

# Large $N$ matrix quantum mechanics as a quantum memory

ChunJun Cao<sup>1,2</sup>, Gong Cheng<sup>3</sup>, and Brian Swingle<sup>4</sup>

<sup>1</sup>*Joint center for Quantum information and Computer science, University of Maryland, College Park, Maryland 20740, USA*

<sup>2</sup>*Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, California 91125, USA*

<sup>3</sup>*Maryland Center for Fundamental Physics, University of Maryland, College Park, Maryland 20740, USA*

<sup>4</sup>*Brandeis University, Waltham, Massachusetts 02453, USA*

 (Received 16 March 2023; accepted 13 September 2023; published 10 October 2023)

In this paper, we explore the possibility of building a quantum memory that is robust to thermal noise using large  $N$  matrix quantum mechanics models. First, we investigate the gauged  $SU(N)$  matrix harmonic oscillator and different ways to encode quantum information in it. By calculating the mutual information between the system and a reference which purifies the encoded information, we identify a transition temperature,  $T_c$ , below which the encoded quantum information is protected from thermal noise for a memory time scaling as  $N^2$ . Conversely, for temperatures higher than  $T_c$ , the information is quickly destroyed by thermal noise. Second, we relax the requirement of gauge invariance and study a matrix harmonic oscillator model with only global symmetry. Finally, we further relax even the symmetry requirement and propose a model that consists of a large number  $N^2$  of qubits, with interactions derived from an approximate  $SU(N)$  symmetry. In both ungauged models, we find that the effects of gauging can be mimicked using an energy penalty to give a similar result for the memory time. The final qubit model also has the potential to be realized in the laboratory.

DOI: [10.1103/PhysRevD.108.086008](https://doi.org/10.1103/PhysRevD.108.086008)

## I. INTRODUCTION

Quantum information is typically fragile when encountered in familiar physical systems, thus giving rise to the intuition that quantum information is inherently hard to preserve at macroscopic scales, e.g. in large systems or at long times. The discovery of quantum error correction showed that this intuition is incomplete: in the right kind of physical system, quantum information can be retained for an arbitrarily long time. Maintaining the information is typically an active process where we keep a vigil for errors and periodically correct them. An error correcting code operating in this mode forms an active quantum memory. Some codes may have the additional capability to preserve the encoded information for a long time naturally, without any active intervention. A code operating in this mode is called a passive or self-correcting quantum memory—it is like a quantum hard drive. In this paper and a forthcoming companion paper, we search for passive quantum memories in the context of quantum gauge theories and models inspired by them.

Why gauge theories? For one thing, some of the most important currently known codes can be viewed as gauge theories. For example, the toric code can be viewed as a  $Z_2$  lattice gauge theory with dynamical electric charges in the extreme deconfined limit. More generally, deconfined lattice discrete gauge theories are associated with a high

degree of quantum entanglement and topological order, both of which are linked with quantum error correction [1–3]. One feature of topological order is a robustness of the ground space to perturbations of the Hamiltonian, which is suggestive of the potential for passive error correction. And while the two-dimensional toric code does not form a passive memory [4], a four-dimensional analog does [5]. Moreover, phases of matter exhibiting topological order associated with a discrete non-Abelian group can be used for fault tolerant quantum computation, a fact that also hints at a degree of intrinsic protection [6].

From another perspective, quantum error correction has been linked with quantum gravity in the context of holographic duality, also known as the AdS/CFT correspondence [7]. The best understood examples of the correspondence involve continuous non-Abelian gauge theories with large gauge groups, thus providing another connection between non-Abelian gauge symmetry and quantum error correction. In this context, one also finds that the boundary conformal field theory with  $SU(N)$  gauge symmetry can possess a low temperature confined phase which is qualitatively distinct from the generic high temperature physics in the deconfined phase. This boundary confinement-deconfinement phase transition is dual to the Hawking-Page transition [8] between a thermal AdS geometry and a AdS black hole geometry in the bulk [9]. It was suggested in [10] that when treating the low energy

sector of AdS/CFT as an approximate quantum error-correcting code under the influence of thermal errors, the encoded information living in the bulk thermal AdS geometry can be disturbed by the scattering of thermal particles with probability  $O(1/N^2)$ . However, in the high temperature phase, the encoded information is scrambled by the black hole, thus inducing a large logical error. In the large  $N$  limit, the distinct phases correspond to the existence of a threshold, such that logical information remains robust in the low temperature phase where the errors occur less frequently.<sup>1</sup> Therefore, it is natural to explore whether this type of robustness also extends to other theories with such gauge symmetries.

With these motivations, we have undertaken an investigation of quantum error correction in non-Abelian gauge theories with large gauge groups. We focus here on  $SU(N)$  groups with  $N$  taken large. This large  $N$  limit is like a thermodynamic limit, and we expect that passive error correction can only emerge in such a thermodynamic limit. We define and study quantum codes in a variety of models, including models inspired by gauge theories but where we relax the gauge constraint. We are seeking to answer two key questions: (1) are there potentially useful lab-constructible quantum memories built from gauge theories or derived models and (2) can naturally occurring systems, e.g. QCD or a hypothetical holographic dual to quantum gravity in our Universe, naturally preserve quantum information for a long time? This paper focuses on quantum mechanical systems with large gauge groups, while the companion paper [12] will consider quantum field theories with large gauge groups.

The key results of the present paper are as follows. First, we discuss broadly how a property called large  $N$  factorization is related to an approximate version of the Knill-Laflamme error correction condition [13,14]. Assuming a code space with the right properties can be defined and coupled to a thermal bath satisfying reasonable assumptions, we then prove a theorem establishing the existence of a memory lifetime which scales polynomially with  $N$ . We then consider a sequence of models of increasing generality and show how the conditions of the main theorem can be satisfied. For a certain fully gauged model, we find that the conditions of the theorem are immediately satisfied; for models where we do not gauge the symmetry or even make the symmetry approximate, an energy penalty scaling like  $\log N$  is needed to meet the conditions of the main theorem. While here we focus on the construction of such quantum

memories and their memory times, there are other aspects of the model that are relevant for practical implementation, such as whether quantum information can be efficiently encoded, recovered, and processed. We will leave these discussions for future work.

In the remainder of the introduction, we first review the basic notions for readers without the necessary background in quantum codes or quantum gauge theories. We then give a high level overview of our results. We also note that a few prior works have considered the role of large  $N$  factorization and/or gauge symmetry in defining quantum error correcting codes [15,16]. There is also a large literature on passive quantum memories from a variety of approaches [2,5,17,18]. See also [19] for a more comprehensive review.

### A. Background on quantum codes

In this subsection, we review the key concepts of quantum error correction needed for our discussion. This section is aimed at readers with a more high energy background.

Mathematically, a quantum error correcting code is a subspace, the *logical space*  $\mathcal{C}$ , of a larger *physical Hilbert space*  $\mathcal{H}$ . Errors are modeled by a quantum process  $\mathcal{E}$  and error correction is given by another process  $\mathcal{R}$  such that

$$\mathcal{R}(\mathcal{E}(\rho)) = \rho \quad \forall \rho \in \mathcal{C}. \quad (1)$$

If the error channel is viewed as probabilistically applying errors from a set  $\{E_a\}$ , then one way to formulate the necessary condition for the existence of an error correction process is the Knill-Laflamme (KL) condition,

$$\langle i|E_a^\dagger E_b|j\rangle = c_{ab}\delta_{ij}, \quad (2)$$

where  $|i\rangle$  and  $|j\rangle$  are any two elements of a basis for  $\mathcal{C}$ . The intuition behind KL is that errors are unable to cause transitions between distinct states in the code. We can also relax these exact equalities to obtain instead an approximate quantum code.

In practice, such a code is instantiated within a concrete physical substrate, with  $\mathcal{H}$  being some model of the physical space of states of the substrate. For example, a code ( $\mathcal{C}$ ) might be implemented as a subspace of the full energy levels ( $\mathcal{H}$ ) of a superconducting circuit (substrate). In some cases, the code space might correspond to the ground space of some Hamiltonian for the substrate, but this is not required in general.

A code is used as follows. Information is first encoded using some encoding operation. This operation might correspond to a quantum circuit which takes a fixed initial state in  $\mathcal{H}$  and produces the desired code state in  $\mathcal{C}$ . After encoding, the code is subject to errors, corresponding to one application of the error channel  $\mathcal{E}$ . Then the recovery procedure  $\mathcal{R}$  is applied to recover the initially encoded information. The condition above guarantees success

<sup>1</sup>Note that unlike the tensor network constructions [11] where one changes the system size by increasing the number of layers, which roughly corresponds to changing the IR cutoff in the bulk or the UV cutoff on the boundary theory when it is put on a lattice, the size of the physical system in [10] is altered by  $N$  of the  $SU(N)$  symmetry while keeping the number of “boundary sites” fixed. Heuristically, the physical Hilbert space can be taken as placing  $O(N^2)$  qudits on each boundary site.

provided the encoding prepares a state in  $\mathcal{C}$ . Finally, the information can be read out by a decoding operation. Note the distinction between recovery, which repairs the information but keeps it encoded, and decoding, which reverses the encoding map when applied to an error free state.<sup>2</sup>

To make these ideas more concrete, we can consider the simple goal of information storage for a period of time (as opposed to any active computation on the encoded data). We think of the code during the error part of the protocol as sitting idle in the presence of an undesirable environment. The effect of the environment is what produces errors in the code, so the channel  $\mathcal{E}$  implicitly contains information about the nature of the environment and how long the code is exposed to the environment. We denote such an error channel by  $\mathcal{E}_t$ , where  $t$  corresponds to the time for which the code is exposed to the environment. Correspondingly, the recovery operation  $\mathcal{R}$  may depend on  $t$  and indeed on the whole structure of  $\mathcal{E}_t$ .

In this paper we consider the environment to be a thermal reservoir at some temperature  $T$ . There is still freedom in determining exactly how the reservoir couples to the system, but from an effective theory perspective we should expect all allowed couplings to be present with some strength. In this context, the need for active error correction can be phrased like this. Given a code  $\mathcal{C}$ , the code requires active error correction if, as  $t \rightarrow \infty$ , there is no  $\mathcal{R}$  that can correct  $\mathcal{E}_t$ . In other words, a code requires active error correction if, given enough time, the error channel will eventually degrade the information beyond the ability of any physically conceivable recovery operation to repair.

More generally, we can define a memory time  $t_{\text{mem}}$  as the time beyond which  $\mathcal{E}_t$  degrades the information beyond any hope of recovery. Of course, this time depends on the code and the error model, but it also depends on what recovery operations we can conceivably perform and on how well we want to be able to recover the information (e.g. the degree of approximation in our approximate code). A perfect passive memory is one for which  $t_{\text{mem}} = \infty$ . Typically, we can only hope for this sort of behavior in a thermodynamic limit, so we should really view a passive memory as a sequence of codes indexed by the physical size  $n$  for which  $\lim_{n \rightarrow \infty} t_{\text{mem}}(n) = \infty$ . Even more practically, what really matters is how  $t_{\text{mem}}$  scales with  $n$  and other parameters of the code.

For the purpose of potentially implementing a quantum memory in the lab, we may also want to impose some additional rules of the game. One variant of such rules are called the Caltech rules [20,21]. For example, we might obtain a long memory time by sending the energy cost of errors to infinity “by hand,” but this could be considered

<sup>2</sup>Note that, somewhat confusingly, decoding in the context of active error correction can also refer to applying some suitably chosen operations to the state based on syndrome measurements. This process typically leave the information encoded, which corresponds to our recovery operation in our current definition.

“cheating” unless there is a good physical reason for this large energy penalty. The idea is that we indeed want to effectively forbid errors, but this should happen naturally without sending any microscopic energy scale to infinity. Of course, if one does have a system where errors are particularly costly, then it might form a good enough passive memory in the sense that  $t_{\text{mem}}$  is long enough for practical purposes. To give another example, to be able to construct codes in physical three-dimensional space, we might want some restrictions on the kinds of codes we consider, e.g. trying to avoid the need for too many long-range interactions [21]. We note that all the models we consider here do have nonlocal interactions between the large  $N$  degrees of freedom.<sup>3</sup>

## B. Background on quantum gauge theories

In this subsection, we review the key concepts of quantum gauge theories needed for our discussion. This section is aimed at readers with a more quantum information background.

At the most basic level, a gauge theory is defined by giving a gauge group  $\mathcal{G}$  and a set of degrees of freedom transforming under  $\mathcal{G}$  with the requirement that all physical states and operators are invariant under  $\mathcal{G}$ . This invariance condition is known as Gauss’ law, in honor of its roots in electromagnetism. An immediate question is why one would use such a redundant description at all? Experience shows that one can always use a gauge-invariant description of the states and operators, but such a description often hides important underlying features of the physics. For example, looplike operators known as Wilson lines provide a class of gauge invariant observables in quantum field theories with gauge fields, but the extended nature of these operators can obscure aspects of locality in the theory. In special cases, one can even have an equivalence between a local gauge theory and another local theory with no explicit gauge structure, for example, the duality between the 3D statistical Ising model and 3D  $Z_2$  gauge theory.

We study such gauge theories in large part because they form a crucial part of our description of nature. The quantum theory of electromagnetism is a gauge theory where the gauge group is related to the Abelian Lie group  $U(1)$ . The more general Standard Model is also a gauge theory with a non-Abelian gauge group of the form  $SU(3) \times SU(2) \times U(1)$ . Gauge theories also arise in condensed matter systems, for example, when describing so-called spin liquid phases,

<sup>3</sup>Note that it is somewhat ambiguous what constitutes the local degrees of freedom in the large  $N$  gauged oscillator model as they are not simply tensor product of qubits. We show in later sections that there is a construction where two such gauged oscillators are spatially local with respect to each other. However, each one contains a large  $N$  degrees of freedom even though there is no sense in which they occupy an extended region in space. It is in this sense that these degrees of freedom are related to each other nonlocally.

and, as we mentioned above, they also have interesting error correction properties.

Apart from the motivations from known models of high-energy particle physics and condensed matter systems, another motivation to consider gauge theories comes from quantum gravity. Since the discovery of AdS/CFT, it is generally believed that certain  $d$ -dimensional  $SU(N)$  gauge theories with large  $N$  admit a dual geometric description in terms of a  $(d + 1)$ -dimensional theory of gravity. From a quantum information perspective, this correspondence provides the possibility of hiding quantum information deep inside the emergent geometry. This has inspired a lot of research in the field of holographic quantum error correcting codes [7], with explicit models such as the HaPPY code [11], the effective field theory code [22], and many other related constructions using tensor networks [23–25]. Current holographic codes often utilize the extra dimension, such that any local/UV errors can not access the information deep inside along the extra radial direction. However, very little of the error correcting power in these constructions come directly from the gauge symmetries in AdS/CFT. In this work, we will focus on a different kind of protection from the large  $N$  gauge symmetries, which can be related to the small Newton’s constant in these holographic theories [10,16]. This opens another possibility to design a new kind of quantum memory.

An interesting structure we exploit in the large  $N$  gauge theory is its sparse density of excitations at low energy. Due to Gauss’ law, the only allowed excitations of a  $SU(N)$  gauge theory on any compact manifold are gluons that combine into gauge singlets. Therefore, the density of states is determined by counting the number of gauge singlets in a given energy level. It is a general feature that the density of states in such model scales exponentially with some power of energy and does not depend on  $N$ . As is pointed out in [26], even for a free gauge theory with large  $N$ , the projection to the gauge invariant subspace induces effective interactions among the gluons, which is negligible at high energy but dominates at the low energy regime. In fact, for a weakly coupled  $SU(N)$  gauge theory, there is a phase transition analogous to the confinement/deconfinement transition in QCD, in the limit of  $N \rightarrow \infty$ . As we show, this induces a critical temperature,  $T_c$ , below which the quantum information can be long lived but above which the information is quickly destroyed. In the geometric description, this is believed to be dual to the Hawking-Page transition and the high temperature phase can be thought of as a black hole that kills the quantum information [10].

To zero in on just the effects of gauge symmetry, we distill only the necessary ingredients from the above gauge theory intuitions and consider a gauge theory with no spatial geometry at all. It turns out that even the simplest  $(0 + 1)$ -dimensional gauge theory of harmonic oscillators is sufficient for our purposes. The inverted harmonic oscillator has been studied as a dual to a two dimensional

string theory and can be considered as toy model of holography [27–30]. Here we start from a similar but noninverted harmonic oscillator model and try to build a quantum memory from it.

