schindler, J. T. Barreiro, T. Monz, V. Nebendahl, D. Nigg, M. Chwalla, M. Hennrich, and R. Blatt, "Experimental repetitive quantum error correction," Science **332**, 1059 (2011).

[53] G. Waldherr *et al.*, "Quantum error correction in a solid-state hybrid spin register," Nature **506**, 204 (2014).

[54] J. Kelly *et al.*, "State preservation by repetitive error detection in a superconducting quantum circuit," Nature **519**, 66 (2015).

[55] Z. Chen *et al.*, "Exponential suppression of bit or phase errors with cyclic error correction," Nature **595**, 383 (2021).

[56] S. J. Devitt, W. J. Munro, and K. Nemoto, "Quantum error correction for beginners," Rep. Prog. Phys. **76**, 076001 (2013).

[57] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[58] D. Gottesman, *Stabilizer codes and quantum error correction*, Ph.D. thesis, California Institute of Technology (1997).

[59] A. Honecker, M. Picco, and P. Pujol, "Universality Class of the Nishimori Point in the 2D $\pm J$ Random-Bond Ising Model," Phys. Rev. Lett. **87**, 047201 (2001).

[60] R. Raussendorf and J. Harrington, "Fault-Tolerant Quantum Computation with High Threshold in Two Dimensions," Phys. Rev. Lett. **98**, 190504 (2007).

[61] M. Rispler, P. Cerfontaine, V. Langrock, and B. M. Terhal, "Towards a realistic GaAs-spin qubit device for a classical error-corrected quantum memory," Phys. Rev. A **102**, 022416 (2020).

[62] 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).

[63] W. Huang *et al.*, "Fidelity benchmarks for two-qubit gates in silicon," Nature **569**, 532 (2019).

[64] D. Greenbaum, "Introduction to quantum gate set tomography," (2015), arXiv:1509.02921 [quant-ph] .

[65] M. Palassini and S. Caracciolo, "Universal finite-size scaling functions in the 3D Ising spin glass," Phys. Rev. Lett. **82**, 5128 (1999).

[66] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," The Journal of Chemical Physics **21**, 1087 (1953).

[67] R. H. Swendsen and J.-S. Wang, "Replica Monte Carlo simulation of spin-glasses," Phys. Rev. Lett. **57**, 2607 (1986).

[68] D. J. Earl and M. W. Deem, "Parallel tempering: Theory, applications, and new perspectives," Phys. Chem. Chem. Phys. **7**, 3910 (2005).

[69] A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using networkx," in *Proceedings of the 7th Python in Science Conference*, edited by G. Varoquaux, T. Vaught, and J. Millman (Pasadena, CA USA, 2008) pp. 11 – 15.

[70] D. Wang, A. Fowler, A. Stephens, and L. Hollenberg, "Threshold error rates for the toric and planar codes," Quant. Inf. Comp **10**, 456 (2010).

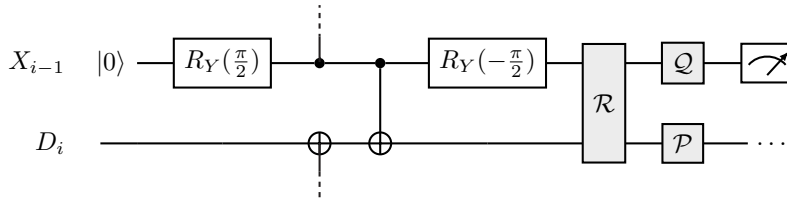# A Effective error model from circuit level noise

Figure 4: Cutout of one ancilla and one data qubit of the QEC circuit. All noise contributions of the circuit-level noise model (see Fig. 1(a) of the main text) can be merged into the three channels $\mathcal{P}$, $\mathcal{Q}$ and $\mathcal{R}$. The effective error rates are given in Eq. (9).

Circuit-level noise in general, by definition, introduces noise channels at all locations inside the circuit. However for the given QEC code, it is possible to rewrite the noise model by exploiting the fact that many errors injected at different locations effectively lead to the same error and are thus equivalent, which allows us to arrive at a much simpler noise model (while keeping the mapping exact). Let us establish these equivalences as follows: on a data qubit, observe that X-errors do not affect the syndrome, hence for the purposes of decoding $X \sim \mathbb{1}$ and thus also $Y \sim Z$. On ancilla qubits, the situation is similar, except that the intermediate rotations $R_Y(\pm\pi/2)$ flip the relevant error: at initialization and read-out, only X-errors are relevant, Z does neither alter the initialization state nor the readout in the computational basis. Between the rotation gates $R_Y(\pm\pi/2)$, the roles are reversed and only Z is relevant. This immediately lets us move all single-qubit channels to one side, since all noise channels commute (Pauli channels commute) and the Z-errors commute past the control of the CNOT. Similarly, the data qubit single noise channels can be collected into four consecutive single qubit noise channels. Turning to the two-qubit error channels, let us observe that by the equivalence of errors there are four possible distinct errors arising from the two-qubit depolarizing channel: $\mathbb{1}\mathbb{1}, \mathbb{1}Z, Z\mathbb{1}$ and ZZ (omitting the tensor product symbol). There is a single new phenomenon encoded here: looking at the earlier CNOT noise channel, observe that ZZ is equivalent to a simple data qubit error that would have preceded the CNOT and $\mathbb{1}Z$ (data-ancilla) is equivalent to a simple measurement error, $Z\mathbb{1}$ generates a new syndrome phenomenon, which we call an r-type error. The effect of the second two-qubit noise channel is the same, just the roles of the terms being permuted. Assuming a depolarizing probability p_2 of the CNOT gate, defining the operators $P = \mathbb{1} \otimes Z$, $Q = X \otimes \mathbb{1}$ and $R = X \otimes Z = P \cdot Q$ and using hats for superoperator notation $\hat{M}(\rho) := M\rho M^\dagger$ we thus have

$$\mathcal{N}_{2qubit} = \left(1 - \frac{12p_2}{15}\right)\hat{\mathbb{1}} + \frac{4p_2}{15}\left(\hat{P} + \hat{Q} + \hat{R}\right) \tag{6}$$

This channel is a mixture of three error events, which we would like to factorize into independent channels. Remarkably, we can make the ansatz of a product of three channels and solve for their error rates.

