Learning transitions in classical Ising models and deformed toric codes

Malte Pütz,[1] Samuel J. Garratt,[2] Hidetoshi Nishimori,[3] Simon Trebst,[1] and Guo-Yi Zhu[4, *]

[1]*Institute for Theoretical Physics, University of Cologne, Zülpicher Straße 77, 50937 Cologne, Germany*
[2]*Department of Physics, University of California, Berkeley, California 94720, USA*
[3]*Institute of Integrated Research, Institute of Science Tokyo,*
Nagatsuta-cho, Midori-ku, Yokohama 226-8503, Japan
[4]*The Hong Kong University of Science and Technology (Guangzhou), Nansha, Guangzhou, 511400, Guangdong, China*
(Dated: May 19, 2025)

Conditional probability distributions describe the effect of learning an initially unknown classical state through Bayesian inference. Here we demonstrate the existence of a *learning transition*, having signatures in the long distance behavior of conditional correlation functions, in the two-dimensional classical Ising model. This transition, which arises when learning local energy densities, extends all the way from the infinite-temperature paramagnetic state down to the thermal critical state. The intersection of the line of learning transitions and the thermal Ising transition is a novel tricritical point. Our model for learning also describes the effects of weak measurements on a family of quantum states which interpolate between the (topologically ordered) toric code and a trivial product state. Notably, the location of the above tricritical point implies that the quantum memory in the entire topological phase is robust to weak measurement, even when the initial state is arbitrarily close to the quantum phase transition separating topological and trivial phases. Our analysis uses a replica field theory combined with the renormalization group, and we chart out the phase diagram using a combination of tensor network and Monte Carlo techniques. Our methods can be extended to study the more general effects of learning on both classical and quantum states.

In classical statistical mechanics the state of a many-body system is a high-dimensional probability distribution, and the effects of learning can be captured through Bayesian inference [1–4]. In this framework, a broad initial probability distribution (for example, a Boltzmann distribution) is updated to a sharper conditional probability distribution which reflects our improved state of knowledge. This setting is dual to one arising in many-body quantum mechanics: the measurement of an observable causes a back-action on a quantum many-body state described by Born's rule. There, a fundamental problem is to determine when it is that quantum information is robust to measurement and decoherence [5–7]. However, in both classical and quantum settings the change in the many-body state depends sensitively on the initial correlations. This sensitivity raises the question of when it is that the effects of learning (or measurements) are universal, depending only on qualitative features of the initial states and learning protocols.
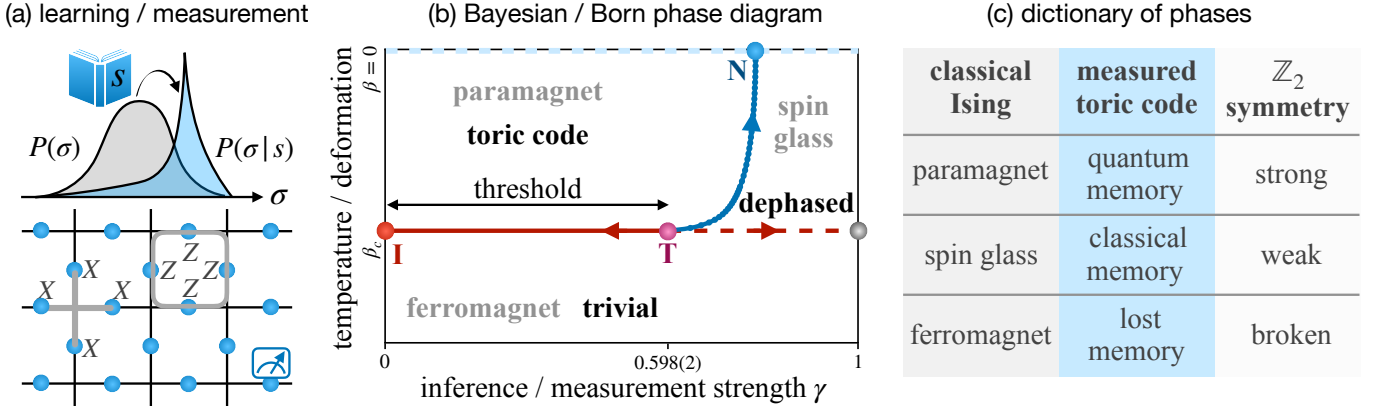
Here, we first study the effects of learning on high-entropy probability distributions describing the states of correlated classical binary degrees of freedom, or 'spins'. Our focus is on the two-dimensional classical Ising model on a square lattice, with our state of knowledge initially described by classical Gibbs states parametrized by inverse temperature β. We show that, throughout the entire high temperature phase $\beta < \beta_c$, and also at the thermal phase transition $\beta = \beta_c$, learning two-point correlations between neighboring spins leads to a sharp transition in our knowledge of long-distance correlations. These transitions occur only when the local entropy reduction, parameterized by a variable γ, exceeds a system-dependent $\gamma_c(\beta)$.

Such a transition has previously been identified [8] at infinite temperature $\beta = 0$, and the fact that the learning transition with finite γ extends down to the thermal phase transition implies a *novel tricritical point* in the $\beta - \gamma$ plane, at $\beta = \beta_c$

and $\gamma_T \equiv \gamma_c(\beta_c) > 0$. By adapting a replica field theory to describe the effects of learning and by leveraging the renormalization group to describe this theory, we provide evidence that the above behavior is universal to two-dimensional systems with global $\mathbb{Z}_2$ symmetry and when learning the values of local $\mathbb{Z}_2$ symmetric observables. We then probe the learning transitions in the Ising model numerically, and we reveal that the critical exponents governing the transition at $\beta = \beta_c$ (i.e. at the tricritical point) are distinct from those at $\beta < \beta_c$.

After characterizing the effects of learning on the classical Ising model, we exploit a duality relation to study the effects of weak measurements on a family of topologically ordered many-body quantum states which include the toric code [9]. Coherent superpositions of locally indistinguishable toric code states, belonging to different topological sectors, define robust quantum memories. These memories are well known to be robust to weak decoherence [10–12], and therefore also to the weak measurements that we consider [13, 14]. A basic question is whether the quantum memory remains robust when it is constructed from initial states that are far from the stabilizer limit of the toric code.

To address this question we study the effects of weak measurements on quantum memories defined by deformed toric code wavefunctions, where now $\beta \geq 0$ parameterizes the wavefunction deformation. In the absence of measurement, these deformations drive a quantum phase transition [15] from topologically ordered ($\beta < \beta_c$) to trivial ($\beta > \beta_c$) wavefunctions. Remarkably, we find that the quantum memory is robust even when constructed from states that are *arbitrarily close* to the quantum phase transition. Using the duality to the classical learning problem this robustness follows from the fact that, at the thermal phase transition of the Ising model, the effects of measurements are (marginally) irrelevant, i.e. that $\gamma_T > 0$. We illustrate this behavior, and its manifestations in

(a) learning / measurement   (b) Bayesian / Born phase diagram   (c) dictionary of phases

classical Ising	measured toric code	$\mathbb{Z}_2$ symmetry
paramagnet	quantum memory	strong
spin glass	classical memory	weak
ferromagnet	lost memory	broken

FIG. 1. **Schematic setup, numerical phase diagram, and duality.** (a) For a classical Ising model with spins σ_i, learning s_{ij} that is correlated with $\sigma_i\sigma_j$ (via a 'strength' γ) changes our state of knowledge from the Boltzmann distribution $P(\boldsymbol{\sigma}) \sim e^{-\beta E(\sigma)}$ to the conditional distribution $P(\boldsymbol{\sigma}|\boldsymbol{s})$. Similarly, when measuring a quantum system initially in a Rokhsar-Kivelson state $|\psi\rangle = \sum_{\boldsymbol{\sigma}} \sqrt{P(\boldsymbol{\sigma})}|\boldsymbol{\sigma}\rangle$, the post-measurement state takes the form $|\psi(\boldsymbol{s})\rangle = \sum_{\boldsymbol{\sigma}} \sqrt{P(\boldsymbol{\sigma}|\boldsymbol{s})}|\boldsymbol{\sigma}\rangle$. Here these states are Wegner-dual to those of deformed toric codes, where β is the strength of the wavefunction deformation. (b) Phase diagram according to the classical Bayesian rule or the quantum Born rule. The paramagnet and 'spin glass' phases correspond to short- and long-range nonlinear correlations $[\langle\sigma_i\sigma_j\rangle_{\boldsymbol{s}}^2] \equiv \sum_{\boldsymbol{s}}[\sum_{\boldsymbol{\sigma}} P(\boldsymbol{\sigma}|\boldsymbol{s})\sigma_i\sigma_j]^2$, respectively, but both have short-range linear correlations $[\langle\sigma_i\sigma_j\rangle_{\boldsymbol{s}}] = [\langle\sigma_i\sigma_j\rangle]$, while the ferromagnet has long-range linear correlations. The three different universality classes of the transitions we study are indicated I (Ising), N (Nishimori) and T (tricritical). (c) Relation between phases of the classical inference problem and the measured and deformed toric code, and symmetries of the Rokhsar-Kivelson classical-quantum state $\sum_{\boldsymbol{s}} P(\boldsymbol{s})|\boldsymbol{s}\rangle\langle\boldsymbol{s}| \otimes |\psi(\boldsymbol{s})\rangle\langle\psi(\boldsymbol{s})|$ consisting of both the measurement record $|\boldsymbol{s}\rangle$ and post-measurement states $|\psi(\boldsymbol{s})\rangle$.