### C. Organization and overview of results

The rest of this paper is organized as follows. In Sec. II, we describe in detail the large  $N$  matrix quantum mechanics model with  $SU(N)$  gauge symmetry. We then discuss how it may be construed as an approximate error correcting code and define various bath models we will use for the rest of the paper. We then discuss the mutual information diagnostic for memory time and distill the core conclusions of our work in the form of Theorem 1. It shows that if a large  $N$  system has a sparse low energy spectrum, couples to the bath uniformly, and is an approximate QECC, then it has a memory time polynomial in  $N$  at low enough temperatures. In the ensuing sections, we deploy this theorem to various quantum memory constructions.

In Sec. III, we study two concrete examples of quantum memories built from two gauged harmonic oscillator models coupled to a thermal reservoir. The memory time-scales as  $N^2$  in both models. For the first model, the  $N^2$  life time persists even at high temperature, but the model requires nonlocal logical operations. We then construct a local model which has a  $N^2$  memory time in the low temperature phase, but requires a large coupling of strength  $\sim \log N$  between the gauged oscillators. At zero temperature limit  $T \rightarrow 0$ , the first model has a constant  $N^2$  life time, while the second model has a divergent life time that scales as  $N^2 e^{\frac{\omega}{T}}$ , where  $\omega$  is some energy scale of this model.

In Sec. IV, we consider what happens if the gauge constraint is not imposed exactly, but only as an energetic cost in the Hamiltonian. A large number of nonsinglet excitations now modify the physics, but the information retention time can be still be quadratic in  $N$  if the energy scale of nonsinglet excitations is large.

Finally, in Sec. V, we ask to what extent does robustness depend on the symmetry being exact and the Hilbert space of quantum systems being infinite dimensional. On the practical level, we want the quantum memories to be implementable on finite dimensional quantum systems like qubits or qudits. To this end, we consider an even more relaxed situation where the  $SU(N)$  symmetry itself is not only ungauged but also approximate. We then build a quantum memory consisting of  $N^2$  number of physical qudits that interact nonlocally with each other. Qualitatively, this model behaves similarly to the one in Sec. IV in the low energy sector, except small correction terms that suppressed to order at most  $\sim \log N/N$ . We find that if we require the total energy of the interaction scales as  $\log N$  and a reasonable pattern of coupling strengths, then Theorem 1 again holds to guarantee a  $N^2$  memory time. The zero temperature limit of life times for the last two models also diverge as  $N^2 e^{\frac{\omega}{T}}$ .

## II. MATRIX MODEL AND SET UP

### A. Matrix quantum mechanics

We wish to study a gauged matrix model of harmonic oscillators. First, recall the classical Lagrangian of a single harmonic oscillator,

$$L = \frac{1}{2}(\dot{x}^2 - \omega^2 x^2). \quad (3)$$

The canonical momentum is  $p = \frac{\partial L}{\partial \dot{x}} = \dot{x}$ , in terms of which we also have the Hamiltonian description,

$$H = \frac{1}{2}(p^2 + \omega^2 x^2). \quad (4)$$

The system is quantized in the usual way by setting  $[x, p] = i\hbar$ .

A generalization can be made by promoting the variable  $x(t)$  to an  $N \times N$  matrix  $X(t)$  and promoting products of variables to matrix multiplication. We hasten to emphasize that this  $N \times N$  matrix structure does not refer to a quantum Hilbert space; rather it is a way of organizing  $N^2$  variables into a single unit. Then one gets an action for a matrix harmonic oscillator,

$$L = \frac{1}{2} \text{Tr}[(\partial_t X)^2 - \omega^2 X^2]. \quad (5)$$

Again, let us emphasize that the trace is not a quantum Hilbert space trace but rather a sum over the matrix degrees of freedom. The diagonal elements of  $X$  are real variables, while the off-diagonal components are complex variables coming in conjugate pairs.

The momentum conjugate to  $X_j^i$  is  $P_j^i = \dot{X}_j^i$ . We quantize the theory by specifying canonical commutators for all pairs of conjugate variables. If we extend quantum Hermitian conjugation to include a transpose of the matrix degrees of freedom, then the quantum variable  $X$  is Hermitian. In terms of its matrix elements  $X_j^i$ , Hermiticity implies that  $X_j^i$  are quantum Hermitian while  $X_j^i$  and  $X_i^j$  are quantum Hermitian conjugates,  $(X_j^i)^\dagger = X_i^j$ . Note that  $X_j^i$  is not quantum Hermitian when  $i \neq j$ , but the above formula is still correct.

After quantization, it is convenient to work in terms of creation and annihilation operators,

$$\begin{aligned} a_j^i &= \sqrt{\frac{\omega}{2}}(X_j^i + iP_j^i), \\ a_i^{\dagger j} &= \frac{1}{\sqrt{2\omega}}(X_i^j - iP_i^j), \end{aligned} \quad (6)$$

which satisfy

$$[a_j^i, a_i^{\dagger k}] = \delta_j^k \delta_i^i. \quad (7)$$

The Hamiltonian written in terms of  $a$  and  $a^\dagger$ 's takes the simple form

$$H = \omega \text{Tr}(a^\dagger a) \quad (8)$$

after subtracting the zero-point energy. There is a unique vacuum state  $|0\rangle$  satisfying

$$a_j^i |0\rangle = 0, \quad \forall i, j. \quad (9)$$

Acting creation operators on this vacuum then generates all the Fock states that span the Hilbert space.

Looking back at the Lagrangian in Eq. (5), we see that it has a symmetry under unitary transformations  $X(t) \rightarrow UX(t)U^\dagger$ , with  $U \in U(N)$ . (This is why we consider complex off-diagonal entries in  $X$ .) This can be made into a local symmetry  $U(t)$ , meaning a symmetry where  $U(t)$  can be different at each different time, as follows. We add a matrix gauge field  $A(t)$  and promote the time derivative to a covariant derivative, defined as

$$D_t X = \partial_t X - [A, X], \quad (10)$$

with  $A(t)$  transforming as  $A(t) \rightarrow U(t)A(t)U(t)^\dagger - iU(t)\partial_t U(t)^\dagger$ . One can check that the covariant derivative transforms in the same way as  $X(t)$ , which is  $D_t X \rightarrow U(t)D_t X U(t)^\dagger$ . Therefore, the Lagrangian is invariant under this transformation. We call this symmetry with respect to the time-dependent transformation the gauge symmetry.

The model equipped with such a symmetry can be regarded as a redundant description of a physical system whose Hilbert space is defined to include only gauge-invariant states. Indeed, in this model,  $A(t)$  can be view as a Lagrange multiplier. Fixing a gauge  $A = 0$ , and integrating out  $A$  in the path integral representation gives the Gauss law constraint:

$$G = [X, \dot{X}] = 0. \quad (11)$$

This condition is implemented by requiring that physical states  $|\psi\rangle$  be annihilated by  $G$ ,  $G|\psi\rangle = 0$ .

One can see that the states satisfying this constraint are the symmetry singlets, generated by traces of operators acting on the vacuum,

$$|\psi\rangle = \text{Tr}(a^{\dagger n_1}) \text{Tr}(a^{\dagger n_2}) \cdots \text{Tr}(a^{\dagger n_k}) |0\rangle. \quad (12)$$

These states, after being normalized, span a subspace of the ungauged matrix model's Hilbert space. They are also eigenstates of the Hamiltonian (8) such that

$$H|\psi\rangle = \omega \sum_{i=1}^k n_i |\psi\rangle. \quad (13)$$

## B. Physical states and the Knill-Laflamme condition

We now describe in broad terms how the properties of these gauge-singlet states can be leveraged to generate approximate codes obeying an approximate version of the KL condition. First some terminology. If a gauge invariant operator contains only a single trace, it is called single trace operator. Operators with multiple traces are called multi-trace. Using the commutation relation, Eq. (7), one can derive a useful commutator rule for single trace operators, with  $n_{1,2}$ ,  $m_{1,2}$  being  $O(1)$  number,

$$\begin{aligned} & [\text{Tr}(a^{n_1} a^{\dagger m_1}), \text{Tr}(a^{n_2} a^{\dagger m_2})] \\ &= (n_1 m_2 - n_2 m_1) \text{Tr}(a^{n_1+n_2-1} a^{\dagger m_1+m_2-1}) + O\left(\frac{1}{N^2}\right). \end{aligned} \quad (14)$$

Using these rules, overlaps of states generated by acting multitrace operators on the ground state can be obtained to leading order in large  $N$ . For example, the squared norm of the state  $\text{Tr}(a^\dagger)|0\rangle$  is

$$\begin{aligned} \langle 0 | \text{Tr}(a^n) \text{Tr}(a^{\dagger n}) | 0 \rangle &= \langle 0 | [\text{Tr}(a^n), \text{Tr}(a^{\dagger n})] | 0 \rangle \\ &= nN^n + O\left(\frac{1}{N^2}\right). \end{aligned} \quad (15)$$

We also find that states with the same energy are almost orthogonal, with overlaps suppressed by a power of  $N$ . For example,

$$\langle 0 | \frac{\text{Tr}(a^m) \text{Tr}(a^{n-m}) \text{Tr}(a^{\dagger n})}{N^{\frac{m}{2}} N^{\frac{n-m}{2}}} | 0 \rangle = O\left(\frac{1}{N}\right). \quad (16)$$

Moreover, one can show that a factorization property holds for products of single trace operators. As an example, given four single trace operators with vanishing expectation value,  $O_i$ ,  $O_j$ ,  $O_k$ , and  $O_l$ , we have

$$\begin{aligned} \langle 0 | O_i O_k O_l^\dagger O_j^\dagger | 0 \rangle &\sim \langle 0 | O_i O_j^\dagger | 0 \rangle \langle 0 | O_k O_l^\dagger | 0 \rangle \\ &+ \langle 0 | O_i O_l^\dagger | 0 \rangle \langle 0 | O_k O_j^\dagger | 0 \rangle + O\left(\frac{1}{N^2}\right), \\ \langle 0 | O_i O_k O_j^\dagger | 0 \rangle &\sim O\left(\frac{1}{N}\right). \end{aligned} \quad (17)$$

If we could find a way to classify  $O_i$  and  $O_j$  as logical operators and  $O_k$  and  $O_l$  as errors, then this equation would be interpreted as an approximate version of the Knill-Laflamme condition (aKL). While one may deviate from KL in many forms, the approximate KL condition we consider in this work could be written as

$$\begin{aligned} \langle i | E_a^\dagger E_b | j \rangle &= f_{ab} \delta_{ij} + \frac{g_{ab}^{ij}}{N^2}, \\ \langle i | E_a | j \rangle &= \frac{e_a^{ij}}{N}. \end{aligned} \quad (18)$$

In the above formulas, we are imagining that  $E_a$  and  $E_b$  are some general nonidentity error operators. In general, we expect them to be any gauge invariant operators that satisfy some constraints, depending on how we couple the model to a thermal environment. We will give concrete examples in Sec. III, and explicitly calculate the functions  $f$ ,  $g$ , and  $e$ . These examples will show such large  $N$  matrix quantum mechanics models can be used to construct approximate quantum error correcting codes [14].

For later use, we also briefly discuss the counting of gauge invariant operators. If we have only one matrix harmonic oscillator, the general invariant states are in the form of Eq. (12). At energy level  $n\omega$ , the number of distinct states (which are also approximately orthogonal) is equal to the number of partitions of integer  $p(n)$ . For example, at level  $n = 3$ . The three partitions are

$$\begin{aligned} 3 &= 3 + 0, \\ 3 &= 2 + 1, \\ 3 &= 1 + 1 + 1. \end{aligned} \quad (19)$$

Correspondingly, we have three gauge invariant states at energy  $3\omega$ , which are

$$\text{Tr}(a^{\dagger 3})|0\rangle, \quad \text{Tr}(a^{\dagger 2})\text{Tr}(a^\dagger)|0\rangle, \quad \text{Tr}(a^\dagger)^3|0\rangle. \quad (20)$$

In general, we can have arbitrary  $k \leq n$  number of matrix harmonic oscillators, each carrying energy that is an integer multiple of  $\omega$ . We label the individual matrix operator by its subindex,  $a_1, a_2, a_3, \dots$ . The counting becomes more complicated with multiple oscillators, so let us focus on the states generated by single trace operators. At energy level  $n\omega$ , these states can be written as

$$\text{Tr}(P\{a_1^{\dagger n_1} a_2^{\dagger n_2} \dots a_k^{\dagger n_k}\})|0\rangle, \quad (21)$$

with  $n = \sum_i n_i$ . The state  $|0\rangle$  is tensor product of  $k$  ground states of each oscillator. We use the symbol  $P$  to denote a particular permutation of all the operators inside the bracket. By counting how many different ways of splitting  $n$  and the number of permutations, one can show that the number of these single trace operators asymptotically scales as  $k^n$  when  $n$  is large enough. For simplicity, in the concrete models that will be discussed in the following sections, we consider no more than two matrix modes.

## C. Bath model

We have seen that if errors and logical operators can be chosen properly, then an approximate Knill-Laflamme

condition holds. Based on this, we expect that a single error operator can only corrupt the logical information at order  $\frac{1}{N^2}$ . At the same time, gauge invariance constrains the number of possible errors that may occur in the form of excitations. Since the density of states for the gauged oscillator grows with energy but is independent of  $N$  below some critical value, we may expect such a system to serve as a good quantum memory even at finite temperature. To show this explicitly, we couple the proposed quantum memory to a thermal reservoir and calculate the memory time  $t_{\text{mem}}$ .

We make the standard assumption that the bath is Markovian and couples to the system locally. More specifically, we model the bath as collection of bosonic modes  $b_l$ , e.g. a collection of oscillators, with the following thermal spectral function [31],

$$\text{Tr}(\rho_B b_k(\nu) b_l^\dagger(-\nu)) = \delta_{kl} \left| \frac{\nu}{1 - e^{-\beta\nu}} \right| := \delta_{kl} \gamma(-\nu). \quad (22)$$

Physically, this means that the probability of having a given error that adds energy  $\epsilon = -\nu$  to the system is suppressed by a thermal factor  $\gamma(\epsilon)$ .

We will discuss three scenarios for the structure of the bath. In particular, there is a degree of ambiguity in what constitutes a generic physical bath for these systems, so we consider a number of models of varying levels of generality.

Case 1: Consider two matrix harmonic oscillators,  $a_1$  and  $a_2$ , separated by a large distance in space. The reason for introducing two such oscillators will become clear in the next section. We model this setup by assuming that these oscillators couple to the thermal bath independently with order one coupling constants. This means we assume that thermal errors can include any multitrace operators provide they are uncorrelated between  $a_1$  and  $a_2$ . By uncorrelated error, we mean that the modes  $a_1$  and  $a_2$  can not appear inside the same trace; instead errors must be of the form

$$E_{L,R} = \prod_{(r_1, r_2) \in L, (s_1, s_2) \in R} \frac{\text{Tr}(P\{a_1^{\dagger r_1} a_1^{r_2}\}) \text{Tr}(P\{a_2^{\dagger s_1} a_2^{s_2}\})}{N^{\frac{r_1+r_2+s_1+s_2}{2}}}, \quad (23)$$

with  $L$  and  $R$  being some sets of integer 2-tuples. The symbol  $P$  denotes the sum over all distinct arrangements of the creation and annihilation operators inside the bracket. The symbol  $::$  means normal ordering and the  $N$ -dependent normalization is to ensure that the single trace operator of each  $a$ ,  $a^\dagger$  arrangement has order one norm [see Eq. (15)]. In terms of these errors, the thermal Hamiltonian is

$$H_{\text{thermal}} = \sum_{L,R} \lambda_{L,R} b_{L,R} E_{L,R} + \text{H.c.}, \quad (24)$$

where H.c. stands for Hermitian conjugate. We will discuss this case in more detail in Sec. III A.

Case 2: We assume the coupling with the bath modes  $b_l$  are through single-trace operators, with order one coupling constants  $\lambda_l$ s. We still consider modes  $a_1$  and  $a_2$  from two different matrix oscillators. The thermal coupling Hamiltonian is in the form of

$$H_{\text{thermal}} = \sum_{\{n_k\}} \lambda_{\{n_k\}} b_{\{n_k\}} \frac{\text{Tr}(P\{a_1^{\dagger n_1} a_2^{\dagger n_2} a_1^{n_3} a_2^{n_4}\})}{N^{\sum_{k=1}^4 \frac{n_k}{2}}} + \text{H.c.} \quad (25)$$

At first sight, it appears that coupling only to single-trace operators is a strong assumption, because in general we should allow all gauge invariant operators to couple with the bath. However, this is still general enough since single-trace errors accumulate into multi-trace errors as time increases. We will discuss this scenario in more detail in Sec. III B.