$$(\mathcal{N}_{2qubit})^2 = \left[(1-\lambda_p)\hat{\mathbb{1}} + \lambda_p\hat{P}\right] \cdot \left[(1-\lambda_q)\hat{\mathbb{1}} + \lambda_q\hat{Q}\right] \cdot \left[(1-\lambda_r)\hat{\mathbb{1}} + \lambda_r\hat{R}\right] \tag{7}$$

Note that the square means applying the channel twice, which comes from the fact that we have two two-qubit noise channels acting subsequently, both of the form (6). By symmetry of Eq. (6), the three rates must be equal, they turn out to be $\lambda_r = \lambda_p = \lambda_q = 8p_2/15$. For the r-type error we are done, but for the effective single-qubit channels $\mathcal{P}$ and $\mathcal{Q}$, it remains to incorporate the four contributions from the single-qubit error channels. This is straightforward in the Pauli-transfer-matrix (PTM) representation [64], where we characterize a channel $\mathcal{E}$ by the matrix $R_{ij} = \text{tr}(P_i\mathcal{E}(P_j))/d$ with $i \in \{0,1,2,3\}$. The strength of this representation is that here the action of a sequence of channels is simply given by matrix multiplication of their PTMs. The PTM of a bit-flip channel is

$$\hat{\mathcal{E}} = (1-\gamma)\hat{\mathbb{1}} + \gamma\hat{X} \cong R_{\mathcal{E}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1-2\gamma & 0 \\ 0 & 0 & 0 & 1-2\gamma \end{pmatrix} \tag{8}$$

and furthermore it is evident that composing bit-flip channels results in a new bit-flip channel, such that we can read off the error rate from computing one of the non-trivial entries on the diagonal. This results in $\mathcal{P} = (1-p)\hat{\mathbb{1}} + p\hat{P}$, $\mathcal{Q} = (1-q)\hat{\mathbb{1}} + q\hat{Q}$ and $\mathcal{R} = (1-r)\hat{\mathbb{1}} + r\hat{R}$ with

$$p = \frac{1}{2}\left[1 - \left(1 - \frac{16p_2}{15}\right)\left(1 - \frac{4p_{\text{id}}}{3}\right)^4\right]$$
$$q = \frac{1}{2}\left[1 - \left(1 - \frac{16p_2}{15}\right)\left(1 - \frac{4p_1}{3}\right)^2\left(1 - \frac{4p_{\text{sp}}}{3}\right)\left(1 - \frac{4p_{\text{m}}}{3}\right)\right] \tag{9}$$
$$r = \frac{8}{15}p_2.$$

These are the three channels depicted in Fig. 4. Let us comment on the structure of these expressions: we recover the standard circuit-level noise by setting $p_2 = p_{\text{id}} = p_{\text{m}} = p_1 = p_{\text{sp}} = \lambda$. In that case, in leading order in λ we recover $r/p = 1/6$, which can be understood also by simply counting the number of locations weighted by their error probability. While distinguishing between different error rates of different components in the circuit is straightforward, the factorization of the two-qubit noise channel into independent channels crucially relied on the symmetry between the different $P_i \otimes P_j$ terms. This suggests that biasing a particular two-qubit noise term would make the prefactor in Eq. (6) non-symmetric and thus hinder the factorization of the channel. As a side note, we remark on a slightly non-intuitive feature in the conventional definition of the depolarizing channel: if we define it (as we do) as a random application of Paulis, complete depolarization (i.e. deterministically receiving the completely mixed state) corresponds to $\lambda = 3/4$ (single qubit depolarizing channel) or $\lambda = 15/16$ (two-qubit depolarizing channel), which shows that no term in the product inside Eq. (9) becomes negative. Additionally, some works in the literature choose to not use a depolarizing noise for every location but rather put a bit-flip channel at the initialization and measurement locations. The two differ by a rescaling-factor of $2/3$, since only two of the three Pauli errors lead to errors at initialization and readout (e.g. in the computational basis a Z acts trivially etc.).

B Derivation of the correlated disordered interacting classical spin model

The technique of the statistical mechanical mapping is used to construct a classical statistical Hamiltonian with quenched disorder for describing the error model of a quantum code. The mapping relies on the identification of the partition function of the Hamiltonian with the probability

$\Pr(\bar{E})$ of the class $\bar{E} = \{E'\}$ of errors E' that are equivalent up to the action of space-time equivalences to a reference error E. Several ways have been introduced for deriving this mapping [16, 23]. Here we will follow mainly the approach of Ref. [23].

Any error E of the repetition code is composed of the fundamental errors e_1, e_2, e_3 introduced in the main text that in turn are generated by the three error processes (i) single-qubit phase flip error on data qubits with probability p, (ii) measurement error with probability q, and (iii) single-qubit phase flip errors happening on the target qubit between two adjacent CNOT gates that trigger a correlated occurrence of a measurement error on one of the neighboring ancilla qubits with probability r.

The probabilities of the errors e_1, e_2, e_3 are easily calculated and are given by

$$
\begin{aligned}
\pi_0 &\equiv \Pr(e_0) = (1-p)(1-q)(1-r) + pqr, \\
\pi_1 &\equiv \Pr(e_1) = p(1-q)(1-r) + rq(1-p), \\
\pi_2 &\equiv \Pr(e_2) = q(1-p)(1-r) + rp(1-q), \\
\pi_3 &\equiv \Pr(e_3) = pq(1-r) + r(1-p)(1-q).
\end{aligned}
\tag{10}
$$

where, for completeness, we added the probability of the trivial error e_0 that corresponds to the absence or the simultaneous presence of all of the three error processes.

In a section of the lattice in Fig. 1(c1) of the main text made only of a horizontal and a vertical link (shown also in Fig. 5), the probability of a generic event E composed by the errors e_j with $j = 0, 1, 2, 3$ can be written as

$$
\Pr(E) = \pi_0^{f_0(v,h)} \pi_1^{f_1(v,h)} \pi_2^{f_2(v,h)} \pi_3^{f_3(v,h)}
\tag{11}
$$

where $v, h \in \{\pm 1\}$ are binary variables that take the negative value -1 if the horizontal (vertical) link belongs to the error E. The Boolean functions $f_j \in \{0, 1\}$ signal if e_j belongs to the error E and they satisfy $f_0(v = +1, h = +1) = f_1(v = -1, h = +1) = f_2(v = +1, h = -1) = f_3(v = -1, h = -1) = 1$, otherwise they are zero. From this, they can be written as

$$
\begin{aligned}
f_0(v,h) &= \frac{1}{4}(1 + h + v + v\,h), & f_1(v,h) &= \frac{1}{4}(1 + h - v - v\,h), \\
f_2(v,h) &= \frac{1}{4}(1 - h + v - v\,h), & f_3(v,h) &= \frac{1}{4}(1 - h - v + v\,h).
\end{aligned}
\tag{12}
$$

