Statistical mechanical models for quantum codes with correlated noise

Christopher T. Chubb[1] and Steven T. Flammia[1, 2]

[1] *Centre for Engineered Quantum Systems, School of Physics, University of Sydney, Sydney NSW, Australia*
[2] *Yale Quantum Institute, Yale University, New Haven, Connecticut 06520, USA*

(Dated: October 30, 2018)

We give a broad generalisation of the mapping, originally due to Dennis, Kitaev, Landahl and Preskill, from quantum error correcting codes to statistical mechanical models. We show how the mapping can be extended to arbitrary stabiliser or subsystem codes subject to correlated Pauli noise models, including models of fault tolerance. This mapping connects the error correction threshold of the quantum code to a phase transition in the statistical mechanical model . Thus, any existing method for finding phase transitions, such as Monte Carlo simulations, can be applied to approximate the threshold of any such code, without having to perform optimal decoding. By way of example, we numerically study the threshold of the surface code under mildly correlated bit-flip noise, showing that noise with bunching correlations causes the threshold to drop to $p_{\text{corr}} = 10.04(6)\%$, from its known iid value of $p_{\text{iid}} = 10.917(3)\%$. Complementing this, we show that the mapping also allows us to utilise any algorithm which can calculate/approximate partition functions of classical statistical mechanical models to perform optimal/approximately optimal decoding. Specifically, for 2D codes subject to locally correlated noise, we give a linear-time tensor network-based algorithm for approximate optimal decoding which extends the MPS decoder of Bravyi, Suchara and Vargo.

Quantum mechanical systems are inherently sensitive to noise. The inability to completely suppress environmental noise and perform noiseless quantum operations therefore provides a significant barrier to scalable quantum information processing. To mitigate this, quantum error correcting codes [? ?] were developed that encode quantum information into a larger system whose redundant degrees of freedom provide protection from physical noise. For a given code family and noise process, quantum information can be encoded and decoded arbitrarily well below a critical noise strength known as the *threshold*. Whilst the threshold is defined with respect to the optimal decoder, it is also often also studied for the case of specific sub-optimal decoding procedures.

The most commonly studied model of noise in quantum codes is that of *iid Pauli noise*, in which each qubit is subjected to an independent, identical Pauli noise process. This is a mathematically convenient model, but it does not account for any possible correlations between errors. Whilst this model can provide a proof-of-principle that an error correction procedure can successfully withstand errors, many of the physical architectures in which we might hope to implement quantum error correction are known to experience correlated noise. Examples include proximity effects [? ?] and bosonic couplings [? ? ? ? ? ? ? ?] in solid state systems. Correlations can also arise when modelling non-Markovian noise processes [? ?]. As the threshold depends on the error model, including the presence and magnitude of correlations, taking these factors into account is important when attempting to specify physically relevant thresholds.

One of the most important correlated noise models is that of *circuit-based noise* [? ? ? ? ? ?], in which elementary gates and measurement are taken to be noisy. Even if we assume the noise introduced in each operation is independent, the operations themselves tend to propagate and accumulate errors, giving an overall correlated noise model. Importantly, there exist codes which have a threshold under this correlated noise model, which is known as the *fault-tolerant threshold* [? ?]. Moreover, it can even be shown that there exist error correction procedures which allow not only for fault tolerant storage of quantum information, but also fault tolerant quantum computation, a result known as the *quantum threshold theorem* [? ? ? ?].

In the context of correlated noise, little is known about optimal decoding or fault tolerance procedures. For this reason, most Monte Carlo estimates of thresholds for correlated noise are given with respect to sub-optimal decoders and fault tolerance schemes [? ? ? ? ? ? ? ? ? ? ?]. Indeed, to our knowledge, no optimal thresholds are known for any interesting quantum code families with a non-trivially correlated noise models.

Remarkably, there exists a method for computing code thresholds with respect to iid Pauli noise called the *statistical mechanical mapping* [?]. In this technique, a classical statistical mechanical system is constructed from a quantum code with the noise model manifesting as a quenched disorder. This mapping is designed in such a way that the thermodynamic properties of the statistical mechanical system relate to the error correction properties of the quantum code, under the optimal decoder. This method has been used [?] to compute thresholds [? ? ? ? ? ? ? ? ?], including fault-tolerance thresholds [? ? ? ? ? ?], for a wide variety of code families based on topological codes [?], i.e. codes with spatially local stabiliser generators on a lattice in Euclidean space.

Importantly, this link implies that the error correction threshold in the code manifests as a phase transition in the resulting stat mech model. This insight implies that one may bring to bear the various numerical and analytical techniques for determining the phases of stat mech systems to indirectly estimate the threshold of our code, *without* having to implement optimal decoding.

Summary of main results

In this manuscript we give an broad generalisation of the stat mech mapping to the case of correlated Pauli noise models acting on any stabiliser or subsystem quantum code. The original mapping for independent bit-flip noise [?], and subsequent generalisation for independent Pauli noise [?], works by showing that, when a special condition known as the Nishimori condition [?] holds, a certain stat mech model with quenched disorder has a partition function that maps directly onto the probability of a logical class given the syndrome of a code. Our first result is that this fundamental theorem of the original stat mech mapping continues to hold in much more generality. This implies that the threshold under a correlated noise model corresponds to a phase transition in the corresponding stat mech system, both for spatially and spatio-temporally correlated noise, including circuit noise. By performing Monte Carlo simulations of the resulting stat mech system, we then use this correspondence to approximate the threshold of the toric code subject to a correlated model of bit-flip errors. Finally we will show how this mapping can also be used to give an efficient approximation of maximum likelihood decoding, specifically we give a linear-time tensor network algorithm, generalising the decoder of Ref. [?].

The paper is organised as follows. After introducing mathematical preliminaries and definitions in Section I, we continue in Section II by reviewing the stat mech mapping in the case of independent noise. In Section III we then extend the mapping to account for noise models with spatial correlations. In Section IV we apply this mapping to correlated bit-flip noise in the toric code, and perform Monte Carlo simulations to estimate the threshold. In Section V we extend our construction to spatio-temporal correlations and circuit-based noise. In Section VI we prove that the phase boundary and the threshold coincide, and using this we show in Section VII how tensor network methods yield efficient approximations of the optimal decoder. We conclude in Section VIII, provide some background on correlated noise in Appendix A, and details of our numerical simulations in Appendix B.

I. PRELIMINARIES

Fix a local dimension $d \in \mathbb{N}$. Let ω denote the fundamental dth root of unity, and let $\mathcal{P}$ denote the set of Pauli operators *modulo phase*. We will associate a copy of the Paulis with each site i for $1 \leq i \leq n$, denoting the Paulis supported on site i by $\mathcal{P}_i$, and the set of global Paulis by $\mathcal{P}^n := \bigtimes_{i=1}^{n} \mathcal{P}_i$.

Pauli operators always commute up to phase. As such, we can capture the non-commutativity of the Paulis via the *scalar commutator* $[\![\cdot,\cdot]\!] : \mathcal{P} \times \mathcal{P} \to \mathbb{C}$, defined by the relation $AB = [\![A,B]\!]\, BA$. It can also be seen as the normalised trace of the group commutator,

$$[\![A,B]\!] := \frac{1}{d} \operatorname{Tr}\left[A,B\right], \qquad (1)$$

where $[A,B] := ABA^{-1}B^{-1}$. We can see that the scalar commutator is a well-defined function on $\mathcal{P} \times \mathcal{P}$, as the group commutator is invariant under phases, $\left[e^{i\theta}A, e^{i\phi}B\right] = [A,B]$. In general the group commutator is not multiplicative[1], but conveniently the scalar commutator of Paulis *is*, in the sense that $[\![A,BC]\!] = [\![A,B]\!] \cdot [\![A,C]\!]$. This also implies that Paulis commute under the scalar commutator, in the sense that $[\![A,BC]\!] = [\![A,CB]\!]$, which will be important for considering subsystem codes.

A. Stabiliser and Subsystem codes

The codes we will be considering in this work are Pauli stabiliser codes [? ?]. A stabiliser code is defined by a subgroup $\mathcal{S}$ of the Paulis acting on n qu$d$its. This subgroup must be Abelian and have trivial overlap with the centre, $\mathcal{S} \cap \langle \omega I \rangle = \{I\}$. This group is typically specified in the form of a set of generators $\{S_k\}_k$. The associated code space is given by all states which are stabilised by every element of the stabiliser group, i.e. all $|\psi\rangle$ such that $s|\psi\rangle = |\psi\rangle$ for all $s \in \mathcal{S}$.

The logical (Pauli) operators of such a code are given by the centraliser of the stabiliser group, that is the set of all $L \in \mathcal{P}^n$ such that $[\![s,L]\!] = +1$ for all stabilisers $s \in \mathcal{S}$. As the stabilisers act trivially on the code space, logicals which differ just by a stabiliser have identical action on the code space, and we will refer to them as *logically equivalent.*

Suppose that we start in a code state $|\psi\rangle$, and our system suffers an error $E \in \mathcal{P}^n$, leaving it in state $E|\psi\rangle$. The first step in the error correction procedure begins by measuring each of the stabiliser generators $\{S_k\}_k$. The Pauli error model has the key feature that any given error yields a deterministic outcome. Specifically, measuring the kth stabiliser generator deterministically gives the outcome $[\![S_k,E]\!]$, and therefore does not disturb the state. Importantly, this outcome does not depend on the initial code state $|\psi\rangle$, depending only on the error E. We refer to the collection of all such measurement outcomes as the *syndrome.*

A decoder for such a code is an algorithm which takes as input the syndrome, and outputs a decoding Pauli D. We then apply D^{-1} to the state, in the hope that this corrects the error. This decoding successfully restores the system back to the original code state if and only if D was logically equivalent to the true error E, i.e. $D^{-1}E \in \mathcal{S}$.

A more general notion of quantum code is that of a subsystem code [?]. Here, some fraction of the logical

[1] For a general group, $[A,BC] = [A,B]\, B\, [A,C]\, B^{-1}$.

qubits are sacrificed to become so-called gauge qubits. Gauge qubits then provide a workspace that simplifies some of the measurements, since instead of measuring a stabiliser directly one can measure combined gauge and stabiliser operators that might have lower weight. In a subsystem code, the gauge group is generated by all of the stabiliser and the gauge operators, and the stabiliser group is the centre of the gauge group (modulo phase). Thus, measuring enough gauge generators to reconstruct the stabilisers is sufficient to perform quantum error correction in a subsystem code. This is true even though general elements of the gauge group don't commute, since the stabiliser elements *do* commute by virtue of living in the centre. We refer the reader to Ref. [**?**] for a more detailed discussion of subsystem codes.

In what follows, we never use the fact that the stabiliser generators commute. We only use the fact that Pauli errors form an abelian error algebra, since they commute modulo the scalar commutator. That is, given two Pauli operators P and Q, the accumulated error on a state is the same if one applies PQ or if one applies QP. Because of this fact, everything that we derive below also applies to subsystem codes where gauge generators are used in place of stabiliser generators. With the exception of Section V, we will use the simpler language of stabiliser codes throughout, but with the understanding that the results can easily be generalised to subsystem codes.

II. THE STATISTICAL MECHANICAL MAPPING

Before considering correlated noise, we start by reviewing the case of independent noise, as first considered in Ref. [**?**]. Although this material is review, our goal is to write the derivation of the independent case in such a way that the correlated case falls out as naturally as possible.