the two problems, in Fig. 1. We note in passing that a similar structure of the phase diagram is found for a measurement version of the all-to-all interacting Ising ferromagnet, which is a 1-replica generalization of the Sherrington-Kirkpatrick model. In particular, there is also a tricritical point at finite $\gamma_c(\beta_c)$ where the three phases meet, see Supplementary Material (SM) [16].

We note that the $\beta = 0$ limit of our quantum problem [13] is dual to those studied in Ref. [17, 18]. More generally, the effects of measurements on quantum ground states can be understood by considering boundary perturbations to Euclidean spacetime path integrals [19–30]. Like those in Refs. [17, 18], the quantum states studied in this work are of Rokhsar-Kivelson (RK) form [15, 31], which is connected to the statistical mechanics of the classical Ising model. The measurement-induced critical points of such RK states are essentially described by non-unitary conformal field theories (CFT) [32], as recently discussed in Ref. [33]. Similarly, non-unitary CFTs describe [34–38] the critical (1+1)D monitored quantum dynamics, whose learnability aspects have been elucidated in Refs. [39–43].

Classical learning problem.– For a classical configuration of spins $\sigma_j = \pm 1$ on the vertices j of a square lattice, the Gibbs distribution is

$$P(\boldsymbol{\sigma}) = e^{-\beta E(\boldsymbol{\sigma})}/\mathcal{Z}, \quad \mathcal{Z} = \sum_{\boldsymbol{\sigma}} e^{-\beta E(\boldsymbol{\sigma})}, \quad (1)$$

where $E(\boldsymbol{\sigma}) = -\sum_{\langle ij\rangle} \sigma_i\sigma_j$ denotes the Ising energy of a full configuration $\boldsymbol{\sigma}$ of spins σ_j. The Gibbs distribution maximizes the entropy subject to the constraint of fixed $E = \sum_{\boldsymbol{\sigma}} P(\boldsymbol{\sigma})E(\boldsymbol{\sigma})$ [1, 2]. The ensemble undergoes a phase transition from the paramagnet phase at high temperature to the

ferromagnetic phase at low temperature across the 2D Ising critical point at $\beta_c = \ln\sqrt{1+\sqrt{2}}$.

We will study how this distribution is modified when we 'learn' the local correlation $\sigma_i\sigma_j$ with i and j nearest neighbors on the square lattice. Here 'learning' means that we extract a random binary variable $s_{ij} = \pm 1$ that is correlated with $\sigma_i\sigma_j$, following a distribution $P(s_{ij}|\sigma_i\sigma_j) = (1+\gamma s_{ij}\sigma_i\sigma_j)/2$ where $\gamma \in [0,1]$ controls the learning precision and hence the local entropy reduction, having an interpretation as a 'measurement strength'. For large γ we learn $\sigma_i\sigma_j$ perfectly, maximally reducing the local entropy, while for $\gamma = 0$ we learn nothing. The full set $\boldsymbol{s}$ of parameters s_{ij} (one for each edge of the square lattice) then has a conditional *product* distribution

$$P(\boldsymbol{s}|\boldsymbol{\sigma}) = \prod_{\langle ij\rangle} \frac{1+\gamma s_{ij}\sigma_i\sigma_j}{2}, \quad (2)$$

and unconditional distribution $P(\boldsymbol{s}) = \sum_{\boldsymbol{\sigma}} P(\boldsymbol{s}|\boldsymbol{\sigma})P(\boldsymbol{\sigma})$. Having learned $\boldsymbol{s}$, the distribution of $\boldsymbol{\sigma}$ is updated according to Bayes' theorem, see Fig. 1(a). The posterior probability distribution is

$$P(\boldsymbol{\sigma}|\boldsymbol{s}) = \frac{1}{\mathcal{Z}(\boldsymbol{s})} \exp\left(\sum_{\langle ij\rangle} \tilde{\gamma}s_{ij}\sigma_i\sigma_j - \beta E(\sigma)\right), \quad (3)$$

with $\tilde{\gamma} \equiv \tanh^{-1}(\gamma)$. Here $\mathcal{Z}(\boldsymbol{s})$ is the normalization constant that is proportional to the probability $P(\boldsymbol{s}) = \mathcal{Z}(\boldsymbol{s})/\sum_{\boldsymbol{s}} \mathcal{Z}(\boldsymbol{s})$. Note that, for $\beta \neq 0$, this is a *gauge asymmetric* random bond Ising model.

To study the effects of learning $\boldsymbol{s}$, we ask how the correlations between physical spins described by $P(\boldsymbol{\sigma}|\boldsymbol{s})$ differ from

FIG. 2. **Finite threshold of the critical state**. Shown is the domain wall entropy I_c (equivalent to the coherent information), which signals the phase transition. (a) I_c for increasing measurement strength at critical Ising temperature $\beta_c = \ln\sqrt{1+\sqrt{2}}$. The pink shade highlights the critical region with finite fraction of I_c, between the Ising critical point "I" and the tricritical point "T", at $\gamma_T \approx 0.598(2)$. It is found that I_c yields a nonzero scaling dimension η/ν at the tricritical point, in contrast to scaling invariant at the Ising critical point. (b) I_c along tuning temperature at a chosen finite measurement strength below the threshold at $\gamma = 0.3 < \gamma_T$. Numerical computation is performed for a square stripe of length d, which is dual to a deformed surface code of code distance d. The error bars of standard deviation are smaller than the markers.

the correlations described by $P(\boldsymbol{\sigma})$. We can develop some intuition by considering the limit $\gamma = 1$, where $s_{ij} = \sigma_i\sigma_j$. In this case, $P(\boldsymbol{\sigma}|\boldsymbol{s}) = 1/2$ for the two spin configurations $\boldsymbol{\sigma}$ that are consistent with the observed $\boldsymbol{s}$ (and that are related to each other by $\boldsymbol{\sigma} \to -\boldsymbol{\sigma}$), while $P(\boldsymbol{\sigma}|\boldsymbol{s}) = 0$ for all other configurations. Therefore, the conditional correlation function $\langle\sigma_i\sigma_j\rangle_{\boldsymbol{s}} \equiv \sum_{\boldsymbol{\sigma}} P(\boldsymbol{\sigma}|\boldsymbol{s})\sigma_i\sigma_j$ is simply $\langle\sigma_i\sigma_j\rangle_{\boldsymbol{s}} = \pm 1$.

More generally, the correlations described by the conditional distributions $P(\boldsymbol{\sigma}|\boldsymbol{s})$ depend sensitively on $\boldsymbol{s}$, and the relationship between $\langle\sigma_i\sigma_j\rangle_{\boldsymbol{s}}$ and $\boldsymbol{s}$ can be quite complex. In order to probe the effects of learning at the level of the ensemble of all possible $\boldsymbol{s}$, we will study averages of correlation functions over this ensemble. We will denote averages with respect to $P(\boldsymbol{s})$ by $[\cdots]$. For example, for $\gamma = 1$ we saw above that $[\langle\sigma_i\sigma_j\rangle_{\boldsymbol{s}}^2] = 1$. Note that, on the other hand, $[\langle\sigma_i\sigma_j\rangle_{\boldsymbol{s}}] = \langle\sigma_i\sigma_j\rangle$ for all γ, where $\langle\cdots\rangle$ is an (unconditional) average with respect to $P(\boldsymbol{\sigma})$.

