CERTIFIED LOWER BOUNDS ON THE OPTIMAL-DECODING THRESHOLD OF THE SURFACE CODE

TEJASVI SINGH TOMAR

Abstract. Under independent bit flips, rigorous lower bounds on the threshold of the surface code from earlier work stay below 0.04, while numerical estimates place the optimal threshold near 0.109. By counting labelled failure paths with exact transfer-matrix certificates, we prove without sampling that every minimum-weight decoder has threshold at least 0.0465 on the planar and rotated codes. We then give statistical certificates that the optimal threshold of the planar code is at least 0.102 under bit flips and at least 0.02 under phenomenological noise. Each certificate bounds an auxiliary decoder that corrects each block from a window around it, and optimal decoding does at least as well. A deterministic lemma reduces the failure of the auxiliary decoder to rare bad blocks, whose probabilities are bounded by a pre-registered Monte Carlo experiment. The confidence holds under a binomial model of the counts and two tested properties of each decoding program. All these results are proved in the proof assistant Lean 4, the statistical ones with these conditions as hypotheses.

1. Introduction

The surface code [20, 7] stores a logical qubit reliably as long as the physical error rate stays below a threshold. In the simplest setting, code capacity, each qubit suffers a bit flip independently with probability p and the syndrome is measured without error. Decoding one error sector is then a problem on a planar graph. For optimal decoding, which picks the most likely logical class given the syndrome, this problem is the random-bond Ising model on the Nishimori line [25, 7], and the optimal threshold is believed to be the Nishimori point, estimated numerically as 0.10917(3) [28]. Minimum-weight matching, the standard efficient decoder, has a threshold near 0.1031 in simulations [34].

Rigorous lower bounds are much smaller. Dennis, Kitaev, Landahl and Preskill [7] showed that minimum-weight decoding, and hence optimal decoding, succeeds with probability tending to one for p below a constant, which their argument puts at 0.0373. For the toric code, a decoding failure produces a homologically nontrivial cycle on which at least half of the edges are errors, and the expected number of such cycles decays for these p. The value 0.0373 uses a numerical estimate of the connective constant μ of self-avoiding walks on the square lattice; with the proven bound $\mu \le 2.662343$ [5] the same argument gives 0.0366. To our knowledge, no larger rigorous code-capacity threshold has been published for either decoder. Other rigorous bounds, from a recent method for general local noise channels [19] and from the bounds for low-density parity-check codes of [8] (see also [21]), are lower. Table 1 collects these values, together with those for measurement errors at the same rate as qubit errors.

On the side of statistical mechanics, Horiguchi and Morita extended the Peierls argument to the $\pm J$ model on the square lattice and showed that a ferromagnetic phase exists when the

Date: November 2026.

2020 *Mathematics Subject Classification.* 81P73; 94B35, 82B44, 62F25, 68V20.

Key words and phrases. Quantum error correction, surface code, toric code, threshold, maximum-likelihood decoding, minimum-weight matching, random-bond Ising model, Nishimori line, statistical certificate, computer-assisted proof, Lean.

	bit flips	bit flips and measurement errors
numerical estimate, optimal decoding	0.10917(3) [28]	about 0.033 [26]
numerical estimate, matching	0.1031 [34]	0.0293(2) [34]
earlier rigorous bound, minimum weight (hence optimal)	0.0366 [5, 7]	0.0112 [7, 30]
earlier rigorous bounds, other methods	0.034 [19]; about 0.029 for the toric code [8]	
this paper, minimum weight (hence optimal)	0.0465 (Theorems 2.1–2.3)	0.016 (Theorem 2.4)
this paper, optimal, statistical	0.102 (Theorem 2.5)	0.02 (Theorem 2.6)

TABLE 1. Thresholds of the surface code under independent noise. The earlier rigorous bounds in the third row are those of the argument of [7] with the proven bounds 2.662343 [5] and 4.739 [30] on the connective constants of self-avoiding walks on the square and the cubic lattice; with numerical estimates of these constants, the authors of [7] report 0.0373 and 0.0114. The value 0.0465 is for the planar and rotated codes; for the toric code we prove 0.046. The statistical results hold with the confidence of Section 2.4.

fraction of antiferromagnetic bonds is below $1 - 0.97394 \approx 0.026$ [16]. Rigorous results on phase transitions along the Nishimori line in finite dimensions remain scarce [27]; for continuous spins in three dimensions see [12]. Numerically, the phase boundary of the square-lattice model is reentrant below the Nishimori point [1, 28], so the matching threshold lies below the optimal one. Reentrance has been proved, with computer assistance, for a decorated square lattice with independent $\pm J$ couplings [29].

Thresholds proved for renormalization-group and cellular-automaton decoders are many orders of magnitude smaller [3, 22]. Under circuit-level noise, the path argument gives 1.7×10^{-4} for the circuit model of [7], matching thresholds of 7.4×10^{-4} [11] and 2×10^{-3} [33] have been proved, each for its own circuit and noise model, and the union-find and greedy decoders are known to have finite thresholds [35].

Critical points have been certified statistically in percolation. The method of Balister, Bollobás and Walters [2] turns Monte Carlo estimates into confidence intervals with exactly computed error probabilities, and it gives intervals of width at most 5×10^{-4} at confidence 99.9999% for the site and bond percolation thresholds of the eleven Archimedean lattices [31]. We are not aware of a comparable certificate for an error-correction threshold.

1.1. **Results.** Without sampling, every minimum-weight decoder, and hence maximum-likelihood decoding, has threshold at least 0.0465 on the planar and rotated codes and at least 0.046 on the toric code (Theorems 2.1–2.3). Under phenomenological noise, in which data qubits and syndrome bits fail independently with the same probability, every minimum-weight decoder of the planar code has threshold at least 0.016 (Theorem 2.4). With sampling, maximum-likelihood decoding of the planar code has threshold at least 0.102 under independent bit flips (Theorem 2.5) and at least 0.02 under phenomenological noise (Theorem 2.6). These two are statistical results. They hold with the confidence defined in Section 2.4, which rests on registered Monte Carlo experiments, on a binomial model of their counts and on two tested properties of the decoding programs. Their bounds become useful only at distances of order 10^5. Table 1 compares these thresholds with earlier ones, and Section 2.3 states all results with their constants.

1.2. **Methods.** The argument of [7] bounds the failure probability by the probability of an increasing event, one that is preserved when errors are added: the presence of a long path on which at least half of the edges are errors. Every argument that applies a union bound to paths, clusters or other sets that the error must contain does the same, and for matching, Monte Carlo estimates suggest that no such argument reaches $p = 0.093$ (Section 7.1). The sampling-free bounds stay in this class of arguments and push it further by exact counting; the statistical certificates leave it.

For the sampling-free bounds (Section 3) we fix a failing minimum-weight correction C^* that minimizes a linear tie-break functional. Its symmetric difference with the error contains a path between the rough boundaries whose edges are labelled as errors or corrections. Exchanging a short stretch of the path for a geodesic, or its correction edges for a shortest pairing of their odd-degree checks, gives another failing correction (Figure 1); since C^* has least weight and, among failing minimizers, least functional, many labelled stretches cannot occur. The remaining labelled walks are counted by a transfer matrix with finite memory, and an integer potential that satisfies every row of the matrix certifies that their growth rate is below one for label weights that bound the probability of each walk (Section 3.4).

For the certificate of Theorem 2.5 we bound optimal decoding by an auxiliary decoder. It divides the code into square blocks (Figure 2). For each block it computes a minimum-weight correction on a window around the block, with edge weights given by the marginals of the window posterior, and keeps the edges that the block owns. A block is bad if the combined first-stage correction leaves a defect in it or a residual path from it to ℓ_∞ distance L; adding errors can remove both, so this event is not increasing. A deterministic lemma (Section 5.3) shows that a failure of the decoder forces a sparse hierarchy of bad blocks, which gives an explicit bound on the failure probability once blocks are bad with probability below a limit, about 4×10^{-4} for the block and window sizes of Theorem 2.5 (Appendix A.4). Under the two tested properties of the program (Section 2.4), whether a block is bad depends only on the errors near it, and blocks in equivalent positions are bad with the same probability, so only one number per kind of block has to be bounded. Monte Carlo sampling bounds these numbers with exact binomial confidence bounds, in the manner of the percolation certificates mentioned above. Optimal decoding does at least as well as the auxiliary decoder, and the failure probability of optimal decoding is non-decreasing in p and in the width and non-increasing in the distance (Section 2.5), so the bound at a single value of p, for codes made of whole blocks, gives a threshold.

Sections 4 and 5.5 carry both methods over to phenomenological noise, on its three-dimensional decoding graph. All theorems of Section 2 are proved in Lean 4 [6] with Mathlib [32] (Section 6), with the conditions of Theorems 2.5 and 2.6 taken as hypotheses; Section 6.4 lists what the formal proofs do not cover.

1.3. **Organization.** Section 2 defines the codes and decoders, states the main results, explains the confidence statement and reduces Theorem 2.5 to a bound for the auxiliary decoder. Section 3 proves the sampling-free bounds for bit flips and Section 4 the sampling-free bound for measurement errors. Section 5 describes the auxiliary decoder, proves the deterministic bound and reports the two certificates. Section 6 describes the formal proofs, and Section 7 discusses the limits of these methods. Appendix A proves Theorem 5.3 in full. Two supplements accompany the paper (see Code and data): the records of the experiments and of the formal checks, and the proof of a criterion for a threshold of optimal decoding in terms of quantities of a single finite window, which Section 7.2 states informally.

2. Setting and main results

2.1. **Decoding graphs.** All results concern one error sector of a surface code under independent errors, so they are statements about a decoding graph. A *decoding graph* G has a set V of checks, in some cases two boundary vertices b_1, b_2 that carry no parity constraint, and a finite set of edges, the error locations; an edge between a check and a boundary vertex is a *half-edge*. Under independent noise of strength p, the error E contains each edge independently with probability p. The syndrome $\partial E \subseteq V$ is the set of checks of odd degree in E. A *correction* is a set C of edges with $\partial C = \partial E$. If G has boundary vertices, C *fails* when $|C \cap \delta(b_1)|$ and $|E \cap \delta(b_1)|$ have different parity, where $\delta(b_1)$ is the set of half-edges at b_1; in particular, $E \oplus C$ then contains a path from b_1 to b_2. For the toric code, C fails when $E \oplus C$ is homologically nontrivial.

We use the following graphs; the rotated code and the toric code are written in coordinates in Section 3.

(a) The *planar code* $\mathcal{P}_{d,w}$ of distance $d \geq 2$ and width $w \geq 1$: checks (x, y) with $0 \leq x \leq w-1$ and $0 \leq y \leq d-2$, an edge between any two checks at Euclidean distance 1, and a half-edge from each check $(x, 0)$ to b_1 and from each check $(x, d-2)$ to b_2. The bottom and top sides are rough and the left and right sides smooth. For $w = d$ this is one error sector of the unrotated planar code of distance d; under independent X and Z errors the other sector is the same graph rotated by a right angle, so for square codes the statements below apply to it by symmetry.

(b) The *rotated code* $\mathcal{R}_d$ of odd distance d: the checks are the X checks of the rotated surface code with d^2 qubits, and the errors are Z errors. Each qubit is an edge between the two X checks that contain it; a qubit that lies in only one X check is a half-edge to b_1 or to b_2, according to the rough side on which it lies.

(c) The *toric code* $\mathcal{T}_m$ of side m: checks $\mathbb{Z}_m^2$, each joined to its right and upper neighbour.

(d) The *phenomenological graph* $\mathcal{Q}_{d,w,T}$ of the planar code with $T \geq 0$ noisy rounds of syndrome measurement and a final perfect round: vertices (t, x, y) with $0 \leq t \leq T$, $0 \leq x \leq w-1$ and $0 \leq y \leq d-2$, an edge between any two vertices at distance 1 (space-like edges are data errors, time-like edges are errors of syndrome bits), and half-edges from $(t, x, 0)$ to b_1 and from $(t, x, d-2)$ to b_2. Every edge has the same probability p, so data qubits and syndrome bits fail at the same rate.

2.2. **Decoders.** A *decoder* maps each syndrome to a logical class, possibly at random; it fails on E when it outputs a class other than that of E. The *maximum-likelihood* (ML) decoder outputs a class of largest posterior probability given the syndrome; it minimizes the failure probability among all decoders, randomized ones included. We write $P_{\mathrm{ML}}(G, p)$ for its failure probability. A *minimum-weight decoder* outputs a correction of least size $|C|$. Our counting bounds concern

$$P_{\mathrm{MW}}(G, p) = \Pr\big(\text{some correction of least size fails}\big),$$

so they hold for every rule of breaking ties. Every minimum-weight decoder with a fixed tie rule is a decoder, so $P_{\mathrm{ML}}(G, p) \leq P_{\mathrm{MW}}(G, p)$. A decoder has threshold at least p_* on a family of codes if its failure probability tends to zero as $d \to \infty$ for every $p < p_*$, along codes whose width w, and for the phenomenological graph also the number T of rounds, are at most polynomial in d. All our bounds give exponential decay in d for every $p \leq p_*$ along such codes.

2.3. **Results.** For $0 < p \leq p_* < \frac{1}{2}$ put

$$\rho(p_*, p) = \sqrt{\frac{p(1-p)}{p_*(1-p_*)}} \ \leq 1.$$

Theorems 2.1–2.4 are proved outright. Theorems 2.5 and 2.6 are statistical: they hold with the confidence defined in Section 2.4. Table 8 in Section 6 lists the Lean theorems.

Theorem 2.1 (planar code)**.** *Let* $0 < p \le 0.0465$*,* $d \ge 11$ *and* $w \ge 1$*, and put* $\lambda = 0.998\,\rho(0.0465, p)$*. Then*

$$P_{\mathrm{ML}}(\mathcal{P}_{d,w}, p) \le P_{\mathrm{MW}}(\mathcal{P}_{d,w}, p) \le w \cdot 398.91 \cdot \frac{\lambda^{d-11}}{1-\lambda}.$$

Theorem 2.2 (rotated code)**.** *Let* $0 < p \le 0.0465$*, let* $d \ge 11$ *be odd, and put* $\lambda = 0.998\,\rho(0.0465, p)$*. Then*

$$P_{\mathrm{ML}}(\mathcal{R}_{d}, p) \le P_{\mathrm{MW}}(\mathcal{R}_{d}, p) \le 2d \cdot 366{,}183{,}771.98 \cdot \frac{\lambda^{d-11}}{1-\lambda}.$$

Theorem 2.3 (toric code)**.** *Let* $0 < p \le 0.046$ *and* $m \ge 1201$*, and put* $\lambda = 0.997\,\rho(0.046, p)$*. Then*

$$P_{\mathrm{ML}}(\mathcal{T}_{m}, p) \le P_{\mathrm{MW}}(\mathcal{T}_{m}, p) \le m^2 \cdot 428{,}572{,}391.6 \cdot \frac{\lambda^{m-10}}{1-\lambda}.$$

For $0 < p \le 0.045$ *and* $m \ge 769$ *also* $P_{\mathrm{MW}}(\mathcal{T}_{m}, p) \le m^2 \cdot 383{,}157{,}811 \cdot {\lambda'}^{m-8}/(1-\lambda')$ *with* $\lambda' = 0.996\,\rho(0.045, p)$*.*

Theorem 2.4 (measurement errors)**.** *Let* $0 < p \le 0.016$*,* $d \ge 9$*,* $w \ge 1$ *and* $T \ge 0$*, and put* $\lambda = 0.998\,\rho(0.016, p)$*. Then*

$$P_{\mathrm{ML}}(\mathcal{Q}_{d,w,T}, p) \le P_{\mathrm{MW}}(\mathcal{Q}_{d,w,T}, p) \le (T+1)\,w \cdot 1.2827 \cdot \frac{\lambda^{d-9}}{1-\lambda}.$$

In each of Theorems 2.1–2.4, $\lambda < 1$ for every p in the stated range, so the failure probability decays exponentially in the distance; in Theorem 2.4 the bound grows linearly in the number $T+1$ of rounds, so the failure probability per round decays exponentially in d. The thresholds of minimum-weight decoding and of ML decoding are therefore at least 0.0465 for the planar and the rotated code, at least 0.046 for the toric code, and at least 0.016 for the planar code under phenomenological noise.

Theorem 2.5 (optimal decoding at $p = 0.102$)**.** *Let* $L = 280$ *and* $R = \frac{3}{2}L - 1 = 419$ *be the block side and the window radius of the registered program. For* $w \ge 1$ *and* $d \ge 41L + 1$ *put* $n_x = \max(6, \lceil w/L \rceil)$ *and* $n_y = \lfloor (d-1)/L \rfloor$*. With confidence* $1 - 10^{-9}$*, under the conditions* (V6)*,* (Hcov) *and* (Bin) *of Section 2.4, the following holds: for every* $p \in [0, 0.102]$*, every* $w \ge 1$ *and every* $d \ge 41L + 1$*,*

$$P_{\mathrm{ML}}(\mathcal{P}_{d,w}, p) \le 35\, n_x\, e^{-0.10\,(n_y - 40)} + 2.5 \times 10^4\, n_x\, e^{-0.02\,(n_y - 40)}. \tag{2.1}$$

In particular, under these conditions and with this confidence, if $w \ge 1$ *is a function of* d *that grows subexponentially, that is,* $w\, e^{-cd} \to 0$ *as* $d \to \infty$ *for every* $c > 0$*, then for every* $p \in [0, 0.102]$ *the probability* $P_{\mathrm{ML}}(\mathcal{P}_{d,w}, p)$ *tends to* 0 *and is at most* $e^{-(d-1)/28000}$ *for all large* d*. Hence the optimal threshold of the planar code under independent bit flips is at least* 0.102*.*

For $n_x = n_y = n$, as for $w = d - 1 = nL$, the right side of (2.1) is below 0.1 from $n = 1008$ and below 10^{-3} from $n = 1249$, that is, from distance 282,241 and 349,721. For square codes, $w = d$, we have $n_x = n_y + 1$, and the two values of n_y are the same, 1008 and 1249. These are the constants of the Lean proof; sharper constants are given in Remark 2.8.

Theorem 2.6 (optimal decoding with measurement errors at $p = 0.02$)**.** *Let* $L = 35$ *be the block side of the registered program of Section 5.5. For* $w \ge 1$*,* $T \ge 0$ *and* $d \ge 5L + 1$ *put* $n_x = \max(5, \lceil w/L \rceil)$*,* $n_t = \max(5, \lceil (T+1)/L \rceil)$ *and* $n_y = \lfloor (d-1)/L \rfloor$*. With confidence* $1 - 10^{-9}$*,*

under the conditions (V6), (Hcov) *and* (Bin) *for this program (Section 2.4), the following holds: for every* $p \in [0, 0.02]$*, every* $w \geq 1$*, every* $T \geq 0$ *and every* $d \geq 5L + 1$*,*

$$P_{\mathrm{ML}}(\mathcal{Q}_{d,w,T}, p) \leq 804\, n_t n_x\, e^{-0.1008\,(n_y-47)} + 5.11 \times 10^6\, n_t n_x\, e^{-0.007\,(n_y-47)}. \tag{2.2}$$

In particular, under these conditions and with this confidence, if $w \geq 1$ *and* T *are functions of* d *that grow subexponentially, that is,* $w\, e^{-cd} \to 0$ *and* $T\, e^{-cd} \to 0$ *as* $d \to \infty$ *for every* $c > 0$*, then for every* $p \in [0, 0.02]$ *the probability* $P_{\mathrm{ML}}(\mathcal{Q}_{d,w,T}, p)$ *tends to* 0 *and is at most* $e^{-(d-1)/28000}$ *for all large* d*. Hence the optimal threshold of the planar code under phenomenological noise is at least* 0.02*.*

For $n_t = n_x = n_y = n$, as for $T + 1 = w = d - 1 = nL$, the right side of (2.2) is below 0.1 from $n = 5018$ and below 10^{-3} from $n = 5713$, that is, from distance 175,631 and 199,956. As for Theorem 2.5, these constants come from the Lean proof; they are not optimized.