For notational convenience, we will restrict out attention to stabiliser codes for this section, but we note that the below construction can be naturally extended to subsystem codes by replacing stabiliser generators by gauge generators.

Consider a code given by a set of stabiliser generators $\{S_k\}_k$. In this section we will be considering an independent Pauli noise model: let $\{p_i\}_i$ be probability distributions on $\mathcal{P}_i$ which describe the probability of a Pauli error (independently) occurring at each site i. The probability of an overall error $E \in \mathcal{P}^n$ is therefore given by

$$\Pr(E) = \prod_{i=1}^{n} p_i(E_i), \tag{2}$$

where E_i denotes the action of E within $\mathcal{P}_i$.

A. Statistical mechanical model

We now want to develop a (classical) spin model whose statistical mechanical properties capture the error correction properties of our quantum code, in such a way that that the threshold in the latter naturally corresponds to a phase transition in the former.

The state space of this model will correspond to the stabiliser group, with the noise model determining the interactions. Specifically, associate a classical spin degree of freedom c_k with each stabiliser generator S_k. We will consider each degree of freedom as a member of $\mathbb{Z}_d$, and so our full state space[2] is given by $\times_k \mathbb{Z}_d$.

The family of Hamiltonians we will consider is defined as follows.

Definition 1 (Stat mech Hamiltonian: *independent noise*)**.** For a Pauli $E \in \mathcal{P}^{\times n}$, and coupling strengths $\{J_i : \mathcal{P}_i \to \mathbb{R}\}_i$, our Hamiltonian H_E is defined as

$$H_E(\vec{c}) = -\sum_{i,\sigma\in\mathcal{P}_i} J_i(\sigma)\, [\![\sigma, E]\!] \prod_k [\![\sigma, S_k]\!]^{c_k}, \tag{3}$$

for any state $\vec{c} \in \times_k \mathbb{Z}_d$, where E forms a (quenched) disorder parameter. Here the sum is over all qudits i and all elements σ in the local Pauli group $\mathcal{P}_i$.

For algebraic manipulation, it will be convenient to note that the above Hamiltonian can also be written as

$$H_E(\vec{c}) = -\sum_{i,\sigma\in\mathcal{P}_i} J_i(\sigma) \left[\!\!\left[\sigma, E\prod_k S_k^{c_k}\right]\!\!\right], \tag{4}$$

due to the multiplicativity of the scalar commutator.

Physically, these Hamiltonians correspond to random bond Ising-type models for $d = 2$. This can be seen by putting the Hamiltonian in terms of the degrees of freedom $s_k := (-1)^{c_k}$,

$$H_E(\vec{s}) = -\sum_{i,\sigma\in\mathcal{P}_i} \overbrace{J_i(\sigma)}^{\text{Strength}}\ \overbrace{[\![\sigma, E]\!]}^{\text{Disorder}}\ \overbrace{\prod_{k:[\![\sigma,S_k]\!]=-1} s_k}^{\text{Interactions}}\,. \tag{5}$$

Similarly for $d > 2$ this can be seen as a type of random-bond vector Potts model (or clock model).

Readers familiar with the prior work on stat mech Hamiltonians (beginning with [**?**]) might be puzzled as to why we write such a convoluted form of H_E for our definition rather than the straightforward Ising-type model in Equation (5). The answer is that, as we will

[2] In the case where d is composite, and some of the generators S_k are not of maximal order, this will actually cause an overcounting in Theorem 1. Each degree of freedom should take on a number of states equal to the order of S_k. For notational convenience we will assume all of our generators S_k are of maximal order, as is always the case for prime d.

see below, this formulation gives the simplest path to generalising these results to the case of correlated noise.

We note that if our stabiliser generators are local, in the sense that each site is only acted upon non-trivially by a finite number of stabiliser generators, then this Hamiltonian is also local, in the sense that each interaction only touches a finite number of sites.

In this construction, each of our interactions corresponded to a site i and a $\sigma \in \mathcal{P}_i$, i.e. a single-site Pauli. As we will see in Section III, this is intimately linked with the restriction to independent noise, and that including larger range interactions will allow us to account for correlated error models.

Before considering specific coupling strengths, we note that this model has been constructed such that it is symmetric under multiplying the disorder E by a stabiliser generator S_k and adding to the corresponding degree of freedom c_k. Specifically,

$$H_{ES_k}(\vec{c}) = H_E(\vec{c} + \hat{k}). \tag{6}$$

This can be seen by noting that our Hamiltonian only depends on each c_k via $ES_k^{c_k}$, which itself has this symmetry.

B. Nishimori condition

We now want to consider the coupling strengths required such that the above model reproduces the statistical properties of our quantum code and noise model. Specifically we want to choose these couplings such that

$$Z_E = \Pr(\overline{E}), \tag{7}$$

where Z_E is the partition function of the Hamiltonian H_E with quenched disorder E, $\Pr(\overline{E})$ denotes the probability of a logical class of errors, and $\overline{E}$ denotes the set of errors which are logically equivalent to E. A sufficient condition for this is given by the *Nishimori conditions*.

Definition 2 (Nishimori conditions: *independent noise*)**.** An inverse temperature β and coupling strengths $\{J_i\}_i$ satisfy the *Nishmori conditions* with respect to distributions $\{p_i\}_i$ if

$$\beta J_i(\sigma) = \frac{1}{|\mathcal{P}|} \sum_{\tau \in \mathcal{P}_i} \log p_i(\tau) \left[\!\left[\sigma, \tau^{-1} \right]\!\right], \tag{8}$$

holds for all i and $\sigma \in \mathcal{P}_i$.

Note that Equation (8) is not defined if any of the $p_i(\tau)$ are exactly zero. However, in those cases, we can formally manipulate $\log 0$ together with the convention that $\exp(\log 0) = 0$ and obtain sensible answers. A more rigorous treatment taking limits is certainly possible, but would obscure the thrust of the argument, so we neglect these details.

This now allows us to prove the critical property of our stat mech mapping, which we refer to as the *fundamental theorem.*

Theorem 1 (Fundamental theorem of stat mech mapping: *independent noise*)**.** *Given the stat mech Hamiltonian* (*Equation* (3)) *associated to a noise model that satisfies the Nishimori condition* (*Equation* (8))*, the probability of a logical class of errors is equal to the corresponding partition function,*

$$Z_E = \Pr(\bar{E}). \tag{9}$$

Proof. First we note that the Nishimori condition takes the form of an (inverse) Fourier transform of the log-probabilities $\log p_i$ with respect to the Pauli group. Using the orthogonality condition $\sum_\sigma [\![\sigma, \rho]\!] = |\mathcal{P}|\delta_{\rho,I}$, we can see that this is equivalent to the requirement that the Fourier transform of βJ_i is $\log p_i$, specifically

$$\sum_{\sigma \in \mathcal{P}_i} \beta J_i(\sigma) [\![\sigma, E]\!] = \log p_i(E_i). \tag{10}$$

Summing over sites i, we see that this means the Gibbs weight in the all-zero state reproduces the error probability,

$$e^{-\beta H_E(\vec{0})} = \exp\left(\sum_{i, \sigma \in \mathcal{P}_i} \beta J_i(\sigma) [\![\sigma, E]\!] \right) \tag{11a}$$

$$= \exp\left(\sum_i \log p_i(E_i) \right) \tag{11b}$$

$$= \prod_i p_i(E_i) = \Pr(E). \tag{11c}$$

Lastly, we can use the symmetry property to write the partition function as a sum over logically equivalent disorders, giving the probability of logical class as required,

$$Z_E = \sum_{\vec{c}} e^{-\beta H_E(\vec{c})} = \sum_{S \in \mathcal{S}} e^{-\beta H_{ES}(\vec{0})} = \Pr(\overline{E}). \tag{12}$$

□

This correspondence between logical class probabilities and partition functions suggests that the regimes in which error correction is and is not possible in our code correspond naturally to phases in our stat mech model. We will show this explicitly in Section VI, and see that the error correcting threshold manifests as a quenched phase transition.

In Section VII we will also see how this correspondence, together with tensor network methods for approximating partition functions, can be used to construct a family of efficient tensor network algorithms which approximate maximum likelihood decoding.

C. Example: Toric code

We now consider several examples of the stat mech systems corresponding to the toric code with iid noise.

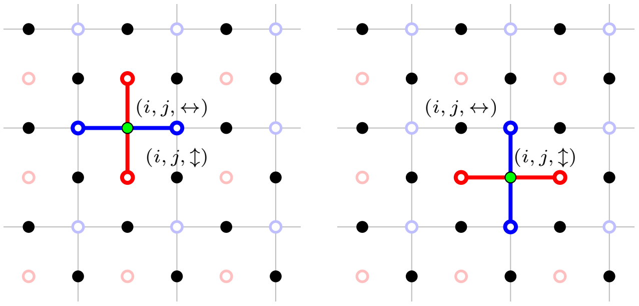

FIG. 1. Stat mech mapping of toric code with iid noise. Solid circles indicate qubits, and hollow circles indicate the spins of the stat mech model. The lines connecting the spins to the highlighted green qubit indicate the spins which are coupled by the interactions corresponding to that qubit. The interaction terms corresponding to a qubit on a horizontal edge $(i,j,\leftrightarrow)$ give a horizontal Ising coupling on the X-sublattice, a vertical Ising coupling on the Z-sublattice, and a four-body term coupling the two sub-lattices. The vertical qubit $(i,j,\updownarrow)$ gives rotated versions of these couplings. The full Hamiltonian is given in Equations (13) and (14).

Let p be the marginal distribution of errors on a single qubit. Given that our stabilisers split into X-type stars and Z-type plaquettes, we will divide the spins into two corresponding sub-lattices, denoting them $\{s^X_k\}_k$, the X-sublattice, or $\{s^Z_k\}_k$, the Z-sublattice, respectively. Applying our mapping to this code gives the Hamiltonian

$$\begin{aligned}H_E = -\sum_e \Big(& J(I) + J(X)\,[\![X, E_e]\!] \prod_{\partial f \ni e} s^Z_f \\ & + J(Z)\,[\![Z, E_e]\!] \prod_{v \in \partial e} s^X_v \\ & + J(Y)\,[\![Y, E_e]\!] \prod_{\partial f \ni e} s^Z_f \prod_{v \in \partial e} s^X_v \Big),\end{aligned} \tag{13}$$

where v, e, f denote vertices, edges, and faces in the lattice, ∂e denote the vertices surrounding an edge, and ∂f the edges surrounding a face. We note that the above Hamiltonian is not only valid for the toric code, but in fact *any* homology code.

Writing this Hamiltonian out more explicitly in Cartesian coordinates, as indicated in Figure 1, we see that our model corresponds to two copies of the 2D random bond Ising model on each sublattice, with a four-body coupling between them, specifically

$$\begin{aligned}H_E = -\sum_{i,j} \Big(& 2J(I) + J(X)\,[\![X, E_{i,j,\leftrightarrow}]\!]\, s^Z_{i,j} s^Z_{i,j+1} \\ & + J(X)\,[\![X, E_{i,j,\updownarrow}]\!]\, s^Z_{i,j} s^Z_{i+1,j} \\ & + J(Z)\,[\![Z, E_{i,j,\leftrightarrow}]\!]\, s^X_{i,j} s^X_{i-1,j} \\ & + J(Z)\,[\![Z, E_{i,j,\updownarrow}]\!]\, s^X_{i,j} s^X_{i,j-1} \\ & + J(Y)\,[\![Y, E_{i,j,\leftrightarrow}]\!]\, s^Z_{i,j} s^Z_{i,j+1} s^X_{i,j} s^X_{i-1,j} \\ & + J(Y)\,[\![Y, E_{i,j,\updownarrow}]\!]\, s^Z_{i,j} s^Z_{i+1,j} s^X_{i,j} s^X_{i,j-1} \Big),\end{aligned} \tag{14}$$

where $E_{i,j,\leftrightarrow}$ and $E_{i,j,\updownarrow}$ indicate the action of the error E on the (i,j)th horizontal and vertical qubits (see Figure 1). The terms inside the parentheses are the interactions corresponding to the qubits labelled $(i,j,\leftrightarrow)$ and $(i,j,\updownarrow)$ as indicated in Figure 1.