Nonlinear probes of the effects of learning.– The above examples illustrate a general feature of learning problems: Averages over linear functions of conditional correlations are insensitive to the effects of learning, i.e. $[\langle\cdots\rangle_{\boldsymbol{s}}] = \langle\cdots\rangle$, where

'$\cdots$' is a function of $\boldsymbol{\sigma}$. To probe the effects of learning at the level of the ensemble of $\boldsymbol{s}$, it is necessary to average *nonlinear* functions of the conditional correlations. In the paramagnetic (PM) phase (at small β and small γ) the Edwards-Anderson correlation function $[\langle\sigma_i\sigma_j\rangle_{\boldsymbol{s}}^2]$ vanishes at large separation between i and j, while in the 'spin glass' (SG) phase (at small β and large γ) this correlation is finite at large separations. At $\beta = 0$, where the model is on the Nishimori line [8, 44], the location $\gamma_c(\beta = 0)$ of the transition between PM and SG has been identified as $\gamma_c(0) \approx 0.782$, which corresponds to the known critical disorder probability 10.9% [45–48].

Here it will be convenient to study an averaged domain wall entropy, and we will use this object to identify and to probe the transition in correlation functions that occurs as γ is increased through $\gamma_c(\beta)$. This domain wall entropy will also turn out to have a clear physical meaning in the context of the deformed toric code where it is equivalent to the coherent information. To define the domain wall entropy, we consider a $d \times d$ square lattice Ising model in a stripe geometry and introduce two additional spins coupled, respectively, to all of the original spins at the two opposite ends of the stripe. The coupling β between these new 'boundary' spins σ_L and σ_R and the bulk spins σ_j is the same as between neighboring bulk spins.

We then study the effect of learning s_{ij} for all $i, j \neq L, R$ on the correlations $C_{\boldsymbol{s}} \equiv \langle\sigma_L\sigma_R\rangle_{\boldsymbol{s}}$ between σ_L and σ_R. Graphically, $C_{\boldsymbol{s}}$ can be represented by

$$\tag{4}$$

Since the probability for the absence/presence of a domain wall is $(1 \pm C_{\boldsymbol{s}})/2$, the conditional domain wall entropy is

$$I_{\boldsymbol{s}} \equiv -\frac{1 + C_{\boldsymbol{s}}}{2}\log_2\frac{1 + C_{\boldsymbol{s}}}{2} - \frac{1 - C_{\boldsymbol{s}}}{2}\log_2\frac{1 - C_{\boldsymbol{s}}}{2}, \tag{5}$$

and the ensemble averaged domain wall entropy is denoted $I_c \equiv [I_{\boldsymbol{s}}]$. In the PM phase, domain wall fluctuations destroy long-range conditional correlations, so that $I_c \to 1$ at large linear dimension d. In the SG phase we instead expect $I_c \to 0$ at large d. On the other hand, in the low temperature phase $\beta > \beta_c$, the prior distribution is already a long-range ordered ferromagnetic (FM) phase, thus for all γ we expect vanishing domain wall entropy $I_c \to 0$ for large L.

We now numerically calculate I_c across the learning transition, extracting $\gamma_c(\beta)$ and the critical exponent $\nu_\gamma = \nu_\gamma(\beta)$ governing the divergence of the correlation length. The exponent $\nu_\gamma(\beta)$ is here defined from the dependence of I_c on γ at fixed β and for $\gamma \approx \gamma_c(\beta)$. We can similarly define the thermal correlation length exponent $\nu_\beta = \nu_\beta(\gamma)$ from the dependence of I_c on β at fixed γ and for $\beta \approx \beta_c$. Our numerical calculations are carried out as follows: First we sample configurations of $\sigma_i\sigma_j$, and hence s_{ij} using Eq. (2), from a standard
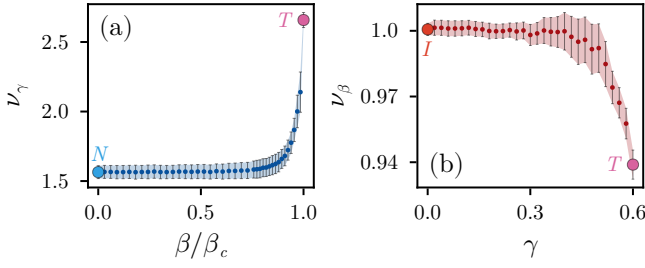
FIG. 3. **Critical scaling and RG flow at the transition lines.** (a) From Nishimori critical to the tricritical point: the critical exponent ν_γ across γ_c defined by $\xi \propto (\gamma - \gamma_c)^{-\nu_\gamma}$, calculated for horizontal cuts of fixed β. (b) From clean Ising critical to tricritical point: the critical exponent ν_β for $\xi \propto (\beta - \beta_c)^{-\nu_\beta}$, calculated for vertical cuts of fixed γ. The critical exponents are extracted from the coherent information I_c ($d = 8, 16, 32, 64$), using a scaling ansatz with two critical exponents ν_γ and η, see the Appendix. In the respective limits, the numerical results agree with the known exponents for Nishimori $\nu_\gamma(0) = 1.564(46)$ (blue circle) and Ising $\nu_\beta(0) = 1$ (red circle) criticality. The tricritical point has scaling exponents $\nu_\gamma(\beta_c) = 2.658(54)$ and $\nu_\beta(\gamma_c) = 0.9389(66)$.

Monte Carlo simulation of the classical Ising model, and we then use tensor network methods [13] to evaluate conditional observables such as I_s. By averaging I_s over observed s we arrive at the estimates for I_c in Fig. 2.

In Fig. 2(a) we vary γ at $\beta = \beta_c$. For small γ the data shows that I_c is approximately independent of d, whereas for large γ the domain wall entropy is suppressed upon increasing d. This suggests that $\gamma_T \equiv \gamma_c(\beta_c)$ is finite. Indeed, we can collapse I_c for various d and γ as shown in the inset, with our results indicating $\gamma_T \approx 0.598(2)$ and a correlation length exponent for the learning transition of $\nu_\gamma(\beta_c) = 2.658(54)$. Interestingly, the exponent $\nu_\gamma(\beta_c)$ is distinct from the $\beta = 0$ value $\nu_\gamma(0) = 1.564(46)$; we will analyze the behavior of $\nu_\gamma(\beta)$ in more detail below.

In Fig. 2(b) we vary β with γ fixed to be below the threshold γ_T. As expected, for high temperature $\beta < \beta_c$, increasing d causes I_c to approach unity, while for low temperature $\beta > \beta_c$ we find that I_c decays rapidly with increasing d. The data collapse (inset) shows that away from the tricritical point the thermal correlation length exponent $\nu_\beta(\gamma) = 0.998(5)$, consistent with the analytical value $\nu_\beta(0) = 1$ for the 2D Ising transition.

In Fig. 3 we show the variation of the critical exponents along the transition lines. Figure 3(a) shows the critical exponent $\nu_\gamma(\beta)$ for the learning transition as a function of β, i.e. moving from the Nishimori point to the tricritical point. The change in $\nu_\gamma(\beta)$ from $\nu_\gamma(0) = 1.564(46)$ to $\nu_\gamma(\beta_c) = 2.658(54)$ occurs over a small interval of β/β_c just below unity. These results suggest that the learning transitions with $\beta < \beta_c$ are in the same universality class as the learning transition at $\beta = 0$, but that the transition at $\beta = \beta_c$ is distinct. That is, we can identify [49] an RG flow from the tricritical point at critical temperature β_c to the Nishimori fixed point at infinite temperature $\beta = 0$, consistent with the perturbative RG analysis of Ref. [50]. Figure 3(b) shows the critical exponent $\nu_\beta(\gamma)$ for the thermal phase transition as a function of γ (i.e. from the pure Ising to the tricritical transition); we again find a fairly sharp change in the exponent on approaching the tricritical point. Together, the numerically extracted critical exponents indicate a tricritical point at finite $\gamma_c(\beta_c) = 0.598(2)$ that is unstable, flowing to Ising criticality along the phase boundary with $\beta = \beta_c$ and $\gamma < \gamma_c(\beta_c)$, and to Nishimori criticality on the phase boundary with $\beta < \beta_c$ and $\gamma = \gamma_c(\beta)$, see Fig. 1(b).

