Spin glass reflection of the decoding transition for quantum error correcting codes

Alexey A. Kovalev
Department of Physics & Astronomy and Nebraska Center for Materials and Nanoscience,
University of Nebraska, Lincoln, Nebraska 68588, USA

Leonid P. Pryadko
Department of Physics & Astronomy, University of California, Riverside, California 92521, USA

We study the decoding transition for quantum error correcting codes with the help of a mapping to random-bond Wegner spin models. Families of quantum low density parity-check (LDPC) codes with a finite decoding threshold lead to both known models (e.g., random bond Ising and random plaquette $\mathbb{Z}_2$ gauge models) as well as unexplored earlier generally non-local disordered spin models with non-trivial phase diagrams. The decoding transition corresponds to a transition from the ordered phase by proliferation of extended defects which generalize the notion of domain walls to non-local spin models. In recently discovered quantum LDPC code families with finite rates the number of distinct classes of such extended defects is exponentially large, corresponding to extensive ground state entropy of these codes. Here, the transition can be driven by the entropy of the extended defects, a mechanism distinct from that in the local spin models where the number of defect types (domain walls) is always finite.

Keywords: spin glass, Ising, Wegner, stabilizer code, decoding transition

Locality in space-time is a great organizing principle for a theoretical physicist who is trying to come up with a model for some phenomenon. It works beautifully both in high energy and in condensed matter physics. Depending on the details, the corresponding techniques can be based on the derivative expansion, minimal gauge coupling, or local lattice Hamiltonians. Often enough, given a few more specific symmetries and natural constraints, and considering only the most local models, one can derive a unique functional form of an effective Hamiltonian.

On the other hand, having concentrated for so long on local models, we remain largely unaware of the physics that may be lurking out there, beyond the familiar locality constraint. Problem is, the space of possible non-local continuum or discrete models is vast. Without this constraint, and given that most interactions in nature are indeed local, what property can we use instead to select a non-trivial model?

In this work we explore disordered spin models associated with maximum likelihood decoding for stabilizer quantum error correction codes[1, 2]. The construction is a generalization of the map between various surface codes and two-dimensional spin models[3–5]. This approach turns out both physically intuitive and useful as a way to choose non-trivial spin models, especially if one concentrates on quantum low density parity-check (LDPC) codes[6, 7].

Unlike the case of the classical error-correcting codes[8] where decoding can be done by minimizing certain energy functional[9, 10], with a quantum stabilizer code large groups of *mutually-degenerate* errors can not and need not be distinguished[1, 2]. To find the most likely error, one has to decide between different equivalence classes; this boils down to minimizing certain free energy functional depending on the relevant error model. We consider a particularly simple error model where this functional can be readily interpreted as the free energy for a disordered Ising spin model of a general form studied by Wegner[11] at some temperature $T \equiv \beta^{-1}$ and individual bonds flipped independently with probability p; the original decoding problem lives on the Nishimori line[10, 12–14] of the phase diagram, which generalizes the result for the surface codes[3]. For a code family where in the limit of large codes the decoding can be done successfully with probability one, the corresponding spin models are non-trivial, meaning that they definitely have an ordered "defect-free" phase at small T and p, and a distinct disordered phase at large T and p. In addition, in the clean limit, $p \to 0$, the spin models associated with quantum codes have exact self-duality, in Wegner's sense[11]. Further, in the special case of quantum LDPC codes[6, 7], the Hamiltonians of the spin models is a sum of generalized Ising bonds, each given by a (generally non-local) product of only a few spin operators.

We show that the decoding transition corresponds to a transition from the ordered phase by proliferation of extended defects which generalize the notion of domain walls to non-local spin models. In the code families where the number of encoded qubits k remains finite in the limit of large n, the transition occurs when the tension λ of one or more of such defects vanish, akin to vanishing line tension of a domain wall in the 2D Ising model. In quantum LDPC code families with finite rates[15–19] the number of distinct classes of such extended defects is exponentially large, corresponding to extensive ground state entropy of these codes. Here, the transition can happen even when all defects have finite energy densities $\lambda \geq \lambda_0 > 0$, driven by the entropy of the extended defects' types. This mechanism is distinct from that in the local spin models where the number of defect types (domain walls) is always finite.

The paper is organized as follows. We first give a brief overview of the **Main results**, namely, formulate all the theorems and briefly describe other results, concentrating on the case of Calderbank-Shor-Steane (CSS) codes[20, 21]. Then, we give a detailed review of the necessary **Background** facts about quantum stabilizer codes. In section **Statistical mechanics of decoding** we describe how a spin model is constructed from a given code family, and relate the possibility of successful decoding with probability one to the existence of an ordered phase in the corresponding spin model. In section **Phase transitions** we discuss the properties of the thermodynamical phase transition corresponding to the maximum-likelihood (ML) decoding threshold, introduce spin correlation functions which can characterize various phases, and give several inequalities on the location of the transition. Finally, we give our **Conclusions**.

I. MAIN RESULTS

We start by listing our main results. To simplify the definitions, here we concentrate on the case of CSS codes; more general results are given later in the text along with the corresponding proofs.

A. Definitions

A classical binary linear code[8] $\mathcal{C}$ with parameters $[n,k,d]$ is a k-dimensional subspace of the vector space $\mathbb{F}_2^n$ of all binary strings of length n. Code distance d is the minimal weight (number of non-zero elements) of a non-zero string in the code. Rows of the binary *generator matrix* G of the code $\mathcal{C}\equiv\mathcal{C}_G$ are formed by its k basis vectors. A linear code can also be specified by the binary *parity check matrix* H, $\mathcal{C}=\{\mathbf{c}\in\mathbb{F}_2^n|H\mathbf{c}^T=0\}$. This implies that H and G are mutually orthogonal, $HG^T=0$, and also

$$\operatorname{rank} H+\operatorname{rank} G=n. \tag{1}$$

Parity check matrix is a generating matrix of the code $\mathcal{C}^\perp=\mathcal{C}_H$ dual to $\mathcal{C}$. Respectively, the matrix H is an exact dual to G, $H\equiv G^*$. Note that here and throughout this work we assume that all linear algebra is done modulo 2, as appropriate for the vector space $\mathbb{F}_2^n$.

Given a binary matrix Θ with dimensions $N_s\times N_\mathrm{b}$, we define a generalized Wegner-type [11] partition/correlation function with multi-spin bonds $R_b\equiv\prod_r S_r^{\Theta_{r,b}}$ corresponding to the columns of Θ and Ising spin variables $S_r=\pm1$, $r=1,\ldots,N_s$:

$$\mathscr{Z}_{\mathbf{e},\mathbf{m}}(\Theta;\{K\})\equiv\frac{1}{2^{N_g}}\sum_{\{S_r=\pm1\}}\prod_{b=1}^{N_\mathrm{b}}R_b^{m_b}\frac{\exp\bigl(K_b(-1)^{e_b}R_b\bigr)}{2\cosh\beta}, \tag{2}$$

where we assume the couplings to be positive, $K_b\equiv\beta J_b>0$, with β being the inverse temperature, the length-N_b binary vectors $\mathbf{e}$, $\mathbf{m}$ respectively specify the electric and magnetic disorder, and $N_g\equiv N_s-\operatorname{rank}\Theta$ is the count of linearly-dependent rows in Θ.

A quantum CSS code[20, 21] with parameters $[[n,k,d]]$ can be specified in terms of two n-column binary generator matrices $\mathcal{G}_X$, $\mathcal{G}_Z$ with mutually orthogonal rows, $\mathcal{G}_X\mathcal{G}_Z^T=0$. Such a code encodes $k=n-\operatorname{rank}\mathcal{G}_X-\operatorname{rank}\mathcal{G}_Z$ qubits in a block of n qubits. A CSS code can be thought of as a couple of binary codes, one correcting X-type errors and the other Z-type errors. However, it turns out that any two errors $\mathbf{e}$ and $\mathbf{e}'$ of, e.g., Z-type differing by a linear combination of rows of $\mathcal{G}_Z$ have exactly the same effect on the quantum code—such errors are called *degenerate*. The corresponding *equivalence* is denoted $\mathbf{e}\simeq\mathbf{e}'$. A *detectable* Z-type error $\mathbf{e}=\mathbf{e}_Z$ has a non-zero *syndrome* $\mathbf{s}_Z=\mathcal{G}_X\mathbf{e}^T$. An undetectable and *non-trivial* Z-type error has a zero syndrome and is not degenerate with an all-zero error; we will call such an error a (non-zero Z-type) *codeword* $\mathbf{c}=\mathbf{c}_Z$. The distance d of a CSS code is the minimal weight of a Z- or an X-type codeword.

For each error type, we introduce a spin glass partition function:

$$Z_0^{(\mu)}(\mathbf{e};\beta)\equiv\mathscr{Z}_{\mathbf{e},\mathbf{0}}(\mathcal{G}_\mu;\{K_b=\beta\}),\ \mu=X,Z. \tag{3}$$

The normalization is such that for a model of independent X or Z errors with equal probability p (probabilities of $e_b=1$ are independent of each other and equal to p), at the Nishimori line [10, 12–14],

$$\beta=\beta_p,\quad e^{-2\beta_p}=p/(1-p), \tag{4}$$

the partition function (3) equals to a total probability of a μ-type error equivalent to $\mathbf{e}$. We also define the partition function with an *extended defect* of additionally flipped bonds at the support of the codeword $\mathbf{c}$,

$$Z_{\mathbf{c}}^{(\mu)}(\mathbf{e};\beta)\equiv Z_0^{(\mu)}(\mathbf{e}+\mathbf{c};\beta), \tag{5}$$

as well as the partition function corresponding to all errors with the same syndrome $\mathbf{s}$ as $\mathbf{e}\equiv\mathbf{e}_\mathbf{s}$,

$$Z_{\mathrm{tot}}^{(\mu)}(\mathbf{s};\beta)\equiv\sum_{\mathbf{c}}Z_{\mathbf{c}}^{(\mu)}(\mathbf{e}_\mathbf{s};\beta)=\mathscr{Z}_{\mathbf{e}_\mathbf{s},\mathbf{0}}(\mathcal{G}_{\bar\mu}^*;\{K_b=\beta\}), \tag{6}$$

where the summation is over all 2^k mutually non-degenerate μ-type codewords $\mathbf{c}$, such that $\mathcal{G}_{\bar\mu}\mathbf{c}^T=0$, and $\bar\mu=X$ if $\mu=Z$ and vice versa. The second form uses a matrix $\mathcal{G}_{\bar\mu}^*$ exactly dual to $\mathcal{G}_{\bar\mu}$, cf. Eq. (1). Note that Eq. (6) at $\beta=\beta_p$ gives the correctly normalized probability to encounter the syndrome $\mathbf{s}$, $\sum_\mathbf{s}Z_\mathrm{tot}(\mathbf{s};\beta)=1$. Here and below we omit the error-type index μ to simplify the notations.

Syndrome-based decoding is a classical algorithm to recover the error equivalence class from the measured syndrome. In maximum-likelihood (ML) decoding, one picks the codeword $\mathbf{c}=\mathbf{c}_{\max}(\mathbf{e})$ corresponding to the largest contribution $Z_{\max}(\mathbf{s};\beta)\equiv Z_{\mathbf{c}_{\max}(\mathbf{e})}(\mathbf{e};\beta)$ to the partition function (6) at $\beta=\beta_p$. Given some unknown error with the syndrome $\mathbf{s}$, the conditional probabilities of successful and of failed ML recovery are, respectively,

$$P_{\rm succ}(\mathbf{s})=\frac{Z_{\rm max}(\mathbf{s};\beta_p)}{Z_{\rm tot}(\mathbf{s};\beta_p)},\quad P_{\rm fail}(\mathbf{s})=1-P_{\rm succ}(\mathbf{s}). \tag{7}$$

The corresponding average over errors can be written as a simple sum over allowed syndrome vectors,

$$P_{\rm succ}\equiv\left[\frac{Z_{\rm max}(\mathbf{s}_\mathbf{e};\beta_p)}{Z_{\rm tot}(\mathbf{s}_\mathbf{e};\beta_p)}\right]=\sum_{\mathbf{s}}Z_{\rm max}(\mathbf{s};\beta_p), \tag{8}$$

where the square brackets $[\,\cdot\,]$ denote an average over the errors $\mathbf{e}$. For a given infinite family of CSS codes, asymptotically certain ML decoding implies $P_{\rm succ}^{(X)}\to 1$ and $P_{\rm succ}^{(Z)}\to 1$ in the limit of large n.

