

Measurement-induced phase transitions in quantum inference problems and quantum hidden Markov models

Sun Woo P. Kim * and Curt von Keyserlingk 

Department of Physics, *King's College London*, Strand, London WC2R 2LS, United Kingdom

Austen Lamacraft 

TCM Group, Cavendish Laboratory, *University of Cambridge*, Cambridge CB3 0HE, United Kingdom



(Received 8 May 2025; accepted 15 January 2026; published 11 May 2026)

Recently, there is interest in coincident “sharpening” and “learnability” transitions in monitored quantum systems. In the latter, an outside observer’s ability to infer properties of a quantum system from measurements undergoes a phase transition. Such transitions appear to be related to the decodability transition in quantum error correction, but the precise connection is not clear. Here, we study these problems under one framework we call the general quantum inference problem. In cases as above where the system has a Markov structure, we say that the inference is on a quantum hidden Markov model. We show a formal connection to classical hidden Markov models and that they coincide for certain setups. For example, we prove this for those involving Haar-random unitaries and measurements. We introduce the notion of Bayes nonoptimality, where parameters used for inference differs from true ones. This allows us to expand the phase diagrams of above models. At Bayes optimality, we obtain an explicit relation between “sharpening” and “learnability” order parameters, explicitly showing that the two transitions coincide. Next, we study concrete examples. We review quantum error correction on the toric and repetition code and their mapping to two-dimensional (2D) random-bond Ising model (RBIM) through our framework. We study the Haar-random $U(1)$ -symmetric monitored quantum circuit and tree, mapping each to inference models that we call the planted simple symmetric exclusion process and planted XOR, respectively, and expanding the phase diagram to Bayes nonoptimality. For the circuit, we deduce the phase boundary numerically and analytically argue that it is of a single universality class. For the tree, we present an exact solution of the entire phase boundary, which displays re-entrance as does the 2D RBIM. We discuss these phase diagrams, with their interpretations for quantum inference problems and rigorous arguments on their shapes.

DOI: [10.1103/b3s8-y1wb](https://doi.org/10.1103/b3s8-y1wb)

I. INTRODUCTION

Recently, there has been considerable attention on measurement-induced phase transitions on certain monitored quantum circuits. Here, as the rate or strength of measurements is increased, there is a coincident “observable sharpening” and “learnability” transition [1–5]. In the former case, the variance of an observable on the final quantum state undergoes a phase transition. While for the latter, the phase transition is for the ability for a “decoder” to correctly predict the observable given the measurement outcomes. The appearance of inference in the latter case suggests a connection to other active areas in physics: quantum error correction and statistical physics of inference; the latter is less discussed in the quantum community.

Sharpening/learnability transition in monitored circuits. More concretely, in the “sharpening” transition, one prepares an initial state $|\psi_0\rangle$ and performs a monitored evolution characterized by measurement strength or rate ϵ , obtaining a particular trajectory of measurements and final state $|\psi_t\rangle$. Then, for typical trajectories and a particular observable $\hat{O}$, the variance $\delta O^2 = \langle \psi_t | \hat{O}^2 | \psi_t \rangle - \langle \psi_t | \hat{O} | \psi_t \rangle^2$ undergoes a phase transition from a finite value to zero as ϵ increases [1]. In the “learnability” transition, we consider “Alice” who chooses an initial state from an ensemble of states with definite values of the observable $\{|\psi_O\rangle\}$ with probability $p(|\psi_O\rangle)$, performs the monitored evolution, then obtains the measurement outcomes Y . An eavesdropper “Eve” intercepts *only* the measurement outcomes and tries to guess O . The classification error of Eve’s guess undergoes a phase transition from a finite value to zero with ϵ [2].

Quantum error correction. A simplified model for quantum error correction (QEC) can be described as follows [6]. First, we prepare a quantum state $|\psi\rangle$ that encode some quantum information. Then, we suppose that, due to the environment, errors E occur with some probability $p(E)$, sending the state to $\hat{E}|\psi\rangle$. Then, we measure some syndromes S that depends on the errors E . We then use a “decoder” to infer the operation to correct the errors and revert the state back to the original

*Contact author: swk34@cantab.ac.uk

$|\psi\rangle$. A natural question is whether there is a transition in the ability to successfully error-correct as the rate of errors varies. The point at which this occurs is the *threshold*. Within this general setting, we can generalize to a case where our readout of the syndromes is unreliable, and/or the case where we accumulate errors e_t and measure syndromes s_t at each time step t before we apply an operation at some final time.

Bayesian inference. In classical inference problems, we are interested in a state of the system $\mathbf{x}$. As the inferrer, we assume a prior distribution $p(\mathbf{x})$ for the state. We do not have direct access to the state $\mathbf{x}$, but instead have access to observations $\mathbf{y}$ that we assume are correlated with $\mathbf{x}$ according to a likelihood/“observation model” $p(\mathbf{y}|\mathbf{x})$. Our goal is to infer $\mathbf{x}$ given $\mathbf{y}$. To do this, we compute the posterior $p(\mathbf{x}|\mathbf{y})$ according to Bayes’ rule,

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \quad (1)$$

where the marginal likelihood is given by $p(\mathbf{y}) = \sum_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$.

We may also assume additional structure, such as that of a hidden Markov model (HMM) [7,8]. Here, the state at time t , $\mathbf{x}_t$ undergoes a Markov process, and therefore the prior for the entire trajectory is given by

$$p(\mathbf{x}_{1:t}) = \left(\prod_{\tau=2}^t p(\mathbf{x}_\tau | \mathbf{x}_{\tau-1}) \right) p(\mathbf{x}_1), \quad (2)$$

where we have introduced the notation $\mathbf{x}_{1:t}$ to denote a sequence of variables $\mathbf{x}_1, \dots, \mathbf{x}_t$. At every time step, information about the current state $\mathbf{x}_t$ is transmitted through observations $\mathbf{y}_t$ according to $p(\mathbf{y}_t|\mathbf{x}_t)$, leading to total likelihood

$$p(\mathbf{y}_{1:t}|\mathbf{x}_{1:t}) = \prod_{\tau=1}^t p(\mathbf{y}_\tau|\mathbf{x}_\tau). \quad (3)$$

With the additional structure, one can consider various inference tasks. One is a “filtering” task, where the posterior for the current state is inferred from a history of measurements, $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. It can be shown that the unnormalized posterior distribution $q(\mathbf{x}_t|\mathbf{y}_{1:t})$ of the *current* state $\mathbf{x}_t$, given the history of measurements $\mathbf{y}_{1:t}$, evolves linearly and as a Markov process according to the *forward algorithm* [9]

$$q(\mathbf{x}_t|\mathbf{y}_{1:t}) = \sum_{\mathbf{x}_{t-1}} p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})q(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}), \quad (4)$$

with the initial condition $q(\mathbf{x}_1|\mathbf{y}_{1:1}) = p(\mathbf{y}_1|\mathbf{x}_1)p(\mathbf{x}_1)$. In fact, the unnormalized posterior is just the joint distribution $p(\mathbf{x}_t, \mathbf{y}_{1:t}) = q(\mathbf{x}_t|\mathbf{y}_{1:t})$. Importantly, the evolution of the posterior for the filtering task is Markovian; i.e., the posterior for the future time step only depends on that of the current time step $q(\cdot|\mathbf{y}_{1:t-1})$ and observation $\mathbf{y}_t$. This is a consequence of the structure of HMMs [7].

Recently, Ref. [10] studied statistical mechanical and hydrodynamical models of Bayesian inference under the condition of Bayes optimality. Its observations have some overlap with ours under the condition of Bayes optimality, while our use of the teacher-student scenario allows nonoptimal inference to be explored. We explain the notion of Bayes optimality and nonoptimality in Secs. II and III.

General quantum inference problem. In this work, we consider an inference problem that we call the general quantum inference problem. Here, we consider a quantum system where certain channels records their outcomes in classical registers. Some of these registers are revealed to the observer, while others are hidden. Then, the task for the observer is to infer the hidden register given the revealed ones. Within this general framework, one can also define quantum hidden Markov models (qHMMs), which have a similar structure as classical HMMs on the level of density matrices. All the aforementioned inference problems are special cases of the general quantum inference problem.

Outline. In the above discussion, we considered a highly idealized scenario where we have perfect information about the distribution that generates the measurements/syndromes/observations. In reality, one does not know the correct model that generates data. For example, in a QEC setting, one may not know the correct error distribution. In the learnability setting, one may incorrectly estimate the circuit parameters. Accounting for this relevant and realistic scenario requires a quantum framework in which the data generating process and observation model can differ. In the classical setting, this is known as the teacher-student framework, where a “student” infers from data provided by the “teacher.” This terminology is taken from the field of *statistical physics of inference* [11] and is introduced in Sec. II. In Sec. III, we discuss the notion of Bayes optimality in inference problems, equivalent to the “Nishimori condition,” which is where the model assumed by the student agrees with the teacher. We show that we have an exact relationship between order parameters of inference (or “learnability”) and uncertainty (or “sharpening”) in this case. Next, we introduce the general quantum inference problem in Sec. IV, which generalizes the teacher-student framework to the quantum case to allow us to study the above examples, among others. In Sec. V, we consider an additional Markov structure similarly to classical inference problems and introduce qHMMs, which can be viewed as a generalization of HMMs, and discuss under what conditions they reduce to HMMs. As concrete examples, in Sec. VI, we prove that inference for qHMMs with Haar-random unitaries under certain conditions reduce to HMMs.

Next, we study three examples through the lens of inference. First, in Sec. VII, we review two QEC problems under bit-flip errors: the toric code, and the quantum repetition code with read-out errors on the syndrome measurements. We show that both setups can be cast as inference problems and map to the two-dimensional (2D) random-bond Ising model (RBIM), and that the repetition code case has a structure of an HMM. Second, in Sec. VIII we study the Haar-random $U(1)$ -symmetric monitored quantum circuit, which maps to a classical inference problem that we call the *planted simple symmetric exclusion process (planted SSEP)*, where the state of an SSEP is inferred from its noisy “images.” Under the teacher-student scenario, we go beyond existing work and expand the phase diagram to Bayes nonoptimality in student and teacher parameter space. Using a combination of numerical and analytic arguments, we deduce its phase boundary and universality class. Third, in Sec. IX, we go the other way and generate a quantum inference problem from a classical one. We start from the *planted XOR* model, a classical inference

problem on an HMM where a Boolean variable, resulting from XOR operations with a tree structure, is inferred from a series of noisy measurements. This maps to a Haar-random $U(1)$ -symmetric monitored quantum tree. We exploit the structure of the tree to analytically obtain the full expanded phase diagram in the space of student and teacher parameters. In particular, we uncover a re-entrance in the phase diagram, which is also present in the 2D RBIM. Finally, in Sec. X, we discuss the phase diagrams of these inference problems, with a rigorous argument on their shape and their interpretation in the quantum case.

II. TEACHER-STUDENT FRAMEWORK AND THE PLANTED ENSEMBLE

In the above discussion, we considered a highly idealized scenario where we have perfect information about the distribution that generates the measurements/syndromes/observations. A more accurate situation would be when they differ. In statistical physics of inference, there is a standard language to describe the inference problem, namely the teacher-student scenario [11]. Here, we discuss the framework in the classical Bayesian inference setting.

We consider the “teacher,” an underlying phenomenon that produces some random variable $\hat{x}^*$ which we call the teacher variable. The teacher variable is picked from $p_*(\cdot)$, which we will call the teacher’s prior for reasons that will follow. Our phenomenon is then observed, giving rise to a random variable $\hat{y}$ which is a noisy function of $\hat{x}^*$; it is encoded by a conditional distribution $p_*(\hat{y} = y | \hat{x}^* = x^*)$; henceforth, we will denote such expressions as $p_*(y | x^*)$ for brevity. This is a convenient abuse of notation, as we are distinguishing different distributions by their arguments. We will refer to this conditional distribution as the likelihood or the observation model.

We, the observer, or “student,” are given $\hat{y}$. We have a probabilistic model, which is not necessarily correct, for how $\hat{y}$ relates to the underlying phenomenon $\hat{x}^*$. We call the corresponding random variable $\hat{x}$, and whose distribution is encoded by the conditional probability $p_s(x | y)$. If the student had a perfect model of the experiment, then we would have $p_s(\cdot | y) = p_*(\cdot | y)$.

As the student, our probabilistic model of the experiment would be the *posterior* assuming a particular likelihood $p_s(x | y)$ and prior $p_s(x)$. For the remainder of the article, we will consider the case where the student has access to the functional forms of the distributions but does not necessarily the correct parameters. Therefore, we will denote the general functional forms without subscripts, for example, likelihood as $p(x | y)$, and use s or $*$ to specify if it is the student’s or the teacher’s.

The joint distribution

$$\begin{aligned} p(x, y, x^*) &= p_s(x | y) p_*(y | x^*) p_*(x^*) \\ &= \frac{p_s(y | x) p_s(x) p_*(y | x^*) p_*(x^*)}{p_s(y)} \end{aligned} \quad (5)$$

is known as the *planted ensemble* [11]. In some cases, it may be beneficial to regard the student’s posterior $p_s(x | y)$ as a Gibbs distribution. This arises, for example, in the study of

spin glasses. In this language, we can map $p_s(y | x) p_s(x) \leftrightarrow e^{-\beta H(x | y)}$ and the marginal likelihood/evidence $p_s(y) \leftrightarrow Z(y)$ as a partition function, dependent on some “disorder” y . In disordered statistical mechanical systems, y is often chosen iid. In the planted ensemble, by contrast, it is distributed as

$$p(y) = \sum_{x^*} p_*(y | x^*) p_*(x^*), \quad (6)$$

so the distribution is affected by the teacher configurations x^* that give rise to it. x^* is sometimes said to be “planted” in y , which gives rise to the name of the ensemble. In the limit where y is independent of x^* , where the observations have zero signal-to-noise ratio (SNR) (e.g., when y is iid), we revert to a typical scenario of disordered stat-mech models.

III. BAYES OPTIMALITY AND MOMENT RELATIONS

In statistical physics of Bayesian inference, when the student and teacher distributions coincide ($*$ = s), the student’s model is said to be “Bayes optimal.” The naming is apt for the following reasons. Consider a classifier $h(y)$ that predicts the hidden state given the observations y . Then, the classification performance is given by the average overlap between $h(y)$ and x^* , $\mathbb{E}_{y, x^*} [\delta_{h(y), x^*}]$, where $\delta_{\cdot, \cdot}$ is the Kronecker delta. Then, the optimal classifier that maximizes the overlap is argmax of the true posterior [12],

$$h_{\text{optimal}}(y) = \text{argmax}_x p_*(x | y). \quad (7)$$

For real-valued state observables $O(x)$, consider an estimator $\tilde{O}(y)$ given y . One possible figure of merit would be the mean squared error (MSE) $\mathbb{E}_{y, x^*} [(\tilde{O}(y) - O(x^*))^2]$. Then, the optimal estimator that minimizes mean squared error of the estimator is the expectation of the true posterior [13],

$$\tilde{O}_{\text{optimal}}(y) = \mathbb{E}_{x \sim p_s(\cdot | y)} [O(x)]. \quad (8)$$

At Bayes optimality, the joint distribution given by the planted ensemble [Eq. (5)] is symmetric with respect to $x \leftrightarrow x^*$. This means that the marginal distribution for x must have the same distribution as that of x^* . In turn, sampling from the student’s posterior distribution is identical to sampling from the true distribution.

In the spin glass and statistical physics of inference literature, Bayes optimality is also known as the Nishimori condition [11].

Given an inferred posterior $p_s(x | y)$ and an observable $O(x)$, there are a few natural “order parameters” that quantify the performance of the inference.

The first natural quantity is the mean squared error of the mean estimator from the inferred posterior (MSEM),

$$\text{MSEM} := \mathbb{E}_{y, x^*} [(\langle O(x) \rangle_s - O(x^*))^2], \quad (9)$$

where $\langle O(x) \rangle_s = \mathbb{E}_{x \sim p_s(\cdot | y)} [O(x)]$ is the expectation on the inferred posterior. It can be viewed as the student’s estimator $\tilde{O}(y)$. Note that at Bayes optimality, the mean on the inferred posterior is the optimal estimator.

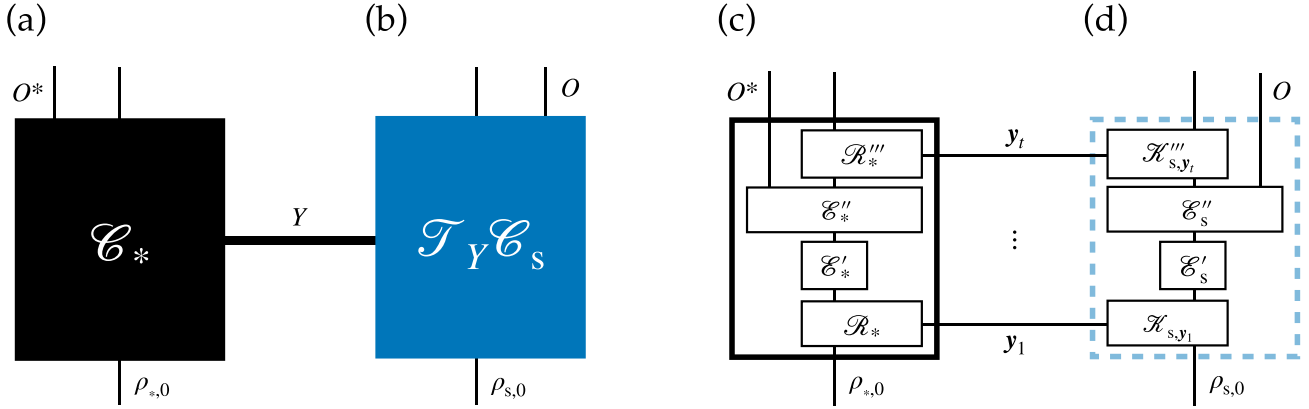


FIG. 1. General quantum inference problems and on qHMMs. (a), (b) In the general quantum inference problem, the “teacher” (a) applies a quantum channel $\mathcal{C}_*$ to their quantum state, which generates a hidden register O^* and revealed registers Y . Given Y , the “student” (b) conditions their channel by projecting on to Y via $\mathcal{T}_Y[\cdot] := \text{tr}_Y[\mathcal{P}_Y(\cdot)]$. The student assumes a channel $\mathcal{C}_s$, which may differ from the teacher’s. Given their conditioned density matrix $\rho_s(Y) = \mathcal{T}_Y \mathcal{C}_s[\rho_{s,0}]$, the student can compute a posterior on O given Y , $p_s(O|Y) = \text{tr}[O\rho_s(Y)]$. (c), (d) Additional causal structure can be put on quantum inference problems. If the channel at each time step does not take previous revealed registers as input, then the evolution of the student’s density matrix is Markovian and inference is on a qHMM.

The second natural quantity in the context of the planted ensemble is the MSE,

$$\text{MSE} := \mathbb{E}_{\mathbf{x}, \mathbf{y}, \mathbf{x}^*} [(O(\mathbf{x}) - O(\mathbf{x}^*))^2], \quad (10)$$

where the expectation is over the planted ensemble [Eq. (5)]. This case can be thought of as the student obtaining their estimator $\tilde{O}(\mathbf{y})$ by sampling from their posterior.

The third quantity is the variance of O on the inferred posterior distribution,

$$\delta O_s^2 := \mathbb{E}_{\mathbf{y}, \mathbf{x}^*} [\langle O(\mathbf{x})^2 \rangle_s - \langle O(\mathbf{x}) \rangle_s^2], \quad (11)$$

which measures the uncertainty of the student on their estimate of the observable.

At Bayes optimality, exploiting the symmetry between the student and teacher replicas, we can derive an exact relation between Eqs. (9) and (10),

$$\begin{aligned} \delta O_s^2(\text{Bayes optimal}) &= \text{MSEM}(\text{Bayes optimal}) \\ &= \frac{1}{2} \text{MSE}(\text{Bayes optimal}). \end{aligned} \quad (12)$$

Therefore, on average, uncertainty in the student’s posterior is equal to its error. The derivation is detailed in Appendix E.

IV. THE GENERAL QUANTUM INFERENCE PROBLEM

In this section, we use the language of the teacher-student scenario discussed in Sec. II to unify inference problems in quantum systems. We call this framework the general quantum inference problem.

Consider the teacher who prepares a quantum state $\rho_{*,0}$. This may be a mixed state, if they were to choose from an ensemble of pure state with some probability, or a pure state. The teacher applies a quantum channel $\mathcal{C}_*$ to the state. The internal structure of the channel may be complicated, but let us assume that the teacher records some of the outcomes in classical registers. For example, they may be outcomes of measurements, or records of application of error channels. Out of all the registers, the teacher will reveal some of these $|Y\rangle\langle Y|$ to the student, while hiding others $|O^*\rangle\langle O^*|$. This is

shown schematically in Fig. 1(a). After the application of the channels the teacher’s density matrix can be written as

$$\mathcal{C}_* \rho_{*,0} = \sum_{O^*, Y} \rho_*(O^*, Y) \otimes |Y\rangle\langle Y| \otimes |O^*\rangle\langle O^*|, \quad (13)$$

$$= \sum_Y \rho_*(Y) \otimes |Y\rangle\langle Y|. \quad (14)$$

In the second line, we have grouped $\rho_*(Y) = \sum_{O^*} \rho_*(O^*, Y) \otimes |O^*\rangle\langle O^*|$. Its trace gives the probability of the teacher recording Y , $p_*(Y)$. Additionally, the joint probability of recording both O^* and Y is given by $p_*(O^*, Y) = \text{tr}[\rho_*(O^*, Y)]$.

The goal of the student is to infer the hidden register O^* given the revealed registers Y . To do this, the student simulates the density matrix conditioned—in a manner we describe below—on the revealed registers. In their simulation, the student applies a channel $\mathcal{C}_s$. This channel may differ from the teacher’s in general. As discussed before, we assume that it has the same functional form as $\mathcal{C}_*$ and only differs in its parameters, unless stated otherwise. After $\mathcal{C}_s$, the student conditions on the revealed register Y by applying $\mathcal{T}_Y[\cdot] := \text{tr}_Y[\mathcal{P}_Y(\cdot)]$. Here, $\mathcal{P}_Y$ is a projector onto the revealed registers and the trace is over the register degrees of freedom $\mathcal{Y}$. This in general will reduce the norm of the student’s density matrix. The student’s procedures are shown in Fig. 1(b). Then, the student’s unnormalized density matrix conditioned on Y is

$$\begin{aligned} \rho_s(Y) &= \mathcal{T}_Y \mathcal{C}_s \rho_{s,0} \\ &= \sum_O \rho_s(O, Y) \otimes |O\rangle\langle O|, \end{aligned} \quad (15)$$

where O denotes the student’s hidden register. The trace of Eq. (15) is the probability that the student’s model produces register Y .

The student can calculate their posterior distribution for the observable as

$$p_s(O|Y) = \frac{\text{tr}[\mathcal{P}_O \rho_s(Y)]}{\text{tr}[\rho_s(Y)]}, \quad (16)$$

where $\mathcal{P}_O$ projects the student's hidden registers onto a particular outcome O .

For the purposes of analysis, it will be convenient to consider the student's density matrix simulation as an actual density matrix. Then, the joint density matrix between the student, the revealed registers, and the teacher is given by

$$\rho = \sum_Y \frac{\rho_s(Y)}{\text{tr}[\rho_s(Y)]} \otimes |Y\rangle\langle Y| \otimes \rho_*(Y), \quad (17)$$

where similarly to the student, $\rho_*(Y) = \mathcal{T}_Y \mathcal{C}_* \rho_{*,0}$. Here, we have normalized the student's density matrix so that the joint density matrix is properly normalized. Equation (17) is the quantum analog of the planted ensemble [Eq. (5)].

At the end, the teacher can “grade” the student's inference. One possibility is that the student calculates an estimator $\tilde{O}(Y)$ by taking the expectation over their posterior, $\tilde{O}(Y) = \sum_O O p_s(O|Y)$, then hands it to the teacher. The teacher then calculates the MSEM. $\tilde{O}(Y)$ can be expressed in terms of Eq. (17) as $\tilde{O}_Y = \text{tr}[O_s \mathcal{P}_Y \rho]$, where the operator O_s acts on the student's register. Then, the MSEM averaged over the experimental runs is found by

$$\text{MSEM} = \sum_Y \text{tr}[(O_* - \tilde{O}(Y))^2 \mathcal{P}_Y \rho], \quad (18)$$

where O_* is an operator acting on the teacher's hidden register. Another way the student could produce an estimator is by sampling from their posterior, $\tilde{O}(Y) \sim p_s(\cdot|Y)$, which is then averaged over the experimental runs. This gives the MSE, which can be written in terms of the joint density matrix as

$$\text{MSE} = \sum_Y \text{tr}[(O_* - O_s)^2 \mathcal{P}_Y \rho]. \quad (19)$$

Within this formalism, we can also study the observable sharpening transition. In this case, we can simply ignore the student by tracing them out. Then, the observable variance of the state can be calculated by

$$\delta O_*^2 = \sum_Y \left(\frac{\text{tr}[O_*^2 \mathcal{P}_Y \rho]}{\text{tr}[\mathcal{P}_Y \rho]} - \frac{\text{tr}[O_* \mathcal{P}_Y \rho]^2}{\text{tr}[\mathcal{P}_Y \rho]^2} \right). \quad (20)$$

One can also consider the observable variance of the student's posterior δO_s^2 . This is given by replacing $*$ $\rightarrow$ s in Eq. (20). Note that at Bayes optimality the joint density matrix Eq. (17) is symmetric with respect to exchange of student and teacher. Therefore,

$$\delta O_*^2 = \delta O_s^2 (\text{Bayes optimal}). \quad (21)$$

That is, the observable variance of the student's posterior is equal to the observable variance of the true quantum state.

Moreover, using the moment relations of Eq. (12), this exactly implies that the “observable-sharpening” transition corresponds to the “learnability” transition at Bayes optimality.

In the field of monitored quantum circuits, “forced measurements” were introduced to make analysis of measurement-induced phase transitions tractable [14]. Here, the measurement outcomes $Y = (y_1, y_2, \dots)$ are chosen to be iid. In this case, no information is provided by the teacher about the quantum state, which in general is the case whenever

the joint distribution $p_*(O^*, Y)$ between the hidden observable O^* and observables Y factorizes.

V. INFERENCE ON QUANTUM HIDDEN MARKOV MODELS

Now consider the case where the whole channel $\mathcal{C}$ can be broken down into a sequence of channels in time, as in Fig. 1(c). There are two kinds of channels in this sequence. For the first kind, denoted as $\mathcal{R}_*$, the outcomes are recorded, while for the second kind, $\mathcal{E}_*$, they are not.

For the recorded channels, the teacher records the outcome in a register as

$$\mathcal{R}_* : \rho_* \rightarrow \sum_a K_{*,a} \rho_* K_{*,a}^\dagger \otimes |a\rangle\langle a|. \quad (22)$$

Here, $\{K_{*,a}\}_a$ are Kraus operators which obey the canonical Kraus map condition. Out of all the registers, some $Y = y_{1:t}$ will be revealed to the student, while others O^* will not.

Now, we assume that the channel applied by the teacher at each step does not depend on the outcomes of the previous *revealed* registers. Then, the student can condition each channel with the revealed register y_t , separately. Recalling the definition $\mathcal{T}_{y_t}[\cdot] = \text{tr}_{\mathcal{Y}_t}[\mathcal{P}_{y_t}(\cdot)]$, where $\mathcal{Y}_t$ is the Hilbert space of the revealed registers at time t in this case,

$$\mathcal{T}_{y_t} \mathcal{R}_s = \mathcal{K}_{s,y_t} : \rho_s \rightarrow K_{s,y_t} \rho_s K_{s,y_t}^\dagger. \quad (23)$$

In this case, we say that the inference is on a qHMM [15].

We can formally connect qHMMs to HMMs as follows. Consider one “time step” as the interval between applying a operator conditioned on a revealed register $\mathcal{K}_{y_t}$. We group all channels between such Kraus maps as $\mathcal{E}_t$. Then, in one time step, the student's density matrix evolves as

$$\rho_s(y_{1:t}) = \mathcal{K}_{s,y_t} \mathcal{E}_{s,t} \rho_s(y_{1:t-1}). \quad (24)$$

Now consider the forward equation [Eq. (4)] for HMMs on a student's model. If we represent the unnormalized posterior distribution as a probability vector, $|q_s(y_{1:t})\rangle$, such that $q_s(\mathbf{x}_t | y_{1:t}) = \langle \mathbf{x}_t | q_s(y_{1:t}) \rangle$, then the forward equation can be written as

$$|q_s(y_{1:t})\rangle = \mathbf{K}_{s,y_t} \mathbf{E}_{s,t} |q_s(y_{1:t-1})\rangle. \quad (25)$$

Here, $\mathbf{K}_{s,y_t}$ is the diagonal measurement model/likelihood matrix defined as $\langle \mathbf{x}_t | \mathbf{K}_{s,y_t} | \mathbf{x}_t \rangle = p_s(y_t | \mathbf{x}_t)$, and $\mathbf{E}_{s,t}$ is the prior Markov kernel matrix defined as $\langle \mathbf{x}_{t+1} | \mathbf{E}_{s,t} | \mathbf{x}_t \rangle = p_s(\mathbf{x}_{t+1} | \mathbf{x}_t)$.

Therefore, the unconditioned channel $\mathcal{E}_t$ and conditioned channel $\mathcal{K}_{y_t}$ of qHMM corresponds to the prior Markov kernels and measurement model/likelihood in HMM, respectively. As mentioned earlier, the only Bayes nonoptimal situations we consider are where the student's model has the same functional form as an optimal model of the system but perhaps the wrong parameters. Note, however, that the student is free to use a qHMM model described by Eq. (24) irrespective of the true structure of the teacher's channel, which could even be non-Markovian. In this case the teacher would have the general structure of Fig. 1(a) while the student uses Fig. 1(d).

In general, the student may need to simulate the full density matrix in order to obtain an accurate estimate for the hidden

register O^* . However, depending on the particular hidden register that the student is trying to infer and the structure of the channels, it may be sufficient for the student to evolve a reduced density matrix on a subset of degrees of freedom whose evolution is still Markovian. Furthermore, it may be that the reduced density matrix is completely diagonal in a certain basis, $\rho_s^{\text{reduced}} = \sum_{\mathbf{x}_t} q_s(\mathbf{x}_t | \mathbf{y}_{1:t}) |\mathbf{x}_t\rangle \langle \mathbf{x}_t|$. In these cases, where the quantum system is accurately captured by a classical HMM, the diagonal matrix elements $q_s(\mathbf{x}_t | \mathbf{y}_{1:t})$ are exactly described by Eq. (25). Therefore, HMMs are a subset of qHMMs. In the remainder of this work, we will describe many such cases where qHMMs reduce to HMMs.

VI. THE CASE OF HAAR-RANDOM UNITARIES

In this section, we give some explicit examples of where classical HMMs provide good, even optimal, predictions for interacting quantum systems. Specifically, we will consider the case where the teacher's evolution involves applying Haar-random gates and measurements to a many-qubit system. It may initially seem surprising that classical models can capture the full complexity of a quantum many-body system. However, in one limit, the ignorance of the student, and in the other (that of large on-site Hilbert space dimension d) the self-averaging qualities of the dynamics give rise to effectively classical behavior.

Consider Haar-random unitary gates [16] with some block diagonal structure in the computational basis. This could be due to an imposed symmetry. For example, taking the local Hilbert space to be $\mathcal{H}_{\text{loc}} = \mathbb{C}^2 \otimes \mathbb{C}^d$ (a qubit-qudit pair), we can consider $U(1)$ -symmetric gates that preserve the particle number in the qubit sector when acting on the overall Hilbert space $\mathcal{H}_{\text{loc}}^{\otimes N}$. A two qubit gate would have the block diagonal structure,

$$u_{U(1)} = u^{(2)} \oplus u^{(1)} \oplus u^{(0)}. \quad (26)$$

Each block acts on the 2, 1, and 0 particle sector and has dimension $d^2 \times d^2$, $2d^2 \times 2d^2$, and $d^2 \times d^2$, respectively.

Then, we combine such Haar-random unitary gates in sequence with measurements of the observables that are also diagonal the computational basis. For example, a weak measurement channel on qubits with continuous outcomes could be

$$\rho \rightarrow \int dy Q(y) \rho Q(y) \otimes |y\rangle \langle y|, \quad (27)$$

where the weak measurement operator (or partial projector) $Q(y)$ is [17]

$$Q(y) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{(y - \epsilon\hat{Z})^2}{4\sigma^2}\right]. \quad (28)$$

Here, y , ϵ , and σ are measurement outcome, strength, and width, respectively.

In this setting, we consider two scenarios of inference problems, and discuss when they can be mapped to classical HMMs.

A. Hidden Haar unitaries

The Haar unitary channel where the teacher records the choice of unitary is given by

$$\rho \rightarrow \int d\mu(u) u \rho u^\dagger \otimes |u\rangle \langle u|. \quad (29)$$

In the first case, we consider when the teacher does not reveal their choice of Haar unitary. In this case, the classical register does not participate in any future operations. Therefore, we are free to trace over these degrees of freedom, leaving us with the memoryless Haar unitary channel,

$$\rho \rightarrow \int d\mu(u) u \rho u^\dagger. \quad (30)$$

Equation (30) is completely depolarizing on each sector and returns an equally weighted ensemble of diagonal density matrices in the sector. The measurements on the qubits are also on diagonal observables. The result is that charge diagonal observables obey a closed set of equations of motion amongst themselves. Therefore other degrees of freedom can be safely traced out, and the circuit can be mapped to a classical HMM.