Marginal irrelevance of weak measurement.– We now use a replica field theory to explain why γ_T is finite. The replica trick involves writing nonlinear correlation functions as, e.g.,

$$[\langle \sigma_i \sigma_j \rangle_{\boldsymbol{s}}^2] = \lim_{n \to 1} \frac{\sum_{\boldsymbol{s}} P^n(\boldsymbol{s}) \left[\sum_{\boldsymbol{\sigma}} P(\boldsymbol{\sigma}|\boldsymbol{s}) \sigma_i \sigma_j \right]^2}{\sum_{\boldsymbol{s}} P^n(\boldsymbol{s})} \quad (6)$$

for bulk correlations. A similar expression holds for I_c. The idea is then to calculate correlation functions for integer $n \geq 2$ and then analytically continue the results to $n = 1$.

From Eq. (6) it is clear that, in the replica theory, the object $\sum_{\boldsymbol{s}} P^n(\boldsymbol{s})$ plays the role of the partition function. Using the expression for $P(\boldsymbol{s})$ as a sum over spin configurations $\boldsymbol{\sigma}$, we can sum over $\boldsymbol{s}$ such that $\sum_{\boldsymbol{s}} P^n(\boldsymbol{s})$ becomes a partition function for n classical Ising models, having degrees of freedom σ_i^a with $a = 1, \dots, n$ replica indices, and 'inter-replica' couplings $\sigma_i^a \sigma_j^a \sigma_i^b \sigma_j^b$ on all nearest-neighbor bonds $\langle ij \rangle$.

Since for $\beta = \beta_c$ and $\gamma = 0$ the long-distance behavior of correlation functions can be calculated using the Ising CFT, it is natural to expect that the effects of a small nonzero γ can be determined using this framework. We therefore consider an n-replica Ising CFT with local perturbations that couple even parity fields (since $\sigma_i \sigma_j$ is even under the parity transformation $\boldsymbol{\sigma} \to -\boldsymbol{\sigma}$). The action for this theory is

$$\mathcal{S}_n = \sum_{a=1}^{n} \mathcal{S}_{\mathrm{I}}^a - g \sum_{\substack{a,b=1 \\ a \neq b}}^{n} \int d^2 x \, \varepsilon^a(\vec{x}) \varepsilon^b(\vec{x}), \quad (7)$$

where $\mathcal{S}_{\mathrm{I}}^a$ describes the ath replica of the unperturbed Ising CFT, with $a = 1, \dots, n$, the parameter $g \propto \tilde{\gamma}^2$, and $\varepsilon^a(\vec{x})$ is the energy density in replica a and at position $\vec{x} = (x, y)$. The spin density is denoted by $\sigma^a(\vec{x})$. Within this theory, the Edwards-Anderson correlation function $[\langle \sigma_i \sigma_j \rangle_{\boldsymbol{s}}^2]$ is given by the $n \to 1$ limit of $\langle \sigma^1(\vec{x}_i) \sigma^1(\vec{x}_j) \sigma^2(\vec{x}_i) \sigma^2(\vec{x}_j) \rangle_{\mathcal{S}_n}$, where $\langle \cdots \rangle_{\mathcal{S}_n}$ is an average with respect to the statistical weight $e^{-\mathcal{S}_n}$, and $\vec{x}_i$ is the position of site i. To determine whether $[\langle \sigma_i \sigma_j \rangle_{\boldsymbol{s}}^2]$ is modified at long distances relative to the behavior $\sim 1/\sqrt{r_{ij}}$ at $g = 0$, we renormalize g. The RG flow under an infinitesimal change of the lattice scale by a factor e^ℓ is

$$\frac{dg}{d\ell} = 4\pi(n-2)g^2 + O(g^3). \quad (8)$$

Sending $n \to 1$ we see that a small nonzero g flows to zero under RG. The marginal irrelevance of the effects of measurements at $\beta = \beta_c$ indicates that $\gamma_T > 0$. This is consistent with our numerical results above.

Deformed toric code.– Here we discuss the connection between our results and the effects of weak measurements on deformed toric code wavefunctions. Ising models in the geometry discussed in connection with Fig 2 and Eq. (5), with boundary conditions $\mu = \sigma_L \sigma_R = \pm 1$, are associated with RK wavefunctions $|\psi_\mu\rangle = \sum_{\boldsymbol{\sigma}} \sqrt{P_\mu(\boldsymbol{\sigma})} |\boldsymbol{\sigma}\rangle$, where $P_\mu(\boldsymbol{\sigma})$ is the Boltzmann weight for the bulk spins $\boldsymbol{\sigma}$ with boundary condition μ. These wavefunctions are dual to the two logical states of deformed toric codes on open surfaces (surface codes), which we denote $|\phi_\mu\rangle \propto \exp\left(\frac{1}{2}\sum_{\langle ij\rangle}\beta Z_{ij}\right)|\text{toric code}_\mu\rangle$. In $|\phi_\mu\rangle$ the qubits reside on edges $\langle ij\rangle$ of the square lattice, Z_{ij} is a Pauli matrix, and $|\text{toric code}_\mu\rangle$ is an ideal toric code ground state with $\mu = \pm 1$ specifying the logical state. These deformed states undergo quantum phase transitions across (2+0)D conformal quantum critical points [15] at β_c, see Fig. 1(b). If, in addition, we apply weak measurement of strength γ to every qubit of the toric code state $|\phi_\mu\rangle$, this results in a statistical ensemble of post-measurement states conditioned on the measurement outcomes $\boldsymbol{s}$

$$|\phi_\mu(\boldsymbol{s})\rangle \propto \prod_{\langle ij\rangle}[1+\tanh(\tilde{\gamma}/2)s_{ij}Z_{ij}]|\phi_\mu\rangle \ . \quad (9)$$

Here the measurement probability is given by Born's rule $P(\boldsymbol{s}) = \sum_\mu \langle\phi_\mu(\boldsymbol{s})|\phi_\mu(\boldsymbol{s})\rangle$ which leads to the exact same result as Bayes' rule (3) in the classical inference model. The coherent information [51, 52] of the toric code [11] under such measurement channel is related to the domain wall entropy I_c in (5), by noting that now $C_{\boldsymbol{s}} = \langle\phi_-(\boldsymbol{s})|\phi_+(\boldsymbol{s})\rangle$ takes physical meaning as the fidelity between two topologically degenerate states [13]. Then it is straightforward to map the phase diagram of the classical inference model to that of the toric code in Fig. 1(b,c). In the SG regime, the toric code is subjected to strong measurement and dephased into a classical loop gas, which loses the quantum memory ($I_c = 0$) but still retains a classical topological memory encoded by the domain wall as a non-contractible loop. Such a phase is characterized by order parameters $[C_{\boldsymbol{s}}] = 0$, $[C_{\boldsymbol{s}}^2] \neq 0$. In the FM regime, the classical topological memory is lost, since the domain wall is exponentially suppressed dictated by $[C_{\boldsymbol{s}}] \neq 0$. The location of the phase boundary between the PM and the SG phases, γ_c, determines the information-theoretic threshold of the toric code states against measurement. According to the phase diagram, when one deforms the toric code away from its $\beta = 0$ fixed point by increasing β, the information-theoretic threshold γ_c decreases from $\gamma_N \approx 0.782$ (Nishimori threshold) to $\gamma_T \approx 0.598(2)$ (tricritical threshold) instead of vanishing as the critical point is approached. Then it suddenly drops to zero when it becomes a trivial phase. In other words, the existence of a tricritical point at finite γ_T means that coherent superpositions of deformed toric code states remain stable to measurement even when arbitrarily close to the critical point. Here, this is a consequence of the marginal irrelevance of $\mathbb{Z}_2$ symmetric measurements in the Ising CFT.

Symmetry perspective.– Let us close by discussing the symmetries of the three phases around the tricritical point. We will focus on the RK wavefunctions $|\psi(\boldsymbol{s})\rangle = \sum_{\boldsymbol{\sigma}} \sqrt{P(\boldsymbol{\sigma}|\boldsymbol{s})}|\boldsymbol{\sigma}\rangle$

associated with the conditional distributions $P(\boldsymbol{\sigma}|\boldsymbol{s})$; for this discussion we can consider an infinite system, and so we omit the boundary condition index μ. Ising criticality is, of course, a well-known universality class for $\mathbb{Z}_2$ symmetry breaking in disorder-free situations (e.g., when moving from paramagnet to ferromagnet). In the presence of disorder, Nishimori criticality – in lieu of Ising – has been found to be ubiquitous in $\mathbb{Z}_2$ mixed quantum states governed by Born's rule [10, 11, 14, 17, 18], where it has been connected to $\mathbb{Z}_2$ strong-to-weak symmetry breaking [53–55]. These two universality classes naturally appear in the problem at hand, and are inevitably bound to join at the tricritical point we have identified, see Figs. 1(b,c). The three phases meeting at the tricritical point distinguish themselves by exhibiting strong, weak, or no $\mathbb{Z}_2$ symmetry, in a sense we now discuss.