2.4. **The confidence statement.** Theorem 2.5 rests on a fixed program, the first stage of the auxiliary decoder of Section 5, and on a random experiment that was registered before it ran. The bound (2.1) is a deterministic statement S. The Lean proof derives S from (V6), (Hcov) and the hypothesis that seven unknown numbers ε_t are at most 3×10^{-4}, where ε_t is the probability that a reference block of type $t \in \{L0, L1, L2, \mathrm{bulk}, R2, R1, R0\}$ is bad under independent bit flips at $p = 0.102$ (Section 5.4). The experiment draws $n = 2 \times 10^5$ independent samples of each type and accepts if at most 17 samples of each type are bad.

Definition 2.7. A statement S holds *with confidence* $1 - \delta$ for a registered experiment if the experiment accepted and $\Pr(\text{the experiment accepts and } S \text{ is false}) \leq \delta$, where the probability is over the randomness of the experiment.

The conditions are the following.

(V6) The first stage of the decoder reads only the syndrome: for every code size and every block, its output on the block depends on the error only through the syndrome in the window of the block.

(Hcov) The first stage commutes with translations: if a code is embedded in a larger one at an offset by whole blocks, with the sides that a block touches aligned, then the larger code's output on the image of a block is the image of the smaller code's output on the block, for every syndrome.

(Bin) For each type t, the number of bad samples of type t has the binomial law with parameters n and ε_t.

Conditions (V6) and (Hcov) are properties of the registered program. The validations V2, V4 and V6 of Section 5.4 tested them and found no mismatch. Table 2 lists, for both certificates, what supports each condition and when it was tested. Validations are finite tests, so (V6) and (Hcov) remain validated hypotheses about the registered program, not proved facts. Condition (Bin) treats the pseudo-random numbers as independent uniform draws. The rounding of the flip probability is handled on paper, outside the formal proof. The sampler compares a uniform number in binary arithmetic with the floating-point number nearest to 0.102, so its flip probability differs from 0.102 by less than 1.2×10^{-16}. A sample uses fewer than 5.7×10^6 edges, so each ε_t changes by less than 6.9×10^{-10}. This change is far inside the margin of the acceptance rule. A type passes with probability at most $10^{-9}/7$ whenever its bad probability is at least 2.9367×10^{-4}. Hence, if some $\varepsilon_t > 3 \times 10^{-4}$, this type still passes with probability at most $10^{-9}/7$, and so does the experiment as a whole. Condition (Bin) also identifies the blocks sampled by the experiment with the reference blocks (Section 5.4); each sampled block lies in a patch of the code that contains all edges on which its bad event depends. Only the marginal law of each count enters.

The experiment accepted (Table 6). Under (V6), (Hcov) and (Bin), if some $\varepsilon_t > 3 \times 10^{-4}$ then the probability of acceptance is at most 10^{-9}, and otherwise S holds. This is the confidence $1 - 10^{-9}$ of Theorem 2.5. It is the confidence of a single experiment: no other experiment was registered for this statement, and during the run the counts were read only by a check that could stop the run without acceptance (Section 5.4).

Remark 2.8 (sharper constants). The seven exact one-sided Clopper–Pearson upper bounds at level $10^{-9}/7$ are at most 2.58406×10^{-4} (rounded up; the largest are those of types $L2$ and $R1$), so they are below $\hat{\varepsilon} = 2.5841 \times 10^{-4}$, with a margin larger than the effect of the rounding of the flip probability. By the Bonferroni inequality, all seven true probabilities lie below their bounds with probability at least $1 - 10^{-9}$. Evaluating the bound of Section 5 at $\hat{\varepsilon}$ in place of 3×10^{-4}, with the level parameter $a_1 = 16$ that minimizes the bound there, as the registration prescribes, gives

$$P_{\mathrm{ML}}(\mathcal{P}_{d,w}, p) \le 5.81 \times 10^2 \, n_x \, e^{-0.186\,(n_y - 40)} + 3.73 \times 10^4 \, n_x \, e^{-0.0376\,(n_y - 40)}$$

under the same conditions and with the same confidence $1 - 10^{-9}$. For $n_x = n_y = n$ it is below 0.1 from $n = 549$ and below 10^{-3} from $n = 678$, that is, from distance 153,721 and 189,841; for square codes the two values are $n_y = 550$ and $n_y = 678$. This bound rests on the numerical evaluation of Appendix A.4, which Lean does not check (Section 6.2).

Theorem 2.6 rests in the same way on a second registered experiment, with the program of Section 5.5. Its blocks are cubes of side $L = 35$ in the coordinates (t, x, y) of $\mathcal{Q}_{d,w,T}$, and its window radius is $R = 34$. For this program, (V6) and (Hcov) are the statements above, with translations by whole blocks in each of the three directions. Condition (Bin) says that for each of the 25 block types t the number of bad samples has the binomial law with parameters $n = 3 \times 10^6$ and ε_t, where ε_t is the probability that the reference block of type t is bad at $p = 0.02$; as above, it also identifies the sampled blocks with the reference blocks. The experiment accepts if no type has more than 14 bad samples. Under (V6), (Hcov) and (Bin), if some $\varepsilon_t > 1.85 \times 10^{-5}$ then the probability of acceptance is at most 10^{-9}, and otherwise (2.2) holds. The experiment accepted (Table 7), which gives the confidence $1 - 10^{-9}$ of Theorem 2.6. Its validations of (V6) and (Hcov) at the registered radius ran only after the analysis (Table 2). The flip probability of its sampler exceeds 0.02 by less than 1.8×10^{-17}, which changes each ε_t by less than 3×10^{-10}. A type passes with probability at most 4×10^{-11} whenever its bad probability is at least the Clopper–Pearson bound at $k = 14$, and this bound lies more than 10^{-7} below 1.85×10^{-5}. So this rounding does not affect the confidence either.

2.5. **From the auxiliary decoder to optimal decoding.** In Section 5 we bound the failure probability of an auxiliary decoder at $p = 0.102$ on codes made of whole blocks, $\mathcal{P}_{n_y L+1,\, n_x L}$ with $n_x \ge 6$ and $n_y \ge 41$, by the right side of (2.1), under the same conditions and with the same confidence. The first stage of this decoder is the registered program; its later stages are minimum-weight corrections with edge weights in $[0.98, 1.02]$. The Lean proof of Theorem 2.5 takes unit weights there, so it has no hypothesis on them; this is allowed because only P_{ML} is bounded, the experiment measures only the first stage, and every such auxiliary decoder bounds P_{ML}. Theorem 2.5 follows from this bound, the optimality of ML decoding and the next two lemmas, which hold for every code size.

Lemma 2.9 (monotonicity in p)**.** *For $0 \le p \le p' \le \frac{1}{2}$, $P_{\mathrm{ML}}(\mathcal{P}_{d,w}, p) \le P_{\mathrm{ML}}(\mathcal{P}_{d,w}, p')$.*

Proof. If $p = \frac{1}{2}$ there is nothing to prove. Otherwise let N contain each edge independently with probability $r = (p' - p)/(1 - 2p) \in [0, \frac{1}{2}]$, independently of E. Then $E' = E \oplus N$ contains each edge independently with probability $p(1 - r) + (1 - p)r = p'$. Consider the randomized decoder that, given ∂E, draws N, applies the ML decoder for p' to the syndrome $\partial E \oplus \partial N = \partial E'$, and

condition	certificate at $p = 0.102$ ($L = 280$, $R = 419$)	certificate at $p = 0.02$ ($L = 35$, $R = 34$)
(V6)	V6 at $(L, R) = (8, 11)$, $(16, 23)$ and $(24, 35)$ before the registration; at the registered $(280, 419)$ after the registration, before the first sample	V6 at $(9, 4)$ before the registration; at $R = L - 1$ only after the analysis, at $(9, 8)$ and $(35, 34)$
(Hcov)	the window-local design (Section 5.1); V2 and V4 at $(8, 11)$, $(16, 23)$ and $(24, 35)$ before the registration	V2 and V4 at $(9, 4)$ before the registration; at $R = L - 1$ only after the analysis, at $(9, 8)$ and $(35, 34)$
(Bin)	fixed n, fresh seeds, every sample counted; samples identified with the reference blocks by construction and by V4	as for $p = 0.102$
the program	V1, also at the registered geometry before the first sample, V3 and V5; the checks of the faster implementation (Section 5.4)	V1 at the registered geometry, V3 at $(9, 4)$

TABLE 2. The conditions of the two certificates and the validations that support them (Section 5.4); pairs (L, R) give the block side and the window radius. No validation had a mismatch. Sections R.1 and R.2 of the records supplement give every run with its number of comparisons.

adds the class of N to its output. Since the class of E' is the class of E plus the class of N, this decoder fails exactly when the ML decoder for p' fails on E'. Its failure probability is therefore $P_{\mathrm{ML}}(\mathcal{P}_{d,w}, p')$, and the ML decoder for p does at least as well. □

Proposition 2.10 (monotonicity in the size)**.** *For fixed $p \in [0, 1]$, $P_{\mathrm{ML}}(\mathcal{P}_{d,w}, p)$ is non-increasing in d and non-decreasing in w.*

Proof. From d to $d+1$: the checks of $\mathcal{P}_{d,w}$ are the checks of $\mathcal{P}_{d+1,w}$ in rows $0, \dots, d-2$. Read the vertical edges of $\mathcal{P}_{d+1,w}$ between rows $d-2$ and $d-1$ as the half-edges of $\mathcal{P}_{d,w}$ at its top side. The restriction of an error of $\mathcal{P}_{d+1,w}$ to these edges and to the edges below them is an independent error of $\mathcal{P}_{d,w}$, with the same class, since the class is the parity at b_1, and its syndrome is the syndrome of the large code in rows $0, \dots, d-2$. Any decoder of $\mathcal{P}_{d,w}$ therefore gives a decoder of $\mathcal{P}_{d+1,w}$ with the same failure probability, and ML decoding of $\mathcal{P}_{d+1,w}$ does at least as well.

From w to $w+1$: a decoder of $\mathcal{P}_{d,w}$ draws an independent error N on the edges of $\mathcal{P}_{d,w+1}$ that are not edges of $\mathcal{P}_{d,w}$, that is, the vertical edges and half-edges of column w and the horizontal edges between columns $w-1$ and w. From ∂E and N it computes the syndrome of $E \cup N$ in $\mathcal{P}_{d,w+1}$, applies ML decoding of $\mathcal{P}_{d,w+1}$ and adds the class of N, the parity of its half-edges at b_1. It fails exactly when ML decoding of $\mathcal{P}_{d,w+1}$ fails on $E \cup N$, an independent error of $\mathcal{P}_{d,w+1}$. Hence $P_{\mathrm{ML}}(\mathcal{P}_{d,w}, p) \le P_{\mathrm{ML}}(\mathcal{P}_{d,w+1}, p)$. □

Proof of Theorem 2.5. Let $0 \le p \le 0.102$, $w \ge 1$ and $d \ge 41L + 1$, and let n_x, n_y be as in the theorem, so that $n_x \ge 6$, $n_y \ge 41$, $w \le n_x L$ and $d \ge n_y L + 1$. Lemma 2.9 and Proposition 2.10 give

$$P_{\mathrm{ML}}(\mathcal{P}_{d,w}, p) \le P_{\mathrm{ML}}(\mathcal{P}_{d,w}, 0.102) \le P_{\mathrm{ML}}(\mathcal{P}_{n_y L+1,\, n_x L}, 0.102).$$

The last probability is at most the failure probability of the auxiliary decoder on $\mathcal{P}_{n_y L+1,\, n_x L}$ at $p = 0.102$, and by Section 5 this is at most the right side of (2.1). For the last statement, $n_y \ge (d-1)/L - 1$ gives $e^{-0.10(n_y - 40)} \le e^{-0.02(n_y - 40)} \le e^{0.82 - (d-1)/14000}$, so the right side of (2.1) is at most $2.6 \times 10^4\, e^{0.82}\, n_x\, e^{-(d-1)/14000}$. If w grows subexponentially in d, so does $n_x \le \max(6, w/L + 1)$, and this is at most $e^{-(d-1)/28000}$ for all large d. □

3. Sampling-free bounds

In this section we prove Theorems 2.1, 2.2 and 2.3 by the scheme outlined in Section 1.2. Two features are specific to this section: the smooth sides of the rotated code, which are not parallel to the lattice axes, and the seam of the tie-break on the torus.

3.1. **Coordinates.** All three decoding graphs consist of points of $\mathbb{Z}^2$, or of $\mathbb{Z}_m^2$, joined by unit steps, together with half-edges. We write $|u-v|$ for the ℓ^1 distance, and a lattice path between u and v with $|u-v|$ steps is a *geodesic*. Each half-edge is regarded as the unit step from its check to an *outer point*, a lattice point outside the graph, as follows.

Planar code. For the check (x', y') of $\mathcal{P}_{d,w}$ in Section 2.1 we write $(x, y) = (y', x')$. The checks are then the points of the box $B = [0, d-2] \times [0, w-1]$, each check $(0, y)$ has a half-edge to b_1 with outer point $(-1, y)$, and each check $(d-2, y)$ has a half-edge to b_2 with outer point $(d-1, y)$. These are the coordinates of the Lean statements, which write $n_x = d-1$ and $n_y = w$.

Rotated code. The d^2 data qubits of $\mathcal{R}_d$ sit at the points (a, b) with $0 \le a, b \le d-1$. For $0 \le i, j \le d-2$ the plaquette with corners (i, j), $(i+1, j)$, $(i, j+1)$, $(i+1, j+1)$ carries an X check when $i + j$ is even. The weight-two X checks act on the qubits $(i, 0), (i+1, 0)$ for odd i and on $(i, d-1), (i+1, d-1)$ for even i, $0 \le i \le d-2$; we place them at the positions $(i, -1)$ and $(i, d-1)$. The X check at position (i, j) gets the lattice coordinates $(x, y) = \big((i+j)/2, (i-j)/2\big)$, so that $x + y = i$ and $x - y = j$. Then:

(a) the checks are exactly the lattice points of $R_d = \{(x, y) \in \mathbb{Z}^2 : 0 \le x + y \le d-2,\ -1 \le x - y \le d-1\}$;
(b) two checks share a qubit if and only if they are lattice neighbours, and then they share exactly one;
(c) the qubits that lie in only one X check are the half-edges. At each check z they correspond one-to-one to the unit steps e for which $z + e$ lies beyond one of the *rough lines* $x + y = -1$ and $x + y = d-1$ and inside the *smooth range* $-1 \le x - y \le d-1$; the outer point of the half-edge is $z + e$. The half-edges across $x + y = -1$ are the qubits of the column $a = 0$ and join b_1; those across $x + y = d-1$ are the qubits of the column $a = d-1$ and join b_2. There are d of each.

These facts follow from the definitions. We also checked them by explicit construction for $d \le 21$, and the Lean development identifies the failure probability of the rotated code with that of this graph (Section 6.1). A check can have two half-edges to the same boundary vertex; they are different edges. We call the lines $x - y = -1$ and $x - y = d-1$ the bottom and the top *smooth sides*, and $h(z) = (d-1) - (x-y)$ the *height* of a point $z = (x, y)$ below the top side.

Toric code. We write the checks of $\mathcal{T}_m$ as (x, y) with $x, y \in \{0, \dots, m-1\}$; each check is joined to $(x \pm 1, y)$ and $(x, y \pm 1)$, modulo m.

3.2. **Tie-break, failure paths and exclusion rules.** For the planar and the rotated code let $\phi(x, y) = x + 9y$. For the toric code let $\phi(x, y) = x$, the representative in $\{0, \dots, m-1\}$; this is a function on $\mathbb{Z}_m^2$, which jumps across the *seam* between the columns $m-1$ and 0. For a set S of edges let $\Phi(S)$ be the sum of $\phi(u) + \phi(v)$ over the edges $\{u, v\}$ of S, where a half-edge contributes its check and its outer point.

Suppose that some correction of least size fails, and let C^* be one with the least value of Φ among the failing corrections of least size. The set $E \oplus C^*$ has even degree at every check. For the planar and the rotated code it has odd degree at b_1 and at b_2, so it contains a simple path γ from b_1 to b_2. Its first and last edges are half-edges, and its *interior* γ', the path without them, joins a check on the side of b_1 to a check on the side of b_2. So γ' has at least $d-2$ edges: in $\mathcal{P}_{d,w}$ the coordinate x runs from 0 to $d-2$, and in $\mathcal{R}_d$ every step changes $x + y$ by ± 1, from 0 to $d-2$. For the toric code $E \oplus C^*$ is not null-homologous, so it contains a simple cycle γ

that is not null-homologous, and γ has at least m edges; we put $\gamma' = \gamma$. In all cases $C^* \oplus \gamma$ is a correction, so $|\gamma \cap E| \geq |\gamma \cap C^*|$ because C^* has least size. We label the edges of γ in E by E and those in C^* by C; every edge of γ gets exactly one label.

A *window* is a sub-path σ of γ' between two checks $u \neq v$. We write n_E and n_C for the numbers of its labels, σ_E and σ_C for its sets of E- and C-edges, and T_σ for the set of checks of odd degree in σ_C. An *alternative* to a set S of edges is a set A of edges of the graph with the same checks of odd degree as S, such that A has an even number of half-edges at b_1 (planar and rotated code), respectively such that $|S| + |A| < m$ (toric code).

Lemma 3.1. *Let σ be a window from u to v.*

(i) *$n_C - n_E \leq |A|$ for every alternative A to σ;*
(ii) *if $n_C - n_E = |A|$ for an alternative A to σ, then $\Phi(\sigma_C) \leq \Phi(\sigma_E) + \Phi(A)$;*
(iii) *$n_C \leq |A|$ for every alternative A to σ_C;*
(iv) *if $n_C = |A|$ for an alternative A to σ_C, then $\Phi(\sigma_C) \leq \Phi(A)$.*

Proof. For an alternative A to σ put $C' = C^* \oplus \sigma \oplus A$. It is a correction, since $\sigma \oplus A$ has even degree at every check, and $|C'| = |C^*| - n_C + n_E + |A| - 2k$, where k is the number of edges of A in $(C^* \setminus \sigma_C) \cup \sigma_E$. It fails, because $C' \oplus C^* = \sigma \oplus A$ has an even number of half-edges at b_1, respectively has fewer than m edges and is therefore null-homologous. Since C^* has least size, (i) follows. In the case of (ii), $k = 0$, so C' is a failing correction of least size; it is the disjoint union of $C^* \setminus \sigma_C$, σ_E and A, so $\Phi(C') = \Phi(C^*) - \Phi(\sigma_C) + \Phi(\sigma_E) + \Phi(A)$, and the choice of C^* gives (ii). Parts (iii) and (iv) follow in the same way from $C'' = C^* \oplus \sigma_C \oplus A$, with $|C''| = |C^*| - n_C + |A| - 2k''$, where k'' is the number of edges of A in $C^* \setminus \sigma_C$. □

The certificates apply Lemma 3.1 with explicit alternatives (Figure 1): in (i) and (ii) a geodesic from u to v, and in (iii) and (iv) the union of geodesics that join the pairs of a perfect matching of T_σ of least total length $c(T_\sigma)$. If these geodesics overlap, the edges used an odd number of times form an alternative with fewer than $c(T_\sigma)$ edges; so when $n_C = c(T_\sigma)$, (iii) forces them to be disjoint, and (iv) applies to their union. In (ii) and (iv) the strongest test uses geodesics of least Φ. For $\phi = x + 9y$ the least Φ among all lattice geodesics from u to v is attained by the *canonical geodesic*, which takes its steps in increasing order of their ϕ-increments, that is, in the order $-y, -x, +x, +y$: exchanging two consecutive steps whose increments are in the wrong order lowers Φ. Two consecutive C-edges that turn a corner form a window with $n_C = 2 = |u - v|$, so by (ii) they turn from the smaller to the larger ϕ-increment whenever the other corner is available. We say that a window *fails a test* if (i)–(iv) are violated for one of these edge sets. By Lemma 3.1, no window of γ' fails a test, as long as every edge set used in a test is an alternative. Whether it is depends on the graph.

3.3. **The tests in each graph.** *Planar code.* Every lattice geodesic between two points of the box B stays in B. So every geodesic between checks of a window is an alternative (it has no half-edge), and the tests use all lattice geodesics, with no condition on the position of the window.