Case 3: We consider the most general coupling to the thermal bath where each of the  $N^2$  harmonic oscillator modes can couple to the bath independently without requiring the coupling terms be gauge invariant. In addition, we assume that the coupling constant decays exponentially with the number of creation and annihilation operators involved. This is a reasonable assumption because the correlated errors that involves more operators occur with smaller probability. We explain this in detail in Sec. IV, where we consider the ungauged model.

#### D. Mutual information diagnostic

Given the assumptions on the bath described above, the time evolution of our system can be described by a Lindblad equation after tracing out the bath,

$$\dot{\rho}(t) = -i[H, \rho(t)] + \sum_a \lambda_a(\nu_a) E_a \rho(t) E_a^\dagger - \frac{1}{2} \sum_a \lambda_a(\nu_a) \{E_a^\dagger E_a, \rho(t)\}, \quad (26)$$

where  $\lambda_a(\nu_a) = \lambda_a \gamma(\nu_a)$ , is the coupling constant multiplied with thermal factor.  $\nu_a$  is the energy of excitation, and the  $\{E_a\}$  depend on the interaction Hamiltonian.

The generic expectation is that the above system dynamics will lead to rapid information loss within the system. We wish to show that this is not the case for the right kind of code constructed from the matrix degrees of freedom. To quantitatively measure the memory time, we entangle the encoded  $d$ -dimensional system with a reference  $R$ . The initial state is thus

$$\frac{1}{\sqrt{d}} \sum_{i=1}^d |\tilde{i}\tilde{i}\rangle := \frac{1}{\sqrt{d}} \sum_{i=1}^d |i\rangle_S |i\rangle_R, \quad (27)$$

$$\rho_0 = \frac{1}{d} \sum_{i,j=1}^d |\tilde{i}\tilde{i}\rangle \langle j\tilde{j}|,$$

and the correlation between the reference is maximal.

The goal is then to understand how the system-reference correlation evolves with time. The state evolves under the Lindblad dynamics into

$$\rho(t) = e^{\mathcal{L}t} \rho_0,$$

where  $\mathcal{L}\mathcal{O} = \sum_k \lambda_k E_k \mathcal{O} E_k^\dagger - \sum_k \frac{\lambda_k}{2} \{E_k^\dagger E_k, \mathcal{O}\}$ . (28)

In terms of the system and reference reduced density matrices  $\rho_S(t) = \text{Tr}_R[\rho(t)]$  and  $\rho_R(t) = \text{Tr}_S[\rho(t)]$ , we may quantify the correlation using the mutual information,

$$I(S:R)(t) = S(\rho_S(t)) + S(\rho_R(t)) - S(\rho(t)). \quad (29)$$

When the mutual information is maximal,  $I = 2 \ln d$ , we say that the quantum information is stored in the quantum memory. However, as  $I(S:R)$  decays, the information of the encoded qubit leaves the memory and dissipates into the environment. For the sake of simplicity, we will focus on quantum memory constructions that store a single encoded qubit in the ensuing sections, so  $d = 2$ .

We now tie together all the ingredients above in a theorem. In general, different couplings with the bath can induce different errors on the system. However, as long as such errors are not too numerous and can be suppressed by the code, i.e. satisfying the aKL condition, then the following theorem guarantees the memory time of the quantum information is relatively long. To be precise, this theorem shows that the mutual information diagnostic decays polynomially with  $t$  at a rate suppressed by  $\frac{1}{N^2}$ , which implies that the memory time can scale polynomially with  $N$ .

**Theorem 1.** Consider a system whose low energy states form an approximate quantum error correcting code that encodes a logical qudit  $S$  and a set of physical errors  $\mathcal{E}$  induced by the system-bath coupling (cf. Sec. II C). Provided the following conditions are satisfied,

- (i) (Sparse spectrum) There exists an energy  $\epsilon_0$  below which the effective number of errors<sup>4</sup> in  $\mathcal{E}$  with energy  $\epsilon \leq \epsilon_0$  is independent of  $N$  and bounded by  $\exp(\mu\epsilon)$  for some  $\mu > 0$ ,
- (ii) (Uniform coupling) The thermal Hamiltonian that couples each  $E \in \mathcal{E}$  to an independent bath operator has coupling constant no larger than  $O(1)$ ,
- (iii) (Approximate error correction)  $\forall E_a, E_b \in \mathcal{E}$ , the approximate Knill-Laflamme condition [Eq. (18)] is satisfied,

then the mutual information  $I(S, R)$  defined in Eq. (29) is given by

$$I(S:R)(t) = 2 \ln d - K \left( \frac{t}{N^2} \right), \quad (30)$$

<sup>4</sup>These are errors that contribute nontrivially in the Lindblad equation.

where  $K(x)$  is a polynomial function of  $x$ , for temperature  $T < \frac{1}{\mu}$ .

*Proof.*—The proof is given in Appendix A. ■

In the following sections, we will construct different systems coupled via different bath models, as outlined in Sec. II C, and prove that they satisfy the assumptions in the theorem, such as the approximate Knill-Laflamme condition (aKL) for a corresponding set of error operators and logical states.

Remark: Note that the above theorem does not require the code subspace to be the ground space of the system. Therefore, the  $N^2$  scaling of memory time holds generally for these models no matter how the code subspace is chosen as long as the requisite conditions are satisfied. Intuitively, this is because the transition rate between states are  $1/N^2$  suppressed. However, as excited states can still slowly decay into lower energy states at the same rate, their memory times need not be longer by lowering the temperature.

On the other hand, for quantum memories where we also choose the code subspace to be the ground space, the memory lifetime can be further extended by a multiplicative factor  $\exp(\omega/T)$ , which is related to the effect of thermal suppression. This renders the total low temperature memory time to be of  $O(N^2 \exp(\omega/T))$  for some energy scale  $\omega$  as we see in some of the examples below.

### III. FULLY GAUGED MODEL

#### A. Nonlocal model

As the first example, consider two gauged harmonic oscillators  $a_1$  and  $a_2$  with  $SU(N)$  gauge symmetry, separated by a large distance. Their interaction is weak, and each couples to the thermal reservoir locally and independently.<sup>5</sup> The Hamiltonian takes the following form:

$$H = \omega \text{Tr}(a_1^\dagger a_1) + \omega \text{Tr}(a_2^\dagger a_2) + \sum_{L,R} \lambda_{L,R} b_{L,R} E_{L,R} + \text{H.c.}, \quad (31)$$

where  $E_{L,R}$ s are the possible errors acting on this model, and  $\lambda_{L,R}$ s are  $O(1)$  coupling constants.

The general gauge invariant states are given in Eq. (21) with  $k = 2$ . However, since we assume that these modes only couple to the thermal bath locally, the set of possible error operators is a proper subset of all physical operators. These errors are uncorrelated in the sense that each error operator contains only a single trace either the mode  $a_1$  or the mode  $a_2$ , but never both. Given two sets of integer 2-tuples  $L$  and  $R$ , we can define the errors as

<sup>5</sup>More realistically, thermal noise can have finite correlation. In such cases, we assume that the distance separating the two oscillators is much larger than the correlation length so our assumptions remain good approximations.

$$E_{L,R} = : \prod_{(r_1, r_2) \in L, (s_1, s_2) \in R} \frac{\text{Tr}(P\{a_1^{\dagger r_1} a_1^{r_2}\}) \text{Tr}(P\{a_2^{\dagger s_1} a_2^{s_2}\})}{N^{\frac{r_1+r_2+s_1+s_2}{2}}} :, \quad (32)$$

where the symbol  $P$  denotes the sum over all distinct arrangements of the  $a_i, a_i^\dagger$  operators.<sup>6</sup> In general,  $P\{a^{\dagger r_1} a^{r_2}\}$  admits  $(r_1 + r_2)!/r_1!r_2!$  terms of distinct arrangements. Each operator is brought to normal ordering, and divided by some power of  $N$  to make sure that these operators have  $O(1)$  norm.

### 1. Logical states

Given this set of error operators, we can construct logical states that are protected against them. In general, we may choose the logical states to be

$$|\mathcal{I}\rangle := \prod_{(m,n) \in \mathcal{I}} \frac{\text{Tr}(a_1^{\dagger m} a_2^{\dagger n})}{N^{\frac{m+n}{2}}} |0\rangle_{12}, \quad (33)$$

with  $m, n \geq 1$  and belonging to some set  $\mathcal{I}$  of integer 2-tuples. The state  $|0\rangle_{12}$  is the tensor product of the vacuum states of the two oscillators. For simplicity, let us focus on a code subspace spanned by two states and treat it as a logical qubit<sup>7</sup>:

$$\begin{aligned} |\tilde{1}\rangle &= \frac{\text{Tr}(a_1^{\dagger 2} a_2^{\dagger 2})}{N^2} |0\rangle_{12}, \\ |\tilde{2}\rangle &= \frac{\text{Tr}(a_1^{\dagger} a_2^{\dagger})^2}{\sqrt{2}N^2} |0\rangle_{12}. \end{aligned} \quad (34)$$

They are not orthogonal states, since  $\langle \tilde{1} | \tilde{2} \rangle = \frac{\sqrt{2}}{N} + O(\frac{1}{N^3})$ . However, it is straightforward to identify an orthonormal basis  $|\tilde{\uparrow}\rangle$  and  $|\tilde{\downarrow}\rangle$ , defined as

$$\begin{aligned} |\tilde{\uparrow}\rangle &= |\tilde{1}\rangle \\ |\tilde{\downarrow}\rangle &= \frac{|\tilde{2}\rangle - \langle \tilde{2} | \tilde{1} \rangle |\tilde{1}\rangle}{\sqrt{1 - \langle \tilde{2} | \tilde{1} \rangle^2}}. \end{aligned} \quad (35)$$

### 2. Errors and aKL

Because we assume errors are gauge invariant, they are expressible as single or multitrace operators [see Eq. (32)]. For the purpose of invoking Theorem 1, it suffices to check that aKL holds for these errors,

<sup>6</sup>For example,  $a^{\dagger 2} a^2$  and  $a^\dagger a a^\dagger a$  are counted as two distinct arrangements.

<sup>7</sup>The logical subspace we use is not the lowest energy subspace of the Hamiltonian, so they have the tendency to decay into lower energy states. However, the decay rate is suppressed by  $\frac{1}{N^2}$ , as discussed in Appendix B 1.

$$\begin{aligned} \langle \tilde{i} | E_{L',R'}^\dagger E_{L,R} | \tilde{j} \rangle &= f_{L',R',LR} \delta_{ij} + \frac{g_{L',R',LR}^{ij}}{N^2} + O\left(\frac{1}{N^4}\right), \\ \langle \tilde{i} | E_{L,R} | \tilde{j} \rangle &= \frac{e_{L',R',LR}^{ij}}{N} + O\left(\frac{1}{N^3}\right). \end{aligned} \quad (36)$$

To build up some intuition, let us first examine a few instructive examples before we arrive at the more general relation (36). To start, one can check that the operator  $\frac{\text{Tr}(a_1^\dagger a_1)}{N}$  has the logical states as eigenstates,

$$\frac{\text{Tr}(a_1^\dagger a_1)}{N} |\tilde{i}\rangle = \frac{2}{N} |\tilde{i}\rangle. \quad (37)$$

Therefore, it is a logical identity operator. We can similarly check that the Hamiltonian that generates the system dynamics is proportional to the logical identity, which is necessary for a memory.

Next, let us examine a class of single-trace errors of the form  $E_n = \frac{\text{Tr}(a_1^{\dagger n})}{\sqrt{n}N^{\frac{n}{2}}}$ . As a potential bit flip error, we check the transition amplitude,

$$\langle \tilde{\downarrow} | E_n^\dagger E_n | \tilde{\uparrow} \rangle = O\left(\frac{1}{N^3}\right). \quad (38)$$

The bit flip error only carries  $n$  dependence at order  $O(\frac{1}{N^3})$ . Similarly, the phase errors are also suppressed,

$$\begin{aligned} \langle \tilde{\uparrow} | E_m^\dagger E_n | \tilde{\uparrow} \rangle &= \delta_{nm} \left[ 1 + \frac{3n}{N^2} + O\left(\frac{1}{N^4}\right) \right], \\ \langle \tilde{\downarrow} | E_m^\dagger E_n | \tilde{\downarrow} \rangle &= \delta_{nm} \left[ 1 + \frac{2n}{N^2} + O\left(\frac{1}{N^4}\right) \right]. \end{aligned} \quad (39)$$

Note that the leading order terms of these equations come from factorized two point functions  $\langle 0 | E_m^\dagger E_n | 0 \rangle \langle \tilde{i} | \tilde{j} \rangle$ , and they represent the nonconnected contribution to the four point correlator. The subleading corrections are from the connected part, shown diagrammatically in Fig. 1. It is this part that induces phase and bit-flip errors between the two logical states. In a similar way, one can check that the aKL condition holds for other single-trace errors.

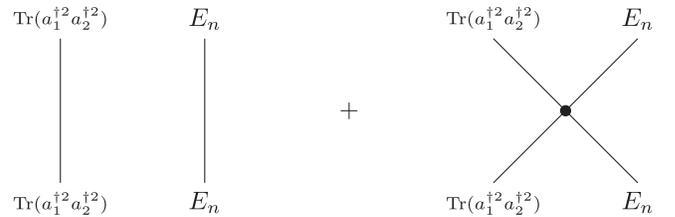


FIG. 1. The inner product can be viewed as scattering between two particles.  $(1, 0)$  represents the logical state  $|\tilde{\uparrow}\rangle$ . The connected part is suppressed by  $\frac{1}{N^2}$ .

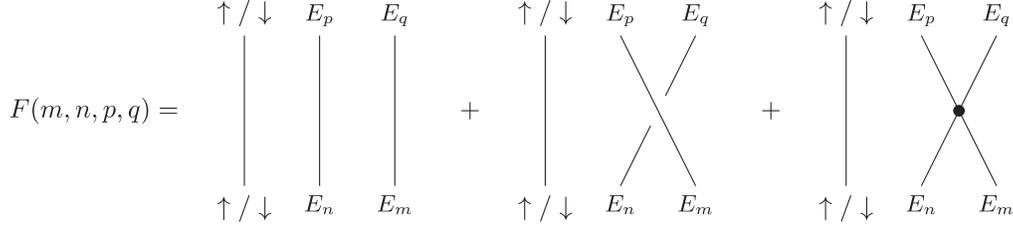


FIG. 2. The first two diagrams are totally disconnected. The third is partly connected and represents the  $O(\frac{1}{N^2})$  term in  $F(m, n, p, q)$ .

We present the more comprehensive calculations in Appendix B 1.

The conclusion of this analysis is that single-trace operators can only produce  $1/N$  suppressed errors, thus giving rise to the aKL. Using the results for single-trace errors, one can also derive the aKL conditions for any multitrace errors. As an example,

$$\begin{aligned} \langle \tilde{\uparrow} | E_p^\dagger E_q^\dagger E_m E_n | \tilde{\uparrow} \rangle &= F(m, n, p, q) \\ &+ \frac{G_\uparrow(m, n, p, q)}{N^2} + O\left(\frac{1}{N^4}\right), \\ \langle \tilde{\downarrow} | E_p^\dagger E_q^\dagger E_m E_n | \tilde{\downarrow} \rangle &= F(m, n, p, q) \\ &+ \frac{G_\downarrow(m, n, p, q)}{N^2} + O\left(\frac{1}{N^4}\right). \end{aligned} \quad (40)$$

The function  $F$  is the same in both correlators, so it represents a correctable error. In contrast, different  $G$  functions appear in the two equations. Their difference,  $G_\uparrow - G_\downarrow$ , represents a small uncorrectable error. They both have simple diagrammatic representation as shown in Figs. 2 and 3, from which we can read off the function  $F(m, n, p, q)$ . It must take the form

$$F(m, n, p, q) = (\delta_m^p \delta_n^q + \delta_m^q \delta_n^p) + \delta_{m+n}^{p+q} \frac{f(p, q, m, n)}{N^2}, \quad (41)$$

where the  $O(\frac{1}{N^2})$  term comes from the partly connected diagram in Fig. 2. Since  $F$  does not depend on the logical information and will not contribute to uncorrectable errors, it is not pertinent to the subsequent discussion. Therefore, we will not determine  $f$  explicitly.

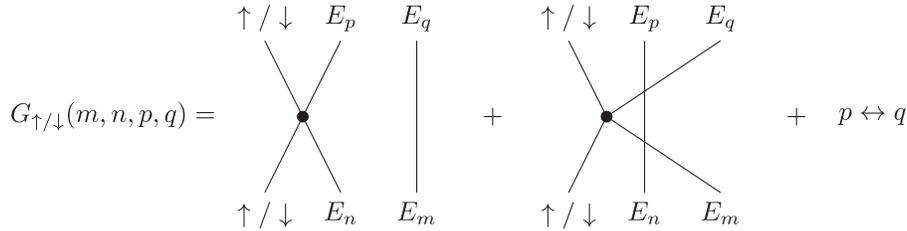


FIG. 3. The  $G$  functions arise from connected diagrams of order  $O(\frac{1}{N^2})$  that involve the state  $\text{Tr}(a_1^\dagger)|0\rangle$ .