The whole statistical ensemble of pure states can be compactly described by a block-diagonal classical-quantum density matrix [33] of the form $\rho = \sum_{\boldsymbol{s}} P(\boldsymbol{s})|\boldsymbol{s}\rangle\langle\boldsymbol{s}|\otimes|\psi(\boldsymbol{s})\rangle\langle\psi(\boldsymbol{s})|$, where $|\boldsymbol{s}\rangle\langle\boldsymbol{s}|$ is the state of the register that records the measurement outcomes. It is then easy to check that the mixed quantum state ρ possesses a strong (exact) $\mathbb{Z}_2$ symmetry $\left(\prod_j X_j\right)\rho = \rho = \rho\left(\prod_j X_j\right)$ throughout the phase diagram. For $\beta > \beta_c$ this symmetry is spontaneously broken to null symmetry resulting in the ferromagnetic phase. For $\beta < \beta_c$ there is no exact long-range order since the ferromagnetic correlation vanishes. Nonetheless, at large γ we can relate the spin-glass correlations to the mixed-state fidelity correlator [53, 54] $\mathcal{F}(\rho, Z_iZ_j\rho Z_jZ_i) = [|\langle\sigma_i\sigma_j\rangle_{\boldsymbol{s}}|]$, indicating the existence of a SW-SSB phase with *statistical average* long-range order. To conclude, the relevant symmetry-preserving perturbations out of the tricritical point could lead to three distinct phases: a strongly symmetric phase (PM), a strong-to-weak symmetry broken phase (SG), or a conventional symmetry broken phase (FM).

* guoyizhu@hkust-gz.edu.cn

[1] E. T. Jaynes, Information Theory and Statistical Mechanics, Phys. Rev. **106**, 620 (1957).

[2] E. T. Jaynes, Information Theory and Statistical Mechanics. II, Phys. Rev. **108**, 171 (1957).

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

[4] L. Zdeborová and F. Krzakala, Statistical physics of inference: thresholds and algorithms, Adv. Phys. **65**, 453 (2016).

[5] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge University Press, 2010).

[6] M. P. Fisher, V. Khemani, A. Nahum, and S. Vijay, Random Quantum Circuits, Annu. Rev. Condens. Matter Phys. **14**, 335 (2023).

[7] A. C. Potter and R. Vasseur, Entanglement dynamics in hybrid quantum circuits, in *Entanglement in Spin Chains: From Theory to Quantum Technology Applications*, edited by A. Bayat, S. Bose, and H. Johannesson (Springer International Publishing, Cham, 2022) pp. 211–249.

[8] Y. Iba, The Nishimori line and Bayesian statistics, J. Phys. A **32**, 3875 (1999).

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

[10] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, Journal of Mathematical Physics **43**, 4452 (2002).

[11] R. Fan, Y. Bao, E. Altman, and A. Vishwanath, Diagnostics of Mixed-State Topological Order and Breakdown of Quantum Memory, PRX Quantum **5**, 020343 (2024).

[12] Y.-H. Chen and T. Grover, Unconventional topological mixed-state transition and critical phase induced by self-dual coherent errors, Phys. Rev. B **110**, 125152 (2024).

[13] F. Eckstein, B. Han, S. Trebst, and G.-Y. Zhu, Robust Teleportation of a Surface Code and Cascade of Topological Quantum Phase Transitions, PRX Quantum **5**, 040313 (2024).

[14] E. H. Chen, G.-Y. Zhu, R. Verresen, A. Seif, E. Bäumer, D. Layden, N. Tantivasadakarn, G. Zhu, S. Sheldon, A. Vishwanath, S. Trebst, and A. Kandala, Nishimori transition across the error threshold for constant-depth quantum circuits, Nat. Phys. **21**, 161 (2025).

[15] E. Ardonne, P. Fendley, and E. Fradkin, Topological order and conformal quantum critical points, Ann. Phys. (N. Y.) **310**, 493 (2004).

[16] See Supplementary Material for numerical method and supplementary data and exactly solved models., .

[17] G.-Y. Zhu, N. Tantivasadakarn, A. Vishwanath, S. Trebst, and R. Verresen, Nishimori's Cat: Stable Long-Range Entanglement from Finite-Depth Unitaries and Weak Measurements, Phys. Rev. Lett. **131**, 200201 (2023).

[18] J. Y. Lee, W. Ji, Z. Bi, and M. P. A. Fisher, Decoding Measurement-Prepared Quantum Phases and Transitions: from Ising model to gauge theory, and beyond, preprint (2022), arXiv:2208.11699.

[19] S. J. Garratt, Z. Weinstein, and E. Altman, Measurements Conspire Nonlocally to Restructure Critical Quantum States, Phys. Rev. X **13**, 021026 (2023).

[20] J. Y. Lee, C.-M. Jian, and C. Xu, Quantum Criticality Under Decoherence or Weak Measurement, PRX Quantum **4**, 030317 (2023).

[21] Y. Bao, R. Fan, A. Vishwanath, and E. Altman, Mixed-state topological order and the errorfield double formulation of decoherence-induced transitions, arXiv 2301.05687.

[22] Z. Weinstein, R. Sajith, E. Altman, and S. J. Garratt, Nonlocality and entanglement in measured critical quantum Ising chains, Phys. Rev. B **107**, 245132 (2023).

[23] S. Murciano, P. Sala, Y. Liu, R. S. K. Mong, and J. Alicea, Measurement-Altered Ising Quantum Criticality, Phys. Rev. X **13**, 041042 (2023).

[24] P. Sala, S. Murciano, Y. Liu, and J. Alicea, Quantum Criticality Under Imperfect Teleportation, PRX Quantum **5**, 030307 (2024).

[25] Z. Yang, D. Mao, and C.-M. Jian, Entanglement in a one-dimensional critical state after measurements, Phys. Rev. B **108**, 165120 (2023).

[26] R. A. Patil and A. W. W. Ludwig, Highly complex novel critical behavior from the intrinsic randomness of quantum mechanical measurements on critical ground states – a controlled renormalization group analysis, preprint (2024), arXiv:2409.02107.

[27] Y. Liu, S. Murciano, D. F. Mross, and J. Alicea, Boundary transitions from a single round of measurements on gapless quantum states, preprint (2024), arXiv:2412.07830.

[28] K. Baweja, D. J. Luitz, and S. J. Garratt, Post-measurement Quantum Monte Carlo, preprint (2024), arXiv:2410.13844.

[29] M. Hoshino, M. Oshikawa, and Y. Ashida, Entanglement swapping in critical quantum spin chains, preprint (2024), arXiv:2406.12377.

[30] Q. Tang and X. Wen, A critical state under weak measurement is not critical, preprint (2024), arXiv:2411.13705.

[31] C. L. Henley, From classical to quantum dynamics at Rokhsar–Kivelson points, Journal of Physics: Condensed Matter **16**, S891 (2004).

[32] V. Gurarie and A. W. W. Ludwig, Conformal field theory at central charge c=0 and two-dimensional critical systems with quenched disorder, in *From Fields to Strings: Circumnavigating Theoretical Physics* (World Scientific, 2005) pp. 1384–1440.

[33] Q. Wang, R. Vasseur, S. Trebst, A. W. W. Ludwig, and G.-Y. Zhu, Decoherence-induced self-dual criticality in topological states of matter, preprint (2025), arXiv:2502.14034.

[34] N. Lang and H. P. Büchler, Entanglement transition in the projective transverse field Ising model, Phys. Rev. B **102**, 094204 (2020).

[35] X. Chen, Y. Li, M. P. A. Fisher, and A. Lucas, Emergent conformal symmetry in nonunitary random dynamics of free fermions, Phys. Rev. Res. **2**, 033017 (2020).

[36] Y. Li, X. Chen, A. W. W. Ludwig, and M. P. A. Fisher, Conformal invariance and quantum nonlocality in critical hybrid circuits, Phys. Rev. B **104**, 104305 (2021).