In terms of the spin glass model (3), this corresponds to a phase where in thermodynamical limit each likely disorder configuration $\mathbf{e}$ corresponds to a unique defect configuration $\mathbf{c}=\mathbf{c}_{\rm max}(\mathbf{e})$:

Definition 1. *(CSS) A* fixed-defect phase *of the spin glass model (3) corresponding to an infinite family of CSS codes has*

$$[Z_{\rm max}^{(\mu)}(\mathbf{s}_\mathbf{e};\beta)/Z_{\rm tot}^{(\mu)}(\mathbf{s}_\mathbf{e};\beta)]\to 1,\quad n\to\infty. \tag{9}$$

It is also useful to define a special case of such a phase where any likely disorder configuration does not introduce any defects:

Definition 2. *(CSS) A* defect-free phase *of the spin glass model (3) corresponding to an infinite family of CSS codes has*

$$[Z_0^{(\mu)}(\mathbf{e};\beta)/Z_{\rm tot}^{(\mu)}(\mathbf{s}_\mathbf{e};\beta)]\to 1,\quad n\to\infty. \tag{10}$$

B. Results: ordered phases

We first prove that the only allowed ordered phase on the Nishimori line is the defect-free phase:

Theorem 1. *For an infinite family of quantum stabilizer codes successful decoding with probability one implies that on the Nishimori line the corresponding spin model is in the defect-free phase, i.e., in any likely configuration $\mathbf{e}$ of flipped bonds the largest $Z_\mathbf{c}(\mathbf{e};\beta_p)$ corresponds to $\mathbf{c}_{\rm max}(\mathbf{e})=\mathbf{0}$.*

Definitions 1 and 2 are formulated in terms of the average ratios of partition functions. As an alternative, we introduce the free energy increment associated with adding an extended defect $\mathbf{c}$ to a most likely configuration at the given disorder $\mathbf{e}$ with the syndrome $\mathbf{s}=G_{\bar\mu}\mathbf{e}^T$,

$$\Delta F_\mathbf{c}^{\max,\mu}(\mathbf{s};\beta)\equiv\beta^{-1}\log\frac{Z_{\rm max}^{(\mu)}(\mathbf{s};\beta)}{Z_{\mathbf{c}_{\rm max}(\mathbf{e})+\mathbf{c}}^{(\mu)}(\mathbf{e};\beta)},\ \ \mu=X,Z. \tag{11}$$

We prove

Theorem 2. *For an infinite family of disordered spin models (3) (or Eq. (33)), in a fixed-defect phase the averaged over the disorder free energy increment for an additional defect corresponding to a non-trivial codeword $\mathbf{c}\not\simeq\mathbf{0}$ diverges at large n, $[\Delta F_\mathbf{c}^{\max}(\mathbf{s}_\mathbf{e};\beta)]\to\infty$.*

In the defect-free phase, the relevant analogous quantity is the free energy increment with respect to a given error $\mathbf{e}$,

$$\Delta F_\mathbf{c}^{(0,\mu)}(\mathbf{e};\beta)\equiv\beta^{-1}\log\frac{Z_0^{(\mu)}(\mathbf{e};\beta)}{Z_\mathbf{c}^{(\mu)}(\mathbf{e};\beta)}. \tag{12}$$

The corresponding average over disorder diverges in the defect-free phase where $\mathbf{c}_{\rm max}(\mathbf{e})=\mathbf{0}$ for every likely error configuration $\mathbf{e}$. Then, the Theorem 1 leads to

Corollary 1. *On the Nishimori line, the disorder-averaged free energy increment $[\Delta F_\mathbf{c}^{(0)}(\mathbf{e};\beta_p)]$ corresponding to any non-trivial codeword $\mathbf{c}\not\simeq\mathbf{0}$ diverges at large n for $p<p_c$, where p_c is the error probability corresponding to the ML decoding transition on the Nishimori line.*

We also introduce a *tension*

$$\lambda_\mathbf{c}\equiv\frac{[\Delta F_\mathbf{c}^{\max}]}{d_\mathbf{c}},\quad d_\mathbf{c}\equiv\min_{\boldsymbol\sigma}\mathrm{wgt}(\mathbf{c}+\boldsymbol\sigma\mathcal{G}), \tag{13}$$

an analog of the domain wall line tension for the extended defects, and prove

Theorem 3. *For disordered spin models (3) (or Eq. (33)) corresponding to an infinite family of quantum codes with asymptotic rate $R=k/n$, in a fixed-defect phase, the defect tension $\overline\lambda$ averaged over all non-trivial defect classes at large n satisfy the inequality $\beta\overline\lambda\ge R\ln 2$.*

C. Results: order parameter

The spin models corresponding to families of quantum codes include the analogs of regular Ising model (e.g., regular Ising model on square lattice for the toric codes) as well as various gauge models, see Example 7. In general, there is no local order parameter that can be used for an alternative definition of the ordered phase. In addition, while an analog of *Wilson loop* operator can be readily constructed for these models and has the usual low- and high-temperature asymptotics, it remains an open question whether it can be used to distinguish between specific disordered phases.

However, we constructed a set of non-local *indicator* spin correlation functions which must all be asymptotically equal to one in the defect-free phase, while some of them change sign in the presence of extended defects. Using these, and the standard inequalities from the gauge theory of spin glasses, we prove the following bound on the location of the defect-free phase (this is an extension of Nishimori's result[12, 13] on possibly reentrant phase diagram for Ising models):

Theorem 4. *Defect-free phase cannot exist at any β for p exceeding that at the decoding transition, $p > p_c$.*

D. Results: phase transition

For zero-R codes, the only mechanism of a continuous transition is for $\lambda_{\mathbf{c}}$ to vanish for some set of codewords $\mathbf{c}$. On the other hand, for finite-rate codes, Theorem 3 implies that there is also a possibility that at the transition point the tension remains finite, $\lambda_{\mathbf{c}} \geq \lambda_{\min} > 0$, for every codeword $\mathbf{c}$. This corresponds to a transition driven by the entropy of extended defects.

While generically the transition in models with multi-spin couplings is of the first order, it is continuous along the Nishimori line since the corresponding internal energy is known exactly and is a continuous function of p. Moreover, the specific heat remains finite at the transition point along the Nishimori line since the same inequality as for regular spin glasses applies[10, 12–14],

$$[C(p;\beta_p] \leq N_{\mathrm{b}} \frac{\beta_p^2}{\cosh^2 \beta_p}, \tag{14}$$

where $N_{\mathrm{b}} = 2n$ for the model (33), and $N_{\mathrm{b}} = n$ for the models (3) corresponding to a half of a CSS code each. Thus, as in the usual spin models, we expect that the transition point $p = p_c$ at the Nishimori line is a multicritical point where several phases come together.

Spin models corresponding to non-CSS zero-rate families of stabilizer codes are exactly self-dual. The same is true for CSS codes where the two generator matrices $\mathcal{G}_X$, $\mathcal{G}_Z$ can be mapped to each other, e.g., by column permutations, as is the case for the toric codes and, more generally, for the hypergraph-product (HP) codes[15]. For many such models, the transition point at the Nishimori line can be obtained to a high numerical accuracy using the strong-disorder self-duality conjecture[22–29]

$$H_2(p_c) = 1/2, \tag{15}$$

where $H_2(p) \equiv -p \log_2 p - (1-p)\log_2(1-p)$ is the binary entropy function. While strictly speaking, there is no exact self-duality in the presence of disorder[30], we have confirmed numerically that this expression is also valid, at least approximately, for several models constructed here, e.g., models with bond structure as in Example 5.

However, for code families with finite rate, the decoding transition must be below the Shannon limit

$$R \leq 1 - H_2(p). \tag{16}$$

Thus, Eq. (15) must be violated for $R \geq 1/2$. On general grounds, we actually expect it to fail for any code family with a finite rate, $R > 0$.

II. BACKGROUND

A. Stabilizer codes

An n-qubit quantum code[1, 2, 31, 32] is a subspace of the n-qubit Hilbert space $\mathbb{H}_2^{\otimes n}$. The idea is to choose a subspace such that a likely error shifts any state from the code to a linearly-independent subspace, to be detected with a suitable set of measurements. Any error, an operator acting on $\mathbb{H}_2^{\otimes n}$, can be expanded as a linear combination of the elements of the n-qubit Pauli group $\mathscr{P}_n$ formed by tensor products of single-qubit Pauli operators X, Y, Z and the identity operator I: $\mathscr{P}_n = i^m\{I, X, Y, Z\}^{\otimes n}$, where $m = 0, 1, 2, 3$. A *weight* of a Pauli operator is the number of non-trivial terms in the tensor product.

An n-qubit quantum *stabilizer code* $\mathcal{Q}$ $[[n, k, d]]$ is a 2^k-dimensional subspace of $\mathbb{H}_2^{\otimes n}$, a common $+1$ eigenspace of all operators in the code's *stabilizer*, an Abelian group $\mathscr{S} \subset \mathscr{P}_n$ such that $-\mathbb{1} \notin \mathscr{S}$. The stabilizer is typically specified in terms of its generators, $\mathscr{S} = \langle S_1, \ldots, S_{n-k}\rangle$. Any operator proportional to an element of the stabilizer $\mathscr{S}$ acts trivially on the code and can be ignored. A non-trivial error proportional to a Pauli operator $E \notin \mathscr{S}$ is detectable iff it anticommutes with at least one stabilizer generator S_i; such an error takes a vector from the code, $|\psi\rangle \in \mathcal{Q}$, to the state $E\,|\psi\rangle$ from an orthogonal subspace $E\mathcal{Q}$ where the corresponding eigenvalue $(-1)^{s_i}$ is negative. Measuring all $n-k$ generators S_i produces the binary syndrome vector $\mathbf{s} \equiv \{s_1, \ldots, s_{n-k}\}$. Two errors (Pauli operators) that differ by an element of the stabilizer and a phase, $E_2 = E_1 S e^{i\phi}$, $S \in \mathscr{S}$, are called mutually degenerate; they have the same syndrome and act identically on the code.

Operators commuting with the stabilizer act within the code; they have zero syndrome. A non-trivial undetectable error E is proportional to a Pauli operator which commutes with the stabilizer but is not a part of the stabilizer. These are the operators that damage quantum information; minimal weight of such an operator is the distance d of the stabilizer code. A quantum or classical code of distance d can detect any error of weight up to $d-1$, and correct up to $\lfloor d/2 \rfloor$.

A Pauli operator $E \equiv i^{m'} X^{\mathbf{v}} Z^{\mathbf{u}}$, where $\mathbf{v}, \mathbf{u} \in \{0,1\}^{\otimes n}$ and $X^{\mathbf{v}} = X_1^{v_1} X_2^{v_2} \ldots X_n^{v_n}$, $Z^{\mathbf{u}} = Z_1^{u_1} Z_2^{u_2} \ldots Z_n^{u_n}$, can be mapped, up to a phase, to a binary vector $\mathbf{e} \equiv (\mathbf{v}, \mathbf{u})$. A product of two Pauli operators corresponds to a sum $(\mathrm{mod}\, 2)$ of the corresponding vectors. Two Pauli operators commute if and only if the *trace inner product* of the corresponding binary vectors is zero, $\mathbf{e}_1 \star \mathbf{e}_2 \equiv \mathbf{u}_1 \cdot \mathbf{v}_2 + \mathbf{v}_1 \cdot \mathbf{u}_2 = 0 \bmod 2$. With this map, generators of a stabilizer group are mapped to rows of the binary generator matrix

$$G = (G_X, G_Z), \tag{17}$$

with the condition that the trace inner product of any two rows vanishes [2]. This commutativity condition can be also written as $G_X G_Z^T + G_Z G_X^T = 0$.