The function  $G$  can be read off from Fig. 3. Using the results for single-trace errors in Eq. (39), we can show that

$$\begin{aligned} G_\uparrow(m, n, p, q) &= (\delta_m^p \delta_n^q + \delta_m^q \delta_n^p)(3n + 3m), \\ G_\downarrow(m, n, p, q) &= (\delta_m^p \delta_n^q + \delta_m^q \delta_n^p)(2n + 2m). \end{aligned} \quad (42)$$

From this explicit form, we see that there is indeed a difference between these functions, so the logical states are connected by multitrace operators, again with  $1/N$  suppressed amplitudes provided the indices  $m, n, p, q$  are not too large.

Now, although we only explicitly calculated a particular class of multitrace errors, the algorithm generalizes directly to more general cases. The analog of the  $G$  function always consists of connected scattering diagram of logical states and single-trace operators. In other words, the uncorrectable part of any multitrace error can be written as a sum over contributions of the single-trace errors, due to the factorization property of large  $N$  matrix models [Eq. (17)]. Therefore, if the aKL condition is satisfied for the single-trace operators, it will also hold for these multitrace errors. Note that if a diagram has four legs connected, then it must be suppressed by  $\frac{1}{N^2}$ . In special cases we also have three-leg diagrams, where the error operators involve  $\frac{\text{Tr}(a_{1,2}^{\dagger 2})}{\sqrt{2N}}$  or  $\frac{\text{Tr}(a_{1,2}^\dagger)}{\sqrt{N}}$ . For example, the  $G$  function in the correlator  $\langle \tilde{\uparrow} | \frac{\text{Tr}(a_1^\dagger)}{\sqrt{2N}} \frac{\text{Tr}(a_2^\dagger)}{\sqrt{2N}} \frac{\text{Tr}(a_1^{\dagger 2})}{\sqrt{2N}} \frac{\text{Tr}(a_2^{\dagger 2})}{\sqrt{2N}} | \tilde{\uparrow} \rangle$  contains the product of two three-leg diagrams, as shown in Fig. 4. Each of them is proportional to  $\frac{1}{N}$ , so the term involving  $G$  that depends on the logical state is still suppressed by  $\frac{1}{N^2}$ .

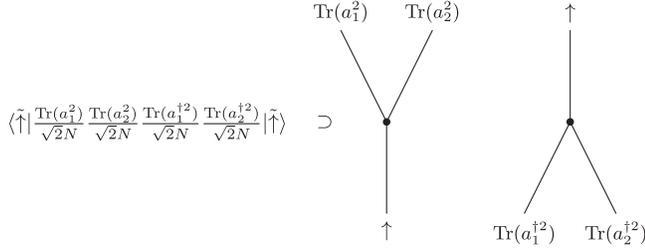


FIG. 4. Contribution of connected diagram with three legs. Each is suppressed by  $\frac{1}{N}$ .

### 3. Mutual information diagnostic

Now we couple this quantum memory to a single-qubit reference system  $R$ , and calculate the mutual information between them [Eq. (29)]. At  $t = 0$ ,  $\rho_0$  is a pure state,

$$\rho_0 = \frac{1}{2} \sum_{i,j=1}^2 |\tilde{i}\tilde{i}\rangle_{SR} \langle \tilde{j}\tilde{j}|. \quad (43)$$

The system and reference are maximally entangled, so the initial mutual information is  $I(S:R)(0) = 2 \ln 2$ .

Next we count the effective number of error operators which will contribute nontrivially to logical errors under the Lindblad dynamics. Operators that contain more than two  $a_1$  or  $a_2$  trivially annihilate the logical states, so we only need to count the error operators with  $r_2, s_2 \leq 2$ . The number of gauge invariant error operators with energy  $n\omega$  that satisfy this requirement is bounded by  $n^2 p(n)$ , where  $p(n)$  is the number of integer partitions of  $n$ . So the number of effective errors is subexponential asymptotically.

By construction, the bath model has uniform coupling as required by condition 2 in Theorem 1. Together with the aKL condition above, we apply Theorem 1 and conclude the mutual information decays in the following way,

$$I(S:R)(t) = 2 \ln 2 - K \left( \frac{t}{N^2} \right), \quad (44)$$

where  $K(t)$  is linear in  $t$  at early time. More specifically, when  $\lambda t \ll N^2$ ,  $K(t) \sim \sum_n n \lambda_n t$ . It develops a dependence on

higher powers of  $t$  at late time  $\lambda t \sim N^2$ , where  $K(t) \sim \sum_p \left( \frac{\sum_n n \lambda_n t}{N^2} \right)^p$ .

Details of the proof for Eq. (44) and more rigorous results can be found in Appendix A, but here we briefly sketch the general ideas behind the proof. The dominant errors in the early time regime are the single trace errors. In contrast, in the late time regime, the dominant errors are multitrace. In fact, the average size  $k$  of errors, defined as the number of traces in an error operator, grows linearly with time. These errors, when acting on logical states, cause a loss of quantum information proportional to  $\frac{k^p}{N^{2p}}$ , leading to the  $\left(\frac{t}{N^2}\right)^p$  growth behavior at late time.

Note that the rate of information loss here is captured by the sum over  $\lambda_n$ , which can be approximated by

$$\sum_n n^q \lambda_n = \sum_n n^q d_n e^{-n\beta\omega} \sim \sum_n n^{q+2} p(n) e^{-n\beta\omega}, \quad (45)$$

where  $q = 1$ , and  $d_n$  is the number of distinct states at the  $n$ th level. In this model,  $d_n \sim n^2 p(n)$ , which grows subexponentially with  $n$ . As the sum is always finite, it follows that the memory time scales like  $t_{\text{mem}} \sim N^2$ . From the above discussion, we conclude that this model serves as a good passive quantum memory.

Now we discuss the zero temperature limit. Since the thermal fluctuation is suppressed, the relevant error operators are only those which lower the energy when acting on logical states. In this model, they are  $\frac{\text{Tr}(a_{1,2})}{\sqrt{N}}$  and  $\frac{\text{Tr}(a_{1,2}^{\dagger 2})}{N}$ , whose aKL conditions are given in Appendix B 1. They lead to a spontaneous decay of quantum information at a constant rate of  $O\left(\frac{1}{N^2}\right)$  at zero temperature. There are also operators that do not change energy of the system, but they have no contribution to the information loss since their thermal spectral functions [Eq. (22)] vanish at zero temperature limit,  $\lim_{\beta \rightarrow \infty} \gamma(\nu = 0) = 0$ .

To modify the encoded information, we need to apply logical operators, which involve  $\text{Tr}(a_1^{\dagger} a_2^{\dagger})$  or  $\text{Tr}(a_1^{\dagger 2} a_2^{\dagger 2})$ . This requires us to perform nonlocal operations (or a series of local operations) by bringing the two modes together and

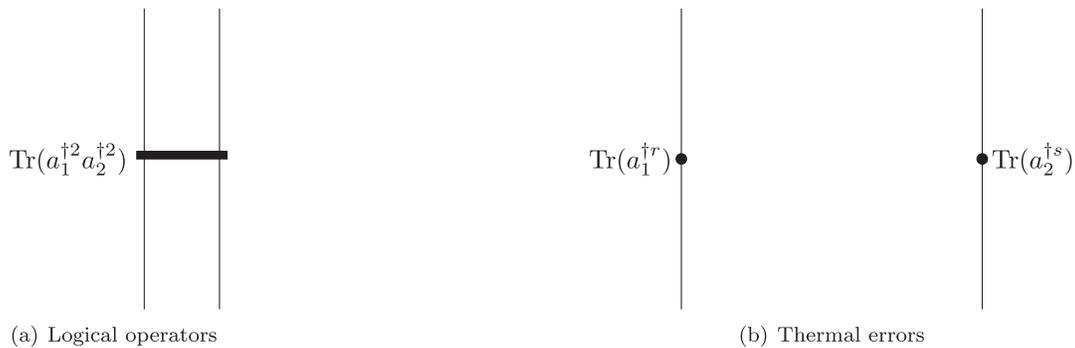


FIG. 5. In (a), two modes are brought together to implement a logical operator. In (b), two modes are separated far away. Thermal fluctuation are sufficiently local so the quantum information is preserved.

turning on some local coupling, as shown in Fig. 5. We assume this procedure to be accurate. After the logical operations, we can separate the two modes so that the system again behaves like a passive memory that protects the encoded information.

We can also implement logical operations without bringing the two modes together. Instead, we may induce a coupling mediated by a third ancillary mode  $c$ . The procedure is that we first bring the mode  $c$  close enough to the mode  $a_1$  and implement the operator  $\text{Tr}(a_1^{\dagger m} c^{\dagger})$ . Then we bring together the intermediate mode and the mode  $a_2$  and implement the operator  $\text{Tr}(a_2^{\dagger n} c)$ . These two moves can be combined together to give the desired logical operation,

$$\text{Tr}(a_2^{\dagger 2} c) \text{Tr}(a_1^{\dagger 2} c^{\dagger}) |0\rangle \propto \text{Tr}(a_1^{\dagger 2} a_2^{\dagger 2}) |0\rangle. \quad (46)$$

In summary, this model achieves the power law scaling of memory time with system size. However, it has some drawbacks. First, it requires a gauge symmetry that acts consistently on the two matrix modes separated by a large distance. Second, the logical operation is nonlocal, and thus can be difficult to implement. For these reasons, we construct another model, where the two matrix modes are put together, but with the added price that a high energy penalty is required.

### B. Local model

In this model, we still consider two matrix harmonic oscillators, but we put them together and turn on a large coupling between them. The full Hamiltonian is

$$H = \omega \text{Tr}(a_1^{\dagger} a_1) + \omega \text{Tr}(a_2^{\dagger} a_2) + J \left[ \frac{\text{Tr}(a_1^{\dagger}) \text{Tr}(a_1)}{N} + \frac{\text{Tr}(a_2^{\dagger}) \text{Tr}(a_2)}{N} - 1 \right]^2 + H_{\text{thermal}} \quad (47)$$

with  $J \sim 2\omega \ln N$ . The ground space of this Hamiltonian is spanned by the following two states,

$$\begin{aligned} |\tilde{\uparrow}\rangle &= \frac{\text{Tr}(a_1^{\dagger})}{\sqrt{N}} |0\rangle_{12}, \\ |\tilde{\downarrow}\rangle &= \frac{\text{Tr}(a_2^{\dagger})}{\sqrt{N}} |0\rangle_{12}, \end{aligned} \quad (48)$$

where  $|0\rangle_{12}$  is again the tensor product of the vacuum states of the two oscillators. These states span the code subspace of the model.

We assume that  $H_{\text{thermal}}$  contains only single trace operators:

$$H_{\text{thermal}} = \sum_{\{n_k\}} \lambda_{\{n_k\}} b_{\{n_k\}} \frac{\text{Tr}(P\{a_1^{\dagger n_1} a_2^{\dagger n_2} a_1^{n_3} a_2^{n_4}\})}{N^{\sum_{k=1}^4 n_k}} + \text{H.c.}, \quad (49)$$

where  $P$  again denotes the sum over distinct arrangements of the operators inside the curly bracket, and all the operators are then normal ordered. In general,  $\text{Tr}(P\{a_1^{\dagger n_1} a_2^{\dagger n_2} a_1^{n_3} a_2^{n_4}\})$  is a sum over  $\frac{(n_1+n_2+n_3+n_4)!}{n_1!n_2!n_3!n_4!}$  terms.

The interaction term in the Hamiltonian Eq. (47) incurs an energy penalty for any logical errors that involve  $\text{Tr}(a_1^{\dagger})$  or  $\text{Tr}(a_2^{\dagger})$ , so their probability to appear are suppressed due to the thermal factor. For example, the probability of the error operator  $\text{Tr}(a_1^{\dagger}) \text{Tr}(a_2)$  is  $\gamma(J) \gamma(-J) \sim e^{-\frac{J}{\omega}}$ . We set the energy scale  $J$  to be  $2\omega \ln N$ , which guarantees that these uncorrectable errors are unlikely to occur when  $N$  is large enough. All the other operators generated by the coupling in Eq. (49) are treated as possible errors that could act on the logical state. To show that these errors are approximately correctable, and to invoke Theorem 1, it suffices to check that aKL holds for single-trace errors,

$$\begin{aligned} E_{\{n_k\}} &:= \frac{\text{Tr}(P\{a_1^{\dagger n_1} a_2^{\dagger n_2} a_1^{n_3} a_2^{n_4}\})}{N^{\sum_{k=1}^4 n_k}}, \\ \langle \tilde{i} | E_{\{n'_k\}}^{\dagger} E_{\{n_k\}} | \tilde{j} \rangle &= f_{\{n'_k\}, \{n_k\}} \delta_{ij} + \frac{g_{\{n'_k\}, \{n_k\}}^{ij}}{N^2} + O\left(\frac{1}{N^4}\right), \\ \langle \tilde{i} | E_{\{n_k\}} | \tilde{j} \rangle &= \frac{e_{\{n_k\}}^{ij}}{N} + O\left(\frac{1}{N^3}\right). \end{aligned} \quad (50)$$

Applying the argument we discussed in Sec. III A, one can again derive the aKL equation for general multitrace operators using the above relations for single-trace operators. This guarantees that the aKL condition holds for all the error operators generated in this model. A proof of the aKL condition for the most general error is given in Appendix B 2. Here we discuss a few helpful examples to show that these errors indeed conform with the aKL condition.

Again, let us first check operators of the form  $\frac{\text{Tr}(a_{1,2}^{\dagger} a_{1,2})}{N}$ . They induce a phase error of size  $O(\frac{1}{N})$ ,

$$\begin{aligned} \langle \tilde{\uparrow} | \frac{\text{Tr}(a_1^{\dagger} a_1)}{N} | \tilde{\uparrow} \rangle &= \frac{1}{N}, \\ \langle \tilde{\downarrow} | \frac{\text{Tr}(a_1^{\dagger} a_1)}{N} | \tilde{\downarrow} \rangle &= 0, \end{aligned} \quad (51)$$

which is consistent with the aKL condition in Eq. (50). Next, we check the aKL conditions for single-trace errors of the form  $E_{(n,m)} = \frac{\text{Tr}(a_1^{\dagger n} a_2^{\dagger m})}{N^{\frac{m+n}{2}}}$ , with  $n \geq 1$ ,  $m \geq 1$ . For the bit flip errors, we can show that

$$\langle \tilde{\downarrow} | E_{(n+1, m-1)}^{\dagger} E_{(n, m)} | \tilde{\uparrow} \rangle = \delta_{nn'} \delta_{mm'} \frac{(n+1)m}{N^2} + O\left(\frac{1}{N^4}\right), \quad (52)$$

and for the phase errors, we have

$$\begin{aligned}\langle \tilde{\uparrow} | E_{(n',m')}^\dagger E_{(n,m)} | \tilde{\uparrow} \rangle &= \delta_{nn'} \delta_{mm'} \left[ 1 + \frac{n^2 - m^2}{N^2} + O\left(\frac{1}{N^4}\right) \right], \\ \langle \tilde{\downarrow} | E_{(n',m')}^\dagger E_{(n,m)} | \tilde{\downarrow} \rangle &= \delta_{nn'} \delta_{mm'} \left[ 1 + \frac{m^2 - n^2}{N^2} + O\left(\frac{1}{N^4}\right) \right].\end{aligned}\quad (53)$$

When  $m$  or  $n$  equals to 0, the equations are slightly different. For instance, when  $m = 0$ ,  $n \geq 2$ , the matrix elements are

$$\begin{aligned}\langle \tilde{\uparrow} | E_{(n',0)}^\dagger E_{(n,0)} | \tilde{\uparrow} \rangle &= \delta_{nn'} \left[ 1 + \frac{n(n-1)}{N^2} + O\left(\frac{1}{N^4}\right) \right], \\ \langle \tilde{\downarrow} | E_{(n',0)}^\dagger E_{(n,0)} | \tilde{\downarrow} \rangle &= \delta_{nn'}, \\ \langle \tilde{\uparrow} | E_{(n',0)}^\dagger E_{(n,0)} | \tilde{\downarrow} \rangle &= 0.\end{aligned}\quad (54)$$

Although  $H_{\text{thermal}}$  only contains single-trace operators, the expansion in Eq. (28) involves multitrace operators acting on the state. To see the effect of these multitrace errors, we evaluate an example:

$$\begin{aligned}\langle \tilde{\uparrow} | E_{(p,0)}^\dagger E_{(q,0)}^\dagger E_{(m,0)} E_{(n,0)} | \tilde{\uparrow} \rangle &= F(m, n, p, q) \\ &\quad + \frac{G(m, n, p, q)}{N^2} + O\left(\frac{1}{N^4}\right), \\ \langle \tilde{\downarrow} | E_{(p,0)}^\dagger E_{(q,0)}^\dagger E_{(m,0)} E_{(n,0)} | \tilde{\downarrow} \rangle &= F(m, n, p, q) + O\left(\frac{1}{N^4}\right).\end{aligned}\quad (55)$$

If the Knill-Laflamme conditions are satisfied, then the above expressions should yield matrix elements that only depend on the errors. We see that the presence of the  $G$  function does allow one to distinguish the logical states, but the effect is suppressed by  $1/N^2$ . These correlation functions both have simple diagrammatic representations similar to those shown in Figs. 2 and 3, but the explicit form of these two functions are complicated,

$$\begin{aligned}F(m, n, p, q) &= (\delta_m^p \delta_n^q + \delta_m^q \delta_n^p) + \delta_{m+n}^{p+q} \frac{f(p, q, m, n)}{N^2}, \\ G(m, n, p, q) &= (\delta_m^p \delta_n^q + \delta_m^q \delta_n^p) [n(n-1) + m(m-1)] \\ &\quad + (\delta_{m\pm 1}^p \delta_{n\mp 1}^q + \delta_{m\pm 1}^q \delta_{n\mp 1}^p) \sqrt{pqmn},\end{aligned}\quad (56)$$

with some function  $f$  whose exact form is not essential for our argument.