Rotated code. Two lemmas make the tests valid in R_d. The first handles the rough sides.

Lemma 3.2. *Let P be a set of unit steps forming paths between checks of R_d, all of whose points lie in the smooth range $-1 \leq x - y \leq d - 1$. Then $\mathcal{R}_d$ contains a set A of edges with the same checks of odd degree as P and an even number of half-edges at b_1 and at b_2, such that $|A| \leq |P|$, and $\Phi(A) = \Phi(P)$ if $|A| = |P|$.*

Proof. A point of P outside R_d lies beyond a rough line. A maximal stretch of a path of P outside R_d leaves R_d by a step from a check z to a point t and returns by a step from a point t' to a check z'; it stays beyond one rough line, since every step changes $x + y$ by ± 1. The steps

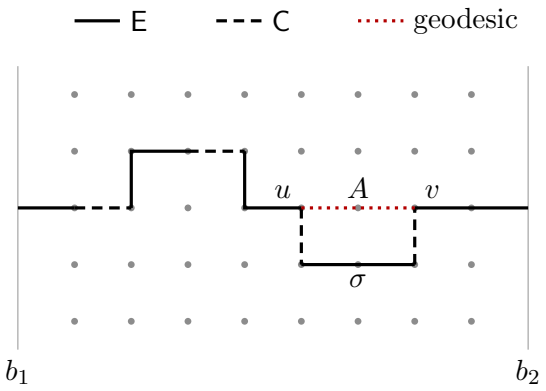


FIGURE 1. An example of a labelled failure path in the planar code $\mathcal{P}_{9,5}$, drawn in the coordinates of Section 3.1. These exchange the two axes of Section 2.1, so the rough sides are on the left and the right, and the half-edges join the checks of the columns $x = 0$ and $x = 7$ to b_1 and b_2. Let C^* be a failing correction of least size that minimizes the tie-break Φ. Then $E \oplus C^*$ contains a path γ from b_1 to b_2, whose edges in E are labelled E (solid) and whose edges in C^* are labelled C (dashed). The window σ of γ from u to v consists of the four edges through $(4,1)$, $(5,1)$ and $(6,1)$. Replacing σ by the geodesic A (dotted) gives another failing correction, $C^* \oplus \sigma \oplus A$. Since C^* has least size, every window satisfies $n_C - n_E \leq |u - v|$, here $2 - 2 \leq 2$. Windows that violate this inequality, or that attain it and lose the tie-break, cannot occur (Lemma 3.1), and the count of Section 3.4 includes only labelled paths without them.

$z \to t$ and $t' \to z'$ cross this line and end in the smooth range, so by (c) they are half-edges of z and of z' to the same boundary vertex, with outer points t and t'. Replace each such stretch by these two half-edges, and let A be the set of edges used an odd number of times. A stretch has at least two steps, and exactly two when $t = t'$; then the two half-edges have the same value of Φ as the two steps they replace. So $|A| \leq |P|$, with equality only if every stretch has two steps and no edge is used twice, in which case $\Phi(A) = \Phi(P)$. Half-edges are added in pairs at one boundary vertex. □

The second handles the smooth sides. Write $\xi = x - y$.

Lemma 3.3 (band lemma)**.** *Let $a, b \in \mathbb{Z}^2$, let $c \geq \max(\xi(a), \xi(b))$, and let P be a geodesic from a to b of least Φ among the geodesics all of whose points satisfy $\xi \leq c$. Then every point of P satisfies $\xi \geq \min\big(\xi(a), \xi(b), c - 1\big)$. Without the constraint $\xi \leq c$, every point of the canonical geodesic satisfies $\xi \geq \min(\xi(a), \xi(b))$.*

Proof. A geodesic uses one horizontal and one vertical direction. If its profile ξ is not monotone, the two directions are $+x, +y$ or $-x, -y$. In both cases one of them, s_1, raises ξ by 1 and has the smaller ϕ-increment ($+x$ against $+y$, and $-y$ against $-x$), while the other, s_2, lowers ξ by 1. Exchanging two consecutive steps s_2, s_1 into s_1, s_2 lowers Φ by $2(\phi(s_2) - \phi(s_1)) > 0$ and raises the point between them by 2 in ξ. Suppose that the least value l of ξ on P is below $\min(\xi(a), \xi(b), c - 1)$. At the first point of P with $\xi = l$, the path arrives by a step s_2 and leaves by a step s_1, from a point at level $l + 1 \leq c - 1$; the exchange keeps the path at $\xi \leq c$ and lowers Φ, a contradiction. Without the constraint the same exchange applies at every level, and the canonical geodesic is the unique geodesic of least Φ. □

For a window σ let ξ_σ be the least value of ξ on its points. The points of σ lie in R_d, so $\xi_\sigma \geq -1$, and since consecutive points differ by 1 in ξ and all satisfy $\xi \leq d-1$, also $\xi_\sigma \leq d-2$. The tests for $\mathcal{R}_d$ use, for each pair of checks of the window, a geodesic of least Φ among those whose points have height $h \geq 0$, that is, $\xi \leq c = d-1$; in (iii) and (iv) these are taken over the pairs of a least-length matching. By Lemma 3.3 and $\xi_\sigma \leq d-2$, every such geodesic stays at $\xi \geq \xi_\sigma \geq -1$, inside the smooth range. Lemma 3.2 turns it, or the union of the matching geodesics, into an edge set A of $\mathcal{R}_d$ with no more edges. If A has fewer edges, the window violates (i) or (iii) for the alternative A; otherwise $\Phi(A)$ equals the Φ of the geodesics. In both cases a window that fails a test violates Lemma 3.1. This holds for every d.

The tests of a window depend on its position only through the ceiling $\xi \leq d-1$. A geodesic between two points a, b of a window rises at most $|a-b| \leq M$ above $\max(\xi(a), \xi(b))$, where M bounds the number of edges of the windows we test. So if all points of a window have height at least M, the ceiling does not bind, and the tests are those of the planar code (all lattice geodesics).

Toric code. The windows we test have at most M edges, and every alternative used has at most M edges, with $2M < m$. So $|\sigma| + |A| < m$ for these edge sets, and the tests use all lattice geodesics on the torus, compared by the true Φ. Measure the position of a point by its *offset* o from the seam, the representative of its column in $\{-\lfloor m/2 \rfloor, \ldots, \lceil m/2 \rceil - 1\}$, so that $\phi = o$ for $o \geq 0$ and $\phi = m + o$ for $o < 0$. A window and its alternatives lie in the range of offsets spanned by the window. If this range does not contain both -1 and 0, the window lies on one side of the seam, ϕ is affine there, and the tests are those of the planar code with $\phi(x, y) = x$, whose canonical order is $-x$, then the vertical steps, then $+x$.

Lemma 3.4. *For a window of at most M edges, the outcome of every test is the same for all $m > 4M^2$.*

Proof. For $m > 2M$ the window and its alternatives do not wrap around the torus, and their points have offsets within M of the offset of the first point of the window. Only (ii) and (iv), and the choice of a geodesic or matching of least Φ, compare values of Φ, and each comparison is between two edge sets with the same number of edges, at most $2M$ in total, hence at most $4M$ endpoints. Each side equals $mN + Q$, where N counts its endpoints at negative offsets and Q is the sum of the offsets of its endpoints. The difference is $m\,\Delta N + \Delta Q$; ΔN does not depend on m, and ΔQ does not change if all offsets are shifted by the same amount, so measuring them from the first point of the window gives $|\Delta Q| \leq 4M \cdot M < m$. Hence the sign of the difference is the lexicographic sign of $(\Delta N, \Delta Q)$, which does not depend on m. □

The certificates evaluate the tests with a fixed large value in place of m, and the Lean statements require $m \geq L_0 = 4M(h_0 + M) + 1$ with $h_0 = K + M + 1$ (that is, $L_0 = 1201$ for $M = 10$ and $L_0 = 769$ for $M = 8$), which is larger than $4M^2$.

3.4. **Counting.** For a fixed labelled path, the event that it occurs as γ with these labels has probability at most $p^{n_E}(1-p)^{n_C}$, where n_E and n_C now count the labels of the whole path: its E-edges lie in E, and its C-edges lie in C^* and therefore not in E. Because $n_E \geq n_C$ on γ, this is at most $\omega_E^{n_E}\omega_C^{n_C}$ with $\omega_E = pX$ and $\omega_C = (1-p)/X$, for every $X \geq 1$.

Fix $M \geq 3$ and $K = M - 1$. A *state* is a walk of K unit steps through $K+1$ distinct points, with a label on each step, none of whose windows fails a test. A *successor* of a state appends one labelled step to a point outside the walk, such that no window of at most M edges that ends with the new step fails a test; the oldest step is then dropped, and $\omega(s \to s')$ is the weight of the new label. If a positive potential H on the states satisfies

$$\sum_{s' \text{ successor of } s} \omega(s \to s')\, H(s') \leq \lambda\, H(s) \qquad \text{for every state } s, \tag{3.1}$$

then by induction on n the labelled continuations of n steps of a state s weigh at most $\lambda^n H(s)/\min H$. Every labelled simple path none of whose windows of at most M edges fails a test is such a continuation of its first K steps, because each of its windows of at most M edges ends at some step and lies in the M-step suffix there.

Planar code. The tests do not depend on the position of the window, so the states are translation classes of labelled walks in $\mathbb{Z}^2$. A state whose points lie in B has no successor inside B that it lacks in $\mathbb{Z}^2$, so the rows (3.1) of the system on $\mathbb{Z}^2$ imply those of all states in B, and no boundary types occur.

Rotated code. Here the tests near the top side depend on the heights. Let $h_0 = 2K + 6$. A state whose last point has height $h < h_0$ carries h; it is a *near* state (s, h). A state whose last point has height $h \geq h_0$ is the translation class s of its walk and gets the potential $H(s)$ of the planar system; the tests of its successors are those of the planar code, because every point of a window that ends at the new point has height at least $h_0 - 1 - M = M + 3$. The rows to check are therefore: the rows of the near states, whose successors are near states or, at height h_0, far states; the rows of the far states at height exactly h_0, whose successors at height $h_0 - 1$ are near states; and the rows of the planar system for the far states above h_0. Near the bottom side and the rough sides the walks of $\mathcal{R}_d$ are among these walks, with the same tests, so these rows suffice for every $d \geq 11$; for small d some modelled heights do not occur.

Toric code. Here the tests of windows that straddle the seam depend on the offsets. Let $h_0 = K + M + 1$. A state whose last point has offset $o \in [-h_0, h_0 - 1]$ carries o; every other state is the translation class of its walk and gets the potential of the system on $\mathbb{Z}^2$ with $\phi(x, y) = x$. Every window of a far state and of its successors lies on one side of the seam, with offsets at least $h_0 - 1 - M = K$ or at most $-h_0 + M = -K - 1$, so its tests are those of the system with $\phi = x$. The rows to check are those of the near states, of the far states at the two interfaces $o = h_0$ and $o = -h_0 - 1$, whose successors at offsets $h_0 - 1$ and $-h_0$ are near states, and of the system with $\phi = x$. By Lemma 3.4 the tests, and hence the rows, are the same for all $m \geq L_0$. We also need $m \geq 2h_0 + 2M + 2$, so that the zone near the seam does not wrap around the torus; this is implied by $m \geq L_0$.

3.5. **The certificates.** Table 3 lists the certificates. Each is a positive integer potential on the states of the system described above, and every row (3.1) is checked in exact integer arithmetic. The potentials are at most 2^{34}.

code	M	h_0	states	transitions	p_*	X	λ	largest ratio
planar	10	–	3,883,082	19,106,700	0.0465	3.091	0.998	0.999306
rotated	10	24	$82{,}148{,}725 + 2 \cdot 3{,}883{,}082$	438,790,694	0.0465	3.091	0.998	$1 - 7.8 \cdot 10^{-10}$
toric	10	20	$157{,}107{,}748 + 3 \cdot 3{,}930{,}982$	832,352,608	0.046	3.12228	0.997	$1 - 1.4 \cdot 10^{-9}$
toric	8	16	$5{,}084{,}396 + 3 \cdot 159{,}004$	27,561,980	0.045	3.1943	0.996	$1 - 1.5 \cdot 10^{-9}$

TABLE 3. The certificates; p_*, X and λ are exact decimals. For the rotated and the toric code the states are the near states plus one copy of the far system for each interface and one for the rows above it, and likewise the transitions. The last column is the largest ratio of the two sides of (3.1), rounded to the digits shown; the exact check is in integers.

The constants of Theorems 2.1–2.3 are start sums of these potentials. With $\omega(s) = \prod \omega$ over the labels of the walk s:

(a) planar: $K_0 = (\omega_E + \omega_C)^2 \sum_s \omega(s) H(s)/\min H \leq 398.91$;

(b) rotated: $K_{\mathrm{rot}} = (\omega_E + \omega_C)^2 \max_h S(h)/\min H \le 366{,}183{,}771.98$, where $S(h)$ is the sum of $\omega(s)H$ over the states whose first point has height h, each taken with the potential of its near or far copy;
(c) toric: $K_{\mathrm{tor}} = (\omega_E + \omega_C) \max_o S(o)/\min H$, with $S(o)$ defined in the same way by the offset o of the first point; $K_{\mathrm{tor}} \le 428{,}572{,}391.6$ for $M = 10$ and $K_{\mathrm{tor}} \le 383{,}157{,}811$ for $M = 8$.

These values are computed exactly from the certificate data, in the weights at $p = p_*$.

3.6. **Proofs.** Below, p_*, X and λ are those of Table 3. For $p < p_*$ put

$$X_p = X\sqrt{\frac{p_*(1-p)}{p(1-p_*)}} \ge X.$$

Then ω_E and ω_C both scale by $\rho(p_*, p)$; the rows are linear in the weights and the tests do not depend on p, so every row holds with $\lambda\rho(p_*, p)$ in place of λ, and the start sums do not increase. It therefore suffices to prove each bound at $p = p_*$, with λ replaced by $\lambda\rho(p_*, p)$ for smaller p. In each case $P_{\mathrm{ML}} \le P_{\mathrm{MW}}$.

Proof of Theorem 2.1. By Lemma 3.1 and Section 3.3, $P_{\mathrm{MW}}(\mathcal{P}_{d,w}, p)$ is at most the total weight of the labelled paths from b_1 to b_2 none of whose windows fails a test. Bound the two half-edges of each path, summed over their labels, by $\omega_E + \omega_C$ each. The interior starts at one of the w checks $(0, y)$ and has at least $d - 2$ edges, so with the certificate and Section 3.4

$$P_{\mathrm{MW}}(\mathcal{P}_{d,w}, p) \le w\, K_0 \sum_{\ell \ge d-2} \lambda^{\ell - K} \le w \cdot 398.91 \cdot \frac{\lambda^{d-11}}{1-\lambda}$$

with $K = 9$ and $\lambda = 0.998\,\rho(0.0465, p)$. The condition $d \ge 11$ makes the interior at least K steps long, so that it begins with a state. □

Proof of Theorem 2.2. As for the planar code, now with the tests and the near states of $\mathcal{R}_d$, valid for every d by Section 3.3. The path γ starts with one of the d half-edges at b_1; its interior starts at the check z of this half-edge, and the first K steps form a state whose first point has height $h(z)$. The interior has at least $d-2$ edges and ends at a check with at most two half-edges to b_2. Hence

$$P_{\mathrm{MW}}(\mathcal{R}_d, p) \le d \cdot 2 \cdot K_{\mathrm{rot}} \sum_{\ell \ge d-2} \lambda^{\ell - K} \le 2d \cdot 366{,}183{,}771.98 \cdot \frac{\lambda^{d-11}}{1-\lambda}$$

with $K = 9$ and $\lambda = 0.998\,\rho(0.0465, p)$; as for the planar code, $d \ge 11$ makes the interior at least K steps long. □

Proof of Theorem 2.3. Every point of $\mathcal{T}_m$ can be the first point of γ. Fix one, and follow γ from it. Its first $\ell - 1$ steps, where $\ell \ge m$ is the length of γ, form a labelled simple path whose windows pass the tests, and the last step closes the cycle; we bound the closing step, summed over its labels, by $\omega_E + \omega_C$, and we do not test the windows that contain it. The first K steps form a state whose first point has offset o. By Lemma 3.4 and Section 3.4, for $m \ge L_0$,

$$P_{\mathrm{MW}}(\mathcal{T}_m, p) \le m^2 K_{\mathrm{tor}} \sum_{\ell \ge m} \lambda^{\ell - 1 - K} = m^2 K_{\mathrm{tor}} \frac{\lambda^{m-1-K}}{1-\lambda}.$$

For $M = 10$ we have $K = 9$, $h_0 = 20$, $L_0 = 1201$, $\lambda = 0.997\,\rho(0.046, p)$ and the bound $m^2 \cdot 428{,}572{,}391.6 \cdot \lambda^{m-10}/(1-\lambda)$. For $M = 8$ we have $K = 7$, $h_0 = 16$, $L_0 = 769$, $\lambda' = 0.996\,\rho(0.045, p)$ and the bound $m^2 \cdot 383{,}157{,}811 \cdot \lambda'^{m-8}/(1-\lambda')$. □

M (longest window)	4	6	8	9	10	11
states	240	6,258	157,618	779,946	3,883,082	19,106,700
p at which the Perron root is 1	0.0420	0.0445	0.0459	0.0462	0.0468	0.0470

TABLE 4. Growth of labelled paths in $\mathbb{Z}^2$ under the tests with $\phi = x + 9y$ on windows of at most M edges, minimized over X (floating-point Perron roots). Only the certificates of Section 3.5 enter the proofs.

Remark 3.5. Without exclusion rules the count above is the argument of Dennis, Kitaev, Landahl and Preskill [7], with the labels in place of the half-weight condition, and it gives about 0.037. The rules remove part of the multiplicity of the union bound: they keep only the crossing and the labels of one canonical minimum-weight correction. Table 4 shows how slowly the threshold of the method grows with the length of the windows. On the torus the tie-break has no vertical term, since a function on $\mathbb{Z}_m^2$ that is linear in both directions would have two seams; with $\phi = x$ the growth rate reaches 1 at about 0.0465 for $M = 10$, slightly below the planar value. The rotated code uses the potential of the planar certificate far from the top side, so the two theorems hold at the same p_*.

4. Measurement errors

In this section we prove Theorem 2.4. The decoding graph $G = \mathcal{Q}_{d,w,T}$ of Section 2.1 has the points of the box $B = [0, T] \times [0, w-1] \times [0, d-2]$ as checks, an edge between any two checks at distance 1, and half-edges from the checks $(t, x, 0)$ to b_1 and from the checks $(t, x, d-2)$ to b_2. We write $|u - v|$ for the ℓ^1 distance of two checks, and a path between checks u, v with $|u - v|$ edges is a *geodesic*. Every geodesic between two checks of B stays in B, since B is a box.

4.1. **Failure paths and exclusion rules.** Let $\phi(t, x, y) = 81t + x + 9y$, and for a set S of edges let $\Phi(S)$ be the sum of $\phi(u) + \phi(v)$ over the edges $\{u, v\} \in S$ between checks. If some correction of least size fails, let C^* be one with the least value of Φ among the failing corrections of least size. The set $E \oplus C^*$ has even degree at every check and odd degree at b_1 and b_2, so it contains a simple path γ from b_1 to b_2. The interior of γ (without its two half-edges) meets all $d - 1$ rows, so it has at least $d - 2$ edges. Since $C^* \oplus \gamma$ is a correction, $|\gamma \cap E| \geq |\gamma \cap C^*|$. We label the edges of γ in E by E and those in C^* by C; every edge of γ gets one label. A *window* is a sub-path σ of γ between checks $u \neq v$ that contains no half-edge; n_E and n_C count its labels, and σ_E, σ_C are its labelled edge sets.