For a more narrow set of CSS codes stabilizer generators can be chosen so that they contain products of only X_i or Z_i single-qubit Pauli operators. The corresponding generator matrix has the form

$$G = \mathrm{diag}(\mathcal{G}_X, \mathcal{G}_Z), \tag{18}$$

where the commutativity condition simplifies to $\mathcal{G}_X \mathcal{G}_Z^T = 0 \bmod 2$. The number of encoded qubits is $k = n - \operatorname{rank} G$; for CSS codes this simplifies to $k = n - \operatorname{rank} \mathcal{G}_X - \operatorname{rank} \mathcal{G}_z$.

Two errors are mutually degenerate iff the the corresponding binary vectors differ by a linear combination of rows of G, $\mathbf{e}' = \mathbf{e} + \boldsymbol{\alpha} G$. It is convenient to define the conjugate matrix $\widetilde{G} \equiv (G_Z, G_X)$ so that $G \star G^T \equiv G \widetilde{G}^T = 0$. Then, the syndrome of an error $\mathbf{e} \equiv (\mathbf{v}, \mathbf{u})$ can be written as the product with the conjugate matrix, $\mathbf{s} = \widetilde{G} \mathbf{e}^T$. A vector with zero syndrome is orthogonal to rows of $\widetilde{G}$; we will call any such vector which is not a linear combination of rows of G a non-zero *codeword* $\mathbf{c} \not\simeq \mathbf{0}$. Two codewords that differ by a linear combination of rows of G are equivalent, $\mathbf{c}_1 \simeq \mathbf{c}_2$; corresponding Pauli operators are mutually degenerate. Non-equivalent codewords represent different cosets of the degeneracy group in the binary code with the check matrix $\widetilde{G}$. For an $[[n, k, d]]$ code, any non-zero codeword has weight $\mathrm{wgt}(\mathbf{c}) \geq d$, and there are exactly $2k$ *independent* codewords which can be chosen to correspond to $2k$ operators $\bar{X}_i$, $\bar{Z}_i$, $i = 1, \ldots, k$ (with the usual commutation relations) acting on the logical qubits.

B. LDPC codes

A binary low density parity-check (LDPC) code is a linear code with sparse parity check matrix[33–36]. These have fast and efficient (capacity-approaching) decoders. Over the last ten years classical LDPC codes have become a significant component of industrial standards for satellite communications, Wi-Fi, and gigabit ethernet, to name a few. *Quantum LDPC codes*[6, 7] are just stabilizer codes[1, 2], but with stabilizer generators which involve only a few qubits each compared to the number of qubits used in the code. Such codes are most often degenerate: some errors have trivial effect and do not require any correction. Compared to general quantum codes, with a quantum LDPC code, each quantum measurement involves fewer qubits, measurements can be done in parallel, and also the classical processing could potentially be enormously simplified.

One apparent disadvantage of quantum LDPC codes is that, until recently[19], there has been no known families of such codes that have finite relative distance $\delta \equiv d/n$ for large n. This is in contrast to regular quantum codes where the existence of "good" codes with finite asymptotic rates $R \equiv k/n$ and finite δ has been proved[20, 37]. With such latter codes, and within a model where errors happen independently on different qubits with probability p, for $p < \delta/2$ all errors can be corrected with probability one. On the other hand, many quantum LDPC code families have a power-law scaling of the distance with n, $d \propto n^\alpha$, with $\alpha \leq 1/2$. Examples include code families in Refs. [15–18]; a single-qubit-encoding code family suggested in Ref. [38] has the distance scaling as $d \propto (n \log n)^{1/2}$.

An infinite quantum LDPC code family with sublinear power-law distance scaling has a finite error correction threshold, including the fault-tolerant case where the measured syndromes may have errors, as long as each stabilizer generator involves a limited number of qubits, and each qubit is involved in a limited number of stabilizer generators[39]. This makes quantum LDPC codes the only code family where finite rate is known to coexist with finite fault-tolerant error-correction threshold, potentially leading to substantial reduction of the overhead for scalable quantum computation[40].

Note that the quantum LDPC codes in Ref. [19] have finite rate and finite relative distance, at the price of stabilizer generator weight scaling like a power-law, $w \propto n^\gamma$, $\gamma \leq 1/2$; it is not known whether a fault-tolerant error-correction protocol exists for such codes.

An example of a large code family containing quantum LDPC codes is the hypergraph-product (HP) codes [15] generalizing the toric code. Such a code can be constructed from two binary matrices, $\mathcal{H}_1$ (dimensions $r_1 \times n_1$) and $\mathcal{H}_2$ (dimensions $r_2 \times n_2$), as a CSS code with the generator matrices [16]

$$\mathcal{G}_X = (E_2 \otimes \mathcal{H}_1, \mathcal{H}_2 \otimes E_1),\ \ \mathcal{G}_Z = (\mathcal{H}_2^T \otimes \widetilde{E}_1, \widetilde{E}_2 \otimes \mathcal{H}_1^T). \tag{19}$$

Here each matrix is composed of two blocks constructed as Kronecker products (denoted with "$\otimes$"), and E_1, $\widetilde{E}_1$, E_2, $\widetilde{E}_2$ are unit matrices of dimensions given by r_1, n_1, r_2 and n_2, respectively. Let us denote the parameters of classical codes using $\mathcal{H}_i$, $\mathcal{H}_i^T$ as parity check matrices, $\mathcal{C}^\perp_{\mathcal{H}_i} = [n_i, k_i, d_i]$, $\mathcal{C}^\perp_{\mathcal{H}_i^T} = [\tilde{n}_i, \tilde{k}_i, \tilde{d}_i]$, $i = 1, 2$, with the convention[15] that the distance $d = \infty$ if the corresponding $k = 0$. Then the parameters of the HP code are $n = n_2 r_1 + n_1 r_2$, $k = k_1 \tilde{k}_2 + \tilde{k}_1 k_2$ while the distance d satisfies[15] a lower bound $d \geq \min(d_1, d_2, \tilde{d}_1, \tilde{d}_2)$ and two upper bounds: if $\tilde{k}_2 > 0$, then $d \leq d_1$; if $\tilde{k}_1 > 0$, then $d \leq d_2$.

Particularly simple is the case when both binary codes are *cyclic*, with the property that all cyclic shifts of a code vector also belongs to the code[8]. A parity check matrix of such a code can be chosen *circulant*, with the first row using the coefficients of the *check* polynomial $h(x) \equiv c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}$ which is a factor of $x^n - 1$. Then, we can choose both circulant matrices $\mathcal{H}_1$ and $\mathcal{H}_2$ in Eq. (19) square $n_i \times n_i$, which gives a CSS code with the parameters $[[2n_1 n_2, 2k_1 k_2, \min(d_1, d_2)]]$. In particular, the toric codes[3, 41] are obtained when the circulant matrices $\mathcal{H}_1$, $\mathcal{H}_2$ are generated by the polynomial $h(x) = 1 + x$, with $k_i = 1$ and $d_i = n_i$, $i = 1, 2$.

III. STATISTICAL MECHANICS OF DECODING.

A. Maximum likelihood decoding

Let us consider one of the simplest error models, where the bit flip and phase flip errors happen independently and with equal probability p. The corresponding transformation of the single-qubit density matrix can be written as

$$\rho \mapsto p_I \rho + p_x X\rho X + p_y Y\rho Y + p_z Z\rho Z\,, \tag{20}$$

where $p_I = (1-p)^2$, $p_x = p_z = p(1-p)$, $p_y = p^2$. After relabeling the axes ($y \leftrightarrow z$) this can be interpreted in terms of the amplitude/phase damping model with some constraint on the decoherence times T_1, T_2. Our goal, however, is not to consider the most general case, but to construct a simple statistical mechanical model.

For the uncorrelated errors described by the completely-positive trace-preserving map (20), the probability of an error described by the binary vector $\mathbf{e} = (\mathbf{v}, \mathbf{u})$ (see the **Background** section) is

$$P(\mathbf{e}) = \prod_{i=1}^{N_\mathrm{b}} p^{e_i}(1-p)^{1-e_i} = p^w (1-p)^{N_\mathrm{b}-w}, \tag{21}$$

where $N_\mathrm{b} = 2n$ and $w \equiv \mathrm{wgt}(\mathbf{e}) = \mathrm{wgt}(\mathbf{v}) + \mathrm{wgt}(\mathbf{u})$ is the regular binary weight. Now, with a stabilizer code, all degenerate errors have the same effect and cannot be distinguished. Thus, one considers the net probability of an error having the same effect as $\mathbf{e}$,

$$P_0(\mathbf{e}) = \frac{1}{2^{N_g}} \sum_{\boldsymbol{\sigma}} p^w (1-p)^{N_\mathrm{b}-w}, \quad w \equiv \mathrm{wgt}(\mathbf{e} + \boldsymbol{\sigma} G), \tag{22}$$

where the generator matrix G (see Eq. (17)) has dimensions $N_s \times N_\mathrm{b}$ and non-zero $N_g \equiv N_s - \operatorname{rank} G$ allows G to have some linearly-dependent rows, cf. Eq. (2). The errors in Eq. (22) are exactly degenerate with $\mathbf{e}$ but they are not all the errors having the same syndrome as $\mathbf{e}$. It is thus convenient to introduce the probability of an error equivalent to $\mathbf{e}$ shifted by a codeword $\mathbf{c}$,

$$P_\mathbf{c}(\mathbf{e}) \equiv P_0(\mathbf{e} + \mathbf{c}), \tag{23}$$

and the total probability of an error with the syndrome $\mathbf{s} \equiv \widetilde{G}\mathbf{e}^T$,

$$P_\mathrm{tot}(\mathbf{s}) = \sum_\mathbf{c} P_\mathbf{c}(\mathbf{e}), \tag{24}$$

where $\mathbf{e}$ is any vector that gives the syndrome $\mathbf{s}$, and the summation is done over all 2^{2k} inequivalent codewords, length N_b zero-syndrome vectors, $\widetilde{G}\mathbf{c}^T = 0$, that are linearly independent from the rows of G, see the **Background** section. When combined with the summation over the degeneracy vectors generated by the rows of G, see Eqs. (22) and (23), the summation in Eq. (24) can be rewritten as that over all zero-syndrome vectors,

$$P_\mathrm{tot}(\mathbf{s}) = \sum_{\mathbf{x}:\widetilde{G}\mathbf{x}^T=0} p^w (1-p)^{N_\mathrm{b}-w}, \;\; w \equiv \mathrm{wgt}(\mathbf{e} + \mathbf{x}). \tag{25}$$

The probability (25) is normalized properly, so that the summation over all allowed syndrome vectors gives 1,

$$\sum_\mathbf{s} P_\mathrm{tot}(\mathbf{s}) = 1. \tag{26}$$

When decoding is done, only the measured syndrome $\mathbf{s}$ is known. For *maximum likelihood* (ML) decoding, the inferred error vector corresponds to the most likely configuration given the syndrome. To find it, we can start with some error configuration $\mathbf{e} \equiv \mathbf{e_s}$ corresponding to the syndrome $\mathbf{s}$, and find a codeword $\mathbf{c} = \mathbf{c}_\mathrm{max}(\mathbf{e})$ such that the corresponding equivalence class $\mathbf{e} + \mathbf{c}$ has the largest probability,

$$P_{\mathbf{c}_\mathrm{max}(\mathbf{e})}(\mathbf{e}) = P_\mathrm{max}(\mathbf{s}) \equiv \max_\mathbf{c} P_\mathbf{c}(\mathbf{e}). \tag{27}$$

Unlike the codeword $\mathbf{c}_\mathrm{max}(\mathbf{e})$ which depends on the choice of $\mathbf{e}$, the maximum probability $P_\mathrm{max}(\mathbf{s})$ depends only on the syndrome $\mathbf{s} \equiv \widetilde{G}\mathbf{e}^T$. The conditional probabilities of successful and of failed recovery given some unknown error with the syndrome $\mathbf{s}$ become

$$P_\mathrm{succ}(\mathbf{s}) = \frac{P_\mathrm{max}(\mathbf{s})}{P_\mathrm{tot}(\mathbf{s})}, \quad P_\mathrm{fail}(\mathbf{s}) \equiv 1 - P_\mathrm{succ}(\mathbf{s}). \tag{28}$$

The net probability of successful recovery averaged over all errors can be written as

$$P_\mathrm{succ} \equiv [P_\mathrm{succ}(\mathbf{s_e})] = \sum_\mathbf{s} P_\mathrm{max}(\mathbf{s}). \tag{29}$$

Here and in the following $[f(\mathbf{e})] \equiv \sum_\mathbf{e} P(\mathbf{e}) f(\mathbf{e})$ denotes the averaging over the errors with the probability (21). The result in the r.h.s. was obtained by partial summation over all errors with the same syndrome, cf. the syndrome probability (24).