For example, our two-site $u_{U(1)}$ Haar-random unitary channel is equal to

$$\rho \rightarrow \mathcal{E}_{\text{SSEP}} \otimes \mathcal{D}_{\text{qudits}}[\rho], \quad (31)$$

where $\mathcal{E}_{\text{SSEP}}$ is the incoherent SSEP [18] channel on the qubits, and $\mathcal{D}_{\text{qudits}}$ is the depolarizing channel on the qudits. On the diagonal entries of density matrices in the qubit computational basis, $\mathcal{E}_{\text{SSEP}}$ acts as an SSEP Markov kernel on two Ising variables s and s' corresponding to on-site charges

$$p_{\text{SSEP}}(s_t, s'_t | s_{t-1}, s'_{t-1}) = (s_t s'_t | \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} | s_{t-1} s'_{t-1}). \quad (32)$$

The measurement operator maps to likelihood/measurement model. For example, the weak measurement [Eq. (28)] maps to the measurement model

$$p(y|s) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{(y - \epsilon s)^2}{2\sigma^2}\right]. \quad (33)$$

Finally, the prior at initial time is $p(s_1) = \text{tr}[\mathcal{P}_{s_1} \rho_0]$. Thus, in the setting where the teacher does not reveal their specific choices of $U(1)$ Haar unitaries, the teacher's experimentally measured outputs are indistinguishable from those generated by an effectively classical HMM. Thus, for the purposes of prediction, the student cannot do any better than to use this classical HMM.

B. Revealed Haar unitaries

In the next case, the teacher stores the choice of Haar unitary in a classical register and reveals it to the student as in Eq. (29).

In this case, coherence is intact for each trajectory. One might imagine that the student will have to simulate the full quantum evolution for an optimal posterior. Now, we consider

that the Haar unitary acts on qubit-qudit pairs (with qudit dimension d), with the same block diagonal structure on the qubits as before. In the limit of $d \gg e^{LT}$, the typical gate u is well approximated by its Haar average. For example, for our two-site $u_{U(1)}$,

$$u_{U(1)}^{\text{typ}}[\cdot](u_{U(1)}^{\text{typ}})^{\dagger} \stackrel{d \gg e^{LT}}{\approx} \mathcal{E}_{\text{SSEP}} \otimes \mathcal{D}_{\text{qudits}}[\cdot]. \quad (34)$$

Precisely, in Appendix G, we rigorously prove that, for any circuit with Haar-random unitary gates acting on k qubit-qudit pairs with block diagonal structure (with each block acting on all qudits) and weak measurements on number of qubits such that there are extensive number ($\sim VT$) of such elements, we have

$$\mathbb{E}_{U,Y} [|p(s|Y) - \bar{p}(s|Y)|] \leq C \frac{V \text{Tr}_1^Y r_2^{YT}}{d^{k/2}}, \quad (35)$$

where C , $r_{1,2}$ are some constants independent of d , V is the volume, T is the total number of time steps, and $\bar{p}(s|Y)$ is the posterior derived from the Haar-averaged channels, i.e., using the channel on the RHS of Eq. (34) in place of the unitaries in order to deduce the posterior.

Previously, that $\bar{p}(s|Y)$ well-approximate $p(s|Y)$ was checked at the level of the replica trick and in the formal limit $d \rightarrow \infty$ for a quantum circuit with two-qubit $u_{U(1)}$ local unitaries with onsite projective measurements, in Ref. [1].

In the $d \gg e^{VT}$ limit, Haar unitaries are self-averaging. For that reason, the student's inferences are not improved further even if the teacher reveals the particular unitaries used; optimal inference is achieved with the ensemble-averaged circuit, which is a classical HMM model.

Finally, note that disregarding (tracing out) the student sector leaves behind a monitored quantum circuit. In the large d limit, any observables on qubits diagonal in the computational basis are given by a classical stochastic model (an HMM), while the “maximally quantum” $d = 1$ limit cannot be reduced to such classical models. Therefore, qudit dimension d can be used to interpolate between the two limits.

VII. RANDOM BOND ISING MODEL AND QUANTUM ERROR CORRECTION

In this section, we review the emergence of the 2D RBIM for two QEC scenarios originally discussed in Ref. [6]: QEC with bit-flip errors on the toric code and one-shot syndrome measurements, and on the repetition code and repeated noisy syndrome measurements with read-out errors. For a detailed introduction, see Ref. [6].

First, we look at these QEC scenarios as quantum inference problems on qHMMs. The toric code case has a trivial Markov structure, as there is only one cycle of errors followed by syndrome measurements. On the other hand, the repetition code case has an extended temporal structure as measurements and errors channels are repeated. The choice of bit-flip errors will mean these scenarios reduce to inference on classical HMMs.

With the mapping to HMMs established, we adopt the language of Bayesian inference in the teacher/student setting for HMMs and discuss each scenario.

A. Quantum error correction as a quantum inference problem

We consider scenarios where the teacher prepares an encoded logical qubit $|\psi\rangle = \alpha|\bar{0}\rangle + \beta|\bar{1}\rangle$ comprised of many physical qubits. For the toric and repetition codes, the physical qubits live on links l of the 2D square lattice and 1D chain, respectively.

The teacher randomly applies bit-flip errors on the physical qubits. Importantly, the teacher keeps a record of links where a bit-flip error is *currently* applied in an “environment” $|\mathbf{f}\rangle\langle\mathbf{f}|$ where $\mathbf{f} = (f_l)_l$. Here, we introduced the compact tuple-builder notation $(f_l)_l := (f_l : l \in \text{all links})$.

When there is a bit-flip is applied on link l , the register is set to $f_l = -1$, and to $f_l = 1$ otherwise. Therefore, the environment is initialized as $|\mathbf{1}\rangle\langle\mathbf{1}|$.

This is modeled by a bit-flip error channel that acts with rate π_* and also updates the record of the currently applied errors, given by

$$\rho_* \rightarrow \pi_*(X_l \otimes X_l^{\text{env}})\rho_*(X_l \otimes X_l^{\text{env}}) + (1 - \pi_*)\rho_* \quad (36)$$

$$= \sum_{e_l} K_{*,e_l} \rho_* K_{*,e_l}^{\dagger}. \quad (37)$$

Here, X_l acts on the physical qubits, whereas X_l^{env} acts on the environment registers. In the following sections, we will relate error rate π_* with fidelity β_* as $\pi_* = e^{-\beta_*}/2 \cosh \beta_*$. In the second line, we wrote the channel in terms of Kraus operators, $K_{e_l} = \sqrt{p(e_l)}(X_l \otimes X_l^{\text{env}})$. There, occurrence of an error and no error is denoted as $e_l = -1$ and $e_l = 1$, respectively. Its probability is given by $p(e_l) = e^{-\beta_* e_l}/2 \cosh(\beta_*)$.

Next, the teacher measures the syndromes and records the outcomes. For the toric code, the syndrome is $\hat{s}_P = \hat{Z}_P = \prod_{l \in P} \hat{Z}_l$, where P is a plaquette on the square lattice and $l \in P$ denotes the links neighboring P . For the repetition code, the syndrome lives on the vertex and is $\hat{s}_v = \hat{Z}_v = \prod_{l \in v} \hat{Z}_l$. The syndrome measurement channel only acts on the quantum state and is given by

$$\rho_* \rightarrow \sum_s \mathbb{P}_s \rho_* \mathbb{P}_s \otimes |s\rangle\langle s|. \quad (38)$$

For the repetition code, we will consider read-out errors. There, there is a further error channel on the syndrome register, i.e., the *true* syndrome register s^* is now a hidden register, and only the (potentially) corrupted syndrome register $s = e \cdot s^*$ is revealed. This corresponds to the channel

$$\rho_* \rightarrow \sum_s \sum_{s^*} \mathbb{P}_{s^*} \rho_* \mathbb{P}_{s^*} \otimes \sqrt{p_*(s|s^*)}|s\rangle\langle s| \sqrt{p_*(s|s^*)}. \quad (39)$$

We consider the case where the probability of occurrence of read-out error is the same as that of the bit-flip error, π_* [6]. Therefore, we have $p(s = e \cdot s^* | s^*) = p(e) = e^{-\beta_* e}/2 \cosh \beta_*$.

Then, the syndromes S are revealed to the student, while the environment register is hidden from them. The goal of the student is to infer the hidden environment register from the revealed syndromes.

Although the above discussion was restricted to toric and repetition codes, the above can be easily generalized to any QEC models, replacing the bit-flip errors, codes, and syndrome measurements to arbitrary ones. As long as the revealed syndrome measurements do not participate in subsequent

channels, as is the case above, the quantum inference problem is on a qHMM.

In the special case of bit-flip errors on toric and repetition codes discussed, the inference problem can be reduced to that on classical HMMs. Application of bit-flip errors will result in a statistical mixture of states $|\tau\rangle := \prod_l X^{(1+\tau_l)/2} |\psi\rangle$. If the environment register faithfully records the errors, the density matrix will be a statistical mixture $\sum_\alpha p_\alpha |\tau^{(\alpha)}\rangle \langle \tau^{(\alpha)}| \otimes |f^{(\alpha)}\rangle \langle f^{(\alpha)}|$, where $\tau_l^{(\alpha)} = f_l^{(\alpha)}$. On such states, tracing out the Hilbert space of the quantum state, we reduce to an HMM on the hidden environment registers. The error channel [Eq. (36)] reduces to

$$p_*(f_{l,t}^* | f_{l,t-1}^*) = \frac{\exp(-\beta_* f_{l,t}^*, f_{l,t-1}^*)}{2 \cosh \beta_*}, \quad (40)$$

while for the syndrome measurement channels [Eqs. (38) and (39)], tracing out the quantum state simply moves the action of projectors to the environment register, resulting in measurement models $p_*(s_P^* | (f_l^*)_{l \in P})$ as in Eq. (42) for the toric code and $p_*(s_v^* | (f_l^*)_{l \in v})$ as in Eq. (46) for the repetition code.

B. Toric code

For the toric code scenario, we consider single round of errors, followed by projective measurements of syndromes, as pictured in Fig. 2(a). Therefore, we can just consider the occurrence of errors $E^* = (e_l^*)_l$. In the mapped classical inference problem, teacher first picks the true error chain E^* from their prior

$$p_*(E^*) = \prod_l p_*(e_l^*) \propto \prod_l \exp(\beta_* e_l^*), \quad (41)$$

then generates syndromes $S = (s_P)_P$ from their likelihood/observation model. As we consider noiseless projective measurements, this is given by

$$\begin{aligned} p_*(S|E^*) &= \prod_P p_*(s_P | (e_l^*)_{l \in P}) \\ &= \prod_P \delta \left\{ s_P, \prod_{l \in P} e_l^* \right\} \\ &= \delta_{S, \partial E^*}, \end{aligned} \quad (42)$$

where $\delta\{\cdot, \cdot\}$, $\delta_{\cdot, \cdot}$ are both Kronecker deltas, and ∂E^* denotes the boundary of the error chain. Here, boundary is defined on the dual lattice and therefore lives on the plaquettes. Note that the process of generating observations (syndromes in this case) is completely deterministic. The failure of perfect inference stems from the fact that the map $E^* \rightarrow S$ is many-to-one.

Given the syndromes $S = (s_P)_P$, the student can infer a posterior distribution assuming student parameters $\pi = e^{-\beta}/2 \cosh(\beta)$,

$$p_s(E|S) = \frac{p_s(S|E)p_s(E)}{p_s(S)}.$$

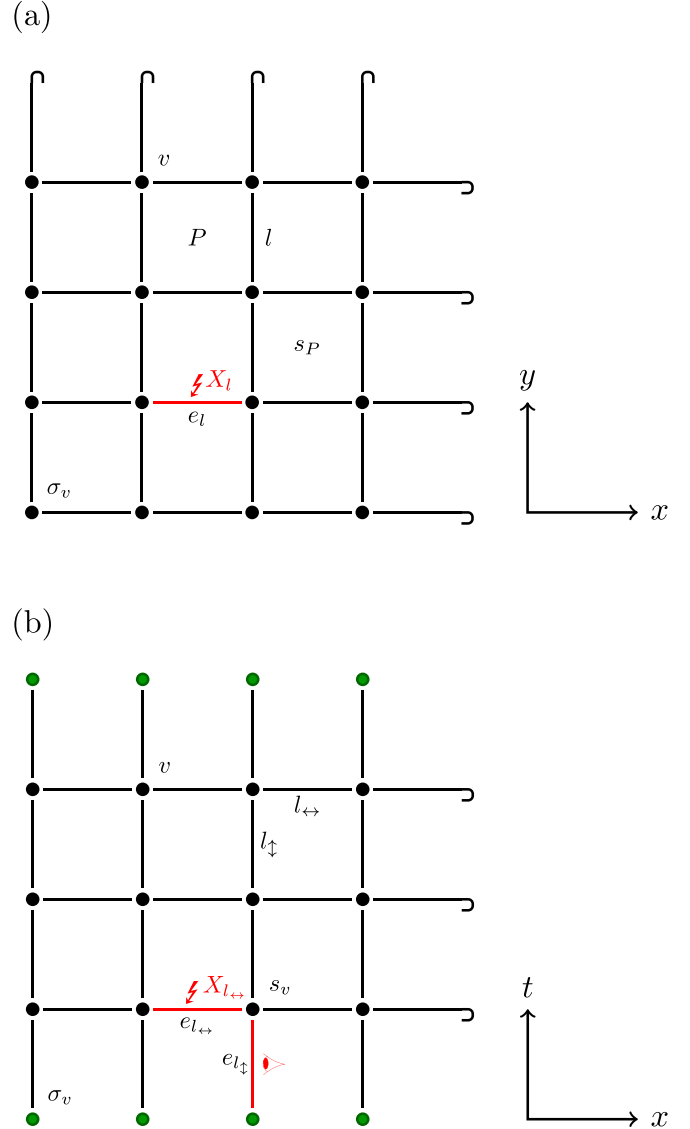


FIG. 2. Schematic of the set up for mapping the quantum error correction on (a) the toric code and (b) quantum repetition code to the 2D random-bond Ising model. The curly links denote periodic boundary conditions. For the toric code, only bit-flip errors (lightning symbol) for a single timeslice is considered. For the repetition code, read-out errors (eye symbol) and multiple time steps are considered, which we view as a space-time lattice. Here, the green vertices denote fixed boundary condition of $\sigma_l = 1$ at $t = 0$ and t_f .

Marginalizing over the syndromes, the conditional probability of the student inferring E given that the teacher picked E^* is

$$\begin{aligned} p(E|E^*) &= \sum_S p_s(E|S) p_*(S|E^*) \\ &= \sum_S \frac{\delta_{S, \partial E} p_s(E) \delta_{S, \partial E^*}}{p_s(S)} \\ &\propto \delta_{\partial E, \partial E^*} p_s(E). \end{aligned}$$

The set of inferred errors E that respect the delta function only contains error chains such that $E = C \odot E^*$, where $\odot$ is elementwise multiplication, such that $e_l = e_l^* c_l$. Here, $C = (c_l)_l$,

$c_l \in \{-1, +1\}$ and links where $c_l = -1$ form a cycle. We can automatically satisfy such configurations by parametrizing $c_l = \sigma_{P_1(l)} \sigma_{P_2(l)}$, where $P_{1,2}(l)$ are the two plaquettes neighbored by link l and σ_P is a fictitious Ising variable living on plaquette P . Finally, we can write down

$$p(E = C(\Sigma) \odot E^* | E^*) = \frac{\exp[-\beta \sum_l e_l^* \sigma_{P_1(l)} \sigma_{P_2(l)}]}{\mathcal{Z}(E^*)}, \quad (43)$$

for $\Sigma = (\sigma_v)_v$ and $\mathcal{Z}(E^*)$ is the partition function that normalizes the distribution. This is the Gibbs distribution for the 2D RBIM. Finally, to calculate expectation values, we do so with respect to the *true/teacher's* prior distribution $p_*(E^*) \propto \exp[\beta_* \sum_l e_l^*]$.

A natural order parameter for successful inference and therefore error correction is the MSE between the true/teacher's error E_l^* and inferred/student's error E_l at the same link averaged over realizations, $\mathbb{E}[(E_l - E_l^*)^2] = 2 - 2\mathbb{E}[E_l E_l^*]$, or equivalently, their correlations,

$$\begin{aligned} \mathbb{E}_{E, S, E^*} [E_l E_l^*] &= \mathbb{E}_{E, S, E^*} [E_l^* \sigma_{P_1(l)} \sigma_{P_2(l)} E_l^*] \\ &= \mathbb{E}_{E, S, E^*} [\sigma_{P_1(l)} \sigma_{P_2(l)}], \end{aligned} \quad (44)$$

which corresponds to ferromagnetic (FM) order for the 2D RBIM. In fact, as long as the inferred error is in the same homological class as that of the true error, the error correction is successful. Therefore, the more relevant order parameter would be the probability of inferring the correct homology class. However, it turns out the transitions for both order parameters coincide [6].

C. Quantum repetition code

In the repetition code case, we consider also read-out errors and multiple time steps. Here, the teacher prepares the initial register as $f_{l,t=0}^* = 1$; therefore, $p_*(f_0^*) = \delta_{f_0^*, 1}$. For subsequent times, f_t^* undergoes a Markov chain as

$$p_*(f_t^* | f_{t-1}^*) = \prod_l p_*(f_{l,t}^* | f_{l,t-1}^*) \propto \prod_l \exp[\beta_* f_{l,t}^* f_{l,t-1}^*],$$

and the prior for entire trajectory $F^* := f_{1:t}^*$, becomes

$$\begin{aligned} p_*(F^*) &= \left(\prod_{\tau=2}^t p_*(f_\tau^* | f_{\tau-1}^*) \right) p_*(f_1^*) \\ &\propto \prod_{l,\tau} \exp[\beta_* f_{l,\tau}^* f_{l,\tau-1}^*], \end{aligned} \quad (45)$$

where we set $f_{l,0}^* = 1$.

Then, they generate unreliable syndromes $s_{v,t}$ at vertex v by randomly flipping the true syndrome $s_{v,t}^* = f_{l_L(v),t}^* f_{l_R(v),t}^*$ as

$$p_*(s_{v,t} | f_{l_L(v),t}^*, f_{l_R(v),t}^*) \propto \exp[\beta_* s_{v,t} f_{l_L(v),t}^* f_{l_R(v),t}^*], \quad (46)$$

where $l_{L(v)}$ and $l_{R(v)}$ are left and right links of vertex v , respectively. Then, the observation model for the entire trajectory is

$$p_s(S | F^*) = \prod_{v,\tau} p_*(s_{v,\tau} | f_{l_L(v),\tau}^*, f_{l_R(v),\tau}^*). \quad (47)$$

The student could then infer a posterior distribution over the whole trajectory through Bayes' rule,

$$p_s(F | S) = \frac{p_s(S | F) p_s(F)}{p_s(S)}, \quad (48)$$

using assumed fidelity $\pi \propto e^{-\beta}$. To infer the current errors in real time, the student may use the forward equation to infer $p_s(f_t | s_{1:t})$, then take its argmax_{f_t} to determine the error correction operation.

The natural order parameter in this case would be the mean-squared error of current bit-flips $\mathbb{E}[(f_{l,t} - f_{l,t}^*)^2] = 2 - 2\mathbb{E}[f_{l,t} f_{l,t}^*]$, or equivalently the equal-bond correlation between that of the student and teacher, $\mathbb{E}[f_{l,t} f_{l,t}^*]$.

To massage the above to the 2D RBIM, we picture space-time as a 2D lattice, as shown in Fig. 2(b). Now the original links on the 1D chain are horizontal links $l_{\rightarrow} = (l, t)$ in the 2D lattice. Vertical links $l_{\uparrow} = (v, t)$ connect the vertex at current time step to that of the previous time step. Note that the lattice is periodic in the horizontal direction, but in the vertical direction, starts with horizontal links and terminates with vertical links. We denote $e_{l_{\rightarrow}}^* = -1$ as a bit-flip error occurring at time t , and $e_{l_{\uparrow}}^* = -1$ to denote if a syndrome measurement is corrupted by a read-out error. In Appendix F, we show that for $E^* = E_{\uparrow}^* + E_{\rightarrow}^*$, and similarly for E , $p(E | E^*)$ is again the 2D RBIM partition function Eq. (43) but with periodic boundary conditions in space, fixed boundary conditions $\sigma = 1$ at both $t = 0$ and $t = t_f$. Furthermore, the correlation order parameter becomes

$$\mathbb{E}[f_{l,t} f_{l,t}^*] = \mathbb{E} \left[\prod_{\tau=1}^t \sigma_{v_L(l),\tau} \sigma_{v_R(l),\tau} \right], \quad (49)$$

a many-body correlation function equal to one if all spins are aligned. Reference [6] claims that the error-correctability transition in this case also coincides with that of ferromagnetism.

D. Implication of the 2D random bond Ising model phase diagram

In the two cases, the information about the true errors E^* is “planted” in the syndromes S . Therefore, RBIMs can also be referred to as “planted Ising models” [11].

Conventionally, the phase diagram of the 2D RBIM is drawn in the $p - T$ parameter space, where $p = \pi_*$ is the probability of ferromagnetic bonds and $T = 1/\beta$ is the temperature. To interpret it in the inference setting, however, it is instructive to plot it in the student-teacher fidelity space $\beta - \beta_*$. In this case, the Nishimori line, equal to the Bayes optimal line, is a straight line $\beta_* = \beta$. The phase diagram is plotted with these variables in Fig. 3 interpolated from numerically accurate datapoints obtained from previous studies [20–22].

Though previously debated, it is now established that there exists a slight re-entrance in the phase diagram, such that the critical line between $\beta_*^c(\beta = \beta_*)$ and $\beta_*^c(\beta = \infty)$ is not horizontal [21,22]. This has implications on the thresholds of different decoders. The “decoder” discussed in quantum error correction literature is equivalent to the classifier $h(Y)$ onto the most likely error chain. As shown in Sec. III, the optimal classifier is the Bayes classifier— argmax of the true

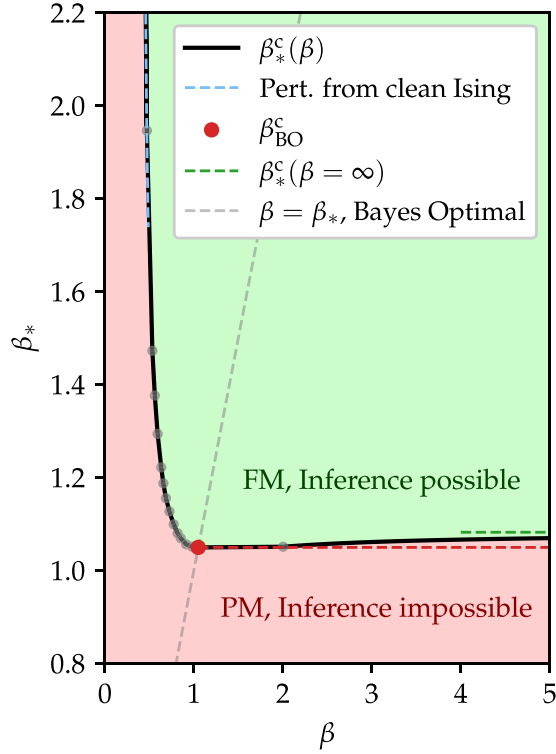


FIG. 3. Phase diagram of the 2D random bond Ising model drawn in student-teacher (assumed-true) fidelity $\beta - \beta_*$ parameter space, which also applies to static QEC problems on the toric code and dynamic QEC the quantum repetition code. Light green region denotes FM phase and the inference- or QEC-possible phase. Light red region denotes the PM phase and inference- or QEC-impossible phase. Black line: Fidelity thresholds, linearly interpolated from the datapoints in the commonly discussed $p - T$ parameters. Gray dashed line: Bayes optimal line $\beta = \beta_*$. Light blue dashed line: exact perturbation from clean 2D Ising model $\beta_* = \infty$, where $\beta^c(\beta_* = \infty) \approx 0.4407$ [19]. Gray markers: fidelity thresholds obtained in Ref. [20]. Red circle and dashed line: fidelity threshold on the Bayes optimal line $\beta_{\text{BO}}^c = 1.048 \pm 0.001$, obtained in Ref. [21]. Green dashed line: fidelity threshold at infinite student fidelity $\beta_*^c(\beta = \infty) = 1.081 \pm 0.001$, obtained in Ref. [22], which implies a slight re-entrance in the phase diagram. Note that the y axis starts at $\beta_* = 0.8$ and that the aspect ratio is not equal between β and β_* , to accentuate the re-entrance.

posterior $h_{\text{optimal}}(S) = \arg\max_E p_*(E|S)$. In a more realistic setting, however, the evaluation of the posterior is too time-consuming. Therefore, the minimum-weight perfect matching (MWPM) decoder is used [23]. In the toric code setting, This is when error chains with the shortest length between the syndromes is assumed to correct the error. It turns out that this is equivalent to the $\arg\max$ of the posterior at $\beta \rightarrow \infty$ [6]. Therefore, this suggests that the threshold for the MWPM decoder is slightly lower in p (or higher in β_*) than the Bayes optimal setting, though not by much.

One order parameter not yet discussed for the 2D RBIM is the observable variance [Eq. (11)] of the error chain in the student's distribution, i.e., δE_i^2 and $\delta f_{i,t}^2$ for the toric and repetition code case, respectively. Although the error chain variance is equal to MSEM and MSE/2 on the Bayes optimal

line, it is not necessarily related off of it. In the FM/inference possible phase, it is clear that the student's variance is small. However, *a priori*, we cannot rule out that it is small in the inference impossible phase. Writing out the variance for both cases results in quantities that only depend on Edwards-Anderson-like (EA) order parameters, $1 - \mathbb{E}[(\sigma_{P_1(l)} \sigma_{P_2(l)})^2]$ and $1 - \mathbb{E}[(\prod_{\tau=1}^t \sigma_{v_L(l),\tau} \sigma_{v_R(l),\tau})^2]$ for the toric and repetition code, respectively (The EA order parameter is $\mathbb{E}[(\sigma_l)^2]$). In the disordered systems literature, the spin glass phase (SG) is characterized by no FM order and presence of EA order [24]. In the 2D RBIM, the SG phase does not exist for any $\beta < \infty$ [25]. Therefore, the “error-chain fuzzy and sharp” phases, corresponding to the uncertainty in the posterior distribution, which in turn correspond to phases where the Edwards-Anderson order parameter does not vanish and vanish, respectively, coincide with the inference impossible and possible phases, respectively. Here, error-chain fuzzy or sharp phases are defined as regions in student-teacher parameter space where the uncertainty of the student's posterior on the error-chains is nonzero or vanishes, respectively.

VIII. THE PLANTED SSEP AND THE HAAR-RANDOM U(1)-SYMMETRIC MONITORED QUANTUM CIRCUIT

In this section, we discuss a weak-measurement variety of the recently studied Haar-random U(1)-symmetric monitored quantum circuit [1]. This is a quantum inference problem on a qHMM, which using the arguments of Sec. VI can be mapped to a classical inference problem on an HMM that we call the *planted SSEP*, where the state of an SSEP is inferred from its noisy “images.” In turn, the original quantum model can be referred to as the *quantum planted SSEP*. If the student is disregarded, the model can be thought of as a standalone circuit operated by an experimentalist/teacher. We introduce the quantum model in Sec. VIII A and the classical counterpart in Sec. VIII B. In Sec. VIII C, we study the planted SSEP. After discussing the general considerations, we develop discuss its replica field theory away from Bayes optimality (Sec. VIII C 1), whose predictions are verified numerically (Sec. VIII C 2) to obtain phase boundary in Fig. 4 and its universality class.

A. Quantum circuit model

Consider a local Hilbert space of qubits and qudits, $\mathcal{H}_{\text{loc}} = \mathbb{C}^2 \otimes \mathbb{C}^d$, with qubit computational basis states $|s\rangle \in \{|-1\rangle, |1\rangle\}$ and qudit basis states $|g\rangle \in \{|0\rangle, \dots, |d-1\rangle\}$. Now, we consider a 1D chain of such local degrees of freedom living at sites x with even system size L and periodic boundary conditions. The teacher (or the experimentalist) prepares an initial state ρ_0 , then runs the quantum circuit which we will describe below.

We study two scenarios for the initial states. The first scenario is where the teacher prepares a pure state $\rho_0 = |\psi_0\rangle\langle\psi_0|$, applies the circuit, then measures the total charge of the system afterwards. The final measured charge is then the hidden register. For simplicity, we take the initial state to be an equal superposition of all qubit and qudit basis states, which can be prepared as a product state $|\psi_0\rangle = |+\rangle^{\otimes L} \otimes |\chi\rangle^{\otimes L} = (2d)^{-L/2} \sum_{s,g} |s\rangle \otimes |g\rangle$. Again, we used the

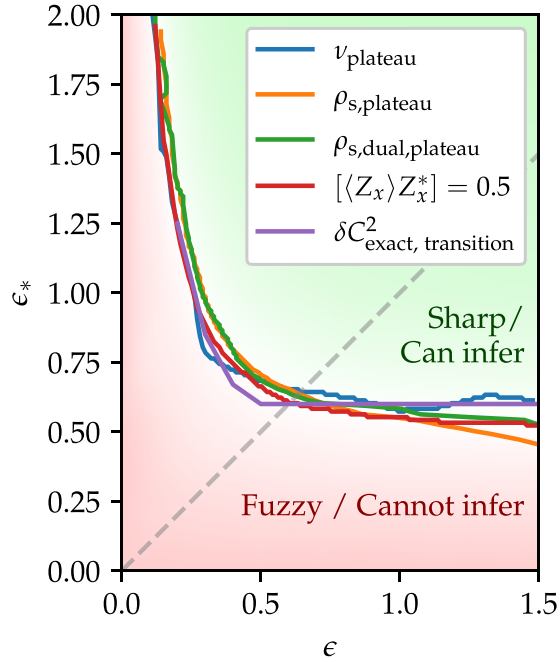


FIG. 4. Phase diagram of the planted SSEP deduced from various methods. From TEBD, we obtain blue, orange, and green lines which delineate where extracted exponent ν and extracted superfluid stiffnesses ρ_s , $\rho_{s,\text{dual}}$ approach plateau values, respectively. Red line is contour for $\mathbb{E}[(s_x)s_x^*] = 0.5$ from TEBD. Purple line is critical point determined by point where $\delta C^2(t)$ turns from a power law in time to exponential decay from exact calculation of the posterior. The 45° gray dashed line in parameter space corresponds to Bayes optimality $\epsilon = \epsilon_*$. Shaded colors are given by the alignment $\mathbb{E}[(s_x)s_x^*]$. The good agreement with phenomenology predicted by replica field theory suggests that the whole phase boundary is modified KT.

bold symbol notation to denote configurations over all space, $|+\rangle := 2^{-1/2}(|-1\rangle + |1\rangle)$ is the Pauli- x eigenstate, and $|\chi\rangle := d^{-1/2} \sum_g |g\rangle$ is the equal superposition of qudit computational basis states.

The second scenario is where the teacher randomly picks from an ensemble of states $\{|\psi_{C^*}\rangle\}$, each with a definite total charge C^* , where charge in the system is defined as $\hat{C} = \sum_x Z_x$. The hidden register then becomes the randomly picked charge. Therefore, the initial ensemble can be written as $\rho_0 = \sum_{C^*} p_*(C^*) |\psi_{C^*}\rangle \langle \psi_{C^*}| \otimes |C^*\rangle \langle C^*|$. In order that $\text{tr}[\mathbb{P}_s \rho_0]$ is the same as the previous scenario, we choose the charge to be binomially distributed $p_*(C^*) = \frac{1}{2^L} \binom{L}{(C+L)/2}$ and $|\psi_{C^*}\rangle \propto (\sum_{s: C[s]=C^*} |s\rangle) \otimes |\chi\rangle^{\otimes L}$.

Although both of these scenarios can be used in the inference and standalone settings, the first scenario is the most natural when the student is disregarded, and the second is the most natural in the teacher-student inference setting.

After preparation of the state, the teacher runs the quantum circuit for time steps in the order of system size $t_f \sim L$. At each time step t , the teacher applies unitary gates on neighboring qubit-qudit pairs in a brickwork fashion,

$$u_t = \bigotimes_{x \in \mathcal{R}_t} u_{x:x+1,t}, \quad (50)$$

where $\mathcal{R}_t$ are the even sites at even times and odd sites at odd times and each $u_{x:x+1,t}$ is a Haar-random two-qubit $u_{U(1)}$ as Eq. (26). We will denote all instances of unitaries as $U := u_{1:t_f}$. This layer of unitary gates is followed by weak measurements on each qubit at every site $Q(y_{x,t})$ as Eq. (28). We will use the notation $y_t = (y_{x,t})_{x \in 1:L}$ and $Y = y_{1:t_f}$ to denote the collection of all measurements gained at time step t and at all time steps, respectively.

In the inference setting, the teacher would then send the student the measurements Y and possibly their sampled Haar unitaries U . The student then infers the total charge in the system C^* with their model which has the same functional form as the teacher's but with the student's assumed parameters ϵ, σ .

Thought of as a standalone circuit, the quantity of interest would be the charge variance δC_*^2 , i.e., Eq. (20) for observable $O = \hat{C}$.

At Bayes optimality, $\delta C_*^2 = \delta C_s^2$. This immediately implies the charge sharpening transition, which is the most naturally formulated in the standalone circuit coincides with the learnability transition in the inference setting here.

B. Classical inference model

Using the arguments of Sec. VI, in cases where the choice of Haar unitaries is hidden or qudits have large Hilbert space dimension $d \gg e^{L t_f}$, the quantum model can be mapped to a classical inference problem on an HMM which we call the *planted SSEP*. Here, the teacher generates trajectories of bitstrings S^* via the SSEP [18]. Then, they generate noisy “images” of the configurations, $Y \sim p_*(\cdot | S^*)$ (which could be done at each time step), which they hand to the student. The student then uses their own model to infer a posterior given the measurements $p_s(S|Y)$.