Lemma 4.1. *Every window σ from u to v satisfies:*

(i) $n_C - n_E \leq |u - v|$;
(ii) *if* $n_C - n_E = |u - v|$, *then* $\Phi(\sigma_C) \leq \Phi(\sigma_E) + \Phi(P')$ *for every geodesic P' from u to v;*
(iii) $n_C \leq c(T_\sigma)$, *the least total length of a perfect matching of the set T_σ of odd-degree checks of σ_C;*
(iv) *if* $n_C = c(T_\sigma)$, *then* $\Phi(\sigma_C) \leq \sum_k \Phi(P_k)$ *for any geodesics P_k joining the pairs of a least-length perfect matching of T_σ.*

Proof. (i) For a geodesic P' from u to v, $C' = C^* \oplus \sigma \oplus P'$ is a correction and $|C'| \leq |C^*| - n_C + n_E + |u - v|$. (ii) In the equality case $|C'| = |C^*| - 2k$, where k is the number of edges of P' in $(C^* \setminus \sigma_C) \cup \sigma_E$; hence $k = 0$, and C' has least size. It fails, because $C' \oplus C^* = \sigma \oplus P'$ has no half-edge, and $\Phi(C') = \Phi(C^*) - \Phi(\sigma_C) + \Phi(\sigma_E) + \Phi(P')$; the choice of C^* gives the inequality. (iii) and (iv) follow in the same way from $C' = C^* \oplus \sigma_C \oplus P''$, where P'' is the set of edges used an odd number of times by the P_k; if the P_k overlap, then $|P''| < c(T_\sigma)$ and $|C'| < |C^*|$. □

The least value of Φ over the geodesics from u to v is attained by the geodesic that takes its unit steps in increasing order of their ϕ-increments. By (ii), every run of C-edges of γ is this staircase; in particular two C-edges that turn a corner turn from the smaller to the larger ϕ-increment.

4.2. **Counting.** The counting of Section 3.4 carries over, with $\mathbb{Z}^3$ in place of $\mathbb{Z}^2$ and Lemma 4.1 in place of the tests. A fixed labelled path occurs as γ with probability at most $\omega_E^{n_E}\omega_C^{n_C}$, with $\omega_E = pX$, $\omega_C = (1-p)/X$ and any $X \geq 1$, since $n_E \geq n_C$ on γ. States and successors are defined as there, with windows that satisfy the lemma, and if a positive potential H on the states satisfies

$$\sum_{s' \text{ successor of } s} \omega(s \to s')\, H(s') \leq \lambda\, H(s) \qquad \text{for every state } s, \tag{4.1}$$

then the labelled continuations of m steps of a state s weigh at most $\lambda^m H(s)/\min H$; they include every labelled simple path whose windows satisfy the lemma.

Lemma 4.2. *For a state whose points lie in B, every successor inside B is a successor in $\mathbb{Z}^3$ with the same steps and labels. Hence the rows* (4.1) *of the translation-invariant system on $\mathbb{Z}^3$ imply the rows of all states in B.*

Proof. Every alternative used in Lemma 4.1, a geodesic between two points of the window or a matching geodesic, lies in the bounding box of the window's points, hence in B. So every window excluded in $\mathbb{Z}^3$ is excluded in B, and in B a state has no successor that it lacks in $\mathbb{Z}^3$. Since $H > 0$, the rows in B follow. □

So one certificate for the bulk system covers every box; no boundary types occur.

4.3. **The certificate.** The certificate is a positive integer potential for the bulk system with $M = 8$: 5,263,174 states and 44,599,560 transitions, $p_* = 2/125$, $X = 231/50$, $\lambda = 499/500$, potentials at most 2^{32}. As in Section 3.5, every row (4.1) is checked in exact integer arithmetic. The largest ratio of the two sides of (4.1) is 0.997805 (the Perron root of the system is 0.995809), and the start sum $K_0 = (\omega_E + \omega_C)^2 \sum_s \omega(s)H(s)/\min H$ is at most 1.2827.

Proof of Theorem 2.4. By Lemma 4.1 and the counting above, $P_{\mathrm{MW}}(\mathcal{Q}_{d,w,T}, p)$ is at most the total weight of the labelled paths from b_1 to b_2 whose windows satisfy the lemma. As in the proof of Theorem 2.1, bound the two half-edges of each path, summed over their labels, by $\omega_E + \omega_C$ each. The interior of the path starts at one of the $w(T+1)$ checks of the rows $y = 0$ and has at least $d-2$ edges. With the certificate, Lemma 4.2 gives at $p = p_* = 2/125$

$$P_{\mathrm{MW}}(\mathcal{Q}_{d,w,T}, p) \leq w(T+1)K_0 \sum_{\ell \geq d-2} \lambda^{\ell-7} \leq (T+1)\, w \cdot 1.2827 \cdot \frac{\lambda^{d-9}}{1-\lambda}.$$

The condition $d \geq 9$ makes the interior at least 7 steps long, so that it begins with a state. For $p < p_*$ take $X_p = X\sqrt{p_*(1-p)/(p(1-p_*))} \geq X$. As in Section 3.6, ω_E and ω_C then both scale by $\rho(p_*, p)$; since the rows are linear in the weights and the rules do not depend on p, every row holds with $\lambda\rho(p_*, p)$, and K_0 does not increase. Finally $P_{\mathrm{ML}} \leq P_{\mathrm{MW}}$. □

Remark 4.3. Dennis, Kitaev, Landahl and Preskill [7] report 0.0114 for this noise model; their argument counts self-avoiding walks on the cubic lattice and uses the numerical value 4.684 of their connective constant. Its condition $4\mu^2 p(1-p) < 1$, which gives the values 0.0373 and 0.0366 of the introduction, gives 0.0115 at $\mu = 4.684$ and 0.0112 at the proven upper bound $\mu \leq 4.739$ [30]. Numerically the threshold of minimum-weight matching is 0.0293(2) [34]. Lemma 4.2 uses that B is a box. We do not treat the rotated or the toric code with measurement errors; there the rules would need the modifications of Section 3.

M (longest window)	4	5	6	7	8
states	968	8,352	72,294	616,322	5,263,174
p at which the Perron root is 1	0.0150	0.0152	0.0157	0.0159	0.0161

TABLE 5. Growth of labelled paths in the bulk under the rules of Lemma 4.1 on windows of at most M edges (floating-point Perron roots). The proof uses only the certificate of Section 4.3.

5. The certified threshold

5.1. **Blocks, windows and the auxiliary decoder.** Fix a block side $L \geq 2$ and a window radius $0 \leq R < \frac{3}{2}L$, and put $g = \lfloor R/L \rfloor + 1$, so that $g = 1$ if $R < L$ and $g = 2$ if $L \leq R < \frac{3}{2}L$. We work on codes made of whole blocks, $\mathcal{P}_{n_yL+1,\,n_xL}$ with $n_y \geq 2g+2$, and with $n_x \geq 5$ if $g = 1$ and $n_x \geq 6$ if $g = 2$: the checks are (x, y) with $0 \leq x < W = n_xL$ and $0 \leq y < H = n_yL$, the bottom and top sides are rough and the left and right sides smooth. Block $c = (a, b)$, with $0 \leq a < n_x$ and $0 \leq b < n_y$, is the set of checks $[aL, aL + L) \times [bL, bL + L)$. An edge between two checks is *owned* by the block of its left or lower endpoint, and a half-edge by the block of its check. Distances between blocks are ℓ_∞ distances of their indices, and $N_1(c)$ is the set of blocks at distance at most 1 from c. The *window* W_c is the set of checks at ℓ_∞ distance at most R from c (Figure 2). The *window graph* G_c has every edge with at least one endpoint in W_c; an edge with exactly one endpoint in W_c (a *cut edge*) and every half-edge end at a boundary vertex, so nothing outside the window constrains it.

Let u be a fixed table of independent uniform numbers in $[-1, 1]$, indexed by a position modulo L and a direction, generated from a seed fixed in the registration. The edge e with left or lower endpoint (x, y) and direction δ gets $u_e = u_\delta(x \bmod L,\, y \bmod L)$ and the weight $w_e = 1 + 0.02\, u_e \in [0.98, 1.02]$; half-edges are indexed in the same way. Since u_e depends on the position only modulo L, translates of a window by whole blocks carry identical weights.

Stage 1. For each block c, the first stage computes from the syndrome in W_c alone a correction X_c of $\mathcal{D} \cap W_c$ on G_c (Section 5.2) and keeps the edges of X_c owned by c. The union of the kept edges is C_0. Its *residual* is $\mathcal{D}' = \mathcal{D} \oplus \partial C_0$, and block c is *flagged* if $\mathcal{D}'$ meets it. Stage 1 is the registered program, and it is the only part of the decoder that the experiment measures.

Levels 0 and 1. Fix integers $a_0, a_1 \geq 1$. Let K range over the components of the flagged blocks under adjacency at distance at most a_0, and let $R_0(K)$ be the set of blocks within distance $\lceil a_0/2 \rceil$ of K. If $|\mathcal{D}' \cap K|$ is even, or $R_0(K)$ contains a check of row 0 or row $H - 1$, or $R_0(K)$ is the whole code, then K is *resolved* by a minimum-weight correction C_K of $\mathcal{D}' \cap K$ that uses only edges inside $R_0(K)$ and half-edges of its checks. Otherwise K is *odd*, and its blocks pass to level 1, which applies the same rule to them with adjacency a_1 and regions within distance $\lceil a_1/2 \rceil$.

Top. Let Ω be the residual $\mathcal{D}'$ restricted to the blocks of the unresolved level-1 components. C_{top} is a minimum-weight edge set of the whole code whose odd checks are exactly Ω (the boundary vertices absorb any parity). The output is $C = C_0 \oplus \bigoplus_K C_K \oplus C_{\mathrm{top}}$.

Proposition 5.1. *For every syndrome, C is a correction, that is, $\partial C = \mathcal{D}$.*

Proof. Blocks at distance at most a have $\lceil a/2 \rceil$-neighbourhoods that share a block, so the region of a component is a connected set of checks. A resolved component has even residual parity or reaches a rough row, so C_K exists. Each flagged block lies in exactly one component at each level until it is resolved or reaches the top, so the odd checks of $\bigoplus_K C_K \oplus C_{\mathrm{top}}$ are exactly $\mathcal{D}'$, and $\partial C = \partial C_0 \oplus \mathcal{D}' = \mathcal{D}$. □

The registered program uses $a_0 = 5$ and $a_1 = 16$. The analysis below holds for every admissible a_1 at once, because the bad events are events of the first stage, which does not depend on a_1; the statement of Theorem 2.5 also uses $a_1 = 16$.

5.2. **The first stage: window marginals.** Stage 1 at block c weights each edge of G_c by the log-odds of its posterior probability given the syndrome in the window, and then matches. It has five steps. Write $\mathcal{D}_c = \mathcal{D} \cap W_c$.

(i) *Reference chain.* A fixed function of $\mathcal{D}_c$ gives an edge set c_0 of G_c with $\partial c_0 = \mathcal{D}_c$ on W_c: each defect is joined by a straight path to the nearest side of the window rectangle along which every check has a cut edge (ties are broken in a fixed order of the sides); if no side has this property, L-shaped paths are used, and in a window without any cut edge the defects are paired in lexicographic order.

(ii) *Window posterior.* Identify all boundary vertices of G_c with one vertex o, and merge the m edges from each check of the window's perimeter to o (cut edges and half-edges; $m \in \{1, 2\}$ in the windows that the analysis uses) into one edge that flips with probability $p_m = \frac{1}{2}(1 - (1 - 2p)^m)$. Under independent flips, the posterior law of the error X on this graph given $\mathcal{D}_c$ is

$$\Pi(X) \propto \prod_e p_e^{X_e}(1 - p_e)^{1 - X_e} \qquad \text{over } X \text{ with } \partial X = \mathcal{D}_c \text{ on } W_c,$$

with $p_e = p$ for an ordinary edge and $p_e = p_m$ for a merged edge. Let $P_e = \Pi(e \in X)$; for an edge of a merged pair, $P_e = \frac{1}{2}P_b + (1 - P_b)\, y^2/(1 + y^2)$ with $y = p/(1 - p)$ and P_b the marginal of the merged edge.

(iii) *Weights.* Each edge of G_c gets

$$w'_e = \min\Big(30, \max\Big(-30,\ \ln\frac{1 - P_e}{P_e}\Big)\Big) + 10^{-3}\, u_e. \tag{5.1}$$

(iv) *Matching.* X_c is a minimum-weight correction of $\mathcal{D}_c$ on G_c for the weights w'. Negative weights are allowed and handled by the standard reduction. It is computed by the sparse-blossom implementation of minimum-weight perfect matching [14, 15] (based on [9]), at a version fixed in the registration, on a canonical ordering of the checks and edges of the window; the term $10^{-3}u_e$ separates ties, and the remaining tie behaviour is part of the fixed program.

(v) *Commit.* The edges of X_c owned by c are kept.

The construction is designed for (V6) and (Hcov). The reference chain, the weights and the matching are computed from $\mathcal{D}_c$ alone, so the output on c should depend on the error only through the syndrome in W_c; the canonical choice of c_0 is what makes this hold bit for bit, since the posterior itself does not depend on c_0 but its floating-point evaluation does. The computation depends on the window only through its shape and the periodic weights u_e, which is what (Hcov) needs.

The marginals as an Ising model on the dual. Write $X = c_0 \oplus Z$. The set Z ranges over the edge sets with even degree at every check of W_c, that is, over the cycles of the planar graph G_c° obtained above, and $\Pi(Z) \propto \prod_{e \in Z} y_e$ with $y_e = p_e/(1 - p_e)$ for $e \notin c_0$ and $y_e = (1 - p_e)/p_e$ for $e \in c_0$. In the plane, the cycles of G_c° are exactly the domain walls of the configurations $\pm\sigma$ of spins on its faces. Hence Π is the law of the domain walls of the Ising model on the dual graph of G_c° with couplings $K_e = -\frac{1}{2}\ln y_e$, that is, with

$$t_e = \tanh K_e = \pm(1 - 2p_e),$$

with the minus sign on the edges of c_0: a random-bond Ising model on the Nishimori line [25], with bond signs given by c_0. By the Kac–Ward formula [17] (equivalently, the Pfaffian method

of [18, 10]), its partition function is a constant times $S = \det(I-\Lambda)^{1/2}$, where Λ acts on the directed edges of the dual graph, $\Lambda_{\vec{e},\vec{e}'} = t_{e'}\, e^{i\theta(\vec{e},\vec{e}')/2}$ when $\vec{e}'$ continues $\vec{e}$ without backtracking, θ being the turning angle, and $\Lambda_{\vec{e},\vec{e}'} = 0$ otherwise. Differentiating,

$$\langle \sigma\sigma' \rangle_e = t_e + (1-t_e^2)\,\frac{\partial \ln S}{\partial t_e}, \qquad \frac{\partial \ln S}{\partial t_e} = -\tfrac{1}{2}\,\mathrm{tr}\Big((I-\Lambda)^{-1}\frac{\partial \Lambda}{\partial t_e}\Big), \tag{5.2}$$

where σ, σ' are the spins on the two faces at e. Then $\Pi(e \in Z) = \frac{1}{2}(1 - \langle \sigma\sigma' \rangle_e)$, and $P_e = \Pi(e \in Z)$ for $e \notin c_0$, $P_e = 1 - \Pi(e \in Z)$ for $e \in c_0$. The marginals therefore need exactly the entries of $(I-\Lambda)^{-1}$ at the nonzero positions of Λ.

Computation. A diagonal unitary similarity, which multiplies the row and the column of each directed edge by a phase determined by its direction, makes Λ real with entries $\pm t_{e'}$; it leaves S and the right side of (5.2) unchanged, so all arithmetic is real, in double precision. Multiplying the real matrix $I - \Lambda$ by a diagonal matrix of fixed signs and the permutation that reverses each directed edge on one side, and by the diagonal matrix of the couplings t_e on the other, gives a real skew-symmetric matrix of the form of Kasteleyn [18], with the same sparsity up to that permutation, from whose inverse the needed entries of $(I-\Lambda)^{-1}$ follow. This matrix is sparse, with the adjacency of the faces of the window. We order its unknowns by nested dissection [13], splitting the rectangle of faces recursively along a line of edges in the middle of its longer side down to pieces of at most 16 faces, factor it by a multifrontal elimination with dense frontal matrices, of which skew symmetry lets us compute only one triangle, and obtain the needed entries of the inverse by the backward recursion of selected inversion [23, 24]. For a window with N edges this takes $O(N^{3/2})$ operations and $O(N \log N)$ memory.

In double precision the computed marginals carry rounding errors, which can be large in some windows. As a diagnostic, which the analysis does not use, the program records for each window the largest residual $|((I-\Lambda)X)_{\vec{e}\vec{e}} - 1|$ over the directed edges $\vec{e}$, where X holds the computed entries of the inverse. Over the 1.14×10^7 windows of the experiment of Section 5.4, its median was 2.7×10^{-8} and its 99th percentile 2.3×10^{-5}, but it exceeded 10^{-4} in 0.4% of the windows and reached 2.9×10^2; in such windows the computed marginals can be far from the true ones. This does not affect the validity of the decoder, since Proposition 5.1 holds for every first stage, and the bound of Theorem 5.3 does not use that the marginals are accurate or that the matching is optimal. Their accuracy affects only how often blocks are bad, and the experiment measures this for the program as it runs.

5.3. **Block events and the deterministic analysis.** Let $F = E \oplus C_0$; its odd checks are $\mathcal{D}'$.

Definition 5.2. Block c is *bad* if it is flagged, or if some path of F inside $N_1(c)$ joins a check of c to a check at ℓ_∞ distance at least L from c (the block is *long*).

By (V6), whether c is bad depends only on the error on the edges with an endpoint in the union of the windows of $N_1(c)$, and these edges lie within ℓ_∞ distance $L+R$ of c. For two blocks at distance n these edge sets are disjoint when $(n-3)L > 2R$. With

$$D = 4 + \lfloor 2R/L \rfloor,$$

the bad events of any set of blocks with pairwise distances at least D are therefore independent; $D = 5$ for $R = L-1$, and $D = 6$ for $L \le R < \frac{3}{2}L$.

Types. Since $R < gL$, the window of a block in column a meets only the columns $a-g, \dots, a+g$, and it is clipped by the left smooth side exactly when $a < g$, since its left cut edges exist if and only if $aL - R - 1 \ge 0$; the same holds on the right. In block rows $g+1, \dots, n_y - g - 2$, a block in one of the columns $g+1, \dots, n_x - g - 2$ is of type *bulk*: neither its window nor those of its neighbours are clipped. The blocks of columns $0, \dots, g$ are of the types $L0, \dots, Lg$, and those of columns $n_x - 1, \dots, n_x - g - 1$ of the types $R0, \dots, Rg$; in each of them the window of the

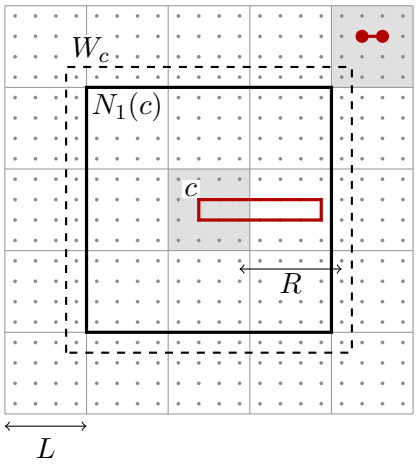


FIGURE 2. The auxiliary decoder of Section 5.1 on a part of a code far from its sides, with block side $L = 4$ and window radius $R = 5$, given by the formula $R = \frac{3}{2}L - 1$ of the registered program of Section 5.4; dots are checks and thin lines bound the blocks. The window W_c of block c (dashed) consists of the checks within ℓ_∞ distance R of c; it reaches beyond $N_1(c)$ (thick), the nine blocks at block distance at most 1 from c. Stage 1 computes a correction for each block from the syndrome in its window and keeps the edges that the block owns; the kept edges form C_0. A block is bad if the residual syndrome has a check in it (flagged; filled dots) or if a path of $F = E \oplus C_0$ (red edges) inside $N_1(c)$ joins it to a check at ℓ_∞ distance at least L (long; the loop through c reaches distance L from c and makes c long, but it stays within distance $3 < L$ of the block to the right of c, which is good). Bad blocks are shaded.

block or of a neighbour in its row is clipped by a smooth side. There are thus $2g + 3$ types: five for $R < L$ and seven for $L \le R < \frac{3}{2}L$. The bounds on n_x ensure that no block has neighbours whose windows reach both smooth sides. By (Hcov), all blocks of one type are bad with the same probability, that of the reference block of the type (Section 5.4). The rough rows $0, \dots, g$ and $n_y - g - 1, \dots, n_y - 1$ are never sampled: the analysis declares their blocks bad.
Hypothesis (H_ε). For every set S of blocks in rows $g+1, \dots, n_y - g - 2$ with pairwise distances at least D, $\Pr(\text{all blocks of } S \text{ are bad}) \le \varepsilon^{|S|}$. Under (V6) and (Hcov) this holds with $\varepsilon = \max_t \varepsilon_t$.