By substituting the Eqs. (12) in Eq. (11), the probability $\Pr(E)$ can be written as

$$
\Pr(E) = \exp\left[-\beta H_E\right],
\tag{13}
$$

where $H_E = -J_0 - J_1 h - J_2 v - J_3 v\,h$ is a noise Hamiltonian [23] and the couplings J_0, J_1, J_2, J_3 satisfy

$$
\begin{aligned}
\beta J_0 &= \frac{1}{4} \log\left(\pi_0 \pi_1 \pi_2 \pi_3\right), & \beta J_1 &= \frac{1}{4} \log \frac{\pi_0 \pi_1}{\pi_2 \pi_3}, \\
\beta J_2 &= \frac{1}{4} \log \frac{\pi_0 \pi_2}{\pi_1 \pi_3}, & \beta J_3 &= \frac{1}{4} \log \frac{\pi_0 \pi_3}{\pi_1 \pi_2}.
\end{aligned}
\tag{14}
$$

These are the Nishimori conditions connecting the Hamiltonian H_E to the error model of Eq. (10). The Hamiltonian H_E can be extended to the whole lattice by introducing binary variables $h_\ell, v_\ell \in \{\pm 1\}$ associated to each link ℓ of the lattice. The variables h_ℓ (v_ℓ) take the negative value -1 if the horizontal (vertical) link ℓ belongs to the error E. The Hamiltonian will then take the form

$$
H_E = -\sum_\ell \left(J_0 + J_1 h_\ell + J_2 v_\ell + J_3 v_\ell\, h_\ell\right).
\tag{15}
$$

The syndrome associated with the error E can also be generated by other error chains $E' = E\hat{\sigma}$ that differ from E by the action of the space-time equivalences $\hat{\sigma}$ that are trivial sequences of data-qubit errors and measurement errors that are not detected by the stabilizers. Figure 6 shows the lattice with two highlighted equivalences: $\hat{\sigma}_0$ and $\hat{\sigma}_1$ belonging to the bulk and the boundary of the lattice, respectively.
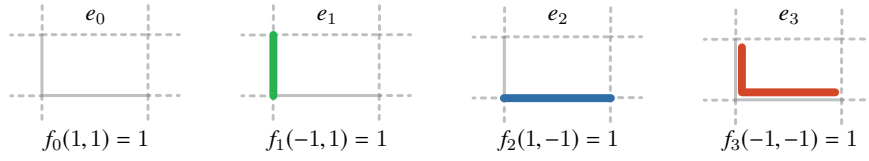
Figure 5: Fundamental errors e_0, e_1, e_2, e_3 parametrized by two binary variables $v, h \in \{-1, +1\}$ for the vertical and horizontal links and four Boolean functions $f_j(v, h) \in \{0, 1\}$. The trivial error e_0 corresponds to $(v, h) = (1, 1)$ and the function f_0 satisfies $f_0(1, 1) = 1$. The error e_1 corresponds to $(v, h) = (-1, 1)$ and the function f_1 satisfies $f_1(-1, 1) = 1$. The error e_2 corresponds to $(v, h) = (1, -1)$ and the function f_2 satisfies $f_2(1, -1) = 1$. The error e_3 corresponds to $(v, h) = (-1, -1)$ and the function f_3 satisfies $f_3(-1, -1) = 1$. On all the other cases not shown in the figure, the functions $f_j(v, h)$ are zero.
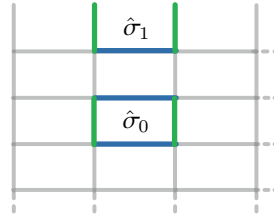

Figure 6: The equivalence $\hat{\sigma}_0$ in the bulk and $\hat{\sigma}_1$ at the boundary of the space-time lattice. The equivalence $\hat{\sigma}_0$ is given by the combined action of two measurement errors on the ancilla qubits (horizontal blue links) preceded and followed by phase flip errors on the data qubit (vertical green links) adjacent to the ancilla qubits. The equivalence $\hat{\sigma}_1$ at the boundary is given by the combined action of one measurement error preceded and followed by phase flip errors on the data qubit.
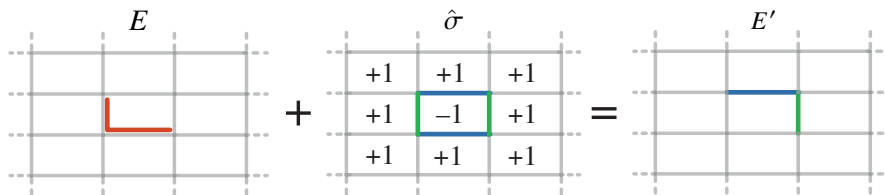
Figure 7: Action of the equivalence $\hat{\sigma}$ on an error $E = e_3$ (red L-shaped link). The equivalence $\hat{\sigma}$ (central panel) is given by two data-qubit errors (vertical green links) and two measurement errors (horizontal blue links). On the lattice the action of σ is represented by a closed loop and it can be parametrized by a configuration of classical Ising spins $\{\sigma = \pm 1\}$ as explained in the text.

We can associate classical Ising spins $\{\sigma\}$ to each of the equivalences such that if a spin σ in this configuration is -1, we are effectively considering the error $E' = E\hat{\sigma}$ that differs from the reference error E by the equivalence $\hat{\sigma}$ (for an example of an equivalence applied to an error E see Fig. 7). These Ising spins $\{\sigma\}$ can be associated to the plaquettes of the lattice in Fig. 1(c1) of the main text (see also Fig. 7). The data-qubit errors (green vertical links in Fig. 7) entering the equivalence $\hat{\sigma}$ are given by the vertical links ℓ of the lattice for which the product $\sigma_\ell \sigma_{\ell'} = -1$ where the equivalences $\sigma_\ell, \sigma_{\ell'}$ belong to the left and the right plaquettes w.r.t ℓ (Fig. 8(a)). Similarly, the measurement errors (blue horizontal links in Fig. 7) entering the equivalence $\hat{\sigma}$ are associated to the horizontal links ℓ of the lattice for which the product $\sigma_\ell \sigma_{\ell'} = -1$ where the equivalences $\sigma_\ell, \sigma_{\ell'}$ belong to the top and the bottom plaquettes w.r.t. ℓ (Fig. 8(b)).