Concretely, we consider the teacher drawing a brickwork SSEP [26] on a 1D lattice of size L with periodic boundary conditions. At $t = 1$ the teacher draws initial Ising (equivalently Boolean $b = (s + 1)/2$) variables $s_1^* = (s_{1,1}^*, \dots, s_{L,1}^*)$ from a separable distribution $p_*(s_1^*) = \prod_x p_*(s_{x,1}^*)$ (the initial state considered in the quantum model maps to an initial distribution of equally likely charge configurations, $p_*(s_1) = 2^{-L}$). For subsequent times, the teacher state evolves as a Markov process

$$p_*(s_t^* | s_{t-1}^*) = \prod_{x \in \mathcal{R}_t} p_*(s_{x:x+1,t}^* | s_{x:x+1,t-1}^*). \quad (51)$$

Denoting the incoming and outgoing states as $(s_{x,t-1}, s_{x+1,t-1})$, $(s_{x,t}, s_{x+1,t})$, respectively, the two-site kernel is given by the SSEP kernel [Eq. (32)]. The probability distribution for the whole trajectory is

$$p_*(S^*) = \left(\prod_{t=2}^{t_f} p_*(s_t^* | s_{t-1}^*) \right) p_*(s_1^*). \quad (52)$$

The “noisy image” generated at each time step, y_t , is given by the pixel values

$$y_{x,t} = \epsilon_* s_{x,t} + \sigma_* \eta_{x,t}, \quad (53)$$

where ϵ_* is the true/teacher's signal strength and σ_* is the true/teacher's noise strength, and $\eta_{x,t} \sim \mathcal{N}(0, 1)$. Therefore,

$$p_*(y_{x,t}|s_{x,t}^*) = \frac{1}{\sqrt{2\pi\sigma_*^2}} \exp\left(-\frac{(y_{x,t} - \epsilon_* s_{x,t}^*)^2}{2\sigma_*^2}\right), \quad (54)$$

and for set of all images, is $p_*(Y|S^*) = \prod_t p_*(y_t|s_t^*) = \prod_{x,t} p_*(y_{x,t}|s_{x,t}^*)$. Again, we will consider the regime where the total number of time steps is on the order of magnitude as the system size $t_f \sim L$.

Given Y , the student infers a posterior distribution $p_s(S|Y)$ using their own model with parameters ϵ and σ , with all other functional forms equal.

The joint distribution is then given by the planted ensemble [Eq. (5)] for $X, X^* \rightarrow S, S^*$. In the Bayes optimal case, where $* = s$, it is symmetric under $S \leftrightarrow S^*$. Since S^* evolves as the SSEP, this means that S must also evolve as the SSEP in this case. Away from this point, $p(S) \neq p_s(S)$ generally.

As in the quantum problem, we define total charge as $C[s_t] := \sum_x s_{x,t}$. Then, a natural quantity to study in the Bayesian inference setting is the MSE and MSEM of the total charge C . As we are interested in observables at the final time, we can marginalize over configurations at previous time and look at the posterior distribution for the filtering task, $p_s(s_t|y_{1:t})$, which will evolve according to the forward equation (4).

This model is natural generalization of the recently studied planted directed polymer [8]. In that case the trajectory of one particle was inferred, equivalent to restricting the prior to the one-charge sector in the planted SSEP.

C. Analysis of the planted SSEP

For simplicity, we consider the case when all initial charge configuration is equally likely, $p_s(s_1) = p_*(s_1^*) = 2^{-L}$. This is half-filling on average, with the total charge distributed binomially. Unless specified, we set $\sigma = \sigma_* = 1$ and consider the behavior of the planted SSEP across the $\epsilon - \epsilon_*$ plane.

Along the $\epsilon = 0$ line, no information is assimilated by the student's posterior. Therefore, the posterior is decoupled from the measurements and evolves as the SSEP $p_s(S|Y) = p_{\text{SSEP}}(S)$. Our uniform initial state is the stationary state of the SSEP, which has no second-order connected correlations, and therefore $\langle s_{x,t} s_{x',t} \rangle_c = 0$. Here, the angled brackets denote expectation on the posterior. At average half filling we also have $\langle s_{x,t} s_{x',t} \rangle = \langle s_{x,t} \rangle = 0$.

On the Bayes optimal line $\epsilon_* = \epsilon$, due to the symmetry between the teacher and student in the planted ensemble, the marginalized distribution of S follows that of the SSEP, $p(S) = p_{\text{SSEP}}(S)$. Therefore, observables linear with the student's posterior, such as the averaged disconnected correlator, is equal to that of the SSEP, i.e., zero due to the choice of average half-filling $\mathbb{E}_{Y,S^*}[\langle s_{x,t} s_{x',t} \rangle] = 0$. However, higher moments, such as the variance of the total charge, can still deviate from those of the SSEP.

Remaining on the Bayes optimal line, existing numerics for the projective measurement case and analytic replica field theory for weak measurements suggest that there is a phase transition between a charge "fuzzy" and "sharp" phase, exhibited in the higher moment observables in the posterior

(or unravelled density matrix of each trajectory) [1,3]. The known phenomenology is as follows: the variance of the total charge of the student's posterior, averaged over Y, S^* , is finite for $t < t_\#$ and is exponentially decaying for $t > t_\#$ where $t_\#$ is the sharpening time. In the charge fuzzy phase, $t_\# \sim L$, whereas it is sublinear in the sharp phase. Our Eq. (12) immediately implies that, as the order parameter of inference exactly equals the order parameter for sharpening, there is also a corresponding phase transition between a hard inference phase, where the time required for the mean squared error of the total charge to exponentially decay grows as $t_{\text{inference}} \sim L$, and an easy phase, where it is sublinear. In the replica field theory, the phase transition is predicted to be a modified-KT transition [3]. Combined, existing works suggest a transition around projective measurement rate $\pi_{\text{BO}}^\# \in [0.2, 0.3]$ [1,3]. From arguments of inferrability on a single site (Appendix D), we expect the transition for weak measurements to be around $\epsilon_{\text{BO}}^\# \in [0.49, 0.63]$.

We now go beyond Refs. [1–3] by moving off the Bayes optimal line. Here, $p(S)$ deviates from that of the SSEP. Treating ϵ as a small variable and $r_\epsilon := \epsilon_*/\epsilon$ as a factor of order 1 and using the identities of the SSEP, we can expand various correlators in orders of ϵ (Appendix B). To smallest nonzero order, we find that the leading order terms scale with ϵ, ϵ^* as

$$\mathbb{E}_{Y,S^*}[\langle s_{x,t} s_{0,t} \rangle_c] \propto -\epsilon^2, \quad (55)$$

$$\mathbb{E}_{Y,S^*}[\langle s_{x,t} s_{0,t} \rangle] \propto \epsilon^4 (r_\epsilon^2 - 1), \quad (56)$$

$$\mathbb{E}_{Y,S^*}[\langle s_{x,t} \rangle s_{0,t}^*] \propto \epsilon \epsilon_*. \quad (57)$$

Along the Bayes optimal line, the averaged disconnected correlation function vanishes as required by the choice of initial condition and that $p(S) = p_{\text{SSEP}}(S)$. We also note that its sign changes as the line is crossed. As expected, the correlation between the student's inferred and teacher's true states is only nonzero and positive only if both $\epsilon, \epsilon_* > 0$.

1. Replica field theory

In this subsection, we outline the derivation of the replica field theory to identify the universality class of the phase transition away from Bayes optimality. We will compute various correlators, and combine these with numerical results, in order to determine a phase diagram for the quantum inference problem. See Appendix A for details.

Observables, both linear and nonlinear in the student's posterior (or student's density matrix in the quantum language), which could also be linear or independent on the teacher distribution (teacher's density matrix), can be written as

$$O = \frac{1}{\mathcal{Z}} \sum_{S^{1:n}, S^*} \int dY O(S^{1:n}, S^*) \times \frac{q_s(S^1, Y) \cdots q_s(S^n, Y)}{q_s(Y)^n} q_*(S^*, Y), \quad (58)$$

where $q_s(Y) = \sum_S q_s(S, Y)$. Here, we replaced the probability distribution with unnormalized ones at the cost of introducing the partition function $\mathcal{Z} = \sum_{S^{1:n}, S^*} \int dY \frac{q_s(S^1, Y) \cdots q_s(S^n, Y)}{q_s(Y)^n} q_*(S^*, Y)$. As examples, $O(S^{1:n}, S^*)$ could be

(1) $\delta C_s^2(s_i^{1:2}) = C^2[s_i^1] - C[s_i^1]C[s_i^2]$ for the student's charge variance,

(2) $\text{MSE}(s_i^1, s_i^*) = (C[s_i^1] - C[s_i^*])^2$ for the MSE of the charge, or

(3) $\text{MSEM}(s_i^{1:2}, s_i^*) = C[s_i^1]C[s_i^2] - 2C[s_i^1]C[s_i^*] + C^2[s_i^*]$ for the MSEM of the charge.

Note that we are free to write $q_s(Y)^{-n} = \lim_{k \rightarrow -1} (\sum_S q_s(S, Y))^{nk}$. Defining $Q = n + nk + 1$, we will adopt the replica trick and switch the limit and integrals, then assume that nk is an integer, resulting in the following:

$$O = \lim_{Q \rightarrow 1} \frac{1}{\mathcal{Z}_Q} \sum_{S^{1:Q}} O(S^{1:n}, S^Q) \times \int dY q_s(S^1, Y) \cdots q_s(S^{Q-1}, Y) q_*(S^Q, Y). \quad (59)$$

Throughout, we will refer to the first $Q - 1$ replicas as the “student replicas,” and the Q th replica as the “teacher replica.” Taking the rate of exchange between neighboring particles in the SSEP kernel as a tunable parameter J , then going to continuous time $\tilde{t} = \mathbf{b}t$, where $\mathbf{b}$ is the temporal lattice constant, and finally removing the tildes, one can find that

$$\mathcal{Z}_Q = \sum_{S^{1:Q}} \int dY q_s(S^1, Y) \cdots q_s(S^{Q-1}, Y) q_*(S^Q, Y), \quad (60)$$

$$\rightarrow \langle + |^{\otimes QL} \exp \left(- \int_0^{t_f} dt H_Q \right) | + \rangle^{\otimes QL}, \quad (61)$$

where $|+\rangle$ is the Pauli- x eigenstate, with replica Hamiltonian

$$H_Q = -J \sum_{A=X,Y,Z} \sum_{a,x} \hat{A}_x^a \hat{A}_{x+1}^a + \sum_{a,b,x} \hat{Z}_x^a M_{ab} \hat{Z}_x^b, \quad (62)$$

interreplica interaction matrix

$$M = \text{diag}[\vec{\epsilon}^2] - \frac{1}{Q} \vec{\epsilon} \vec{\epsilon}^T, \quad (63)$$

and $\vec{\epsilon}^m = (\epsilon^m \cdots \epsilon^m \epsilon_*^m)$ is a Q -dimensional vector in replica space. Throughout this subsection, we will adopt the notation that arrowed vectors refer to vectors in replica space.

We now adopt the spin-coherent path integral [27] and take the continuum limit. Introducing compact fields $(\theta_x^a(t), \phi_x^a(t))_{x=1:L}^{a=1:Q}$ with correspondence $(\hat{X}, \hat{Y}, \hat{Z}) \rightarrow (\cos \phi \sin \theta, \cos \phi \sin \theta, \cos \theta)$ turns the partition function into a path integral $\mathcal{Z} = \int D[\theta_x^a, \phi_x^a] e^{-\mathcal{S}}$ with action $\mathcal{S} = \int_t (-\frac{i}{2} \sum_{a,x} \cos(\theta_x^a) (\partial_t \phi_x^a) + H_Q(\theta, \phi))$. As we are working at half-filling, we expand $\theta = \theta_0 + \delta\theta$ at $\theta_0 = \pi/2$ and treat $\delta\theta \approx -\hat{Z}$ as a Gaussian variable, neglecting its compactness. Introducing continuum position $\tilde{x} = \mathbf{a}x$, where $\mathbf{a}$ is the spatial lattice constant, expanding the action to smallest relevant orders in $\mathbf{a}$, and removing the tildes, results in the partition function $\mathcal{Z} = \int D[\theta^a, \phi^a] e^{-\int_{t,x} \mathcal{L}}$ with Lagrangian density

$$\mathcal{L} = \frac{i}{2\mathbf{a}} \sum_{a=1}^Q \delta\theta^a \partial_t \phi^a + \frac{1}{2\mathbf{a}} \sum_{a,b=1}^Q \delta\theta^a M_{ab} \delta\theta^b + \frac{J\mathbf{a}}{2} \sum_{a=1}^Q [(\partial_x \delta\theta^a)^2 + (\partial_x \phi^a)^2], \quad (64)$$

The interreplica interaction matrix can be diagonalized as $M = V \Lambda V^T$. There is one zero mode $\lambda_0 = 0$ with $\vec{v}_0 \propto \epsilon_*^{-1}$ and one mode with eigenvalue $\lambda_K = (\epsilon/\sigma)^2 \frac{1+Kr_*^2}{1+Kr_*^2}$ and eigenvector $\vec{v}_K \propto (\epsilon \cdots \epsilon - K\epsilon_*)$ where $K := Q - 1$ is the number of student replicas and $r_* := \sigma_*/\sigma$. The rest of the modes are massive with eigenvalues $\lambda_{1 \leq \alpha \leq Q-2} = (\epsilon/\sigma)^2$. Integrating out the massive modes of $\delta\theta_\alpha$, doing a spacetime transformation to make the action isotropic $\tilde{x} = x/\sqrt{v}$ and $\tilde{t} = \sqrt{v}t$ with $v = 2\mathbf{a}(\epsilon/\sigma)\sqrt{J}$, removing the tildes, applying a Hubbard-Stratonovich transformation, separating the azimuthal angle field into smooth and vortex parts $\phi^a = \phi_s^a + \phi_v^a$ then integrating out the smooth part, we retrieve the dual Lagrangian with degrees of freedom χ^a , related to ϕ as $\rho_s \partial_\mu \phi^a \leftrightarrow \frac{\epsilon^{\mu\nu}}{2\pi} \partial_\nu \chi^a$ (just in this case, $\epsilon^{\mu\nu}$ is the Levi-Civita symbol for $\mu, \nu \in \{t, x\}$), and vortex density $\rho_v^a = \sum_\mu \frac{(-1)^\mu}{2\pi} (\partial_\mu \partial_{\tilde{\mu}} \phi_v^a)$,

$$\mathcal{L}_d = -i \sum_{a=1}^Q \chi^a \rho_v^a + \mathcal{L}_G, \quad (65)$$

where the Gaussian part is

$$\mathcal{L}_G = \frac{1}{8\pi^2 \rho_s} [(\partial_t \chi_0)^2 + D^2 (\partial_x^2 \chi_0)^2] + \frac{1}{8\pi^2 \rho_s} \sum_{\mu=t,x} \sum_{a,b=1}^Q (\partial_\mu \chi^a) \Phi_{ab} (\partial_\mu \chi^b) + \frac{1}{8\pi^2 \rho_s} \left[(\partial_t \chi_K)^2 + \frac{1+r_*^2 K}{1+r_*^2 K} (\partial_x \chi_K)^2 \right], \quad (66)$$

$\Phi = \mathbb{1}_Q - \vec{v}_0 \vec{v}_0^T - \vec{v}_K \vec{v}_K^T$ is a projector out of the zeroth and K th modes, the noncompact dual fields related to the original variables as $\hat{Z} \approx -(\mathbf{a}/\pi) \partial_x \chi$, the vortex densities are $\rho_v^a = \sum_{i=1}^{N_a} m_i^a \delta(\mathbf{r} - \mathbf{r}_i^a)$ with vorticities m_i^a , space-time coordinate $\mathbf{r} = (t, x)$, “superfluid stiffness” $\rho_s \sim \sqrt{J\sigma^2/\epsilon^2}$ and diffusion coefficient $D^2 \sim \sqrt{J\epsilon^2/\sigma^2}$.

We would like to identify the most relevant term after integrating out ρ_v^a . At Bayes optimality, it was shown in Ref. [3] that the minimal vortex configurations are interreplica vortex-antivortex pairs between all replicas, leading to the minimal Lagrangian $\mathcal{L}_{\text{BO}} = \mathcal{L}_G + \lambda \sum_{a>b} \cos(\chi_a - \chi_b)$. From renormalization analysis, it can be shown that the minimal interaction becomes relevant at $\rho_s^\# = \pi^{-1}$, corresponding to a modified Kosterlitz-Thouless (KT) transition [28].

Away from Bayes optimality, we use a “Landau approach” to determine the minimal interaction term. Writing down the most general term satisfying the symmetries then identifying the most relevant minimal vortex term [28], we find that away from Bayes optimality, the teacher vorticity is forced to be zero, as adding terms that correspond to introducing teacher vorticity are irrelevant away from commensurate points, but also always less relevant than the term that follows. The minimal vortex configurations are interreplica vortex-antivortex pairs involving *only* the student replicas. This leads to the minimal Lagrangian $\mathcal{L}_{\text{BnO}} = \mathcal{L}_G + \lambda \sum_{a>b: a,b \leq K} \cos(\chi_a - \chi_b)$. However, as vortex anti-vortex pairs are still the minimal configurations, we still predict a transition at the identical $\rho_s^\# = \pi^{-1}$.

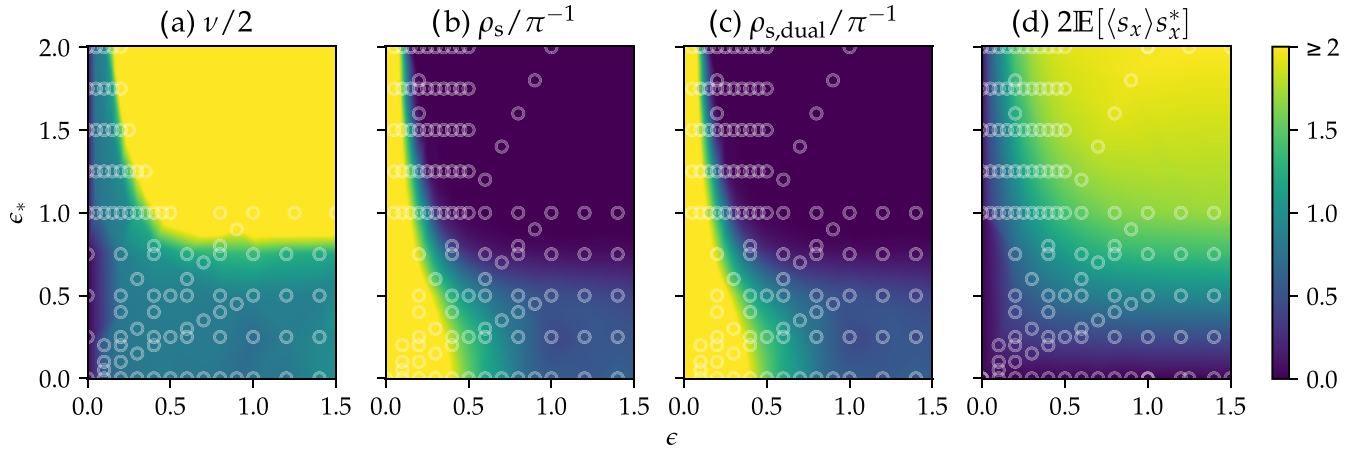


FIG. 5. Heatmap of observables computed via TEBD, for $L = 60$, $t_f = 4L$, and $\#_{\text{samples}} \geq 41$. (a) Decay of connected correlator, $\mathbb{E}[\langle s_x s_0 \rangle_c] \sim |x|^{-\nu}$, expected to be $\nu/2 \approx 1$ in the fuzzy phase. Note that vanishing ν as the $\epsilon = 0$ axis is approached reflects vanishing correlations in the SSEP. (b) Superfluid stiffness extracted from local charge variance $\delta C_{[0,x]}^2 \sim (8\rho_s/\pi) \ln |x|$, (c) from string operator $W_{[0,x]}^2 \sim |x|^{-2\pi\rho_{s,\text{dual}}}$, expected $\rho_s^\#/\pi^{-1} = 1$ at the transition. (d) Student-teacher alignment $\mathbb{E}[\langle s_x \rangle s_x^*]$, expected to be ~ 1 in the easy inference phase and ~ 0 in the hard phase. White circles indicate points in parameter space obtained, while the rest of the plot is interpolated. To make the features visible, we cap the dynamic range of the heatmap at 2.

Inside the charge-fuzzy phase, where the vortex terms are irrelevant, we can use the Gaussian theory to predict the scaling of correlators. Away from the Bayes optimal point, the connected correlator can be nonzero. Using the aforementioned operator relations we find

$$\mathbb{E}_{Y,S^*} [\langle s_x s_0 \rangle_c] = -4\pi\rho_s \left(\frac{a}{x}\right)^2, \quad (67)$$

$$\mathbb{E}_{Y,S^*} [\langle s_x s_0 \rangle] = 4\pi\rho_s(r_\epsilon^2 - 1) \left(\frac{a}{x}\right)^2, \quad (68)$$

$$\mathbb{E}_{Y,S^*} [\langle s_x \rangle s_0^*] = 4\pi\rho_s r_\epsilon \left(\frac{a}{x}\right)^2. \quad (69)$$

Although the exact scaling does not agree with the lowest-order expansion [Eqs. (55)–(57)], the sign of the quantities do. In particular, both methods agree that the sign of the disconnected correlator switches sign at the Bayes optimal point $r_\epsilon = 1$.

In the charge-sharp phase, connected correlators decay exponentially, $\mathbb{E}_{Y,S^*}[\langle s_x s_0 \rangle_c] \sim -e^{x/x_0}$. As for the disconnected correlator, in the crudest approximation, since we expect the student to closely follow the teacher, we expect $\mathbb{E}_{Y,S}[\langle s_x s_0 \rangle] \approx \mathbb{E}_{Y,S}[\langle s_x^* s_0^* \rangle]$. Since the teacher configuration evolves as the SSEP and due to the choice of the prior at initial time, we expect $\mathbb{E}_{Y,S}[\langle s_x s_0 \rangle] \approx 0$ and have no sign on average.

The above field theory analysis is only valid for $\epsilon > 0$. At the exceptional line $\epsilon = 0$, all three of the correlators should vanish for $x \neq 0$.

2. Numerical results

Time-evolving block decimation (TEBD). To test the long-distance predictions of the replica field theory, we must study larger system sizes. We therefore simulate the posterior using TEBD [29] for $L = 60$ and $t_f = 4L$ using the iTensor library [30].

We fit a power-law exponent ν to the decay of the averaged connected correlator, $\mathbb{E}_{Y,S^*}[\langle s_x s_0 \rangle_c] \sim |x|^{-\nu}$. We expect ν to

plateau to a value ≈ 2 inside of the charge fuzzy phase with $\epsilon > 0$. In the charge sharp phase, we expect an exponential decay. Therefore, a fit to power-law decay will give $\nu > 2$. In Fig. 5(a), we plot a heatmap of $\nu/2$ in the $\epsilon - \epsilon_*$ plane. We can clearly see two regions where ν plateaus to value $\nu_{\text{plateau}} \approx 0.9 \times 2$, indicating a charge-fuzzy phase, and one where $\nu \gg 2$, indicating a charge-sharp phase.

We can extract superfluid stiffness ρ_s from two different observables. The first is via the local charge variance, which, in the charge-fuzzy phase, can be predicted to be

$$\delta C_{[0,x]}^2 = \mathbb{E}_{Y,S^*} \sum_{x',x'' \in [0,x]} [\langle s_{x'} s_{x''} \rangle_c] \quad (70)$$

$$\approx \frac{8\rho_s}{\pi} \ln |x| + \dots \quad (71)$$

This is shown in Fig. 5(b). The second is via the string order parameter

$$W_{[0,x]}^2 = \mathbb{E}_{Y,S^*} [\langle s_0 s_1 \dots s_x \rangle^2], \quad (72)$$

$$\propto |x|^{-2\pi\rho_{s,\text{dual}}}, \quad (73)$$

shown in Fig. 5(c). In the charge sharp phase, both quantities are exponentially decaying, so we expect both of the extracted stiffnesses to be small, $\rho_s \ll 1$. We see good agreement between the two methods of extraction. At Bayes optimality, we see good evidence of phase transition at $\epsilon_{\text{BO}}^\# \approx 0.6$, consistent with the expected location when mapping from projective to weak measurements as discussed in the beginning of the section. At fixed $\epsilon_* < \epsilon_{\text{BO}}^\# \approx 0.6$, as ϵ is increased, we see that ρ_s plateaus instead of decaying to a value $\rho_s \ll 1$ as is the case for $\epsilon_* > \epsilon_{\text{BO}}^\#$. Presumably due to finite size effects, the plateau value is smaller than critical value predicted by the field theory $\rho_{s,\text{plateau}}(\epsilon_* < \epsilon_{\text{BO}}^\#) \approx 0.25\rho_s^\#$, but is much larger than $\rho_s(\epsilon_* > \epsilon_{\text{BO}}^\#) \approx 10^{-3}\rho_s^\#$.

In the Bayes optimal case, due to Eq. (12), the transition in the charge variance also corresponds to the transition in the mean squared error of the charge. We expect the phase

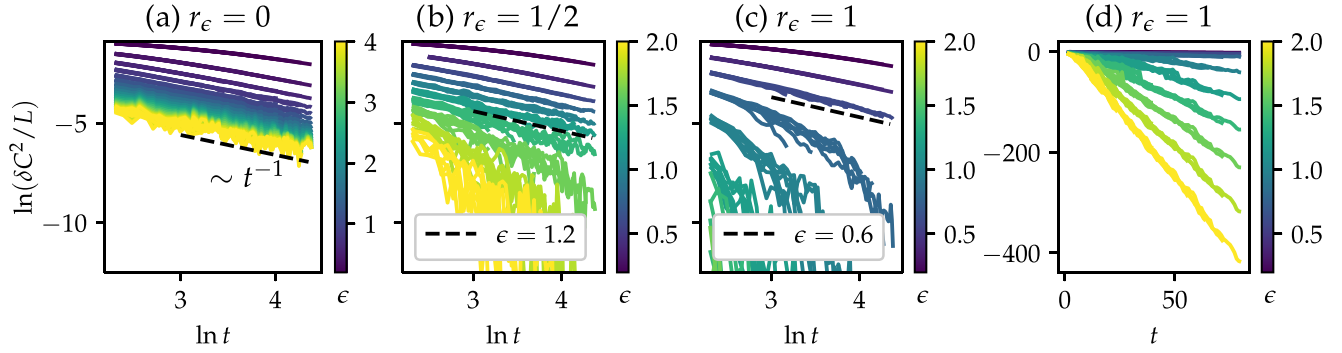


FIG. 6. Time evolution of normalized charge variance $\delta C^2/L$ as a function of time via exact calculation of the posterior, for $6 \leq L \leq 22$, $\#_{\text{samples}} = 100$. Data from all system sizes L are overlaid on top of each other, showing good collapse. Log-log plot for (a) $r_\epsilon := \epsilon_*/\epsilon = 0$, (b) $r_\epsilon = 1/2$, and (c) Bayes optimal case $r_\epsilon = 1$. For $r_\epsilon = 0$, we see data consistent with a power law $\sim t^{-1}$ even for high $\epsilon = 4$. For $r_\epsilon = 1/2, 1$, the scaling switches from $\sim t^{-1}$ to exponential decay e^{-t} at critical $\epsilon^\#(r_\epsilon)$, as shown by semilog plot in panel (d).

diagram of the easy and hard regimes of inference to continue away from the Bayes optimal point. As a measure of the quality of the inference, we look at the alignment between the teacher and the student, $\mathbb{E}_{Y,S^*}[\langle s_x \rangle s_x^*]$ in Fig. 5(d). We see that it follows the general profile of the other three observables. This indicates that the charge-fuzzy and sharp regimes also corresponds to the inference impossible and possible regimes, respectively.

To determine the phase boundary from each of the computed observables, in Fig. 4, we overlay the boundaries extracted from the four observables. For ν , ρ_s , and $\rho_{s,\text{dual}}$, we plot the points when the observables reach the plateau value. For $\mathbb{E}[\langle s_x \rangle s_x^*]$, we plot the contour when it is equal to 0.5. This is for the following reason. If s_x is guessed randomly, the order parameter is 0. On the other hand, if the inference is perfect, it is 1. 0.5 is a simple choice of midpoint between these two limits. In principle, assuming that there's a phase transition, any other point between these two limits should work. We see that the phase boundary suggested by the four observables has good agreement. First, along the Bayes optimal line, we see a phase transition at around $\epsilon^\#(r_\epsilon = 1) = \epsilon_{\text{BO}}^\# \approx 0.6$. Looking at fixed ϵ_* and varying ϵ , we see that there appears to be a threshold again at around $\epsilon_* = \epsilon_{\text{BO}}^\# \approx 0.6$.

Exact calculation of the posterior. The remaining question is whether the phase boundary hits the $\epsilon_* = 0$ line as ϵ is further increased. The replica field theory cannot give predictions on whether the phase transition occurs at finite ϵ , since we do not have an explicit map to ρ_s . From TEBD, we see no evidence it for $\epsilon < 1.5$. However, higher values of ϵ are unachievable at this system size using TEBD. This can be understood in the “forced” measurement limit ($r_\epsilon = 0$) [31,32], where the measurements are iid Gaussian noise. When ϵ is large, consider receiving the first measurement y_1 is compatible with total charge $C_1 = \sum_r \text{sgn}(y_{r,1})$. Consider when ϵ is large enough such that a TEBD algorithm with a finite cutoff for truncation (or even exact calculation of the posterior with finite precision) will discard other charge sectors not equal to C_1 . In the next time step however, we may receive y_2 with $C_2 \neq C_1$ that has higher likelihood than the first image, $\exp(\epsilon \sum_r y_{r,2}) \gg \exp(\epsilon \sum_r y_{r,1})$. However, due to the truncation at the first time step, the weight in charge sector C_2 is zero and stays zero even after applying the likelihood

with large weight. The C_1 sector on the other hand may be killed as the likelihood has very small weight. Normalization then leads to numerical instability due to division by zero.

We therefore resort to exact calculation of the prior for small system sizes $6 \leq L \leq 22$ using quadruple-precision floating numbers, capable of expressing floating points of the range $[10^{-4932}, 10^{4932}]$ compared to normal double-precision floats with range $[10^{-308}, 10^{308}]$. This allows us to reach higher values of ϵ without suffering from numerical instabilities.

We first study the charge variance. In Fig. 6, we plot the dynamics of the normalized charge variance $\delta C^2/L$ up to $t_f = 4L$. We overlay data for all system sizes, which shows good collapse, indicating that $\delta C^2(t \sim L) \sim L$. We see that for $r_\epsilon = 1/2, 1$, in the charge fuzzy phase, the decay of the charge variance is consistent with $\delta C^2 \propto Lt^{-1}$, while in the charge-sharp phase $\epsilon > \epsilon^\#(r_\epsilon)$, the dynamics switches to an exponential decay $\delta C^2 \propto Le^{-t}$ as evidenced by Fig. 6(d). In the forced measurement case $r_\epsilon = 0$, we see no deviation from the power law up to $\epsilon \leq 4$.

A simple tractable model to get an intuition on the power-law/exponential-decay transition is when the charges are either static or always swap, i.e., $J = 0, \infty$ (see Appendix C for details). In this case, the asymptotic behavior for charge variance can be calculated exactly via saddle-point approximation. For $\epsilon > 0$ and $\epsilon_* = 0$, $\delta C^2 \propto L\epsilon^{-1}t^{-1/2}$, while for any $\epsilon_* > 0$, we have exponential decay with time, with $\delta C^2 \propto L\epsilon^{-1}t^{-1/2}e^{-r_\epsilon^2 t/2}$ when $r_\epsilon \leq 2\epsilon$ and $\delta C^2 \propto L \exp(-2\epsilon(r_\epsilon - 2\epsilon)t)$ when $r_\epsilon > 2\epsilon$. The reason for the change in scaling with t is due to the fact that the saddle point is at zero for $\epsilon_* = 0$ and therefore no leading order term is present. When $\epsilon_* > 0$, the saddle point shifts away from zero resulting in leading order factor $e^{-r_\epsilon^2 t/2}$. Therefore, the simple model has no phase transition in the forced measurement case and transitions into the exponentially decaying phase as soon as $\epsilon_* > 0$. In Fig. 4, we also overlay the boundary where the charge variance switches from a power law decay in time to an exponential one. We see good agreement with the boundaries found from TEBD.

Next we look at the sign of the averaged disconnected charge correlator, $\frac{1}{L-1} \sum_{x \neq 0} \text{sgn} \mathbb{E}_{Y,S^*}[\langle s_x s_0 \rangle]$. At $\epsilon = 0$, it is exactly zero due to it corresponding to SSEP at half filling. In the charge-fuzzy phase, both perturbation theory and

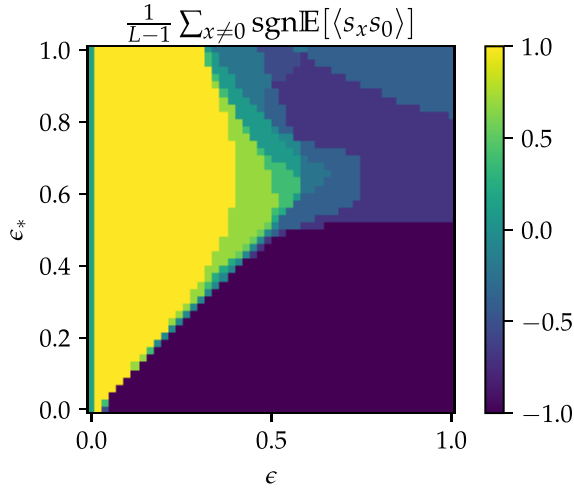


FIG. 7. Average sign of the disconnected charge correlator for $L = 14$ and $\#_{\text{samples}} = 100$, showing clear delineation between $r_\epsilon \leq 1$ charge-fuzzy regions.

replica field theory predicts that sign is positive for $r_\epsilon > 1$ and negative for $r_\epsilon < 1$. In the charge-sharp phase, we expect it to be zero. Therefore, the sign of the disconnected correlator distinguishes between the three regions. The predictions are well supported by the numerics as shown in Fig. 7, showing a sharp delineation between $r_\epsilon \leq 1$ charge-fuzzy regions, even for smaller system sizes.

IX. THE PLANTED XOR AND THE HAAR-RANDOM $U(1)$ -SYMMETRIC MONITORED QUANTUM TREE

In the previous section, we started with a monitored quantum circuit and derived a classical inference problem. Here, we go the other way around. In Sec. IX A, we start with a classical inference problem on an HMM we call *the planted XOR*. In Sec. IX B, we convert it to the Haar-random $U(1)$ monitored quantum tree with qubits and qudits of dimension d , which can also be referred to as *the quantum planted XOR*. For the task of inferring the final (qu)bit, the $d = \infty$ quantum case is equivalent to the classical HMM model. For $d < \infty$, however, the quantum model is not reducible to the classical HMM. For both the original HMM (Sec. IX C) and the fully quantum $d = 1$ limit (Sec. IX D), we exploit the structure of the tree and analytically deduce the full phase diagram in the student-teacher parameter space as Fig. 9. Remarkably, we uncover a re-entrance in the phase diagram in the quantum case, which has been previously seen numerically in the 2D RBIM (cf. Fig. 3).