Theorem 5.3 (failure bound of the auxiliary decoder)**.** *Let $L \ge 2$, $0 \le R < \frac{3}{2}L$, $D = 4 + \lfloor 2R/L \rfloor$, $a_0 \ge D - 1$ and $a_1 \ge \max(a_0, 2D - 2)$ with a_1 even, and put $H_0 = 2a_1 + D + 2$. Let stage 1 be any program whose output on each block depends only on the syndrome in the block's window, and let the later stages be as in Section 5.1, with weights in $[0.98, 1.02]$. Assume (H_ε) with ε below the convergence limits of the two series of Theorem A.8. Then there are explicit constants $C_1, s_1, C_2, s_2 > 0$, which depend only on ε, D, a_0, a_1, the weight ratio $\kappa = 1.05$ and L, and do not increase with L, such that for all n_x as in Section 5.1 and all $n_y > H_0$*

$$\Pr\big(\textit{the decoder fails on } \mathcal{P}_{n_y L+1,\, n_x L}\big) \le C_1\, n_x\, e^{-s_1(n_y - H_0)} + C_2\, n_x\, e^{-s_2(n_y - H_0)}.$$

In the instance $(\varepsilon, a_0, a_1) = (3 \times 10^{-4}, 5, 16)$ with $L \le R < \frac{3}{2}L$, so that $g = 2$, $D = 6$ and $H_0 = 40$, the bound holds for every $L \ge 2$, $n_x \ge 6$ and $n_y \ge 41$ with

$$C_1 = 35, \quad s_1 = 0.10, \quad C_2 = 2.5 \times 10^4, \quad s_2 = 0.02,$$

the constants of (2.1). At the same (ε, a_0, a_1), the numerical evaluation of the series in Appendix A.4 gives the sharper bound $5.52 \times 10^2\, n_x\, e^{-0.171(n_y-40)} + 2 \times 10^5\, n_x\, e^{-0.032(n_y-40)}$, and at the certified level $\hat{\varepsilon}$ it gives the constants of Remark 2.8.

The proof is given in full in Appendix A; its idea is as follows. Call the components of the bad blocks under adjacency at distance at most a_0 the *clusters*. A path of F that starts in a good block c ends before it reaches distance L from c, since otherwise c would be long. Hence every component of F meets at most one cluster, and the residual defects, which lie in flagged blocks, lie in clusters. Group the clusters into *families* linked by level-1 components. If the decoder fails, then either the part of $E \oplus C$ attached to a single family has odd parity at both boundary vertices, or the top stage joins a chain of families from one rough side to the other. Parity inside interior clusters forces witnesses, flagged blocks far apart within a cluster or a family, and keeps consecutive families of a chain more than a_1 apart; an exchange argument bounds the total weight of the top-stage paths of a chain by the distances between the entry and exit defects within its families, since otherwise re-pairing would lower the weight of the top stage. Among the bad blocks of a cluster, a greedy selection keeps blocks at mutual distance at least D, which by (H_ε) are all bad with probability at most ε to the number of blocks. Counting the two alternatives with exponential tilts then gives two geometric series, and hence the two terms of the bound.

5.4. **The certificate at $p = 0.102$.** In the code of 7×7 blocks, $\mathcal{P}_{7L+1,\,7L}$, the blocks $(a, 3)$ with $a = 0, \ldots, 6$ are the *reference blocks* of the types $L0$, $L1$, $L2$, bulk, $R2$, $R1$, $R0$, and ε_t is the probability that the reference block of type t is bad at $p = 0.102$. A sample of type t draws independent flips on a *patch*, which consists of the checks within ℓ_∞ distance $R+1$ of $N_1(c)$ and all edges between them; for the types other than bulk the patch is clipped at the smooth side. The sample then runs stage 1 on the windows of the blocks of $N_1(c)$, nine of them, or six for the types $L0$ and $R0$, and evaluates Definition 5.2. Every edge that the bad event depends on lies in the patch, so a sample is bad with probability ε_t; this identification is part of (Bin).

The registration was frozen before the first sample was drawn. It fixes the program: $p = 0.102$, $L = 280$, $R = 419$, the posterior weights (5.1) on every edge of every window with their tie term and clipping, the marginal computation of Section 5.2, and the matching implementation at a fixed version. It fixes the later stages, with $a_0 = 5$ and $a_1 = 16$, and the rule that the sharper constants of Remark 2.8 use the admissible a_1 that minimizes the bound at the certified level. It fixes the experiment: $n = 2 \times 10^5$ samples of each of the seven types, with seeds derived from entropy drawn at the freeze; every sample is counted, and the only stopping rule ends the run without acceptance as soon as some type has more than 17 bad samples. The experiment accepts if $k_t \le 17$ for every type t, where k_t is the number of bad samples of type t. Equivalently, for every type the exact one-sided Clopper–Pearson [4] upper bound for ε_t at level $10^{-9}/7$ is at most 3×10^{-4}: this bound is below 2.9367×10^{-4} at $k = 17$ and above 3.0222×10^{-4} at $k = 18$. The command of the analysis was written down before the run started. The block side, the window radius and the level 3×10^{-4} were chosen before the registration, on the basis of pilot runs with separate seeds.

The registration names two implementations of stage 1, which differ in how the window marginals are computed, and a rule that chose between them before the first sample: the faster one was to be used if it passed five checks, among them bit-identical outputs on two machines at the registered geometry and, on 100 samples of each type there, block events equal to those of the other implementation on at least 99% of the samples. It passed them, with the same first-stage corrections and block events as the other implementation on all 700 samples, and every sample of the experiment used it; Section 5.2 describes it.

Six validations test the registered program. Each uses its own random inputs, and none reads a certificate sample. V6 compares the stage-1 outputs, matching weights and flags of all nine windows for errors with the same syndrome: E; $E \oplus S$, for plaquette boundaries S inside the windows; $E \oplus S \oplus \Gamma$, for a relative cycle Γ across them; and the canonical chain of the syndrome. It tests (V6) directly. V2 compares stage-1 outputs, in block coordinates, for errors that are periodic with period L. V4 compares the block events of the sampler with the full decoder on planar codes of 8×8 blocks, for all seven types. V2 and V4 test (Hcov) and the identification of samples with reference blocks. V1 (two runs of the program on the same samples), V3 (an independent implementation of the marginals and the matching) and V5 (exact transfer-matrix marginals) test the implementation. All six passed before the registration was frozen, on blocks of side 8, 16 and 24 with window radius $R = \frac{3}{2}L - 1$, the formula of the registered program, and V1 and V6 passed again at the registered geometry $(280, 419)$ after the freeze and before the first sample. Table 2 gives the outcomes, and Table R.1 of the records supplement lists every run.

During the run, an automatic check, made at every collection of results, was the only reader of the counts: it would have ended the run without acceptance if some type had more than 17 bad samples, and it never did. This was the only experiment registered for this statement (Section 2.4). Five earlier versions of the registration were replaced before any of them was time-stamped or run, and no sample was drawn under them; their seed ranges are disjoint from that of the experiment.

Table 6 shows the result of the analysis. Every k_t is at most 17, so the experiment accepted, and by Sections 2.4 and 2.5 this gives Theorem 2.5. The largest upper bounds, those of $L2$ and $R1$, are 2.58406×10^{-4} rounded up; they give the level $\hat{\varepsilon}$ of Remark 2.8. Every bad sample was decoded again from its seed on a second machine, with records identical apart from the fields of time and memory, and for 5,600 samples whose full first-stage data the registration required to be stored, the flags and long paths were recomputed from these data with identical results (Section R.1).

type	n	bad samples k_t	upper bound for ε_t
$L0$	200000	8	2.111×10^{-4}
$L1$	200000	6	1.905×10^{-4}
$L2$	200000	13	2.585×10^{-4}
bulk	200000	11	2.400×10^{-4}
$R2$	200000	12	2.493×10^{-4}
$R1$	200000	13	2.585×10^{-4}
$R0$	200000	3	1.566×10^{-4}

TABLE 6. The certificate at $p = 0.102$, $L = 280$, $R = 419$. The last column is the exact one-sided Clopper–Pearson upper bound at level $10^{-9}/7$, rounded up.

5.5. **The certificate under measurement errors.** Theorem 2.6 rests on a certificate of the same kind on the phenomenological graph. The codes are $\mathcal{Q}_{n_y L+1,\, n_x L,\, n_t L-1}$ with $n_x, n_t, n_y \geq 5$, whose checks form the box $[0, n_t L) \times [0, n_x L) \times [0, n_y L)$ in the coordinates (t, x, y). The blocks are cubes of side $L = 35$, and the window of a block is the set of checks at ℓ_∞ distance at most $R = 34$ from it. Distances between blocks are ℓ_∞ distances of their indices, so $N_1(c)$ has at most 27 blocks. Only the sides $y = 0$ and $y = n_y L - 1$ carry half-edges; the two sides in x and the first and last layers in t behave like smooth sides. Every weight is 1. For each block, stage 1 computes a minimum-weight correction of the syndrome in the block's window on the graph of

the window. In this graph, every check with a neighbour outside the window has one edge to a boundary vertex. Stage 1 uses a matching program at a fixed version with a fixed order of nodes and edges, and it keeps the edges owned by the block. Levels 0 and 1, with $a_0 = 4$ and $a_1 = 20$, and the top stage are as in Section 5.1. A block is bad if it is flagged or long, as in Definition 5.2, and the bad events of blocks at pairwise distances at least $D = 5$ are independent. Outside the declared rows 0, 1, $n_y - 2$, $n_y - 1$, a block has one of 25 types. Its block column gives one of $L0$, $L1$, bulk, $R1$, $R0$, as in Section 5.3, and its block layer in time gives one of $B0$, $B1$, bulk, $E1$, $E0$, defined in the same way.

The deterministic lemma and the counting of Theorem 5.3 have three-dimensional analogues, with $D = 5$, $a_0 = 4$, $a_1 = 20$, $H_0 = 47$ and $\kappa = 1$, since all weights are equal. Lean proves them in full, for the decoder whose later stages use any minimum-weight corrections for these unit weights. If every type is bad with probability at most $\varepsilon^* = 1.85 \times 10^{-5}$, the decoder fails on $\mathcal{Q}_{n_yL+1,\, n_xL,\, n_tL-1}$ with probability at most the right side of (2.2); the rates 0.1008 and 0.007 there are the tilts of the two counts. Lemma 2.9 and Proposition 2.10 hold for $\mathcal{Q}_{d,w,T}$ with the same proofs, where a new layer, together with the time-like edges into it, is added like a new column. As in the proof of Theorem 2.5, these lemmas and the optimality of ML decoding turn the bound for codes made of whole blocks into (2.2) for all $d \geq 5L + 1$, $w \geq 1$ and $T \geq 0$. When w and T grow subexponentially in d, so does $n_t n_x$, while the exponentials of (2.2) decay at least like $e^{-(d-1)/5000}$; this gives the last statement of Theorem 2.6.

As in Section 5.4, no sample was drawn before the registration was frozen; this registration fixes the program described above, the level ε^* and a joint design. Each of $N = 3 \times 10^6$ simulations draws independent flips at $p = 0.02$ on a patch of $5 \times 5 \times 3$ blocks (in x, t and y) with their windows and a margin. It runs stage 1 on the 75 windows and evaluates the bad events of the 25 blocks of the middle block row, one of each type. The seeds are derived from fresh entropy; every simulation is counted, and there is no stopping rule. The experiment accepts if $k_t \leq 14$ for every type t. Equivalently, for every type the exact one-sided Clopper–Pearson upper bound for ε_t at level $4 \times 10^{-11} = 10^{-9}/25$ is at most ε^*.

Before the registration, V1 ran at the registered geometry. Also before the registration, V2, V3, V4 and V6 of Section 5.4, adapted to three dimensions, ran at $(L, R) = (9, 4)$, not at the registered radius $R = L - 1$. V2, V4 and V6 ran at $R = L - 1$ only after the analysis, at $(L, R) = (9, 8)$ and at the registered $(35, 34)$. The outcomes are in Table 2, with the individual runs in Section R.2 of the records supplement.

Table 7 shows the counts. There were 32 bad samples, all flagged and none long, and at most 4 of any type, so the experiment accepted. The largest exact Clopper–Pearson upper bound at level 4×10^{-11}, that of the count 4, is 1.1706×10^{-5}, rounded up.

	$L0$	$L1$	bulk	$R1$	$R0$
$B0$	0	0	3	3	0
$B1$	2	1	1	1	2
bulk	2	1	1	2	0
$E1$	1	0	0	2	2
$E0$	0	4	1	2	1

TABLE 7. The certificate under measurement errors at $p = 0.02$, $L = 35$, $R = 34$: the number of bad samples among the 3×10^6 samples of each type, by block layer in time (rows) and block column (columns). Every count is at most 14.

result	content	Lean theorems	checked by
Thm. 2.1	0.0465	`planar_A10_kernel`, `planar_A10_kernel_ML`	kernel
	same statement	`planar_A`, `planar_A_ML`	compiled evaluation
Thm. 2.2	0.0465	`rotated_B_kernel`, `rotated_B_kernel_ML`	kernel
	same statement	`rotated_B`, `rotated_B_ML`	compiled evaluation
Thm. 2.3	0.046	`toric_A'_10_kernel`, `toric_A'_10_kernel_ML`	kernel
	same statement	`toric_A'_10`, `toric_A'_10_ML`	compiled evaluation
	0.045	`toric_A'_kernel`, `toric_A'_kernel_ML`	kernel
	same statement	`toric_A'`, `toric_A'_ML`	compiled evaluation
Thm. 2.4	0.016	`phenom_D2M8_kernel`, `phenom_D2M8_kernel_ML`	kernel
	same statement	`phenom_D2`, `phenom_D2_ML`	compiled evaluation
Thm. 2.5	bound (2.1) and threshold	`theoremB_q102_main`, `pMLPlanar_tendsto_zero_subexp_q102`, `pMLPlanar_exp_decay_subexp_q102`	kernel
	confidence statements	`theoremB_certified_q102`, `threshold_q102_general_certified`	kernel
	counts of Table 6 and the rule	`obs_accept_q102`, `obs_cp_q102`, `cp_accept_q102`, `cp_reject_q102`, and `cp_upper_q102_k`, `cp_tight_q102_k` for $k = 3, 6, 8, 11, 12, 13$	kernel
Thm. 5.3	the instance with $D = 6$	`S3Fam.h2d_K2`	kernel
Prop. 5.1	validity of the decoder	`S3dec_valid`	kernel
Thm. 2.6	bound (2.2)	`d1B_num`, through `fail_bound_prog3`	kernel
	threshold	`pML3_tendsto_zero_subexp`, `pML3_exp_decay_subexp`	kernel
	confidence statements	`d1B_certified`, `threshold_d1_general_certified`	kernel
	counts of Table 7 and the rule	`d1_accept`, `d1_rule_tail`, `d1_cp_upper_k4`, `d1_cp_tight_k4`	kernel

TABLE 8. The main results and their formal proofs. Names ending in `_ML` are the statements for P_{ML}. "Kernel" means that every step, including every finite computation, is checked by Lean's kernel; "compiled evaluation" means that the certificate checks are run as compiled code inside Lean (Section 6.1).

6. Formal verification in Lean 4

The theorems of Section 2 are also proved in the proof assistant Lean 4, with the Mathlib library; Table 8 lists the formal theorems. A proof assistant checks every step of a proof against the rules of its logic, and the part of Lean that performs this check is its kernel.

6.1. **The sampling-free bounds.** The Lean statements of Theorems 2.1–2.4 concern the decoding graphs of Section 2.1, which are defined in Lean from their checks and edges. In Lean, $\mathcal{P}_{d,w}$ is the box $[0, d-1) \times [0, w)$ of checks with half-edges at the two ends of the first coordinate; the Lean statements write $n_x = d-1$ and $n_y = w$, so the exponent $n_x - 10$ in Lean is the exponent $d-11$ of Theorem 2.1. The phenomenological graph $\mathcal{Q}_{d,w,T}$ is the box $[0, T+1) \times [0, w) \times [0, d-1)$

with half-edges at the two ends of the last coordinate. The rotated code is defined by the positions of its d^2 qubits and of its X checks, and the logical class of an error is the parity of its flips on the first column; a Lean lemma identifies its failure probability with that of the graph used in the proof. The toric code $\mathcal{T}_m$ has the checks $\mathbb{Z}_m^2$, and a correction fails when $E \oplus C$ crosses one of the two seams of the torus an odd number of times. In Lean, P_{MW} is the probability that some correction of least size fails, and P_{ML} is the least failure probability of a map from syndromes to logical classes; a randomized decoder is a mixture of such maps, so it does no better. Each theorem has a version for P_{MW} and one for P_{ML}; the second follows from the first and $P_{\mathrm{ML}} \le P_{\mathrm{MW}}$. The constants and rates are those of Section 2.3, entered as exact rationals, for example $398.91 = 39891/100$ and $0.998 = 499/500$. The statements have no hypotheses beyond the ranges of p and of the size.

The general argument. Each proof consists of a general argument, proved once for each family of graphs, and a finite check of a certificate. The general argument has four parts: the probability estimates (independence of the edges, the union bound over failure paths, and the change of measure from p_* to p), a decoder lemma, a counting lemma, and the soundness of the checks. The decoder lemma states that if some correction of least size fails, then a failing correction of least size with the least value of the tie-breaking potential Φ exists; $E \oplus C$ then contains a path between the two boundaries (for the toric code, a cycle that is not null-homologous), and no window of this path can be replaced by a witness that passes the check, which covers the exclusion rules of Section 3 in one statement. The counting lemma turns the row inequalities of a potential into a bound on the weighted number of legal labelled walks of each length. The soundness theorems derive the hypotheses of the counting lemma from the Boolean checks of a certificate. For the rotated code the walk carries its height below the top side while it is near that side, and for the toric code its offset from the seam of Φ while it is near the seam; a translation lemma shows that the checks of states far from the seam do not depend on the offset, so the far states are checked at a single offset.

Certificates. A certificate lists the states of the walk, that is, its last K steps with their labels (for the rotated and the toric code also the height or the offset near the boundary or the seam), gives each state a positive integer potential, and fixes p_*, λ and the constant of the bound. The check functions verify, for every listed state, that the weighted sum of the potentials of its legal successors is at most λ times its own potential and that every successor that is not listed ends in an excluded window; they verify that every legal start of a walk is listed or excluded, and they bound the start sum by the stated constant. An excluded window counts only if a witness for it passes the witness check. Hence neither the search that finds the witnesses nor the certificate files need to be trusted: a wrong value can only make a check fail. Tables 3 and 5 give the sizes of the certificates.

Compiled evaluation and the kernel. A generator turns the certificate files into Lean data modules and writes the sha256 of each input file into the module. The checks are then evaluated in one of two ways. With compiled evaluation, the check functions are compiled to machine code and run through `native_decide`; each call adds one axiom, which `#print axioms` lists, and the compiler and the runtime join the trusted base. On the kernel route, the certificate is a binary search tree whose nodes hold a state, its potential and the successors that the certificate claims to be excluded; the check functions recurse on this tree, and every check is reduced by the kernel (`decide +kernel`), which adds no axiom. The checks are split into pieces of moderate size (ranges of rows or subtrees, and prefixes of the start walks), each piece is a separate theorem, and lemmas proved once combine the pieces. The planar, the rotated, the phenomenological and the two toric bounds are proved in both ways from the same certificates, with identical statements (Table 8). The Lean sources also contain kernel-checked bounds from two smaller certificates, for the planar code at $p \le 0.045$ with $d \ge 9$ and for the phenomenological graph at

$p \le 0.0155$ with $d \ge 7$, with 157,618 and 72,294 states. The data modules can be regenerated from the certificate files, whose sha256 they record; the reproduction manifest lists the input files of each theorem with their hashes. For the kernel certificates these include the transition or hint files, from which the generator reads the windows that the certificate claims to be excluded. Section R.3 of the records supplement gives the cost of the checks behind Table 8 and lists how many such axioms each compiled-evaluation theorem uses besides the standard ones (Section 6.4).