The Nishimori conditions give coupling strengths of the form

$$J(I) = \frac{1}{4\beta} \log p(I)p(X)p(Y)p(Z), \tag{15a}$$

$$J(X) = \frac{1}{4\beta} \log \frac{p(I)p(X)}{p(Y)p(Z)}, \tag{15b}$$

$$J(Z) = \frac{1}{4\beta} \log \frac{p(I)p(Z)}{p(X)p(Y)}, \tag{15c}$$

$$J(Y) = \frac{1}{4\beta} \log \frac{p(I)p(Y)}{p(X)p(Z)}. \tag{15d}$$

We now consider this model for specific iid error models of interest.

1. Depolarising noise (Bombin et.al.)

If we consider the depolarising channel, with $p(I) = 1 - p$ and $p(X) = p(Y) = p(Z) = p/3$, then our model reduces to

$$\begin{aligned}H_E = -\sum_{i,j} \Big(& -K + J\,[\![X, E_{i,j,\leftrightarrow}]\!]\, s^Z_{i,j} s^Z_{i,j+1} \\ & + J\,[\![X, E_{i,j,\updownarrow}]\!]\, s^Z_{i,j} s^Z_{i+1,j} \\ & + J\,[\![Z, E_{i,j,\leftrightarrow}]\!]\, s^X_{i,j} s^X_{i-1,j} \\ & + J\,[\![Z, E_{i,j,\updownarrow}]\!]\, s^X_{i,j} s^X_{i,j-1} \\ & + J\,[\![Y, E_{i,j,\leftrightarrow}]\!]\, s^Z_{i,j} s^Z_{i,j+1} s^X_{i,j} s^X_{i-1,j} \\ & + J\,[\![Y, E_{i,j,\updownarrow}]\!]\, s^Z_{i,j} s^Z_{i+1,j} s^X_{i,j} s^X_{i,j-1} \Big),\end{aligned} \tag{16}$$

with Nishimori conditions

$$\beta K = \frac{1}{2} \log \frac{27}{p^3(1-p)}\,, \tag{17a}$$

$$\beta J = \frac{1}{4} \log \frac{3(1-p)}{p}\,. \tag{17b}$$

This corresponds to a disordered variant of the eight-vertex model [? ? ?], which was considered in Ref. [?].

2. Independent X and Z

Consider now a model with independent X and Z errors. Specifically let p_X, p_Z denote the probability of each *generator*, such that

$$p(I) = (1 - p_X)(1 - p_Z), \tag{18a}$$

$$p(X) = p_X(1 - p_Z), \tag{18b}$$

$$p(Z) = (1 - p_X)p_Z, \tag{18c}$$

$$p(Y) = p_X p_Z. \tag{18d}$$

Importantly, this means that $p(X)p(Z) = p(I)p(Y)$, and therefore $J(Y) = 0$. This has the effect of decoupling the two sub-lattices into two non-interacting random-bond Ising models, such that the Hamiltonian can be decomposed $H_E = H_E^X + H_E^Z$, where

$$\begin{aligned} H_E^X &= -\sum_{i,j}\Big(-K^X + J^X \, [\![Z, E_{i,j,\leftrightarrow}]\!] \, s^X_{i,j}s^X_{i-1,j} \qquad (19a) \\ &\qquad + J^X \, [\![Z, E_{i,j,\updownarrow}]\!] \, s^X_{i,j}s^X_{i,j-1}\Big), \\ H_E^Z &= -\sum_{i,j}\Big(-K^Z + J^Z \, [\![X, E_{i,j,\leftrightarrow}]\!] \, s^Z_{i,j}s^Z_{i,j+1} \qquad (19b) \\ &\qquad + J^Z \, [\![X, E_{i,j,\updownarrow}]\!] \, s^Z_{i,j}s^Z_{i+1,j}\Big), \end{aligned}$$

and

$$K^A = \frac{1}{2\beta}\log\frac{1}{p_A(1-p_A)}\,, \qquad (20a)$$

$$J^A = \frac{1}{2\beta}\log\frac{1-p_A}{p_A}\,, \qquad (20b)$$

for $A \in \{X, Z\}$. This corresponds to two decoupled copies of the random-bond Ising model, with disorder probabilities p_X and p_Z. The decoupled nature of these two models is generic for CSS codes (i.e., codes whose stabilisers split into separate X and Z type) under independent X and Z noise. A further consequence of this is that the optimal decoder can decode the X and Z errors independently.

3. Pure bit-flip noise (Dennis et.al.)

As mentioned when we introduced the Nishimori condition above, care must be taken whenever the noise model does not have full support, as our Hamiltonian becomes divergent on certain states. An important example of such a model is that of iid bit-flip noise, as was considered in the seminal paper Ref. [?]. Here we take $p(I) = 1 - p$, $p(X) = p$, and $p(Y) = p(Z) = 0$.

We know that the Nishimori condition implies

$$H_E(\vec{c}) = -\frac{1}{\beta}\log\Pr\left(E\prod_k S_k^{c_k}\right). \qquad (21)$$

Given that the probability of any error containing any non-trivial Z contribution is zero in this error model, this tells us that any state which is *not* entirely magnetised on the Z-sub-lattice ($s^Z_{i,j}s^Z_{i+1,j} = -1$ or $s^Z_{i,j}s^Z_{i,j+1} = -1$ for some i, j) has infinite energy. We can interpret this as the degrees of freedom in this sub-lattice being *frozen out*, into the one of the two entirely magnetised states. Restricting to these (degenerate) Z-magnetised states ($s^Z_{i,j}s^Z_{i+1,j} = s^Z_{i,j}s^Z_{i,j+1} = +1$ for all i, j), our Hamiltonian therefore reduces to a single copy of the random bond Ising model

$$\begin{aligned} H_E = -\sum_{i,j}\Big(&-2K + J\,[\![Z, E_{i,j,\leftrightarrow}]\!]\, s^X_{i,j}s^X_{i-1,j} \qquad (22) \\ &+ J\,[\![Z, E_{i,j,\updownarrow}]\!]\, s^X_{i,j}s^X_{i,j-1}\Big), \end{aligned}$$

where

$$K = \frac{1}{2\beta}\log\frac{1}{p(1-p)}, \qquad (23a)$$

$$J = \frac{1}{2\beta}\log\frac{1-p}{p}. \qquad (23b)$$

We can see now that J and K are both finite for any $p \in (0, 1)$, meaning that our Hamiltonian is no longer divergent on the remaining degrees of freedom. Moreover, this is the Hamiltonian that was considered in Ref. [?]. This shows that, taking appropriate care, errors models without full support can also be considered using this construction, after considering frozen out degrees of freedom.

III. CORRELATED NOISE

Now that we have reviewed the stat mech construction for *independent* Pauli noise, we now want to consider extending this to *correlated* noise models.

We will consider a model with local spatial correlations, which forms a natural and systematic way of relaxing the independence condition above. Specifically, we want to consider models in which errors at sufficiently distant locations are conditionally independent, but in which they may be arbitrarily correlated at short range. The conditional dependences described above are naturally represented through *probabilistic graphical models*, such as Markov random fields and Bayesian networks. Below we will focus on a simpler and more mathematically convenient model of *factored distributions*. We discuss the relationship between these models in more detail in Appendix A.

A. Factored distributions

For a global error $E \in \mathcal{P}^n$, let E_i denote the action on a single site i. Similarly, for some set of sites $R = \{i_1, i_2, \dots\}$, let E_R denote action of E on sites in R.

For the previous mapping, we leveraged the fact that independence of two random variables A and B implies their joint distribution factors, $\Pr(A, B) = \Pr(A)\cdot\Pr(B)$. We now want to consider more general distributions which can also be locally factored.

Definition 3 (Factored distribution)**.** A distribution *factors* over sets $\{R_j\}_j$ if there exists non-negative functions $\{\phi_j : \mathcal{P}_{R_j} \to \mathbb{R}^+\}_j$ such that

$$\Pr(E) = \prod_j \phi_j(E_{R_j}). \qquad (24)$$

If the regions $\{R_j\}_j$ are disjoint, then these just correspond to distributions which are independent after appropriate coarse-graining. We, however, *do not* require that these regions are disjoint, which allows us to consider genuinely correlated noise models, which remain correlated even after coarse-graining. In another extreme limit, if there is only a single region R and it contains *all* of the random variables, then we can take ϕ_R to be simply the complete joint probability distribution. This shows that the structure of the individual regions R_j can interpolate between the case of independent noise (when $R_j = \{j\}$) and a general distribution (when $R = \{1, \ldots, n\}$). Intermediate cases correspond to probability distributions with differing ranges of correlation. Thus the factored distribution formalism forms a natural ansatz for describing finite-range correlations efficiently in a probabilistic model. We refer the reader to Appendix A for more discussion of factored distributions.

B. Correlated statistical mechanical mapping

We now extend the stat mech mapping to the case of correlated noise. Specifically we will consider an error model which factors over $\{R_j\}_j$, with factors $\{\phi_j\}_j$. In the mapping for independence noise, each interaction term corresponding to a single-site Pauli. As hinted at earlier, we can account for correlated noise models by including interaction terms corresponding to multi-site σ, specifically σ living on a single region R_j. Recall that S_k are the stabilisers of the quantum code and c_k are the associated d-level classical degrees of freedom taking values in $\mathbb{Z}_d$. Then the stat mech Hamiltonian takes the following form.

Definition 4 (Stat mech Hamiltonian: *correlated noise*)**.** For a Pauli $E \in \mathcal{P}^n$, and coupling strengths $\{J_j : \mathcal{P}_{R_j} \to \mathbb{R}\}_j$, the stat mech Hamiltonian H_E is defined as

$$H_E(\vec{c}) = -\sum_{j,\sigma\in\mathcal{P}_{R_j}} J_j(\sigma) \, [\![\sigma, E]\!] \prod_k [\![\sigma, S_k]\!]^{c_k}. \tag{25}$$

We notice that we can once again use the multiplicativity of the scalar commutator to rewrite this in the more mathematically convenient form

$$H_E(\vec{c}) = -\sum_{j,\sigma\in\mathcal{P}_{R_j}} J_j(\sigma) \left[\!\!\left[\sigma, E\prod_k S_k^{c_k}\right]\!\!\right], \tag{26}$$

and see that our Hamiltonian retains the symmetry we leveraged in the independent case,

$$H_{ES_k}(\vec{c}) = H_E(\vec{c} + \hat{k}). \tag{27}$$

In order for the definition of the stat mech Hamiltonian to connect to the error probabilities of the quantum code for a given noise model, we need to find an analogue of the Nishimori conditions. Fortunately, our formulation of this condition in the independent case gives an immediate generalisation.