A. Classical inference model

Consider the teacher, who prepares 2^N hidden Boolean variables $\mathbf{b}_1^* = b_{1:2^N,1}^*$, each chosen randomly with equal probability. At each time step, they apply the XOR gate on each neighboring pair, yielding $\mathbf{b}_2^* = b_{1:2^{N-1},2}^*$ at the next time step. This is repeated until there is only one Boolean $b_{1,N+1}^*$ left at final time step $N + 1$. Throughout this section, we will interchange between Ising and Boolean variables as $s_{i,t} = 2b_{i,t} - 1$.

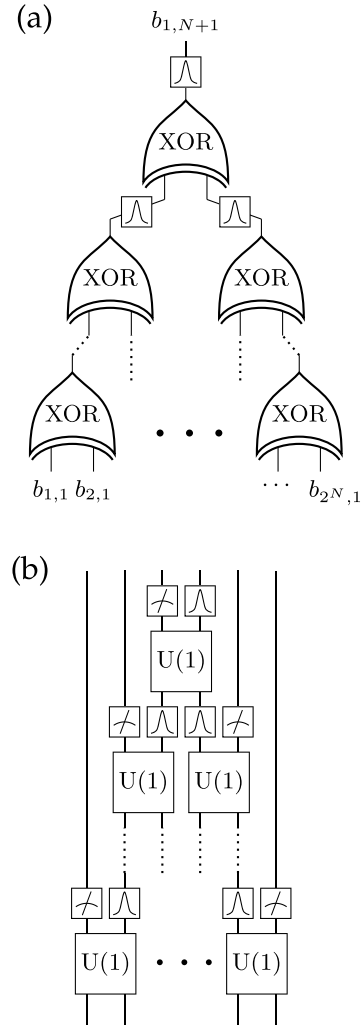


FIG. 8. (a) The planted XOR model and (b) The Haar-random $U(1)$ -symmetric quantum tree. Box with a meter represents projective measurements. Box with a bell curve represents a noisy observations in the classical inference case, and continuous weak measurement in the quantum case. Box with $U(1)$ denotes a Haar unitary acting on qubit-qudit pairs that conserves charge on the qubit sector.

Although the evolution for the Booleans is deterministic, it can still be described as an HMM. The prior for the entire trajectory $B^* := \mathbf{b}_{1:t}^*$ would be $p_*(B^*) = [\prod_t p_*(\mathbf{b}_t^* | \mathbf{b}_{t-1}^*)] p_*(\mathbf{b}_1^*)$. The initial choice of Booleans is still random, $p_*(\mathbf{b}_1^*) = 2^{-N}$, but the “Markov” kernel is deterministic and separable as $p_*(\mathbf{b}_t^* | \mathbf{b}_{t-1}^*) = \prod_i p_*(b_{i,t}^* | b_{i_L,t-1}^*, b_{i_R,t-1}^*)$, where

$$p_*(b_{i,t}^* | b_{i_L,t-1}^*, b_{i_R,t-1}^*) = \delta\{b_{i,t}^*, \text{XOR}(b_{i_L,t-1}^*, b_{i_R,t-1}^*)\}. \quad (74)$$

Here, we recall that $\delta\{\cdot, \cdot\}$ is the Kronecker delta.

At each time step, the teacher also generates observations on each Boolean $b_{i,t}^*$ $y_{i,t} = \epsilon_* s_{i,t}^* + \sigma_* \eta_{i,t}$, yielding the measurement model [Eq. (54)] as in the planted SSEP case.

The goal of the student is to determine the last bit $b_{1,N+1}^*$, or equivalently, the parity of the starting bits. They can do this by inferring a posterior for the filtering task, $p_s(b_{1,N+1} | Y)$. The

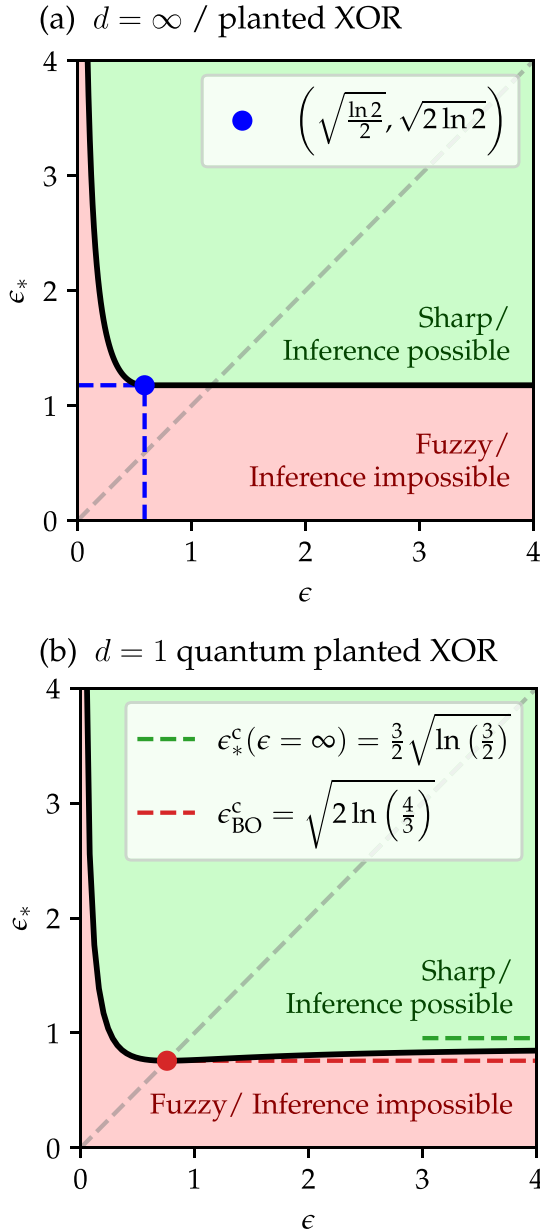


FIG. 9. Exact expanded phase diagrams for the (a) classical planted XOR or the $d = \infty$ Haar-random U(1)-symmetric monitored quantum tree and (b) quantum XOR or the $d = 1$ Haar-random U(1)-symmetric monitored quantum tree, the latter of which is given by Eq. (102). Note the re-entrance of the phase diagram, similar to that of 2D RBIM Fig. 3.

natural order parameter then would be the MSEM or the MSE of the final bit.

B. Quantum tree model

To map the above classical inference problem into a quantum one, we first note that given input bits b_L, b_R , the SSEP kernel [Eq. (32)] allows for the possible transitions $00 \rightarrow 00$, $01 \rightarrow 01$, $01 \rightarrow 10$, $10 \rightarrow 01$, $10 \rightarrow 10$, and $11 \rightarrow 11$. Then, we observe that if we condition on the left output bit to be a 0, the right output bit is $\text{XOR}(b_L, b_R)$. Conversely, if we condition the left output bit to be a 1, then the right output

bit is $\text{NOTXOR}(b_L, b_R)$. Then, applying the SSEP gate then giving the information about the left output bit is equivalent to applying an XOR or NOTXOR and revealing which was used. As this information is completely specified, the model can be exactly mapped to the above situation of applying the XOR each time (and the student knowing this fact).

To construct the quantum tree model, we do the opposite procedure of Sec. VI. Since the SSEP kernel [Eq. (32)] is obtained from $u_{U(1)}$, we replace each XOR gate with a $u_{U(1)}$ gate with one output qubit projectively measured. The observation model is replaced by weak gaussian measurements [Eq. (28)]. This results in a quantum tree as sketched in Fig. 8(b).

As for the initial state, we will select the scenario where teacher randomly chooses from an equally weighted ensemble of product states of qubit-strings and equal superposition of qudit strings in the computational basis, $|\chi\rangle$ (introduced in Sec. VIII A),

$$\{|\psi_0\rangle\} = \{|b_{1:2^N}^*\rangle \otimes |\chi\rangle^{\otimes 2^N} : b_{1:2^N}^* \in \{0, 1\}^{2^N}\}. \quad (75)$$

The teacher evolves the quantum tree and sends the choice of unitary gates sampled from the Haar distribution and the projective and weak measurement outcomes to the student. At the end, the teacher measures the final qubit in the computational basis to obtain Z_* . The student then tries to guess the final projective qubit measurement performed at the end by the teacher.

In the case of general local qudit dimension $d < \infty$ and revealed choice of Haar unitaries, the inference problem on this quantum tree is not reducible to a classical HMM. A similar model with both measurements being projective with a finite rate was studied in Ref. [33] at Bayes optimality.

In the inference setting, the natural order parameter to study would be the mean squared errors or the alignment between the true and inferred final qubit,

$$\begin{aligned} \text{MSE} &= \mathbb{E}_{U, Y, S^*} [(s_{1,N+1} - s_{1+N+1}^*)^2] \\ &= 4 - 4 \mathbb{E}_{U, Y, S^*} [p_s(s_{1,N+1} = s_{1+N+1}^* | Y)]. \end{aligned} \quad (76)$$

$$\begin{aligned} \text{MSEM} &= \mathbb{E}_{U, Y, S^*} [(\langle s_{1,N+1} \rangle - s_{1+N+1}^*)^2], \\ &= 4 - 4 \mathbb{E}_{U, Y, S^*} [p_s(s_{1,N+1} = s_{1+N+1}^* | Y)] \\ &\quad + 4 \mathbb{E}_{U, Y, S^*} [p_s(s_{1,N+1} = -1 | Y)p_s(s_{1,N+1} = 1 | Y)], \end{aligned} \quad (77)$$

and the student's final qubit variance,

$$\delta Z_s^2 = \mathbb{E}_{U, Y} [\langle Z^2 \rangle - \langle Z \rangle^2], \quad (79)$$

$$= 4 \mathbb{E}_{U, Y} [p_s(s_{1,N+1} = -1 | Y)p_s(s_{1,N+1} = 1 | Y)]. \quad (80)$$

In the standalone quantum tree case, the natural order parameter would be the teacher's final qubit variance δZ_*^2 , equal to Eq. (79) at Bayes optimality.

C. Analysis of the classical ($d = \infty$ quantum) planted XOR

Due to the tree structure and Booleanity of the model, the model can be studied without resorting to the replica trick. The key observation is that the student's posterior, or the density

matrix, can be treated as a random variable dependent on the observations Y . Furthermore, due to the forward equation and the tree structure, the posterior for a particular bit at a particular time $p(b_{i,t}|\mathbf{y}_{1:t})$ only depends on the posterior for the left and right bits at previous times, $p(b_{i_L,t-1}|\mathbf{y}_{1:t-1})$, $p(b_{i_R,t-1}|\mathbf{y}_{1:t-1})$, which are independent, and likewise for the density matrix. Due to Booleanity, one only needs to keep track of the posterior for one value of the Boolean, since $p(b_{i,t}|\mathbf{y}_{1:t}) = 1 - p(\bar{b}_{i,t}|\mathbf{y}_{1:t})$.

1. Finding the critical parameters

Then, we can apply the method of traveling waves used to study the directed polymer on a tree in Ref. [34]. In that case, the partition function at time t , Z_t , evolved linearly with independent partition functions at previous times as $Z_t = e^{-\beta V} [Z_{t-1}^{(1)} + Z_{t-1}^{(2)}]$, where $V \sim \mathcal{N}(0, 1)$ in their case. Then, a generating function $G_t(x) := \mathbb{E}_V[\exp(-e^{\beta x} Z_t)]$ is defined and evolves as the FKPP (Fisher-Kolmogorov-Petrovsky-Piskunov) equation. In our case, the posterior at time t evolves nonlinearly as a function of the posterior at previous time $t-1$. However, if there exists a transition as parameters are varied, we expect the posterior of the student evaluated at the true Boolean to be close to one, $p_s(b_{i,t} = b_{i,t}^*|\mathbf{y}_{1:t}) \approx 1$, or equivalently, the posterior evaluated at the opposite value

$\bar{b}_{i,t}^*$ to be small $p_s(b_{i,t} = \bar{b}_{i,t}^*|\mathbf{y}_{1:t}) \ll 1$. Therefore, we can linearize the evolution equation and continue with the same analysis. This is the same strategy used in Ref. [32].

Concretely, consider the posterior for a boolean $b = b_{i,t}$ on a particular branch at a given time step, $p(b|\mathbf{y}_{1:t})$. Due to the structure of the tree, it is in fact only dependent on observations from branches connected to it from the past, i.e., $p(b|\mathbf{y}_{1:t}) = p(b|\mathbf{y}_{1:t}^c)$. Using the forward equation, this can be related to the posterior for booleans connected from left $b_L = b_{i_L,t-1}$ and the right $b_R = b_{i_R,t-1}$ from the previous time step,

$$p(b|\mathbf{y}_{1:t}^c) = \frac{p(y_t|b) \sum_{b_{L,R}} p(b|b_{L,R}) p(b_L|\mathbf{y}_{1:t-1}^L) p(b_R|\mathbf{y}_{1:t-1}^R)}{p(\mathbf{y}_{1:t}^c)}. \quad (81)$$

Here, the marginalized likelihood is $p(\mathbf{y}_{1:t}^c) = \sum_{b,b_{L,R}} p(y_t|b) p(b|b_{L,R}) p(b_L|\mathbf{y}_{1:t-1}^L) p(b_R|\mathbf{y}_{1:t-1}^R)$. $y = y_t$ is not iid but is distributed as $y = \eta + \epsilon_* s^*$, with $s^* = -s_L^* s_R^*$. Near the phase transition, we assume that $p(b = b^*|\mathbf{y}_{1:t}^c)$ is large and $p(b = \bar{b}^*|\mathbf{y}_{1:t}^c)$ is small, and likewise for the posteriors for b_L and b_R . If $b^* = 1$, we can either have $(b_L, b_R) = (0, 1)$ or $(1, 0)$. If $b^* = 0$, we can have $(1, 1)$ or $(0, 0)$. For every case, denoting the smaller posterior value as $p_t = p(b = \bar{b}^*|\mathbf{y}_{1:t}^c)$ and likewise for $p_t^{L,R} = p(b_{L,R} = \bar{b}_{L,R}^*|\mathbf{y}_{1:t-1}^{L,R})$, the posterior evolution equation is the same for every four possibilities mentioned above (up to changing sign of η),

$$p_t = \frac{(p_{t-1}^L + p_{t-1}^R - 2p_{t-1}^L p_{t-1}^R) e^{\epsilon(\eta - \epsilon_*)}}{((p_{t-1}^L + p_{t-1}^R - 2p_{t-1}^L p_{t-1}^R) e^{\epsilon(\eta - \epsilon_*)} + (1 - p_{t-1}^L - p_{t-1}^R + 2p_{t-1}^L p_{t-1}^R) e^{-\epsilon(\eta - \epsilon_*)})}, \quad (82)$$

which can be expanded to linear order as

$$p_t = A(p_{t-1}^L + p_{t-1}^R), \quad (83)$$

with $A = e^{2\epsilon(\eta - \epsilon_*)}$, which is importantly iid.

To understand the evolution of typical p_t with increasing t , we define the generating function $G_t(x) = \langle \exp(-e^{-x} p_t) \rangle$. The linearized recursion can be expressed as

$$G_{t+1}(x) = \mathbb{E}_\eta [G_t(x - \ln A) G_t(x - \ln A)], \quad (84)$$

with traveling wave ansatz $G_t(x) = G^{(\lambda)}(x - v_{\epsilon, \epsilon_*}(\lambda)t)$. $\lambda \in [0, 1]$ parametrizes a family of possible traveling wave solutions and will be chosen to be λ_c which minimizes the velocity $v_{\epsilon, \epsilon_*}(\lambda)$, subject to the initial condition $G_0(x) = \exp(-e^{-x}/2) \xrightarrow{x \gg 1} 1 - (1/2)e^{-x}$ [31,34]. Using the exponential decay of the wave ansatz at large arguments $G^{(\lambda)}(u) \simeq 1 - e^{-\lambda u}$, and that $\mathbb{E}_\eta[A^\lambda] = e^{2\lambda\epsilon(\lambda\epsilon - \epsilon_*)}$, we can find the wavefront velocity as

$$v_{\epsilon, \epsilon_*}(\lambda) = \frac{1}{\lambda} \ln \left(2 \mathbb{E}_\eta[A^\lambda] \right) = \frac{\ln 2}{\lambda} + 2\epsilon(\lambda\epsilon - \epsilon_*). \quad (85)$$

minimizing it respect to $\lambda \in [0, 1]$, we find that

$$\lambda_c = \min \left(1, \sqrt{\frac{\ln 2}{2} \frac{1}{\epsilon}} \right). \quad (86)$$

Therefore, the velocity is given by $v_{\epsilon, \epsilon_*} = v_{\epsilon, \epsilon_*}(\lambda_c)$. If the velocity is positive, then it means that p_t is exponentially

decaying as we increase t . If velocity is negative, then p_t is exponentially growing with time and the nonlinear part of the evolution becomes important. Then, the critical point is the velocity changes sign, i.e., when $v_{\epsilon, \epsilon_*} = 0$. Combining the two conditions, we find

$$\epsilon_*^c(\epsilon) = \begin{cases} \frac{\ln 2}{2\epsilon} + \epsilon \approx 0.346574\epsilon^{-1} + \epsilon, & \text{for } \epsilon < \sqrt{\frac{\ln 2}{2}}, \\ \sqrt{2 \ln 2} \approx 1.17741, & \text{for } \epsilon \geq \sqrt{\frac{\ln 2}{2}}. \end{cases} \quad (87)$$

Using this result, we can deduce that in the regime where $\epsilon > 0$, $\epsilon_* > \epsilon_*^c$, as the student correctly deduces the final Boolean, both MSE, MSEM and δZ_s^2 would decay exponentially with time and tend to zero as $t \rightarrow \infty$. On the other hand, for $\epsilon > 0$, $\epsilon_* < \epsilon_*^c$, we would expect to converge to a nonzero value. The derived phase diagram is plotted in Fig. 9(a).

2. Power-law scaling of bit variance in the $\epsilon_* = 0$ limit

The scaling of the long-time bit variance can be analyzed in the $\epsilon_* \rightarrow 0$ and large ϵ limit. Here, we cannot *a priori* expect either of $p_s(1|Y)$ or $p_s(0|Y)$ to be larger than the other. Instead, we will exploit the iid nature of the disorder to extract the distribution of the posterior.

Let us arbitrarily choose one of the posterior $\pi := p_s(1|\mathbf{y}_{1:t}^c)$ to study. We will denote $\pi' := p_s(1|\mathbf{y}_{1:t-1}^L)$ and

$\pi'' := p_s(1|y_{1:t-1}^R)$. Then, the evolution is

$$\pi = \frac{e^{\epsilon y} A_+}{e^{\epsilon y} A_+ + e^{-\epsilon y} A_-}, \quad (88)$$

where $A_+ = \pi'(1 - \pi'') + (1 - \pi')\pi''$ and $A_- = \pi'\pi'' + (1 - \pi')(1 - \pi'')$. For forced measurements, y is iid and uncorrelated with π' and π'' .

The probability distribution for π then evolves as

$$p(\pi) = \int dy \frac{e^{-y^2/2}}{\sqrt{2\pi}} \int_0^1 dp(\pi') dp(\pi'') \times \delta\left(\pi - \frac{e^{\epsilon y} A_+}{e^{\epsilon y} A_+ + e^{-\epsilon y} A_-}\right). \quad (89)$$

Define $g(y) := \pi - \frac{e^{\epsilon y} A_+}{e^{\epsilon y} A_+ + e^{-\epsilon y} A_-}$. The delta function of a function is given by $\delta(g(y)) = \sum_i \frac{1}{|\partial_y g(y_i)|} \delta(y - y_i)$, where y_i 's are zeros of $g(y)$. First solving for $g(y) = 0$, we get $y_0 = \frac{1}{2\epsilon} \ln\left(\frac{\pi_1 A_-}{(1 - \pi_1) A_+}\right)$. Differentiating $g(y)$ and then substituting y_0 , we get $\partial_y g(y_0) = 2\epsilon\pi(1 - \pi)$. Therefore, we have

$$p(\pi) = \frac{1}{2\epsilon\pi(1 - \pi)\sqrt{2\pi}} \int_0^1 dp(\pi') dp(\pi'') e^{-\frac{1}{8\epsilon^2} \ln^2\left(\frac{\pi A_-}{(1 - \pi) A_+}\right)}. \quad (90)$$

We now look for the steady state, that is, $p(\pi) = p(\pi' = \pi) = p(\pi'' = \pi)$. Then, the above equation becomes a self-consistent equation/Neumann-type equation and we have a systematic way of expanding the series, if desired. Expanding with respect to large ϵ or small $1/\epsilon$ we find that

$$p(\pi) = \frac{1}{2\epsilon\pi(1 - \pi)\sqrt{2\pi}} + \mathcal{O}(\epsilon^{-5}). \quad (91)$$

Although the series is not normalizable, we can calculate the leading order of the student's bit variance, $\delta Z_s^2 = 4p_0 p_1$. Using that $p_0 p_1 = \pi(1 - \pi)$, we find that

$$\begin{aligned} \delta Z_s^2 &= 4\mathbb{E}[p_0 p_1] = 4 \int_0^1 d\pi \cdot \pi(1 - \pi)p(\pi) \\ &= \frac{2}{\epsilon\sqrt{2\pi}} + \mathcal{O}(\epsilon^{-5}), \end{aligned} \quad (92)$$

a power law in the student's assumed SNR, $\sim \epsilon^{-1}$. This indicates that the bit variance does not vanish at long times on the $\epsilon_* = 0$ line.

D. Analysis of the $d = 1$ quantum planted XOR

Approaching the sharpening/inferrability transition from the qubit "sharp" phase, we expect the reduced density matrix of the student's density matrix of the final qubit to be collapsed to one of the two qubit values. Therefore, we parametrize the student's reduced density matrix at a particular node as

$$\rho_{s,t} = (1 - z_t)|b\rangle\langle b| + z_t|\bar{b}\rangle\langle \bar{b}|, \quad (93)$$

where we take z_t to be a small parameter, and therefore b to be the most likely qubit. The student's density matrix then evolves as

$$\rho_{s,t+1} = \frac{(Q_{s,y} \otimes P_{b'})U(\rho_{s,t}^L \otimes \rho_{s,t}^R)U^\dagger(Q_{s,y} \otimes P_{b'})^\dagger}{\text{tr}[(Q_{s,y} \otimes P_{b'})U(\rho_{s,t}^L \otimes \rho_{s,t}^R)U^\dagger(Q_{s,y} \otimes P_{b'})^\dagger]}, \quad (94)$$

where $Q_{s,y}$ is the student's partial projector on (say) the left qubit given weak measurement outcome y , and $P_{b'}$ is the projector on the right qubit given strong measurement outcome b' , and U is a Haar-random $U(1)$ -symmetric gate. The measurement outcomes have probability the teacher's system,

$$p_*(y, s'|U) = \text{tr}[(Q_{s,y} \otimes P_{s'})U(\rho_{*,t}^L \otimes \rho_{*,t}^R)U^\dagger(Q_{s,y} \otimes P_{s'})^\dagger]. \quad (95)$$

Here, we recall that we will freely change notation between bits and Pauli- z eigenvalues as $s = 2b - 1$. We will assume that $\rho_{*,t}^{L/R}$ has the same most likely qubit as that of the student, a sensible assumption in the inferrable phase. We would like to find the linearized evolution of z_t , and thereby construct the evolution for the generating function $G_t(x) = \mathbb{E}[\exp(-e^{-x} z_t)]$. Here, we must keep track of both (z_t, b) and consider all the cases $(b_L, b_R, b') = (0, 0, 0), (0, 0, 1), \dots, (1, 1, 1)$, and appropriately weight them by their probabilities. Here, $b_{L/R}$ is the most likely qubit on the density matrices at time step t connected on the left/right site.

There are three kinds of cases. The first is $b_L = b_R = b'$. In this case, we have

$$z_{t+1} = e^{-2s'y\epsilon} (|u|^2 z_t^L + |u'|^2 z_t^R), \quad (96)$$

where u and u' are elements from different columns of the Haar unitary on the $|01\rangle, |10\rangle$ sector. This occurs with probability up to zeroth order in z_* as

$$p_*(y, s' | U, s_L = s_R) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y - s'\epsilon_*)^2}{2}\right). \quad (97)$$

We neglect the case $b_L = b_R \neq b'$ as it occurs with probability $p_*(y, b'|U) \sim \mathcal{O}(z_*)$. For the $b_L \neq b_R$ case, we find

$$z_{t+1} = \frac{1}{|u|^2} e^{2s'y\epsilon} z_t^{L/R}, \quad (98)$$

occurring with probability up to zeroth order in z_* as

$$p_*(y, s' | U, s_L \neq s_R) = \frac{|u|^2}{\sqrt{2\pi}} \exp\left(-\frac{(y + s'\epsilon_*)^2}{2}\right). \quad (99)$$

Then, the generating function is given by

$$\begin{aligned} G_{t+1}(x) &= \int_{U, y, z_t^{L/R}} \sum_{s_L, s'} \exp(-e^{-x} z_{t+1}(s_L, s')) \\ &\quad \times p_*(y, s', U | s_L, s_R) p(s_L, z_t^L) p(s_R, z_t^R). \end{aligned} \quad (100)$$

Substituting the linearized evolution and their probability for the two different cases and using the asymptotic form $G^{(\lambda)}(u) \approx 1 - e^{-\lambda u}$, we find the velocity

$$v_{\epsilon, \epsilon_*}(\lambda) = 2\epsilon(\lambda - \epsilon_*) + \frac{1}{\lambda} \ln(|u|^{2\lambda} + |u|^{2-2\lambda})_u. \quad (101)$$

For a particular (ϵ, ϵ_*) , the velocity is given by $\lambda \in [0, 1]$ that minimizes the velocity given λ . The phase transition occurs when $v_{\epsilon, \epsilon_*}(\lambda_c) = 0$. Using that $\langle |u|^{2n} \rangle = 1/(1 + n)$, the exact

phase boundary is then given by

$$\epsilon_*^c(\epsilon) = \min_{\lambda \in [0,1]} \left[\lambda \epsilon + \frac{1}{2\lambda} \ln \left(\frac{3}{2 + \lambda - \lambda^2} \right) \right], \quad (102)$$

while λ_c is given by the argmin. In general, this requires a solution to a transcendental equation. It is plotted in Fig. 9(b). However, in the Bayes optimal case $\epsilon_* = \epsilon$, Eq. (102) can be found in closed form,

$$\epsilon_*^c(\epsilon = \epsilon_*) = \sqrt{2 \ln \left(\frac{4}{3} \right)} \approx 0.758\,528, \quad (103)$$

with $\lambda_c = 1/2$, as well as in the limit of infinite student signal,

$$\epsilon_*^c(\epsilon = \infty) = \frac{3}{2} \sqrt{\ln \left(\frac{3}{2} \right)} \approx 0.955\,142, \quad (104)$$

where $\lambda_c \rightarrow 0$. Finally, in the $\epsilon \rightarrow 0$ limit, $\lambda_c \approx 0.916\,732$ [35] and the critical teacher signal strength diverges as $\epsilon_*^c \approx 0.200\,717\epsilon^{-1}$.

Similarly to the classical planted XOR case, we can deduce that in the regime where $\epsilon > 0$, $\epsilon_* > \epsilon_*^c$, for a typical trajectory, the student's density matrix collapses onto one of the qubit and correctly deduces the final qubit. Therefore all of MSE, MSEM and δZ_s^2 would decay exponentially with time and tend to zero as $t \rightarrow \infty$. On the other hand, for $\epsilon > 0$, $\epsilon_* < \epsilon_*^c$, we would expect to converge to a nonzero value.

The phase diagram for $d = 1$ has two notable features. First, at fixed ϵ , the SNR threshold is lower than that of $d = \infty$. Recall that the $d = \infty$ case also corresponds to the inference problem where U is not provided (for any d), whereas in the $d = 1$ case considered here, it is provided. Therefore, the lowering of the threshold is expected, as the student has more information. Second, the phase diagram features a re-entrance. This is qualitatively similar to the re-entrance seen in the QEC/2D RBIM (see Fig. 3), but found exactly in our case.

X. DISCUSSION OF THE PHASE DIAGRAMS

In this section, we discuss the obtained phase diagrams. First, we discuss their general features followed by a rigorous argument as to why the phase boundary cannot curve downwards after hitting the Bayes optimal line. Then, focusing on quantum inference problems, we discuss the operational consequences of these phase diagrams.

In Sec. III, we introduced order parameters that pertain to inference (or “learnability”), such as MSEM and MSE/2, and an order parameter related to the student's uncertainty (or “sharpness”) in their posterior, δO_s^2 . Though these order parameters are equal at Bayes optimality, they need not be away from it. Therefore, there could be one phase boundary between observable sharp/fuzzy phases, and another between inference possible/impossible phases. However, in the three examples of (quantum) inference problems studied, they coincided, i.e., the sharp phase corresponded to the inference possible phase, and the fuzzy phase corresponded to the inference impossible phase. Recall from Sec. VII that in the RBIM case, quality of inference is quantified by FM order, while sharpness of the student's posterior is quantified by EA-like order parameters. The SG phase is characterized by the lack

of FM order but the presence of EA order. Therefore, the SG phase can be intuitively thought of as a phase where the student is confident about their *wrong* answer. In the language of spin glasses, the fact that the phases of inference coincide with those of sharpness means that there is no SG phase. Note that there are examples of inference models where SG-like phases exist, such as the planted RBIM on random graphs [11], or the planted directed polymer on the tree [8].

The phase boundaries in these examples delineate when the student can infer the state despite the nonoptimality/inaccuracy of their measurement model. Note the similarity of these phase diagrams (cf. Figs. 3, 4, and 9). In all three cases, the studied parameters quantify the amount of true signal strength (quantified by ϵ_* or β_*) or assumed signal strength (quantified by ϵ or β) passed through the measurements. Below the Bayes optimal line, we find two scenarios: either the boundary plateaus (as for the classical planted XOR model or the planted directed polymer on a tree [8]) or it curves slightly upwards to display a re-entrance (as shown numerically for the 2D RBIM or exactly for our $d = 1$ quantum planted XOR model).

That the phase boundary for the MSEM cannot curve down has the status of a theorem. It follows from the Bayes optimal estimator theorem [Eq. (8)] that for a given ϵ_* the MSEM is minimized at Bayes optimality ($\epsilon = \epsilon_*$). Thus, if a given Bayes optimal point ($\epsilon = \epsilon_*$, ϵ_*) is in the inference impossible phase, then so are all points on the horizontal line passing through it. A corollary is that the phase boundary cannot pass through values of ϵ_* below the critical value $\epsilon_{\text{BO}}^\#$ on the Bayes optimal line.

In measurement-induced phase transitions, forced measurements represent an often-tractable limit where measurement outcomes are chosen to be iid instead of following the Born Rule. In our expanded phase diagram, this has the simple interpretation of $\epsilon_* = r_\epsilon = 0$. Then, from the above discussion, a transition for forced measurements corresponds to a spin-glass-like transition. In particular, we have demonstrated that there is no transition for this limit for the planted SSEP and XOR models.

Although not discussed earlier, other problems on quantum systems can also be viewed through the lens of the general quantum inference problem, such as quantum state preparation [36] or quantum state tomography [37]. In the former case, an experimentalist wants to know whether it is possible to infer the correct operation to prepare the desired quantum state given measurements. The same teacher-student framework can be used to describe the underlying process generating the measurements and the experimentalists' measurement model.

What are the operational consequences of these phase diagrams? Remaining within the quantum state preparation setting, let us consider a situation where ϵ and ϵ^* vary together as the assumed signal strength is varied, meaning that we will move along a line $\epsilon^*(\epsilon)$ in a typical inference phase diagram like Fig. 10. An example of such situation may be one where you (an experimentalist) prepare and evolve the quantum circuit with parameter ϵ , and receive measurement outcomes Y . Then, you would assume that you prepared the state $\rho_\epsilon(\epsilon, Y)$. However, due to some systemic and/or random errors, the parameters are actually ϵ^* and in fact the state you

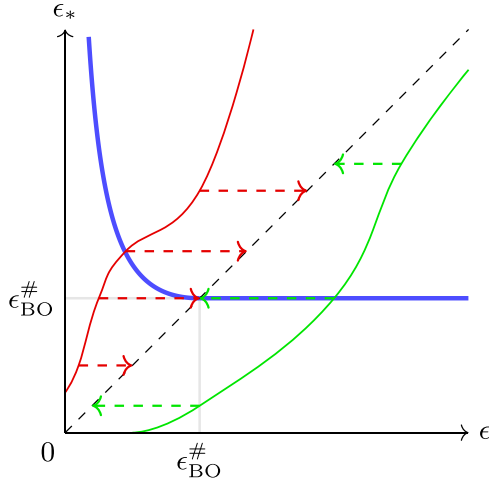


FIG. 10. Scenario where ϵ_* vary together with ϵ for a typical phase diagram of a quantum inference problem (red and green solid lines). Here, we approximate that the phase boundary (blue line) flattens at the Bayes optimal transition point with no re-entrance. The phase that an experimentalist detects using their classical decoder is given by the phase on the red or green lines. The true state of the prepared state is given by the phase on the diagonal black dashed line horizontal to the point on the red or green lines, as connected by the red and green dashed lines. When measurement strength is overestimated (green line, $r_\epsilon < 1$), the detected phase on the decoder is the same as that of the prepared state. When the measurement strength is underestimated (red line, $r_\epsilon > 1$), as the measurement strength is increased, the phase transition is detected on the decoder *after* the real one occurs.

prepared is $\rho_t(\epsilon^*, Y)$. Nevertheless, you may hope that ϵ and ϵ^* are in some way correlated, and in the simplest scenario ϵ^* is a simple function of ϵ . Suppose then that you repeat the experiment for a range of ϵ 's. You receive measurement outcomes from the system with parameter $\epsilon^*(\epsilon)$. Using your classical decoder, you detect a phase transition at $\epsilon^\#$.

Let us consider the case when $\epsilon^*(\epsilon)$ is monotonic, and $r_\epsilon < 1$, which corresponds to overestimating the measurement strength/rate in the quantum circuit. For simplicity, let us also approximate that the phase boundary flattens at the Bayes optimal transition point with no re-entrance. As shown in Fig. 10, the relationship between the phase of the assumed and the true quantum states is given by drawing a horizontal line from the assumed point to the 45° Bayes optimal line. In this case, we see that when the the classical decoder detects the phase transition at $\epsilon^\#$, the true quantum state is also prepared in the critical state. In fact, the phase detected by the classical decoder always coincides with that of the true quantum state.