6.2. **The certified threshold.** The development defines the planar code $\mathcal{P}_{d,w}$ of Section 2.1, in which every edge, half-edges included, is flipped independently with probability p. It defines the class of an error, the parity of its half-edges at b_1, and the failure probability $P_{\mathrm{ML}}(\mathcal{P}_{d,w}, p)$ of a decoder that maps each syndrome to a class of largest probability; every tie rule gives the same value. The auxiliary decoder of Section 5 is defined on every code size, with block side L and window radius R; $n_x = \max(6, \lceil w/L \rceil)$ and $n_y = \lfloor (d-1)/L \rfloor$ are as in Theorem 2.5. The probabilities ε_t that the reference blocks are bad are defined from the first stage of the decoder on a code of 7×7 blocks, as in Section 5.4.

Let the first stage have block side $L \ge 2$ and window radius $L \le R < \frac{3}{2}L$, so that $D = 6$, and let it satisfy (V6) and (Hcov). Lean derives the bound (2.1) of Theorem 2.5 from the hypothesis that $\varepsilon_t \le 3 \times 10^{-4}$ for the seven types. From the bound it derives the threshold statement for every width $w \ge 1$ that grows subexponentially in d, in particular for every width bounded by a polynomial in d: for every $p \in [0, 0.102]$, the optimal failure probability tends to 0 and is at most $e^{-(d-1)/28000}$ for all large d. The proof bounds P_{ML} by the failure probability of the auxiliary decoder, using Lemma 2.9 and Proposition 2.10, which are also proved in Lean. The per-block hypothesis (H_ε) follows from three facts: from (V6), through the independence of blocks at block distance at least 6; from (Hcov), through the equality of the bad probabilities of blocks of the same type, for every $n_x \ge 6$; and from the bounds on the ε_t. Lean proves Theorem 5.3 in the instance described after that theorem, $(\varepsilon, a_0, a_1) = (3 \times 10^{-4}, 5, 16)$ with $D = 6$, for every $L \ge 2$ and every first stage, with weights in $[0.98, 1.02]$ on the edges and in $[0, 1.02]$ on the half-edges in the later stages, and with both terms of (2.1) and the constants shown; Theorem 2.5 uses it with unit weights (Section 2.5). The deterministic part of this proof is proved for every $g = \lfloor R/L \rfloor + 1$, with $a_0 \ge \max(2, D-1)$ and $a_1 \ge a_0$, and so is the reduction, by (Hcov), to the $2g+3$ reference blocks for every $n_x \ge 2g+2$.

The experiment enters through (Bin). Lean proves the confidence statement of Section 2.4, both for the bound and for the threshold statement, with one acceptance event for all widths; it does not assume a bound on the ε_t. Lean also checks three facts about the counts of Table 6: they satisfy the acceptance rule; each type satisfies $F(k; 2 \times 10^5, 3 \times 10^{-4}) \le 10^{-9}/7$ for the binomial lower tail F, so that each exact one-sided Clopper–Pearson upper bound at level $10^{-9}/7$ is at most 3×10^{-4}; and the rounded upper bounds of the table are valid and correctly rounded. Together these give Theorem 2.5. The hypotheses (V6), (Hcov) and (Bin) concern the registered program and the experiment, and Lean cannot check them; Table 2 gives the evidence for them. The kernel evaluates the arithmetic of the acceptance rule, $F(17; 2 \times 10^5, 3 \times 10^{-4}) \le 10^{-9}/7 < F(18; 2 \times 10^5, 3 \times 10^{-4})$, and the checks of the quoted upper bounds in exact integer arithmetic. No theorem of this development uses compiled evaluation.

6.3. **The certified threshold under measurement errors.** A separate development proves Theorem 2.6 in the same way. It imports only Mathlib. Up to names, it repeats the parts of the development of Section 6.2 that it needs: ML decoding and the Clopper–Pearson arithmetic. Its graph is $\mathcal{Q}_{d,w,T}$, the box $[0, T+1) \times [0, w) \times [0, d-1)$ of checks with half-edges at the two ends of the last coordinate, and the class of an error is the parity of its half-edges at b_1 over all layers. The decoder is that of Section 5.5, with any minimum-weight corrections in its later stages, and

the 25 probabilities ε_t are defined on the code of $5 \times 5 \times 5$ blocks. Lean derives (2.2) from (V6), (Hcov) and $\varepsilon_t \le 1.85 \times 10^{-5}$, with no other hypothesis. The derivation goes through the failure bound of the decoder, whose deterministic lemma and counts Lean proves in full; Lean evaluates the constants of the proof, 803.44 and about 5.106×10^6, and rounds them up. It also derives the threshold statement for all subexponential w and T. It derives the confidence statement of Section 2.4 for both statements, here too with a single acceptance event, valid for all w and T. In exact integer arithmetic, the kernel checks that the counts of Table 7 satisfy the rule and that $F(14; 3 \times 10^6, 1.85 \times 10^{-5}) \le 4 \times 10^{-11}$, and it checks the quoted bound 1.1706×10^{-5}. Its axiom report covers 849 declarations, among them every theorem of it in Table 8, and each of them depends only on the standard axioms of Section 6.4.

6.4. **Scope and trusted base.** The formal proofs do not cover the following:

(i) the identification of the planar, toric and phenomenological codes with their decoding graphs, which is standard; for the rotated code Lean proves it;
(ii) the symmetry that carries the planar bounds to the other sector of square codes (Section 2.1);
(iii) the argument of Section 2.4 that the rounding of the samplers' flip probabilities does not change the confidence;
(iv) for Theorems 2.5 and 2.6, the three conditions under which each holds, which enter the Lean proofs as hypotheses.

Proved on paper only are Theorem 5.3 in its general form, whose full proof is given in Appendix A; the sharper constants that follow it and those of Remark 2.8; and the single-scale criterion of Section 7.2, which the criterion supplement proves.

Lean reports, for every theorem, the axioms it depends on. The trusted base is Lean's kernel, including its built-in arithmetic on natural numbers, together with three standard axioms of Lean's foundations: propositional extensionality, the axiom of choice and the soundness of quotients. The theorems of Sections 6.2 and 6.3 and the kernel-route theorems of Section 6.1 depend on nothing else. The theorems marked "compiled evaluation" in Table 8 also depend on one axiom for each large certificate check: Lean evaluates the check, a Boolean test, with compiled code and records its outcome `true` as an axiom that names the test. Beyond the kernel, they rely on Lean's compiler and runtime, on the C compiler that builds the precompiled modules, and on these modules having been built from the checked sources. Programs outside Lean generate the certificate data, and the searches for exclusion witnesses are Lean functions; neither needs to be trusted (Section 6.1).

7. Discussion

7.1. **Monotone arguments.** Call an error set *failing* for minimum-weight matching if some minimum-weight correction of it fails. An event is *increasing* if it is preserved when errors are added. The argument of [7], the bounds of Sections 3 and 4, and every union bound over sets that the error must contain, such as paths or clusters, bound the failure probability by the probability of an increasing event that contains the failing sets. Every such event contains the event U that some subset of the error is failing, and U is itself increasing, so no such bound is smaller than $\Pr(U)$. For the planar code at $p = 0.093$, Monte Carlo estimates of $\Pr(U)$, with U decided by integer programming, grow from 0.264 at $d = 7$ to at least 0.357 at $d = 19$, while the failure probability of matching itself does not grow. These estimates are not rigorous, and the behaviour for large d is an extrapolation, but they indicate that arguments of this kind cannot prove a threshold for matching above 0.093, and in particular that the counting bounds cannot be pushed to the numerical threshold 0.1031 of matching. The certificates of Theorems 2.5

and 2.6 are not of this kind: a block is bad when window solutions leave a defect or a long residual path, and either can disappear when errors are added (Section 1.2).

Within the counting method, the value of p at which the growth rate of the labelled walks reaches one increases only slowly with the length of the windows that the exclusion rules inspect, from about 0.042 for windows of four edges to about 0.047 for windows of eleven edges (Table 4). We do not know how to go well beyond 0.05 without sampling. A natural next step would also use the correction away from the crossing, but that is a non-local function of the error.

7.2. **Statistical certification.** Theorem 2.5 is a theorem about unknown numbers, the seven bad-block probabilities, combined with a registered experiment that bounds them. Its confidence statement concerns the sampling only: it bounds the probability that the experiment accepts while the bound is false, under the binomial model, and it does not depend on any prior on the unknown numbers. Anyone can recompute the counts from the archived records, and a new run with fresh seeds would be an independent test of the same statement. The same holds for Theorem 2.6 and its 25 block types, except that, as Section 5.5 records, its tests of (V6) and (Hcov) at the registered radius were run once the analysis was complete.

A deterministic proof at $p = 0.102$ would need the bad-block probabilities themselves, which are probabilities of events depending on about 5.6×10^6 independent bits, or a criterion whose inputs can be computed exactly. The criterion supplement (see Code and data) proves such a criterion for optimal decoding, at a single scale. Draw a sample from the posterior law of the errors in a window, given the syndrome in the window. If, for each of four window shapes, in the bulk, at a rough side, at a smooth side and at a corner, the probability that this sample and the true error differ by a long path near the centre of the window, plus twice the square root of a measure of how strongly the errors on the edges that leave the window influence the posterior law near its centre, is less than $1/16$ at p_*, then the optimal threshold of the planar code is at least p_*. The inputs need exact posterior samples and partition functions with certified numerics on windows of side near 10^3. The criterion is not checked in Lean, and we have not certified it at any p.

7.3. **Beyond $p = 0.102$.** The certified decoder takes its weights from the window posteriors. With matching weights instead, the bad-block probability at $p = 0.10$ stayed at several per cent in pilot runs even for blocks of side 256 and windows of radius 320, because the residual loops of matching approach the block size near the matching threshold. Posterior weights remove this obstacle, but the window side needed to make bad blocks rare still grows as p approaches the Nishimori point, and the cost of computing the window marginals grows like the cube of the window side. At $p = 0.102$ the windows have side $L + 2R = 1118$, and the experiment took about 4.9×10^4 core-hours. Larger windows also cost in the deterministic analysis. A window radius $R \geq L$ raises the independence distance D, so that bad blocks count as independent only at larger distances, and the series of Theorem A.8 converge only for smaller ε: below about 4×10^{-4} at $D = 6$, against about 6×10^{-4} at $D = 5$ (Appendix A.4).

7.4. **Measurement errors.** Theorem 2.6 carries the method of Theorem 2.5 over to measurement errors; its first stage is plain matching with uniform weights. The value 0.02 lies below the numerical thresholds for this noise, 0.0293(2) for matching [34] and near 0.033 for optimal decoding [26], and above the value 0.016 that Theorem 2.4 gives without sampling for every minimum-weight decoder.

7.5. **Open problems.**

(1) A deterministic proof that the optimal threshold of the surface code exceeds 0.05, for instance by certifying the single-scale criterion of Section 7.2 with exact arithmetic.

(2) A rigorous threshold for minimum-weight matching above 0.0465. By the discussion above, such a proof cannot rest on an increasing event alone near the numerical threshold of matching.
(3) Certified bounds under measurement errors that approach the numerical thresholds, and bounds for circuit-level noise beyond the matching thresholds proved so far.

Code and data

The material behind this paper is archived in four parts, each with a file listing the SHA-256 checksums of its files:

- the Lean development of Theorem 2.5: record 10.5281/zenodo.23994431; SHA-256 of its checksum file: `e23a17d0c77ac6b17350ed86569ef3b7bfae7148a1a47ccf0930964cf4e8198c`;
- the Lean development of Theorem 2.6: record 10.5281/zenodo.23994431; SHA-256 of its checksum file: `834111ee330382643e873ca897c60dc79d887a2316b2e5357c8f2706fa70e133`;
- the Lean development of Theorems 2.1–2.4: record 10.5281/zenodo.23994431; SHA-256 of its checksum file: `76f02636f65959e936d2aae8cc75f700ae6b78b809e534bffbb4848278be31cd`;
- the registrations, the programs, the certificate data, the outputs of all checks, the records supplement (records of the experiments and of the formal checks, cited here as Sections R.1–R.3) and the criterion supplement (the single-scale criterion of Section 7.2 with its proof): record 10.5281/zenodo.23994430; SHA-256 of its checksum file: `739f509ec93cebd39b0bb7a615a4e6a1705e8b0c78590d68d1e2790bb9297e9a`.

The Lean developments use Lean 4 version 4.35.0-rc3 and Mathlib at commit `c55e6e78`. Details of the computing setup in the archived files, such as file paths, host, machine, account and process names, processor and machine types, cloud zones and internal working names, were normalized for release; in the files of the certificate at $p = 0.102$, every quoted SHA-256 of a normalized file as written was replaced as well. No measured value was changed. The record files of the certificate at $p = 0.102$ are released byte for byte, and the archive lists every normalized file with the SHA-256 of its released version. The records are released under CC BY 4.0, with the Lean sources and the programs under Apache-2.0.

Appendix A. Proof of Theorem 5.3

This appendix proves Theorem 5.3, in the explicit form of Theorem A.8, and records how its constants are evaluated. We use the decoder of Section 5.1, with $0 \le R < \frac{3}{2}L$ and $g = \lfloor R/L \rfloor + 1 \in \{1, 2\}$, on $\mathcal{P}_{H+1,W}$ with $W = n_x L$, $H = n_y L$, n_x as there and $n_y \ge 2g + 2$, and the bad blocks, types, declared rows and hypothesis (H_ε) of Section 5.3. For a check x, $\beta(x)$ denotes its block; block displacements are measured with the norms $|\cdot|_\infty$ and $|\cdot|_1$ of their indices, $N_k(c)$ is the set of blocks at distance at most k from c, and $d(x, y)$ is the least weight of a path between checks x and y. Stage 1 is any program as in Theorem 5.3. Let $F = E \oplus C_0$; its odd checks are $\mathcal{D}'$.

A.1. **Block events.**

Lemma A.1 (dependence)**.** *Whether c is bad depends only on the error on the edges that have an endpoint in the union $U(c)$ of the windows of the blocks of $N_1(c)$. With $D = 4 + \lfloor 2R/L \rfloor$, the bad events of any set of blocks with pairwise distances at least D are independent.*

Proof. Whether c is flagged depends on $\mathcal{D}$ and on C_0 on the edges at the checks of c; whether c is long depends on F on the edges with both endpoints in $N_1(c)$. Each of these edges has its left or lower endpoint in $N_1(c)$, so it is owned by a block of $N_1(c)$, and its value in C_0 is a function of the syndrome in that block's window, which is a function of E on the edges with an endpoint in the window. These edges, and the edges at the checks of $N_1(c)$ themselves, have an endpoint in $U(c)$; so bad(c) depends only on E on the edges with an endpoint in $U(c)$. In block coordinates, if c is in column a, the checks of $U(c)$ have first coordinate in $[aL - L - R,\, aL + 2L + R - 1]$, and similarly for the rows. For blocks c, c' at distance n, say in the first coordinate, an edge with an endpoint in $U(c)$ and an endpoint in $U(c')$, or a check in both, exists only if $aL + 2L + R \geq (a+n)L - L - R$, that is, if $(n-3)L \leq 2R$. Hence the two edge sets are disjoint when $(n-3)L > 2R$, which holds for $n \geq 4 + \lfloor 2R/L \rfloor$, and disjoint edge sets carry independent errors. □

If every block of type t is bad with probability ε_t, then (H_ε) holds with $\varepsilon = \max_t \varepsilon_t$, by Lemma A.1.

A.2. **The deterministic lemma.** From now on, $a_0 \geq D - 1$, $a_1 \geq a_0$, $a_1 \geq 2D - 2$, and a_1 is even. (The condition $a_1 \geq a_0$ is used only in the strip argument below.) The declared rows are $0, \ldots, g$ and $n_y - g - 1, \ldots, n_y - 1$. Since $R \geq (g-1)L$, we have $D \geq 2g + 2$; hence $a_0 \geq 2g + 1$, so that $\lceil a_0/2 \rceil \geq g + 1$, and $h_s = a_1/2 + 1 \geq D > g$ for the strip depth h_s used below. Besides the declared rows, this is all that the proof uses about g. The lemmas of this section carry the letters (A), (B), (C), (W) and (X) of the outline of the proof of Theorem 5.3.
Clusters. With the blocks of the declared rows counted as bad, the *clusters* are the components of the bad blocks (declared or not) under adjacency at distance at most a_0. Distinct clusters are at distance at least $a_0 + 1 \geq D$. G_{bot} and G_{top} are the clusters that contain the declared rows at the bottom and at the top (possibly the same cluster). Every other cluster is *interior*; its blocks lie in the rows $g + 1 + a_0, \ldots, n_y - g - 2 - a_0$.

Lemma A.2 (A)**.** *Take the components of F on the checks (a half-edge belongs to the component of its check; b_1 and b_2 join nothing). A component of F that meets no bad block has no odd check and lies within distance $L - 1$ of each block it meets. Every component of F meets at most one cluster.*

Proof. Let c be a good block. Then c is in the rows $g + 1, \ldots, n_y - g - 2$, and no path of F from a check of c reaches ℓ_∞ distance L from c: every edge changes each coordinate by at most 1, so the first check of such a path at distance at least L from c is at distance exactly L, and the path up to that check lies in $N_1(c)$, which would make c long. (Since c is in rows $g + 1, \ldots, n_y - g - 2$ and $g \geq 1$, the blocks of $N_1(c)$ contain no check of row 0 or $H - 1$, so the path meets no half-edge before that check.) The odd checks of F are the residual defects, which lie in flagged, hence bad, blocks; so a component meeting no bad block has none.

Suppose a component meets two clusters $K_1 \neq K_2$. A path in it from K_1 to K_2 leaves K_1 into a block c' adjacent to K_1, which is good because $a_0 \geq 1$. The whole component then lies within distance $L - 1$ of c', so it meets only blocks within distance 1 of c'; in particular K_2 has a block within distance 1 of c', hence within distance $2 \leq a_0$ of K_1, a contradiction. □

Since flagged blocks are bad and the level-0 adjacency equals the cluster adjacency, every level-0 component lies in one cluster. Two clusters are *linked* if a level-1 component (resolved or not) meets both, and the *families* are the classes of clusters under the equivalence generated by linking. For a family $\mathcal{F}$ let $Y_{\mathcal{F}}$ be the sum of the components of F that meet $\mathcal{F}$ and of the corrections C_K of the resolved components inside $\mathcal{F}$.

Lemma A.3 (B, decomposition)**.** *$\partial Y_{\mathcal{F}} = \Omega \cap \mathcal{F}$ for every family, and*

$$E \oplus C = \Big(\sum_{\text{components of } F \text{ meeting no bad block}} \Big) \oplus \bigoplus_{\mathcal{F}} Y_{\mathcal{F}} \oplus C_{\text{top}}.$$

The components of F that meet no bad block are closed and do not change the class.

Proof. $E \oplus C = F \oplus \bigoplus_K C_K \oplus C_{\text{top}}$, and F is the sum of its components; by Lemma A.2, each component meeting a bad block meets exactly one cluster, hence one family. The components of F meeting $\mathcal{F}$ have as odd checks the residual defects in the flagged blocks of $\mathcal{F}$, and each resolved C_K inside $\mathcal{F}$ has odd checks $\mathcal{D}' \cap K$. Every flagged block of $\mathcal{F}$ is resolved exactly once, or its residual belongs to Ω; this gives $\partial Y_{\mathcal{F}} = \Omega \cap \mathcal{F}$. A component meeting no bad block has no odd check and lies within distance $L-1$ of a good block, so it touches no half-edge and its height is less than $3L \le H$; it is therefore a cycle of checks with no half-edge, which does not change the parity at b_1. □

Lemma A.4 (C, failure path)**.** *If the decoder fails, then one of the following holds:*

(T1) some family $\mathcal{F}$ has $Y_{\mathcal{F}}$ with odd degree at both b_1 and b_2;

(T2) there is a sequence $b_1 \to \mathcal{F}_1 \to \cdots \to \mathcal{F}_m \to b_2$ of distinct families in which consecutive families are joined by paths of C_{top}, and the first (last) link is a path of C_{top} from b_1 (to b_2) or an odd degree of $Y_{\mathcal{F}_1}$ at b_1 (of $Y_{\mathcal{F}_m}$ at b_2).