Definition 5 (Nishimori conditions: *correlated noise*)**.** An inverse temperature β and coupling strengths $\{J_j\}_j$ satisfy the *Nishimori conditions* with respect to factors $\{\phi_j\}_j$ if

$$\beta J_j(\sigma) = \frac{1}{|\mathcal{P}_{R_j}|} \sum_{\tau\in\mathcal{P}_{R_j}} \log\phi_j(\tau) \, [\![\sigma, \tau^{-1}]\!], \tag{28}$$

Using a proof analogous to the independent case Theorem 1, we also get a *fundamental theorem* for this correlated model.

Theorem 2 (Fundamental theorem of stat mech mapping: *correlated noise*)**.** *Given the stat mech Hamiltonian (Equation (25)) associated to a noise model that satisfies the Nishimori condition (Equation (28)), the probability of a logical class of errors is equal to the corresponding partition function,*

$$Z_E = \Pr(\bar{E}). \tag{29}$$

As with the independent case, we will see that this similarly implies that the error-correction threshold manifests as a quenched phase transition (Section VI), and that this allows us to construct efficient tensor network approximations to the maximum likelihood decoder (Section VII).

C. Noise Hamiltonian

Consider the case where each factor ϕ_j is strictly positive. Suppose we define the local Hamiltonian

$$\tilde{H}(E) := -\sum_j \log\phi_j(E), \tag{30}$$

which we refer to as the *noise Hamiltonian*. The states of this Hamiltonian are labelled by Pauli errors, elements of $\mathcal{P}^n$. The noise model then corresponds to thermal distribution of this Hamiltonian at inverse-temperature $\beta = 1$,

$$\Pr(E) = e^{-\tilde{H}(E)}. \tag{31}$$

Expressed in this way, the Nishimori condition, Definition 5, can be seen as a relationship between the stat mech Hamiltonian of Definition 4 and the noise Hamiltonian of Equation (30). Specifically, it takes the form

$$\beta H_E(\vec{c}) = \tilde{H}\left(E\prod_k S_k^{c_k}\right). \tag{32}$$

In this sense we see that, when the Nishimori condition is satisfied, the interactions within the stat mech Hamiltonian H_E naturally correspond to those in the noise Hamiltonian $\tilde{H}$, but with a change in the underlying state space.

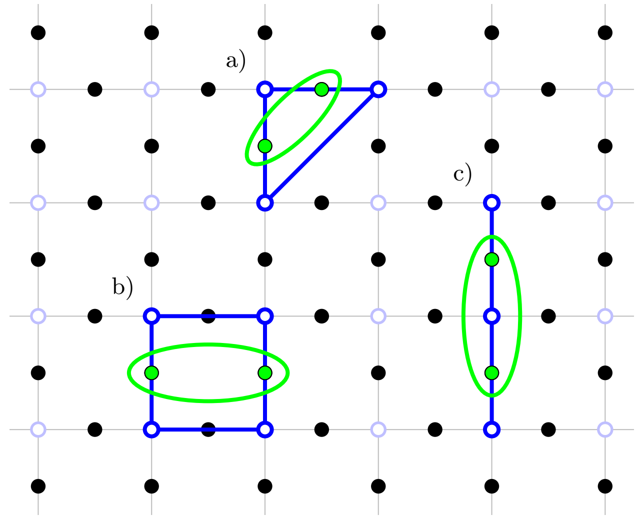

FIG. 2. Stat mech couplings induced by pairwise correlated bit-flips in the toric code. Solid circles indicate qubits, and hollow circles the spins of our stat mech model, corresponding to stabiliser generators. Green ellipses denote the type of correlations, specifically the regions over which the error model factors. These correspond to the interactions in the noise Hamiltonian Equation (33). Blue edges connecting spins indicate the couplings induced in the statistical mechanical model by such correlations. The labelled couplings a), b) and c) respectively correspond to: a) nearest-neighbour correlations induce a two-body diagonal coupling, b) next-nearest neighbour (across-plaquette) correlations induce a four-body face coupling, c) next-nearest-neighbour (across-vertex) induce two-body distance-2 couplings.

D. Example: Toric code with correlated bit flips

We now consider an example of a correlated noise model: correlated bit-flips in the toric code. Consider a noise model defined by an Ising noise Hamiltonian with coupling $\tilde{J}$ and field strength $\tilde{h}$,

$$\tilde{H} = -\sum_{e} \tilde{h}\, x_e - \sum_{e\sim e'} \tilde{J}\, x_e x_{e'}, \tag{33}$$

where we have chosen the convention that $x_e = -1$ corresponds to $E_e = X$ and $x_e = +1$ to $E_e = I$. Here $\tilde{J}$ controls the magnitude (and sign) of the correlations, with $\tilde{J} = 0$ corresponding to independent errors, $\tilde{J} > 0$ to bunching errors, and $\tilde{J} < 0$ to anti-bunching errors.

For a given error E (and the corresponding values of the spin variables x_e), the stat mech model is of the form

$$H_E = -\sum_{e} (h_e x_e) \prod_{v\in\partial e} s_v - \sum_{e\sim e'} (J x_e x_{e'}) \prod_{v\in\partial(ee')} s_v, \tag{34}$$

where $\partial(ee')$ denotes the vertices that surround either e or e', but not both. Here the variables s_v form the degrees of freedom, and the x_e form the quenched disorder. The Nishimori conditions for this model reduce to $\beta J = \tilde{J}$ and $\beta h = \tilde{h}$.

As in the independent case, we can see that the $\tilde{h}$ field term in our noise Hamiltonian has induced a 2D random-bond Ising model. The addition of the $\tilde{J}$ term has induced additional longer range couplings. The geometry of these couplings is shown is Figure 2 for the case where the noise correlations couple nearest-neighbour and next-nearest-neighbour qubits. The corresponding stat mech Hamiltonian has a similar locality, and contains at most 4-body interactions among the s_v degrees of freedom.

IV. NUMERICS

A key advantage of the stat mech mapping is that it allows us to reappropriate techniques for determining the phase diagrams of classical spin systems for approximating the thresholds of quantum codes. By way of example, we consider using Monte Carlo simulations to determine the threshold of the toric code under a correlated model of bit-flip noise.

We shall consider the model of 'across-plaquette' correlated bit-flips described in Section III D (see Figure 2b). We will restrict our attention to noise which obeys certain natural symmetries, namely, that the correlations are site-independent and symmetric between correlated errors. This is equivalent to saying that the noise Hamiltonian has symmetric, site-independent interaction terms.

Instead of expressing this model in terms of an Ising noise Hamiltonian, it will be convenient to parameterise this model in terms of the marginal error rate. Let p denote the marginal error probability on any site, i.e. $\Pr(E_e = X) = p$ on any edge e. Suppose that the error probability given that a neighbouring error has or has not occurred is $p_\pm$, so that

$$\Pr(E_e = X | E_{e'} = X) = p_+, \tag{35a}$$

$$\Pr(E_e = X | E_{e'} = I\ ) = p_-, \tag{35b}$$

where e and e' lie on opposite sides of a plaquette. For the marginal probability to be p, these probabilities are subject to the consistency condition $p = pp_+ + (1-p)p_-$.

If $p_- < p < p_+$ the errors tend to bunch together, whereas they tend to anti-bunch if $p_- > p > p_+$. A natural way of parameterising these correlations is in terms of the *correlation parameter*,

$$\eta := p_+/p_-, \tag{36}$$

where $\eta = 1$ corresponds to uncorrelated noise, $\eta > 1$ to bunched, and $\eta < 1$ to anti-bunched.

In the limit of infinite bunching ($\eta \to \infty$) the model produces exclusively logical errors, meaning that the threshold vanishes, $p_\mathrm{t} \to 0\%$. Similarly, for infinite anti-bunching ($\eta \to 0$) the model cannot produce non-trivial logical errors, and we expect that $p_\mathrm{t} \to 50\%$ [?]. We will be considering the case of mild bunching correlations, $\eta = 2$, which one would expect to lower the threshold.

Applying the stat mech mapping, we get a random-bond Ising-type model on a square lattice, containing

2-body edge terms and 4-body face terms. The Hamiltonian takes the form

$$H_E(\vec{s}) = -\sum_{e}(J_2 x_e)\prod_{v\in\partial e} s_v - \sum_{e\sim e'}(J_4 x_e x_{e'})\prod_{v\in\partial(ee')} s_v, \tag{37}$$

where $x_e = +1$ if $E_e = I$ and $x_e = -1$ if $E_e = X$, and $e \sim e'$ denotes edges lying across a plaquette from each other.

Normalising our Hamiltonian such that $J_2 \equiv 1$, our system has two Nishimori conditions. When $\eta = 2$, the first of these conditions is

$$J_4 = \frac{\log\frac{1-p}{2}}{\log 4p^2}. \tag{38}$$

We will impose this condition, and determine the phase diagram of the system in the two remaining parameters: the error probability p and temperature T. The threshold probability can be found by finding the intersection of this phase boundary with the second remaining Nishimori condition,

$$\beta_{\text{Nish}} = -\log 4p^2. \tag{39}$$

In Figure 3 we show the phase diagrams of the $\eta = 1$ uncorrelated model, and our numerical results for the $\eta = 2$ correlated model. The details of our numerical simulation are found in Appendix B. We find that that under these mild correlations the threshold drops to $p_{\mathrm{t}}(\eta = 2) = 10.04(6)\%$, from the uncorrelated threshold of $p_{\mathrm{t}}(\eta = 1) = 10.917(3)\%$ [?], confirming the earlier intuition that correlated errors will indeed reduce the threshold.

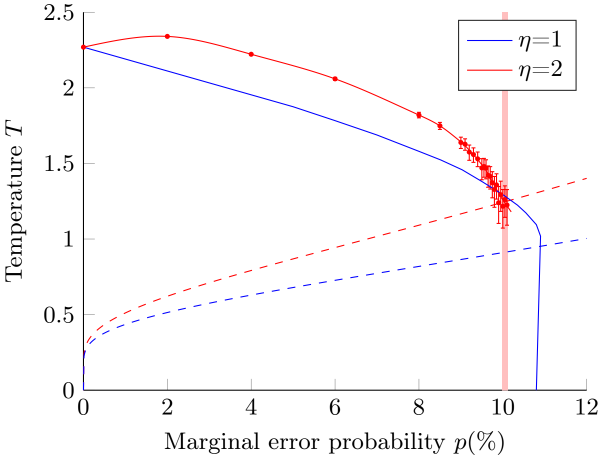

FIG. 3. Phase boundary of the stat mech models corresponding to uncorrelated ($\eta = 1$) and correlated ($\eta = 2$) bit-flip noise in the toric code. The solid lines indicate the phase boundaries of the two models, with error bar indicating statistical uncertainty in our numerical results. The $\eta = 1$ data is taken from Refs. [? ? ? ? ?], and the $\eta = 2$ comes from our numerics, as detailed in Appendix B. The dashed lines indicate the corresponding Nishimori conditions. The shaded red region indicates our estimate of $p_{\mathrm{t}} = 10.04(6)\%$ for the threshold of the correlated model.

V. SPATIO-TEMPORAL CORRELATIONS

In Section III we considered the stat mech mapping for error models which factor. This model applies in the important case of spatially correlated errors, followed by ideal measurements. We would like to generalise this construction to the case of multiple rounds of syndrome measurements, subject to *spatio-temporal* correlations in the noise. We will do this by constructing a subsystem code which, subject to purely spatially correlated noise, reproduces the measurement statistics of our original stabiliser code under a spatio-temporally correlated noise model. We refer to this construction as the *history code.* Doing this in such a way that the locality of our code and noise model are preserved, we can then apply the construction of Section III to the history code to give a corresponding stat mech model.