Instead, assuming all of the above but for $r_\epsilon > 1$, which corresponds to underestimating the measurement strength, as ϵ is increased, the real quantum state undergoes the phase transition into the charge-sharp phase before the classical decoder detects it. When the classical decoder detects a phase transition, the real quantum state is already deep in the charge-sharp phase.

This would suggest that if the phase boundary is horizontal to the right of the Bayes optimal transition point with no re-entrance, then you should always overestimate the mea-

surement strength if you want to prepare the true quantum state in the correct phase. Even if the phase boundary is not strictly horizontal, it is better to overestimate the measurement strength than underestimate it, because the curvature of the phase boundary means you have a better chance of preparing the quantum state in the intended phase.

If the precise numerical value of the phase transition at Bayes optimality is already known (ex. through simulations), then you would see the phase transition detected by the classical decoder at the wrong value of ϵ . This would provide the systematic error between the assumed parameters and true parameters. Finally, in the special case of the planted SSEP, within the charge-fuzzy phase, the average sign of the disconnected correlator [Eq. (56)] could also be used to detect Bayes nonoptimality (Fig. 7).

XI. OUTLOOK

In this work, we introduced *the general quantum inference problem*, a generalization of classical Bayesian inference to quantum systems. We also defined scenarios where the quantum inference model has the structure of qHMMs, which generalized HMMs in classical Bayesian inference. We discussed the requirements for qHMMs to reduce to classical HMMs and proved such cases for qHMMs with Haar-random unitary gates and measurements. Using the teacher-student scenario, we introduced the notion of Bayes nonoptimality, where the model assumed by the student differs from the true model of the teacher. This allows us to go beyond existing work and expand the phase diagram in the student-teacher parameters. Under this lens we studied three explicit examples: (1) quantum error correction on the toric and repetition code under bit-flip errors, which map onto the 2D random-bond Ising model, (2) Haar-random U(1)-symmetric monitored quantum circuit or the planted SSEP, and (3) Haar-random U(1)-symmetric monitored quantum tree or the planted XOR. Then, we discussed the obtained phase diagrams, with a rigorous argument on their shape and their interpretation in quantum inference problems.

Many avenues for future research exist. One could reverse the quantum to classical inference problem mapping discussed Sec. VI to study “quantum” versions of classical inference problems. One could explore the more general Bayes nonoptimality where even the functional form of the student and teacher’s models differ. This was done for the planted directed polymer on a tree, where the teacher could have any arbitrary noise model while the student assumes a Gaussian one [8]. It is also a common case for many QEC problems where approximate decoders are used [38]. One could also explore non-Markovian quantum inference problems, such as non-Markovian errors in QEC problems [39].

ACKNOWLEDGMENTS

S.W.P.K. acknowledges support from EPSRC DTP International Studentship Grant Ref. No. EP/W524475/1. C.v.K. was supported by a UKRI FLF MR/Z000297/1. A.L. acknowledges support from EPSRC Critical Mass Grant Ref. No. EP/V062654/1.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX A: DERIVATION OF THE REPLICA FIELD THEORY FOR THE PLANTED SSEP

1. Setup

Consider a 1D lattice of Ising variables $\mathbf{s} = (s_1, \dots, s_L)$, picked from some separable initial distribution $p(\mathbf{s}) = \prod_x p(s_x)$, with $p(s_x) = \pi \delta_{s_x, 1} + (1 - \pi) \delta_{s_x, -1}$.

Let it evolve through charge-conserving Markovian dynamics, $p(\mathbf{s}_t | \mathbf{s}_{t-1}) = \prod_{x \in \mathcal{R}(t)} p(s_{x:t+1,t} | s_{x:t+1,t-1})$, where $\mathcal{R}(t)$ are even/odd sites for even/odd t , and the transition are given by

$$p(s_{x:t+1,t} | s_{x:t+1,t-1}) = \langle s_{x:t+1,t} | u_{x:t+1,t} | s_{x:t+1,t-1} \rangle, \quad (\text{A1})$$

where

$$u_{x:t+1,t} := \left(\frac{1 - \sigma_x^z \sigma_{x+1}^z}{2} \right) \left(\frac{1}{2} (\sigma_x^+ \sigma_{x+1}^- + \sigma_x^- \sigma_{x+1}^+) + \frac{1}{2} \right) + \frac{1 + \sigma_x^z \sigma_{x+1}^z}{2}. \quad (\text{A2})$$

Let us denote the entire trajectory as $S := s_{1:t}$. Then, we have $p(S) = p(\mathbf{s}_t | \mathbf{s}_{t-1}) \cdots p(\mathbf{s}_2 | \mathbf{s}_1) p(\mathbf{s}_1)$.

Given the dynamics, we generate images over all times $Y := \mathbf{y}_{1:t}$, where each image consisting of pixels $\mathbf{y}_t := (y_{1,t}, \dots, y_{L,t})$ from Ising configurations is given by $p(Y | S) = \prod_{x,t} p(y_{x,t} | s_{x,t})$, where

$$p(y_{x,t} | s_{x,t}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_{x,t} - \epsilon s_{x,t})^2}{2\sigma^2}\right) = \langle s_{x,t} | P_x(y_{x,t}) | s_{x,t} \rangle, \quad (\text{A3})$$

where $P_x(y_{x,t}) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_{x,t} - \epsilon s_{x,t})^2}{2\sigma^2}\right)$.

Let the parameters of the evolution process and imaging process be unknown, and denote them via $(\pi_*, \epsilon_*, \sigma_*)$. Assuming some other set of parameters (π, ϵ, σ) , the posterior distribution over S given the images Y is given by

$$p(S | Y) = \frac{p(Y | S) p(S)}{p(Y)}, \quad (\text{A4})$$

where $p(Y) = \sum_S p(Y | S) p(S)$ is given by the assumed parameters. The distribution of Y and true states S^* is given by $p_*(Y, S) = p_*(Y | S^*) p_*(S^*)$ and $p_*(Y) = \sum_{S^*} p_*(Y | S^*) p_*(S^*)$, where the star denotes that the true parameters are used.

We have

$$p(Y, S) = p(Y | S) p(S) = \left(\prod_{t=2}^L (s_t | P(\mathbf{y}_t) \mathbf{u}_t | s_{t-1}) \right) \langle s_1 | P(\mathbf{y}_1) | \boldsymbol{\pi} \rangle, \quad (\text{A5})$$

where $P(\mathbf{y}_t) = \bigotimes_x P_x(y_{x,t})$ and $\mathbf{u}_t = \bigotimes_{x \in \mathcal{R}(t)} u_{x:t+1,t}$, and $|\boldsymbol{\pi}\rangle = \bigotimes_x |\pi\rangle_x$ and $|\pi\rangle_x = \pi |1\rangle_x + (1 - \pi) |-1\rangle_x$.

Although we will sometimes keep track of normalization, we will show now that the normalization factor *should* not matter in the replica calculation. For normalized probabilities, we have

$$1 = \sum_{S^{1:n}, S^*} \int dY p(S^{1:n}, Y, S^*) = \sum_{S^{1:n}, S^*} \int dY \frac{p(S^1, Y) \cdots p(S^n, Y)}{p(Y)^n} p_*(S^*, Y). \quad (\text{A6})$$

Let's consider unnormalized distributions $q(S, Y) := c \cdot p(S, Y)$ and $q_*(S^*, Y) := c^* \cdot p_*(S^*, Y)$, and define

$$\mathcal{Z} = \sum_{S^{1:n}, S^*} \int dY \frac{q(S^1, Y) \cdots q(S^n, Y)}{q(Y)^n} q_*(S^*, Y) = c^*. \quad (\text{A7})$$

Then, any mixed moment observable can be written as

$$O = \frac{1}{\mathcal{Z}} \sum_{S^{1:n}, S^*} \int dY O(S^{1:n}, S^*) \frac{q(S^1, Y) \cdots q(S^n, Y)}{q(Y)^n} q_*(S^*, Y) = \lim_{Q \rightarrow 1} \frac{1}{\mathcal{Z}_Q} \sum_{S^{1:Q}} O(S^{1:Q}) \times \int dY q(S^1, Y) \cdots q(S^{Q-1}, Y) q_*(S^Q, Y), \quad (\text{A8})$$

where $Q = nk + 1$, with the Q th replica index being the teacher index. Throughout we will use $Q \leftrightarrow Q$ interchangeably, and $\mathcal{Z}_Q = \sum_{S^{1:Q}} \int dY q(S^1, Y) \cdots q(S^{Q-1}, Y) q_*(S^Q, Y)$.

For example, student's variance or the “precision” on come observable C_t can be written as

$$\delta C^2 = \mathbb{E}_{Y \sim p^*(Y)} \left[\mathbb{V}_{S \sim p(\cdot | Y)} [C(s_t)] \right] = \int dY p^*(Y) \sum_{S, S'} (C_t C_t - C_t C_t') \frac{p(S, Y) p(S', Y)}{p(Y)^2} = \lim_{Q \rightarrow 1} \frac{1}{\mathcal{Z}_Q} \sum_{S^{1:Q}} (C_t^1 C_t^1 - C_t^1 C_t^2) \times \int dY q(S^1, Y) \cdots q(S^{Q-1}, Y) q_*(S^Q, Y). \quad (\text{A9})$$

$$= \lim_{Q \rightarrow 1} \frac{1}{\mathcal{Z}_Q} \sum_{S^{1:Q}} (C_t^1 C_t^1 - C_t^1 C_t^2) \times \int dY q(S^1, Y) \cdots q(S^{Q-1}, Y) q_*(S^Q, Y). \quad (\text{A10})$$

The mean squared error or the “accuracy,” averaged over the joint distribution can also be written in above replica form as

$$\text{MSE}_B = \mathbb{E}_{S, Y, S^* \sim p(S, Y, S^*)} [(C_t - C_t^*)^2] = \sum_S \int dY \sum_{S^*} p(S, Y, S^*) (C(s_t) - C(s_t^*))^2 = \lim_{Q \rightarrow 1} \frac{1}{\mathcal{Z}_Q} \sum_{S^{1:Q}} (C_t^1 C_t^1 - 2C_t^1 C_t^Q + C_t^Q C_t^Q) \times \int dY q(S^1, Y) \cdots q(S^{Q-1}, Y) q_*(S^Q, Y). \quad (\text{A11})$$

2. Continuum limit in time

We can write (unnormalized) expectation values of observables as

$$\sum_S C_t \cdot p(S, Y) = \langle \tilde{\dagger} | C_t T_{1:t}(Y) | \pi \rangle \quad (\text{A13})$$

and the “evidence” as

$$p(Y) = \langle \tilde{\dagger} | T_{1:t}(Y) | \pi \rangle, \quad (\text{A14})$$

where $T_{1:t}(Y) := T_t(y_t) \cdots T_1(y_1)$,

$$T_t(y_t) = \begin{cases} P(y_t) \mathbf{u}_t, & t > 1, \\ P(y_t), & t = 1, \end{cases} \quad (\text{A15})$$

$|\tilde{\dagger}\rangle := \bigotimes_x |\tilde{\dagger}\rangle_x$, $|\tilde{\dagger}\rangle_x = |-1\rangle_x + |1\rangle_x$.

To go to the continuum limit, we slightly modify the six-vertex transition kernel to the following:

$$u_{x,x+1} \rightarrow \left(\frac{1 - \sigma_x^z \sigma_{x+1}^z}{2} \right) (4J\mathbf{b}(\sigma_x^+ \sigma_{x+1}^- + \sigma_x^- \sigma_{x+1}^+) + (1 - 4J\mathbf{b})) + \frac{1 + \sigma_x^z \sigma_{x+1}^z}{2}, \quad (\text{A16})$$

$$\approx \exp(2J[\vec{\sigma}_x \cdot \vec{\sigma}_{x+1} - 2]\mathbf{b}), \quad (\text{A17})$$

where $\mathbf{b}$ is the temporal lattice spacing. The above corresponds to setting a rate to the exchange. We can perform integration over each measurement $y_{x,t}$. We consider

$$\bar{P} = \int dy P^1(y) \cdots P^Q(y), \quad (\text{A18})$$

$$= \int dy \frac{1}{(2\pi\sigma^2)^{(Q-1)/2}} \frac{1}{(2\pi\sigma_*^2)^{1/2}} \exp \left\{ - \sum_a \frac{1}{2\sigma_a^2} (y - \epsilon^a \hat{Z}^a)^2 \right\}, \quad (\text{A19})$$

$$= \int dy \frac{1}{(2\pi\sigma^2)^{(Q-1)/2}} \frac{1}{(2\pi\sigma_*^2)^{1/2}} \exp \left\{ - \sum_a \frac{\sum_c \sigma_c^{-2}}{2} y + \sum_a \frac{\epsilon_a}{\sigma_a^2} \hat{Z}^a - \frac{1}{2} \sum_a \frac{\epsilon_a^2}{\sigma_a^2} (\hat{Z}^a)^2 \right\}, \quad (\text{A20})$$

$$= \int dy \frac{1}{(2\pi\sigma^2)^{(Q-1)/2}} \left(\frac{\sigma_*^{-2}}{\sum_c \sigma_c^{-2}} \right)^{1/2} \exp \left\{ \frac{1}{2 \sum_c \sigma_c^{-2}} \left(\sum_a \frac{\epsilon_a}{\sigma_a^2} \hat{Z}^a \right)^2 - \frac{1}{2} \sum_a \frac{\epsilon_a^2}{\sigma_a^2} (\hat{Z}^a)^2 \right\}. \quad (\text{A21})$$

Recall that we were allowed to replace the students' distributions with unnormalized ones. Since

$$\exp\{-(y - \epsilon \hat{Z})^2 / 2\sigma^2\} = \exp\{-y^2 / 2\sigma^2 + y\epsilon \hat{Z} / \sigma^2 - \epsilon^2 / 2\sigma^2\}, \quad (\text{A22})$$

and only $\exp\{y\epsilon \hat{Z} / \sigma^2\}$ depends on the state $\hat{Z}$, we could have ignored all factors independent of $\hat{Z}$ and just kept $\exp\{y\epsilon \hat{Z} / \sigma^2\}$, for example. In fact, we could have had any factor of form $\exp\{-\alpha y^2 / 2 + y\epsilon \hat{Z} / \sigma^2 - \beta / 2\}$ for each student replica. On the teacher replica, we can have add any constant terms, so $\exp\{-y^2 / 2\sigma_*^2 + y\epsilon_* \hat{Z}^Q / \sigma_* - \gamma / 2\}$.

In general, we are free to substitute

$$\bar{P} \rightarrow \int dy \left(\frac{\sigma_*^{-2}}{(Q-1)\alpha + \sigma_*^{-2}} \right)^{1/2} \exp \left\{ \frac{1}{2([Q-1]\alpha + \sigma_*^{-2})} \left(\sum_a \frac{\epsilon_a}{\sigma_a^2} \hat{Z}^a \right)^2 - \frac{1}{2} \left(\sum_{a=1}^{Q-1} \beta (\hat{Z}^a)^2 + \gamma (\hat{Z}^Q)^2 \right) \right\}, \quad (\text{A23})$$

$$= \int dy \left(\frac{\sigma_*^{-2}}{(Q-1)\alpha + \sigma_*^{-2}} \right)^{1/2} \exp \left\{ -\frac{1}{2} \sum_{ab} M_{ab} \hat{Z}^a \hat{Z}^b \right\}, \quad (\text{A24})$$

where

$$M = \begin{pmatrix} \beta & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & \beta & 0 \\ 0 & 0 & 0 & \gamma \end{pmatrix} - \frac{1}{\sigma_*^{-2} + (Q-1)\alpha} \begin{pmatrix} \epsilon / \sigma^2 \\ \vdots \\ \epsilon / \sigma^2 \\ \epsilon_* / \sigma_*^2 \end{pmatrix} (\epsilon / \sigma^2 \quad \dots \quad \epsilon / \sigma^2 \quad \epsilon_* / \sigma_*^2). \quad (\text{A25})$$

The natural choice of $\alpha = \sigma^{-2}$ and $\beta = \epsilon^2 / \sigma^2$, $\gamma = \epsilon_*^2 / \sigma_*^2$ retrieves the result above. The choice of $\beta = \epsilon_*^2 / \sigma_*^2$ and $\alpha = \sigma_*^2 \epsilon^2 / \sigma^4 \epsilon_*$, $\gamma = \epsilon_*^2 / \sigma_*^2$ turns M proportional to a projector.

However, we will choose the “natural” choice of $\alpha = \sigma^{-2}$ and $\beta = \epsilon^2/\sigma^2$, $\gamma = \epsilon_*^2/\sigma_*^2$ as Eq. (A18). Above we used that $\hat{Z}\hat{Z} = 1$ in order to argue that we can add any constant terms. However, below we will use the spin-coherent path integral formalism, which is valid for large spin, where $\hat{S}^z\hat{S}^z \propto 1$. Considering the model for general spin then interpolating to spin 1/2 corresponds to fixing β and γ as above. Second, if we are studying the model near the Bayes optimal line, we must respect the symmetry between the student and teacher replica at the Bayes optimal point. Choosing a different α corresponds to breaking symmetry and altering original measurement operator on the student replicas, which also introduces additional $\hat{Z}\hat{Z}$ terms.

We renormalize $M_{ab} \rightarrow \mathbf{b}M_{ab}$, where $\mathbf{b}$ is the temporal lattice spacing. This can be also seen as renormalizing the noise and signal $\sigma^2 \rightarrow \mathbf{b}\sigma^2$ and $\epsilon \rightarrow \mathbf{b}\epsilon$.

For small time steps, all operators commute so we can combine the even and odd gates into gates over all sites $\tilde{U}_t$ with half strength as

$$\bar{P}U_t^{\otimes Q}\bar{P}U_{t-1}^{\otimes Q} \propto \exp \left\{ -\mathbf{b} \left[- \sum_{A=X,Y,Z} \sum_{x,a} 2J\hat{A}_x^a \cdot \hat{A}_{x+1}^a + \sum_{xab} M_{ab}\hat{Z}_x^a\hat{Z}_x^b \right] \right\}, \quad (\text{A26})$$

$$\approx \prod_{t'=t,t-1} \exp \left\{ -\mathbf{b} \left[-J \sum_{A=X,Y,Z} \sum_{x,a} \hat{A}_x^a\hat{A}_{x+1}^a + \frac{1}{2} \sum_{xab} M_{ab}\hat{Z}_x^a\hat{Z}_x^b \right] \right\}, \quad (\text{A27})$$

$$\approx \bar{P}\tilde{U}_t\bar{P}\tilde{U}_{t-1}. \quad (\text{A28})$$

We define continuous time $\tilde{t} = \mathbf{b}t$. We send $t_f \rightarrow \infty$ and $\mathbf{b} \rightarrow 0$ while keeping $\tilde{t}_f = t_f/\mathbf{b}$ constant. Doing so then removing the tildes,

$$\int dY \bigotimes_{a=1}^Q T_{1:t_f}^{(a)} \propto \exp \left(- \int_0^{t_f} dt H_Q \right), \quad (\text{A29})$$

where

$$H_Q = -J \sum_{A=X,Y,Z} \sum_{ar} \hat{A}_x^a\hat{A}_{x+1}^a + \frac{1}{2} \sum_{abr} \hat{Z}_x^a M_{ab} \hat{Z}_x^b. \quad (\text{A30})$$

As we are interested in boundary observables at t_f , we will shift time, such that starting time is at $-t_f$ and end time at 0. Therefore, the partition function can be written as

$$\mathcal{Z}_Q = \langle \# | \exp \left(- \int_{-t_f}^0 dt H_Q \right) | \mathbf{\Pi} \rangle, \quad (\text{A31})$$

and variance of an observable can be written as

$$\delta C^2 = \lim_{Q \rightarrow 1} \frac{1}{\mathcal{Z}_Q} \langle \# | (C_0^1 C_0^1 - C_0^1 C_0^2) \exp \left(- \int_{-t_f}^0 dt H_Q \right) | \mathbf{\Pi} \rangle, \quad (\text{A32})$$

where superscript above C denotes the replica index, $|\# \rangle := |+\rangle^{\otimes Q}$, and $|\mathbf{\Pi} \rangle := |\pi \rangle^{\otimes Q}$. Similarly, the mean-squared error can be written as

$$\text{MSE} = \lim_{Q \rightarrow 1} \frac{1}{\mathcal{Z}_Q} \langle \# | (C_0^1 C_0^1 - 2C_0^1 C_0^* + C_0^* C_0^*) \exp \left(- \int_{-t_f}^0 dt H_Q \right) | \mathbf{\Pi} \rangle. \quad (\text{A33})$$

3. Coherent state path integral

Doing the coherent state path integral involves replacing $(\hat{X}, \hat{Y}, \hat{Z}) \rightarrow (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$ in the Hamiltonian and adding the “kinetic” term $\frac{i}{2a} \sum_a \theta^a \partial_t \phi^a$. For simplicity, we will look at half filling [$\cos(\theta_0) = 0$, $\sin(\theta_0) = 1$]. We define continuous space $\tilde{x} := \mathbf{a}x$, expand $\theta = \theta_0 + \delta\theta$ around $\theta_0 = \pi/2$ and keeping the lowest order terms, then removing the tildes, we have

$$\mathcal{L} = \frac{i}{2a} \sum_{a=1}^Q \delta\theta^a \partial_t \phi^a + \frac{Ja}{2} \sum_{a=1}^Q ((\partial_x \delta\theta^a)^2 + (\partial_x \phi^a)^2) + \frac{1}{2a} \sum_{a,b=1}^Q \delta\theta^a M_{ab} \delta\theta^b. \quad (\text{A34})$$

Then, we have the identification that $\hat{Z} \approx \cos \theta \approx -\delta\theta$. To keep track of the operators, we will add a source term $\sum_a h^a(x, t) \delta\theta^a(x, t)$ to the Lagrangian density and send $h^a \rightarrow 0$. We will not keep track of terms quadratic in h .

4. Diagonalizing M

We can diagonalize M as follows. For notational convenience, let's denote $K := Q - 1$. We can write M as

$$M = \begin{pmatrix} \frac{\epsilon^2}{\sigma^2} \mathbb{1}_K & \vec{0}_K \\ \vec{0}_K^T & \frac{\epsilon_*^2}{\sigma_*^2} \end{pmatrix} - \frac{1}{2(K\sigma^{-2} + \sigma_*^{-2})} \begin{pmatrix} \frac{\epsilon^2}{\sigma^4} \vec{1}_K \vec{1}_K^T & \frac{\epsilon\epsilon_*}{\sigma^2\sigma_*^2} \vec{1}_K \\ \frac{\epsilon\epsilon_*}{\sigma^2\sigma_*^2} \vec{1}_K^T & \frac{\epsilon_*^2}{\sigma_*^4} \end{pmatrix}. \quad (\text{A35})$$

On an ansatz of form $(v \ w)^T$, one can see that the solution only depends on $\vec{1}_K^T v$ and w . In the case $\vec{1}_K^T v = 0$, ex. nonzero wave vector Fourier modes on the K student modes $\hat{v}_1, \dots, \hat{v}_{Q-2}$, one finds that the eigenvalue is ϵ^2/σ^2 . In the other case, we have a 2×2 eigenvalue of the following form:

$$\begin{pmatrix} \frac{\epsilon^2}{\sigma^2} - \frac{K\epsilon^2}{\sigma^4(\frac{K}{\sigma^2} + \frac{1}{\sigma_*^2})} & -\frac{K\epsilon\epsilon_*}{(\sigma^2\sigma_*^2)(\frac{K}{\sigma^2} + \frac{1}{\sigma_*^2})} \\ -\frac{\epsilon\epsilon_*}{(\sigma^2\sigma_*^2)(\frac{K}{\sigma^2} + \frac{1}{\sigma_*^2})} & \frac{\epsilon_*^2}{\sigma_*^2} - \frac{\epsilon_*^2}{\sigma_*^4(\frac{K}{\sigma^2} + \frac{1}{\sigma_*^2})} \end{pmatrix} \begin{pmatrix} \vec{1}_K^T v \\ w \end{pmatrix} = \lambda \begin{pmatrix} \vec{1}_K^T v \\ w \end{pmatrix}, \quad (\text{A36})$$

where eigenvalues and eigenvectors are given by

$$\lambda_0 = 0, \quad v_0^T = \frac{1}{\sqrt{K\epsilon^{-2} + \epsilon_*^{-2}}} (\epsilon^{-1} \dots \epsilon^{-1} \epsilon_*^{-1})^T, \quad (\text{A37})$$

$$\lambda_{Q-1} = \frac{K\epsilon_*^2 + \epsilon^2}{K\sigma_*^2 + \sigma^2}, \quad v_{Q-1}^T = \frac{1}{\sqrt{K[\epsilon^2 + K\epsilon_*^2]}} (\epsilon, \dots, \epsilon, -K\epsilon_*)^T, \quad (\text{A38})$$

So, we have $V^T M V = \Lambda$, where $V = [\hat{v}_0 \ \hat{v}_1 \ \dots \ \hat{v}_{Q-2} \ \hat{v}_{Q-1}]$ such that $\hat{V}_{\alpha\alpha} = \hat{v}_\alpha(a)$ and

$$\Lambda = \frac{\epsilon^2}{\sigma^2} \text{diag} \left[0 \ 1 \dots 1 \ \frac{1 + (\epsilon_*/\epsilon)^2 K}{1 + (\sigma_*/\sigma)^2 K} \right]^T.$$

Let $\delta\theta_\alpha = \sum_{a=1}^Q [\hat{V}^\dagger]_{\alpha a} \delta\theta^a = \hat{v}_\alpha^*(a) \delta\theta^a$. We similarly define the transformed fields for ϕ . As $\hat{V}$ is unitary, the already-diagonal terms stay diagonal. Then, the Lagrangian density is (in diagonal basis)

$$\mathcal{L} = \frac{i}{2a} \sum_{\alpha=0}^{Q-1} \delta\theta_\alpha \partial_t \phi_\alpha + \frac{Ja}{2} \sum_{\alpha=0}^{Q-1} [(\partial_x \delta\theta_\alpha)^2 + (\partial_x \phi_\alpha)^2] + \frac{\epsilon^2}{2a\sigma^2} \left(\sum_{\alpha=1}^{Q-2} \delta\theta_\alpha^2 + \frac{1 + (\epsilon_*/\epsilon)^2 K}{1 + (\sigma_*/\sigma)^2 K} \delta\theta_{Q-1}^2 \right) \quad (\text{A39})$$

$$+ \sum_{\alpha=0}^{Q-1} h_\alpha \delta\theta_\alpha. \quad (\text{A40})$$

We use lower indices for the α basis and upper indices for a basis throughout. Note that as long as we don't attempt to *integrate* over ϕ , we can write any linear combination of them such as above, as long as we transform back to the original basis when we consider its compactness.

5. Integrating out massive modes of θ_α

As we are treating $\delta\theta_\alpha$'s to be Gaussian fluctuations we can integrate over them as long as they are massive. Integrating over the $\alpha = 1, \dots, Q-1$ modes, we get

$$\mathcal{L} = \frac{i}{2a} \delta\theta_0 (\partial_t \phi_0) + \frac{\sigma^2}{8a\epsilon^2} \left(\sum_{\alpha=1}^{Q-2} (\partial_t \phi_\alpha)^2 + \frac{1 + (\sigma_*/\sigma)^2 K}{1 + (\epsilon_*/\epsilon)^2 K} (\partial_t \phi_{Q-1})^2 \right) + \frac{Ja}{2} \left[(\partial_x \delta\theta_0)^2 + \sum_{\alpha=0}^{Q-1} (\partial_x \phi_\alpha)^2 \right] \quad (\text{A41})$$

$$+ h_0 \delta\theta_0 - \frac{ia\sigma^2}{2\epsilon^2} \left[\sum_{\alpha=1}^{Q-2} h_\alpha (\partial_t \phi_\alpha) + \frac{1 + (\sigma_*/\sigma)^2 K}{1 + (\epsilon_*/\epsilon)^2 K} h_{Q-1} (\partial_t \phi_{Q-1}) \right]. \quad (\text{A42})$$

We now rescale space and time to make the action as isotropic as possible. We transform $\partial_t \rightarrow \sqrt{v} \partial_t$ and $\partial_x \rightarrow (1/\sqrt{v}) \partial_x$ with $v = 2a(\epsilon/\sigma)\sqrt{J}$. Defining $\rho_s = \sqrt{J}/2(\epsilon/\sigma)$, we have

$$\mathcal{L} = \frac{i\sqrt{v}}{2a} \delta\theta_0 (\partial_t \phi_0) + \frac{\rho_s}{2} (\partial_x \phi_0)^2 + \frac{\rho_s}{2} (\partial_x \delta\theta_0)^2 + \frac{\rho_s}{2} \sum_{\mu=t,x} \sum_{\alpha=1}^{Q-2} (\partial_\mu \phi_\alpha)^2 + \frac{\rho_s}{2} \frac{1 + (\sigma_*/\sigma)^2 K}{1 + (\epsilon_*/\epsilon)^2 K} (\partial_t \phi_{Q-1})^2 + \frac{\rho_s}{2} (\partial_x \phi_{Q-1})^2 \quad (\text{A43})$$

$$+ h_0 \delta\theta_0 - \frac{ia\sigma^2 \sqrt{v}}{2\epsilon^2} \left[\sum_{\alpha=1}^{Q-2} h_\alpha (\partial_t \phi_\alpha) + \frac{1 + (\sigma_*/\sigma)^2 K}{1 + (\epsilon_*/\epsilon)^2 K} h_{Q-1} (\partial_t \phi_{Q-1}) \right]. \quad (\text{A44})$$

6. Duality transformation

In preparation to the duality transformation, we will define $j_0^t := (\sqrt{v}/2a)\theta_0$. Defining $D^2 := v/4a^2$, we have

$$\mathcal{L} = i j_0^t (\partial_t \phi_0) + \frac{\rho_s}{2D^2} (\partial_x j_0^t)^2 + \frac{\rho_s}{2} (\partial_x \phi_0)^2 + \frac{\rho_s}{2} \sum_{\mu=t,x} \sum_{\alpha=1}^{Q-1} (\partial_\mu \phi_\alpha)^2 + \frac{\rho_s}{2} \frac{1 + (\sigma_*/\sigma)^2 K}{1 + (\epsilon_*/\epsilon)^2 K} (\partial_t \phi_{Q-1})^2 + \frac{\rho_s}{2} (\partial_x \phi_{Q-1})^2 \quad (\text{A45})$$

$$+ \frac{2a}{\sqrt{v}} h_0 j_0^t - \frac{ia\sigma^2 \sqrt{v}}{2\epsilon^2} \left[\sum_{\alpha=1}^{Q-2} h_\alpha (\partial_t \phi_\alpha) + \frac{1 + (\sigma_*/\sigma)^2 K}{1 + (\epsilon_*/\epsilon)^2 K} h_{Q-1} (\partial_t \phi_{Q-1}) \right]. \quad (\text{A46})$$

To disentangle the $(\partial_\mu \phi_\alpha)^2$ terms for $\alpha = 1, \dots, Q-1$, we will introduce the Hubbard-Stratonovich currents j_α^μ . To disentangle the $(\partial_x \phi_0)^2$ term, we introduce j_0^x . Then, we get

$$\mathcal{L} = i \sum_{\mu=t,x} \sum_{\alpha=0}^{Q-1} j_\alpha^\mu (\partial_\mu \phi_\alpha) + \frac{1}{2\rho_s} \sum_{\mu=t,x} \sum_{\alpha=1}^{Q-2} (j_\alpha^\mu)^2 + \frac{1}{2\rho_s} (j_{Q-1}^x)^2 + \frac{1}{2\rho_s} \frac{1 + (\epsilon_*/\epsilon)^2 K}{1 + (\sigma_*/\sigma)^2 K} (j_{Q-1}^t)^2 + \frac{1}{2\rho_s} (j_0^x)^2 + \frac{D^2}{2\rho_s} (\partial_x j_0^t)^2 \quad (\text{A47})$$

$$+ \frac{2a}{\sqrt{v}} h_0 j_0^t - \frac{ia\sigma^2 \sqrt{v}}{2\epsilon^2} \left[\sum_{\alpha=1}^{Q-2} h_\alpha (\partial_t \phi_\alpha) + \frac{1 + (\sigma_*/\sigma)^2 K}{1 + (\epsilon_*/\epsilon)^2 K} h_{Q-1} (\partial_t \phi_{Q-1}) \right]. \quad (\text{A48})$$

Now, we can separate $\phi^a = \phi_s^a + \phi_v^a$ for smooth and vortex parts, respectively, and therefore $\phi_\alpha = \phi_\alpha^s + \phi_\alpha^v$, being careful that the vortex part is defined in the original a basis. Integrating by parts to move the derivative to j_μ^a , then integrating over ϕ_s^a , we impose divergence-free condition on the current, $\sum_\mu \partial_\mu j_\mu^a = 0 \forall a$ without the source term.