Asymptotically successful recovery with probability one for an infinite family of QECCs implies that in the limit of large n, $P_\mathrm{succ} \to 1$ while $P_\mathrm{fail} \to 0$. Alternatively, in this limit Eqs. (28) and (29) give

$$\left[\frac{P_\mathrm{max}(\mathbf{s_e})}{P_\mathrm{tot}(\mathbf{s_e})}\right] \to 1. \tag{30}$$

Comparing Eqs. (26) and (29), we see that asymptotically, for each error that is likely to happen, the sum (24) is dominated by a single term with $\mathbf{c} = \mathbf{c}_\mathrm{max}(\mathbf{e})$. We can state this formally as

Lemma 1. *For an infinite family of quantum codes, successful decoding with probability one implies that asymptotically at large n, the ratio*

$$r(\mathbf{e}) \equiv \frac{P_\mathrm{max}(\mathbf{s_e})}{P_\mathrm{tot}(\mathbf{s_e})} = \frac{P_\mathrm{max}(\mathbf{e})}{\sum_\mathbf{c} P_\mathbf{c}(\mathbf{s_e})} \to 1.$$

for any error configuration $\mathbf{e}$ *likely to happen.*

Proof. Note that $r(\mathbf{e}) < 1$. Indeed, the summation in the denominator is over all $\mathbf{c}$, one of them equals $\mathbf{c}_{\max}(\mathbf{e})$ while the remaining terms are positive. Now, let us choose an arbitrarily small $\epsilon > 0$ and separate the errors into "good" where $1 - r(\mathbf{e}) < \epsilon$ and "bad" where $1 - r(\mathbf{e}) \geq \epsilon$. Use the following Bayesian expansion for the successful decoding probability:

$$P_{\text{succ}} = (1 - P_{\text{bad}}) \left[r(\mathbf{e})\right]_{\text{good}} + P_{\text{bad}} \left[r(\mathbf{e})\right]_{\text{bad}} \tag{31}$$

where the averaging in each term is limited to a particular type of errors as indicated. The first term can be bounded from above by $1 - P_{\text{bad}}$, while the second one by $P_{\text{bad}}(1-\epsilon)$, which gives

$$P_{\text{succ}} \leq 1 - \epsilon P_{\text{bad}}. \tag{32}$$

Since $P_{\text{fail}} = 1 - P_{\text{succ}} \to 0$ at large n, the probability P_{bad} can be made arbitrarily small by choosing large enough n. □

Generally, given an infinite family of codes, asymptotically certain recovery is possible with sufficiently small $p < p_c \leq 1/2$, as well in the symmetric region $p > 1 - p_c$, while it may not be a sure thing in the remaining interval $p_c \leq p \leq 1 - p_c$. This defines the ML decoding transition.

B. Random bond spin model

Given the well-established parallel between Wegner's models and binary codes[9, 10], it is straightforward to come up with a spin model matching the probabilities defined in the previous section. We use the binary error $\mathbf{e}$ to introduce the bond disorder using $J_b = (-1)^{e_b}$, and consider Wegner's partition function (2) with $\Theta = G$,

$$Z_0(\mathbf{e};\beta) \equiv \mathscr{Z}_{\mathbf{e},\mathbf{0}}(G, \{K_b = \beta\}). \tag{33}$$

The normalization is such that the probability in Eq. (22) is recovered on the Nishimori line (4),

$$P_0(\mathbf{e}) = Z_0(\mathbf{e};\beta_p), \quad e^{-2\beta_p} = p/(1-p). \tag{34}$$

To shorten the notations, we will omit the inverse temperature β whenever it is not likely to cause a confusion, $Z_0(\mathbf{e}) \equiv Z_0(\mathbf{e};\beta)$, and use $P_0(\mathbf{e})$ at the Nishimori line, $\beta = \beta_p$.

We also define the partition function with an *extended defect* of flipped bonds at the support of the codeword $\mathbf{c}$, $Z_{\mathbf{c}}(\mathbf{e};\beta) \equiv Z_{\mathbf{0}}(\mathbf{e}+\mathbf{c};\beta)$ [cf. Eq. (23)], the corresponding maximum $Z_{\max}(\mathbf{s};\beta) \equiv Z_{\mathbf{c}_{\max}}(\mathbf{e};\beta)$ [the maximum is reached at $\mathbf{c}_{\max} \equiv \mathbf{c}_{\max}(\mathbf{e};\beta)$ which may differ from that in Eq. (27) depending on the temperature], as well as an analog of $P_{\text{tot}}(\mathbf{s})$ [Eq. (24)],

$$Z_{\text{tot}}(\mathbf{s};\beta) = \mathscr{Z}_{\mathbf{e},\mathbf{0}}(\widetilde{G}^*, \{K_b = \beta\}), \tag{35}$$

where the binary matrix $\widetilde{G}^*$ is exactly dual to $\widetilde{G}$, namely $\widetilde{G}^* \widetilde{G}^T = 0$ and $\operatorname{rank} \widetilde{G} + \operatorname{rank} \widetilde{G}^* = N_{\text{b}}$ (cf. Eq. (1)), and we used the fact that $\widetilde{G}^*$ is a generating matrix for all vectors $\mathbf{x}$ in Eq. (25).

Except for disorder, the partition function (35) is related to Eq. (33) by Wegner's duality transformation [11],

$$\frac{2^{(N_g - N_s)/2} \mathscr{Z}_{\mathbf{e},\mathbf{0}}(\Theta, \{K\})}{\prod_b \sqrt{(\tanh K_b)^2 + 1}} = \frac{2^{(N_g^* - N_s^*)/2} \mathscr{Z}_{\mathbf{0},\mathbf{e}}(\Theta^*, \{K^*\})}{\prod_b \sqrt{(\tanh K_b^*)^2 + 1}}, \tag{36}$$

where bonds are defined by the columns of a $N_s^* \times N_{\text{b}}$ binary matrix Θ^* exactly dual to Θ, see Eqs. (1) and (2). The dual model has the same number of bonds, $N_{\text{b}}^* = N_{\text{b}}$, N_s^* spins, and its ground state degeneracy parameter $N_g^* = N_s^* - \operatorname{rank} \Theta^*$. The coupling parameters of mutually dual bonds are related by $\tanh K_b = \exp(-2K_b^*)$. The conjugation in Eq. (35) just rearranges the order of bonds and therefore leaves the partition/correlation function invariant, except for corresponding permutation of bond-specific variables: coupling parameters K_b and electric and magnetic charges,

$$\mathscr{Z}_{\mathbf{e},\mathbf{m}}(\widetilde{G}^*, \{K\}) = \mathscr{Z}_{\widetilde{\mathbf{e}},\widetilde{\mathbf{m}}}(G^*, \{\widetilde{K}\}). \tag{37}$$

We note in passing that the binary matrices Θ and Θ^* defining the mutually dual partition functions in Eq. (36) can be also thought of as the generating matrices of the two dual binary codes [Eq. (1)], with some additional linearly dependent rows. In fact, Wegner's duality has been long known in the coding theory as the MacWilliams identities between weight generating polynomials of dual codes[8, 42].

For a CSS code with the generator matrix in the form (18) the partition function (33) splits into a product of those for two non-interacting models corresponding to matrices $\mathcal{G}_X$ and $\mathcal{G}_Z$, see Eq. (3). In addition, two models defined by $\mathcal{G}_X$ and $\mathcal{G}_Z$ are dual to each other modulo logical operators. We can find the ground state degeneracies $2^{N_g^\mu}$, $\mu = X, Z$, of the corresponding models from $N_g^\mu = N_s^\mu - \operatorname{rank} \mathcal{G}_\mu$, where N_s^μ, $\mu = X, Z$ defines the number of rows in the matrix $\mathcal{G}_\mu$. For hypergraph-product codes in Eq. (19) the ground state degeneracy is given by[16] $N_g^X = \tilde{k}_1 \tilde{k}_2$ and $N_g^Z = k_1 k_2$.

Example 1. *For a CSS code with the check matrix (18), the partition function (33) is a product of those for two mutually decoupled spin models defined by matrices* $\Theta = \mathcal{G}_X$ *and* $\Theta = \mathcal{G}_Z$*, respectively, see Eq. (3). Since* $\mathcal{G}_X \mathcal{G}_Z^T = 0$*, in the absence of disorder these models are mutually dual, modulo logical operators.*

Example 2. *HP codes in Eq. (19) are CSS codes. In the special case* $\mathcal{H}_1 = \mathcal{H}_2^T$*, the matrices* $\mathcal{G}_X$ *and* $\mathcal{G}_Z$ *can be mapped to each other by permutations of rows and columns; the two spin models (3) are identical. In the absence of disorder both models are self-dual, modulo logical operators.*

Example 3. *Suppose matrices* $\mathcal{H}_1$ *and* $\mathcal{H}_2$ *in Eq. (19) are square and circulant, corresponding to two cyclic codes with generally different check polynomials* $h_1(x)$

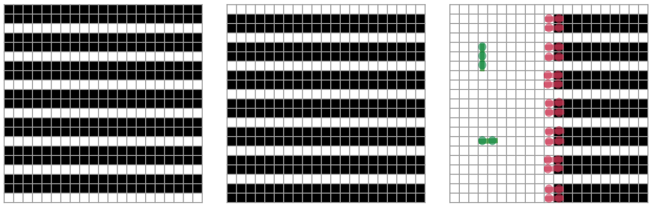

FIG. 1: Left and Center: two basis ground states of the spin model in Example 4, with black squares corresponding to flipped spins. An arbitrary ground state of this spin model is a linear combination of these two. Right: a domain wall between two such ground states. Green squares show the pattern of vertical and horizontal bonds involving interactions of two or three spins, respectively. A column of "unhappy" bonds forming the domain wall is shown with red.

and $h_2(x)$. Then the matrices $\mathcal{G}_X$ and $\mathcal{G}_Z$ can be mapped to each other by permutations of rows and columns, and thus in the absence of disorder the corresponding spin models (3) are self-dual modulo logical operators. This map is generally different from that in the previous example. This case has a nice layout on square lattice with periodic boundary conditions, with the horizontal and vertical bonds R_b in Eq. (2) formed according to the pattern of coefficients in the polynomials $h_1(x)$ and $h_2(x)$. In particular, with $h_1(x) = h_2(x) = 1 + x$, the hypergraph-product code is a toric code, while Eq. (3) gives two mutually decoupled Ising models.