As well as presenting the construction of the history code, we will also discuss an important family of spatio-temporally correlated noise models which possess a correlation structure which make them amenable to this construction: circuit-based noise. In this model we consider a syndrome measurement procedure composed of faulty gates. The key distinction between this and purely spatial correlations is that errors incurred at an earlier time can be spread around by subsequent measurement circuits. Considering Clifford measurement circuits subject to Pauli noise, this gives a spatio-temporally correlated Pauli noise model. By applying the construction of Section III to the corresponding history code, we get a stat mech model with a phase transition corresponding to the *fault-tolerant threshold* of the original code.

A. History code

The idea behind the history code is to convert time into an additional spatial dimension, allowing us to naturally convert spatio-temporal correlations into purely spatial correlations, albeit living in one dimension higher. Specifically, for each site i in our original code, and measurement round t, the history code will have a corresponding site (i, t).

1. Noise model

The error models we shall consider are those which factor in a spatio-temporal sense. Specifically, let $E^{(t)}$ denote the error to which our code is subjected *prior* to the tth round of measurements, and $E := (E^{(1)}, \ldots, E^{(T)})$ denote the *error history.* We will consider models in which the distribution of error histories factors, i.e.

$$\Pr(E) = \prod_l \phi_l(E), \tag{40}$$

where the supports of ϕ_l are local in both space *and* time.

As the name may suggest, we will take our error model on the history code to be given by the error *histories* of our noise with spatio-temporal correlations. Specifically let $\hat{E}^{(t)}$ denote the action of $E^{(t)}$ on the tth layer of the history code, and correspondingly let the action of an error history be denoted $\hat{E} := \bigotimes \hat{E}^{(t)}$. Given this action, the distribution of errors is simply given by the original error model, and so

$$\Pr(\hat{E}) = \prod_l \phi_l(\hat{E}). \tag{41}$$

As $\hat{E}$ can be interpreted as a purely spatial error in the history code, instead of an error history, this now constitutes a noise model with purely spatial correlations.

2. Gauge generators

Let $M_j^{(t)}$ denote the jth Pauli measurement occurring in the tth round of syndrome measurements. For error correction to remain well-defined, the final measurement round will need to consist of ideal measurements of all the stabiliser generators[3]. As such if we let T be the number of measurement rounds, this assumption means that $\left\{M_j^{(T)}\right\}_j = \{S_k\}_k$. We will require the measurements within each round have disjoint support with the exception of the special final round.

Let $\hat{M}_j^{(t)}$ denote the action of $M_j^{(t)}$ on the corresponding layer of the history code. By construction, we have that the measurement statistics of these operators reproduce those of the measurements in the original code, as

$$\left[\!\left[\hat{M}_j^{(t)}, \hat{E}\right]\!\right] = \left[\!\left[M_j^{(t)}, E^{(t)}\right]\!\right] \tag{42}$$

for any j and t.

We now want to consider the history code itself, which takes the form of a subsystem code. To find the gauge generators, we can consider when two error histories are logically equivalent. As it is only the final accumulated error which determines whether or not a logical error has occurred, any two error histories are logically equivalent if and only if they possess the same syndrome, and do not differ by a non-trivial logical operator in the original code on the final time-slice. This implies that the gauge group is generated locally, and that these generators come in two forms: for $t < T$, the generators are given by the generators of the centraliser of $\hat{M}_j^{(t)}$ on the support of $\hat{M}_j^{(t)}$, and for $t = T$ they are just given by the measurement operators themselves $\{\hat{M}_j^{(T)}\}_j$.

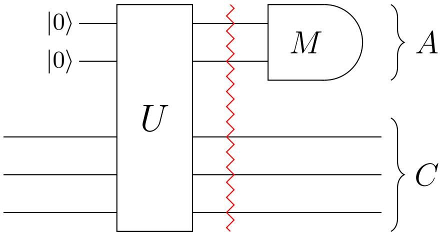

FIG. 4. The form of measurement circuits we shall be considering. For each stabiliser generator S, several ancillae are prepared in the state $|0\rangle$, an entangling Clifford U is performed, and then a Pauli measurement M is performed on the ancillae. Sets C and A are used to denote the code and ancillary qubits involved in this measurement.

B. Circuit-based model

We now turn our attention to a spatio-temporally correlated noise model which arises in the study of fault-tolerant error correction—circuit-based noise. In this noise model we explicitly take into account the circuits used to implement our syndrome measurements, and consider the faults within the constituent gates.

1. Measurement circuits

We will consider measurement circuits of the form shown in Figure 4: several ancillae are prepared in the state $|0\rangle$, a Clifford gate is applied to these and the qudits to be measured, and finally a Pauli measurement is performed on the ancillae. For the measurement circuit corresponding to $S_j^{(t)}$, we let $U_j^{(t)}$ denote the Clifford gate applied and $M_j^{(t)}$ the Pauli being measured. We denote the set of code qudits and ancillae involved in this syndrome measurement by

$$C_j^{(t)} := \operatorname{supp}\left(S_j^{(t)}\right), \tag{43a}$$

$$A_j^{(t)} := \operatorname{supp}\left(M_j^{(t)}\right), \tag{43b}$$

and let $R_j^{(t)} := C_j^{(t)} \cup A_j^{(t)}$ denote the full set of qudits involved, such that $\operatorname{supp}\left(U_j^{(t)}\right) = R_j^{(t)}$. The assumption that our measurements in each round are non-overlapping is equivalent $\left\{C_j^{(t)}\right\}_j$ being non-overlapping for each t.

[3] This is required to preclude errors occurring after measurements, which are clearly uncorrectable. This assumption is equivalent to only requiring that an error correction procedure correct errors prior to the final measurement round for it to be deemed successful.

As noted earlier, the final error correction round will correspond to an ideal measurement of all of the stabiliser generators. As such we will take $\{M_j^{(T)}\}_j = \{S_k\}_k$.

2. Noise model

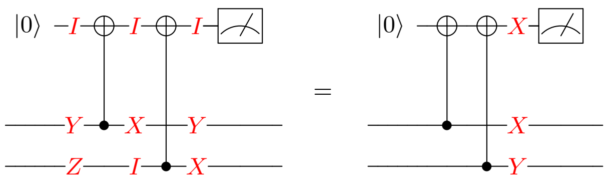

FIG. 5. General Pauli errors being pushed through a measurement circuit. Notice that the Y error on the upper code qubit spreads onto the measurement qudit.

Consider now performing the above syndrome measurement circuits, subject to Pauli noise. For simplicity we will assume that the errors experienced within different measurement circuits are independent[4], but will allow arbitrary correlations within each circuit. Without loss of generality, we can push the error through the circuit, giving an effective error in $\mathcal{P}_{R_j^{(t)}}$ acting *after* the application of $U_j^{(t)}$ (see Figure 5). Since we have assumed that our circuits are Clifford, this updated Pauli error can be computed efficiently. Let $p_j^{(t)}$ be the distribution on $\mathcal{P}_{R_j^{(t)}}$ of these pushed-through errors.

There are two error sources which contribute to $E^{(t)}$: errors which accrued from previous syndrome measurement rounds, and those which occurred in the tth round itself. This means that our errors satisfy the recurrence relation

$$E^{(t)} = \epsilon^{(t)} \cdot U_t E^{(t-1)} U_t^\dagger, \tag{44}$$

where $\epsilon^{(t)}$ denotes the freshly introduced errors which occurred in the tth round, and $U^{(t)}$ to the unitary action of the syndrome measurement circuits in the tth round.

Decomposing this further into the individual measurement circuits, we have

$$U^{(t)} = \bigotimes_j U_j^{(t)} \quad \text{and} \quad \epsilon^{(t)} = \bigotimes_j \epsilon_j^{(t)}, \tag{45}$$

where $\epsilon_j^{(t)}$ and $U_j^{(t)}$ are supported solely on $R_j^{(t)}$, and $\epsilon_j^{(t)}$ is distributed according to $p_j^{(t)}$. As our code is assumed to be error-free prior to the beginning of the syndrome measurements, we have the initial condition $E^{(0)} = I$, and so this recurrence relation entirely describes the error model.

3. Factorising circuit noise

We start by noticing that in Equation (44) the errors at time t are entirely determined by the errors at time $t-1$, and the newly incurred errors $\epsilon^{(t)}$. As we have assumed that these new errors incurred in each round of syndrome measurements are independent, our error model is therefore Markovian, allowing us to factorise our noise model in the temporal direction

$$\Pr(E^{(1)}, \ldots, E^{(T)}) = \prod_t \Pr(E^{(t)} | E^{(t-1)}). \tag{46}$$

Next we can decompose this into the individual measurement circuits using Equation (45). Specifically we know that the individual new errors $\epsilon_j^{(t)}$ are independently distributed according to $p_j^{(t)}$, which allows us to express the conditional probability as

$$\Pr\left(E^{(t)} \middle| E^{(t-1)}\right) = \prod_j p_j^{(t)}\left(\epsilon_j^{(t)}\right). \tag{47}$$

We now want to put this expression back purely in terms of the accumulated errors $E^{(t)}$, instead of the fresh errors $\epsilon^{(t)}$. Inverting the recurrence relation Equation (44), we have that

$$\epsilon^{(t)} = E^{(t)} \cdot U^{(t)} E^{(t-1)\dagger} U^{(t)\dagger}. \tag{48}$$

Recalling that $\epsilon^{(t)}$ and $U^{(t)}$ factor across $\{R_{j,t}\}_j$ (see Equation (45)), this implies that

$$\epsilon_j^{(t)} = \left[E^{(t)}\right]_{R_j^{(t)}} \cdot U_j^{(t)} \left[E^{(t-1)\dagger}\right]_{C_j^{(t)}} U_j^{(t)\dagger}, \tag{49}$$

where $[P]_R$ denotes the restriction of a Pauli P to a region R. Putting this together with Equations (46) and (47), we have that the distribution on error histories $\Pr(E^{(1)}, \ldots, E^{(T)})$ factors both in space and time, allowing us to apply the history code construction of Section V A.

VI. ERROR CORRECTION AS A STATISTICAL MECHANICAL PHASE

The fundamental theorem of the stat mech mapping, Theorem 2, links the equilibrium thermodynamic properties of our stat mech model and the error correction properties of our code via the disordered partition function and the error class probabilities. In fact, a much stronger connection is true, as has already been noted by previous authors in the case of independent noise, beginning with Dennis *et al.* [?]. Previous work has shown that the regions in parameter space for independent noise in which a code can and cannot be decoded are phases in the associated stat mech system [? ? ?]. This implies that the error correction threshold of our code manifests as a phase transition in the associated

[4] More generally any model which produces factored spatio-temporally correlated noise could be used.

stat mech model. Moreover, there is an explicit order parameter that captures this phase transition, although unfortunately it actually a "disorder" parameter in that it involves the amount of disorder in the stat mech model.

In this section, we review and extend this connection to the general case of correlated noise. We define appropriate notions of "above" and "below" the threshold and prove that the disorder parameter exhibits non-analytic behaviour in the sense that it converges or diverges if and only if the code is above or below the threshold, respectively. This shows a precise sense in which the phase boundary of the stat mech model is exactly the threshold of the corresponding code.

In Section IV we saw how, using this correspondence, numerical techniques to analyse phase transition in stat mech systems can be used to give approximations of code thresholds. We also note that this correspondence also opens the door to applying analytic techniques for studying phase transitions, such as the duality method [**? ? ?**].