With the source term, we can use the ansatz

$$j_\alpha^t = \frac{1}{2\pi} \partial_x \chi_\alpha + \begin{cases} 0, & \alpha = 0 \\ \frac{a\sigma^2 \sqrt{v}}{2\epsilon^2} h_\alpha, & \alpha = 1, \dots, Q-2 \\ \frac{1 + (\sigma_*/\sigma)^2 K}{1 + (\epsilon_*/\epsilon)^2 K} \frac{a\sigma^2 \sqrt{v}}{2\epsilon^2} h_{Q-1}, & \alpha = Q-1, \end{cases} \quad (\text{A49})$$

and $j_\alpha^x = (-1/2\pi) \partial_t \chi_\alpha$. Substituting, we find

$$\mathcal{L} = -i \sum_{\mu=t,x} \sum_{a=1}^Q \chi^a \frac{(-1)^\mu}{2\pi} (\partial_\mu \partial_\mu \phi_v^a) \quad (\text{A50})$$

$$+ \frac{1}{8\pi^2 \rho_s} \left[(\partial_t \chi_0)^2 + D^2 (\partial_x^2 \chi_0)^2 + \sum_{\mu=t,x} \sum_{\alpha=1}^{Q-2} (\partial_\mu \chi_\alpha)^2 + \frac{1 + (\epsilon_*/\epsilon)^2 K}{1 + (\sigma_*/\sigma)^2 K} (\partial_x \chi_{Q-1})^2 + (\partial_t \chi_{Q-1})^2 \right] \quad (\text{A51})$$

$$+ \frac{a}{\pi \sqrt{v}} \left[h_0 (\partial_x \chi_0) + \sum_{\alpha=1}^{Q-2} h_\alpha (\partial_x \chi_\alpha) + \left(\frac{1 + (\epsilon_*/\epsilon)^2 K}{1 + (\sigma_*/\sigma)^2 K} \right)^2 h_{Q-1} \chi_{Q-1} \right]. \quad (\text{A52})$$

Noticing that $\sum_\mu \frac{(-1)^\mu}{2\pi} (\partial_\mu \partial_\mu \phi_v^a) = \rho_v^a$ is the vortex density for the a th replica, and partially transforming back to the original replica basis a ,

$$\mathcal{L} = -i \sum_{a=1}^Q \chi^a \rho_v^a + \frac{1}{8\pi^2 \rho_s} \sum_{\mu=t,x} \sum_{a,b=1}^Q (\partial_\mu \chi^a) \Phi_{ab} (\partial_\mu \chi^b) + \frac{1}{8\pi^2 \rho_s} [(\partial_t \chi_0)^2 + D^2 (\partial_x^2 \chi_0)^2] + \frac{1}{8\pi^2 \rho_s} \left[(\partial_t \chi_{Q-1})^2 + \frac{1 + (\epsilon_*/\epsilon)^2 K}{1 + (\sigma_*/\sigma)^2 K} (\partial_x \chi_{Q-1})^2 \right], \quad (\text{A53})$$

where χ^a are noncompact variables, $\rho_s \sim \sqrt{J}/(\epsilon/\sigma)$, $\rho_0 \sim \epsilon^2$, $\Phi = \mathbb{1}_Q - \hat{v}_0 \hat{v}_0^T - \hat{v}_{Q-1} \hat{v}_{Q-1}^T$, and

$$\hat{v}_0 = \frac{1}{\sqrt{K\epsilon^{-2} + \epsilon_*^{-2}}} \begin{bmatrix} \epsilon^{-1} \\ \vdots \\ \epsilon^{-1} \\ \epsilon_*^{-1} \end{bmatrix} \Rightarrow \chi_0 = \frac{1}{\sqrt{K\epsilon^{-2} + \epsilon_*^{-2}}} \sum_{a=1}^Q \epsilon_a^{-1} \chi^a, \quad (\text{A54})$$

$$\hat{v}_{Q-1} = \frac{1}{\sqrt{K[\epsilon^2 + K\epsilon_*^2]}} \begin{bmatrix} \epsilon \\ \vdots \\ \epsilon \\ -K\epsilon_* \end{bmatrix} \Rightarrow \chi_{Q-1} = \frac{1}{\sqrt{K[\epsilon^2 + K\epsilon_*^2]}} \left(\epsilon \sum_{a=1}^{Q-1} \chi^a - \epsilon_* K \chi^Q \right). \quad (\text{A55})$$

7. Renormalization group analysis

We have the partition function

$$\mathcal{Z}_Q = \int D[\chi^a]_{a=1}^Q D[\rho^a]_{a=1}^Q \exp \left\{ - \int_{t,x} \mathcal{L} \right\}, \quad (\text{A56})$$

where $D[\rho^a] = \sum_{N_a=0}^{\infty} \sum_{i_a=1}^{N_a} \prod_{i_a} (\sum_{m_a=-\infty}^{\infty} \int_{x_{i_a}})$ for each replica index a and $\rho_v^a = \sum_{i_a} m_{i_a} \delta(r - r_{i_a})$.

We can see that the action is invariant under:

- (1) $\vartheta^a \rightarrow \vartheta^a + 2\pi \forall a$,
- (2) Replica permutation, $a \leftrightarrow b \forall a, b \neq Q$, i.e., for student replicas only.
- (3) Symmetric under $\chi^a \rightarrow -\chi^a$.

We will try the “Landau approach” to determining relevant terms. We will try adding terms to the Lagrangian $\mathcal{L} = \mathcal{L}_G + \mathcal{V}$ that respects the symmetry of the problem. The minimal such terms are as follows:

$$\mathcal{V} = -\lambda_0 \mathcal{S}_{Q-1} \cos \left(\sum_a n^a \chi^a \right) = -\lambda_0 \frac{1}{(Q-1)!} \sum_{\tau \in S_{Q-1}} \cos \left(\sum_a n^{\tau(a)} \chi^a \right), \quad (\text{A57})$$

with the understanding that $\tau(Q) = Q$, and $\{n^a\}_{a=1}^Q$ are integers.

Then, we will apply renormalization and see if which terms are relevant under renormalization group.

We will go through the following recipe.

- (1) Integrate out the fast moving + modes of wave vector $k \in (\Lambda/\zeta, \Lambda)$, where Λ is the UV cutoff. Since $S_G = S_G^+ + S_G^-$,

$$Z = \int D[\chi^-] e^{-\int_{tx} \mathcal{L}_G^-} \left[\int D[\chi^+] e^{-\int_{tx} (\mathcal{L}_G^+ + \mathcal{V})} \right] = \int D[\chi^-] e^{-\int_{tx} \mathcal{L}_G^-} \langle e^{-\int_{tx} \mathcal{V}} \rangle_{G^+} =: \int D[\chi^-] e^{-\int_{tx} (\mathcal{L}_G^- + \Delta \mathcal{L}^-)}, \quad (\text{A58})$$

where we equated the expectation value as additional terms in the Lagrangian.

- (2) Rescale (t, x) , such that $(t', x') = \frac{1}{\zeta} (t, x)$.
- (3) If need be, rescale χ to keep the quadratic term equal.

The interaction term integrated over the fast modes is

$$\left\langle \exp \left\{ -\lambda \mathcal{S}_{Q-1} \int_{tx} \cos \left(\sum_a n^a \vartheta^a(r) \right) \right\} \right\rangle_{G^+} \approx 1 + \frac{\lambda}{2} \mathcal{S}_{Q-1} \int_{tx} \sum_{\sigma=\pm 1} e^{i\sigma \sum_a n^a \vartheta^a} \langle e^{i\sigma \sum_a n^a \vartheta^a} \rangle_{G^+}. \quad (\text{A59})$$

Noting that $\vartheta_+^a = \sum_{\alpha} U_{a\alpha} \vartheta_{\alpha}^+$ and using that $\langle e^{ib\vartheta_{\alpha}(r)} \rangle = e^{-\frac{b^2}{2} \langle \vartheta_{\alpha}(r) \vartheta_{\alpha}(r) \rangle}$,

$$\langle e^{i\sigma \sum_a n^a \vartheta_+^a} \rangle_{G^+} = \prod_{\alpha} \langle e^{i\sigma \sum_a n^a U_{a\alpha} \vartheta_{\alpha}^+} \rangle_{G^+} = \prod_{\alpha} \left\langle \exp \left\{ -\frac{1}{2} \left(\sum_a n^a U_{a\alpha} \right)^2 \langle \vartheta_{\alpha}^a(r) \vartheta_{\alpha}^a(r) \rangle \right\} \right\rangle_{G^+}. \quad (\text{A60})$$

We have

$$\langle \vartheta_{\alpha}^+(r) \vartheta_{\alpha}^+(r) \rangle_{G^+} = \int_{\Lambda/\zeta}^{\Lambda} d^2 k G_{\alpha}(k) = \begin{cases} 4\pi^2 \rho_s \sqrt{4D} \bar{f} \zeta, & \alpha = 0, \\ 2\pi \rho_s \ln \zeta, & \alpha = 1, \dots, Q-2, \\ \sqrt{\frac{1+K(\sigma_*/\sigma)^2}{1+K(\epsilon_*/\epsilon)^2}} 2\pi \rho_s \ln \zeta, & \alpha = Q-1. \end{cases} \quad (\text{A61})$$

Therefore

$$\begin{aligned} \langle e^{i\sigma \sum_a n^a \vartheta_+^a} \rangle_{G^+} &= \exp \left\{ -\frac{1}{2} \left(\sum_a n^a V_{a0} \right)^2 4\pi^2 \rho_s \sqrt{4D} \bar{f} \zeta \right\} \exp \left\{ -\frac{1}{2} \sum_{\alpha=1}^{Q-2} \left(\sum_a n^a V_{a\alpha} \right)^2 2\pi \rho_s \ln \zeta \right\} \\ &\times \exp \left\{ -\frac{1}{2} \left(\sum_a n^a V_{a,Q-1} \right)^2 \sqrt{\frac{1+K(\sigma_*/\sigma)^2}{1+K(\epsilon_*/\epsilon)^2}} 2\pi \rho_s \ln \zeta \right\} \end{aligned} \quad (\text{A62})$$

$$= \exp \left\{ -\frac{1}{2[K\epsilon^{-2} + \epsilon_*^{-2}]} \left(\sum_a n^a (\epsilon^a)^{-1} \right)^2 4\pi^2 \rho_s \sqrt{4Df\zeta} \right\} \exp \left\{ -\frac{1}{2} \sum_{ab} n^a n^b \Phi_{ab} 2\pi \rho_s \ln \zeta \right\} \quad (\text{A63})$$

$$\times \exp \left\{ -\frac{1}{2K[1 + K(\epsilon_*/\epsilon)^2]} \left(\sum_{a=1}^{Q-1} n^a - K(\epsilon_*/\epsilon) n^Q \right)^2 \sqrt{\frac{1 + K(\sigma_*/\sigma)^2}{1 + K(\epsilon_*/\epsilon)^2}} 2\pi \rho_s \ln \zeta \right\}. \quad (\text{A64})$$

Since the expectation value does not depend on σ , we see that

$$\left\langle \exp \left\{ -\int_r \mathcal{V} \right\} \right\rangle_{G+} \approx 1 + \frac{\lambda \langle e^{i \sum_a n^a \vartheta_+^a} \rangle_{G+}}{2} \mathcal{S}_{Q-1} \int_r \sum_a \sum_{\sigma=\pm 1} e^{i\sigma \vartheta_-^a} \quad (\text{A65})$$

$$\approx \exp \left\{ \lambda \langle e^{i \sum_a n^a \vartheta_+^a} \rangle_{G+} \mathcal{S}_{Q-1} \int_r \cos \left(\sum_a n^a \vartheta_-^a \right) \right\}. \quad (\text{A66})$$

Rescaling space-time, we get a factor of ζ^2 and so the coupling constant flows as

$$\lambda \rightarrow \lambda \zeta^2 \langle e^{i \sum_a n^a \vartheta_+^a} \rangle_{G+}. \quad (\text{A67})$$

Looking at Eq. (A62), we see that in order to prevent the operator from being exponentially irrelevant, we have a hard constraint that we require $\sum_a n^a \epsilon_a^{-1} = 0$. Away from commensurate points, this forces the teacher vorticity to be zero, $n^Q = 0$. Then, the minimal relevant configuration is the inter-replica vortex configuration $(1, -1, 0, \dots, 0, 0)$, such that

$$\langle e^{i\sigma \sum_a n^a \vartheta_+^a} \rangle_{G+} = \exp \{-2\pi \rho_s (\ln \zeta)\}. \quad (\text{A68})$$

Let us consider the best case scenario for commensurate points. For $r_\epsilon > 1$, we can consider $\epsilon^* = 2\epsilon$. In this case, we can consider the configuration $(1, 0, \dots, 0, -2)$. Here, we have

$$\langle e^{i\sigma \sum_a n^a \vartheta_+^a} \rangle_{G+} = \exp \left\{ -\frac{1}{2} 2\pi \rho_s (\ln \zeta) (-2) (\Phi_{1,Q} + \Phi_{Q,1}) \right\} \quad (\text{A69})$$

$$\times \exp \left\{ -\frac{1}{2K[1 + K(\epsilon_*/\epsilon)^2]} (1 + 2K(\epsilon_*/\epsilon))^2 \sqrt{\frac{1 + K(\sigma_*/\sigma)^2}{1 + K(\epsilon_*/\epsilon)^2}} 2\pi \rho_s \ln \zeta \right\}. \quad (\text{A70})$$

For $r_\epsilon < 1$, we can consider $\epsilon^* = \frac{1}{2}\epsilon$. In this case, we can consider the configuration $(2, 0, \dots, 0, -1)$. Then, we have

$$\langle e^{i\sigma \sum_a n^a \vartheta_+^a} \rangle_{G+} = \exp \left\{ -\frac{1}{2} 2\pi \rho_s (\ln \zeta) (-2) (\Phi_{1,Q} + \Phi_{Q,1}) \right\} \quad (\text{A71})$$

$$\times \exp \left\{ -\frac{1}{2K[1 + K(\epsilon_*/\epsilon)^2]} (2 - K(\epsilon_*/\epsilon))^2 \sqrt{\frac{1 + K(\sigma_*/\sigma)^2}{1 + K(\epsilon_*/\epsilon)^2}} 2\pi \rho_s \ln \zeta \right\}. \quad (\text{A72})$$

Therefore even at the commensurate points, the minimal relevant configuration is the inter-replica vortex configuration $(1, -1, 0, \dots, 0, 0)$.

8. Observables in the charge-fuzzy phase

In the charge-fuzzy phase, where the vortex terms are irrelevant, we can calculate the expectation values using the Gaussian theory.

As we are interested in the quantities at the end time 0 (with starting time $-t_f$), we need consider the end boundary conditions. This is given by $\theta^a(x, 0) = \pi/2 \forall a, x$, which corresponds to $\delta\theta^a(x, 0) = 0$, $\partial_x \phi^a(x, 0) = 0$ and $\partial_t \chi^a(x, 0) = 0$. Using the method of images and solving for the Green's functions, one finds an overall factor of 2 compared to bulk Green's function. The $(Q-1)$ th mode has the same boundary Green's function due to the log multiply rule, so

$$\langle \chi_\alpha(x, 0) \chi_\alpha(x', 0) \rangle_G = \begin{cases} -2 \times 2\pi^2 \rho_s |x - x'| + \dots, & \alpha = 0, \\ -2 \times 2\pi \rho_s \ln |x - x'| + \dots, & \alpha = 1, \dots, Q-1. \end{cases} \quad (\text{A73})$$

Looking at the source term, we can deduce that $\hat{Z}_\alpha \approx -\frac{a}{\sqrt{v\pi}} \partial_x \chi_\alpha$. Recall that we rescaled space, with $\sqrt{v} x_{\text{iso.}} = x$, and that the relations above are for $x_{\text{iso.}}$. In the nonisotropic coordinates, we have $\hat{Z}_\alpha \approx -\frac{a}{\pi} \partial_x \chi_\alpha$. Using the diagonalization transformation

$Z^a = \sum_{\alpha} V_{a\alpha} Z_{\alpha}$, we have

$$\langle \hat{Z}_x^a \hat{Z}_0^b \rangle_G = -4\pi \rho_s \Psi_{ab} \left(\frac{a}{x} \right)^2, \quad (\text{A74})$$

where $\Psi = \mathbb{1}_Q - \hat{v}_0 \hat{v}_0^T$ is the projector out of the zero mode. Then, we have

$$\mathbb{E}_{Y, S^*} [\langle s_x s_0 \rangle_c] = \lim_{Q \rightarrow 1} [\langle \hat{Z}_x^1 \hat{Z}_0^1 \rangle_G - \langle \hat{Z}_x^1 \hat{Z}_0^2 \rangle_G] = -4\pi \rho_s \left(\frac{a}{x} \right)^2, \quad (\text{A75})$$

$$\mathbb{E}_{Y, S^*} [\langle s_x s_0 \rangle] = \lim_{Q \rightarrow 1} [\langle \hat{Z}_x^1 \hat{Z}_0^1 \rangle_G] = 4\pi \rho_s (r^2 - 1) \left(\frac{a}{x} \right)^2, \quad (\text{A76})$$

$$\mathbb{E}_{Y, S^*} [\langle s_x \rangle s_0^*] = \lim_{Q \rightarrow 1} [\langle \hat{Z}_x^1 \hat{Z}_0^* \rangle_G] = 4\pi \rho_s r \left(\frac{a}{x} \right)^2. \quad (\text{A77})$$

APPENDIX B: PERTURBATIVE ANALYSIS OF THE PLANTED SSEP

We would like to calculate the various correlators at the limit of small ϵ while keeping $r_{\epsilon} := \epsilon_*/\epsilon$ in the order of 1. In tensor network/dirac notation, posterior distribution can be written as

$$p_s(s_{t_f} | y_{1:t_f}) = \frac{\langle s_{t_f} | M(y_{t_f}) \mathbf{u} \cdot \cdot M(y_1) \mathbf{u} | +' \rangle}{\langle \tilde{+} | M(y_{t_f}) \mathbf{u} \cdot \cdot M(y_1) \mathbf{u} | +' \rangle}, \quad (\text{B1})$$

where $|\tilde{+}\rangle := \otimes_x |\tilde{+}\rangle_x$, $|\tilde{+}\rangle_x = |-1\rangle_x + |1\rangle_x$, $|+\rangle := \otimes_x |+\rangle_x$, $|+\rangle_x = \frac{1}{2}(|-1\rangle_x + |1\rangle_x)$, and the measurement operator will be expanded as

$$M(y_{t_f}) = \otimes_x \exp(\epsilon y_{x,t} Z_x) \approx 1 + \epsilon y_{x,t} Z_x + \frac{1}{2} \epsilon^2 y_{x,t}^2 + \frac{1}{6} \epsilon^3 y_{x,t}^3 Z_x + \mathcal{O}(\epsilon^4). \quad (\text{B2})$$

Similarly, the joint distribution of measurement and true/teacher trajectories can also be written as

$$p_*(y_{1:t_f}, s_{t_f}^*) = e^{-L_{t_f} r_{\epsilon}^2 \epsilon^2 / 2} \langle s_{t_f} | M_{t_f}^*(y_{t_f}) \mathbf{u} \cdot \cdot M_1^*(y_1) \mathbf{u} | +' \rangle \pi(y_{1:t_f}), \quad (\text{B3})$$

where $\pi(y_{1:t_f})$ denotes the Gaussian measure and M^* denotes M but with teacher parameters $\epsilon \rightarrow \epsilon_* = r_{\epsilon} \epsilon$.

We will use the fact that any state $\propto |+\rangle$ is an invariant state under the SSEP, that is, $\mathbf{u} |+\rangle = |+\rangle$, and that the choice of half-filling initial condition has vanishing disconnected correlators, $\langle + | Z_x | + \rangle = \langle + | Z_x Z_0 | + \rangle = \langle + | Z_{x_1} \cdots Z_{x_n} | + \rangle = 0$.

We have

$$M(y_{t_f}) \mathbf{u} \cdot \cdot M(y_1) \mathbf{u} | +' \rangle = \left\{ 1 + \underbrace{\epsilon \sum_{x,t} y_{x,t} Z_{x,t}}_{:= (1)} + \epsilon^2 \left(\underbrace{\frac{1}{2} \sum_{x,t} y_{x,t}^2}_{:= (2)} + \underbrace{\sum_{(x,t) \neq (x',t')} y_{x,t} y_{x',t'} Z_{x,t} Z_{x',t'}}_{:= (11)} \right) \right\} \mathbf{u} \cdot \cdot \mathbf{u} | +' \rangle, \quad (\text{B4})$$

where $Z_{x,t}$ denotes inserting the Z_r operator in the middle of the string of $\mathbf{u}$'s at time t , and we introduced the bracket notation, where the each number n in the bracket denotes including the n th order term of M . Continuing the series, we get

$$\begin{aligned} M(y_{t_f}) \mathbf{u} \cdot \cdot M(y_1) \mathbf{u} | +' \rangle &= (1 + \epsilon(1) + \epsilon^2((2) + (11)) + \epsilon^3((3) + (21) + (111)) + \epsilon^4((4) + (31) + (22) + (211) + (1111)) \\ &\quad + \epsilon^5((5) + (41) + (32) + (311) + (22) + (211) + (11111)) + \mathcal{O}(\epsilon^6)) \mathbf{u} \cdot \cdot \mathbf{u} | +' \rangle. \end{aligned} \quad (\text{B5})$$

1. Disconnected correlator

The disconnected correlator can be written as

$$\mathbb{E} [\langle s_{x,t_f} s_{0,t_f} \rangle] = e^{-L_{t_f} r_{\epsilon}^2 \epsilon^2 / 2} \left[\frac{AB}{C} \right], \quad (\text{B6})$$

where squared bracket will denote average over Y under the Gaussian measure and

$$A = \langle \tilde{+} | Z_x Z_0 M(y_{t_f}) \mathbf{u}_{t_f} \cdot \cdot M(y_1) \mathbf{u}_1 | +' \rangle, \quad (\text{B7})$$

$$B = \langle \tilde{+} | M^*(y_{t_f}) \mathbf{u}_{t_f} \cdot \cdot M^*(y_1) \mathbf{u}_1 | +' \rangle, \quad (\text{B8})$$

$$C = \langle \tilde{+} | M(y_{t_f}) \mathbf{u}_{t_f} \cdot \cdot M(y_1) \mathbf{u}_1 | +' \rangle. \quad (\text{B9})$$

Note that in the Dirac expectation, each Z would flip $|+\rangle \rightarrow |-\rangle$, and $\mathbf{u}$ would generate a superposition of $|-\rangle$ the diffuses. Since at the end, we bra with a $\langle +|$, each Z must be able to be paired up with another Z from previous time steps. Expanding each factor as $A = A_0 + \epsilon A_1 + \epsilon^2 A_2$ and similarly for B, C , This implies that $A_{\text{odd}} = B_{\text{odd}} = C_{\text{odd}} = 0$. Due to the choice of the initial condition and properties of the SSEP, we also have that $A_0 = 0$, and $B_0 = C_0 = 1$. Also, since we put dependence on ϵ and ϵ_* outside the definition of A_n, B_n, C_n , we have that $B_n = C_n$. Denote $(n \cdots m)_A := \langle \tilde{+} | Z_x Z_0 (n \cdots m) \mathbf{u} \cdots \mathbf{u} | + \rangle$ and similarly for B, C . Then, any lone $(n_{\text{odd}})_{A,B,C} = 0$ there are odd numbers of Z 's, $(n_{\text{even}})_A = \sum_{x,t} y_{x,t}^{n_{\text{even}}} \langle Z_x Z_0 \rangle_{\text{SSEP}} = 0$ due to the choice of the initial condition, and $(n_{\text{even}})_{B,C} = \sum_{x,t} y_{x,t}^{n_{\text{even}}}$.

Therefore,

$$\left[\frac{AB}{C} \right] = \left[\frac{(\epsilon^2 A_2 + \epsilon^4 A_4 + \cdots)(1 + r_\epsilon^2 \epsilon^2 B_2 + r_\epsilon^4 \epsilon^4 B_4 + \cdots)}{1 + \epsilon^2 B_2 + \epsilon^4 B_4 + \cdots} \right] \\ = [(\epsilon^2 A_2 + \epsilon^4 A_4 + \cdots)(1 + r_\epsilon^2 \epsilon^2 B_2 + r_\epsilon^4 \epsilon^4 B_4 + \cdots)(1 - (\epsilon^2 B_2 + \epsilon^4 B_4 + \cdots) + (\epsilon^2 B_2 + \epsilon^4 B_4 + \cdots)^2 + \cdots)] \quad (\text{B10})$$

$$= [(\epsilon^2(11)_A + \epsilon^4\{(31) + (211) + (1111)\}_A + \mathcal{O}(\epsilon^6)) \quad (\text{B11})$$

$$\times (1 + r_\epsilon^2 \epsilon^2 \{(2) + (11)\}_B + \mathcal{O}(\epsilon^4)) \quad (\text{B12})$$

$$\times (1 - \epsilon^2 \{(2) + (11)\}_B + \mathcal{O}(\epsilon^4))]. \quad (\text{B13})$$

When averaging over each $y_{x,t}$, $[y_{x,t}^{n_{\text{odd}}}] = 0$, which means that each $y_{x,t}$ must pair up. Therefore,

$$\left[\frac{AB}{C} \right] = \epsilon^4 (r_\epsilon^2 - 1) [(11)_A (11)_B]. \quad (\text{B14})$$

The exponential factor cannot pair up with any term to create factors of order smaller or equal to ϵ^4 , so we have

$$\mathbb{E} [s_{x,t_f} s_{0,t_f}] = \epsilon^4 (r_\epsilon^2 - 1) [(11)_A (11)_B]. \quad (\text{B15})$$

Now

$$[(11)_A (11)_B] = \sum_{(x,t) \neq (x',t')} \langle Z_{0,t_f} Z_{l,t_f} Z_{x,t} Z_{x',t'} \rangle_{\text{SSEP}} \langle Z_{x,t} Z_{x',t'} \rangle_{\text{SSEP}} \quad (\text{B16})$$

$$= \left(\sum_{t' \neq t, x, x'} + \sum_{t' = t, x' \neq x} \right) \langle Z_{0,t_f} Z_{l,t_f} Z_{x,t} Z_{x',t'} \rangle_{\text{SSEP}} \langle Z_{x,t} Z_{x',t'} \rangle_{\text{SSEP}}. \quad (\text{B17})$$

The second sum is zero again due to the half-filling initial condition. So,

$$[(11)_A (11)_B] = 2 \sum_{t < t', x, x'} \langle Z_{0,t_f} Z_{l,t_f} Z_{x,t} Z_{x',t'} \rangle_{\text{SSEP}} \langle Z_{x,t} Z_{x',t'} \rangle_{\text{SSEP}}. \quad (\text{B18})$$

$$= 2 \sum_{t < t', x, x'} p(\mathbf{0}, \mathbf{l} \leftarrow \mathbf{r}, \mathbf{r}' \text{ or } \mathbf{0}, \mathbf{l} \leftarrow \mathbf{r}', \mathbf{r}; \mathbf{r}, \mathbf{r}' \text{ do not meet}) p(\mathbf{r}' \leftarrow \mathbf{r}), \quad (\text{B19})$$

where $\mathbf{r} = (x, t)$, $\mathbf{0} = (0, t_f)$, $\mathbf{l} = (l, t_f)$, and the probabilities refers to inserting a particle at $\mathbf{r}$ then at $\mathbf{r}'$ and evolving them according to the SSEP. This is a sum of probabilities, and therefore is greater than zero. In fact, all terms $[(n \cdots n)_X \cdots (m \cdots m)_Y]$ are non-negative quantities similarly to above. Therefore,

$$\mathbb{E} [s_{x,t_f} s_{0,t_f}] = 2\epsilon^4 (r_\epsilon^2 - 1) \sum_{t < t', x, x'} p(\mathbf{0}, \mathbf{l} \leftarrow \mathbf{r}, \mathbf{r}' \text{ or } \mathbf{0}, \mathbf{l} \leftarrow \mathbf{r}', \mathbf{r}; \mathbf{r}, \mathbf{r}' \text{ do not meet}) p(\mathbf{r}' \leftarrow \mathbf{r}) + \mathcal{O}(\epsilon^6). \quad (\text{B20})$$

2. Connected correlator

To calculate the connected correlator, we look at the disconnected part,

$$\mathbb{E} [s_{0,t_f} s_{l,t_f}] = e^{-L t_f r^2 \epsilon^2 / 2} \left[\frac{DD'C}{B^2} \right], \quad (\text{B21})$$

where

$$D = \langle \tilde{+} | Z_l M(\mathbf{y}_{t_f}) \mathbf{u}_{t_f} \cdots M(\mathbf{y}_1) \mathbf{u}_1 | + \rangle, \quad (\text{B22})$$

$$D' = \langle \tilde{+} | Z_0 M(\mathbf{y}_{t_f}) \mathbf{u}_{t_f} \cdots M(\mathbf{y}_1) \mathbf{u}_1 | + \rangle. \quad (\text{B23})$$

In this case, for $D = D_0 + \epsilon D_1 + \epsilon^2 D_2 + \dots$, only odd terms survive, since there odd numbers of Z 's must be inserted to have even numbers in the Dirac expectation. So we have

$$\left[\frac{DD'C}{B^2} \right] = \left[\frac{(\epsilon D_1 + \epsilon^3 D_3 + \dots)(\epsilon D'_1 + \epsilon^3 D'_3 + \dots)(1 + r^2 \epsilon^2 B_2 + r^4 \epsilon^4 B_4 + \dots)}{(1 + \epsilon^2 B_2 + r^4 \epsilon^4 B_4 + \dots)(1 + \epsilon^2 B_2 + \epsilon^4 B_4 + \dots)} \right] \quad (\text{B24})$$

$$= [\epsilon^2 D_1 D'_1 + \epsilon^4 r^2 D_1 D'_1 B_2 - 2\epsilon^4 D_1 D'_1 B_2 + \dots] \quad (\text{B25})$$

$$= \epsilon^2 [(1)_D (1)_{D'}] + \dots \quad (\text{B26})$$

$$= \epsilon^2 \sum_{(x,t)} \langle Z_{0,t_f} Z_{x,t} \rangle_{\text{SSEP}} \langle Z_{l,t_f} Z_{x,t} \rangle_{\text{SSEP}} + \dots \quad (\text{B27})$$

$$= \epsilon^2 \sum_{\mathbf{r}} p(\mathbf{0} \leftarrow \mathbf{r}) p(\mathbf{l} \leftarrow \mathbf{r}) + \dots \quad (\text{B28})$$

Again, any term from the expontial factor $e^{L t_f r^2 \epsilon^2 / 2}$ only creates higher order terms, so

$$\mathbb{E} [\langle s_{0,t_f} \rangle \langle s_{l,t_f} \rangle] = \epsilon^2 \sum_{\mathbf{r}} p(\mathbf{0} \leftarrow \mathbf{r}) p(\mathbf{l} \leftarrow \mathbf{r}). \quad (\text{B29})$$

Since the disconnected correlator is zero for order ϵ^2 , we have

$$\mathbb{E} [\langle s_{0,t_f} s_{l,t_f} \rangle_c] = -\epsilon^2 \sum_{\mathbf{r}} p(\mathbf{0} \leftarrow \mathbf{r}) p(\mathbf{l} \leftarrow \mathbf{r}) + \mathcal{O}(\epsilon^6). \quad (\text{B30})$$

3. Teacher-student alignment

We have

$$\mathbb{E} [\langle s_{l,t_f} \rangle s_{0,t_f}^*] = e^{-L t_f r^2 \epsilon^2 / 2} \left[\frac{DE}{B} \right], \quad (\text{B31})$$

where

$$E = \langle \tilde{\mathbf{r}} | Z_0^* M^*(\mathbf{y}_{t_f}) \mathbf{u}_{t_f} \dots M^*(\mathbf{y}_1) \mathbf{u}_1 | +' \rangle. \quad (\text{B32})$$

Again only $E_{n_{\text{odd}}}$ survive. We also have that $E_n = D'_n$ as we put dependence on r_ϵ, ϵ outside, so

$$\left[\frac{DE}{B} \right] = \left[\frac{(\epsilon D_1 + \epsilon^3 D_3 + \dots)(r_\epsilon \epsilon D'_1 + r_\epsilon^3 \epsilon^3 D'_3 + \dots)}{1 + \epsilon^2 B_2 + r^4 \epsilon^4 B_4 + \dots} \right], \quad (\text{B33})$$

$$= [\epsilon^2 D_1 D'_1 + \dots] = r_\epsilon \epsilon^2 [(1)_D (1)_{D'}] + \dots \quad (\text{B34})$$

$$= r_\epsilon \epsilon^2 \sum_{(x,t)} \langle Z_{0,t_f} Z_{x,t} \rangle_{\text{SSEP}} \langle Z_{l,t_f} Z_{x,t} \rangle_{\text{SSEP}} + \dots$$

$$= r_\epsilon \epsilon^2 \sum_{\mathbf{r}} p(\mathbf{0} \leftarrow \mathbf{r}) p(\mathbf{l} \leftarrow \mathbf{r}) + \dots, \quad (\text{B35})$$

so we have

$$\mathbb{E} [\langle s_{l,t_f} \rangle s_{0,t_f}^*] = r_\epsilon \epsilon^2 \sum_{\mathbf{r}} p(\mathbf{0} \leftarrow \mathbf{r}) p(\mathbf{l} \leftarrow \mathbf{r}) + \mathcal{O}(\epsilon^4). \quad (\text{B36})$$

APPENDIX C: SADDLE POINT ANALYSIS OF THE STATIC CHARGE INFERENCE PROBLEM

We consider a simple model of inference, where the charges are static. In this case, the posterior is given by

$$p_s(\mathbf{s} | \mathbf{y}_{1:t_f}) = \frac{\exp \{ \epsilon \sum_{x,t} y_{x,t} s_x \}}{Z(\mathbf{y}_{1:t_f})}, \quad (\text{C1})$$

where

$$Z(\mathbf{y}_{1:t_f}) = \sum_{s_{t_f}} \exp \left\{ \epsilon \sum_{x,t} y_{x,t} s_r \right\}. \quad (\text{C2})$$

As the partition function is separable we can sum over each site separately. Hence, we have

$$\langle s_x \rangle = \tanh \left(\epsilon \sum_t y_{x,t} \right), \quad \langle s_x s_{x'} \rangle = \tanh \left(\epsilon \sum_t y_{x,t} \right) \tanh \left(\epsilon \sum_t y_{x',t} \right) \quad \text{for } x \neq x'. \quad (\text{C3})$$

Therefore, we have that $\langle s_x s_{x'} \rangle_c = 0$. Using translation invariance, the charge variance is given by

$$\delta C^2 = \mathbb{E} \left[\sum_{x,y} \langle s_x s_y \rangle_c \right] = \mathbb{E} \left[L(1 - \langle s_0 \rangle^2) + 2L \sum_{d=1}^{L/2} \langle s_0 s_d \rangle_c \right] = \mathbb{E} [L(1 - \langle s_0 \rangle^2)] = L \times \mathbb{E} \left[\text{sech}^2 \left(\epsilon \sum_t y_{0,t} \right) \right]. \quad (\text{C4})$$

The teacher's distribution is given by

$$p_*(\mathbf{s}^*, \mathbf{y}_{1:t_f}) = \frac{1}{2^L} \prod_{x,t} \left(\frac{\exp \left\{ -\frac{1}{2} (y_{x,t} - \epsilon_* s_x^*)^2 \right\}}{\sqrt{2\pi}} \right). \quad (\text{C5})$$

To calculate the expectation value, first note that we can integrate over all $y_{x \neq 0,t}$. Then, we shift $y_{0,t} \rightarrow y_{0,t} + \epsilon_* s_0^*$. We are left with

$$\delta C^2 = \frac{L}{2} \sum_{s_0^*} \int \prod_t \left(dy_{0,t} \frac{e^{-y_{0,t}^2/2}}{\sqrt{2\pi}} \right) \text{sech}^2 \left(\epsilon \sum_t y_{0,t} + \epsilon_* t_f s_0^* \right). \quad (\text{C6})$$

Next, we shift $y_{0,t} \rightarrow s_0^* y_{0,t}$. Using the evenness of sech^2 , we get

$$\delta C^2 = L \int \prod_t \left(dy_{0,t} \frac{e^{-y_{0,t}^2/2}}{\sqrt{2\pi}} \right) \text{sech}^2 \left(\epsilon \sum_t y_{0,t} + \epsilon_* t_f \right). \quad (\text{C7})$$

We can treat $\sum_t y_{0,t} =: y$ as a single random variable with variance t_f . Making this transformation then transforming $y \rightarrow t_f y$, we get

$$\delta C^2 = L \sqrt{\frac{t_f}{2\pi}} \int dy \exp [-t_f f(y)], \quad (\text{C8})$$

where $f(y) = \frac{y^2}{2} - \frac{2}{t_f} \ln \text{sech}(t_f(y + \epsilon_*))$. Then, the saddle point approximation is

$$\delta C^2 = L \frac{1}{\sqrt{f''(y_0)}} e^{-t_f f(y_0)} (1 + \dots). \quad (\text{C9})$$

Now $-(1/t_f) \ln \text{sech}(t_f x) \rightarrow |x|$ as $t_f \rightarrow \infty$, which we will use to identify the saddle point $y_0 + \text{sgn}(\epsilon y_0 + \epsilon_*) = 0$, though we will keep the exact function for the curvature $f''(y) = 1 + 2t_f \epsilon^2 \text{sech}^2(t_f(y + \epsilon_*))$.