The probability $\Pr(E')$ of the error E' can also be written as $\Pr(E') = \exp\left[-\beta H_{E'}\right]$ where in the noise Hamiltonian $H_{E'} = -\sum_\ell \left(J_0 + J_1 h'_\ell + J_2 v'_\ell + J_3 v'_\ell h'_\ell\right)$ the binary parameters h'_ℓ and v'_ℓ defining the error E' are fixed by the parameters h_ℓ and v_ℓ of E and by the equivalence σ as follows:

- for a vertical link ℓ, $v'_\ell = v_\ell \sigma_\ell \sigma_{\ell'}$ where the equivalences $\sigma_\ell, \sigma_{\ell'}$ are on the left and on the right of ℓ (Fig. 8(a));

- for a horizontal link ℓ, h'_ℓ is given by $h'_\ell = h_\ell \sigma_\ell \sigma_{\ell'}$ where the equivalences σ_ℓ and $\sigma_{\ell'}$ are on top and at the bottom of ℓ (Fig. 8(b)).

Substituting the variables h'_ℓ and v'_ℓ in the Hamiltonian $H_{E'}$ allows us to compute the probability $\Pr(\bar{E})$ of the class $\bar{E} = \{E'\}$ composed by errors E' that are equivalent up to the action of equivalences to a reference error E as

$$\Pr(\bar{E}) = \sum_{\{\hat{\sigma}\}} \Pr(E\hat{\sigma}) = \sum_{\{\sigma\}} \exp\left[-\beta H_E(\sigma)\right], \tag{16}$$

where

$$H_E(\sigma) = -\sum_{\ell \updownarrow \ell'} J_1 h_\ell \sigma_\ell \sigma_{\ell'} - \sum_{\ell \leftrightarrow \ell'} J_2 v_\ell \sigma_\ell \sigma_{\ell'} - \sum_{\ell \nwarrow \ell'} J_3 h_\ell v_\ell \sigma_\ell \sigma_{\ell'} \tag{17}$$

is the Ising Hamiltonian with correlated disorder of the main text and we have neglected the unimportant additive constant J_0. The couplings J_1, J_2, J_3 and the inverse temperature β are fixed from the error model describing E by the Nishimori conditions previously obtained (see Eq. (14)). The type of interactions (ferromagnetic or antiferromagnetic) is given by the signs of v_ℓ and h_ℓ that are fixed by the reference error E. In the first sum of the r.h.s. of Eq. (17), σ_ℓ and $\sigma_{\ell'}$ are the spins horizontally adjacent to the link ℓ (e.g. spins σ_1 and σ_2 in Fig. 1(c1) of the main text); in the second sum, σ_ℓ and $\sigma_{\ell'}$ are the spins vertically adjacent to the link ℓ (e.g. spins σ_3 and σ_4 in Fig. 1(c1) of the main text); in the third sum, σ_ℓ and $\sigma_{\ell'}$ are the two spins adjacent to the two arms of the L-shaped error e_3 (e.g. spins σ_5 and σ_6 in Fig. 1(c1) of the main text). Equation (16) shows that the probability $\Pr(\bar{E})$ of the class $\bar{E}$ can be written as the partition function $\mathcal{Z}_E = \sum_{\{\sigma\}} \exp\left[-\beta H_E(\sigma)\right]$ of a statistical mechanical Hamiltonian $H_E(\sigma)$. The limiting behavior of the complementary probabilities $\Pr(\bar{E})$ and $\Pr(\overline{EZ_L})$ with increasing system size allows us to distinguish two regimes: below threshold, either of the probabilities converges to unity and applying a correction from the same error class will undo the error with probability approaching unity, above threshold all probabilities remain asymptotically finite [21].

C   Monte-Carlo simulation study of the phase diagram

In this section we present the details on the numerical simulations of the Hamiltonian $H_E(\sigma)$. In particular, we describe how we choose the couplings J_1, J_2, J_3 of $H_E(\sigma)$ and the methods for obtaining the phase transition points.

C.1   Couplings of the random bond-Ising model

The Hamiltonian of Eq. (17) that we analyse corresponds to a random-bond Ising model on a triangular lattice with couplings J_1, J_2 and J_3 (see Fig. 1(c2) of the main text). Relative ratio
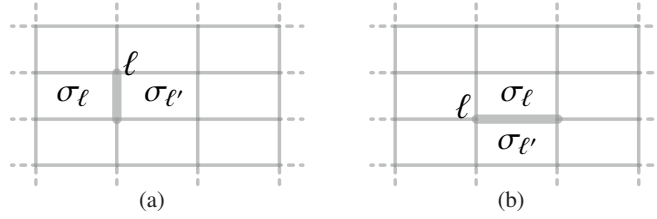
Figure 8: Parametrization of the equivalences and the error E'. For the equivalences: Given an Ising spin configuration $\{\sigma\}$ where each spin is associated to the plaquette of the lattice (see Fig. 7 central panel), (a) a data-qubit error occurring on the vertical link ℓ enters the equivalence $\hat\sigma$ if the product $\sigma_\ell \sigma_{\ell'} = -1$ where the equivalences $\sigma_\ell, \sigma_{\ell'}$ are on the left and on the right of ℓ. (b) Measurement errors occurring on the horizontal link enters the equivalence $\hat\sigma$ if the product $\sigma_\ell \sigma_{\ell'} = -1$ where the equivalences $\sigma_\ell, \sigma_{\ell'}$ are at the top and the bottom of ℓ. For the error E': given a reference error E described by the binary variables v_ℓ and h_ℓ and the equivalences $\hat\sigma$, the error E' equivalent to E up to $\hat\sigma$ is parametrized by v'_ℓ and h'_ℓ with (a) $v'_\ell = v_\ell \sigma_\ell \sigma_{\ell'}$ where the equivalences $\sigma_\ell, \sigma_{\ell'}$ are on the left and on the right of ℓ and (b) for a horizontal link ℓ, h'_ℓ is given by $h'_\ell = h_\ell \sigma_\ell \sigma_{\ell'}$ where the equivalences σ_ℓ and $\sigma_{\ell'}$ are on top and at the bottom of ℓ.