Proof. Decompose C_{top} into paths that pair the checks of Ω with each other or with a boundary vertex, and form the multigraph whose vertices are the families and b_1, b_2: it has an edge for each such path (between the families, or the family and boundary vertex, containing its ends) and an edge $\mathcal{F}b_i$ whenever $Y_{\mathcal{F}}$ has odd degree at b_i. By the handshake lemma for $Y_{\mathcal{F}}$ and Lemma A.3, every family vertex has even degree. If the decoder fails, $E \oplus C$ has odd degree at b_1, and by Lemma A.3 so has b_1 in the multigraph; hence b_1 and b_2 lie in one component of the multigraph, and a simple path joins them. If this path is $b_1 \to \mathcal{F} \to b_2$ and both of its edges are odd degrees of $Y_{\mathcal{F}}$, this is case (T1); otherwise the path is as in (T2). (A minimum-weight C_{top} contains no path between two boundary vertices, since weights are positive.) □

Lemma A.5 (W, parity witnesses)**.**

(W0) An odd level-0 component K in an interior cluster has, in the same cluster, a flagged block at distance at least $a_0 + 1$ from K.

(W1) Let K_1 be an odd level-1 component such that every cluster meeting K_1 is interior. Then one of these clusters contains a flagged block of another level-1 component, at distance at least $a_1 + 1$ from K_1.

(R) Every check of Ω lies in a flagged block of an odd level-0 component whose blocks belong to an odd level-1 component. Consequently, a path of C_{top} between two different families of a (T2) chain joins blocks at distance at least $a_1 + 1$.

Proof. The components of F that meet an interior cluster reach no rough side, since the cluster's blocks lie in the rows $g + 1 + a_0, \ldots, n_y - g - 2 - a_0$ and the components stay within distance $L-1$ of good blocks outside it. Their odd checks are the residual defects in the cluster's flagged blocks, so the number of these defects is even. For (W0): $|\mathcal{D}' \cap K|$ is odd, so the cluster has residual defects outside K, that is, in another level-0 component, which is at distance at least $a_0 + 1$ from K. For (W1), add to the components of F meeting an interior cluster the resolved level-0 corrections inside it; the odd checks of the sum are the residual defects of the cluster's odd level-0 components, again an even number. Summing over the clusters that meet K_1, all interior by assumption, the total number of such defects is even, while K_1 contains an odd number of them; so some cluster has level-1 residual outside K_1, that is, in a flagged block of

another level-1 component, at distance at least a_1+1 from K_1. (R) holds by the construction of Ω. Blocks of distinct level-1 components are more than a_1 apart, and the two ends of a path of C_{top} between different families lie in different level-1 components, because a level-1 component lies in a single family. □

Strips and the sub-path of a chain. Put $h_s = a_1/2+1$. The bottom strip S_{bot} is the set of block rows $\le h_s$, the top strip S_{top} the set of rows $\ge n_y - 1 - h_s$. A *bottom (top) strip family* is a family with a block in S_{bot} (S_{top}); a family containing G_{bot} is a bottom strip family. A family $\mathcal{F}$ with $Y_{\mathcal{F}}$ of odd degree at b_1 is also a bottom strip family: a half-edge of $Y_{\mathcal{F}}$ at b_1 lies in a component of F meeting block row 0, whose blocks are declared and belong to G_{bot}, or in a correction C_K whose region reaches row 0, so that K has a block within $\max(\lceil a_0/2\rceil, a_1/2) \le h_s$ rows of row 0. The same holds at the top. On a (T2) chain, let $\mathcal{F}_0$ be the last bottom strip family (b_1 if there is none) and $\mathcal{F}_{m+1}$ the first top strip family after it (b_2 if there is none); the families $\mathcal{F}_1, \dots, \mathcal{F}_m$ between them are *intermediate*. If $\mathcal{F}_0$ is also a top strip family, it has blocks in both strips; this case is counted with (T1) in Section A.3, so we assume $\mathcal{F}_0 \neq \mathcal{F}_{m+1}$ there. An intermediate family contains no block of the strips and hence, since $g < h_s$, no declared block; its level-1 regions $N_{a_1/2}(K)$ stay in the rows ≥ 2 and its clusters are interior, so (W0) and (W1) apply to all of them. Residual defects lie in unresolved components K, whose regions $N_{\lceil a_0/2\rceil}(K)$ contain no check of row 0 or $H-1$; since $\lceil a_0/2\rceil \ge g+1$, the entry and exit defects of every family on the chain lie in the rows $g+2, \dots, n_y - g - 3$, never in a declared row.

For $1 \le i \le m$ let u_i and v_i be the checks of Ω at which the chain enters and leaves $\mathcal{F}_i$, let v_0 be the check at which it leaves $\mathcal{F}_0$ and u_{m+1} the check at which it enters $\mathcal{F}_{m+1}$, and let P_i be the path of C_{top} from v_i to u_{i+1}.

Lemma A.6 (X, exchange)**.** $\sum_{i=0}^{m} w(P_i) \le \sum_{i=1}^{m} d(u_i, v_i) + d(v_0, b_1) + d(u_{m+1}, b_2)$*, where* $d(x, b_j)$ *is the least weight of a path from* x *to a half-edge at* b_j*, the half-edge included.*

Proof. Remove $P_0, \dots, P_m$ from C_{top}, join u_i to v_i by a shortest path for $1 \le i \le m$, and join v_0 to b_1 and u_{m+1} to b_2 by shortest paths. The result is again an edge set whose odd checks are Ω, and C_{top} has minimum weight among such sets. □

A.3. **The probability bound.**

A.3.1. *Metric facts.* Write $|z|_1$ for the ℓ_1 norm. Since every edge changes $|\cdot|_1$ by at most 1 and weighs at least 0.98, every path between checks x and y weighs at least $0.98\,|x-y|_1$. A monotone lattice path inside the rectangle of checks gives $d(x,y) \le 1.02\,|x-y|_1$, and a straight path down to row 0 followed by a half-edge gives $d(x, b_1) \le 1.02\,(y_x + 1)$ for a check x in row y_x, and similarly at the top. For checks x, y in blocks c, c',

$$L\,(|c'-c|_1 - \nu) \le |x-y|_1 \le L\,(|c'-c|_1 + \nu), \qquad \nu = 2,$$

so the offset between a check and its block costs at most ν in units of L. Put $\kappa = 1.05 \ge 1.02/0.98$.

A.3.2. *Selection trees.* For a cluster $\mathcal{C}$ and a block of it chosen as its *root*, process the blocks of $\mathcal{C}$ in breadth-first order for the adjacency at distance at most a_0, starting from the root, and *select* a block if it is at distance at least D from every block selected before.

Lemma A.7. *(i) The selected blocks are pairwise at distance at least* D*. (ii) Every block of* $\mathcal{C}$ *is within distance* $D-1$ *of a selected block. (iii) Every selected block other than the root has a selected block processed before it at a distance in* $[D, A_0]$*,* $A_0 = D - 1 + a_0$*; so the selected blocks*

form a tree $\tau_{\mathcal{C}}$ whose steps lie in $S_0 = \{v \in \mathbb{Z}^2 : D \le |v|_\infty \le A_0\}$. Selected blocks of distinct clusters are at distance at least $a_0 + 1 \ge D$.

Proof. (i) and (ii) hold by construction. For (iii), a block b other than the root is a_0-adjacent to a block b'' processed before it, and b'' is within $D-1$ of a block b' selected before b; then $|b - b'|_\infty \le a_0 + D - 1 = A_0$, and $|b - b'|_\infty \ge D$ because b was selected. The last claim holds because distinct clusters are at distance at least $a_0 + 1$. □

Hence every set of selected blocks, of one or several clusters, that avoids the declared rows satisfies the hypothesis of (H_ε).

A.3.3. *Generating functions.* Put $A_1 = D - 1 + a_1$, $S_1 = \{v \in \mathbb{Z}^2 : D \le |v|_\infty \le A_1\}$, $T_0 = |S_0|$, $T_1 = |S_1|$ and $\sigma_w = a_1 + 2 - D$. For $\lambda > 0$ and $s \ge 0$ let

$$t(v) = e^{\lambda\kappa|v|_1 + s\,v_y}, \qquad \tilde{T}_0 = \sum_{v\in S_0} t(v), \qquad \tilde{T}_1 = \sum_{v\in S_1} t(v),$$

where v_y is the signed row component of v. Define

$$U_0 = \varepsilon T_0, \quad \tilde{U}_0 = \varepsilon\tilde{T}_0, \quad \tilde{Q} = \varepsilon T_0 + \frac{\tilde{U}_0}{1-\tilde{U}_0}, \quad \tilde{V} = \varepsilon\tilde{T}_1\tilde{Q}, \quad V_0 = \varepsilon T_1\Big(\varepsilon T_0 + \frac{U_0}{1-U_0}\Big),$$

$$W_0 = \frac{1}{1-\tilde{U}_0}, \qquad B_0 = \frac{1}{1-U_0}, \qquad Q_m = \big((1-\tilde{U}_0)^{-2} - 1\big)B_0 + \varepsilon T_0 + B_0 - 1.$$

For $a, c \ge 0$ let $M_{a,c} = \sum_{x:\,|x|_\infty \ge \sigma_w} \big(w^{*a} * \mathbf{1}_{S_0}^{*c}\big)(x)$, where $w = \mathbf{1}_{S_0}\, t$ and $*$ is convolution on $\mathbb{Z}^2$: it counts tree paths with a tilted and then c untilted steps in S_0 that end at ℓ_∞ distance at least σ_w. Put

$$\mathrm{Wit}_{\mathrm{on}} = W_0 \sum_{a,c\ge 0} \varepsilon^{a+c} M_{a,c}, \qquad \mathrm{Wit}_{\mathrm{off}} = \sum_{c\ge 1} \varepsilon^c M_{0,c}\big|_{\lambda=s=0},$$

$\mathrm{off} = (1+\kappa)D\nu$, and

$$\Phi = e^{\lambda\,\mathrm{off}}\,\varepsilon\Big[\mathrm{Wit}_{\mathrm{on}} + \frac{Q_m\,\varepsilon T_1\,\mathrm{Wit}_{\mathrm{off}}}{1-V_0}\Big](1-\tilde{V})^{-2}, \qquad J = \sum_{\Delta\in\mathbb{Z}^2,\,|\Delta|_\infty\ge\sigma_w} e^{-\lambda|\Delta|_1 + s\Delta_y}, \qquad r = J\,\Phi.$$

J is finite if and only if $s < \lambda$. For the spanning case let $T_1(s) = \sum_{v\in S_1} e^{sv_y}$, $U_0(s) = \varepsilon\sum_{v\in S_0} e^{sv_y}$ and

$$V_1(s) = \varepsilon T_1(s)\Big(\varepsilon T_0 + \frac{U_0(s)}{1-U_0(s)}\Big).$$

A.3.4. *The theorem.*

Theorem A.8. *Let $L \ge 2$, $0 \le R < \frac{3}{2}L$, $D = 4 + \lfloor 2R/L \rfloor$, $a_0 \ge D - 1$, and let a_1 be even with $a_1 \ge a_0$ and $a_1 \ge 2D - 2$; put $h_s = a_1/2 + 1$ and $H_0 = 2h_s + A_1 + 1 = 2a_1 + D + 2$. Assume (H_ε).*

(a) If $V_1(0) < 0.9$, then with $s_1 = \sup\{s \ge 0 : V_1(s) \le 0.9,\ U_0(s) \le 0.9\}$,

$$\Pr(\mathrm{T1},\ \textit{or (T2) with } \mathcal{F}_0 = \mathcal{F}_{m+1}) \le C_1\, n_x\, e^{-s_1(n_y - H_0)}, \qquad C_1 = \frac{(h_s+1)\,W_1\,(1+\varepsilon T_1(s_1)W_1)}{1-V_1(s_1)},$$

where $W_1 = 1/(1-U_0(s_1))$.

(b) If at some $0 \le s < \lambda$ we have $\tilde{U}_0 < 1$, $\tilde{V} < 1$, $V_0 < 1$ and $r < 1$, then

$$\Pr(\mathrm{T2},\ \mathcal{F}_0 \ne \mathcal{F}_{m+1}) \le C_2\, n_x\, e^{-s(n_y - H_0)}, \qquad C_2 = \frac{4(h_s+1)\,G_{\mathrm{bot}}\,J\,E_{\mathrm{top}}\,e^{\lambda c_{\mathrm{ends}}}}{1-r},$$

where $G_{\mathrm{bot}} = W_0/(1-\tilde{V})$, $E_{\mathrm{top}} = \varepsilon W_0(1+\varepsilon\tilde{T}_1 W_0)/(1-\tilde{V})$ and $c_{\mathrm{ends}} = D\nu + \kappa D\nu + \kappa\nu + \kappa(2h_s + A_1) + 2.04\,\kappa/L$.

Consequently, the failure probability of the decoder on $\mathcal{P}_{n_yL+1,n_xL}$ is at most $C_1 n_x e^{-s_1(n_y-H_0)} + C_2 n_x e^{-s_2(n_y-H_0)}$ with $s_2 = s$.

This is Theorem 5.3 with explicit constants; the constants depend on L only through the last term of c_{ends}, which decreases in L. The rest of this section proves it. Throughout, a *configuration* is an error E, and every sum over block tuples counts each tuple once; tuples that cannot arise only add nonnegative terms.

A.3.5. *Routes through a family.* Fix a family $\mathcal{F}$ of a (T2) chain, with entry check u and exit check v (for $\mathcal{F}_0$ only an exit, for $\mathcal{F}_{m+1}$ only an entry). Join two clusters of $\mathcal{F}$ if some level-1 component has odd level-0 components o in one and o' in the other at distance at most a_1. This graph on the clusters of $\mathcal{F}$ is connected, because each level-1 component is a_1-connected and each odd level-0 component lies in one cluster. Take a simple path $c_0 = \mathcal{C}(u), \dots, c_t = \mathcal{C}(v)$ in it, where $\mathcal{C}(x)$ is the cluster containing $\beta(x)$, and for each step a pair (o, o') at distance at most a_1 with blocks $b_o \in o$ and $b_{o'} \in o'$ at distance at most a_1. Root c_0 at $\beta(u)$ and c_{j+1} at $b_{o'}$, and select blocks as in Lemma A.7. The *route* of $\mathcal{F}$ is the sequence of selected blocks formed by the tree path in c_j from its root to the selected block s_j that covers b_o (in-cluster steps in S_0), the *link* $s_j \to \text{root}(c_{j+1})$, and so on, ending in c_t with the tree path from its root to the selected block $\sigma(\mathcal{F})$ that covers $\beta(v)$. A link has length at least $a_0 + 1 \geq D$ (the clusters are distinct) and at most $(D-1) + a_1 = A_1$, so its step lies in S_1.

Leaves. Every root is a block of an odd level-0 component: $\beta(u)$ by (R), and $b_{o'}$ because o' belongs to a level-1 component. In an interior cluster, by (W0) and Lemma A.7(ii) the cluster has at least two selected blocks. For every route cluster whose route part is its root alone, other than the cluster $\mathcal{C}^*$ below, add one tree neighbour of the root (a *leaf*, with displacement in S_0, not on the route).

Witness. Let K_j and K'_j be the level-1 components of the links that enter and leave c_j, with $K_0 = K(u)$ and $K'_t = K(v)$, the level-1 components of u and v.

(α) If $K_j \neq K'_j$ for some j, put $\mathcal{C}^* = c_j$ and $e^* = \text{root}(c_j)$, a block of K_j. The block b_o of the leaving link (or $\beta(v)$ if $j = t$) belongs to K'_j, so it is at distance at least $a_1 + 1$ from e^*, and the selected block b' that covers it satisfies $|b' - e^*|_\infty \geq a_1 + 1 - (D-1) = \sigma_w$.

(α') Otherwise every link belongs to $K = K(u) = K(v)$, which is odd because it carries the residual check u, and by (W1) a cluster $\mathcal{C}^*$ meeting K contains a block of a level-1 component $K' \neq K$. If $\mathcal{C}^*$ is a route cluster, its root e^* is a block of K, and the selected block b' covering the block of K' satisfies $|b' - e^*|_\infty \geq \sigma_w$.

(β) If $\mathcal{C}^*$ is not a route cluster, follow a chain of odd level-0 components of K from a route cluster $\mathcal{C}_x$ to $\mathcal{C}^*$, shortened so that the clusters after $\mathcal{C}_x$ are distinct and off the route, and root each of them at the block reached by the link into it (an *excursion*); again $e^* = \text{root}(\mathcal{C}^*)$ is a block of K and $|b' - e^*|_\infty \geq \sigma_w$.

In all cases the tree path π in $\mathcal{C}^*$ from e^* to b' has at least one step, so $\mathcal{C}^*$ needs no leaf. The *skeleton* of $\mathcal{F}$ consists of the route, the leaves, π and, in case (β), the excursion: the tree path in $\mathcal{C}_x$ from a route block to the start of the first excursion link, the excursion links, and the excursion clusters with their own leaves. All skeleton blocks are distinct selected blocks, so by Lemma A.7 and (H_ε) the probability that all of them are bad is at most ε to the number of them. (The skeleton of an intermediate family contains no declared block.)

A.3.6. *Proof of part (a).* Let $\mathcal{F}$ be a family with blocks in both strips; this covers (T1), where some $Y_{\mathcal{F}}$ has odd degree at b_1 and at b_2, and the case $\mathcal{F}_0 = \mathcal{F}_{m+1}$. Take, as in Section A.3.5, a route through the clusters of $\mathcal{F}$ from a block in S_{bot} or in a declared bottom row to a block in S_{top} or in a declared top row, with leaves on the link-entered clusters. If $\mathcal{F}$ contains G_{bot}, choose the route to start at a declared block of G_{bot}; this is possible because the graph on the

clusters of $\mathcal{F}$ is connected, and then the route, a simple path in that graph, never returns to G_{bot}. Cut it at the last block b_s that lies in S_{bot} or in a declared row and at the first later block that lies in S_{top} or in a declared row; the counted part runs from b_s to the block b_e before that cut. There are at most $(h_s+1)\,n_x$ choices for b_s, which is not counted with a factor ε; every later route block and every leaf is bad, selected and not declared. Every later link enters an interior cluster, which has (W0), except possibly one link into the non-declared part of G_{top}, whose target may lack (W0). The block b_e is within A_1 rows of a block of S_{top} or of a declared row, so the signed row progress from b_s to b_e is at least $(n_y-1-h_s-A_1)-h_s = n_y-H_0$. Multiplying the indicator by $e^{s(\mathrm{progress}-(n_y-H_0))}\geq 1$ and distributing $e^{s\,\mathrm{progress}}$ over the steps gives, summed over all counted routes,

$$(h_s+1)\,n_x\cdot W_1\cdot\sum_{k\geq 0}V_1(s)^k\cdot(1+\varepsilon T_1(s)W_1)\cdot e^{-s(n_y-H_0)},$$

since the in-cluster steps from b_s contribute $\sum_k U_0(s)^k = W_1$, each link into an interior cluster contributes $\varepsilon T_1(s)$ times εT_0 (root with a leaf, the route staying at the root) plus $U_0(s)/(1-U_0(s))$ (one or more route steps), which is $V_1(s)$, and the one exceptional link contributes at most $1+\varepsilon T_1(s)W_1$. At $s=s_1$ this is $C_1 n_x e^{-s_1(n_y-H_0)}$.