At $\epsilon_* = 0$, the saddle point is at $y_0 = 0$, giving $f(y_0) = 0$ with curvature $f''(0) = 1 + 2\epsilon^2 t_f$.

For $0 < r_\epsilon < 2\epsilon$, the saddle point is at $y_0 = -\epsilon_*/\epsilon$ so $f(y_0) = r_\epsilon^2/2 - (2/t_f) \ln \text{sech}(t_f(0)) = r_\epsilon^2/2$ and $f''(y_0) = 1 + 2t_f \epsilon^2$.

For $r_\epsilon > 2\epsilon$, $y_0 = -2\epsilon$ and $f''(y_0) = 1 + 2t_f \epsilon^2 \text{sech}^2(t_f(\epsilon_* - 2\epsilon^2)) \rightarrow 1$.

Therefore for large t_f , we have (for $\epsilon > 0$)

$$\delta C^2 = L \times \begin{cases} \epsilon^{-1} t_f^{-1/2}, & r_\epsilon = 0, \\ \epsilon^{-1} t_f^{-1/2} e^{-r_\epsilon^2 t_f/2}, & r_\epsilon < 2\epsilon, \\ e^{-2\epsilon(r_\epsilon - 2\epsilon)t_f}, & r_\epsilon > 2\epsilon. \end{cases} \quad (\text{C10})$$

APPENDIX D: INFERRABILITY ARGUMENT FOR MAPPING FROM PROJECTIVE TO WEAK MEASUREMENTS

In projective measurement case, when a measurement does occur, then we know the true spin with probability 1. When it does not occur, then the probability of guessing the true spin is 1/2. Therefore, with rate π the probability of guessing the correct spin is $p(\pi) = \frac{1}{2} \times (1 - \pi) + 1 \times \pi$.

For the weak measurement case, the probability of guessing the correct spin should be given by $p(s = s' | s^* = s')$. Now

$$\begin{aligned} p(s = s' | s^* = s') &= \int_\phi p(s = s' | \phi) p(\phi | s^* = s') = p(s | s^*) \\ &= \int_\phi \frac{p(\phi | s = s') p(\phi | s^* = s')}{p(\phi)} \end{aligned} \quad (\text{D1})$$

$$= \int_\phi \frac{e^{\epsilon \phi}}{2 \cosh \epsilon \phi} \frac{e^{-(\phi - \epsilon)^2/2}}{\sqrt{2\pi}} =: p(\epsilon). \quad (\text{D2})$$

If we assume that only the probability of guessing the correct spin is important in the transition, then the measurement rate and signal strength variables should relate as $p(\epsilon) = p(\pi)$.

In Ref. [3], they approximate the critical rate to be around $\pi_{\text{BO}}^{\#} \in [0.2, 0.3]$ for the U(1) quantum circuit. Using above relations, this corresponds to $p_{\text{BO}}^{\#} \in [0.60, 0.65]$, which corresponds to $\epsilon_{\text{BO}}^{\#} \in [0.49, 0.63]$.

APPENDIX E: EXACT RELATIONS BETWEEN “SHARPENING” AND “INFERENCE” MOMENT ORDER PARAMETERS

Let us use $\langle \cdot \rangle$ to denote the average over the student's posterior. The first natural quantity is the mean squared error of the mean of the student's inferred charge distribution,

$$\text{MSEM} = \mathbb{E}_{Y, X^*} [\langle C \rangle - C^*]^2 = \mathbb{E}_{Y, X^*} [\langle C \rangle^2 - 2\langle C \rangle C^* + (C^*)^2]. \quad (\text{E1})$$

The second is the mean squared error on the joint distribution,

$$\text{MSE} = \mathbb{E}_{Y, X^*} [\langle (C - C^*)^2 \rangle] = \mathbb{E}_{Y, X^*} [\langle C^2 \rangle - 2\langle C \rangle C^* + (C^*)^2]. \quad (\text{E2})$$

The third is the charge variance of the student's inferred charge distribution,

$$\delta C^2 = \mathbb{E}_{Y, X^*} [\langle C^2 \rangle - \langle C \rangle^2]. \quad (\text{E3})$$

In the Bayes optimal case, distribution of X and X^* is identical in the joint distribution. Therefore, $[\langle C \rangle^2] = [\langle C \rangle C^*]$. This means that, at Bayes optimality,

$$\text{MSEM}(\text{Bayes optimal}) = \frac{1}{2} \text{MSE}(\text{Bayes optimal}) = \delta C^2(\text{Bayes optimal}). \quad (\text{E4})$$

APPENDIX F: DERIVATION OF THE REPETITION CODE CASE

We denote $e_{l_{\rightarrow}}^* = -1$ as a bit-flip error occurring at time t , and $e_{l_{\uparrow}}^* = -1$ to denote if a syndrome measurement is corrupted by a read-out error. Then, we have $e_{l_{\rightarrow}}^* = f_{l,t}^* f_{l,t-1}^*$ and $e_{l_{\uparrow}}^* = s_{v,t} s_{v,t}^* = s_{l_{\uparrow}} s_{l_{\uparrow}}^*$, where $s_{v,t}^* = f_{l_{\text{L}}(v),t}^* f_{l_{\text{R}}(v),t}^*$ is the “true” syndrome. Then, we can decouple the prior as

$$p_*(F^*) \rightarrow p_*(E_{\rightarrow}^*) \propto \exp \left[\beta_* \sum_{l_{\rightarrow}} e_{l_{\rightarrow}}^* \right].$$

The likelihood for each syndrome can also be written as

$$p_*(s_{v,t} | f_{l_{\text{L}}(v),t}^*, f_{l_{\text{R}}(v),t}^*) \rightarrow p_*(s_{l_{\uparrow}}, e_{l_{\uparrow}}^* | s_{l_{\uparrow}}^*) \propto \delta[s_{l_{\uparrow}}, e_{l_{\uparrow}}^* s_{l_{\uparrow}}^*] \exp[\beta_* e_{l_{\uparrow}}^*] = p_*(s_{l_{\uparrow}} | e_{l_{\uparrow}}^*, s_{l_{\uparrow}}^*) p_*(e_{l_{\uparrow}}^*).$$

Denoting all occurrences of error as $E^* = E_{\rightarrow}^* \odot E_{\uparrow}^*$, we then have

$$p_*(S|F^*) \rightarrow p_*(S|E_{\uparrow}^*, S^*(E_{\rightarrow}^*)) p_*(E_{\uparrow}^*) = p_*(S|E^*) p_*(E^*).$$

Let W be the true “worldlines,” which includes the “true” syndromes S^* on the vertical links and the true bit-flips $E_{\rightarrow}^*$ on the horizontal links. Then, $W = S^* \odot E_{\rightarrow}^*$. For the whole set of syndromes, we have $E_{\uparrow}^* = S \odot S^*$. Then, we have that

$$S = E^* \odot W.$$

Now, W does not have a boundary, if we assert that vertical links at final time has no boundary above. Then this means that S and E^* must share a boundary. Therefore, likelihood for syndrome trajectory given all errors is

$$p_*(S|E^*) = \delta_{\partial S, \partial E^*}.$$

In this case, the boundary of S or E^* now lives on *vertices* of the 2D space-time lattice.

Marginalizing over the syndromes, the conditional probability of inferring E given E^* is

$$p(E|E^*) = \sum_S p_s(E|S) p_*(S|E^*) = \sum_S \frac{\delta_{\partial S, \partial E} p_s(E) \delta_{\partial S, \partial E^*}}{p_s(S)} = \sum_S \frac{\delta_{\partial S, \partial E} p_s(E) \delta_{\partial S, \partial E^*}}{\sum_{E'} \delta_{\partial S, \partial E'} p_s(E')}.$$

Since in the sum over E' , only S such that $\partial S = \partial E^*$ survive, we can replace $\delta_{\partial S, \partial E'} \rightarrow \delta_{\partial E^*, \partial E'}$. Then, this term does not depend on S and can be pulled out of the sum. Therefore,

$$\begin{aligned} p(E|E^*) &= \sum_S \frac{\delta_{\partial S, \partial E} p_s(E) \delta_{\partial S, \partial E^*}}{\sum_{E'} \delta_{\partial E^*, \partial E'} p_s(E')} = \frac{p_s(E)}{\sum_{E'} \delta_{\partial E^*, \partial E'} p_s(E')} \sum_S \delta_{\partial S, \partial E} \delta_{\partial S, \partial E^*} \\ &= \frac{p_s(E)}{\sum_{E'} \delta_{\partial E^*, \partial E'} p_s(E')} \sum_S \delta_{\partial E, \partial E^*} \delta_{\partial S, \partial E^*} = p_s(E) \delta_{\partial E, \partial E^*} \times \frac{\sum_S \delta_{\partial S, \partial E^*}}{\sum_{E'} \delta_{\partial E^*, \partial E'} p_s(E')}. \end{aligned}$$

Then, one notes that the second factor is independent of E ; therefore,

$$p(E|E^*) \propto p_s(E) \delta_{\partial E, \partial E^*}.$$

Similarly to the toric code case, to respect the delta function, we only consider $E = E^* \odot C$, where C is a cycle up to the boundary. Therefore, we can parametrize $c_l = \sigma_{v_l(l)} \sigma_{v_{l+1}(l)}$ except at the final time where $c_{l_t} = \sigma_{v_{\text{below}}(l_t)}$, and the vertex lives below l_t . We could also set the boundary condition that the vertex that lives above has $\sigma_{v_{\text{above}}} = 1$. Therefore, we again have the 2D RBIM partition function [Eq. (43)] with periodic boundary conditions in space, free open boundary at $t = 1$, and fixed boundary condition $\sigma_{v_{\text{above}}} = 1$ at $t = t_f$. To calculate expectation values, we do so with respect to the *true/teacher's* prior distribution $p_*(E^*) \propto \exp[\beta_* \sum_l e_l^*]$.

Using that $f_{l,t} = e_{l,t} \cdots e_{l,1}$, $f_{l,t}^* = e_{l,t}^* \cdots e_{l,1}^*$, and $e_{l,t} = c_{l,t} e_{l,t}^* = \sigma_{v_L(l),t} \sigma_{v_R(l),t} e_{l,t}^*$, the correlation order parameter becomes

$$\mathbb{E}[f_{l,t} f_{l,t}^*] = \mathbb{E} \left[\prod_{\tau=1}^t \sigma_{v_L(l),\tau} \sigma_{v_R(l),\tau} \right]. \quad (\text{F1})$$

APPENDIX G: PROOF OF CONCENTRATION OF MEASURE FOR HAAR UNITARIES

Consider a quantum circuit given by the evolved state $\rho_t = \mathcal{U} \cdots \mathcal{U} \mathcal{M} \cdots \mathcal{M} \cdots \rho_0$, where each $\mathcal{U}$ is an instance Haar-random distributed local unitary gate, acting on (say) sites $\mathbf{i} = (i_1, i_2, \dots, i_k)$, i.e., on k qubit-qudit pairs. It could have further block-diagonal structure, and its average is given by $\mathbb{E}_u \mathcal{U} = \bar{\mathcal{U}}$. $\mathcal{M}(y)$ is a local measurement map

$$\mathcal{M}(y) : \rho \rightarrow \frac{Q(y) \rho Q(y)}{\text{tr}[Q(y) \rho Q(y)]}, \quad (\text{G1})$$

where $Q(y)$ are partial projectors acting on qubits on sites $\mathbf{i}'$. We will suppress the argument y in what follows, unless required.

We will denote state evolved by averaged Haar-unitary elements as $\bar{\rho}_t = \bar{\mathcal{U}} \cdots \bar{\mathcal{U}} \mathcal{M} \cdots \mathcal{M} \cdots \rho_0$.

We would like to show that the Bayes-optimal posterior distribution is concentrated around that given by the Haar-averaged elements, i.e.,

$$\mathbb{E}_{U,Y} |p(s|Y) - \bar{p}(s|Y)| \xrightarrow{d \rightarrow \infty} 0, \quad (\text{G2})$$

where the posterior is given by $p(s|Y) = \text{tr}[\mathbb{P}_s \rho_t]$ and the posterior from averaged elements, and $\bar{p}(s|Y) := \text{tr}[\mathbb{P}_s \bar{\rho}_t]$. $Y = \{y\}_y$ is the set of all measurements. Precisely, we will prove the following.

Theorem G1. Consider a quantum circuit composed of Haar-randomly distributed local unitaries and local measurements. The unitaries acts on k qudit-qubit pairs, and there is a computational qubit basis such that it is block diagonal, $u = u^{(1)} \oplus \cdots \oplus u^{(B)}$. Each sector has dimensions $\propto d^k$. The local measurements $\mathcal{M}$ measure k' qubits and are also diagonal in the same computational qubit basis. Assume that there is an extensive number of such elements $\sim VT$, where VT is the space-time volume, and that measurements are performed after each layer of unitaries. Then, there exists some constant r , such that

$$\mathbb{E}_{U,Y} |p(s|Y) - \bar{p}(s|Y)| \leq C \frac{VT r_1^V r_2^{VT}}{d^{k/2}}. \quad (\text{G3})$$

Corollary G1. For a circuit with Haar-random $u_{U(1)}$ gates in brickwork fashion with single-site monitoring, we have

$$\mathbb{E}_{U,Y} |p(s|Y) - \bar{p}(s|Y)| \leq \frac{LT\phi}{2d} r_{\text{unitary}}^{L/2} (r_{\text{unitary}}^{1/2} r_{\text{meas}})^{LT}, \quad (\text{G4})$$

with $\phi = \sqrt{3}$, $r_{\text{unitary}} = 10$, and $r_{\text{meas}} = 2(\sqrt{2} + 1)$.

Proof of Theorem 1 and Corollary 1. Our strategy is as follows. We would like to bound

$$c := \mathbb{E}_{U,Y} |p(s|Y) - \bar{p}(s|Y)| = \mathbb{E}_{U,Y} |\text{tr}[\mathbb{P}_s(\rho - \bar{\rho})]|. \quad (\text{G5})$$

We define an inner product between operators as $\langle\langle A|B \rangle\rangle := \text{tr}[A^\dagger B]$. By Cauchy-Schwartz inequality, we have that $|\langle x, y \rangle| \leq \|x\| \|y\|$. Since we have $\langle\langle A|A \rangle\rangle = \text{tr}[A^\dagger A] = \|A\|_F^2$, the appropriate norm is the Frobenius norm. Therefore, we have

$$c \leq \mathbb{E}_{U,Y} [\|\mathbb{P}_s\|_F \|\rho - \bar{\rho}\|_F] = \|\mathbb{P}_s\|_F \mathbb{E}_{U,Y} [\|\rho - \bar{\rho}\|_F]. \quad (\text{G6})$$

Looking at $\|\mathbb{P}_s\|_F$, it will grow as d^L . However, we also have that

$$p(s|Y) = \text{tr}[\mathbb{P}_s^{\mathbf{S}} \text{tr}_{\mathbf{S}}[\rho]], \quad (\text{G7})$$

where $\mathbb{P}_s^{\mathbf{S}}$ only acts on a reduced Hilbert space of qubits $\mathbf{S}$ and $\text{tr}_{\mathbf{S}}$ denotes the trace over all degrees of freedom other than qubits. Denoting the space of qudits as $\mathbf{G}$, we have $\mathbf{S} = \mathbf{G}$. Following through the line of reasoning as before, we have

$$c = \mathbb{E}_{U,Y} |\text{tr}[\mathbb{P}_s^{\mathbf{S}} \text{tr}_{\mathbf{G}}(\rho - \bar{\rho})]| = \mathbb{E}_{U,Y} [\|\mathbb{P}_s^{\mathbf{S}}\|_F \|\text{tr}_{\mathbf{G}}(\rho - \bar{\rho})\|_F] = \mathbb{E}_{U,Y} \|\text{tr}_{\mathbf{G}}(\rho - \bar{\rho})\|_F. \quad (\text{G8})$$

Suppose that the last operation was an application of $\mathcal{U}$, supported on $\mathbf{SG}(i)$, the Hilbert space of qubits and qudits on sites i . Then,

$$c = \mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{G}}(\rho_t - \bar{\rho}_t)\|_{\mathrm{F}} = \mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{G}}(\rho_{t-1} - \bar{\rho}_{t-1})\|_{\mathrm{F}} = \mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{G}}(\mathcal{U}\rho_{t-1} - \bar{\mathcal{U}}\bar{\rho}_{t-1})\|_{\mathrm{F}} \quad (\text{G9})$$

$$= \mathbb{E}_{U,Y} [\|\mathrm{tr}_{\mathbf{G}}\bar{\mathcal{U}}(\rho_{t-1} - \bar{\rho}_{t-1}) + \mathrm{tr}_{\mathbf{G}}(\mathcal{U} - \bar{\mathcal{U}})\rho_{t-1}\|_{\mathrm{F}}] \leq \mathbb{E}_{U,Y} [\|\mathrm{tr}_{\mathbf{G}}\bar{\mathcal{U}}(\rho_{t-1} - \bar{\rho}_{t-1})\|_{\mathrm{F}}] + \mathbb{E}_{U,Y} [\|\mathrm{tr}_{\mathbf{G}}(\mathcal{U} - \bar{\mathcal{U}})\rho_{t-1}\|_{\mathrm{F}}]. \quad (\text{G10})$$

In Lemmas 3 and 4, we will prove that the above is bounded by

$$c \leq r_{\text{unitary}} \mathbb{E}_{U,Y} [\|\mathrm{tr}_{\mathbf{G} \setminus \mathbf{S}(i) \cup \mathbf{G}(i)}(\rho_{t-1} - \bar{\rho}_{t-1})\|_{\mathrm{F}}] + \frac{\phi}{d^{k/2}} = r_{\text{unitary}} \mathbb{E}_{U,Y} [\|\mathrm{tr}_{\mathbf{G}}(\rho_{t-1} - \bar{\rho}_{t-1})\|_{\mathrm{F}}] + \frac{\phi}{d^{k/2}}, \quad (\text{G11})$$

for some constant r_{unitary} and ϕ . If the last operation was instead an application of $\mathcal{M}$, with support on qubits on sites i , we would have

$$c = \mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{G}}[\mathcal{M}[\rho_{t-1}] - \mathcal{M}[\bar{\rho}_{t-1}]]\|_{\mathrm{F}}. \quad (\text{G12})$$

We will also prove in Lemma 5 that

$$\mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{G}}[\mathcal{M}[\rho_{t-1}] - \mathcal{M}[\bar{\rho}_{t-1}]]\|_{\mathrm{F}} \leq r_{\text{meas}}^{(1)} \mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{G} \setminus \mathbf{S}(i)}(\rho_{t-1} - \bar{\rho}_{t-1})\|_{\mathrm{F}} + r_{\text{meas}}^{(2)} \mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{S}(i)}(\rho_{t-1} - \bar{\rho}_{t-1})\|_{\mathrm{F}} \quad (\text{G13})$$

$$= r_{\text{meas}}^{(1)} \mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{G}}(\rho_{t-1} - \bar{\rho}_{t-1})\|_{\mathrm{F}} + r_{\text{meas}}^{(2)} \mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{G} \cup \mathbf{S}(\setminus i)}(\rho_{t-1} - \bar{\rho}_{t-1})\|_{\mathrm{F}}. \quad (\text{G14})$$

Then, we can recursively apply the same procedure as above for $\mathbb{E}_{U,Y} [\|\mathrm{tr}_{\mathbf{K}}(\rho_{\tau} - \bar{\rho}_{\tau})\|_{\mathrm{F}}]$ for subspaces of the form $\mathbf{K}$. Lemmas 3 and 4 also prove that, for subspaces of the form $\mathbf{K} = \mathbf{G} \cup \mathbf{S}(\cdot)$, i.e., contains *all* qudits,

$$\mathbb{E}_{U,Y} [\|\mathrm{tr}_{\mathbf{K}}(\rho_{\tau} - \bar{\rho}_{\tau})\|_{\mathrm{F}}] \leq r_{\text{unitary}} \mathbb{E}_{U,Y} [\|\mathrm{tr}_{\mathbf{K} \setminus \mathbf{S}(i) \cup \mathbf{G}(i)}(\rho_{\tau-1} - \bar{\rho}_{\tau-1})\|_{\mathrm{F}}] + \frac{\phi}{d^{k/2}}, \quad (\text{G15})$$

if the previous operation was a unitary. On the other hand, Lemma 5 proves that

$$\mathbb{E}_{U,Y} [\|\mathrm{tr}_{\mathbf{K}}(\rho_{\tau} - \bar{\rho}_{\tau})\|_{\mathrm{F}}] \leq r_{\text{meas}}^{(1)} \mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{K} \setminus \mathbf{S}(i)}(\rho_{\tau-1} - \bar{\rho}_{\tau-1})\|_{\mathrm{F}} + r_{\text{meas}}^{(2)} \mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{G} \cup \mathbf{S}(\setminus i)}(\rho_{\tau-1} - \bar{\rho}_{\tau-1})\|_{\mathrm{F}}, \quad (\text{G16})$$

if the previous operation was a measurement. Note that starting with $\mathbf{K} = \mathbf{G}$, we only keep generating subspaces of form $\mathbf{G} \cup \mathbf{S}(\cdot)$. We can continue these bounds recursively until $\tau = 0$, where we have that $\rho_0 = \bar{\rho}_0$, and therefore $\mathbb{E}_{U,Y} [\mathrm{tr}_{\mathbf{K}}(\rho_0 - \bar{\rho}_0)] = 0$.

Then, if we assume that a layer of measurement follows each layer of unitaries, and that each layer of unitary contains αV unitaries and each layer of measurement contains measurements, then we have

$$c \leq \frac{\phi}{d^{k/2}} \left(\sum_{x=1}^{\alpha V} (r_{\text{unitary}})^x \right) \left(\sum_{t=0}^{T-1} ((r_{\text{unitary}})^{\alpha} (r_{\text{meas}})^{\gamma})^{Vt} \right) \leq \frac{\phi}{d^{k/2}} \times \alpha V \times r_{\text{unitary}}^{L/2} \times T \times ((r_{\text{unitary}})^{\alpha} (r_{\text{meas}})^{\gamma})^{VT}. \quad (\text{G17})$$

For example, for single site measurements interleaved between Haar-random brickwork circuit,

$$c \leq \frac{\phi}{d} \left(\sum_{x=1}^{L/2} r_{\text{unitary}}^x \right) \left(\sum_{t=0}^{T-1} (r_{\text{unitary}}^{1/2} r_{\text{meas}})^{Lt} \right) \leq \frac{\phi}{d} \times \frac{L}{2} \times r_{\text{unitary}}^{L/2} \times T \times (r_{\text{unitary}}^{1/2} r_{\text{meas}})^{LT}, \quad (\text{G18})$$

with $\phi = \sqrt{3}$, $r_{\text{unitary}} = 10$, and $r_{\text{meas}} = r_{\text{meas}}^{(1)} + r_{\text{meas}}^{(2)} = 2(\sqrt{2} + 1)$. ■

As a first example, we consider a full block Haar unitary gate.

Lemma G1. Consider a full block Haar unitary gate supported on Hilbert space $\mathbf{N}$ with dimension N , acting on full Hilbert space that is potentially larger. Then,

$$\mathbb{E}_U [\|\mathrm{tr}_{\mathbf{K}}(\mathcal{U} - \mathcal{D})\rho\|_{\mathrm{F}}] \leq \frac{1}{\sqrt{|\mathbf{N} \cap \mathbf{K}|}}$$

for any state ρ , and the trace is on any subspace $\mathbf{K}$, which includes the trivial case where there is no partial trace.

Proof. Using Jensen's inequality, we have that

$$\mathbb{E}_U [\|\mathrm{tr}_{\mathbf{K}}(\mathcal{U} - \mathcal{D})\rho\|_{\mathrm{F}}] \leq \sqrt{\mathbb{E}_{U,Y} [\|\mathrm{tr}_{\mathbf{K}}(\mathcal{U} - \mathcal{D})\rho\|_{\mathrm{F}}^2]}. \quad (\text{G19})$$

Now consider

$$\mathbb{E}_{U,Y} [\|\mathrm{tr}_{\mathbf{K}}(\mathcal{U} - \mathcal{D})\rho\|_{\mathrm{F}}^2] = \mathbb{E}_{U,Y} [\|\mathrm{tr}_{\mathbf{K}}(u\rho u^{\dagger}) - \mathrm{tr}_{\mathbf{K}}\mathcal{D}[\rho]\|_{\mathrm{F}}^2] = \mathbb{E}_{U,Y} [\mathrm{tr}[(\mathrm{tr}_{\mathbf{K}}(u\rho u^{\dagger}) - \mathrm{tr}_{\mathbf{K}}\mathcal{D}[\rho])(\mathrm{tr}_{\mathbf{K}}(u\rho u^{\dagger}) - \mathrm{tr}_{\mathbf{K}}\mathcal{D}[\rho])]] \quad (\text{G20})$$

$$= \mathbb{E}_{U,Y} [\mathrm{tr}[\mathrm{tr}_{\mathbf{K}}(u\rho u^{\dagger})\mathrm{tr}_{\mathbf{K}}(u\rho u^{\dagger})] - 2\mathrm{tr}[\mathrm{tr}_{\mathbf{K}}\mathcal{D}(\rho)\mathrm{tr}_{\mathbf{K}}(u\rho u^{\dagger})] + \mathrm{tr}[\mathrm{tr}_{\mathbf{K}}\mathcal{D}(\rho)\mathrm{tr}_{\mathbf{K}}\mathcal{D}(\rho)]]. \quad (\text{G21})$$

Now taking the expectation value for u only, we have that $\mathbb{E}_{U,Y}[u\rho u^\dagger] = \mathcal{D}[\rho]$, such that

$$\mathbb{E}_{U,Y}[\|\mathrm{tr}_K(\mathcal{U} - \mathcal{D})\rho\|_F^2] = \mathbb{E}_{U,Y}[\mathrm{tr}[\mathrm{tr}_K(u\rho u^\dagger)\mathrm{tr}_K(u\rho u^\dagger)] - \mathrm{tr}[\mathrm{tr}_K\mathcal{D}(\rho)\mathrm{tr}_K\mathcal{D}(\rho)]] . \quad (\text{G22})$$

Let us consider the first term.

The following identities will prove to be useful, for an $N \times N$ Haar random unitary [16]:

$$\mathbb{E}_u[u_{ij}u_{kl}^*] = \frac{1}{N}\delta_{(ij),(kl)}, \quad (\text{G23})$$

$$\mathbb{E}_u u_{ij}u_{mn}^*u_{kl}u_{pq}^* = \frac{1}{N^2}\left(1 + \frac{1}{N^2 - 1}\right)(\delta_{(ij),(mn)}\delta_{(kl),(pq)} + \delta_{(ij),(pq)}\delta_{(kl),(mn)}) \quad (\text{G24})$$

$$- \frac{1}{N(N^2 - 1)}(\delta_{im}\delta_{jq}\delta_{kp}\delta_{\ell n} + \delta_{ip}\delta_{jn}\delta_{km}\delta_{\ell q}), \quad (\text{G25})$$

$$= \frac{1}{N}\delta_{(ij),(mn)}\frac{1}{N}\delta_{(kl),(pq)} + \frac{1}{N^2}\delta_{(ij),(pq)}\delta_{(kl),(mn)} \quad (\text{G26})$$

$$+ \frac{1}{N^2(N^2 - 1)}(\delta_{(ij),(mn)}\delta_{(kl),(pq)} + \delta_{(ij),(pq)}\delta_{(kl),(mn)} - \delta_{im}\delta_{jq}\delta_{kp}\delta_{\ell n} - \delta_{ip}\delta_{jn}\delta_{km}\delta_{\ell q}). \quad (\text{G27})$$

These can be written diagrammatically as the following:

$$\mathbb{E}_u \left[\begin{array}{c} | \\ \boxed{u} \\ | \end{array} \begin{array}{c} | \\ \boxed{u^*} \\ | \end{array} \right] = \begin{array}{c} \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \end{array}, \quad (\text{G28})$$

$$\mathbb{E}_u \left[\begin{array}{c} |j \\ \boxed{u^*} \\ |i \\ |k \\ \boxed{u} \\ |l \end{array} \begin{array}{c} |n \\ \boxed{u} \\ |m \\ |p \\ \boxed{u^*} \\ |q \end{array} \right] = \begin{array}{c} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \end{array} + \begin{array}{c} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \end{array} \quad (\text{G29})$$

$$+ \frac{1}{N^2 - 1} \left(\begin{array}{c} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \end{array} + \begin{array}{c} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \end{array} - \begin{array}{c} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \end{array} - \begin{array}{c} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \boxed{\chi_N} \text{---} \\ \text{---} \end{array} \right). \quad (\text{G30})$$

Here, χ_N is the maximally mixed state on Hilbert space acted on by the Haar unitary, N . Now consider the partial trace over K . Part of it will be within the N space, denoted by $K \cup N$ while the rest will be over $\setminus N$ space, denoted by $K \setminus N$. To denote $K \cup N$

and $N \setminus (K \cup N)$ separately, we can double the lines as following:

$$|\rho\rangle\rangle \leftrightarrow \text{diagram}, \quad (\text{G31})$$

where gray lines are in K , black lines in $\setminus K$, and solid lines are in N and dotted lines are in $\setminus N$. All the lines in the unitary relations can also be doubled to accomodate this, i.e.,

$$\text{diagram} \rightarrow \text{diagram}. \quad (\text{G32})$$

For example, we have

$$\text{tr}_K \mathcal{U}[\rho] \leftrightarrow \text{tr}_K \mathcal{U} |\rho\rangle\rangle \leftrightarrow \text{diagram}, \quad (\text{G33})$$

such that we have

$$\text{tr}[(\text{tr}_K \mathcal{U}[\rho])^2] \leftrightarrow \text{diagram}. \quad (\text{G34})$$

Then, we can apply the identities. For example, the first two terms in is equal to

$$\text{diagram} + \text{diagram} \leftrightarrow \text{tr}[\text{tr}_K \mathcal{D}[\rho] \text{tr}_K \mathcal{D}[\rho]] + \text{tr}_{N \cap K}[(\text{tr}_{N \setminus K}[\chi_N])^2] \cdot \text{tr}[\rho^2]. \quad (\text{G35})$$

Additionally using that $\rho^\dagger = \rho$ on the other terms, we find

$$\mathbb{E}_{U,Y} [\text{tr}[\text{tr}_K(u\rho u^\dagger) \text{tr}_K(u\rho u^\dagger)]] = \text{tr}[\text{tr}_K \mathcal{D}[\rho] \text{tr}_K \mathcal{D}[\rho]] + \text{tr}_{N \cap K}[(\text{tr}_{N \setminus K}[\chi_N])^2] \cdot \text{tr}[\rho^2] \quad (\text{G36})$$

$$+ \frac{1}{N^2 - 1} (\text{tr}[\text{tr}_K \mathcal{D}[\rho] \text{tr}_K \mathcal{D}[\rho]] + \text{tr}_{N \cap K}[(\text{tr}_{N \setminus K}[\chi_N])^2] \cdot \text{tr}[\rho^2]) \quad (\text{G37})$$

$$- \frac{1}{N^2 - 1} (\text{tr}[\text{tr}_K \mathcal{D}[\rho] \text{tr}_K \mathcal{D}[\rho]] + \text{tr}_{N \cap K}[(\text{tr}_{N \setminus K}[\chi_N])^2] \cdot \text{tr}[\rho^2]). \quad (\text{G38})$$

The last two terms cancel out, and $\text{tr}_{\mathbf{N} \cap \mathbf{K}}[(\text{tr}_{\mathbf{N} \setminus \mathbf{K}}[\chi_{\mathbf{N}}])^2] = \text{tr}_{\mathbf{N} \cap \mathbf{K}}[\chi_{\mathbf{N} \cap \mathbf{K}}] = 1/|\mathbf{N} \cap \mathbf{K}|$. Furthermore, $\text{tr}_{\mathbf{K} \setminus \mathbf{N}}[\rho] = \rho'$ is another state, and the trace of square of a density matrix is upper bounded by 1. The first term exactly cancels out the second term in $\mathbb{E}_{U,Y}[\|\text{tr}_{\mathbf{K}}(\mathcal{U} - \mathcal{D})\rho\|_{\text{F}}^2]$, and therefore we may upper bound

$$\mathbb{E}_{U,Y}[\|\text{tr}_{\mathbf{K}}(\mathcal{U} - \mathcal{D})\rho\|_{\text{F}}^2] \leq \frac{1}{|\mathbf{N} \cap \mathbf{K}|} \quad (\text{G39})$$

with no further terms. Using Jensen's inequality, we have

$$\mathbb{E}_{U,Y}[\|\text{tr}_{\mathbf{K}}(\mathcal{U} - \mathcal{D})\rho\|_{\text{F}}] \leq \frac{1}{|\mathbf{N} \cap \mathbf{K}|}. \quad (\text{G40})$$

Corollary G2. For two-site full block Haar unitary acting on a qubit-qudit pair with $\mathbf{K} = \mathbf{G} \cup \mathbf{S}(\cdot)$, we have $\mathbf{N} = \mathbf{G}(i_1, i_2) \cup \mathbf{S}(i_1, i_2)$, and therefore the bound

$$\mathbb{E}_{U,Y}[\|\text{tr}_{\mathbf{K}}(\mathcal{U} - \mathcal{D})\rho\|_{\text{F}}] \leq \frac{1}{d}. \quad (\text{G41})$$

Lemma G2. For the depolarizing channel $\mathcal{D}$ acting on $\mathbf{N}$ and any matrix A living in a potentially larger Hilbert space,

$$\|\text{tr}_{\mathbf{K}}\mathcal{D}A\|_{\text{F}} \leq \frac{1}{\sqrt{|\mathbf{N} \setminus \mathbf{K}|}} \|\text{tr}_{\mathbf{N} \setminus \mathbf{K}}A\|_{\text{F}}, \quad (\text{G42})$$

for any subspace $\mathbf{K}$.