Again all the errors discussed above are approximately correctable errors, because they can only corrupt the logical information up to  $O(\frac{1}{N^2})$ . The uncorrectable errors or logical operators are those that contain  $\text{Tr}(a_1^\dagger)$  and  $\text{Tr}(a_2^\dagger)$ . However, recall that we have tuned the coupling  $J$  to be

proportional to  $2\omega \ln N$  so that these uncorrectable errors have a similar  $1/N$  suppressed effect comparable in size to the approximately correctable errors.

By construction, the bath model has uniform coupling as required by condition 2 of Theorem 1. In addition, the number of gauge invariant error operators with energy  $n\omega$  that are induced by Eq. (49) grows asymptotically as  $2^n$ . In particular, the number of effective error operators that act nontrivially on the logical state, i.e. ones contributing to the information leak in the Lindblad equation, must have  $n_2, n_4 \leq 1$ . Therefore, the total number of effective errors with energy  $\omega n$  is bounded by an exponential function, as required by condition 1.

Because the aKL condition holds and the bath model only generates errors which are protected by the code, we again invoke the Theorem 1. The mutual information between the encoded qubit  $S$  and a reference  $R$  is

$$I(S:R)(t) = 2 \ln 2 - K_{\text{local}} \left( \frac{t}{N^2} \right). \quad (57)$$

The expression  $K_{\text{local}}(t)$  in this model is slightly different from  $K(t)$ , since the  $O(\frac{1}{N^2})$  term in the aKL condition depends quadratically on  $n$  [see Eq. (52)]. When  $\lambda t \ll N^2$ ,  $K_{\text{local}}(\frac{t}{N^2}) \sim \sum_n \frac{n^2 \lambda_n t}{N^2}$ . At late time  $\lambda t \sim N^2$ ,  $K_{\text{local}}(t) \sim (\sum_n n^2 \lambda_n t)^2$ .

More importantly, note that the rate of information loss is given by

$$\begin{aligned}\sum_n n^q \lambda_n &= \sum_n n^q d_n e^{-n\beta\omega}, \\ &\sim \sum_n n^q 2^n e^{-n\beta\omega},\end{aligned}\quad (58)$$

where  $q = 2$ , and  $d_n \sim 2^n$  is the number of single trace states at the  $n$ th energy level. A crucial distinction from the previous model is the Hagedorn-like transition with a critical temperature  $T_c = \frac{\omega}{\ln 2}$  [26]. In the low temperature phase where  $T < T_c$ , the rate of information loss (58) is finite and  $N$  independent. Therefore the large  $N$  gauged matrix model has a long memory time  $t_{\text{mem}} \sim N^2$ . However, for  $T > T_c$ , (58) exhibits a Hagedorn-like behavior.<sup>8</sup> As a result,  $K_{\text{local}}(t)$  now also scales with  $N$ . This  $N$  dependence cancels out the large  $N$  suppression, and the memory time no longer scales with the system size. This behavior is similar to the holographic code [10], where the temperature needs to be lower than a critical value to maintain the robustness of the encoded information.

<sup>8</sup>Note that for finite  $N$ , the Hagedorn behavior does not continue to arbitrary energies, but only to order  $N^2$ , the total number of gauge invariant degrees of freedom. The counting of gauge invariant operators can be found in Sec. II D.

Since the logical states are within the ground space of this model, there are no operators that can lower the energy. The zero energy error operators do not contribute to information loss at zero temperature limit, due to the vanishing thermal spectral function  $\lim_{\beta \rightarrow \infty} \gamma(\nu = 0)$ . Therefore, when  $\beta \rightarrow \infty$ , we expect the information decay rate to scale as  $\frac{1}{N^2} e^{-\beta\omega}$  according to Eq. (58).

#### IV. RELAXING THE GAUGE CONSTRAINT

We now relax the gauge constraint to a global symmetry, which we then impose energetically. The purpose of this is twofold—(i) we want to understand how the stability of quantum memory is tied to the gauge symmetry specifically, and (ii) there are very few models in nature that are known to have gauge symmetries. For example, quantum electrodynamics has a  $U(1)$  symmetry and QCD has an  $SU(3)$  symmetry. However, there is no known  $SU(N)$  gauged model that can be realized in the laboratory, especially for arbitrarily large  $N$ . Therefore, in this section, we consider a single oscillator model with only a global  $SU(N)$  symmetry. The system of interest has Hamiltonian

$$H_{\text{Gauss}} = \frac{\mathcal{J}}{N} \sum_{ij} G_j^i G_i^j, \quad (59)$$

$$G_j^i = \sum_k a_k^{\dagger i} a_j^k - a_j^{\dagger k} a_k^i,$$

where  $G_j^i |\psi\rangle = 0$  is the Gauss law constraint associated with the gauge symmetry. The energy penalty term plays the role of a Casimir operator in the group  $SU(N)$ , with eigenvalues proportional to  $\mathcal{J}$  and being the same for all states in an irreducible representation. This term in the Hamiltonian forces the low energy states to be singlets of the symmetry group.

In the absence of gauge symmetry, we allow the modes  $a_j^i$ s to couple independently with the thermal bath operator. Let us consider the following thermal coupling Hamiltonian,

$$H_{\text{thermal}} = \lambda_1 \sum_{ij} b_j^i a_i^{j(\dagger)} + \lambda_2 \sum_{ijkl} b_{jl}^{ik} a_i^{j(\dagger)} a_k^{l(\dagger)} + \lambda_3 \sum_{ijklpq} b_{jlpq}^{ikp} a_i^{j(\dagger)} a_k^{l(\dagger)} a_p^{q(\dagger)} + \dots + \text{H.c.} \quad (60)$$

One can show that the Lindblad Master equation takes the form of

$$\dot{\rho}(t) = \lambda_1 \frac{\text{Tr}(a^\dagger)}{\sqrt{N}} \rho \frac{\text{Tr}(a)}{\sqrt{N}} + \lambda_1 e^{-\frac{\mathcal{J}}{T}} \sum_{ij} \tilde{a}_j^{\dagger i} \rho \tilde{a}_i^j + \dots, \quad (61)$$

where  $\tilde{a}_j^i$  represents the component that transforms in the adjoint representation of  $SU(N)$ , such that

$$\tilde{a}_j^{\dagger i} = a_j^{\dagger i} - \frac{\text{Tr}(a^\dagger)}{N} \delta_j^i. \quad (62)$$

We show in Appendix C that it is a consequence of the  $H_{\text{Gauss}}$  term in the Hamiltonian.

These non-singlet errors acting on the ‘‘gauge invariant’’ logical states, i.e. logical states satisfying the Gauss law constraint, only corrupt quantum information at  $O(\frac{1}{N^2})$ , which we show in Appendix B 3. The summation in Eq. (61) is over  $N^2$  number of such errors and therefore the total effect of these nonsinglet errors is proportional to  $N^2 e^{-\frac{\mathcal{J}}{T}}$ . Hence the coupling strength  $\mathcal{J}$  has to be at least  $\sim 2T \log N$  in order to suppress their effects to the same order as those from the singlet errors. Note that this model effectively contains one such matrix oscillator and does not exhibit a Hagedorn-like transition. Nevertheless, it is clear that stability is only ensured when  $T \lesssim \mathcal{J}$ , hence can also depend on the system size.<sup>9</sup> In the following, we will assume their suppression with a strong enough coupling. As for the singlet error operators, they are defined as

$$E_{(\{n_k, m_k\})} := \prod_k \frac{\text{Tr}(P\{a^{\dagger n_k} a^{m_k}\})}{N^{\sum_k \frac{n_k + m_k}{2}}}. \quad (63)$$

The logical states we choose in this model are (divided by proper normalization constant),

$$|\tilde{1}\rangle = \frac{\text{Tr}(a^{\dagger L})^2}{\mathcal{N}(L)} |0\rangle,$$

$$|\tilde{2}\rangle = \frac{\text{Tr}(a^{\dagger L+1}) \text{Tr}(a^{\dagger L-1})}{\mathcal{N}(L)} |0\rangle. \quad (64)$$

We also define the orthonormal basis

$$|\tilde{\uparrow}\rangle = \frac{|\tilde{1}\rangle + |\tilde{2}\rangle}{2(1 + \langle \tilde{1} | \tilde{2} \rangle)},$$

$$|\tilde{\downarrow}\rangle = \frac{|\tilde{1}\rangle - |\tilde{2}\rangle}{2(1 - \langle \tilde{1} | \tilde{2} \rangle)}. \quad (65)$$

In order to recast the code subspace as a lower energy subspace, let us consider a shifted version of the system Hamiltonian and write the full Hamiltonian as

<sup>9</sup>This should be contrasted with the models with exact gauge symmetry, where the temperature that sets apart a stable and an unstable memory only depends on the natural frequency of the system. Note that in the local model with gauge symmetry, there is an additional scale introduced by  $J$  which has to satisfy  $J \gtrsim \omega \log N$  to ensure stability. However, as long as we have tuned  $J > T_c$ , then  $T_c$  will be the relevant energy scale for the low energy stability.

$$H = \frac{\mathcal{J}}{N} \sum_{ij} G_j^i G_i^j + \omega(\text{Tr}(a^\dagger a) - L)^2 + H_{\text{thermal}}, \quad (66)$$

where  $L > 0$  is some arbitrary integer.

One can check that the aKL condition,

$$\begin{aligned} \langle \tilde{i} | E_{\{(n'_k, m'_k)\}}^\dagger E_{\{(n_k, m_k)\}} | \tilde{j} \rangle &= f_{\{(n'_k, m'_k)\}, \{(n_k, m_k)\}} \delta_{ij} \\ &\quad + \frac{g_{\{(n'_k, m'_k)\}, \{(n_k, m_k)\}}^{ij}}{N^2} + O\left(\frac{1}{N^4}\right), \\ \langle \tilde{i} | E_{\{(n_k, m_k)\}} | \tilde{j} \rangle &= \frac{e_{\{(n_k, m_k)\}}^{ij}}{N} + O\left(\frac{1}{N^3}\right), \end{aligned} \quad (67)$$

holds for all relevant errors induced by (63) as long as they do not contain  $\text{Tr}(a^{\dagger L})$  or  $\text{Tr}(a^{\dagger L \pm 1})$  as a factor. For example, consider the class of error operators  $E_n = \frac{\text{Tr}(a^{\dagger n})}{\sqrt{n} N^{\frac{n}{2}}}$ . The relevant matrix elements are

$$\begin{aligned} \langle \tilde{\uparrow} | E_m^\dagger E_n | \tilde{\downarrow} \rangle &= \delta_{mn} \left[ \frac{2n}{N^2} + O\left(\frac{1}{N^4}\right) \right], \\ \langle \tilde{\uparrow} | E_m^\dagger E_n | \tilde{\uparrow} \rangle - \langle \tilde{\downarrow} | E_m^\dagger E_n | \tilde{\downarrow} \rangle &= O\left(\frac{1}{N^4}\right). \end{aligned} \quad (68)$$

Note that although the Hamiltonian and logical states depend on  $L$ , the error induced by a low energy operator has no such dependence [the  $O(\frac{1}{N^2})$  term does not depend on  $L$ ]; ergo, the aKL condition in the standard form Eq. (18) holds for such singlet errors (63). We present the details of the proof in Appendix B 3].

Similar to the discussion in Sec. III A, the number of singlet errors with the energy  $n^2 \omega$  scales as the number of integer partitions  $p(n)$ . This is clearly slower than exponential growth. Therefore, if the thermal coupling in (60) does not contain  $\text{Tr}(a^{\dagger L})$  and  $\text{Tr}(a^{\dagger L \pm 1})$  as factors, and  $\lambda \sim O(1)$ , then Theorem 1 applies and we again attain a memory time quadratic in  $N$ . Since the logical states are within the ground space, there are no spontaneous decay of quantum information. So the memory time is proportional to  $N^2 e^{\frac{\mathcal{J}}{\omega}}$  at zero temperature limit.

However, the low energy errors like  $\text{Tr}(a^{\dagger L}) \text{Tr}(a^L)$  do appear, which means that the aKL condition does not hold for these operators. Nevertheless, it is possible to retain this long memory time in this model as long as we also suppress the singlet errors containing such terms. One way to achieve this is to assume that the coupling constants are exponentially suppressed with respect to the number of modes involved, i.e.  $\lambda_k \sim \lambda^k$  for some  $\lambda < 1$ . This gaps out the dangerous error for large  $L \sim \log N$  by suppressing the error probability to order  $\frac{1}{N^2}$ . In the meantime, it is also a natural requirement in this model because the thermal bath mode couples to each oscillator individually.

One potential drawback in this model (and also the fully gauged model in Sec. III B) is that the coupling constants

$\mathcal{J}, J \propto \ln N$  depend on the system size. However, this is not as unreasonable as it looks because the interaction energy per harmonic oscillator mode is small<sup>10</sup> and is proportional to  $\frac{\ln N}{N^2} \omega$ . For reference, let us contrast this  $N$  dependence with that of the toric code model with system size equal to  $N$ . Recall that the toric code is a  $Z_2$  lattice gauge theory and the Hamiltonian also involves terms that are proportional to the Gauss's law operators, with coupling constant  $\Delta$ . When it is coupled to a bath with fixed temperature  $T$ , the memory time is proportional to  $\frac{1}{N} e^{\frac{\Delta}{T}}$ . It becomes a viable quantum memory if one can tune the coupling  $\Delta$  to  $T \ln N$ , which corresponds to an energy density much larger than that required in our matrix model.

## V. LARGE $N$ SPIN MODEL

In the previous sections, we saw that exact  $SU(N)$  symmetries in oscillator models can lead to robust quantum memories, both when the symmetry is fully gauged and when the singlet restriction is enforced energetically. It is natural to ask whether such exact symmetries and infinite dimensional Hilbert spaces are truly necessary for robustness. In other words, can we realize an analogous phenomenon with finite dimensional systems, e.g. involving qubits, that only have an approximate  $SU(N)$  symmetry? In this section, we propose one such model and show that the requisite large  $N$  factorization properties are mostly preserved up to small corrections in the low energy sector. As a result, the general conclusions of Theorem 1 still apply to this system in a way that is similar to Sec. IV.