[37] A. Zabalo, M. J. Gullans, J. H. Wilson, R. Vasseur, A. W. W. Ludwig, S. Gopalakrishnan, D. A. Huse, and J. H. Pixley, Operator Scaling Dimensions and Multifractality at Measurement-Induced Transitions, Phys. Rev. Lett. **128**, 050602 (2022).

[38] A. Kumar, K. Aziz, A. Chakraborty, A. W. W. Ludwig, S. Gopalakrishnan, J. H. Pixley, and R. Vasseur, Boundary transfer matrix spectrum of measurement-induced transitions, Phys. Rev. B **109**, 014303 (2024).

[39] F. Barratt, U. Agrawal, A. C. Potter, S. Gopalakrishnan, and R. Vasseur, Transitions in the Learnability of Global Charges from Local Measurements, Phys. Rev. Lett. **129**, 200602 (2022).

[40] Y. Bao, S. Choi, and E. Altman, Theory of the phase transition in random unitary circuits with measurements, Phys. Rev. B **101**, 104301 (2020).

[41] U. Agrawal, J. Lopez-Piqueres, R. Vasseur, S. Gopalakrishnan,

and A. C. Potter, Observing quantum measurement collapse as a learnability phase transition, Phys. Rev. X **14**, 041012 (2024).

[42] M. Ippoliti and V. Khemani, Learnability Transitions in Monitored Quantum Dynamics via Eavesdropper's Classical Shadows, PRX Quantum **5**, 020304 (2024).

[43] H. Singh, R. Vasseur, A. C. Potter, and S. Gopalakrishnan, Mixed-state learnability transitions in monitored noisy quantum dynamics, preprint (2025), arXiv:2503.10308.

[44] H. Nishimori, Internal Energy, Specific Heat and Correlation Function of the Bond-Random Ising Model, Prog. Theor. Phys. **66**, 1169 (1981).

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

[46] F. Merz and J. T. Chalker, Two-dimensional random-bond Ising model, free fermions, and the network model, Phys. Rev. B **65**, 054425 (2002).

[47] M. Ohzeki and J. L. Jacobsen, High-precision phase diagram of spin glasses from duality analysis with real-space renormalization and graph polynomials, J. Phys. A **48**, 095001 (2015).

[48] T. Chen, E. Guo, W. Zhang, P. Zhang, and Y. Deng, Tensor network Monte Carlo simulations for the two-dimensional random-bond Ising model, Phys. Rev. B **111**, 094201 (2025).

[49] M. Pütz, R. Vasseur, A. W. W. Ludwig, S. Trebst, and G.-Y. Zhu, to appear, (2025).

[50] P. Le Doussal and A. B. Harris, ϵ expansion for the Nishimori multicritical point of spin glasses, Phys. Rev. B **40**, 9249 (1989).

[51] B. Schumacher and M. A. Nielsen, Quantum data processing and error correction, Phys. Rev. A **54**, 2629 (1996).

[52] M. J. Gullans and D. A. Huse, Scalable Probes of Measurement-Induced Criticality, Phys. Rev. Lett. **125**, 070606 (2020).

[53] P. Sala, S. Gopalakrishnan, M. Oshikawa, and Y. You, Spontaneous strong symmetry breaking in open systems: Purification perspective, Phys. Rev. B **110**, 155150 (2024).

[54] L. A. Lessa, R. Ma, J.-H. Zhang, Z. Bi, M. Cheng, and C. Wang, Strong-to-Weak Spontaneous Symmetry Breaking in Mixed Quantum States, PRX Quantum **6**, 010344 (2025).

[55] C. Zhang, Y. Xu, J.-H. Zhang, C. Xu, Z. Bi, and Z.-X. Luo, Strong-to-weak spontaneous breaking of 1-form symmetry and intrinsically mixed topological order, preprint (2024), arXiv:2409.17530.

[56] A. Nahum and J. Lykke Jacobsen, Bayesian critical points in classical lattice models, preprint (2025), arXiv:2504.01264.

[57] M. Pütz, S. J. Garratt, H. Nishimori, S. Trebst, and G.-Y. Zhu, Data for "Learning transitions in classical Ising models and deformed toric codes", Zenodo 10.5281/zenodo.15227834 (2025).

Supplemental Material

Numerical method: sampling and random tensor network

In our numerical simulation, we first generate the samples s for the precise learning limit $\gamma = 1$. This can be done by standard Monte Carlo method sampling of Ising configurations σ with the Boltzmann weight $P(\sigma)$. We can then convert the Ising configurations to domain wall configurations via the duality transformation: $\sigma_i \sigma_j = s_{ij}$, for any bond $\langle ij \rangle$. When the measurement strength is decreased to $\gamma < 1$, we randomly flip the sign of each bond variable s_{ij} with a probability $(1-\gamma)/2$, according to $P(s_{ij}|\sigma_i\sigma_j) = (1 + \gamma s_{ij}\sigma_i\sigma_j)/2$. In this way we obtain the samples s for arbitrary measurement strength γ. Then we use the samples s to construct the random tensor network for the partition function $\mathcal{Z}(s)$. As shown in Fig. 4, we can contract out the tensor network by performing a (1+1)D matrix product state evolution from left to right. Note that this single layer Ising tensor network can also be fermionized into a (1+1)D monitored Gaussian fermion chain, or a 2D Chalker-Coddington network model [33, 46] for efficient simulation, but the tensor network contraction can be applied to more generic statistical model that cannot be mapped to free fermion models.
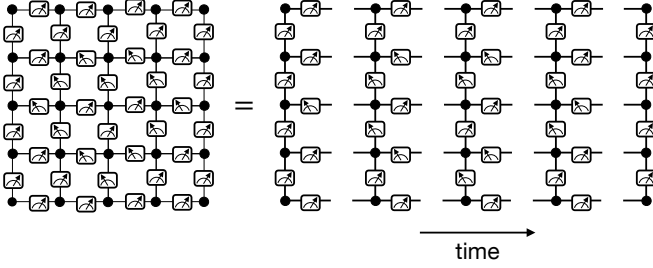
FIG. 4. **Random tensor network for the measurement problems** can be decomposed into slices of random matrix product operator as transfer matrices. A (1+1)D matrix product state evolves from left to right by the transfer matrices.

Supplementary numerical data

Moments of domain wall correlation

Let us supplement the numerical data in the main text by different moments of the domain wall correlation functions $[C_s]$, $[C_s^2]$, and $[|C_s|]$ weeping γ along the same critical line $\beta = \beta_c$, see Fig. 5. The ferromagnetic correlation as a linear average of the density matrix does not depend on the measurement strength γ, as expected. The disorder average of the absolute value is related to the fidelity correlator [53, 54] of the classical-quantum mixed state [33]

$\rho = \sum_s P(s)|s\rangle\langle s| \otimes |\psi(s)\rangle\langle\psi(s)|$:

$$[|C_s|] = \sum_s P(s)\sqrt{C_s^2} = \mathcal{F}(\rho, Z_iZ_j\rho Z_jZ_i), \quad (10)$$

where $\mathcal{F}(\rho_1,\rho_2) = \mathrm{tr}\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}$ is the mixed state fidelity. $[C_s^2]$ is the Edwards-Anderson order parameter often employed to detect the spin glass order. The two non-linear probes here signal the phase transition at γ_c that are roughly consistent with the entropy.