Example 4. *Debierre and Turban [43] suggested a model that corresponds to a CSS code in the previous example with the check polynomials $h_1(x) = 1 + x$ and $h_2(x) = 1 + x + \ldots + x^{l-1}$ for some positive integer l. The two binary codes have $k_1 = 1$ (codewords are all-one or all-zero vectors), and, with n_2 divisible by l, $k_2 = l - 1$ (2^{l-1} codewords given by the repetitions of all length-l even-weight vectors). With $l = 3$, each of the two equivalent spin models (3) have four ground states in a pattern of stripes given by the repetitions of the vectors $[1, 1, 0]$, $[0, 1, 1]$, $[1, 0, 1]$ or $[0, 0, 0]$. A boundary between two distinct ground states produce a pattern of "unhappy" bonds that corresponds to an extended defect $\mathbf{c}$ in Eq. (24), see Fig. 1, Right.*

Example 5. *Spin models corresponding to quantum hypergraph-product codes $[[98s^2, 6s, 4s]]$, $s = 1, 2, \ldots$. The model is constructed from $7s \times 7s$ circulant matrices $\mathcal{H}_i$ corresponding to $h_i(x) = 1 + x + x^3$, $i = 1, 2$. A ground state of such a model is a linear combination of the nine basis states with the unit cell in Fig. 2, Left. Fig. 2, Right: a boundary between two ground states.*

C. Ordered state

In contrast to spin glass theory of classical binary codes where it is generally possible to apply a gauge transformation so that perfect decoding corresponds to a uniform magnetization [9, 10], this is not necessarily possible in the setting corresponding to a quantum code. For example, the models with the partition function (33) include those with exact $S \to -S$ symmetry. In such a case it appears natural to introduce the average spin as an order parameter. On the other hand, such a symmetry is not generic; the partition function (2) may not even have any degeneracy if Θ is a full-row-rank matrix. Also, except for the toric and related codes local in 2D [3], it is not at all clear what would be the relation of such an order parameter to the decoding transition in a given code.

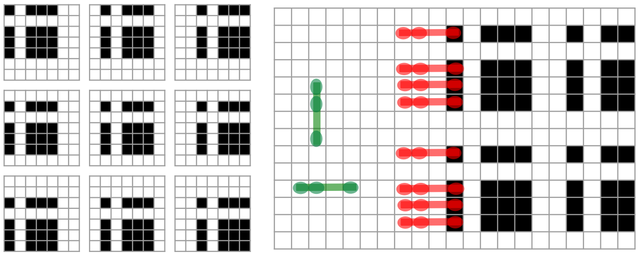

FIG. 2: Left: nine ground states of the spin model corresponding to the $\mathcal{G}_X$ matrix of the HP code (19) generated by circulant matrices $\mathcal{H}_i$ corresponding to $h_1(x) = h_2(x) = 1+x+x^3$, where $n_1 = n_2 = 21$ (they both must be factors of 7), see Example 5. An arbitrary ground state of the spin model is a linear combination of these nine states. Right: a domain wall formed between two such ground states. Green squares on white background show the patterns of horizontal and vertical bonds, each involves three spins. A column of "unhappy" bonds forming the extended defect is shown with red.

Here, we define an *ordered* phase as an analog of the region of parameters where asymptotically certain decoding is possible. We start with two definitions describing different phases:

Definition 1. *A* fixed-defect phase *of the spin glass model (33) corresponding to an infinite family of stabilizer codes has*

$$[Z_{\max}(\mathbf{s_e}; \beta)/Z_{\text{tot}}(\mathbf{s_e}; \beta)] \to 1, \quad n \to \infty. \tag{38}$$

Definition 2. *A* defect-free phase *of the spin glass model (33) corresponding to an infinite family of stabilizer codes has*

$$[Z_0(\mathbf{e}; \beta)/Z_{\text{tot}}(\mathbf{s_e}; \beta)] \to 1, \quad n \to \infty. \tag{39}$$

We note that analogs of Lemma 1 apply for the ratios in Eqs. (38) and (39). Thus, both in the fixed-defect and the defect-free phases, for any error $\mathbf{e}$ likely to happen, the partition function $Z_{\text{tot}}(\mathbf{s_e}; \beta)$ is going to be dominated by a single defect configuration, $\mathbf{c}_{\max}(\mathbf{e})$. In the defect-free phase, $\mathbf{c}_{\max}(\mathbf{e}) = \mathbf{0}$, while in a fixed-defect phase one may have a non-trivial defect $\mathbf{c}_{\max}(\mathbf{e}) \not\simeq \mathbf{0}$.

D. No fixed-defect phase on the Nishimori line

On the Nishimori line, the definition of a fixed-defect phase matches that of a region with asymptotically certain successful decoding, see Eq. (30). The latter region terminates at the decoding transition at the single-bit error probability $p = p_c$. On the other hand, the proof of the lower bound on the decoding threshold from Ref. [39] actually establishes the existence of a zero-defect phase on the Nishimori line, for small enough p. With both phases present, one would expect an additional transition between these phases at some $p < p_c$. Theorem 1 on p. 3 shows that this does not happen because there is no fixed-defect phase along the Nishimori line.

Proof. of Theorem 1. Below the decoding transition, $p < p_c$, according to Lemma 1, the probability $P_{\rm tot}(\mathbf{s})$ to obtain each likely syndrome is dominated by a single disorder configuration $\mathbf{e}_0(\mathbf{s})$. This is also the configuration most likely to happen, as opposed to any other configuration corresponding to the same syndrome. □

In comparison, for $\beta \neq \beta_p$, the disorder probability distribution $P_0(\mathbf{e})$ is different from the partition function $Z_0(\mathbf{e};\beta)$. In general, the dominant contribution to $Z_{\rm tot}(\mathbf{s}_\mathbf{e};\beta)$ may come from some other defect configuration $\mathbf{c}_{\rm max}(\mathbf{e};\beta) \not\simeq \mathbf{0}$.

In practical terms, when designing a decoding algorithm, we can concentrate on the portion of the free energy corresponding to $Z_\mathbf{0}(\mathbf{e};\beta_p)$ and ignore the possibility of any non-trivial defects without affecting the decoding probability in the limit of large n.

E. Free energy of a defect

In a fixed-defect phase: Let us introduce the free energy cost of flipping the bonds corresponding to an nonzero bits of the codeword $\mathbf{c}$ on top of the flipped bond pattern in the most likely configuration $\mathbf{c}_{\rm max}(\mathbf{e})$ corresponding to an error $\mathbf{e}$ with the syndrome $\mathbf{s} = \widetilde{G}\mathbf{e}^T$,

$$\Delta F_\mathbf{c}^{\rm max}(\mathbf{s};\beta) \equiv \beta^{-1} \log \frac{Z_{\rm max}(\mathbf{s})}{Z_{\mathbf{c}_{\rm max}(\mathbf{e})+\mathbf{c}}(\mathbf{e})}. \qquad (40)$$

Proof. of Theorem 2. In the fixed-defect phase each syndrome $\mathbf{s}$ likely to happen must be characterized by a unique configuration of defects, with the other configurations strongly suppressed. Version of Lemma 1 appropriate for this phase (see Def. 1) implies that $\Delta F_\mathbf{c}^{\rm max}(\mathbf{s};\beta) \to \infty$ asymptotically at large n. The corresponding disorder average must also diverge at large n. □

If we introduce the minimum weight $d_\mathbf{c}$ of a bit string in the degeneracy class of $\mathbf{c}$, $d_\mathbf{c} \equiv \min_{\boldsymbol{\sigma}} \mathop{\rm wgt}(\mathbf{c} + \boldsymbol{\sigma} G)$, we can formulate the following bounds

Lemma 2. *For any error* $\mathbf{e}$ *which gives the syndrome* $\mathbf{s}$*, any codeword* $\mathbf{c}$*, and any temperature* β^{-1}*,* $0 \leq \Delta F_\mathbf{c}^{\rm max}(\mathbf{s};\beta) \leq 2 d_\mathbf{c}$.

Proof. The lower bound follows trivially from the fact that $Z_{\rm max}(\mathbf{s})$ is the largest of $Z_\mathbf{c}(\mathbf{e})$. To prove the upper bound, use the Gibbs-Bogoliubov inequality in the form:

$$\beta^{-1} \log \frac{Z_0(\mathbf{e}')}{Z_\mathbf{c}(\mathbf{e}')} \leq \langle E_{\mathbf{c}+\mathbf{e}'} - E_{(\mathbf{e}')} \rangle = \sum_{b:\mathbf{c}_b \not\simeq \mathbf{0}} 2\langle (-1)^{\mathbf{e}'_b} R_b \rangle, \qquad (41)$$

where $\mathbf{e}' \equiv \mathbf{e} + \mathbf{c}_{\rm max}(\mathbf{e})$ is the same-syndrome disorder configuration such that the maximum is reached at $\mathbf{c} = \mathbf{0}$, $E_\mathbf{e} \equiv \sum_b (-1)^{e_b} R_b$ is the energy of a spin configuration, see Eq. (2), and the averaging is done over all spin configurations contributing to $Z_0(\mathbf{e}';\beta)$. Each term in the r.h.s. of Eq. (41) is uniformly bounded from above, $2(-1)^{\mathbf{e}_b} R_b \leq 2$; this gives $\Delta F_\mathbf{c}^{\rm max}(\mathbf{e};\beta) \leq 2 \mathop{\rm wgt} \mathbf{c}$. Minimizing over the vectors degenerate with $\mathbf{c}$ gives the stated result. □

Note that at zero temperature and in the absence of disorder, $\mathbf{e} = \mathbf{0}$, the upper bound in Lemma 2 is saturated. We conjecture that a similar asymptotic scaling, with some finite

$$\lambda_\mathbf{c} \equiv \frac{[\Delta F_\mathbf{c}^{\rm max}(\mathbf{s}_\mathbf{e};\beta)]}{d_\mathbf{c}}, \qquad (42)$$

should be valid for the free energy increments averaged over disorder, with the defect *tension* $\lambda_\mathbf{c}$ analogous to the domain wall tension in the 2D Ising model. In the fixed-defect phase, where $\Delta F_\mathbf{c}$ is expected to diverge, we thus expect the tensions (42) to be non-zero, $\lambda_\mathbf{c} > 0$.

In the defect-free phase: In such a phase, the total partition function (35) is entirely dominated by that without any extended defects, see Eq. (33). Instead of Eq. (40), it is convenient to consider the free energy increment for flipping the bonds corresponding to the codeword $\mathbf{c}$ starting with a given defect configuration $\mathbf{e}$,

$$\Delta F_\mathbf{c}^{(0)}(\mathbf{e};\beta) \equiv \beta^{-1} \log \frac{Z_0(\mathbf{e};\beta)}{Z_\mathbf{c}(\mathbf{e};\beta)}. \qquad (43)$$

Similar to the upper bound in Lemma 2, we can state

$$\Delta F_\mathbf{c}^{(0)}(\mathbf{e};\beta) \leq 2 d_\mathbf{c}; \qquad (44)$$

however, the corresponding lower bound might be violated for some disorder configurations $\mathbf{e}$ where $\mathbf{c}_{\rm max}(\mathbf{e}) \not\simeq \mathbf{0}$. In the defect-free phase, the total probability of such configurations, $P_{\rm defect}$, as well as the configurations where $F_\mathbf{c}^{(0)}(\mathbf{e};\beta)$ remains bounded, $P_{\rm finite}$, should be vanishingly small at large n, $P_{\rm defect} + P_{\rm finite} \to 0$. The corresponding bounds can be readily formulated by analogy with Lemma 1. As a result, while in general the increments in Eqs. (40) and (43) have both the initial and the final states different and cannot be easily compared, in the defect-free phase the corresponding averages should coincide asymptotically at $n \to \infty$. In particular, this implies $[\Delta F_\mathbf{c}^{(0)}(\mathbf{e};\beta)] \to \infty$ at large n in the defect-free phase.

On the Nishimori line: According to Theorem 1, the only ordered phase at the Nishimori line is the defect-free phase. This immediately gives Corollary 1.