	Errors					Couplings	
z_p	z_q	z_r	v_ℓ	h_ℓ	$J_1 h_\ell$	$J_2 v_\ell$	$J_3 v_\ell h_\ell$
$+1$	$+1$	$+1$	$+1$	$+1$	F	F	F
$+1$	$+1$	-1	-1	-1	AF	AF	F
$+1$	-1	$+1$	$+1$	-1	AF	F	AF
$+1$	-1	-1	-1	$+1$	F	AF	AF
-1	$+1$	$+1$	-1	$+1$	F	AF	AF
-1	$+1$	-1	$+1$	-1	AF	F	AF
-1	-1	$+1$	-1	-1	AF	AF	F
-1	-1	-1	$+1$	$+1$	F	F	F

Table 1: The binary variables z_p, z_q, z_r, drawn randomly with probabilities p, q, r, represent the fundamental errors and fix the sign of the variables v_ℓ, h_ℓ and thus of the couplings (ferromagnetic (F) or antiferromagnetic (AF)) of the noise Hamiltonian of the random bond Ising model Eq. (17).

of these couplings are fixed by the Nishimori conditions (see Eq. (14)) while the type of interactions (ferromagnetic or antiferromagnetic) is assigned by drawing three random binary variables $z_p, z_q, z_r \in \{\pm 1\}$ with probability p, q, r for each link of the triangular lattice. These variables take the negative sign if one of fundamental processes (data-qubit error, measurement error, correlated error) is present in the reference error E. Therefore, they fix the signs of the variables v_ℓ and h_ℓ and thus of the couplings entering the Hamiltonian of Eq. (17) according to Table 1.

C.2 Finite-size scaling and transition points

Determining the code threshold requires in general the computation of the scaling with the system size of the free energy cost of a domain wall as reported in Ref. [16]. However since in [23] the threshold of the quantum code has been proven to correspond to the phase transition point of the statistical mechanics model, in this work we locate the code threshold by looking at the critical point of the random Ising Hamiltonian. In the case of zero disorder the system is completely magnetized and a convenient order parameter is given by the total magnetization: $M = \sum_{\vec{x}} \sigma_{\vec{x}}$ where $\vec{x}$ denotes a site in the 2D lattice of linear size L (here we consider lattices with $L = 16, 24$ and 32). Instead of looking at the behavior of M for different system sizes, we define the Fourier transform of the spin correlation function $\langle \sigma_0 \sigma_{\vec{x}} \rangle$ as

$$\hat{G}_L(\vec{k}) = \sum_{\vec{x}} \langle \sigma_0 \sigma_{\vec{x}} \rangle_L \, e^{i \vec{k} \cdot \vec{x}} \tag{18}$$

and we extract the correlation length ξ_L in the unit of lattice spacing from

$$\xi_L = \frac{1}{2\sin(q_{\min}/2)}\sqrt{\frac{\langle \hat{G}_L(0)\rangle_{\mathrm{av}}}{\langle \hat{G}_L(q_{\min})\rangle_{\mathrm{av}}} - 1} \tag{19}$$

where $q_{\min} = (2\pi/L, 0)$. The brackets $\langle\cdots\rangle_{\mathrm{av}}$ denote averages over the quenched disorder distribution (in addition to the thermal average). Near the critical temperature T_c, the correlation length ξ_L is expected to obey a scaling relation [65] of the form $\xi_L/L \sim f(L^{1/\nu}(T - T_c))$, where f is an unknown scaling function and ν is the critical exponent related to the correlation length. Therefore, at the critical temperature T_c, the quantity ξ_L/L becomes independent of the temperature and T_c can be found by locating the temperature at which the lines representing ξ_L/L for different L intersect. Some examples of the comparisons of ξ_L/L are given in Fig. 9 for the case $p = q = 2r$. In each figures, the vertical line indicates the temperature at which ξ_L/L cross for three different lattice sizes L. The figure without the vertical line shows the case in which there is no phase transition.

For computing the spin correlation function and thus the correlation length we use the standard Metropolis algorithm [66] interspersed with parallel tempering between adjacent temperatures to allow efficient sampling of the low temperature phase [67, 68]. Moreover, for a given quenched disorder distribution, the spin configurations are swapped after $N_{\mathrm{met}} = 800$ Metropolis steps following the parallel tempering algorithm (except the data point for $p = q, r = 0$ with $p = 0.11$ (open blue circle in Fig. 2 of the main text) where $N_{\mathrm{met}} = 8000$ Metropolis steps are performed). Metropolis steps/spin-swap combination are repeated 10000 times. The spin correlation function is measured during the Metropolis steps. Averages are taken over the Metropolis steps and then these averages serve as one of the jack-knife bins used to determine the statistical error. We use 250 different quenched disorder samples to compute the average of the correlation length over the disorder distribution. Table 2 lists the details of the numerics: for every case shown in Fig. 2 of the main text, when the probability p is fixed to the values reported in the second column of the table, we found the critical temperatures reported in the third column when the simulations are carried with a number of metropolis step N_{met}.

For assessing whether the random bond Ising model has thermalized we perform additional checks near the quenched probabilities at which the thermal transition crosses the Nishimori lines (Fig. 2 of the main text). In Fig. 10, we compare the effects of increasing the Metropolis steps between parallel tempering spin-swap steps. Comparing the two top panels of Fig. 10, we observe that there is no noticeable difference in the crossing of the lines ξ_L/L when the number of Metropolis steps is increased from 800 (panel (a)) to 8000 (panel (b)) for the probability $p = q = r = 0.070$. On the other hand, the comparison of the two bottom panels in Fig. 10 ($p = q = r = 0.075$) shows that the crossing of ξ_L/L for the three different lattice sizes, which seems to suggest a phase transition in panel (c) for 800 Metropolis steps, almost disappears for 8000 Metropolis steps (panel (d)) and the lines of ξ_L/L overlap for all the temperatures $T \lesssim 1.35$. Therefore we conclude that there is no transition at the quenched disorder probability $p = q = r = 0.075$ and the threshold probability should be larger than 0.070 but smaller than 0.075.

In Table 3, we list the largest p's at which MC simulations (run for $N_{\mathrm{met}} = 8000$) show a phase transition and the smallest p's at which MC simulations do not show a phase transition for the cases (I)-(IV). Figure 3 in the main text is produced with these data. The central value of these two p's is taken as the estimated threshold probability and a half of the difference is given as the error estimate. A more precise determination of the threshold probability will require even longer Metropolis updates between parallel tempering steps.