To motivate the model, let us first review how one can approximately embed the low-energy states of an oscillator into a spin Hilbert space. Consider a spin- $j$  representation of  $SU(2)$ . When acting on states with  $S^z$  eigenvalue close to maximal,  $m \lesssim j$ , the spin commutator becomes

$$[S^x, S^y] = i\hbar S^z \approx i\hbar^2 j + \dots \quad (69)$$

This suggests that  $S^x$  and  $S^y$  can be approximately mapped to a canonical pair. Setting

$$Q = \frac{\ell}{\hbar \sqrt{j}} S^x, \quad P = \frac{1}{\ell \sqrt{j}} S^y \quad (70)$$

for some as yet undetermined length  $\ell$ , we see that

$$[Q, P] \approx i\hbar + \dots \quad (71)$$

Now considering the harmonic oscillator Hamiltonian

$$H = \frac{1}{2} (P^2 + \omega^2 Q^2) = \frac{1}{2} \left( \frac{1}{\ell^2 j} (S^y)^2 + \frac{\omega^2 \ell^2}{\hbar^2 j} (S^x)^2 \right), \quad (72)$$

<sup>10</sup>The energy for thermal state at temperature  $T \sim \omega \ll \mathcal{J}$  is  $O(1)$  times  $\mathcal{J}$ . See Appendix D for more detail.

we may set  $\ell^2 = \hbar/\omega$  to obtain

$$H = \frac{\omega}{2\hbar j} ((S^x)^2 + (S^y)^2) = \frac{\omega}{2\hbar j} (S^2 - (S^z)^2). \quad (73)$$

Using the standard spectrum of  $S^2$  and  $S^z$  with  $m = j - \Delta$ , the spectrum of  $H$  is

$$\frac{\hbar\omega}{2j} (j(j+1) - (j-\Delta)^2) = \hbar\omega \left( \Delta + \frac{1}{2} + O(\Delta^2/j) \right). \quad (74)$$

From this discussion, we expect that the physics of our oscillator models may be suitably extended by replacing the oscillators with qudits of dimension  $d$  ( $d = 2j + 1$  in the above mapping). At large  $d$  we have a nearly identical low-energy spectrum, but the models make sense for any  $d$  and we will show that even  $d = 2$  ( $j = 1/2$ ) suffices.

Specifically, we can construct a system of  $N^2$  spins, which can be thought of as qudits with local dimension  $d$ . It is tempting to write each spin operator in the adjoint representation of  $SU(N)$ , meaning each spin operator carries indices  $i, j$  running from 1 to  $N$  and transforms as

$$S_j^{i\pm} \rightarrow \sum_{kl} U_k^i S_l^{k\pm} U_j^{\dagger l}. \quad (75)$$

Each spin also satisfies the following  $SU(2)$  algebra,

$$\begin{aligned} [S_i^{j+}, S_j^{i-}] &= 2S_i^{jz}, \\ [S_i^{jz}, S_i^{j+}] &= S_i^{j+}, \\ [S_i^{jz}, S_j^{i-}] &= -S_j^{i-}, \end{aligned} \quad (76)$$

where  $i, j$  again label the different spins or qudits. However, the  $SU(N)$  symmetry does not act consistently on the spin algebra due to the first line above: we cannot interpret  $S_i^{jz}$  as being in the adjoint of  $SU(N)$  while also obeying the commutation relation. Moreover, even if we grant that  $S_i^{jz}$  is not in the adjoint representation, it is impossible to construct  $SU(N)$  generators  $G_i^j$  with the desired commutation relations with  $S_i^{\pm}$ , i.e. there is no  $G_i^j$  obeying<sup>11</sup>

$$[G_i^j, S_k^{\pm}] = \delta_i^k S_k^{j\pm} - \delta_k^j S_i^{\pm}. \quad (77)$$

However, it is still possible to realize an approximate symmetry in terms of the pseudogenerators,

<sup>11</sup>The inconsistency of  $SU(N)$  and the  $SU(2)$  algebra can be seen from the equation  $\sum_k [[S_k^+, S_j^-], S_h^+] = \delta_j^i \delta_h^+ S_h^+$ , where the right-hand side transforms differently from the left-hand side.

$$\tilde{G}_i^j = \sum_k S_k^{j+} S_i^{k-} - S_i^{k+} S_k^{j-},$$

$$[\tilde{G}_i^j, S_k^{\pm}] = (\delta_i^k S_k^{j\pm} - \delta_k^j S_i^{\pm}) S_k^z. \quad (78)$$

Indeed, if  $S^z$  could be approximated as a constant operator, then we would have the desired commutators. This approximate algebra thus connects back with the oscillator models of the previous sections where the symmetry becomes exact. In what follows, we work with the pseudogenerators  $\tilde{G}_i^j$  and set  $d = 2$ .

Similar to our recipe in the previous sections, one can construct the analogs of the gauge invariant states by acting multitrace operators on the tensor product of spin down states

$$\begin{aligned} &\text{Tr}(S^{+n_1}) \text{Tr}(S^{+n_2}) \dots \text{Tr}(S^{+n_k}) |0\rangle, \\ |0\rangle &:= |\downarrow\downarrow \dots \downarrow\rangle. \end{aligned} \quad (79)$$

For convenience, we still refer to these analogs of the gauge invariant operators (states) singlet operators (states). Indeed, one can see that the desired symmetry is only approximate because the algebra of such operators is not closed. For example,

$$\begin{aligned} \text{Tr}(S^{-2}) \text{Tr}(S^{+2}) \text{Tr}(S^{+2}) |0\rangle &= (4N^2 - 4N - 8) \text{Tr}(S^{+2}) |0\rangle \\ &+ 32 \sum_i S_i^{+} S_i^{+} |0\rangle. \end{aligned} \quad (80)$$

Generally, the product of such singlet terms can produce nonsinglet terms that are only subleading in  $N$ .<sup>12</sup>

Now let us consider a system Hamiltonian  $H_S = \hbar H_0 + \mathcal{J} H_G$ , where

$$H_0 = \left( \sum_{ij} S_j^{iz} - S_{\text{tot}} \right)^2, \quad (81)$$

and  $-\frac{N^2}{2} \leq S_{\text{tot}} \leq \frac{N^2}{2}$  is some  $N$ -dependent offset.  $H_G$  is an energy penalty term similar to the Gauss law constraint; it corresponds to the Casimir operator that enforces the approximate  $SU(N)$  symmetry for low energy states,<sup>13</sup>

$$H_G = \frac{1}{N} \tilde{G}^2 := \frac{1}{N} \sum_{ij} \tilde{G}_i^j \tilde{G}_j^i. \quad (82)$$

<sup>12</sup>If we further restrict ourselves to only consider spin- $\frac{1}{2}$ , i.e.  $d = 2$ , then the nonsinglet terms also vanishes. This can be a convenient simplification when the local systems are exactly qubits; however, our analysis of robustness will also hold for systems with  $d > 2$ .

<sup>13</sup>The Casimir operator  $C = \sum_{ij} G_i^j G_j^i$  for an exact  $SU(N)$  symmetry has a spectrum of the form  $NwJ$ , where  $w$  is the highest weight of given representation. Analogously, the Hamiltonian  $H_G$  also has an energy spectrum of the form of  $wJ$  up to  $\frac{1}{N}$  corrections.

When  $\mathcal{J}$  is large, it creates a large gap in the spectrum. We define the low energy sector as the states with eigenvalues suppressed by  $\frac{1}{N}$ . The energy eigenstates of  $H_S$  are generally superpositions of the singlet and nonsinglet states, but it can be shown that the low energy sector is isomorphic to the singlet subspace. For example, the following are four states created by product of three  $S^+$ s, which are simultaneous eigenstates of both  $H_0$  and  $H_G$ ,

$$\begin{aligned} |\psi_1\rangle &= \text{Tr}(S^{+3})|0\rangle, \\ |\psi_2\rangle &= [\text{Tr}(S^{+3}) - 3\text{Tr}(S^{+2})\text{Tr}(S^+)]|0\rangle, \\ |\psi_3\rangle &= \left[ \text{Tr}(S^{+2})\text{Tr}(S^+) + \left(1 + \frac{2}{N}\right)\text{Tr}(S^+)^3 \right. \\ &\quad \left. - \frac{19}{2N} \sum_i (S^{+2})^i S_i^{+i} \right] |0\rangle, \\ |\psi_4\rangle &= \left[ \sum_i (S^{+2})^i S_i^{+i} - \frac{1}{N} \text{Tr}(S^{+2})\text{Tr}(S^+) \right] |0\rangle. \end{aligned} \quad (83)$$

The first three states are in the low energy sector, since they are eigenstates of  $H_G$  with eigenvalues  $E_1^G = E_2^G = 0$  and  $E_3^G = \frac{2\mathcal{J}}{2N}$ . The fourth state has eigenvalue  $E_4^G = \mathcal{J}$ , and hence a high energy excitation.

There are two potential causes for this model to deviate from the one in Sec. IV—effects from the symmetry being approximate and finite dimensionality of each spin. Formally, we should still expect the low energy states, i.e. ones with energy  $\ll \mathcal{J}$ , to have a one-to-one correspondence with the multitrace operators formed by  $S^+$ s. More precisely, their low energy spectra should coincide up to corrections that are  $1/N$  suppressed and the multiplicity, when we ignore such corrections, should be identical.

We see this by noting that their energy only depend on number of  $S^+$  involved, with small corrections proportional to  $\frac{\mathcal{J}}{N}$  (Fig. 6). Therefore, up to these small corrections introduced by the deviations from the exact symmetry, the operator counting at each energy level is the same as that in

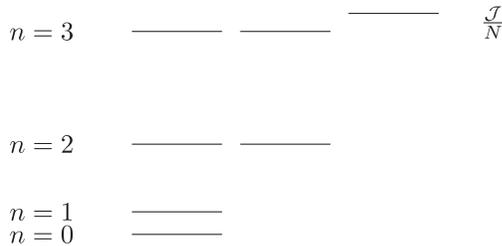


FIG. 6. Spectrum of the low energy states. The ground state has energy zero. Excited states have energy proportional to  $n^2 h$ . Note that due to the nonexactness of symmetry, there are small energy splittings proportional to  $\frac{\mathcal{J}}{N}$ . For sufficiently large  $N$ , this small gap can be neglected and the degeneracy at each level is the same as in harmonic oscillator model. Nonsinglet excitations have energy  $O(\mathcal{J})$ , which we do not count as low energy.

Sec. IV. These corrections can be connected to the breakdown of the exact symmetry.

As for the correction from finite dimensionality, thanks to the ‘‘Gauss law’’ constraint  $H_G$  and the form of  $H_0$ , these corrections would only significantly modify this counting when the total number of  $S^+$  involved is of  $O(N)$ . Therefore, in the low energy regime where we operate, both effects are suppressed by having a large enough  $N$ .

Similar to the harmonic oscillator model, we consider the full Hamiltonian,

$$H = hH_0 + \mathcal{J}H_G + H_{\text{thermal}}, \quad (84)$$

with  $H_{\text{thermal}}$  being

$$\begin{aligned} H_{\text{thermal}} &= \lambda_1 \sum_{ij} b_j^i S_i^{j\pm} + \lambda_2 \sum_{ijkl} b_{ji}^{ik} S_i^{j\pm} S_k^{l\pm} \\ &\quad + \lambda_3 \sum_{ijklpq} b_{jlpq}^{ikp} S_i^{j\pm} S_k^{q\pm} + \dots + \text{H.c.} \end{aligned} \quad (85)$$

The lowest energy states must have total spin equal to  $S_{\text{tot}}$ . Let us define  $L = \frac{S_{\text{tot}} + N^2/2}{2}$  and pick a two-dimensional subspace as the code subspace. With proper normalization,

$$\begin{aligned} |\tilde{1}\rangle &= \frac{\text{Tr}(S^{+L})^2}{\mathcal{N}} |0\rangle, \\ |\tilde{2}\rangle &= \frac{\text{Tr}(S^{+L+1})\text{Tr}(S^{+L-1})}{\mathcal{N}} |0\rangle. \end{aligned} \quad (86)$$

Again, we define the orthonormal basis in this logical subspace as

$$\begin{aligned} |\tilde{\uparrow}\rangle &= \frac{|\tilde{1}\rangle + |\tilde{2}\rangle}{2(1 + \langle \tilde{1} | \tilde{2} \rangle)}, \\ |\tilde{\downarrow}\rangle &= \frac{|\tilde{1}\rangle - |\tilde{2}\rangle}{2(1 - \langle \tilde{1} | \tilde{2} \rangle)}, \end{aligned} \quad (87)$$

with the overlap  $\langle \tilde{1} | \tilde{2} \rangle = \frac{f(L)L\sqrt{L^2-1}}{\sqrt{2N^2}}$ , where  $f(L)$  is a complicated function defined in Appendix B 4.

Let us couple this system to a thermal bath at temperature  $T$ . First note that the nonsinglet excitations are again suppressed by a factor of  $e^{-\frac{\mathcal{J}}{T}}$  because of the energy penalty term  $H_G$ , while the singlet interaction has no such suppression. Then let us reorganize the system-bath interactions by the approximate symmetry such that

$$\begin{aligned} H_{\text{thermal}} &= \lambda_1 \sum_{ij} b_j^i S_i^{j\pm} + \lambda_2 \sum_{ijkl} b_{ji}^{ik} S_i^{j\pm} S_k^{l\pm} + \dots + \text{H.c.}, \\ &= \lambda_1 \sum_{ij} b_j^i \left( S_i^{j+} - \frac{\text{Tr}(S^+)}{N} \delta_i^j \right) + \lambda_1 \frac{\text{Tr}(b)}{\sqrt{N}} \frac{\text{Tr}(S^+)}{\sqrt{N}} \\ &\quad + \dots \end{aligned} \quad (88)$$

Similar to the discussion in last section, this interaction induces an order  $N^2$  number of nonsinglet errors. Therefore, the total error contribution from nonsinglet excitations is  $\sim N^2 e^{-\frac{\mathcal{J}}{T}}$  times that of singlet excitations. To suppress both types of errors to the same order, we require the coupling  $\mathcal{J}$  to be greater than  $2h \log N$ .

More precisely, the singlet error operators are defined as

$$E_{\{(n_k, m_k)\}} := \prod_k \frac{\text{Tr}(P\{S^{+n_k} S^{-m_k}\})}{N^{\sum_k \frac{n_k + m_k}{2}}}. \quad (89)$$

Again, we show that the aKL condition

$$\begin{aligned} \langle \tilde{i} | E_{\{(n'_k, m'_k)\}}^\dagger E_{\{(n_k, m_k)\}} | \tilde{j} \rangle &= f_{\{(n'_k, m'_k)\}, \{(n_k, m_k)\}} \delta_{ij} \\ &\quad + \frac{g_{\{(n'_k, m'_k)\}, \{(n_k, m_k)\}}^{ij}}{N^2} + O\left(\frac{1}{N^4}\right), \\ \langle \tilde{i} | E_{\{(n_k, m_k)\}} | \tilde{j} \rangle &= \frac{e_{\{(n_k, m_k)\}}^{ij}}{N} + O\left(\frac{1}{N^3}\right) \end{aligned} \quad (90)$$

is satisfied in Appendix B 4. For example, one can see that the operator  $\frac{\text{Tr}(S^+ S^-)}{N} = \frac{\sum_{ij} S_j^z}{N}$  acts as the logical identity, since the logical states have the same total spin in  $z$  direction. For another example, the aKL conditions for errors  $E_n = \frac{\text{Tr}(S^{+n})}{\sqrt{n} N^{\frac{n}{2}}}$  are

$$\begin{aligned} \langle \tilde{\uparrow} | E_m^\dagger E_n | \tilde{\downarrow} \rangle &= \delta_{mn} \left[ \frac{2n}{N^2} + O\left(\frac{1}{N^4}\right) \right], \\ \langle \tilde{\uparrow} | E_m^\dagger E_n | \tilde{\uparrow} \rangle - \langle \tilde{\downarrow} | E_m^\dagger E_n | \tilde{\downarrow} \rangle &= O\left(\frac{1}{N^4}\right). \end{aligned} \quad (91)$$

Note that, in these examples, and more generally in Appendix B 4, deviations from the KL condition only depend on  $n$  but not on  $L$ . Therefore, the aKL condition remains valid even for very large  $L$ . Similar to the last section, we can pick  $L \gtrsim \log N$  to make sure uncorrectable errors are suppressed for large  $N$ . For instance, one may consider choosing  $L = N^2/4$  such that  $S_{\text{tot}} = 0$ .

Since this spin model has a structure very similar to the ungauged harmonic oscillator model, we can follow the same argument in Sec. IV. In the low energy sector with  $\epsilon \ll \mathcal{J} \sim \log N$ , (approximate) singlet errors are the dominant error modes. The number of such error operators grows subexponentially with energy because these operators still have the multitrace structure whose counting are the same as the one in Sec. IV. There are small corrections because of the approximate symmetry, however they are proportional to  $\frac{\mathcal{J}}{N}$ , and therefore for large  $N$ , they do not affect the overall structure of the spectrum for the purpose of our analysis (Fig. 6). Similarly, we also need the thermal coupling strength to decay exponentially with number of spins to suppress the singlet error operators such as

$\text{Tr}(S^{-L})\text{Tr}(S^{+L})$ . As argued in the previous section, this is a natural assumption since the bath mode couples to each spin individually. Since the rest of the argument is identical to those in Sec. IV, we can then invoke Theorem 1 to conclude that the memory time  $t_m$  scales quadratically with  $N$ . There is no operator that can lower the logical states' energy, so the zero temperature limit of memory time is still proportional to  $N^2 e^{\frac{\epsilon}{T}}$ . We remark that this model also has the same drawback as the ungauged harmonic oscillator model, that is, the interactions remain nonlocal and we need the coupling strength  $\mathcal{J}$  to scale as at least  $\log N$ .