On the Nishimori line, it is convenient to consider the free energy $\Delta F_{\mathbf{c}}(\mathbf{s};\beta)$ of a defect $\mathbf{c}$ averaged over the errors $\mathbf{e}$ with the same syndrome, $\mathbf{s}=\widetilde{G}\mathbf{e}^T$,

$$\Delta F_{\mathbf{c}}(\mathbf{s};\beta)\equiv\left[\Delta F_{\mathbf{c}}^{(0)}(\mathbf{e};\beta)\right]_{\mathbf{s}}, \tag{45}$$

where the average is extended over all non-equivalent codewords $\mathbf{c}$,

$$[f(\mathbf{e})]_s\equiv\sum_{\mathbf{c}}\frac{P_0(\mathbf{e}+\mathbf{c})}{P_{\rm tot}(\mathbf{s})}f(\mathbf{e}+\mathbf{c}). \tag{46}$$

For the average (45), we prove the following version of Lemma 2:

Lemma 3. *At the Nishimori line, for every allowed syndrome* $\mathbf{s}$ *and every codeword* $\mathbf{c}$*, the free energy averaged over the errors with the same syndrome satisfies* $0\le\Delta F_{\mathbf{c}}(\mathbf{s};\beta_p)\le 2d_{\mathbf{c}}$.

Proof. The upper bound is trivial since it applies for every term in the average, see Eq. (44). The lower bound follows from the Gibbs inequality. Explicitly, introduce two normalized distribution functions of codewords $\mathbf{b}$: $f_{\mathbf{b}}\equiv P_{\mathbf{0}}(\mathbf{e}')/P_{\rm tot}(\mathbf{s})$, $g_{\mathbf{b}}\equiv P_{\mathbf{c}}(\mathbf{e}')/P_{\rm tot}(\mathbf{s})$, where $\mathbf{e}'\equiv\mathbf{e}+\mathbf{b}$; then, using the map (34) on the Nishimori line,

$$\beta\Delta F_{\mathbf{c}}(\mathbf{s};\beta_p)=\sum_{\mathbf{b}}f_{\mathbf{b}}\log\frac{f_{\mathbf{b}}}{g_{\mathbf{b}}}\ge\sum_{\mathbf{b}}f_{\mathbf{b}}\left(1-\frac{g_{\mathbf{b}}}{f_{\mathbf{b}}}\right)=0,$$

where the summation is done over all non-equivalent codewords $\mathbf{b}$ and we used $\log(x)\ge 1-1/x$. □

Note that this Lemma gives an alternative proof of Theorem 1.

F. Self-averaging

Conditions of Theorem 2 guarantee that the disordered system is not in a spin glass phase. A self-averaging for the partition functions $Z_{\mathbf{c}}(\mathbf{e};\beta)$ would immediately imply the statement of the theorem. Note however, that (**i**) in the presence of disorder self-averaging is not expected for the partition function even in the case of the toric codes as fluctuations could be exponentially large, and (**ii**) spin models corresponding to general families of quantum codes, whether LDPC or not, are expected to involve highly non-local interactions. Thus, without additional conditions, one cannot guarantee self-averaging even for the free energy.

However, we did not rely on self-averaging in any of the proofs. In particular, results in this section apply to spin models corresponding to finite-rate quantum hypergraph-product and related codes[15, 16] that can be obtained from random binary LDPC codes:

Example 6. *This is a special case of the model in Example 2. Consider* a random *binary matrix* $\mathcal{H}$ *with* h *non-zero entries per row and* v *per column, with* $h<v$*, e.g., see Ref. [33]. The rate of the corresponding binary code* $\mathcal{C}_{\mathcal{H}}^{\perp}$ *with parameters* $[n_{\rm c},k_{\rm c},d_{\rm c}]$ *is limited,* $R_{\rm c}\equiv k_{\rm c}/n_{\rm c}\ge 1-h/v$*. With high probability at large* $n_{\rm c}$*, the classical code will have the relative distance in excess of* $\delta_{\rm c}\equiv\delta_{\rm c}(h,v)$ *given in Ref. [33]. Such an* $[n_{\rm c},k_{\rm c},d_{\rm c}]$ *code produces a quantum HP code (19) with* $\mathcal{H}_1=\mathcal{H}_2^T=\mathcal{H}$*, which is a quantum LDPC code with the asymptotic rate* $k/n\ge(v-h)^2/(h^2+v^2)$ *and the distance scaling as* $d/\sqrt{n}=\delta_{\rm c}v/\sqrt{h^2+v^2}$*. Such a code has a decoding transition at a finite* p*, see Ref. [39]. Our present results indicate that each of the corresponding spin models (3) has non-local bonds involving up to* v *spins, exponentially large number of mutually inequivalent extended defects, and an ordered state where such defects do not appear. In addition, as already stated in Example 2, the two models are self-dual modulo logical operators.*

IV. PHASE TRANSITIONS

A. Transition to a disordered phase

Transition mechanism: An ordered phase (whether fixed-defect or defect-free) of the model (35) is characterized by a unique defect pattern $\mathbf{c}_{\rm max}(\mathbf{e})$ for every likely configuration of flipped bonds $\mathbf{e}$. In the case of a code family where k remains fixed, for the stability of such a phase it is sufficient that non-trivial defects $\mathbf{c}\not\simeq\mathbf{0}$ have divergent free energies, as in Theorem 2. On the other hand, defects can proliferate if at least one of the free energies $\Delta F_{\mathbf{c}}^{\rm max}$ remains bounded in the asymptotic $n\to\infty$ limit.

The situation is different in the case of a code family with divergent k, e.g., with fixed rate $R\equiv k/n$, as in Example 6. Here, the number of different defects, $2^{2k}-1$, diverges exponentially at large n; in an ordered phase the free energies of individual defects must be large enough to suppress this divergence. This implies, in particular, that for a typical defect the tension (42) must exceed certain limit. The statement of Theorem 3 concerns the corresponding average tension,

$$\overline{\lambda}\equiv(2^{2k}-1)^{-1}\sum_{\mathbf{c}\not\simeq\mathbf{0}}\lambda_{\mathbf{c}}. \tag{47}$$

Proof. of Theorem 3. Let us start with a version of Lemma 1 for the fixed-defect phase (Def. 1): for any likely disorder configuration $\mathbf{e}$,

$$\sum_{\mathbf{c}\not\simeq\mathbf{0}}\frac{Z_{\mathbf{c}+\mathbf{c}_{\rm max}(\mathbf{e})}(\mathbf{e};\beta)}{Z_{\rm max}(\mathbf{s}_{\mathbf{e}};\beta)}\to 0, \tag{48}$$

asymptotically at $n\to\infty$. Note that we cannot just average this expression term-by-term, since unlikely errors could potentially dominate the sum which involves an exponentially large number of terms. Instead, we fix some

$\epsilon > 0$ and first consider the average of Eq. (48) only over the "good" errors where the sum does not exceed ϵ. Using the standard inequality $\exp\langle f\rangle \le \langle \exp f\rangle$, we obtain the following expression involving the averages of the free energies (40) over "good" errors only:

$$\sum_{\mathbf{c}\not\simeq\mathbf{0}} \exp\left(-\beta[\Delta F_{\mathbf{c}}^{\max}(\mathbf{s}_{\mathbf{e}};\beta)]_{\rm good}\right) \le \epsilon. \qquad (49)$$

Rewriting this sum in terms of an average over non-trivial defects which we denote as $\langle\,\cdot\,\rangle_{\mathbf{c}\not\simeq\mathbf{0}}$, and using the same inequality, we get

$$(2^{2k}-1)\exp\left(-\beta\left\langle[\Delta F_{\mathbf{c}}^{\max}(\mathbf{s}_{\mathbf{e}};\beta)]_{\rm good}\right\rangle_{\mathbf{c}\not\simeq\mathbf{0}}\right) \le \epsilon. \qquad (50)$$

It is convenient to introduce an analog of the tension (42) for finite ϵ,

$$\lambda_{\mathbf{c}}^{(\epsilon)} \equiv \frac{[\Delta F_{\mathbf{c}}^{\max}]_{\rm good}}{d_{\mathbf{c}}}, \qquad (51)$$

along with the corresponding average $\overline{\lambda}_{(\epsilon)}$ over non-trivial defects $\mathbf{c} \not\simeq \mathbf{0}$, defined as in Eq. (47). According to Lemma 2, each of the tensions satisfy $0 \le \lambda_{\mathbf{c}}^{(\epsilon)} \le 2$, which means the same bounds for the defects-average, $0 \le \overline{\lambda}_{(\epsilon)} \le 2$. With the help of the trivial upper bound $d_{\mathbf{c}} \le N_{\rm b} = 2n$, Eq. (50) gives

$$(2^{2k}-1)\exp(-2n\beta\overline{\lambda}^{(\epsilon)}) \le \epsilon, \qquad (52)$$

which implies for large n, k

$$\beta\overline{\lambda}^{(\epsilon)} \ge \frac{k}{n}\log 2 = R\log 2. \qquad (53)$$

We can now introduce the full average tension $\overline{\lambda}$ which involves both "good" and "bad" errors by writing a Bayesian expansion similar to Eq. (31). The key observation leading to the statement of the Theorem is that the contribution of "bad" errors disappears in the large-n limit since for each error configuration the tension is limited, while the total probability of "bad" errors $P_{\rm bad} \to 0$. □

As a consequence, for any code family with a finite rate R, we expect one of the two possibilities at the transition to a disordered phase: (**i**) Transition driven by proliferation of some (e.g., finite) subset of the defects whose tensions $\lambda_{\mathbf{c}}$ vanish at the transition, with the average in Theorem 3 still finite; and (**ii**) Transition driven by the entropy of some macroscopic number of the defects, in which case tensions of all defects remain bounded at the transition, $\lambda_{\mathbf{c}} \ge \lambda_0 > 0$. In the case (**i**), one gets to a phase with "limited disorder" where only some of all possible defects $\mathbf{c}$ may happen with non-zero probability at large n.

Continuity of the transition: At the Nishimori line, the average energy is known exactly[10, 12, 13], it is a continuous function of parameters. This guarantees the continuity of the decoding transition. The same conclusion can be drawn from the bound (14) on the specific heat along the Nishimori line—the derivation is identical to the standard case[10, 12–14].

On the other hand, away from the Nishimori line, the transition from an ordered to a disordered phase can be (and often is) discontinuous. In particular, mean field analysis using the TAP equations (named for Thouless, Anderson, and Palmer, see Ref. [44]) generically gives a discontinuous transition for local magnetization whenever the bonds R_b couple more than two spins.

Self-duality in the absence of disorder: In the absence of errors, we can use Wegner's duality (36) to relate the partition functions of the models with the generator matrices G and G^*, that is, Eqs. (33) and (35), since the matrices G^* and $\widetilde{G}^*$ differ by an inessential permutation of columns (bonds). Assuming the transition is unique, whether continuous or not, it must happen at the self-dual point, $\sinh(2\beta_{\rm s.d.}) = 1$. Here Eq. (36) gives $Z_{\rm tot}(\mathbf{0};\beta_{\rm s.d.}) = 2^k Z_0(\mathbf{0};\beta_{\rm s.d.})$, or, equivalently,

$$\sum_{\mathbf{c}\not\simeq\mathbf{0}} e^{-\beta_{\rm s.d.}\Delta F_{\mathbf{c}}^{(0)}(\mathbf{0};\beta_{\rm s.d.})} = 2^k - 1. \qquad (54)$$

This equation is exact since no disorder is involved. The summation over $\mathbf{c}$ here includes $2^{2k}-1$ terms, and the result is independent of the distance of the code. For a finite-R code family, arguments similar to those in the proof of Theorem 3 give a lower bound $\overline{\lambda}_{\rm s.d.} \ge (R/2)\ln 2$, which is smaller by half of the corresponding bound deep inside an ordered phase.