D Minimum-weight-perfect-matching decoding

To motivate this decoding strategy, observe that the probability of some error pattern E consisting of n_i errors of the $i-$th type, where $i = p, q$ or r) is proportional to

$$p(E) \propto \left(\frac{p}{1-p}\right)^{n_p(E)} \left(\frac{q}{1-q}\right)^{n_q(E)} \left(\frac{r}{1-r}\right)^{n_r(E)}. \tag{20}$$

case	p	T_c	N_{met}
(I) $p = q = 2r$	0.001	3.86(2)	800
	0.020	3.72(2)	800
	0.040	3.38(2)	800
	0.060	2.93(2)	800
	0.080	2.00(2)	800
	0.090	1.43(2)	800
(II) $p = q = r$	0.000	3.64(2)	800
	0.020	3.21(2)	800
	0.040	2.74(2)	800
	0.060	2.19(2)	800
	0.070	1.67(2)	800
(III) $p = q = r/2$	0.001	3.38(2)	800
	0.020	2.75(2)	800
	0.030	2.44(2)	800
	0.040	2.07(2)	800
	0.045	1.80(2)	800
(IV) $p = q, r = 0$	0.000	2.266(2)	800
	0.060	1.760(5)	800
	0.080	1.565(5)	800
	0.100	1.32(1)	800
	0.105	1.16(1)	800
	0.110	0.93(5)	8000

Table 2: Critical temperatures T_c for the different cases (I)-(IV) of quenched disorder from the Monte Carlo simulations. The columns report the case analyzed, the quenched disorder probability p, the corresponding critical temperature T_c obtained by performing N_{met} Metropolis steps.
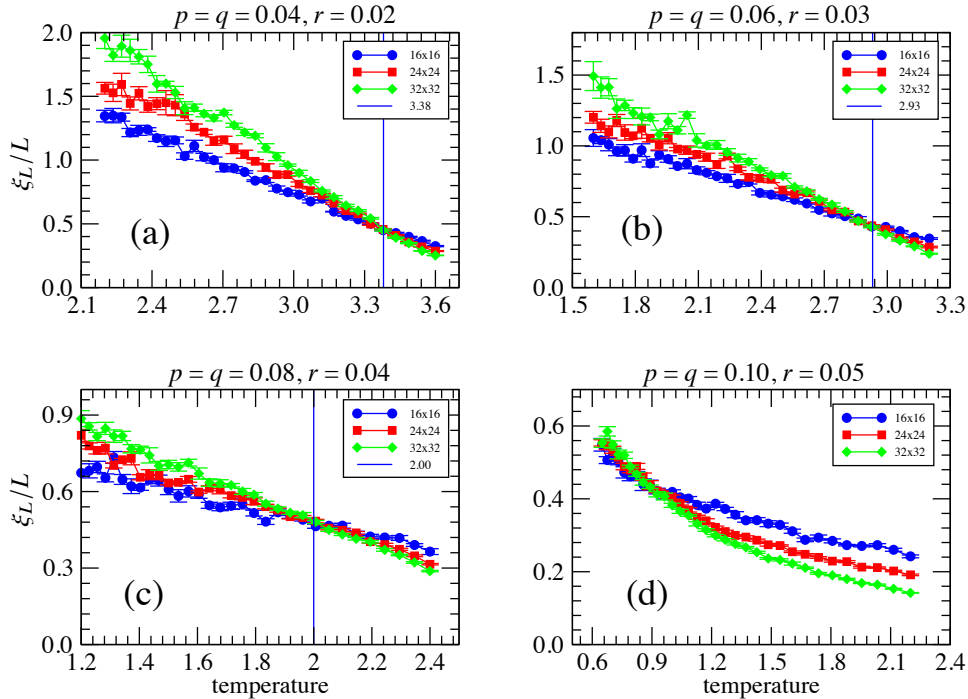

Figure 9: Critical temperatures from finite size scaling. The critical temperatures are found by locating the points where the lines representing the quantities ξ_L/L cross each other for different systems sizes L. Examples of the occurrence of a phase transition are shown in (a) for $p = q = 0.04, r = 0.02$, in (b) for $p = q = 0.06, r = 0.03$, in (c) for $p = q = 0.08, r = 0.04$. If it is not possible to identify a temperature where the lines with the quantities ξ_L/L cross, the transition does not occur. This is shown in panel (d) for $p = q = 0.10, r = 0.05$.

case	largest p showing a critical temperature	smallest p not showing a critical temperature
(I) $p = q = 2r$	0.090	0.095
(II) $p = q = r$	0.070	0.075
(III) $p = q = r/2$	0.045	0.050
(IV) $p = q, r = 0$	0.110	0.115

Table 3: Lower and upper bounds of the probability p between which the system exhibits a critical temperature for the cases (I)-(IV). The threshold probabilities and the associated errors shown in Fig. 3 of the main text for the statistical mechanics model are estimated by the mean and the deviation of these bounds for p.
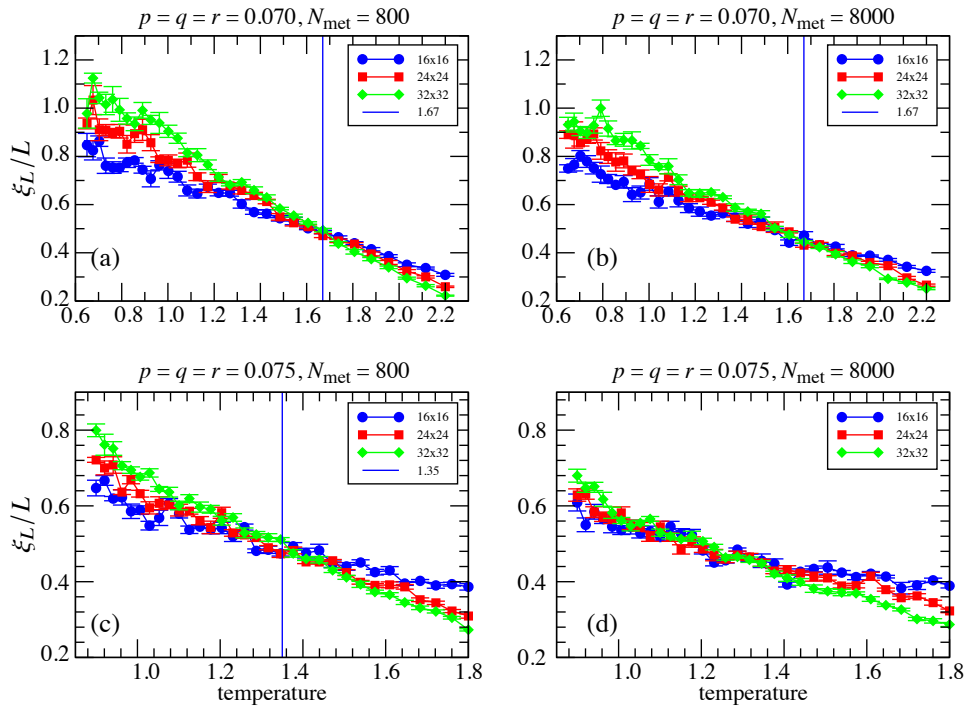