A. Maximum probability and maximum likelihood decoding

A decoder is an algorithm which takes as input the syndrome and attempts to estimate which error occurred. A natural starting point would be a decoder which simply outputs the most likely error in this error model among those which are consistent with the observed syndrome. We will refer to this as the *maximum probability* (MP) decoder.

For degenerate codes, where several logically equivalent errors can have the same syndrome, the MP decoder is generally sub-optimal. This stems from the fact that successful decoding does not require the decoder to output precisely the error which occurred, but just an error which is logically equivalent. As such, the ideal decoding will not correspond to the single error with the highest probability, but the *error class* with the highest probability. The optimal decoder therefore outputs an error from this most likely class, and we will refer to this as the *maximum likelihood* (ML) decoder.

To see the sub-optimality of the maximum probability decoder, consider the following example. Suppose a syndrome s is measured which is consistent with three errors, E_1, E_2, E_3, which occur with probabilities

$$\Pr(E_1) = 4\%, \quad \Pr(E_2) = 3\%, \quad \Pr(E_3) = 3\%. \tag{50}$$

If E_2 and E_3 are logically equivalent, then the error *class* $\overline{E_2}$ is more likely than $\overline{E_1}$, even though E_1 itself is the single most likely error. In this case we can see that the MP decoder will be sub-optimal, and the conditional success probabilities for the two decoders are

$$\Pr\left(\text{MPD success}|s\right) = 40\%, \tag{51a}$$
$$\Pr\left(\text{MLD success}|s\right) = 60\%. \tag{51b}$$

For each syndrome s, let C_s denote an arbitrary Pauli with syndrome s. Similarly, for each logical l, let L_l denote a Pauli corresponding to logical l. The sets $\{\overline{C_s L_l}\}_{s,l}$ correspond to the logical error classes, and form a partition of the Paulis $\mathcal{P}^n$.

For a decoder to always return the code back to the code space, we require that when a syndrome s is input, it always returns an error with this same syndrome s. As such, without loss of generality, a decoder can be taken to be of the form $s \mapsto C_s L_{\delta(s)}$, where δ is a map from syndromes to logicals. In this notation, the maximum probability decoder takes the form

$$\delta_{\mathrm{MP}}(s) := \arg\max_{l} \max_{E\in\overline{C_s L_l}} \Pr(E), \tag{52}$$

and the maximum likelihood decoder the form

$$\delta_{\mathrm{ML}}(s) := \arg\max_{l} \Pr(\overline{C_s L_l}). \tag{53}$$

B. Minimum free energy and minimum energy decoding

Let H_E denote the stat mech model corresponding to our code and error model (see Sections II, III and V), which satisfies the Nishimori conditions at inverse temperature β_{N}.

Similar to the maximum probability and likelihood decoders defined above, consider the β-*minimum free energy* (β-MFE) decoder, which is given by minimising the free energy at inverse temperature β,

$$\delta_{\beta\text{-MFE}}(s) := \arg\min_{l} F_{C_s L_l}(\beta), \tag{54}$$

where $F_E(\beta) := -\frac{1}{\beta}\ln Z_E(\beta)$.

When satisfying the Nishimori conditions, $\beta = \beta_{\mathrm{N}}$, we can apply the fundamental theorem of the stat mech mapping (Theorem 2), to give

$$F_E(\beta_{\mathrm{N}}) = -\frac{1}{\beta_{\mathrm{N}}} \ln \Pr(\overline{E}). \tag{55}$$

As this is a monotonically decreasing function of the error class probability, this tells us that the β_{N}-MFE decoder is precisely the maximum likelihood decoder,

$$\delta_{\beta_{\mathrm{N}}\text{-MFE}} \equiv \delta_{\mathrm{ML}}. \tag{56}$$

Similarly, if we take $\beta \to \infty$, then the free energy reduces to the minimum energy,

$$\lim_{\beta\to\infty} F_E(\beta) = \min_{\vec{s}} H_E(\vec{s}). \tag{57}$$

Theorem 2 gives us that the energies of our stat mech model correspond to error probabilities, and so

$$\lim_{\beta\to\infty} F_E(\beta) = \min_{E'\in\overline{E}} -\frac{1}{\beta_{\mathrm{N}}} \ln \Pr(E'). \tag{58}$$

As this is a monotonically decreasing function of the error probabilities, this tells us that the ∞-MFE decoder—which one could refer to as the *minimum energy* decoder—is precisely the MP decoder,

$$\delta_{\infty\text{-MFE}} \equiv \delta_{\text{MP}}. \tag{59}$$

This reduction to MFE decoding implies that any of the plentiful methods for approximating partition functions [? ?], or free energy differences [?] can be used to implement approximate ML decoding. In Section VII we will expand upon this connection, giving a tensor network algorithm which approximates ML decoding.

C. The error correction threshold as a phase transition

An important way of quantifying the resilience of a quantum error correction procedure to an error model is the *error threshold*. Specifically, consider a family of quantum codes, with a logical algebra of dimension K, and an error model which depends on a parameter $\theta \geq 0$. We define the notion of threshold as follows. A code family has a *threshold* if there exists a $\theta_t > 0$ such that the asymptotic success probability is maximal for $\theta < \theta_t$, and minimal for $\theta > \theta_t$, i.e.

$$\lim_{n\to\infty} \Pr(\text{Decoder success}) = \begin{cases} 1 & \text{if } \theta < \theta_t, \\ 1/K & \text{if } \theta > \theta_t. \end{cases} \tag{60}$$

These regimes we refer to as being *below threshold* and *above threshold* respectively.

Clearly this is a rather strong notion of threshold. One can imagine codes or noise models where there are intermediate regimes where some, but not all, of the codewords are decodable. As an example, having asymmetric noise between independent X and Z error rates in a surface code will generate such a gap. Therefore, not all codes necessarily possess a threshold in this sense. However, the presence of a threshold in our sense is often taken as a desirable property in engineered quantum codes.

We also use the notion of threshold with respect to a particular choice of decoder, not just the optimal decoder. However, unless otherwise specified, we will be default to considering the threshold with respect to the optimal decoder, and refer to this as *the* threshold.

Given that the threshold corresponds to a dramatic jump in the success probability of decoding, one might naturally suspect that this is precipitated by a corresponding dramatic jump in the error class probabilities. As these probabilities equal the partition function of our stat mech models, this would also suggest a (disordered) phase transition. To see that this is indeed the case, we will consider the disorder parameter given by the free energy cost of a non-trivial logical operator L_m,

$$\begin{aligned} \Delta_m(E) &:= F_{EL_m}(\beta_{\text{N}}) - F_E(\beta_{\text{N}}) && \text{(61a)} \\ &= -\frac{1}{\beta_{\text{N}}}\log Z_{EL_m} + \frac{1}{\beta_{\text{N}}}\log Z_E. && \text{(61b)} \end{aligned}$$

We note that for topological codes which have string-like (or sheet-like in higher dimensions) logical operators, this corresponds to the free energy cost of a domain wall, as noted by Dennis *et al.* [?]. Let θ_p denote the point at which the stat mech model undergoes a phase transition corresponding to this parameter.

We now want to show that the threshold corresponds to a phase transition in our stat mech model, $\theta_t = \theta_p$. As a first step, we consider the quenched average, $\Delta_m := \langle \Delta_m(E) \rangle_E$. We would expect that $\Delta_m \to \infty$ below threshold and $\Delta_m \to 0$ above. The former limit was shown in Ref. [?], which implies that $\theta_p > \theta_t$. We present a simplified proof below.

Lemma 1 (Divergence in mean)**.** *If the code is below threshold $\theta < \theta_t$, then the quenched average free energy cost diverges,*

$$\lim_{n\to\infty} \Delta_m = \infty. \tag{62}$$

Proof. By recalling that $Z_E = \Pr(\overline{E})$, we can see that $\Delta_m(E)$ only depends on E up to logical equivalence. A convenient set of representatives from each error class are given by the Paulis $\{D_s L_l\}_{s,l}$, where s and l correspond to the syndrome and logical degrees of freedom, and $\{D_s\}_s$ are the decoding Paulis associated with the MLD. As shorthand, we will denote $\Pr(s,l) := \Pr(\overline{D_s L_l})$.

Using this, we can see that Δ_m can be expanded as a Kullback-Leibler divergence,

$$\Delta_m = \frac{1}{\beta_{\text{N}}} \sum_{s,l} \Pr(s,l) \log \frac{\Pr(s,l)}{\Pr(s,l+m)}. \tag{63}$$

Applying the log sum inequality to the summation over syndromes, we get a bound in terms of the KL divergence of the marginal $\Pr(l)$,

$$\Delta_m \geq \frac{1}{\beta_{\text{N}}} \sum_{l} \Pr(l) \log \frac{\Pr(l)}{\Pr(l+m)}. \tag{64}$$

We note that $\Pr(l=0)$ is the probability of the MLD succeeding. Since we are below threshold we have that $\Pr(l=0) \to 1$, and so $\Delta_m \to \infty$. □

Above the threshold one cannot simply consider the quenched average Δ_m. To see this, consider a code in which a single unlikely syndrome can be perfectly decoded with certainty, and all others provide a uniform logical error. The correctable syndrome gives that $\Delta_m = \infty$, despite the fact that the code is above threshold. As such, we cannot necessarily conclude that $\Delta_m \to 0$ above threshold.

Therefore, to provide a converse statement we must consider more than just the quenched average. Here we will formalise an argument first sketched in Ref. [?] that shows that the way to obtain a converse is to change our notion of convergence.

Suppose for the moment that $\Delta_m(E)$ concentrates around a value $\tilde{\Delta}_m$, in the sense that

$$\lim_{n\to\infty} \Pr\Big(|\Delta_m(E) - \tilde{\Delta}_m| \leq \epsilon\Big) = 1 \quad \forall \epsilon > 0. \tag{65}$$

A phase transition corresponds to a jump in this 'typical' value $\tilde{\Delta}_m$. Because $\Delta_m(E)$ is not bounded from above, we notice that this typical value $\tilde{\Delta}_m$ need not necessarily correspond to the mean Δ_m. Indeed, in the above counterexample, where $\Delta_m = \infty$ above threshold, we nonetheless see that $\tilde{\Delta}_m \to 0$ as expected. As such, we see that the mean is not necessarily the correct figure-of-merit. If instead we look to this typical value, then we can see that $\tilde{\Delta}_m \to \infty$ and $\tilde{\Delta}_m \to 0$ below *and* above threshold respectively.