Location of the multicritical point: In many types of local spin glasses on self-dual lattices the transition from the ordered phase on the Nishimori line happens at a multicritical point whose location to a very good accuracy has been predicted by the strong-disorder self-duality conjecture[22–29]. In case of the Ising spin glasses, the corresponding critical probability $p_c \approx 0.110$ satisfies Eq. 15. The derivation of this expression[22] uses explicitly only the probability distribution of allowed energy values for a single bond. Our limited simulations indicate that for several quasi-local models (see Example (3)) with finite k the multicritical point is indeed located at $p_c \approx 0.11$, also very close to the Gilbert-Varshamov existence bound for zero-rate codes. However, for code families with finite rates k/n, see Example 6, the threshold probability must be below the Shannon limit (16), which means the self-duality conjecture must be strongly violated for $R > 1/2$.

B. Transition between defect-free and fixed-defect phases

Theorem 1 states that on the Nishimori line below the decoding transition the spin model (33) is in the defect-free phase. If a distinct fixed-defect phase exists somewhere on the phase diagram, there is a possibility for a transition between these phases.

More generally, defect-free phase is a special case of an ordered fixed-defect phase. One can imagine a transitions between two such phases. However, at least in the case of a temperature-driven transition, the spin model (33) must become disordered at the transition point. Indeed, for a transition to happen at $T = T_0(p)$, at least for some of the likely disorder configurations, for $T < T_0(p)$, $Z_{\mathbf{c}_1}(\mathbf{e};\beta)$ must dominate, while for $T > T_0(p)$, some of errors $\mathbf{e}$ will be dominated by $Z_{\mathbf{c}_2}(\mathbf{e};\beta)$ with $\mathbf{c}_2 \not\simeq \mathbf{c}_1$. This implies that at the actual transition point some codewords must become degenerate with non-zero probability, which would violate the condition in Def. 1. Once the system becomes disordered at some p, one would generically expect it to remain disordered at larger p. By this reason, we expect that non-trivial fixed-defect phases are not common.

C. Absence of a local order parameter

In Examples 1 to 6 we considered some spin models which do not have any gauge-like symmetries. However, the same approach can be also used to construct non-local spin models which have "local" gauge symmetries and at the same time highly non-trivial phase diagrams.

The following example is a generalization of the mutually dual three-dimensional Ising model and a random plaquette $\mathbb{Z}_2$ gauge.

Example 7. *Consider a CSS code (18) with the generators:*

$$\mathcal{G}_X = (E_1 \otimes G, \quad R \otimes E_2), \tag{55}$$

$$\mathcal{G}_Z = \begin{pmatrix} R \otimes \widetilde{E}_2, & \widetilde{E}_1 \otimes G \\ E_1 \otimes \widetilde{G}, & 0 \end{pmatrix}, \tag{56}$$

where R is a square circulant matrix corresponding to the polynomial $h(x) = 1+x$ and $G \equiv (G_X, G_Z)$ is the generator matrix (17) of an arbitrary quantum code. This construction follows the hypergraph-product code construction (19), and the unit matrices E_1, $\widetilde{E}_1$, E_2, $\widetilde{E}_2$ are chosen accordingly. The additional block involving the conjugate matrix $\widetilde{G} = (G_Z, G_X)$ differentiates this construction from the hypergraph-product code construction. This code defines two non-interacting, mutually dual spin models (3). In particular, when G corresponds to a toric code, we recover a three dimensional Ising model for $\mu = X$, and a three dimensional random plaquette $\mathbb{Z}_2$ gauge model for $\mu = Z$.

A spin model with a local gauge symmetry cannot have a local order parameter[11]. Thus, one cannot hope to construct a local order parameter that would describe the transition from a defect-free phase and be applicable to all of the models (33).

The same result can be obtained by noticing that the transition from the defect-free phase can be driven by delocalization of any of $2^{2k} - 1$ non-trivial defects. For a finite-R code family this number scales exponentially with n; we find it not likely that an order parameter defined locally can distinguish this many possibilities.

D. Spin correlation functions

The average of any product of spin variables which cannot be expressed as a product of the bond variables in the Hamiltonian is zero [11]. Thus, we consider two most general non-trivial spin correlation functions:

$$Q^{\mathbf{m}}_{\text{tot}}(\mathbf{e};\beta) \equiv \frac{\mathscr{Z}_{\mathbf{e},\mathbf{m}}(\widetilde{G}^*; \{K_b = \beta\})}{\mathscr{Z}_{\mathbf{e},\mathbf{0}}(\widetilde{G}^*; \{K_b = \beta\})}, \tag{57}$$

$$Q^{\mathbf{m}}_{\mathbf{c}}(\mathbf{e};\beta) \equiv \frac{\mathscr{Z}_{\mathbf{e}+\mathbf{c},\mathbf{m}}(G; \{K_b = \beta\})}{\mathscr{Z}_{\mathbf{e}+\mathbf{c},\mathbf{0}}(G; \{K_b = \beta\})}; \tag{58}$$

both correlation functions satisfy $-1 \le Q^{\mathbf{m}}(\mathbf{e};\beta) \le 1$. The thermal average in Eq. (58) corresponds to summation over spin configurations in $Z_{\mathbf{c}}(\mathbf{e};\beta)$, while that in Eq. (57) corresponds to the same defect and spin configurations that enter $Z_{\text{tot}}(\mathbf{s};\beta)$, cf. Eq. (35). Using the explicit form (2), definitions of Z_{tot} and $Z_{\mathbf{c}}$, and the fact that additional linearly-independent rows in $\widetilde{G}^*$ form a basis of non-equivalent codewords $\mathbf{c}$, we can write the following expansion

$$Q^{\mathbf{m}}_{\text{tot}}(\mathbf{e};\beta) = \sum_{\mathbf{c}} (-1)^{\mathbf{c}\cdot\mathbf{m}} \frac{Z_{\mathbf{c}}(\mathbf{e};\beta)\, Q^{\mathbf{m}}_{\mathbf{c}}(\mathbf{e};\beta)}{Z_{\text{tot}}(\mathbf{s}_{\mathbf{e}};\beta)}. \tag{59}$$

The correlation functions contain the products of $\prod_b R_b^{m_b} = \prod_r (S_r)^{G_{rb} m_b}$, or the product of spin variables in the support of the syndrome vector $\mathbf{s}_{\widetilde{\mathbf{m}}} \equiv G\mathbf{m}^T = \widetilde{G}\widetilde{\mathbf{m}}^T$ corresponding to $\mathbf{m}$. Thus, the defined correlation functions are trivially symmetric with respect to any gauge symmetries, $S_r \to S_r(-1)^{\alpha_r}$, $\boldsymbol{\alpha} G = 0$ (present whenever there are $N_g > 0$ linearly dependent rows of G), as well as the transformations of $\mathbf{m}$ leaving the syndrome invariant, $\mathbf{m} \to \mathbf{m} + \boldsymbol{\gamma}\widetilde{G}$.

Wilson loop: In lattice gauge theory, in the absence of a local order parameter, the deconfining transition can be characterized by the average of the Wilson loop operator[45], with the thermal and disorder average scaling down as an exponent of the area in the high-temperature phase, and an exponent of the perimeter in the low-temperature phase. In the case of the three-dimensional $\mathbb{Z}_2$ gauge model[11, 46], see Example 7, the corresponding correlator is a product of plaquette operators covering certain surface. The correlation function (58) is a natural generalization to non-local Ising models, with the minimum weight $\mathbf{d}_{\mathbf{m}} \equiv \min_{\boldsymbol{\gamma}} \operatorname{wgt}(\mathbf{m} + \boldsymbol{\gamma}\widetilde{G})$ of $\mathbf{m}$ corresponding to the area, and the binary weight of the syndrome $\mathbf{s}_{\widetilde{\mathbf{m}}}$ corresponding to the perimeter. Indeed, taking $\mathbf{e} = \mathbf{c} = \mathbf{0}$, at high temperatures, independent bond variables R_b fluctuate independently, and one can write $Q^{\mathbf{m}}_{\mathbf{0}}(\mathbf{0};\beta) = \langle \prod R_b^{m_b} \rangle \propto \beta^{d_{\mathbf{m}}}$, which corresponds to the area law. The same quantity at low temperatures can be evaluated in leading order by substituting average spin

$S_b \to \langle S_b \rangle \sim M$, with the result $Q_{\mathbf{0}}^{\mathbf{m}}(\mathbf{0};\beta) \propto M^{\operatorname{wgt} \mathbf{s}_{\widetilde{\mathbf{m}}}}$, the perimeter law. We expect such a behavior to persist in a finite range of temperatures below the transition from the ordered phase, at least in the case of LDPC codes.

However, in general there is no guarantee that the spin model (33) has a unique transition, and the functional form of the spin correlation function (58) with generic $\mathbf{m}$ cannot be easily found at intermediate temperatures. By this reason, it remains an open question whether the scaling of the analog of the Wilson loop can be used to distinguish between specific disordered phases.

Indicator correlation functions. Consider the correlation function (59) for $\mathbf{m}$ such that the corresponding syndrome is zero $\mathbf{s}_{\widetilde{\mathbf{m}}} = \mathbf{0}$. Then the spin products in each term of the expansion disappear, and $Q_{\mathbf{c}}^{\mathbf{m}}(\mathbf{e};\beta) = 1$ for any $\mathbf{c}$. The corresponding $\mathbf{m}$ are just the dual codewords $\widetilde{\mathbf{b}}$. In general, for a pair of codewords $\mathbf{b}$, $\mathbf{c}$, the scalar product $\mathbf{c} \cdot \widetilde{\mathbf{b}} = 0$ iff the corresponding logical operators commute, see the **Background** section. For each codeword $\mathbf{c} \not\simeq \mathbf{0}$ there is at least one codeword $\mathbf{c}'$ such that $\mathbf{c} \cdot \widetilde{\mathbf{c}}' = 1$, and the $2k$ scalar products $\mathbf{c} \cdot \widetilde{\mathbf{b}}$ with the basis codewords $\mathbf{b}$ are sufficient to recover the equivalence class of $\mathbf{c}$.

We further note that in the defect-free phase, for any likely disorder $\mathbf{e}$, $Z_{\text{tot}}(\mathbf{s}_{\mathbf{e}};\beta)$ is dominated by the term with $\mathbf{c} = 0$, thus at large n the average $[Q_{\text{tot}}^{\widetilde{\mathbf{b}}}(\mathbf{s}_{\mathbf{e}};\beta)] = 1$ for any codeword $\mathbf{b}$. Similarly, in a fixed-defect phase, there is only one dominant term $Z_{\mathbf{c}}(\mathbf{e};\beta)$, and $[Q_{\text{tot}}^{\widetilde{\mathbf{b}}}(\mathbf{s}_{\mathbf{e}};\beta)] = \pm 1$; the patterns of signs for different $\mathbf{b}$ can be used to find out which of the codewords $\mathbf{c}$ dominates the partition function.

E. Bound on the location of the defect-free phase

In order to prove the Theorem 4, we first need to extend identities of Nishimori's gauge theory of spin glasses[10, 12, 47] to the averages of the spin correlation functions (57). We prove the following

Lemma 4. *The disorder average of the spin correlation function (57) for any* $\mathbf{m}$ *satisfies* $[Q_{\text{tot}}^{\mathbf{m}}(\mathbf{e};\beta)] = [Q_{\text{tot}}^{\mathbf{m}}(\mathbf{e};\beta)\, Q_{\text{tot}}^{\mathbf{m}}(\mathbf{e};\beta_p)]$.