Figure 10: Effect of increasing the Metropolis steps. The quantity ξ_L/L computed with (a) 800 Metropolis steps and (b) 8000 Metropolis steps is shown for $p = q = r = 0.07$ before the Nishimori line is crossed. In this case a transition is identified at a temperature $T_c = 1.67$. When instead $p = q = r = 0.075$ after the Nishimori line is crossed, a transition at $T_c = 1.35$ seems to appear when ξ_L/L is computed with 800 Metropolis steps (panel (c)). However, when we compute ξ_L/L with 8000 Metropolis steps (d) the lines do not cross any more and overlap for the temperature less than 1.35. This implies that a transition does not occur when $p = q = r = 0.075$.
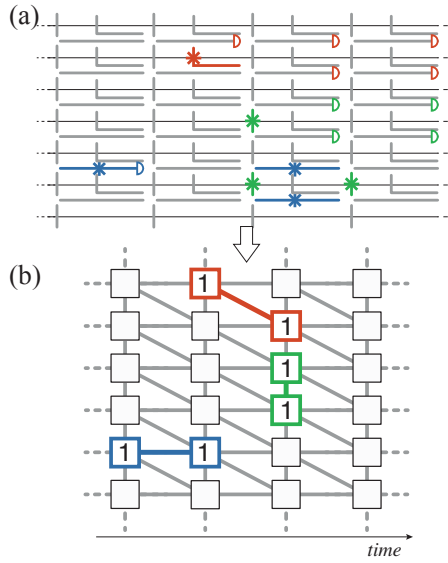
Figure 11: (a) Example of an error graph generated by physical errors. Positions where qubit, measurement and correlated data phase-flip and measurement errors can happen are represented by vertical, horizontal and L-shaped edges, respectively. Colored links represent occured error events while colored semicircles represent the ancilla qubits triggered by an error event. From this graph we derive syndrome volume which is the input for the MWPM decoder. (b) Syndrome volume graph derived by the error graph in (a). The boxes labeled with 1s represent defects, i.e. ancilla qubits where the measurement outcome differs from the previous round of measurements. The colored diagonal red, vertical green and horizontal edges represent a possible matching of the defects.
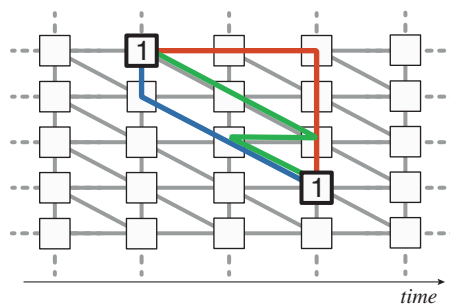

Figure 12: Simplification in computing the distance of two defects (boxes labeled with 1s) in the triangular lattice. Depending on the values of the weights w_p, w_q, w_r the shortest path between any two vertices can always be found by considering error patterns consisting of two types of errors only. If $w_r > w_p + w_q$ (pq case) the shortest path is the regular Manhattan distance of the square lattice, counting the number of p and q edges between the two defects (red path). If $w_r + w_p < w_q$ (pr case) the shortest path is composed by only p and r edges (blue path). Similarly, if $w_r + w_p > w_q$ (qr case) the shortest path is composed by only q and r (green path).

Given some observed collection of stabilizer measurement outcomes (the syndrome), repairing that syndrome amounts to finding a configuration of errors where every violated syndrome (a defect) is the endpoint of one error string, an assignment known as a (perfect) matching in graph theory (Fig. 11). Now Eq. (20) implies that finding the most likely among all these is equivalent to finding an error configuration with the lowest weight $w_{\text{tot}} = -\log(p(E)) = n_p w_p + n_q w_q + n_r w_r$, hence *minimum weight* perfect matching. By creating a graph with all defects as vertices and the weight between every pair of nodes given by the number of error locations times error type weight ($w_p = -\log \frac{p}{1-p}$ etc.) between the respective stabilizers, this is what is solved by the above mentioned Blossom algorithm (implemented e.g. in [69]). As a side remark, we note that the QEC code at hand does have a boundary, i.e. error strings extending to a boundary qubit only create a single defect, which is not immediately amenable to the above but can be incorporated by creating a copy of the matching graph with weights set to zero and then putting an edge between every defect and its virtual partner with the corresponding weight of matching to the code boundary [70]. Due to the addition of r-type edges the calculation of weights going into the matching problem has to be adapted as well. While in general we would have to compute the shortest path on a triangular lattice with the three types of edges being weighted respectively, which can be done by a Dijkstra-type algorithm, we can make a vast simplification by observing that the shortest path between any two vertices in our setting can always be found by considering error patterns consisting of two types of errors only. To see this, observe that any single edge in a certain path can be replaced by the other two edges in the triangle, which are hence of the other two types, e.g. a particular r-edge can be circumvented by moving across the neighboring p- and q-edge. Furthermore all vertices are connected to at least one edge of each type and the weights are globally the same (up to the edge type). Let us assume that $w_r < w_p + w_q$ (if this was not the case the metric would not change compared to the case without r-edges). Now the two possible cases are that either $w_r + w_p < w_q$ or $w_r + w_p > w_q$. In the former case, we replace all q-moves by the same amount of r- and p- moves, in the latter case we do the opposite and replace all p-moves by q- and r-moves, which shows that given a path, we can always find a path of lower weight by eliminating one of the edge types. In the borderline case $w_r + w_p = w_q$ both replacements are admissible, such that the statement holds that we can eliminate one edge type (see Fig. 12). This shows that a minimum weight path between any two vertices necessarily lies on one of the three sublattices and the path weight on a sublattice is given by the Manhattan distance on the respective sublattice:

$$wt(s_1, s_2) = \min \begin{cases} ||s_1 - s_2||_1^{pq} \\ ||s_1 - s_2||_1^{pr} \\ ||s_1 - s_2||_1^{qr} \end{cases} . \tag{21}$$