## VI. DISCUSSION AND CONCLUSION

Motivated by connections between error correction and phenomena in gauge theory and quantum gravity, we have taken a first step in understanding how  $SU(N)$  gauge symmetries at large  $N$  can be used to construct stable quantum memories. In particular, we construct a gauged matrix quantum mechanics toy model with nonlocal logical operations that is a self-correcting quantum memory. The memory time of this model is polynomial in  $N$  when coupled to a thermal bath at nonzero temperature. The presence of gauge symmetry induces approximate quantum error correcting codes that become exact in the large  $N$  limit, thanks to the large  $N$  factorization property. Furthermore, gauge invariance limits the number of error patterns that can be excited up to a certain energy scale. The combination of these factors suppresses the proliferation of memory errors under thermal noise. A similar model with local logical operation is also possible, but as a trade-off, we impose an additional energy penalty with a coupling of strength  $\sim \log N$ . The memory time is again polynomial in  $N$  but only for temperatures below a critical value.

As large  $N$  gauge symmetries are not known to occur in nature, we relax the gauge symmetries and impose only large  $N$  global symmetries energetically on a single matrix oscillator. We also construct a nonlocal finite dimensional spin model which further relaxes the symmetry constraint where we only enforce approximate large  $N$  symmetries energetically. Interestingly, these constructions remain robust in both scenarios under the assumed system-bath coupling as long as the symmetry constraints can be energetically enforced with a penalty that scales as  $O(\log N)$ .

Stepping back from these particular examples, our work provides one set of guidelines for building large  $N$  passive quantum memories in which the encoded information remains robust when coupled to a thermal bath. Sufficient conditions on the quantum system and its coupling to the bath are captured in Theorem 1. We hope to extend our current work in a number of future directions. Let us roughly categorize these potential generalizations in the direction of optimality, dimensionality, symmetries, variety, and practicality.

Although we can construct models that have  $\text{poly}(N)$  memory time, they also often come with other features that

are practically difficult, such as nonlocal interactions and  $\log N$  coupling strengths. However, we by no means claim that these are the optimal constructions. Indeed, it is possible that a more careful design can produce better self-correcting memories. For example, an interesting theoretical problem is whether there exists a gauged  $SU(N)$  oscillator model with local coupling for which the quantum memory is stable against thermal noise.

It is also natural to generalize these findings to systems with gauge symmetries in higher dimensions by considering geometrically coupled oscillators, lattice gauge theories, and concatenations with other quantum error correcting codes. We will take a first step towards tackling a subset of these problems in an upcoming work [12]. Although we enforce global symmetries energetically in our work, one can similarly ask whether there are advantages in considering systems with emergent gauge symmetries at low energies, which is better understood in higher dimensional systems. Relatedly, one can examine what aspects of the  $SU(N)$  symmetry are actually essential. For instance, does a similar result hold for discrete, instead of continuous, non-Abelian symmetries?

Drawing intuition from gauge theories, we can also examine how to generalize such models by coupling the “pure gauge” degrees of freedom to fundamental degrees of freedom like matter fields and defects. In particular, how does the configuration of these other degrees of freedom impact our conclusions? One may, for instance, introduce a variety of fields that may interact under these gauge couplings. Extensions in this direction draw heavy parallels with traditional theoretical particle physics, and can open up new connections with quantum information in this area.

Lastly, as a practical aspect of any quantum memory, information needs to be encoded and extracted relatively easily. This is yet unexplored for both the infinite dimensional oscillator and the finite dimensional spin system. As a standalone quantum code, the gauged matrix oscillator is a continuous variable system that resembles bosonic codes [32] in many ways. However, we have barely begun to analyze its code properties. It is desirable to understand its similarities and differences with existing bosonic codes, its robustness against different kinds of errors, and how active error correction like syndrome extraction can be performed.

## ACKNOWLEDGMENTS

We thank Alexey Milehkin, John Preskill for helpful comments and discussions. C. C. acknowledges the support by the U.S. Department of Defense and National Institute of Standards and Technology (NIST) through the Hartree Postdoctoral Fellowship at QuICS, the Air Force Office of Scientific Research (Grant No. FA9550-19-1-0360), and the National Science Foundation (Grant No. PHY-1733907). G. C. acknowledges support from the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Accelerated

Research for Quantum Computing program “FAR-QC.” B. G. S. acknowledges support from the AFOSR under Grant No. FA9550-19-1-0360.

## APPENDIX A: PROOF OF THEOREM 1

In this appendix, we will prove that the mutual information decays polynomially in  $t$  when coupled to a thermal bath of suitably low temperature, as long as the conditions in Theorem 1 are satisfied. We first set up some useful notations for later use.

Let  $\mathcal{E}$  be the set of operators that generate the errors. For all operators  $E_a \in \mathcal{E}$  satisfying the aKL equation, we can always diagonalize the matrix  $f_{ab}$  by a basis transformation inside  $\mathcal{E}$ , such that  $\langle \tilde{i} | E_b^\dagger E_a | \tilde{i} \rangle = f_a \delta_{ab} + O(\frac{1}{N^2})$ . The matrix elements here are taken with respect to the  $d$  dimensional code words. Then the aKL equation can be written as

$$\langle \tilde{j} | E_b^\dagger E_a | \tilde{i} \rangle = f_a \delta_{ab} \delta_{ij} + \frac{g_{ab}^{ij}}{N^2}, \quad (\text{A1})$$

where  $f_a \geq 0$  is some  $O(1)$  number that depends on the error indices.

Now we define two classes of errors according to whether  $f_a$  is 0 or not. *The first class of errors*, labeled by  $E_\alpha$  and  $E_\beta$ , satisfies  $f_\alpha, f_\beta \neq 0$ . *The second class of errors*, labeled by  $E_\sigma$  and  $E_\nu$ , satisfies  $f_\sigma = f_\nu = 0$ . Note that we will label the errors with greek indices,  $\alpha, \beta$  for the first class,  $\sigma, \nu$  for the second class, when we distinguish their classes. When we do not need to treat them in different classes, we use Latin indices like  $a, b$  to label these errors. For the first class of errors, we can define an orthonormal basis  $\{|\mu_\alpha^i\rangle\}$  by acting error operators  $E_\alpha$  on the logical states  $|\tilde{i}\rangle$  such that

$$\begin{aligned} E_\alpha |\tilde{i}\rangle &= \sqrt{f_\alpha} |\mu_\alpha^i\rangle + \sum_\beta \frac{c_{\alpha\beta}}{2\sqrt{f_\beta} N^2} |\mu_\beta^i\rangle + \sum_\beta \frac{\Delta_{\alpha\beta}^i}{2\sqrt{f_\beta} N^2} |\mu_\beta^i\rangle \\ &+ \sum_{\beta, j \neq i} \frac{h_{\alpha\beta}^{ij}}{2\sqrt{f_\beta} N^2} |\mu_\beta^j\rangle. \end{aligned} \quad (\text{A2})$$

The states  $|\mu_\alpha^i\rangle$  s can be read off from the  $O(1)$  term in above equations. All the other coefficients can be read off from the aKL equations as

$$\begin{aligned} c_{\alpha\beta} &= \frac{1}{d} \sum_{i=1}^d g_{\alpha\beta}^{ii}, \\ \Delta_{\alpha\beta}^i &= g_{\alpha\beta}^{ii} - c_{\alpha\beta}, \\ h_{\alpha\beta}^{ij} &= \begin{cases} g_{\alpha\beta}^{ij}, & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases}. \end{aligned} \quad (\text{A3})$$

One can see that there is a clear separation between the phase errors that are captured by  $\Delta_{\alpha\beta}^i$ , and the bit flip errors

that are captured by  $h_{\alpha\beta}^{ij}$ . So far we have discussed the first class of errors with nonzero  $O(1)$  norm  $\sqrt{f_\alpha}$ . However, the second class of error operators have vanishing  $O(1)$  norm and let us label them by  $E_\sigma$  and  $E_\nu$ , so  $\langle E_\sigma^\dagger E_\sigma \rangle = O(\frac{1}{N^2})$ . Generally, the aKL equations for these operators take the form of

$$\langle \tilde{j} | E_\sigma^\dagger E_\nu | \tilde{i} \rangle = \frac{g_{\sigma\nu}^{ij}}{N^2} + O\left(\frac{1}{N^3}\right). \quad (\text{A4})$$

These operators in the second class of errors map the logical states to states with norms proportional to  $O(\frac{1}{N})$ . More generally, they are operators that satisfy the following equation:

$$E_\sigma | \tilde{i} \rangle = \sum_\nu \frac{s_{\sigma\nu}^i}{N} | \mu_\nu^i \rangle, \quad (\text{A5})$$

for some  $O(1)$  coefficients  $ss$ . The states  $|\mu_\nu^i\rangle$ s are another set of normalized basis, orthogonal to the states  $|\mu_\alpha^i\rangle$  s up to  $O(\frac{1}{N})$  overlap. The errors that satisfy

$$\langle \tilde{i} | E_\sigma | \tilde{j} \rangle = \frac{e_{\sigma}^{ij}}{N} + O\left(\frac{1}{N^3}\right) \quad (\text{A6})$$

also belong to the second class of errors, since they are consistent with Eq. (A5).

For some calculations, we do not need to differentiate the two types of errors. In that case, we use latin indices and simply label all of them by  $E_a$ . Let us denote the KL matrix element by  $\langle \tilde{j} | E_b^\dagger E_a | \tilde{i} \rangle = G_{ab}^{ij}$ ; one can write down a general relation

$$E_a | \tilde{i} \rangle = \sum_A u_{aA}^i | \mu_A \rangle, \quad (\text{A7})$$

where  $\{|\mu_A\rangle\}$  is a set of orthonormal states, and  $u_{aA}^i$  are coefficients that satisfy

$$\sum_A u_{aA}^i u_{bA}^j = G_{ab}^{ij}. \quad (\text{A8})$$

### 1. Early time approximation

Recall that we model our open system dynamics as a Markovian process. At early time, we approximate the evolution by only including the first order perturbation of errors.<sup>14</sup>

<sup>14</sup>The coupling  $\lambda_a(\epsilon_a)$  for each error operator  $E_a$  is defined as  $\lambda_a \gamma(\epsilon_a)$ , with  $\gamma(\epsilon_a)$  defined in Eq. (22).

$$\begin{aligned} \rho(t) \approx & \rho_0 - \sum_a \frac{\lambda_a(\epsilon_a)t}{2} \{E_a^\dagger E_a, \rho_0\} + \sum_a \lambda_a(\epsilon_a)t E_a \rho_0 E_a^\dagger \\ & + \dots, \end{aligned} \quad (\text{A9})$$

where the sum over  $E_a$  includes both the first and second class of errors, and  $\rho_0$  is the maximally entangled state between reference  $R$  and system  $S$  at  $t = 0$ ,

$$\rho_0 = \frac{1}{d} \sum_{ij} | \tilde{i} \rangle_{RS} \langle \tilde{j} |. \quad (\text{A10})$$

We proceed to compute the mutual information  $I(R:S) = S(\rho^R) + S(\rho^S) - S(\rho)$  at early times. Since the nontrivial unitary dynamics that induces the above Markovian evolution only occurs between the system and the bath, the entropy of  $R$  should remain invariant over time, i.e.  $S(\rho^R) = \ln d$ . Therefore, it suffices to calculate the entropies of  $S$  and the  $RS$  joint system.

Generally, the density matrix is block diagonal. We can generate each block matrix using the operators  $E_a$ , with  $a \in \mathcal{B}_n$ , such that for any  $a, b \in \mathcal{B}_n$ ,  $G_{ab} \neq 0$ . Here  $\mathcal{B}_n$  is an index set that captures the error operators used to generate a particular block matrix. Sometimes we will also refer to this set as a block, since it defines the corresponding block matrix. We write the density matrix associated to the block  $\mathcal{B}_n$  as

$$\begin{aligned} \rho_n &= \sum_{a \in \mathcal{B}_n} E_a \rho_0 E_a^\dagger, \\ &= \frac{1}{d} \sum_{aABij} u_{aA}^i u_{aB}^{j*} | \mu_A \rangle_S \langle \mu_B | \otimes | i \rangle_R \langle j|. \end{aligned} \quad (\text{A11})$$

The eigenvalues of matrix  $\frac{1}{d} \sum_a u_{aA}^i u_{aB}^{j*}$  is the same as that of the matrix  $\frac{1}{d} \sum_{iA} u_{aA}^i u_{bA}^{i*} = \frac{1}{d} \sum_i G_{ab}^{ii}$ . After tracing out the reference  $R$ , the density matrix of system becomes

$$\rho_n^S = \frac{1}{d} \sum_{aiAB} u_{aA}^i u_{aB}^{i*} | \mu_A \rangle_S \langle \mu_B|. \quad (\text{A12})$$

The eigenvalues of matrix  $\frac{1}{d} \sum_{ai} u_{aA}^i u_{aB}^{i*}$  is the same as that of the matrix  $\frac{1}{d} \sum_A u_{aA}^i u_{bA}^{i*} = \frac{1}{d} G_{ab}^{ij}$ .

Therefore, to evaluate the entropy, we need to calculate the eigenvalues of the matrix  $\frac{1}{d} G_{ab}^{ij}$ . Generally, inside a block, we have both the first and the second class of errors. Their overlap  $G_{\alpha\sigma}$  scales as  $\frac{1}{N^2}$  by the aKL condition. However, we can always diagonalize the matrix  $G$ , where the diagonal form  $\tilde{G}$  has vanishing matrix elements for the overlaps  $\tilde{G}_{\alpha\sigma} = 0$  and  $\tilde{G}_{\alpha\beta} - G_{\alpha\beta} = O(\frac{1}{N^4})$ ,  $\tilde{G}_{\sigma\nu} - G_{\sigma\nu} = O(\frac{1}{N^4})$  such that the correction is at subleading order. Therefore, in the following, we can neglect the overlap between the first and the second class of errors, and always treat them as different blocks.