Lemma 2 (Divergence/convergence in probability [?])**.** *The code is below threshold if and only if the free energy cost of every non-trivial logical m diverges in probability,*

$$\lim_{n\to\infty} \Pr\big(\Delta_m(E) \geq t\big) = 1 \quad \forall t, m \neq 0. \tag{66}$$

Similarly, the code is above threshold if and only if the free energy cost of every logical converges to zero in probability,

$$\lim_{n\to\infty} \Pr\big(\,|\Delta_m(E)| \leq \epsilon\,\big) = 1 \quad \forall \epsilon > 0, m. \tag{67}$$

Proof. The success probability of the MLD corresponds to $\Pr(l = 0)$. We first note that $\Delta_m(E)$ can be written in terms of conditional probabilities,

$$\Delta_m(E) = \frac{1}{\beta_{\mathrm{N}}} \log \frac{\Pr(l|s)}{\Pr(l+m|s)}. \tag{68}$$

Next we recall that the maximum likelihood condition $\Pr(l = 0|s) \geq \Pr(m|s)$ holds for all s and m, and as such $\Pr(l = 0) \geq 1/d^K$. The code is below threshold if the probability of successful decoding approaches 1—this is equivalent to the probability of successful decoding *conditioned upon the syndrome* approaching 1 almost surely,

$$\lim_{n\to\infty} \Pr\Big[\Pr(l = 0|s) \geq 1 - \gamma\Big] = 1 \quad \forall \gamma > 0. \tag{69}$$

To be clear, in this expression the inner probability is a probability over logicals l and the outer probability is over syndromes. Thus, the below threshold condition says informally that this distribution peaks at one in the limit. This can be seen to be equivalent to being in the ordered phase,

$$\lim_{n\to\infty} \Pr\big(\Delta_m(E) \geq t\big) = 1 \quad \forall t, m \neq 0. \tag{70}$$

The forward and reverse directions are given by taking the choices of parameter

$$t := \frac{1}{\beta_{\mathrm{N}}} \log \frac{1-\gamma}{\gamma}, \tag{71a}$$

$$\gamma := 1 - \frac{1}{1 + e^{-\beta_{\mathrm{N}} t}(K-1)}, \tag{71b}$$

respectively.

In a similar vein, being above threshold implies that the conditional probability of decoding almost surely approaches $1/K$,

$$\lim_{n\to\infty} \Pr\Big[\Pr(l = 0|s) \leq 1/K + \eta\Big] = 1 \quad \forall \eta > 0. \tag{72}$$

This is equivalent to being in the disordered phase,

$$\lim_{n\to\infty} \Pr\big(|\Delta_m(E)| \leq \epsilon\big) = 1 \quad \forall \epsilon > 0. \tag{73}$$

The forward and reverse directions here follow from the choices of parameter

$$\epsilon := \frac{1}{\beta_{\mathrm{N}}} \log \frac{1+\eta K}{1 - \eta K(K-1)}, \tag{74a}$$

$$\eta := \frac{1}{1 + (K-1)e^{-\beta_{\mathrm{N}}\epsilon}} - \frac{1}{K}, \tag{74b}$$

respectively. □

We add more remark about the notions of above and below threshold that we have adopted here. As noted in the case of the surface code with independent X and Z noise with different strengths, there can be a gap between the cases of above and below threshold. In that case our theorem does not apply. However, it is still the case that the steps of our proof could be followed to establish that there are a sequence of phase transitions in the disorder parameters Δ_m in the case where there are potentially differing thresholds for each logical subalgebras.

D. Reentrance

Above we have shown that, along the Nishimori line, there exists a phase transition which corresponds to the error correction threshold of the optimal decoder. By considering non-optimal decoders, we can extend this to show the equivalence between phase transitions and thresholds away from the Nishimori line.

By way of example, we consider the β-MFE decoders. We could now consider the phase diagram of our stat mech system in (θ, T)-space, where θ was the parameter of our noise model, and $T = 1/\beta$ is the temperature.

Lemma 3 (Reentrance of the ordered phase)**.** *For any β such that the β-MFE decoder has a threshold, this threshold $\theta_t(\beta)$ is always less than or equal to that seen on the Nishimori line,*

$$\theta_t(\beta) \leq \theta_t(\beta_N). \tag{75}$$

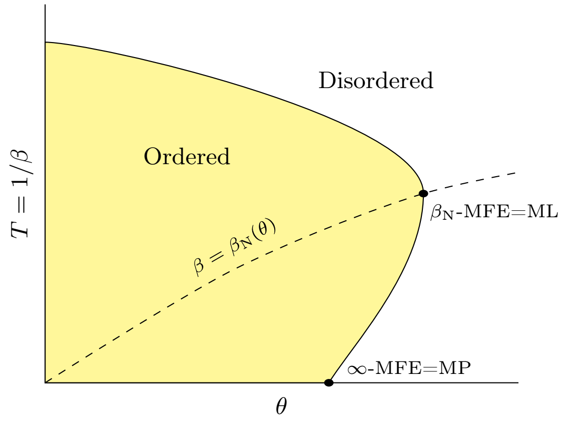

FIG. 6. A sketch of the generic phase diagram for the stat mech model. The solid line indicates the phase boundary, and the shaded region the phase in which error correction is possible. The dashed line indicates the Nishimori condition, along which the MFE decoder corresponds to the ML decoder. The thresholds of the MP and ML decoder are both indicated by black dots. An explicit example, for the case of bit-flip noise in the toric code, is presented in Figure 3.

Proof. As the β-MFE decoder is directly defined in terms of minimising the free energy, Lemmas 1 and 2 both naturally extend to show that the stat mech model has a phase transition at inverse temperature β and $\theta = \theta_t(\beta)$.

As the ML decoder is optimal, it necessarily has the highest threshold of any decoder. Recalling that the ML decoder is precisely the $\beta_{\rm N}$-MFE decoder, where $\beta_{\rm N}$ is the inverse Nishimori temperature, this implies that $\theta_t(\beta) \leq \theta_t(\beta_{\rm N})$ for all β. □

VII. TENSOR NETWORK MAXIMUM LIKELIHOOD DECODING

In Section VI we showed that the problem of maximum likelihood decoding can be reduced to calculating partition functions/free energy differences. Specifically it reduces to calculating

$$\delta_{\rm ML}(s) := \arg\max_l Z_{C_s L_l}(\beta_{\rm N}), \tag{76}$$

where $\beta_{\rm N}$ is the inverse Nishimori temperature. This relationship was first discussed in Ref. [?].

The problem of calculating the partition function of a stat mech model can be naturally expressed as the contraction of a tensor network [? ?], as we will review below. This suggests a general method for approximating the ML decoder by finding a tractable (possibly approximate) contraction sequence for the tensor network associated to the partition function.

The idea of a such a *tensor network decoder* is not new. In Ref. [?], Bravyi, Suchara and Vargo (BSV) consider an explicit tensor network for the ML decoder for the surface code with independent Pauli noise, and numerically study the threshold for some parameter choices. It is clear from their paper that this method generalises to other codes with independent Pauli noise models [?], though the specific methods they use for (approximately) contracting the tensor network would need to be modified if the code were not a planar code. Other authors have also considered tensor network decoders which can account for non-Pauli noise [? ?].

In fact, the tensor network considered by BSV [?] is exactly the natural tensor network that one obtains from the partition function using the mapping of Refs. [? ?]. Thus, the tensor network decoders that we define below are generalisations of the BSV decoder that, by virtue of Theorem 2, apply even in the case of correlated noise models on arbitrary codes. If the code has low-weight stabiliser (or gauge) generators, then storing these tensors is efficient. Contracting a general tensor network is unfortunately #P-hard [? ?], however, we will discuss efficient approximate contraction strategies that could be employed in the interesting cases of spatially local codes.

A. Tensor network algorithm for approximate maximum likelihood decoding

To see how calculating the partition function of a stat mech model can be naturally expressed as the contraction of a tensor network, we begin with an arbitrary Hamiltonian

$$h(\vec{s}) = \sum_j h_j(\vec{s}), \tag{77}$$

with a corresponding partition function

$$Z = \sum_s e^{-\beta h(\vec{s})} = \sum_s \prod_j e^{-\beta h_j(\vec{s})}. \tag{78}$$

This expression almost takes the form of a tensor contraction. To make this into a tensor network, consider indices $\{\alpha_{i,j}\}_{i,j}$, labelled by both spins and interactions. We want to require that $\alpha_{i,j} = s_i$ for all j, which we can impose by including Kronecker delta factors. In terms of these indices, the partition function can be expressed as

$$Z = \sum_\alpha \left[\prod_i \delta\left(\{\alpha_{i,j}\}_j\right) \prod_j e^{-\beta h_j(\{\alpha_{i,j}\}_i)} \right]. \tag{79}$$

The virtue of this expression is that each $\{\alpha_{i,j}\}_{i,j}$ now occurs precisely twice in each term above, and as such the above expression forms a tensor network. Obviously, for local Hamiltonians in which h_j only depends non-trivially on a finite set of sites i, we can drop any index $\alpha_{i,j}$ for which interaction h_j is independent of s_i. After doing so, the tensor network simplifies to a network which has the same connectivity as the underlying Hamiltonian. Specifically, there is a tensor corresponding to each site (δ) and

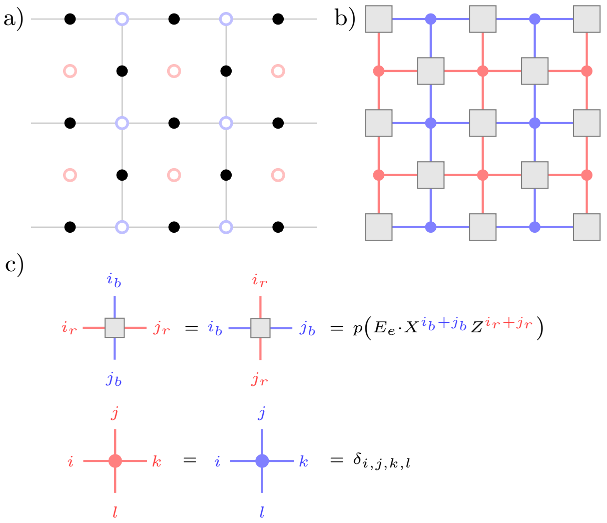

FIG. 7. Tensor network ML decoder of the surface code under iid noise. a) surface code and the associated statistical mechanical model. Black circles indicate qubits, and the red/blue open circles the Z/X-type spins. See Figure 1 for the interactions within this model. b) Tensor network for the probability of an error class $\Pr(\overline{E})$, or equivalently for the partition function Z_E, for iid noise distributed according to $p : \mathcal{P} \to \mathbb{R}$. We note that this matches Fig. 6 of Ref. [?]. c) Values of the two tensors involved. Red/blue dots are δ tensors, corresponding to the Z/X-type spins, and grey tensors correspond to Gibbs weights. Once again this matches the tensors in Equations. (39–41) of Ref. [?].

each bond (Gibbs weight), with the natural connectivity. In Figure 7 we show the corresponding diagram for the surface code under iid noise, which reduces to that found in Ref. [?].

Applying this construction to the stat mech mappings of Sections II, III and V, we see that contracting this tensor network gives an algorithm to compute the ML decoding. Because contracting tensor networks is generally hard, we will also need *approximate* tensor contraction schemes that provide parameterised approximations.

In the case of 2D topological codes under spatially correlated noise, three approximate contraction schemes suggest themselves:

- MPS-MPO contraction [? ?]. For the toric code, this method exactly reproduces the MPS decoder of Ref. [?] (see Figure 7).
- Transfer matrix [?] and corner transfer matrix methods [? ?].
- Methods that involve renormalisation on the virtual level, e.g. tensor network renormalisation (TNR) [? ?] or (higher-order) tensor renormalisation group (HOTRG) [? ?]. These examples remain efficient when extended to more than two dimensions.

Each of these methods contain an approximation parameter in the form of the bond dimension. For a fixed bond dimension, these methods all provide polynomial time approximations to the ML decoder, though we do not know of any general results which control the approximation as a function of the bond dimension in any non-trivial way. Actually implementing these decoders (beyond what was done in Refs. [? ?]) or finding provable guarantees on the approximation to the ML decoder both remain open problems.