Proof.

$$\|\text{tr}_{\mathbf{K}}\mathcal{D}A\|_{\text{F}}^2 = \text{tr}[(\text{tr}_{\mathbf{K} \cap \mathbf{N}}\mathcal{D}\text{tr}_{\mathbf{K} \setminus \mathbf{N}}A^\dagger)(\text{tr}_{\mathbf{K} \cap \mathbf{N}}\text{tr}_{\mathbf{K} \setminus \mathbf{N}}A)], \quad (\text{G43})$$

$$= \text{tr}[(\text{tr}_{\mathbf{K} \cap \mathbf{N}}\chi_{\mathbf{N}} \otimes \text{tr}_{\mathbf{N} \setminus \mathbf{K}}[A^\dagger])(\text{tr}_{\mathbf{K} \cap \mathbf{N}}\chi_{\mathbf{N}} \otimes \text{tr}_{\mathbf{N} \setminus \mathbf{K}}[A])], \quad (\text{G44})$$

$$= \text{tr}_{\mathbf{N} \setminus \mathbf{K}}[\text{tr}_{\mathbf{K} \cap \mathbf{N}}[\chi_{\mathbf{N}}]\text{tr}_{\mathbf{K} \cap \mathbf{N}}[\chi_{\mathbf{N}}]] \times \text{tr}[\text{tr}_{\mathbf{N} \setminus \mathbf{K}}[A^\dagger]\text{tr}_{\mathbf{N} \setminus \mathbf{K}}[A]], \quad (\text{G45})$$

$$= \frac{1}{|\mathbf{N} \setminus \mathbf{K}|} \|\text{tr}_{\mathbf{N} \setminus \mathbf{K}}A\|_{\text{F}}^2. \quad (\text{G46})$$

Remark G1. For any block-diagonal Haar unitary $u = u^{(0)} \oplus \dots \oplus u^{(B)}$ acting on $\mathbf{N}$, with $\sum_{\alpha=1}^B N_\alpha = N$, we have

$$\mathbb{E}_u[\mathcal{U}] = \mathcal{D}_{\mathbf{N}_1} \oplus \dots \oplus \mathcal{D}_{\mathbf{N}_B} = \sum_{\alpha} \frac{\mathbb{P}_{\alpha}^{\mathbf{N}}}{N_{\alpha}} \otimes \text{tr}_{\mathbf{N}}\mathcal{P}_{\alpha}. \quad (\text{G47})$$

Lemma G3. For a block-diagonal Haar unitary $u = u^{(0)} \oplus \dots \oplus u^{(B)}$ with $\mathbf{N} = \mathbf{m} \otimes \mathbf{n}$, $\mathbf{n} = \bigoplus_{\alpha} \mathbf{n}_{\alpha}$, $\sum_{\alpha} N_{\alpha} = N$, where $N_{\alpha} = m n_{\alpha}$ and $\mathbf{m} \subset \mathbf{K}$,

$$\mathbb{E}_U[\|\text{tr}_{\mathbf{K}}(\mathcal{U} - \mathcal{D}_{\mathbf{N}_1} \oplus \dots \oplus \mathcal{D}_{\mathbf{N}_B})\rho\|_{\text{F}}] \leq \sum_{\alpha} \text{tr}_{\mathbf{N} \cap \mathbf{K}} \left[\text{tr}_{\mathbf{N} \setminus \mathbf{K}} \left[\frac{\mathbb{P}_{\alpha}^{\mathbf{N}}}{N_{\alpha}} \right]^2 \right] \leq \sqrt{\frac{B}{m}}.$$

Proof. The block diagonal Haar unitary can be written as

$$u = \sum_{\alpha} \mathbb{P}_{\alpha} u_{\alpha} \mathbb{P}_{\alpha} \leftrightarrow \begin{array}{c} | \\ \boxed{u} \\ | \end{array} = \sum_{\alpha} \begin{array}{c} | \\ \boxed{v_{\alpha}} \\ | \end{array} = \sum_{\alpha} \begin{array}{c} \boxed{\mathbb{P}_{\alpha}} \\ | \\ \boxed{u_{\alpha}} \\ | \\ \boxed{\mathbb{P}_{\alpha}} \end{array}. \quad (\text{G48})$$

Here, $u_\alpha = u^{(\alpha)} \oplus \mathbb{1}_{N \setminus N_\alpha}$. Then, we have

$$[\|\text{tr}_K(\mathcal{U} - \mathcal{D}_{N_1} \oplus \dots \oplus \mathcal{D}_{N_B})\rho\|_F] = \sum_{\alpha\beta\gamma\delta} \left[\begin{array}{c} \rho \\ u^* \quad u \\ u \quad u^* \\ \rho \end{array} \right] = \sum_{\alpha\beta\gamma\delta} \left[\begin{array}{c} \rho \\ v_\gamma^* \quad v_\delta \\ v_\alpha \quad v_\beta^* \\ \rho \end{array} \right]. \quad (\text{G49})$$

Since there are projectors in v_α , we have

$$= \sum_{\alpha\beta\gamma\delta} \left[\begin{array}{c} \rho \\ v_\gamma^* \quad \delta_{\gamma\delta} \quad v_\delta \\ \delta_{\alpha\gamma} \quad \delta_{\beta\delta} \\ v_\alpha \quad \delta_{\alpha\beta} \quad v_\beta^* \\ \rho \end{array} \right]. \quad (\text{G50})$$

From Haar averaging, one of (1) $\alpha = \beta$, $\gamma = \delta$, $\alpha \neq \gamma$, (2) $\alpha = \gamma$, $\beta = \delta$, $\alpha \neq \beta$, or (3) $\alpha = \beta = \gamma = \delta$ must be satisfied. (1) and (2) are disallowed by projectors. Therefore only (3) is allowed, and we have

$$\xrightarrow{\mathbb{E}[\cdot]} \sum_{\alpha} \mathbb{E} \left[\begin{array}{c} \rho \\ v_\alpha^* \quad v_\alpha \\ v_\alpha \quad v_\alpha^* \\ \rho \end{array} \right]. \quad (\text{G51})$$

Therefore, we have

$$\mathbb{E}_U [\|\text{tr}_K(\mathcal{U} - \mathcal{D}_{m_1} \oplus \dots \oplus \mathcal{D}_{m_B})\rho\|_F]^2 \leq \sum_{\alpha} \text{tr}_{N \cap K} [\text{tr}_{N \setminus K} [\chi_{N_\alpha}]^2] \cdot \text{tr}[\mathcal{P}_\alpha[\rho]^2] = \sum_{\alpha} \text{tr}_{N \cap K} \left[\text{tr}_{N \setminus K} \left[\frac{\mathbb{P}_\alpha^N}{N_\alpha} \right]^2 \right] \cdot \text{tr}[\mathcal{P}_\alpha[\rho]^2] \quad (\text{G52})$$

$$\leq \sum_{\alpha} \text{tr}_{N \cap K} \left[\text{tr}_{N \setminus K} \left[\frac{\mathbb{P}_\alpha^N}{n_\alpha m} \right]^2 \right] \leq \sum_{\alpha} m \frac{1}{m^2} \leq \frac{B}{m}. \quad (\text{G53})$$

■

Corollary G3. For the $U(1)$ -symmetric gate acting on two qubit-qudit pairs and $\mathbf{K} = \mathbf{G} \cup \mathbf{S}(\cdot)$, we have

$$\mathbb{E}_u [\|\mathrm{tr}_{\mathbf{K}}(\mathcal{U}_{U(1)} - \mathcal{E}_{\mathrm{SSEP}} \otimes \mathcal{D}_{\mathrm{qudits}})\rho\|_{\mathrm{F}}] \leq \frac{\sqrt{3}}{d}. \quad (\mathrm{G54})$$

Lemma G4. Consider $(\mathcal{D}_{N_1} \oplus \cdots \oplus \mathcal{D}_{N_B})$ that decomposes as $\mathcal{D}_m \otimes (\mathcal{D}_{n_1} \oplus \cdots \oplus \mathcal{D}_{n_B})$, such that $\sum_{\alpha} n_{\alpha} = n$ and $nm = N$, $n \otimes m = \mathbf{N}$, and $\mathcal{D}_m$ acts on $\mathbf{m}$. Then, we have

$$\left\| \mathrm{tr}_{\mathbf{K}} \left(\mathcal{D}_m \otimes \bigoplus_{\alpha} \mathcal{D}_{n_{\alpha}} \right) A \right\|_{\mathrm{F}} \leq \frac{\sqrt{n|\mathbf{K} \cap \mathbf{n}|}}{|\mathbf{m} \setminus \mathbf{K}|} \left(\sum_{\alpha} \frac{1}{n_{\alpha}} \right) \|\mathrm{tr}_m \mathrm{tr}_{\mathbf{K} \setminus \mathbf{n}}[A]\|_{\mathrm{F}}. \quad (\mathrm{G55})$$

Proof. Then, we have that

$$\left\| \mathrm{tr}_{\mathbf{K}} \left(\mathcal{D}_m \otimes \bigoplus_{\alpha} \mathcal{D}_{n_{\alpha}} \right) A \right\|_{\mathrm{F}}^2 = \left\| \mathrm{tr}_{\mathbf{K}} \left(\mathcal{D}_m \otimes \bigoplus_{\alpha} \mathcal{D}_{n_{\alpha}} \right) [A] \right\|_{\mathrm{F}}^2 \quad (\mathrm{G56})$$

$$= \left\| \mathrm{tr}_{\mathbf{K} \cap \mathbf{m}} \chi_m \otimes \mathrm{tr}_{\mathbf{K} \cap \mathbf{n}} \left(\bigoplus_{\alpha} \mathcal{D}_{n_{\alpha}} \right) \mathrm{tr}_m \mathrm{tr}_{\mathbf{K} \setminus \mathbf{n}}[A] \right\|_{\mathrm{F}}^2 \quad (\mathrm{G57})$$

$$= \frac{1}{|\mathbf{m} \setminus \mathbf{K}|} \left\| \mathrm{tr}_{\mathbf{K} \cap \mathbf{n}} \sum_{\alpha} \frac{\mathbb{P}_{\alpha}^n}{n_{\alpha}} \otimes \mathrm{tr}_n \mathcal{P}_{\alpha} \mathrm{tr}_m \mathrm{tr}_{\mathbf{K} \setminus \mathbf{n}}[A] \right\|_{\mathrm{F}}^2 \quad (\mathrm{G58})$$

$$\leq \frac{1}{|\mathbf{m} \setminus \mathbf{K}|} \left(\sum_{\alpha} \left\| \mathrm{tr}_{\mathbf{K} \cap \mathbf{n}} \frac{\mathbb{P}_{\alpha}^n}{n_{\alpha}} \otimes \mathrm{tr}_n [\mathcal{P}_{\alpha} \mathrm{tr}_m \mathrm{tr}_{\mathbf{K} \setminus \mathbf{n}}[A]] \right\|_{\mathrm{F}} \right)^2 \quad (\mathrm{G59})$$

$$\leq \frac{1}{|\mathbf{m} \setminus \mathbf{K}|} \left(\sum_{\alpha} \frac{\sqrt{|\mathbf{K} \cap \mathbf{n}|}}{n_{\alpha}} \|\mathrm{tr}_n [\mathbb{P}_{\alpha} \mathrm{tr}_m \mathrm{tr}_{\mathbf{K} \setminus \mathbf{n}}[A]]\|_{\mathrm{F}} \right)^2 \quad (\mathrm{G60})$$

$$\leq \frac{1}{|\mathbf{m} \setminus \mathbf{K}|} \left(\sum_{\alpha} \frac{\sqrt{|\mathbf{K} \cap \mathbf{n}|}}{n_{\alpha}} \sqrt{n} \|\mathbb{P}_{\alpha} \mathrm{tr}_m \mathrm{tr}_{\mathbf{K} \setminus \mathbf{n}}[A]\|_{\mathrm{F}} \right)^2. \quad (\mathrm{G61})$$

Now observe that $\|\mathbb{P}_{\alpha} A\|_{\mathrm{F}}^2 = \mathrm{tr}[A^{\dagger} \mathbb{P}_{\alpha} A] \geq 0$. Also, we have that $\|A^{\dagger} A\|_{\mathrm{F}}^2 = \mathrm{tr}[A^{\dagger} A] = \sum_{\alpha} \mathrm{tr}[A^{\dagger} \mathbb{P}_{\alpha} A] = \sum_{\alpha} \|\mathbb{P}_{\alpha} A\|_{\mathrm{F}}^2$. Therefore, $\|\mathbb{P}_{\alpha} A\|_{\mathrm{F}} \leq \|A^{\dagger} A\|_{\mathrm{F}}$. Therefore, we have

$$\left\| \mathrm{tr}_{\mathbf{K}} \left(\mathcal{D}_m \otimes \bigoplus_{\alpha} \mathcal{D}_{n_{\alpha}} \right) A \right\|_{\mathrm{F}} \leq \frac{1}{|\mathbf{m} \setminus \mathbf{K}|} \sum_{\alpha} \frac{\sqrt{|\mathbf{K} \cap \mathbf{n}|}}{n_{\alpha}} \sqrt{n} \|\mathrm{tr}_m \mathrm{tr}_{\mathbf{K} \setminus \mathbf{n}}[A]\|_{\mathrm{F}} \quad (\mathrm{G62})$$

$$= \frac{\sqrt{n|\mathbf{K} \cap \mathbf{n}|}}{|\mathbf{m} \setminus \mathbf{K}|} \left(\sum_{\alpha} \frac{1}{n_{\alpha}} \right) \|\mathrm{tr}_m \mathrm{tr}_{\mathbf{K} \setminus \mathbf{n}}[A]\|_{\mathrm{F}}. \quad (\mathrm{G63})$$

■

Corollary G4. For the SSEP gate, we have

$$\left\| \mathrm{tr}_{\mathbf{K}} \mathbb{E}_u [\mathcal{U}] A \right\|_{\mathrm{F}} \leq 10 \|\mathrm{tr}_m \mathrm{tr}_{\mathbf{K} \setminus \mathbf{n}}[A]\|_{\mathrm{F}}, \quad (\mathrm{G64})$$

and therefore $r_{\mathrm{unitary}} = 10$.

Lemma G5. Consider a measurement map $\mathcal{M}$ that acts on $\mathbf{S}(i)$. Then, we have

$$\mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{K}}[\mathcal{M}[\rho_{\tau}] - \mathcal{M}[\bar{\rho}_{\tau}]]\|_{\mathrm{F}} \leq |\mathbf{S}(i)| \left(\sqrt{|\mathbf{K} \cap \mathbf{S}(i)|} \mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{K} \setminus \mathbf{S}(i)}(\rho_{\tau} - \bar{\rho}_{\tau})\|_{\mathrm{F}} + \mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{S}(i)}(\rho_{\tau} - \bar{\rho}_{\tau})\|_{\mathrm{F}} \right). \quad (\mathrm{G65})$$

For $\mathbf{K} = \emptyset$, $|\mathbf{K} \cap \mathbf{S}(i)| \rightarrow 1$.

Proof. For $\mathcal{Q}[\rho] := \mathcal{Q}\rho\mathcal{Q}$,

$$\mathbb{E}_{U,Y} \|\mathrm{tr}_{\mathbf{K}}[\mathcal{M}[\rho_{\tau}] - \mathcal{M}[\bar{\rho}_{\tau}]]\|_{\mathrm{F}} \quad (\mathrm{G66})$$

$$= \mathbb{E}_{U,Y} \left\| \frac{\mathrm{tr}_{\mathbf{K}}[\mathcal{Q}\rho_{\tau}]}{\mathrm{tr}[\mathcal{Q}\rho_{\tau}]} - \frac{\mathrm{tr}_{\mathbf{K}}[\mathcal{Q}\bar{\rho}_{\tau}]}{\mathrm{tr}[\mathcal{Q}\bar{\rho}_{\tau}]} \right\|_{\mathrm{F}} = \mathbb{E}_{U,Y} \left[\frac{1}{\mathrm{tr}[\mathcal{Q}\rho_{\tau}]\mathrm{tr}[\mathcal{Q}\bar{\rho}_{\tau}]} \|\mathrm{tr}[\mathcal{Q}\bar{\rho}_{\tau}]\mathrm{tr}_{\mathbf{K}}[\mathcal{Q}\rho_{\tau}] - \mathrm{tr}[\mathcal{Q}\rho_{\tau}]\mathrm{tr}_{\mathbf{K}}[\mathcal{Q}\bar{\rho}_{\tau}]\|_{\mathrm{F}} \right] \quad (\mathrm{G67})$$

$$= \mathbb{E}_{U,Y} \left[\frac{1}{\mathrm{tr}[\mathcal{Q}\rho_{\tau}]\mathrm{tr}[\mathcal{Q}\bar{\rho}_{\tau}]} \|\mathrm{tr}[\mathcal{Q}\bar{\rho}_{\tau}]\mathrm{tr}_{\mathbf{K}}[\mathcal{Q}(\rho_{\tau} - \bar{\rho}_{\tau})] - (\mathrm{tr}[\mathcal{Q}\rho_{\tau}] - \mathrm{tr}[\mathcal{Q}\bar{\rho}_{\tau}])\mathrm{tr}_{\mathbf{K}}[\mathcal{Q}\bar{\rho}_{\tau}]\|_{\mathrm{F}} \right] \quad (\mathrm{G68})$$

$$\leq \mathbb{E}_{U,Y} \left[\frac{1}{\text{tr}[\mathcal{Q}\rho_\tau]\text{tr}[\mathcal{Q}\bar{\rho}_\tau]} (\text{tr}[\mathcal{Q}\bar{\rho}_\tau] \|\text{tr}_K[\mathcal{Q}(\rho_\tau - \bar{\rho}_\tau)]\|_F + |\text{tr}[\mathcal{Q}(\rho_\tau - \bar{\rho}_\tau)]| \|\text{tr}_K[\mathcal{Q}\bar{\rho}_\tau]\|_F) \right] \quad (\text{G69})$$

$$\leq \mathbb{E}_{U,Y} \left[\frac{1}{\text{tr}[\mathcal{Q}\rho_\tau]\text{tr}[\mathcal{Q}\bar{\rho}_\tau]} (\text{tr}[\mathcal{Q}\bar{\rho}_\tau] \|\text{tr}_K[\mathcal{Q}(\rho_\tau - \bar{\rho}_\tau)]\|_F + \|\mathcal{Q}_{\text{loc}}\|_F \|\text{tr}_{S(i)}(\rho_\tau - \bar{\rho}_\tau)\|_F \|\text{tr}_K[\mathcal{Q}\bar{\rho}_\tau]\|_F) \right]. \quad (\text{G70})$$

We first look at $\|\text{tr}_K[\mathcal{Q}\bar{\rho}_\tau]\|_F^2$. Note that $\bar{\rho}_\tau$ is a completely diagonal density matrix. Therefore,

$$\|\text{tr}_K[\mathcal{Q}\bar{\rho}_\tau]\|_F^2 = \text{tr}[(\text{tr}_K \mathcal{Q} \bar{\rho}_\tau)(\text{tr}_K \mathcal{Q} \bar{\rho}_\tau)] = \sum_{s,s',s'',s'''} \sqrt{p(y|s)p(y|s')p(y|s'')p(y|s''')} \text{tr}[\text{tr}_K[\mathbb{P}_s \bar{\rho}_\tau \mathbb{P}_{s''}] \text{tr}_K[\mathbb{P}_{s'} \bar{\rho}_\tau \mathbb{P}_{s'''}]], \quad (\text{G71})$$

where s, s', s'', s''' are the degrees of freedom involved in the measurement. Now $\bar{\rho}_\tau$ is completely diagonal. Therefore, this gives $\delta_{s,s'}, \delta_{s'',s'''}:$

$$\|\text{tr}_K[\mathcal{Q}\bar{\rho}_\tau]\|_F^2 = \sum_{s,s''} p(y|s)p(y|s'') \text{tr}[\text{tr}_K[\mathbb{P}_s \bar{\rho}_\tau \mathbb{P}_s] \text{tr}_K[\mathbb{P}_{s''} \bar{\rho}_\tau \mathbb{P}_{s''}]]. \quad (\text{G72})$$

Now $\mathbb{P}_s = \mathbb{P}_{s \setminus K} \mathbb{P}_{s \cap K}$, and $\mathbb{P}_{s \setminus K}$ commutes through the trace. Therefore,

$$\|\text{tr}_K[\mathcal{Q}\bar{\rho}_\tau]\|_F^2 = \sum_{s,s''} p(y|s)p(y|s'') \text{tr}[\mathbb{P}_{s \setminus K} \text{tr}_K[\mathbb{P}_{s \cap K} \bar{\rho}_\tau \mathbb{P}_{s \cap K}] \mathbb{P}_{s \setminus K} \mathbb{P}_{s'' \setminus K} \text{tr}_K[\mathbb{P}_{s'' \cap K} \bar{\rho}_\tau \mathbb{P}_{s'' \cap K}] \mathbb{P}_{s'' \setminus K}]. \quad (\text{G73})$$

Using that $\bar{\rho} = \sum_{s,g} \bar{\rho}(s,g) |s,g\rangle \langle s,g|$, where $|s\rangle \in \mathbf{S}$, $|g\rangle \in \mathbf{G}$,

$$\|\text{tr}_K[\mathcal{Q}\bar{\rho}_\tau]\|_F^2 = \sum_{s,s''} p(y|s)p(y|s'') \delta_{s \setminus K, s'' \setminus K} \bar{\rho}(s) \bar{\rho}(s'') \leq \sum_{s,s''} p(y|s)p(y|s'') \bar{\rho}(s) \bar{\rho}(s'') = \bar{\rho}(y)^2 = \text{tr}[\mathcal{Q}\bar{\rho}_\tau]^2. \quad (\text{G74})$$

Therefore,

$$\mathbb{E}_{U,Y} \|\text{tr}_K[\mathcal{M}[\rho_\tau] - \mathcal{M}[\bar{\rho}_\tau]]\|_F \quad (\text{G75})$$

$$\leq \mathbb{E}_{U,Y} \left[\frac{1}{\text{tr}[\mathcal{Q}\rho_\tau]\text{tr}[\mathcal{Q}\bar{\rho}_\tau]} (\text{tr}[\mathcal{Q}\bar{\rho}_\tau] \|\text{tr}_K[\mathcal{Q}(\rho_\tau - \bar{\rho}_\tau)]\|_F + \|\mathcal{Q}^{S(i)}\|_F \|\text{tr}_{S(i)}(\rho_\tau - \bar{\rho}_\tau)\|_F \text{tr}[\mathcal{Q}\bar{\rho}_\tau]) \right] \quad (\text{G76})$$

$$= \mathbb{E}_{U,Y} \left[\frac{1}{\text{tr}[\mathcal{Q}\rho_\tau]} (\|\text{tr}_K[\mathcal{Q}(\rho_\tau - \bar{\rho}_\tau)]\|_F + \|\mathcal{Q}^{S(i)}\|_F \|\text{tr}_{S(i)}(\rho_\tau - \bar{\rho}_\tau)\|_F) \right]. \quad (\text{G77})$$

Now consider

$$\|\text{tr}_K[\mathcal{Q}(\rho_\tau - \bar{\rho}_\tau)]\|_F = \|\text{tr}_{K \cap S(i)}[\mathcal{Q}^{S(i)} \text{tr}_{K \setminus S(i)}(\rho_\tau - \bar{\rho}_\tau)]\|_F \leq \|\text{tr}_{K \cap S(i)}[\mathcal{Q}^{S(i)} \text{tr}_{K \setminus S(i)}(\rho_\tau - \bar{\rho}_\tau)]\|_F \quad (\text{G78})$$

$$\leq \sqrt{|K \cap S(i)|} \|\mathcal{Q}^{S(i)} \text{tr}_{K \setminus S(i)}(\rho_\tau - \bar{\rho}_\tau)\|_F \leq \sqrt{|K \cap S(i)|} p_{\max} \|\text{tr}_{K \setminus S(i)}(\rho_\tau - \bar{\rho}_\tau)\|_F, \quad (\text{G79})$$

where $p_{\max} = \max_s p(y|s)$, and the third line is due to Rastegin '12. Similarly, we have $\|\mathcal{Q}^{S(i)}\|_F \leq p_{\max}$.

Notice that $p(y_\tau) = \text{tr}[\mathcal{Q}\rho_\tau]$, such that $\mathbb{E}_{y_\tau} [\frac{1}{\text{tr}[\mathcal{Q}\rho_\tau]}] = \int dy_\tau [\cdot]$. Therefore,

$$\mathbb{E}_{U,Y} \|\text{tr}_K[\mathcal{M}[\rho_\tau] - \mathcal{M}[\bar{\rho}_\tau]]\|_F \leq \left(\int_y p_{\max}(y) \right) \mathbb{E}_{U,Y:1:\tau-1} [\sqrt{|K \cap S(i)|} \|\text{tr}_{K \setminus S(i)}(\rho_\tau - \bar{\rho}_\tau)\|_F + \|\text{tr}_{S(i)}(\rho_\tau - \bar{\rho}_\tau)\|_F]. \quad (\text{G80})$$

Now $p_{\max}(y) = \max_s p(y|s) \leq \sum_s p(y|s)$. Therefore, $\int_y p_{\max}(y) \leq \sum_s \int_y p(y|s) = \sum_s 1 = |\mathbf{S}(i)|$. Therefore,

$$\mathbb{E}_{U,Y} \|\text{tr}_K[\mathcal{M}[\rho_\tau] - \mathcal{M}[\bar{\rho}_\tau]]\|_F \leq |\mathbf{S}(i)| \left(\sqrt{|K \cap S(i)|} \mathbb{E}_{U,Y} \|\text{tr}_{K \setminus S(i)}(\rho_\tau - \bar{\rho}_\tau)\|_F + \mathbb{E}_{U,Y} \|\text{tr}_{S(i)}(\rho_\tau - \bar{\rho}_\tau)\|_F \right). \quad (\text{G81})$$

■

Corollary G5. For on-site measurements, we have

$$\mathbb{E}_{U,Y} \|\text{tr}_K[\mathcal{M}[\rho_\tau] - \mathcal{M}[\bar{\rho}_\tau]]\|_F \leq 2 \left(\sqrt{2} \mathbb{E}_{U,Y} \|\text{tr}_{K \setminus S(i)}(\rho_\tau - \bar{\rho}_\tau)\|_F + \mathbb{E}_{U,Y} \|\text{tr}_{S(i)}(\rho_\tau - \bar{\rho}_\tau)\|_F \right), \quad (\text{G82})$$

and therefore $r_{\text{meas}} = 2(\sqrt{2} + 1)$.

[1] U. Agrawal, A. Zabalo, K. Chen, J. H. Wilson, A. C. Potter, J. Pixley, S. Gopalakrishnan, and R. Vasseur, Entanglement and charge-sharpening transitions in U(1) symmetric monitored quantum circuits, *Phys. Rev. X* **12**, 041002 (2022).

[2] 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).
[3] F. Barratt, U. Agrawal, S. Gopalakrishnan, D. A. Huse, R. Vasseur, and A. C. Potter, Field theory of charge sharpening

- in symmetric monitored quantum circuits, *Phys. Rev. Lett.* **129**, 120604 (2022).
- [4] 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).
- [5] Y. Bao, S. Choi, and E. Altman, Theory of the phase transition in random unitary circuits with measurements, *Phys. Rev. B* **101**, 104301 (2020).
- [6] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, *J. Math. Phys.* **43**, 4452 (2002).
- [7] L. Rabiner and B. Juang, An introduction to hidden Markov models, *Ieee Assp Magazine* **3**, 4 (1986).
- [8] S. W. P. Kim and A. Lamacraft, Planted directed polymer: Inferring a random walk from noisy images, *Phys. Rev. E* **111**, 024135 (2025).
- [9] R. L. Stratonovich and V. Tikhonov, *Non-Linear Transformations of Stochastic Processes* (Pergamon Press, Oxford, London, 1965).
- [10] A. Nahum and J. L. Jacobsen, Bayesian critical points in classical lattice models, *Phys. Rev. B* **112**, 235113 (2025).
- [11] L. Zdeborová and F. Krzakala, Statistical physics of inference: Thresholds and algorithms, *Adv. Phys.* **65**, 453 (2016).
- [12] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning* (Springer Science+Business Media, LLC, New York, NY, 2006), Vol. 4.
- [13] H. Pishro-Nik, *Introduction to Probability, Statistics, and Random Processes* (Kappa Research, LLC, Blue Bell, PA, 2014), Chap. 9.1.5.
- [14] B. Skinner, J. Ruhman, and A. Nahum, Measurement-induced phase transitions in the dynamics of entanglement, *Phys. Rev. X* **9**, 031009 (2019).
- [15] A. Monras, A. Beige, and K. Wiesner, Hidden quantum Markov models and non-adaptive read-out of many-body states, *arXiv:1002.2337*.
- [16] B. Collins, S. Matsumoto, and J. Novak, The weingarten calculus, *Not. Ame. Math. Soc.* **69**, 1 (2022).
- [17] A. Barchielli, L. Lanz, and G. M. Prosperi, A model for the macroscopic description and continual observations in quantum mechanics, *Nuovo Cim. B* **72**, 79 (1982).
- [18] K. Binder and D. W. Heermann, Theoretical foundations of the Monte Carlo method and its applications in statistical physics, in *Monte Carlo Simulation in Statistical Physics*, edited by W. T. Rhodes and H. E. Stanley, Graduate Texts in Physics (Springer, Berlin, Heidelberg, 2010), pp. 5–67.
- [19] L. Onsager, Crystal statistics. I. A two-dimensional model with an order-disorder transition, *Phys. Rev.* **65**, 117 (1944).
- [20] F. Merz and J. T. Chalker, Two-dimensional random-bond Ising model, free fermions, and the network model, *Phys. Rev. B* **65**, 054425 (2002).
- [21] 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: Math. Theor.* **48**, 095001 (2015).
- [22] K. Fujii and Y. Tokunaga, Error and loss tolerances of surface codes with general lattice structures, *Phys. Rev. A* **86**, 020303(R) (2012).
- [23] J. Edmonds, Paths, trees, and flowers, *Can. J. Math.* **17**, 449 (1965).
- [24] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Clarendon Press, Oxford, UK, 2001), Vol. 111, p. 17.
- [25] D. S. Fisher and D. A. Huse, Equilibrium behavior of the spin-glass ordered phase, *Phys. Rev. B* **38**, 386 (1988).
- [26] T. M. Liggett, *Interacting Particle Systems* (Springer Science & Business Media, Berlin, Heidelberg, 1985), Vol. 2, pp. 361–417.
- [27] A. Altland and B. D. Simons, *Condensed Matter Field Theory* (Cambridge University Press, Cambridge, UK, 2010), pp. 134–142.
- [28] D. Tong, *Lectures on Statistical Field Theory* (Cambridge University, Cambridge, UK, 2017), p. 125.
- [29] G. Vidal, Efficient simulation of one-dimensional quantum many-body systems, *Phys. Rev. Lett.* **93**, 040502 (2004).
- [30] M. Fishman, S. R. White, and E. M. Stoudenmire, The ITensor software library for tensor network calculations, *SciPost Phys. Codebases* **4** (2022).
- [31] A. Nahum, S. Roy, B. Skinner, and J. Ruhman, Measurement and entanglement phase transitions in all-to-all quantum circuits, on quantum trees, and in landau-ginsburg theory, *PRX Quantum* **2**, 010352 (2021).
- [32] X. Feng, B. Skinner, and A. Nahum, Measurement-induced phase transitions on dynamical quantum trees, *PRX Quantum* **4**, 030333 (2023).
- [33] X. Feng, N. Fishchenko, S. Gopalakrishnan, and M. Ippoliti, Charge and spin sharpening transitions on dynamical quantum trees, *Quantum* **9**, 1692 (2025).
- [34] B. Derrida and H. Spohn, Polymers on disordered trees, spin glasses, and traveling waves, *J. Stat. Phys.* **51**, 817 (1988).
- [35] This is the solution to $(1 - 2\lambda)\lambda + (2 + \lambda - \lambda^2) \ln(3/[2 + \lambda - \lambda^2])$, which comes from minimizing Eq. (102) while ignoring the first term in the RHS.
- [36] 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).
- [37] M. Christandl and R. Renner, Reliable quantum state tomography, *Phys. Rev. Lett.* **109**, 120403 (2012).
- [38] A. deMarti iOlius, P. Fuentes, R. Orús, P. M. Crespo, and J. Etchezarreta Martinez, Decoding algorithms for surface codes, *Quantum* **8**, 1498 (2024).
- [39] J. F. Kam, S. Gicev, K. Modi, A. Southwell, and M. Usman, Detrimental non-Markovian errors for surface code memory, *Quantum Sci. Technol.* **10**, 035060 (2025).