Data collapse and critical exponents

Next we show a detailed exposition of the finite-size scaling data collapse that we employ to extract the critical exponents from I_c:

$$I_c(\gamma) = L^{\eta_\beta/\nu_\gamma} f(d^{1/\nu_\gamma}(\gamma - \gamma_c)), \quad (11)$$

where $f(x)$ is a scaling function. Fig. 6 shows such s data collapse for the learning transition at the critical temperature β_c, when sweeping the measurement strength γ.
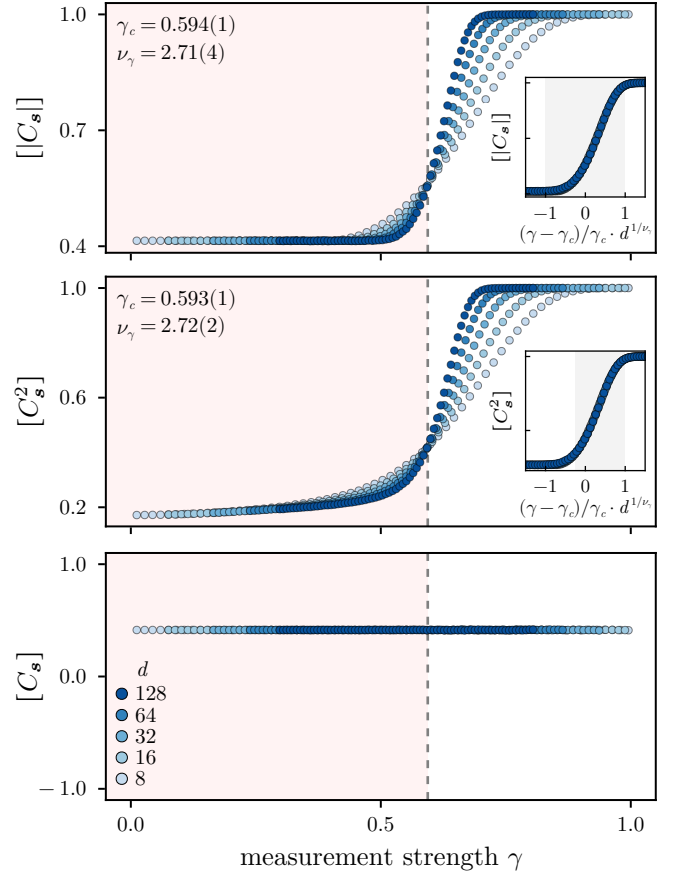

FIG. 5. **Varying moments of domain wall correlation functions**, averaged according to the Bayes / Born rule, along the critical line $\beta = \beta_c$ of sweeping γ.
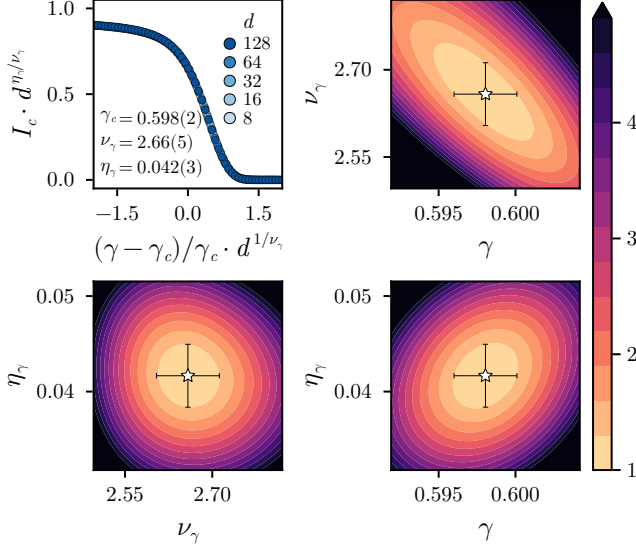
FIG. 6. **Finite-size data collapse** of the average domain wall entropy/coherent information I_c on the critical line $\beta = \beta_c$ Besides the critical measurement strength γ_c the collapse allows to extract the two critical exponents ν_γ and η_γ, with the three contour plots indicating the quality of the fits in relative coordinates.

Critical exponents along the critical lines

Let us supplement Fig. 3 of the main text with detailed fit results for both critical exponents and the critical threshold value from the data collapse along the critical lines are shown in Fig. 7.

Exactly solved models

Here we show two variants of the models discussed in the main text that are exactly solvable. The first one is an all-to-all interacting Ising model bearing similarity to the Sherrington-Kirkpatrick model, both of which are exactly solved by mean field theory. The second one is a 2-replica variant of the model, which can be mapped to the Ashkin-Teller model, see Fig. 8.

Exactly solved all-to-all model

In this Appendix we discuss the effects of measurements on all-to-all ferromagnets. The probability to find the system in configuration $\boldsymbol{\sigma}$ is

$$P(\boldsymbol{\sigma}) \sim e^{\frac{\beta}{N} \sum_{i<j} \sigma_i \sigma_j}, \qquad (12)$$

where we have rescaled the coupling by $1/N$ so that the energy is extensive. Using a Gaussian measurement model, the
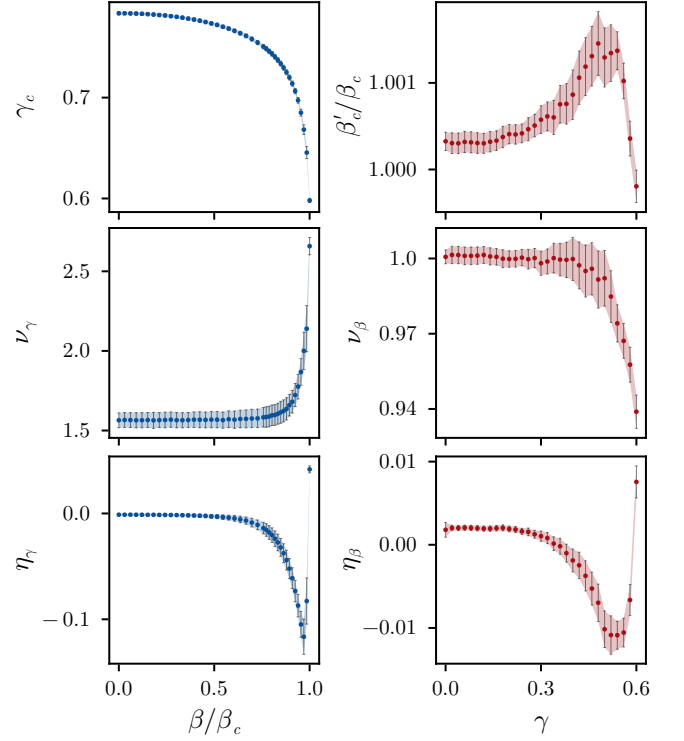

FIG. 7. **Fit critical exponents along the critical lines**. Left panel: scaling exponents of the inference transition (horizontal cuts of the phase diagram in the main text), for the critical line between the Nishimori criticality at infinite temperature $\beta = 0$ and the tricritical point at β_c. Right panel: scaling exponents of the thermal phase transition (vertical cuts of the phase diagram in the main text), for the critical line between the tricritical point at γ_c and the clean Ising point at $\gamma = 0$. β'_c not being exactly β_c is a numerical artifact from the finite size scaling, similar to η not being exactly 0 at the Nishimori point (left plot $\beta = 0$) and the clean Ising point (right plot $\gamma = 0$).

probability density to find outcomes $\boldsymbol{s} = \{s_{ij}\}$ is

$$P(\boldsymbol{s}) \sim \sum_{\boldsymbol{\sigma}} e^{\frac{\beta}{N} \sum_{i<j} \sigma_i \sigma_j - \frac{\gamma^2}{2N} \sum_{i<j} [s_{ij} - \sigma_i \sigma_j]^2}, \qquad (13)$$

where the distribution is normalized as $\int_{-\infty}^{\infty} \prod_{i<j} ds_{ij} P(\boldsymbol{s}) = 1$. Increasing γ corresponds to increasing the correlation between s_{ij} and $\sigma_i \sigma_j$. To calculate averages of (nonlinear) correlations over s, we use a replica trick. The partition function for the n replica theory is $\mathcal{Z}_n = \int d\boldsymbol{s} P^n(s)$ and $d\boldsymbol{s} = \prod_{i<j} ds_{ij}$. Integrating out $\boldsymbol{s}$ we have

$$\mathcal{Z}_n \sim \sum_{\boldsymbol{\sigma}^1 \cdots \boldsymbol{\sigma}^n} e^{\frac{\beta}{N} \sum_\alpha \sum_{j<k} \sigma_j^\alpha \sigma_k^\alpha + \frac{\gamma^2}{nN} \sum_{\alpha<\beta} \sum_{j<k} \sigma_j^\alpha \sigma_k^\alpha \sigma_j^\beta \sigma_k^\beta}, \qquad (14)$$

where the superscript $\alpha = 1, \cdots, n$ is a replica index. We can decouple the interactions between different spins j in two steps. First we introduce variables m^α and $q^{\alpha\beta}$ (for $\alpha < \beta$):

$$\mathcal{Z}_n \sim \int \prod_\alpha dm^\alpha \prod_{\beta>\alpha} dq^{\alpha\beta} e^{-\frac{1}{2} N\beta \sum_\alpha (m^\alpha)^2 + \beta m^\alpha \sum_j \sigma_j^\alpha} \qquad (15)$$

$$\times e^{-\frac{1}{2n} N\gamma^2 \sum_{\alpha<\beta} (q^{\alpha\beta})^2 + \frac{\gamma^2}{n} \sum_{\alpha<\beta} q^{\alpha\beta} \sum_j \sigma_j^\alpha \sigma_j^\beta}.$$
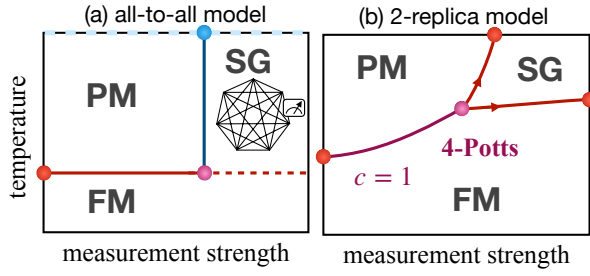
FIG. 8. **Exactly solved models**. (a) Phase diagram of measured all-to-all interacting Ising model, exactly solved by mean field theory akin to the Sherrington-Kirkpatrick model. (b) 2-replica variant of the model, by mapping the 2-replicated partition function $P(s)^2$ to the Ashkin-Teller model.