A.3.7. *The counting lemma for intermediate families.*

Lemma A.9. *Let $\mathcal{F}_1,\dots,\mathcal{F}_m$ be the intermediate families of a (T2) chain, $D'_i=\mathrm{root}(\mathcal{F}_{i+1})-\sigma(\mathcal{F}_i)$ the jump displacements ($0\leq i\leq m$, with $\sigma(\mathcal{F}_0)$ the block of $\mathcal{F}_0$'s route that covers $\beta(v_0)$) and $R_i=\sigma(\mathcal{F}_i)-\mathrm{root}(\mathcal{F}_i)$ the route displacements ($1\leq i\leq m$). Then:*

(i) $|D'_i|_\infty\geq\sigma_w$ *for every jump;*

(ii) $\sum_{i=0}^m|D'_i|_1\leq\kappa\sum_{i=1}^m|R_i|_1+m\,\mathrm{off}+c_{\mathrm{ends}}+\kappa\,\rho_{\mathrm{bot}}+\kappa\,\rho_{\mathrm{top}}$, *where ρ_{bot} and ρ_{top} are the ℓ_1 lengths, in blocks, of the end routes of Section A.3.8;*

(iii) summed over all skeletons of $\mathcal{F}_1,\dots,\mathcal{F}_m$ and all landing roots, with the weight ε per skeleton block, $e^{\lambda\kappa|v|_1+sv_y}$ per route step v, and $e^{-\lambda|D'|_1+sD'_y}$ per jump, each intermediate family contributes at most Φ and each jump at most J.

Proof. (i) By (R), v_i and u_{i+1} lie in odd level-0 components of odd level-1 components, in $\mathcal{F}_i$ and $\mathcal{F}_{i+1}$; a level-1 component lies in one family, so these level-1 components differ, and blocks of different level-1 components are at distance at least a_1+1. Since $|\sigma(\mathcal{F}_i)-\beta(v_i)|_\infty\leq D-1$, we get $|D'_i|_\infty\geq a_1+1-(D-1)=\sigma_w$.

(ii) By Section A.3.1, the path P_i weighs at least $0.98\,L\,(|D'_i|_1-D\nu)$: the root of $\mathcal{F}_{i+1}$ is $\beta(u_{i+1})$, and $\sigma(\mathcal{F}_i)$ is within $D-1$ of $\beta(v_i)$, which costs $(D-1)\nu$, plus ν for the offsets of the checks in their blocks. Likewise $d(u_i,v_i)\leq 1.02\,L\,(|R_i|_1+D\nu)$. Inserting these and the end bounds of Section A.3.8 into Lemma A.6 and dividing by $0.98\,L$ gives (ii), since $1.02/0.98\leq\kappa$ and $\mathrm{off}=(1+\kappa)D\nu$.

(iii) For every configuration in (T2) choose one skeleton for each intermediate family as in Section A.3.5. On that configuration (ii) holds for the data of the chosen skeletons, so

$$1\leq\exp\Big(\lambda\Big(\kappa\sum_i|R_i|_1+m\,\mathrm{off}+c_{\mathrm{ends}}+\kappa\rho_{\mathrm{bot}}+\kappa\rho_{\mathrm{top}}-\sum_i|D'_i|_1\Big)\Big),$$

and $|R_i|_1$ is at most the sum of the ℓ_1 norms of the route steps of $\mathcal{F}_i$. The signed row progress of the block path from b_s to b_e (Section A.3.8) is the sum of the row components of all route steps and jumps, in-block offsets telescoping, and it is at least n_y-H_0; so also $1\leq e^{s(\mathrm{progress}-(n_y-H_0))}$. Taking expectations and the union over skeletons, the probability is at most the sum, over all skeleton tuples, of the product of these factors and ε per skeleton block. The sum factorizes. Per jump, the landing root of $\mathcal{F}_{i+1}$ ranges over the $\Delta=D'_i$ with $|\Delta|_\infty\geq\sigma_w$: factor J. Per

family in cases (α), (α'): the route clusters $c_0, \dots, c_t$, one of which is $\mathcal{C}^*$. Each link contributes $\varepsilon\tilde{T}_1$ (its target block and its tilted step), and each route cluster other than $\mathcal{C}^*$ contributes $\tilde{Q}$: either the root with a leaf (εT_0, the route staying at the root) or $k \geq 1$ tilted route steps ($\tilde{U}_0^k$). $\mathcal{C}^*$ contributes $\mathrm{Wit}_{\mathrm{on}}$: its route part from e^* shares an initial segment of a tilted steps with π, π continues with c untilted steps and ends at distance at least σ_w from e^* (the count $M_{a,c}$), and the rest of the route part from the branching block contributes W_0. The position of $\mathcal{C}^*$ among the $t+1$ route clusters gives $\sum_t (t+1)\tilde{V}^t = (1-\tilde{V})^{-2}$, and the landing root gives ε. In case (β), the marked route cluster $\mathcal{C}_x$ contributes Q_m: a route part with $k \geq 1$ steps, $k+1$ choices of the branching block and the excursion's in-cluster walk (B_0), or the root with a leaf and an empty excursion walk (εT_0), or a non-empty excursion walk from the root ($B_0 - 1$); the excursion links are untilted (εT_1 each), the excursion clusters contribute $\sum_j V_0^j = (1-V_0)^{-1}$, and $\mathcal{C}^*$ contributes $\mathrm{Wit}_{\mathrm{off}}$ through π from its root. Together with $e^{\lambda\,\mathrm{off}}$ per family this gives at most Φ per intermediate family. □

A.3.8. *The ends, and proof of part (b).* Assume $\mathcal{F}_0 \neq \mathcal{F}_{m+1}$, and first that both are strip families.

Bottom end. Choose a route in $\mathcal{F}_0$ that ends at the selected block $\sigma(\mathcal{F}_0)$ covering $\beta(v_0)$ and that starts in a declared row of G_{bot}, if $\mathcal{F}_0$ contains G_{bot} (the tree path of G_{bot} from a declared block, then links into interior clusters; the route never returns to G_{bot}), and otherwise at a block of S_{bot}. Let b_s be the last block of the route that is declared or lies in S_{bot}. There are at most $(h_s+1)\,n_x$ positions for b_s, each with weight at most 1; every later route block has its factor ε, and every later link enters an interior cluster, which has (W0). The route from b_s therefore contributes at most $G_{\mathrm{bot}} = W_0/(1-\tilde{V})$. Its length $\rho_{\mathrm{bot}} = |\sigma(\mathcal{F}_0) - b_s|_1$ enters (ii) above through

$$d(v_0, b_1) \leq 1.02\,L\,\big(\rho_{\mathrm{bot}} + D\nu\big) + 1.02\,\big(h_s L + 1.02\big),$$

a path from v_0 to a check of b_s, then straight down at most $h_s L$ rows, then a half-edge of weight at most 1.02. Divided by $0.98\,L$, with $1.02/0.98 \leq \kappa$, the last term gives $\kappa h_s + 1.02\,\kappa/L$; the two ends together give the term $2.04\,\kappa/L$ of c_{ends}.

Top end. The route of $\mathcal{F}_{m+1}$ starts at the landing root $\beta(u_{m+1})$, which has its factor ε (it is never declared), and stops before the first block that is declared or lies in S_{top}; let b_e be its last block. It may enter the non-declared part of G_{top} once by a link whose target lacks (W0), with weight $\varepsilon\tilde{T}_1$ instead of $\tilde{V}$, and then continues by in-cluster steps; so it contributes at most $E_{\mathrm{top}} = \varepsilon W_0(1+\varepsilon\tilde{T}_1 W_0)/(1-\tilde{V})$. Both of its ends are selected blocks, and b_e lies within $h_s + A_1$ block rows of the top row, so with $\rho_{\mathrm{top}} = |b_e - \beta(u_{m+1})|_1$,

$$d(u_{m+1}, b_2) \leq 1.02\,L\,\big(\rho_{\mathrm{top}} + \nu\big) + 1.02\,\big((h_s + A_1)L + 1.02\big).$$

Assembly. The chain from $\mathcal{F}_0$ to $\mathcal{F}_{m+1}$ has m intermediate families and $m+1$ jumps; the extra jump adds $D\nu$ in (ii). Adding the two end bounds, divided by $0.98\,L$, gives the constant $c_{\mathrm{ends}} = D\nu + \kappa D\nu + \kappa\nu + \kappa(2h_s + A_1) + 2.04\,\kappa/L$ of the theorem. The signed row progress from b_s (row at most h_s) to b_e (row at least $n_y - 1 - h_s - A_1$) is at least $n_y - H_0$. By Lemma A.9(iii), the probability of a (T2) chain with strip families at both ends is at most

$$(h_s+1)\,n_x \cdot G_{\mathrm{bot}} \cdot J \sum_{m \geq 0} r^m \cdot E_{\mathrm{top}} \cdot e^{\lambda c_{\mathrm{ends}}} \cdot e^{-s(n_y - H_0)}.$$

If $\mathcal{F}_0 = b_1$ or $\mathcal{F}_{m+1} = b_2$, the chain starts or ends with a path of C_{top} to the boundary. The landing sum of such a boundary match over the n_x block columns is dominated by the corresponding strip factor $(h_s+1)\,n_x\,G_{\mathrm{bot}}\,J$, respectively $J\,E_{\mathrm{top}}$, and a factor 4 covers the four variants. With $\sum_m r^m = (1-r)^{-1}$ this is $C_2\,n_x\,e^{-s(n_y - H_0)}$. This proves part (b), and with part (a) and Lemma A.4 the theorem.

A.4. **Evaluation of the constants.**

Method. All counts are exact integers: T_0, T_1 and the sums over S_0 and S_1 are finite sums. The witness counts $M_{a,c}$ for $a+c \le 9$ are computed by discrete Fourier convolution on a grid large enough that no wrap-around occurs, which reproduces the direct convolution up to rounding; the remaining terms with $a+c \ge 10$ are bounded by the sum over all walks, $\sum_{n\ge 10} \sum_{a=0}^{n} \tilde{U}_0^{\,a} U_0^{\,n-a}$. The landing sum has the closed form

$$J = \frac{1+x}{1-x}\Big(1 + \frac{y_+}{1-y_+} + \frac{y_-}{1-y_-}\Big) - \sum_{|\Delta|_\infty < \sigma_w} e^{-\lambda|\Delta|_1 + s\Delta_y}, \qquad x = e^{-\lambda},\ y_\pm = e^{-(\lambda \mp s)},$$

valid for $s < \lambda$. In the numerical evaluation the offset is taken as $\mathrm{off} + \kappa/2$, which only increases r. For given ε, the rate s of part (b) is the largest s (by bisection) for which $\min_\lambda r(\lambda, s) \le 0.9$ over a geometric grid of λ, possibly multiplied by a fixed fraction in $\{1, 0.9, 0.8, 0.7, 0.6\}$ chosen to minimize the code size at which the bound drops below 0.1; then λ is the minimizer at that s. The rate s_1 of part (a) is found by bisection. All arithmetic is in double precision. The stated constants are rounded: rates down, prefactors up, so that the stated bound is implied by the computed one.

Convergence limits. For $D = 6$, $a_0 = 5$, $a_1 = 16$ and $\kappa = 1.05$, one has $T_0 = 320$, $T_1 = 1728$, $\sigma_w = 12$, $h_s = 9$ and $H_0 = 40$. The spanning term converges ($V_1(0) \le 0.9$) for $\varepsilon \le 8.301 \times 10^{-4}$, and the chain term ($\min_\lambda r(\lambda, 0) \le 0.9$) for $\varepsilon \le 3.982 \times 10^{-4}$. For comparison, with $D = 5$, $a_0 = 4$ and $a_1 = 12$, which serve window radii $\frac{1}{2}L \le R < L$, the two limits are 1.341×10^{-3} and 6.176×10^{-4}.

The level of the registered experiment. At $\varepsilon = 3 \times 10^{-4}$, the level that the experiment of Section 5.4 certifies (with $D = 6$, $a_0 = 5$, $a_1 = 16$, $L \le R < \frac{3}{2}L$), the computation gives $s_1 = 0.1711083$ and $s_2 = 0.0325974$ (fraction 1), with prefactors $C_1 = 551.43$ and $C_2 = 1.9556 \times 10^5$ at these exact rates, and $C_1(0.171) = 539.55$ and $C_2(0.032) = 1.1091 \times 10^5$ at the rates rounded down. The stated prefactors below are at least both, so the stated bound

$$\Pr(\text{failure}) \le 5.52 \times 10^2\, n_x\, e^{-0.171\,(n_y - 40)} + 2 \times 10^5\, n_x\, e^{-0.032\,(n_y - 40)}$$

is implied by the computed one. At the smaller rates $s_1 = 0.10$ and $s_2 = 0.02$ the same computation gives $C_1 = 33.04$ and $C_2 = 7.32 \times 10^3$, so it also implies the bound

$$\Pr(\text{failure}) \le 35\, n_x\, e^{-0.10\,(n_y - 40)} + 2.5 \times 10^4\, n_x\, e^{-0.02\,(n_y - 40)}$$

with the constants of (2.1), for $n_x \ge 6$ and $n_y \ge 41$, under the hypotheses of Section 5.3. These evaluations take the term $2.04\,\kappa/L$ of c_{ends} at $L = 192$, so they hold for every $L \ge 192$; Lean proves the last bound for every $L \ge 2$ (Section 6.2). At the certified level $\hat{\varepsilon}$, the constants of Remark 2.8 are computed in the same way with the fraction 0.9.

A.5. **Remarks on the proof.**

(1) The proof never uses that the window corrections are optimal or that stage 1 uses a particular weight; this is why one certificate of stage-1 events serves every admissible a_1.
(2) Lemma A.2 uses that every edge changes each coordinate by at most one; this is where the square lattice enters.
(3) The later stages enter only through Lemma A.6 and the metric facts of Section A.3.1, that is, through the weight bounds $0.98 \le w_e \le 1.02$.
(4) The declared rows and the strips make the end families explicit: residual defects never lie in a declared row, so no landing sum over declared rows is needed.

References

1. C. Amoruso and A. K. Hartmann, *Domain-wall energies and magnetization of the two-dimensional random-bond Ising model*, Phys. Rev. B **70** (2004), 134425.

2. Paul Balister, Béla Bollobás, and Mark Walters, *Continuum percolation with steps in the square or the disc*, Random Structures Algorithms **26** (2005), 392–403.
3. Sergey Bravyi and Jeongwan Haah, *Analytic and numerical demonstration of quantum self-correction in the 3D Cubic Code*, 2011, arXiv:1112.3252.
4. C. J. Clopper and E. S. Pearson, *The use of confidence or fiducial limits illustrated in the case of the binomial*, Biometrika **26** (1934), 404–413.
5. Olivier Couronné, *New upper bound for the connective constant for square-lattice self-avoiding walks*, 2022, arXiv:2211.16146.
6. Leonardo de Moura and Sebastian Ullrich, *The Lean 4 theorem prover and programming language*, Automated Deduction – CADE 28, Lecture Notes in Comput. Sci., Springer, 2021, pp. 625–635.
7. Eric Dennis, Alexei Kitaev, Andrew Landahl, and John Preskill, *Topological quantum memory*, J. Math. Phys. **43** (2002), 4452–4505.
8. Ilya Dumer, Alexey A. Kovalev, and Leonid P. Pryadko, *Thresholds for correcting errors, erasures, and faulty syndrome measurements in degenerate quantum codes*, Phys. Rev. Lett. **115** (2015), 050502.
9. Jack Edmonds, *Paths, trees, and flowers*, Canad. J. Math. **17** (1965), 449–467.
10. Michael E. Fisher, *Statistical mechanics of dimers on a plane lattice*, Phys. Rev. **124** (1961), 1664–1672.
11. Austin G. Fowler, *Proof of finite surface code threshold for matching*, Phys. Rev. Lett. **109** (2012), 180502.
12. Christophe Garban and Thomas Spencer, *Continuous symmetry breaking along the Nishimori line*, J. Math. Phys. **63** (2022), 093302.
13. Alan George, *Nested dissection of a regular finite element mesh*, SIAM J. Numer. Anal. **10** (1973), no. 2, 345–363.
14. Oscar Higgott, *PyMatching: a Python package for decoding quantum codes with minimum-weight perfect matching*, ACM Trans. Quantum Comput. **3** (2022), no. 3, 1–16.
15. Oscar Higgott and Craig Gidney, *Sparse Blossom: correcting a million errors per core second with minimum-weight matching*, Quantum **9** (2025), 1600.
16. T. Horiguchi and T. Morita, *Existence of the ferromagnetic phase in a random-bond Ising model on the square lattice*, J. Phys. A **15** (1982), L75–L80.
17. M. Kac and J. C. Ward, *A combinatorial solution of the two-dimensional Ising model*, Phys. Rev. **88** (1952), 1332–1337.
18. P. W. Kasteleyn, *The statistics of dimers on a lattice*, Physica **27** (1961), 1209–1225.
19. Sun Woo P. Kim and Max McGinley, *Mixed-state topological order and error-correction thresholds in non-Abelian codes: rigorous results*, 2026, arXiv:2607.21706.
20. A. Yu. Kitaev, *Fault-tolerant quantum computation by anyons*, Ann. Physics **303** (2003), 2–30.
21. Alexey A. Kovalev and Leonid P. Pryadko, *Fault tolerance of quantum low-density parity check codes with sublinear distance scaling*, Phys. Rev. A **87** (2013), 020304(R).
22. Aleksander Kubica and John Preskill, *Cellular-automaton decoders with provable thresholds for topological codes*, Phys. Rev. Lett. **123** (2019), 020501.
23. Lin Lin, Jianfeng Lu, Lexing Ying, Roberto Car, and Weinan E, *Fast algorithm for extracting the diagonal of the inverse matrix with application to the electronic structure analysis of metallic systems*, Commun. Math. Sci. **7** (2009), no. 3, 755–777.
24. Lin Lin, Chao Yang, Juan C. Meza, Jianfeng Lu, Lexing Ying, and Weinan E, *SelInv—an algorithm for selected inversion of a sparse symmetric matrix*, ACM Trans. Math. Software **37** (2011), no. 4, 1–19.
25. H. Nishimori, *Internal energy, specific heat and correlation function of the bond-random Ising model*, Progr. Theoret. Phys. **66** (1981), 1169–1181.
26. Takuya Ohno, Gaku Arakawa, Ikuo Ichinose, and Tetsuo Matsui, *Phase structure of the random-plaquette Z_2 gauge model: accuracy threshold for a toric quantum memory*, Nuclear Phys. B **697** (2004), 462–480.
27. Manaka Okuyama and Masayuki Ohzeki, *Existence of a phase transition in the one-dimensional Ising spin glass model with long-range interactions on the Nishimori line*, J. Phys. A **59** (2026), 305204.
28. Francesco Parisen Toldin, Andrea Pelissetto, and Ettore Vicari, *Strong-disorder paramagnetic-ferromagnetic fixed point in the square-lattice $\pm J$ Ising model*, J. Statist. Phys. **135** (2009), 1039–1061.
29. Yan Ru Pei, *Vertical and reentrant ferromagnetic boundaries below the Nishimori point*, 2026, arXiv:2609.37479.
30. André Pönitz and Peter Tittmann, *Improved upper bounds for self-avoiding walks in $\mathbb{Z}^d$*, Electron. J. Combin. **7** (2000), R21.
31. Oliver Riordan and Mark Walters, *Rigorous confidence intervals for critical probabilities*, Phys. Rev. E **76** (2007), 011110.

32. The mathlib Community, *The Lean mathematical library*, Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs, ACM, 2020, pp. 367–381.
33. Tejasvi Singh Tomar, *A matching threshold of* 2×10^{-3} *for the rotated surface code under circuit-level noise*, Zenodo, 2026, https://doi.org/10.5281/zenodo.23725694.
34. Chenyang Wang, Jim Harrington, and John Preskill, *Confinement-Higgs transition in a disordered gauge theory and the accuracy threshold for quantum memory*, Ann. Physics **303** (2003), 31–58.
35. Satoshi Yoshida, Ethan Lake, and Hayata Yamasaki, *Proof of a finite threshold for the union-find decoder*, 2026, arXiv:2602.20238.

ORCID: https://orcid.org/0000-0003-4668-3639