VIII. CONCLUSIONS

In this paper we have provided a broad extension of the stat mech mapping of Ref. [?] to consider arbitrary stabiliser and subsystem codes, subject to correlated noise, including circuit-based noise. As an application, we applied Monte Carlo simulation to this construction to show how positive correlations can push down the bit-flip threshold of the toric code. Finally, we showed how the stat mech mapping gives a natural family of efficient tensor network algorithms which approximate maximum likelihood decoding, generalising the decoder of Ref. [?].

ACKNOWLEDGEMENTS

We thank Michael Beverland, Jacob Bridgeman, Benjamin Brown and Thomas Smith for helpful discussions. Special thanks go to Robin Harper for assistance with the numerics. The authors acknowledge the University of Sydney HPC service at The University of Sydney for providing HPC resources that have contributed to the research results reported within this paper.

This research was supported by the Australian Research Council Centre of Excellence for Engineered Quantum Systems (Project ID CE170100009). CTC also acknowledges support Sydney Nano Postgraduate Scholarship (John Makepeace Bennett Gift).

Appendix A: Graphical models

An important family of correlated probability distributions is that of *probabilistic graphical models* [?]. In this appendix we briefly review the relationship between certain classes of graphical models and the factored distributions we consider in Section III. Specifically we will see that both undirected graphical models (Markov random fields) and directed acyclic graphical models (Bayesian networks) are both special examples of factored distributions. As such, the statistical mechanical mapping given in Section III also naturally extends to noise models of these forms as well.

1. Markov random field

We start by considering undirected graphical models, known as Markov random fields.

Definition 6 (Markov random field)**.** A distribution $\Pr(E) = \Pr(E_1, \ldots, E_n)$ is a *Markov random field* (MRF) with respect to an undirected graph G, if it satisfies the local Markov condition

$$\Pr(E_i|E_{G\setminus i}) = \Pr(E_i|E_{\partial i}), \tag{A1}$$

for all i, where ∂i denotes the neighbours of i within G.

For such a graphical model, Hammersley-Clifford theorem tells us that it can be expressed as a factored distribution.

Lemma 4 (Hammersley-Clifford theorem)**.** *A strictly positive distribution which is a MRF with respect to G, can be factored over the set of all maximal cliques of G.*

Proof sketch. For any subset $Y \subseteq G$, let

$$\phi_Y(E_Y) := \prod_{X:X\subseteq Y} \Pr(E_X \otimes I_{G\setminus X})^{(-1)^{|Y\setminus X|}}, \tag{A2}$$

where I was chosen as an arbitrary, but fixed, reference value of E. Noting that $\mu(X,Y) = (-1)^{|Y\setminus X|}$ forms a Möbius function, we can apply Möbius inversion theorem to give

$$\Pr\left(E_Z \otimes I_{G\setminus Z}\right) = \prod_{Y\subseteq Z} \phi_Y(E_Y), \tag{A3}$$

and more specifically

$$\Pr\left(E\right) = \prod_{Y\subseteq G} \phi_Y(E_Y). \tag{A4}$$

Applying the local Markov property, we find that $\phi_Y \equiv 1$ for any Y which is *not* a clique. As such we find that

$$\Pr(E) = \prod_j \phi_{C_j}(E_{C_j}), \tag{A5}$$

where $\{C_j\}_j$ are the set of *all* cliques in G. By grouping together factors, this can straightforwardly reduced to a factorisation over just *maximal* cliques. □

2. Bayesian Networks

We now consider directed cyclic graphical models, known as Bayesian Networks.

Definition 7 (Bayesian network)**.** A distribution $\Pr(E) = \Pr(E_1, \ldots, E_n)$ is a *Bayesian network* (BN) with respect to an directed acyclic graph G, if it satisfies the local Markov condition

$$\Pr(E_i|E_{G\setminus\partial^- i}) = \Pr(E_i|E_{\partial^+ i}), \tag{A6}$$

for all i, where $\partial^- i$ and $\partial^+ i$ denote the descendants and parents of i within G respectively.

As with Markov random fields, we see that Bayesian networks also naturally factorise.

Lemma 5 (Factorisation of Bayesian Networks)**.** *A distribution* $\Pr(\cdot)$ *which is a Bayesian network with respect to a directed acyclic graph G, can be factored over* $\{\{i\} \cup \partial^+ i\}_i$. *Specifically,*

$$\Pr(E) = \prod_{i=1}^{n} \Pr(E_i|E_{\partial^+ i}). \tag{A7}$$

Proof. As our directed graph G is acyclic, it possesses a topological order. Consider labelling our indices according to that ordering, such that $\partial^+ i > i > \partial^- i$ elementwise. Next we consider expanding our joint distribution using the chain rule of conditional probabilities,

$$\Pr(E_1, \ldots, E_n) = \prod_{i=1}^{n} \Pr(E_i|E_{i+1}, \ldots, E_n). \tag{A8}$$

Applying the local Markov condition, we get the desired factorisation,

$$\Pr(E) = \prod_{i=1}^{n} \Pr(E_i|E_{\partial^+ i}). \tag{A9}$$

□

Appendix B: Numerical simulation details

In this appendix we cover the details of how the statistical mechanical simulations of Section IV were performed, specifically how the data presented in Figure 3 was collected.

The simulations we present follow closely the techniques used in Ref. [?] to study the toric code under a depolarising noise model.

1. Order parameter

As discussed in Section IV, the system we are studying corresponds to a random-bond Ising model, with an addition four-body coupling. In the case of zero disorder, $p = 0$, this system simply limits to the standard square-lattice Ising model. Noticing this, a convenient order parameter for this system is simply given by the average magnetisation,

$$m := \frac{1}{L^2} \sum_i s_i, \tag{B1}$$

where L is the linear size of our system, such that we have L^2 sites.

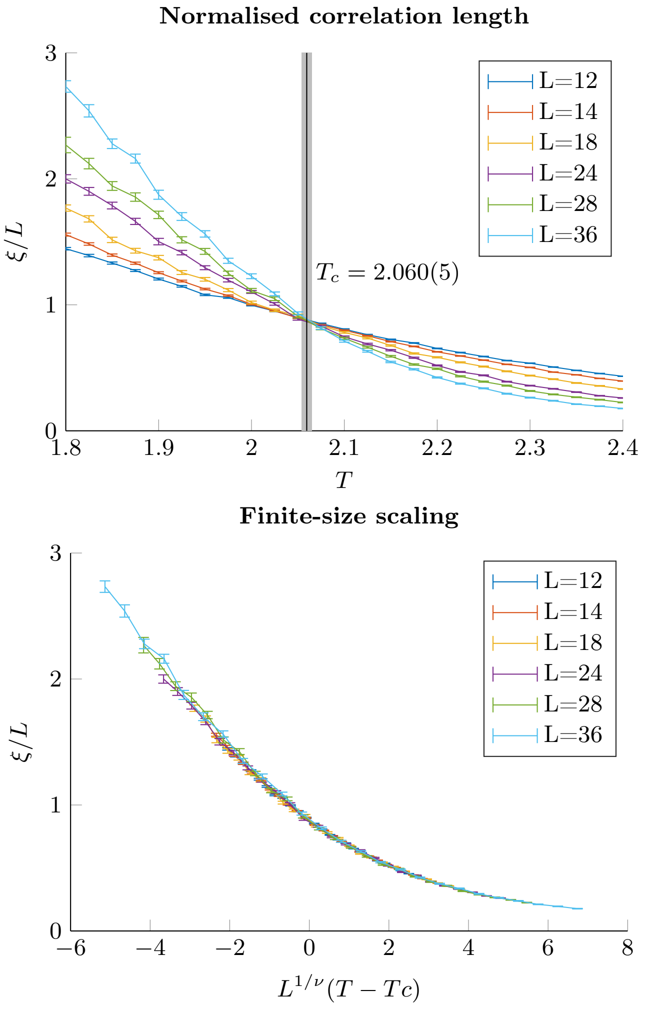

FIG. 8. Crossing diagrams for $p = 6\%$. a) Normalised correlation length versus temperature, for several system sizes, with critical temperature indicated. b) Finite-size scaling of normalised correlation length. All error bars indicated are 95% confidence intervals given by bootstrapping.

2. Finite-size scaling

For very large system sizes, we could sample the magnetisation directly, and find the phase transition by detecting when the system spontaneously magnetises. This naïve method would however require prohibitively large system sizes, especially for large disorder. Instead we will utilise this order parameter indirectly, by considering the finite-size scaling of the corresponding correlation length.

Following Ref. [?], we define the wave-vector-dependent magnetic susceptibility

$$\chi(\vec{k}) := \frac{1}{L^2}\left\langle \left| \sum_i s_i e^{i\vec{k}\cdot\vec{r_i}} \right|^2 \right\rangle, \tag{B2}$$

where $\vec{r_i}$ denotes the spatial position of site i, and $\langle\cdot\rangle$ denotes the thermal and disorder average,

$$\langle X \rangle := \sum_E \Pr(E) \cdot \sum_s \frac{e^{-\beta H_E(\vec{s})}}{Z_E} X(\vec{s}). \tag{B3}$$

Using this, we can now define the two-point finite-size correlation function

$$\xi := \frac{1}{2\sin(k_{\min}/2)}\sqrt{\frac{\chi(\vec{0})}{\chi(\vec{k}_{\min})} - 1}, \tag{B4}$$

where $\vec{k}_{\min} = (2\pi/L, 0)$ is the minimal non-zero wave-vector. Near the phase transition this correlation length is expected to have the finite-size scaling [?]

$$\xi/L \approx f\left[L^{1/\nu}(T - T_c)\right], \tag{B5}$$

where f is a dimensionless scaling function.

At the critical temperature $T = T_c$, the normalised correlation length ξ/L becomes independent of temperature. We will determine the temperature of the phase transition by plotting ξ/L as a function of T for several different system sizes L, and fitting to Equation (B5). If these curves do not cross, then this will be taken as indication that no phase transition is present.

3. Simulation parameters

p (%)	T	τ_{14}	τ_{24}	τ_{36}
2.00	2.00 – 2.50	14	16	18
4.00	2.00 – 2.50	14	16	18
6.00	1.80 – 2.40	14	16	18
8.00	1.60 – 2.30	16	18	22
8.50	1.50 – 2.10	17	21	23
9.00 – 9.10	1.40 – 2.00	18	22	24
9.20 – 9.50	1.30 – 2.00	18	22	25
9.55 – 10.20	1.25 – 2.00	19	23	26
10.30 – 10.50	1.25 – 2.00	20	24	27

TABLE I. Temperature ranges and equilibration times. An equilibration time of τ corresponds to 2^τ Metropolis-Hastings steps (a number of single spin updates equal to the system size) before the correlation lengths recorded in the three last logarithmic bins are statistically consistent.

We simulate our system for system sizes of $L \in \{12, 14, 18, 24, 28, 36\}$. Equilibration of our system is found by logarithmically binning the correlation length, and requiring that three successive bins agree to within error bars. To save time, the equilibration testing was only done on $L = 14, 24, 36$, and these times were reused for $L = 12, 18, 28$. The times and temperature ranges are given in Appendix B 3.

For $L = 12, 14$ the correlation length is recorded for 5000 disorder samples, 1000 samples are taken for $L = 18, 24$, and 500 for $L = 28, 36$. The temperature ranges are swept over with a resolution of $\Delta T = 0.025$. Error bars are generated for all the data is given by the 95% confidence intervals given by bootstrapping, using 10,000 resamples. An example crossing plot is given for $p = 6\%$ in Figure 8. By performing this analysis for each p, we can estimate the phase boundary, as seen in Figure 3. All the simulations took a total of 1.1×10^6 CPU-hours, or 120 CPU-years.