At large N we use a saddle point approximation: $q^{\alpha\beta} = q$ and $m^\alpha = 1$. Then $\sum_{\alpha<\beta} q^{\alpha\beta}\sigma_j^\alpha\sigma_j^\beta = \frac{1}{2}q[(\sum_\alpha \sigma_j^\alpha)^2 - n]$, and we can absorb the term quadratic in spins using

$$\sum_{\sigma_j^1 \cdots \sigma_j^n} e^{\beta m \sum_\alpha \sigma_j^\alpha - \frac{\gamma^2 q}{2} + \frac{\gamma^2 q}{2n}(\sum_\alpha \sigma_j^\alpha)^2}$$

$$\sim \int dz\, e^{\beta m \sum_\alpha \sigma_j^\alpha - \frac{\gamma^2 q}{2} - \frac{1}{2}z^2 + z\gamma(q/n)^{1/2}\sum_\alpha \sigma_j^\alpha} \qquad (16)$$

$$= 2^n e^{-\frac{\gamma^2 q}{2}} \int dz\, e^{-\frac{1}{2}z^2} \cosh^n\left[\beta m + \gamma(q/n)^{1/2}z\right].$$

In our saddle point approximation, we then have $\mathcal{Z}_n \sim \int dm\,dq\, e^{-Nf_n[m,q]}$ with free energy density

$$f_n[m,q] = \frac{1}{4}(n-1)\gamma^2 q^2 + \frac{1}{2}n\beta m^2 + \frac{1}{2}\gamma^2 q \qquad (17)$$
$$- \ln \int dz\, e^{-z^2/2} \cosh^n\left[\beta m + \gamma(q/n)^{1/2}z\right].$$

To identify the PM-SG boundary we set $m = 0$ and expand the free energy density in powers of q:

$$f_n[0,q] = \frac{1}{4}(n-1)(1 - \gamma^2/n)\gamma^2 q^2 + O(q^3). \qquad (18)$$

For $n = 1 + \epsilon$ the prefactor of q^2 changes sign at $\gamma^2 = 1 + \epsilon$, indicating that for $\epsilon \to 0$ the phase boundary is at $\gamma = 1$ independent of β. To identify the PM-FM boundary we minimize $f_n[m,q]$ with respect to m. The result, for $n \to 1$, is

$$m = \frac{\int dz\, e^{-z^2/2} \sinh[\beta m + \gamma q^{1/2}z]}{\int dz\, e^{-z^2/2} \cosh[\beta m + \gamma q^{1/2}z]}. \qquad (19)$$

Expanding the right-hand side in powers of γ, it can be verified that this self-consistency equation is $m = \tanh[\beta m]$ at all orders. As required, the PM-FM boundary (as well as the SG-FM boundary) is independent of γ.

Exactly solved 2-replica model

Although it is more natural to study conditional expectation values $\langle \cdots \rangle_s$ weighted according to the distribution $P(s)$, for example $\sum_s P(s)\langle\sigma_i\sigma_j\rangle_s^2$, we can develop more insight into the full ensemble of conditional states by weighting averages by powers of $P(s)$. For example, we can ask about the behavior of $\sum_s P^n(s)\langle\sigma_i\sigma_j\rangle_s^2/\sum_s P^n(s)$ with separation as β and γ are varied. For $n > 1$ such averages bias the distribution towards more likely outcomes s. Note also that, through Eq. (6), we can understand these theories as corresponding to a general number n of replicas.

Here, a key advantage of considering such objects is that for $n = 2$ we can solve the resulting statistical mechanics model exactly. This is because the $n = 2$ theory is simply the *Ashkin-Teller model*. Denoting $\tilde{\gamma} = \tanh^{-1}(\gamma)$ we have

$$\sum_s P^2(s) \propto \sum_{s,\sigma,\tau} e^{\sum_{\langle ij\rangle}(\beta+\tilde{\gamma}s_{ij})(\sigma_i\sigma_j+\tau_i\tau_j)}$$

$$\propto \sum_{\sigma,\tau} e^{\sum_{\langle ij\rangle}\beta(\sigma_i\sigma_j+\tau_i\tau_j)+\tanh^{-1}(\gamma^2)\sigma_i\sigma_j\tau_i\tau_j}. \qquad (20)$$

Here we use the notation $\sigma_j = \sigma_j^1$, $\tau = \tau_j^2$ for the spins in the two replicas $a = 1, 2$. The analogue of the spin glass order parameter in this limit is represented by the linear observable $\langle\sigma\tau\rangle$ weighted according the $n = 2$ partition function $\sum_s P^2(s)$.

Note that the solution of the Ashkin-Teller (AT) model provides us with predictions for the behavior of Rényi-2 fidelity correlator [53, 54] in the classical-quantum Rokhsar-Kivelson state $\rho = \sum_s P(s)|s\rangle\langle s| \otimes |\psi(s)\rangle\langle\psi(s)|$. This correlator can be used to detect SW-SSB in the absence of conventional SSB, and is defined by

$$\mathcal{F}_2(\rho, Z_iZ_j\rho Z_jZ_i) = \frac{\mathrm{tr}(Z_iZ_j\rho Z_jZ_i\rho)}{\mathrm{tr}\rho^2}. \qquad (21)$$

Inserting our expression for ρ it is clear that

$$\mathcal{F}_2(\rho, Z_iZ_j\rho Z_jZ_i) = \langle\sigma_i\tau_i\sigma_j\tau_j\rangle_{\mathrm{AT}}. \qquad (22)$$

For comparison, conventional SSB is detected by

$$\mathcal{F}_2(\rho, Z_iZ_j\rho) = \frac{\sum_s P(s)^2\langle Z_iZ_j\rangle}{\sum_s P(s)^2} = \langle\sigma_i\sigma_j\rangle_{\mathrm{AT}}. \qquad (23)$$

Thus the "partial order" in the Ashkin-Teller model i.e. $\langle\sigma\tau\rangle \neq 0$ is equivalent to replica correlation and, in our case, corresponds 'spin glass' order in the $n = 2$ theory.

The phase diagram of the Ashkin-Teller model is schematically shown in Fig. 8b. A qualitative distinction with the 1-replica case is that the ferromagnetic phase is enhanced by the measurement, which raises the critical temperature. The transition line between ferromagnetic and paramagnetic phases is pinpointed exactly by the Kramers Wannier self-duality $\sinh(2\beta) = \exp(-2\tanh^{-1}\gamma^2)$, which exhibits continuously varying critical exponents and are described by the compactified boson conformal field theory (CFT) with central charge $c = 1$. This $c = 1$ critical line starts from the $\gamma = 0$, $\beta = \beta_c$ point, which is described by two decoupled Ising CFTs with $c = 1/2$ respectively. The self-dual critical line terminates at the S_4 symmetric Potts point $\beta_p = \tanh^{-1}\gamma_p^2 =$

$\tanh^{-1}(1/2)/2 = 0.2746...$, namely, $\gamma_p = 0.5176...$. Beyond this point, it splits into two Ising critical lines sandwiching a phase with the so-called "partial order" $\langle \sigma_i \tau_i \sigma_j \tau_j \rangle_{\mathrm{AT}} \neq 0$ in the context of the Ashkin-Teller model. Put in our context, this corresponds to the "spin glass" phase in the 2-replica theory.