Here the two superscript labels indicate the sublattice of the triangular lattice, that consists of the two types of edges. In the pq case this is just the regular Manhattan distance, counting the number of p and q edges between the two vertices at hand. In case of pr and qr this is the Manhattan distance on a skewed square lattice, but by simply rotating the basis vectors we can still use the Euclidean position vectors to compute the Manhattan distance on the respective sublattices. The weight function assigning the weight between any two defects then simply takes the minimum over the three weighted Manhattan distances. Having assigned the edge weights, we can then solve the minimum weight matching problem, which amounts to finding the most likely collection of errors explaining the observed syndrome in a trial run, from which we can therefore deduce the recovery operation we should apply to the code (or in which way we should update the "Pauli frame"). The figure of merit we are after is the logical error rate, the probability that the decoding strategy fails. In order to cleanly decide in our simulation whether a logical error happened or did not happen, we initialize a perfect codeword, we then simulate d subsequent QEC-cycles, each consisting of injecting phase-flip errors with probability p on each data qubit followed by a noisy syndrome extraction (the syndrome bit being flipped with probability q). We furthermore inject the third error type needed, namely a correlated flip of the data-qubit accompanied by a flip of one of the syndrome measurement outcomes adjacent to it (depending on the order of the two-qubit gates this deterministically happens on the syndrome bit sitting either to the left or to the right of the data qubit). This type of error is injected with probability r. In order to make sure that we can declare the binary outcome of the trial run logical error/no logical error, we let the final

measurement be perfect, i.e. in the final round the syndrome bits never flip (with q or r). The final measurement being perfect ensures that matching up all defects will with certainty put the data qubits back into a code state. The simulation results are shown in Fig. 13.
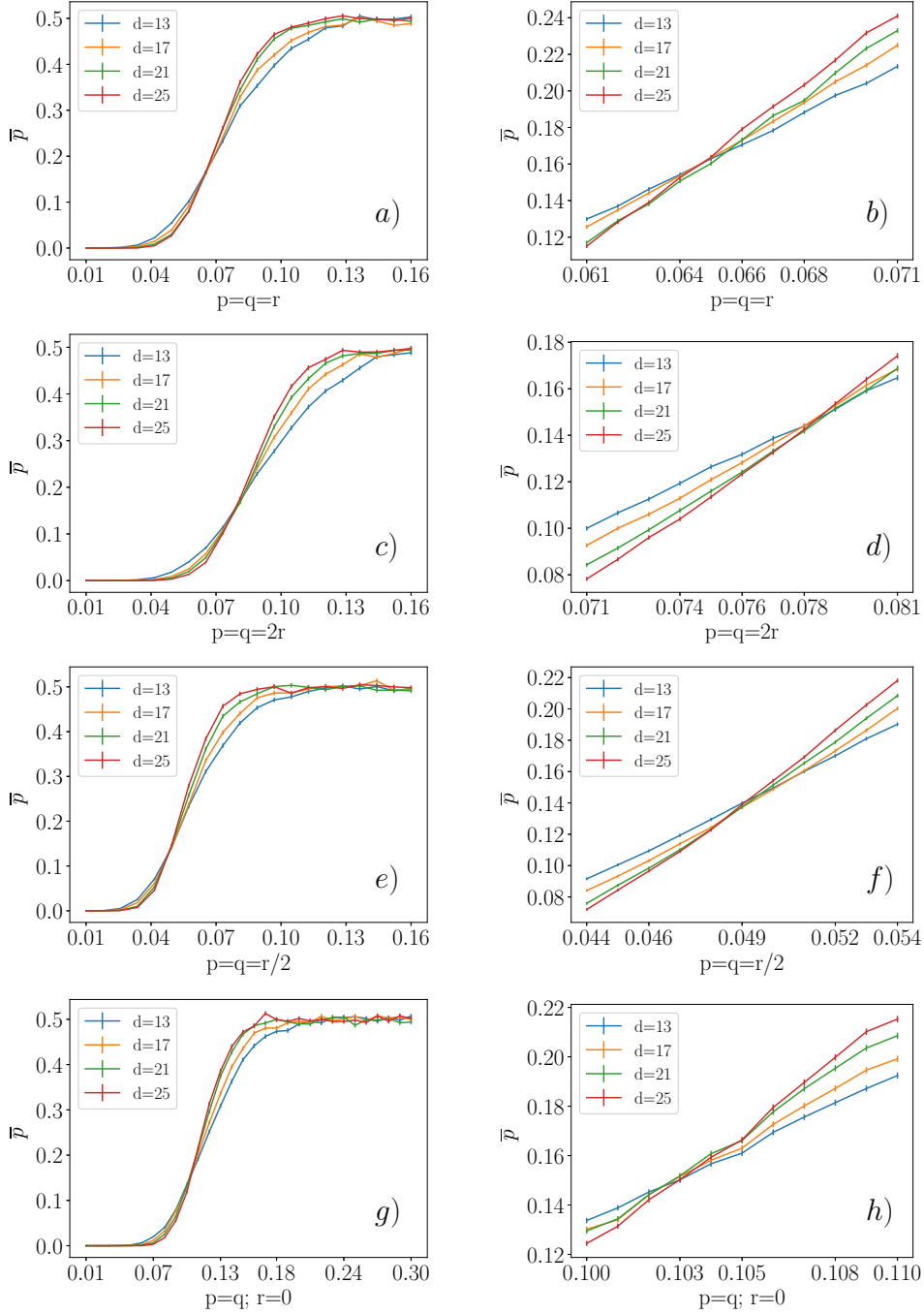
Figure 13: The logical error rate obtained with the minimum weight perfect matching decoder on the triangular syndrome lattice. Left column is a broad sweep, right column is a closeup around the transition point. We plot the average logical error rate as a function of the input error rates p, q, and r. Shown are the three cases $p = q = r$ in a)+b), $p = q = 2r$ in c)+d), $p = q = r/2$ in e)+f) and $p = q, r = 0$ in g)+h) . On the left for the broader scan, we use 10^4 samples, data points in the right colum are averaged over 10^5 samples. In each case, we observe a transition from error suppression to error enhancement with increasing error rate. This transition is signified by the behavior of the logical error rate when increasing the code size (the distance d): for small p, increasing the distance leads to a suppression of the logical error rate, whereas for error rates beyond an inflection point increasing the code size instead leads to an increased logical error rate. We estimate the threshold from the region where we can not distinguish the logical error rates within errorbars.