Proof. Follows exactly the proof in the usual case, if we observe

$$\sum_{\boldsymbol{\alpha}} P_0(\mathbf{e} + \boldsymbol{\alpha} \widetilde{G}^*) = 2^{N_r - N_g + N_g^*} Z_{\text{tot}}(\mathbf{s}_{\mathbf{e}};\beta_p),$$

where N_r is the number of rows of the matrix G. □

Proof. of Theorem 4. To shorten the notations, denote the correlation function in Lemma 4 as $A \equiv Q_{\text{tot}}^{\mathbf{m}}(\mathbf{e};\beta)$ and B the same correlation function at the Nishimori temperature, $\beta = \beta_p$. Lemma 4 gives

$$[A] = [AB], \quad [B] = [B^2]. \tag{60}$$

Now, for any real-valued t, the inequality

$$0 \le [(A - tB)^2] = [A^2] + t^2[B^2] - 2t[AB] \tag{61}$$

must be valid. This is equivalent to $[AB]^2 \le [A^2][B^2]$. Using the identities (60), we obtain $[A]^2 \le [A^2][B] \le [B] = [B^2]$, which is equivalent to

$$[Q_{\text{tot}}^{\mathbf{m}}(\mathbf{e};\beta)]^2 \le [Q_{\text{tot}}^{\mathbf{m}}(\mathbf{e};\beta_p)]. \tag{62}$$

A different derivation of this inequality can be found in Ref. [48]. If sum both sides of Eq. (62) over all dual codewords $\mathbf{m} = \widetilde{\mathbf{c}}$, using the expansion (59), we obtain

$$\sum_{\mathbf{c}} [Q_{\text{tot}}^{\mathbf{m}=\widetilde{\mathbf{c}}}(\mathbf{e};\beta)]^2 \le 2^{2k} \left[\frac{Z_0(\mathbf{e};\beta_p)}{Z_{\text{tot}}(\mathbf{s}_{\mathbf{e}};\beta_p)} \right]. \tag{63}$$

The r.h.s. equals exactly the average probability of successful decoding times 2^{2k}; for large n it equals 2^{2k} below the decoding transition, $p < p_c$, and it is smaller than 2^{2k} above the decoding transition. On the other hand, we saw that in in the defect-free phase, at large n, all correlation functions $[Q_{\text{tot}}^{\widetilde{\mathbf{m}}}(\mathbf{e};\beta)] = 1$. According to Eq. (63), this is only possible for $p < p_c$. □

This implies that the phase boundary below the Nishimori line is either vertical or reentrant as a function of temperature. Recent numerical studies suggest that the second option is true for the random bond Ising model[49].

V. CONCLUDING REMARKS

In this work we considered spin glass models related to the decoding transition in stabilizer error correcting codes. Generally, these are non-local models with multi-spin couplings, with exact Wegner-type self-duality at zero disorder, but no $S \to -S$ symmetry or other sources of ground state degeneracy. Nevertheless, we show that for models corresponding to code families with maximum-likelihood decoding (ML) transition at a finite bit error probability p_c, there is a region of an ordered phase which must be limited to $p \le p_c$, and a line of non-trivial phase transitions.

The models support generally non-topological extended defects which generalize the notion of domain walls in local spin models. For a quantum code that encodes k qubits, there are $2^{2k} - 1$ different types of extended defects. A disordered phase is associated with proliferation of at least one of such defects. In an ordered phase, the free energy of each defect must diverge at large n. Moreover, for a code family with finite rate k/n, the average defect tension, an analog of domain wall line tension, must exceed some finite threshold (Theorem 3).

The original decoding problem corresponds to the Nishimori line at the phase diagram of the disordered spin model, with the maximum-likelihood (ML) decoding transition located exactly at the multicritical point of the spin model. The ML decoding threshold is the maximum possible threshold for any decoder. Thus, exploring this connection with statistical mechanics of spin glasses, one can compare codes irrespectively of the decoder efficiency, and get an absolute measure of performance for any given, presumably suboptimal, decoder.

There are a number of open question in relation to the models we studied. In particular, is there some sort of universality for transitions with nonlocal spin couplings? If yes, what determines the universality class, and is there an analog of the hyperscaling relation?

Acknowledgments

This work was supported in part by the U.S. Army Research Office under Grant W911NF-11-1-0027, by the NSF under Grant 1018935, and Central Facilities of the Nebraska Center for Materials and Nanoscience supported by the Nebraska Research Initiative. We acknowledge hospitality of the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center with support of the Gordon and Betty Moore Foundation. We also thank Professor Hidetoshi Nishimori for useful comments on early version of the manuscript.

[1] D. Gottesman, Ph.D. thesis, Caltech (1997), URL `http://arxiv.org/abs/quant-ph/9705052`.

[2] A. R. Calderbank, E. M. Rains, P. M. Shor, and N. J. A. Sloane, IEEE Trans. Inf. Th. **44**, 1369 (1998), URL `http://dx.doi.org/10.1109/18.681315`.

[3] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, J. Math. Phys. **43**, 4452 (2002), URL `http://dx.doi.org/10.1063/1.1499754`.

[4] A. J. Landahl, J. T. Anderson, and P. R. Rice (2011), unpublished, arXiv:1108.5738.

[5] H. G. Katzgraber and R. S. Andrist (2013), unpublished, arXiv:1306.0540, URL `http://arxiv.org/abs/1306.0540`.

[6] M. S. Postol (2001), unpublished, arXiv:quant-ph/0108131v1, URL `http://arxiv.org/abs/quant-ph/0108131`.

[7] D. J. C. MacKay, G. Mitchison, and P. L. McFadden, IEEE Transactions on Information Theory **59**, 2315 (2004), URL `http://dx.doi.org/10.1109/TIT.2004.834737`.

[8] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes* (North-Holland, Amsterdam, 1981).

[9] N. Sourlas, Nature **339**, 693 (1989).

[10] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Clarendon Press, Oxford, 2001).

[11] F. Wegner, J. Math. Phys. **2259**, 12 (1971), URL `http://dx.doi.org/10.1063/1.1665530`.

[12] H. Nishimori, Progress of Theoretical Physics **66**, 1169 (1981), http://ptp.oxfordjournals.org/content/66/4/1169.full.pdf+ht URL `http://ptp.oxfordjournals.org/content/66/4/1169.abstract`.

[13] H. Nishimori, Journal of Physics C: Solid State Physics **13**, 4071 (1980), URL `http://stacks.iop.org/0022-3719/13/i=21/a=012`.

[14] T. Morita and T. Horiguchi, Physics Letters A **76**, 424 (1980), ISSN 0375-9601, URL `http://www.sciencedirect.com/science/article/pii/0375960180907525`.

[15] J.-P. Tillich and G. Zemor, in *IEEE Int. Symp. on Inf. Th., 2009. ISIT 2009.* (2009), pp. 799–803.

[16] A. A. Kovalev and L. P. Pryadko, in *Proc. 2012 IEEE Int. Symp. Inf. Th. (ISIT)* (2012), pp. 348–352, ISSN 2157-8095, arXiv:1202.0928.

[17] A. A. Kovalev and L. P. Pryadko, Phys. Rev. A **88**, 012311 (2013), arXiv:1212.6703, URL `http://link.aps.org/doi/10.1103/PhysRevA.88.012311`.

[18] I. Andriyanova, D. Maurice, and J.-P. Tillich (2012), unpublished, arXiv:1202.3338.

[19] S. Bravyi and M. B. Hastings (2013), unpublished, arXiv:1311.0885.

[20] A. R. Calderbank and P. W. Shor, Phys. Rev. A **54**, 1098 (1996).

[21] A. M. Steane, Phys. Rev. A **54**, 4741 (1996), URL `http://dx.doi.org/10.1103/PhysRevA.54.4741`.

[22] H. Nishimori, Journal of Physics C: Solid State Physics **12**, L905 (1979), URL `http://stacks.iop.org/0022-3719/12/i=23/a=007`.

[23] H. Nishimori and K. Nemoto, Physica A: Statistical Mechanics and its Applications **321**, 108 (2003), ISSN 0378-4371, URL `http://www.sciencedirect.com/science/article/pii/S0378437102017910`.

[24] H. Nishimori, Journal of Statistical Physics **126**, 977 (2007), ISSN 0022-4715, URL `http://dx.doi.org/10.1007/s10955-006-9156-1`.

[25] H. Nishimori and M. Ohzeki, J. Phys. Soc. Jpn. **75**, 034004 (2006), cond-mat/0601356.

[26] M. Ohzeki, H. Nishimori, and A. N. Berker, Phys. Rev. E **77**, 061116 (2008), URL `http://pre.aps.org/abstract/PRE/v77/i6/e061116`.

[27] M. Ohzeki, Physical Review E **79**, 021129 (2009), URL `http://pre.aps.org/abstract/PRE/v79/i2/e021129`.

[28] H. Bombin, R. S. Andrist, M. Ohzeki, H. G. Katzgraber, and M. A. Martin-Delgado, Phys. Rev. X **2**, 021004 (2012), URL `http://link.aps.org/doi/10.1103/PhysRevX.2.021004`.

[29] M. Ohzeki and K. Fujii, Phys. Rev. E **86**, 051121 (2012), URL `http://pre.aps.org/abstract/PRE/v86/i5/e051121`.

[30] A. Aharony and M. J. Stephen, Journal of Physics

C: Solid State Physics **13**, L407 (1980), URL http://stacks.iop.org/0022-3719/13/i=16/a=001.

[31] P. W. Shor, Phys. Rev. A **52**, R2493 (1995), URL http://link.aps.org/abstract/PRA/v52/pR2493.

[32] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Infomation* (Cambridge Unive. Press, Cambridge, MA, 2000).

[33] R. Gallager, IRE Trans. Inf. Th. **8**, 21 (1962).

[34] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, Information Theory, IEEE Transactions on **47**, 619 (2001).

[35] T. J. Richardson and R. L. Urbanke, Information Theory, IEEE Transactions on **47**, 599 (2001).

[36] S.-Y. Chung, G. D. Forney Jr, T. J. Richardson, and R. Urbanke, Communications Letters, IEEE **5**, 58 (2001).

[37] K. Feng and Z. Ma, Information Theory, IEEE Transactions on **50**, 3323 (2004).

[38] M. H. Freedman, D. A. Meyer, and F. Luo, in *Computational Mathematics* (Chapman and Hall/CRC, 2002), pp. –, URL http://dx.doi.org/10.1201/9781420035377.ch12.

[39] A. A. Kovalev and L. P. Pryadko, Phys. Rev. A **87**, 020304(R) (2013), arXiv:1208.2317, URL http://link.aps.org/doi/10.1103/PhysRevA.87.020304.

[40] D. Gottesman (2013), unpublished, arXiv:1310.2984.

[41] A. Y. Kitaev, Ann. Phys. **303**, 2 (2003), URL http://arxiv.org/abs/quant-ph/9707021.

[42] J. MacWilliams, Bell System Tech. J **42**, 79 (1963).

[43] J.-M. Debierre and L. Turban, Journal of Physics A: Mathematical and General **16**, 3571 (1983), URL http://stacks.iop.org/0305-4470/16/i=15/a=022.

[44] D. J. Thouless, P. W. Anderson, and R. G. Palmer, Philosophical Magazine **35**, 593 (1977), http://www.tandfonline.com/doi/pdf/10.1080/14786437708235992, URL http://www.tandfonline.com/doi/abs/10.1080/14786437708235992.

[45] K. G. Wilson, Phys. Rev. D **10**, 2445 (1974), URL http://link.aps.org/doi/10.1103/PhysRevD.10.2445.

[46] J. B. Kogut, Rev. Mod. Phys. **51**, 659 (1979), URL http://link.aps.org/doi/10.1103/RevModPhys.51.659.

[47] T. Horiguchi and T. Morita, Journal of Physics A: Mathematical and General **14**, 2715 (1981), URL http://stacks.iop.org/0305-4470/14/i=10/a=024.

[48] H. Nishimori, Journal of Physics A: Mathematical and General **35**, 9541 (2002), URL http://stacks.iop.org/0305-4470/35/i=45/a=304.

[49] C. K. Thomas and H. G. Katzgraber, Phys. Rev. E **84**, 040101 (2011), URL http://link.aps.org/doi/10.1103/PhysRevE.84.040101.