For our following discussion, we use  $\rho_n^{(1)}, \rho_n^{(2)}$  to denote the state created purely by the first and the second class of errors, respectively. The entropy for  $\rho_n^{(1)}$  is

$$\begin{aligned} S(\rho_n^{(1)}) &= -\text{Tr}(\rho_n^{(1)} \ln \rho_n^{(1)}), \\ &= -\text{Tr} \left[ \left( f + \frac{c}{N^2} \right) \ln \left( f + \frac{c}{N^2} \right) \right], \\ &= -\text{Tr}(f \ln f) - \frac{1}{N^2} \text{Tr}(c \ln f) - \frac{1}{N^2} \text{Tr}(c) - \frac{1}{2N^4} \text{Tr}(cf^{-1}c) + O\left(\frac{1}{N^6}\right). \end{aligned} \quad (\text{A13})$$

In the second step we used that the eigenvalues of the density matrix are the same as those of the KL overlap matrix  $\frac{1}{d} \sum_i G^{ii}$ .  $f$  is the diagonal matrix  $f_{\alpha\alpha} \delta_{\alpha\beta}$  and the matrix elements of  $c$  are given by Eq. (A3). For the reduced density matrix of system  $S$ ,  $\rho_n^S = \text{Tr}_R(\rho_n)$ , its entropy is

$$\begin{aligned} S(\rho_n^{(1)S}) &= -\text{Tr} \left[ \frac{1}{d} \left( f \otimes I + \frac{g}{N^2} \right) \ln \left( f \otimes I + \frac{g}{N^2} \right) \right] + \frac{1}{d} \text{Tr} \left( f \otimes I + \frac{g}{N^2} \right) \ln d, \\ &= \text{Tr}(\rho_n) \ln d - \frac{1}{d} \text{Tr}[(f \otimes I) \ln(f \otimes I)] - \frac{1}{dN^2} \text{Tr}[g \ln(f \otimes I)] - \frac{1}{dN^2} \text{Tr}(g) - \frac{1}{2dN^4} \text{Tr}[g(f^{-1} \otimes I)g] + O\left(\frac{1}{N^6}\right), \end{aligned} \quad (\text{A14})$$

where  $f$  is the matrix with element being  $f_{\alpha\beta}$ , and  $I$  is the  $d \times d$  identity matrix. After subtracting  $S(\rho_n^{(1)})$  from  $S(\rho_n^{(1)S})$ , we finally obtain

$$\begin{aligned} S(\rho_n^{(1)S}) - S(\rho_n^{(1)}) &= \text{Tr}(\rho_n^{(1)}) \ln d - \frac{1}{2dN^4} \text{Tr}[g(f^{-1} \otimes I)g] + \frac{1}{2N^4} \text{Tr}(cf^{-1}c) + O\left(\frac{1}{N^6}\right), \\ &= \text{Tr}(\rho_n^{(1)}) \ln d - \frac{1}{2dN^4} \left( \sum_i \text{Tr}[\Delta^i f^{-1} \Delta^i] + \sum_{ij} \text{Tr}[h_{ij} f^{-1} h_{ij}^T] \right) + O\left(\frac{1}{N^6}\right), \end{aligned} \quad (\text{A15})$$

where matrices  $\Delta^i$  and  $h_{ij} := h^{ij}$  are defined in Eq. (A3). We see that the decay rate of mutual information is suppressed by  $\frac{1}{N^4}$ . Repeating the same subtractions but for the second class of errors, the difference in entropy is

$$\begin{aligned} S(\rho_n^{(2)S}) - S(\rho_n^{(2)}) &= \frac{1}{N^2} \left[ S\left(\frac{1}{d} g_{\sigma\nu}^{ij}\right) - S\left(\frac{1}{d} \sum_i g_{\sigma\nu}^{ii}\right) \right] \Big|_{\sigma,\nu \in \mathcal{B}_n} + O\left(\frac{1}{N^4}\right), \\ &= \text{Tr}(\rho_n^{(2)}) \ln d - \frac{1}{N^2} \left[ S\left(\frac{1}{d} \sum_i g_{\sigma\nu}^{ii}\right) - \frac{1}{d} S(g_{\sigma\nu}^{ij}) \right] \Big|_{\sigma,\nu \in \mathcal{B}_n} + O\left(\frac{1}{N^4}\right), \end{aligned} \quad (\text{A16})$$

where  $S(g_{\sigma\nu}^{ij})$  means that we are calculating entropy of the matrix whose elements are  $g_{\sigma i, \nu j}$ . On the other hand, in  $S(\frac{1}{d} \sum_i g_{\sigma\nu}^{ii})$ , we compute the entropy of the matrix with elements proportional to  $\frac{1}{d} \sum_i g_{\sigma\nu}^{ii}$ . The decay rate is only suppressed in  $\frac{1}{N^2}$ . The entropy difference in the second term depends on the matrix elements  $g_{\sigma\nu}^{ij}$  in a complicated way. Instead of calculating it explicitly, we just provide an upper bound for this expression. Suppose that in each block  $\mathcal{B}_n$ , the matrix  $g_{\sigma\nu}^{ij}$  has a rank of  $r(n)$  with its elements bounded by a function  $f(n)$ , then the second term is bounded by  $\frac{r(n)f(n)\log d}{N^2}$ .

For  $n \geq 1$ , define  $\alpha_n = \lambda_n t$ . The early time density matrix can be written as<sup>15</sup>

$$\rho(t) = \bigoplus_{n=0}^{\infty} (\alpha_n \rho_n). \quad (\text{A17})$$

<sup>15</sup>To keep the notation tractable, we absorbed any  $a$  dependence of the coupling  $\lambda_a$  into the operators  $E_a$  such that the couplings  $\lambda$  only retain the  $n$  dependence. To be more precise, one can redefine the couplings and errors such that  $\lambda(\epsilon_a)$  only depends on the energy of each error operator  $\epsilon_a$ . Because  $\epsilon_a = \epsilon_n$  for all  $a \in \mathcal{B}_n$ , we can rewrite  $\lambda(\epsilon_a)$  as  $\lambda_n$ .

Applying the relation  $\sum_n \alpha_n \text{Tr}(\rho_n) = 1$ , one find that the difference of total entropy is

$$S(\rho^S(t)) - S(\rho(t)) = \ln d - \sum_{n=1}^{\infty} \frac{\lambda_n t}{N^2} \left[ S\left(\frac{1}{d} \sum_i g_{\sigma\nu}^{ii}\right) - \frac{1}{d} S(g_{\sigma\nu}^{ij}) \right] \Big|_{\sigma,\nu \in \mathcal{B}_n} \\ - \sum_{n=1}^{\infty} \frac{\lambda_n t}{2dN^4} \sum_{\alpha\beta \in \mathcal{B}_n} \left[ \sum_i \Delta_{\alpha\beta}^{i2} + \sum_{i \neq j} (h_{\alpha\beta}^{ij})^2 \right] + O\left(\frac{1}{N^6}\right). \quad (\text{A18})$$

For early time, i.e.  $\lambda_n t \ll 1$ , the dominant contribution to information decay is from the second class of errors.

## 2. Finite time

In this section, we look at the finite time behavior of density matrix up to  $O(\frac{1}{N^2})$  contributions. We will see that, although the early time effect of the first class of errors is subdominant, it becomes more relevant at later times since it grows quadratically in time. Recall that the overlap between the different classes of errors are suppressed, we can split the Lindbladian operator into two parts associated with the first class and the second class of errors,

$$\rho(t) = e^{\mathcal{L}t} \rho(0) = e^{\mathcal{L}^{(1)t + \mathcal{L}^{(2)t}} \rho(0) + O\left(\frac{1}{N^4}\right)}. \quad (\text{A19})$$

From Eq. (A5), the second class of Lindblad operator creates a term of  $O(\frac{1}{N^2})$  each time it acts on the density matrix. Therefore, the state can be approximated by

$$\rho(t) = \rho^{(1)}(t) + t \mathcal{L}^{(2)}[\rho^{(1)}(t)] + O\left(\frac{1}{N^4}\right), \quad (\text{A20})$$

with  $\rho^{(1)}(t) = e^{\mathcal{L}^{(1)t} \rho(0)$  being the state evolved under the first class of errors. Here we only keep the commuting contributions from  $\mathcal{L}^{(1)}$  and  $\mathcal{L}^{(2)}$ , since their commutator only contributes at higher orders of  $\frac{1}{N}$ .

To simplify the discussion, we will look at an approximation, which we claim is sufficient to capture the behavior of information decay in our discussion. Let us first look at the case with only one type of error  $E$  in the first class,

$$\dot{\rho}(t) = \lambda E \rho E^\dagger + \lambda E^\dagger \rho E - \frac{\lambda}{2} \{E^\dagger E, \rho\} - \frac{\lambda}{2} \{E E^\dagger, \rho\}. \quad (\text{A21})$$

For low energy operators  $E$ , we can make the approximation where the coupling of  $E$  and its conjugate has the same strength  $\lambda$ , which in general should be different due to the thermal factor. In this large  $N$  model, we observe that  $E^\dagger E$  acts trivially in the leading order,

$$E^\dagger E E^p \rho = (p+1) E^p \rho + O\left(\frac{1}{N^2}\right), \\ E E^\dagger E^p \rho = p E^p \rho + O\left(\frac{1}{N^2}\right), \quad (\text{A22})$$

where  $p$  is some integer power.

The order  $\frac{1}{N^2}$  corrections can be captured by an operator  $Z$ , which acts on the code state in the following way:

$$Z|\mu_\alpha^i\rangle = \left(1 + \frac{g^{ii}}{2N^2}\right) |\mu_\alpha^i\rangle. \quad (\text{A23})$$

It is not hard to find the solution up to order  $\frac{1}{N^2}$ ,

$$\rho(t) = \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{(1 + \lambda Z t)^{k+1}} \frac{E^k \rho_0 E^{\dagger k}}{k!} + O\left(\frac{1}{N^4}\right). \quad (\text{A24})$$

One might generalize this equation to include all the first class error operators  $E_\alpha$ . We denote the density matrix dressed by the first class of errors as  $\rho^{(1)}(t)$ , which is

$$\rho^{(1)}(t) = \sum_{\{k_\alpha\}} w(\{k_\alpha\}) \prod_\alpha \frac{E_\alpha^{k_\alpha} \rho(0) E_\alpha^{\dagger k_\alpha}}{k_\alpha!} + O\left(\frac{1}{N^4}\right), \quad (\text{A25})$$

where the weight function can be upper bounded by

$$w(\{k_\alpha\}) = \prod_\alpha w(k_\alpha), \\ \text{where } w(k_\alpha) \leq \frac{(\lambda_\alpha t)^{k_\alpha}}{(1 + \lambda_\alpha Z_\alpha t)^{k_\alpha+1}}, \quad \text{for } k_\alpha \geq 1. \quad (\text{A26})$$

The bound is only saturated if we could neglect the difference between the coupling constants of  $E_\alpha$  and  $E_\alpha^\dagger$ . While said difference is negligible for low energy errors, it is not for high energy ones. In fact, the coupling constant goes down exponentially with the energy  $\epsilon_\alpha$  of the operator  $E_\alpha$ ,

$$\lambda_\alpha \propto \gamma(\epsilon_\alpha) \sim e^{-\beta \epsilon_\alpha}, \quad (\text{A27})$$

for  $\epsilon_\alpha > 0$ . However, the coupling constant  $\lambda'_\alpha$  for  $E_\alpha^\dagger$  barely depends on energy, since  $\gamma(\epsilon_\alpha) \sim 1$ , for  $\epsilon_\alpha < 0$ . We only bound the weight function for  $k_\alpha \geq 1$ , since the  $k_\alpha = 0$  component does not contribute to the mutual information decay, and it is determined by the constraint  $\sum_{k_\alpha=0}^{\infty} w(k_\alpha) = 1$  of the weight function.

Given the density matrix, we calculate the entropy difference by the equation

$$S(\rho^{(1)S}) - S(\rho^{(1)}) = \text{Tr}(\rho^{(1)}) \ln 2 - \frac{1}{2dN^4} \sum_{AB} w_{\mathcal{A}} \left( \sum_i (\Delta_{AB}^i)^2 + \sum_{i \neq j} (h_{AB}^{ij})^2 \right) + O\left(\frac{1}{N^6}\right), \quad (\text{A28})$$

where we use  $\mathcal{A}$  and  $\mathcal{B}$  as a shorthand notation for the product of first class operators, so that  $E_{\mathcal{A}} := \prod_{\alpha_i \in \mathcal{A}} E_{\alpha_i}$ .  $w_{\mathcal{A}}$  is a number obtained by setting all the  $Z_{\alpha}$ s to identity in Eq. (A26). The  $d$  function and  $h$  function scale linearly with  $k_n$ , and are proportional to some power  $q$  of  $n_{\alpha}$ , with  $n_{\alpha} = \frac{\epsilon_{\alpha}}{\Omega}$  being the energy level of  $E_{\alpha}$ , where  $\Omega$  has dimension of energy. For example,  $\Omega = \omega$  for our oscillator model. Thus we can upper bound the information decay rate by (up to  $\frac{1}{N^4}$ ),

$$K^{(1)}(t) = \frac{1}{N^4} \sum_{k_{\alpha}=1}^{\infty} w(\{k_{\alpha}\}) \left( \sum_{\alpha} n_{\alpha}^q k_{\alpha} \right)^2, \leq \frac{1}{N^4} \left[ 2 \left( \sum_{\alpha} \lambda_{\alpha} n_{\alpha}^q t \right)^2 + \sum_{\alpha} \lambda_{\alpha} n_{\alpha}^{2q} t \right]. \quad (\text{A29})$$

Note that the coefficient  $\lambda_{\alpha}$  scales as  $e^{-\beta \epsilon_{\alpha}}$ , and the density of errors with energy  $\epsilon$  is bounded by  $e^{\mu \epsilon}$  by our sparsity condition. By the same argument we have for Eq. (58) in Sec. III B, one can show that the quantity inside the square bracket does not scale as  $N$  as long as  $T < \frac{1}{\mu}$ .

The decay rate due to the first class of errors is suppressed by  $N^4$  and grows quadratically with time for  $\lambda t \gg 1$ . Similar arguments also apply for higher orders of  $\frac{1}{N^2}$ . In general, the decay rate at order  $\frac{1}{N^{2m}}$  is proportional to  $\Delta^m$  and  $h^m$  [defined in Eq. (A3)] of the corresponding error operators, which scales as  $k^m$  for the error operator  $E_{\alpha}^k$ . Therefore, the general decay rate takes the form of  $\frac{t^m}{N^{2m}}$  at each order with  $m \geq 2$ .

The second class of errors of the form  $E_{\nu}^k$  is suppressed by  $\frac{t^k}{N^{2k}}$  in the Lindblad equation, where the leading order term is linear in  $t$ ,

$$\rho^{(2)}(t) := t \mathcal{L}^{(2)}[\rho^{(1)}(t)], = \sum_{\nu} \lambda_{\nu} t \left[ E_{\nu} \rho^{(1)}(t) E_{\nu}^{\dagger} - \frac{1}{2} \{ \rho^{(1)}(t), E_{\nu}^{\dagger} E_{\nu} \} \right]. \quad (\text{A30})$$

Recall that  $\rho^{(1)}(t)$  is the density matrix induced by the first class of errors, explicitly given by Eq. (A25). For the sake of clarity, let us rewrite  $\rho^{(1)}(t)$  as

$$\rho^{(1)}(t) = \sum_{\mathcal{A}} w_{\mathcal{A}}(t) E_{\mathcal{A}} \rho_0 E_{\mathcal{A}}^{\dagger}. \quad (\text{A31})$$

Like our discussion in Appendix A 1, we convert the entropy of the state to the entropy of the matrix  $g_{\sigma_{\mathcal{A}}, \nu_{\mathcal{B}}}^{ij}$ , which can be read off from

$$\langle \tilde{j} | E_{\mathcal{B}}^{\dagger} E_{\nu}^{\dagger} E_{\sigma} E_{\mathcal{A}} | \tilde{i} \rangle = \frac{g_{\mathcal{A}\sigma, \mathcal{B}\nu}^{ij}}{N^2} + O\left(\frac{1}{N^4}\right). \quad (\text{A32})$$

Being the product of a second class error and a first class error, this composite operator belongs to the second class of errors, but it has  $O(\frac{1}{N})$  overlap with some operators in the first class,

$$\langle \tilde{j} | E_{\mathcal{B}}^{\dagger} E_{\sigma} E_{\mathcal{A}} | \tilde{i} \rangle = \frac{\delta_{ij} g_{\sigma_{\mathcal{A}}, \mathcal{B}}}{N} + O\left(\frac{1}{N^3}\right). \quad (\text{A33})$$

Due to the factorization property,  $\langle \tilde{j} | E_{\mathcal{B}}^{\dagger} E_{\nu}^{\dagger} E_{\sigma} E_{\mathcal{A}} | \tilde{i} \rangle = \langle \tilde{j} | E_{\nu}^{\dagger} E_{\sigma} | \tilde{i} \rangle \langle 0 | E_{\mathcal{B}}^{\dagger} E_{\mathcal{A}} | 0 \rangle + \langle \tilde{j} | \tilde{i} \rangle \langle 0 | E_{\mathcal{B}}^{\dagger} E_{\nu}^{\dagger} E_{\sigma} E_{\mathcal{A}} | 0 \rangle$ , the matrix element can be written as sum of two terms,

$$g_{\mathcal{A}\sigma, \mathcal{B}\nu}^{ij} = g_{\sigma\nu}^{ij} f_{\mathcal{A}\mathcal{B}} + \delta^{ij} g_{\sigma_{\mathcal{A}}, \nu_{\mathcal{B}}}, \quad (\text{A34})$$

where  $f_{\mathcal{A}\mathcal{B}} := \langle 0 | E_{\mathcal{B}}^{\dagger} E_{\mathcal{A}} | 0 \rangle$  and  $|0\rangle$  is the vacuum state of our system. The matrix  $g_{\sigma\nu}^{ij}$  is identical to the one in Eq. (A4).  $g_{\sigma_{\mathcal{A}}, \nu_{\mathcal{B}}}$  is positive definite where each element is equal to the four point function  $\langle 0 | E_{\mathcal{B}}^{\dagger} E_{\nu}^{\dagger} E_{\sigma} E_{\mathcal{A}} | 0 \rangle$ , modulo the  $\frac{1}{N^2}$  factor. The overlap between the composite operators and some first class operators [Eq. (A33)] can be removed by a basis transformation, which also cancels the second term in Eq. (A34). Therefore, similar to Eq. (A16), we conclude that the entropy difference is

$$S(\rho^{(2)S}(t)) - S(\rho^{(2)}(t)) = \frac{1}{N^2} \left[ S\left(\frac{1}{d} g_{\sigma\nu}^{ij}(t) f_{\mathcal{A}\mathcal{B}}(t)\right) - S\left(\frac{1}{d} \sum_i g_{\sigma\nu}^{ii}(t) f_{\mathcal{A}\mathcal{B}}(t)\right) \right], = \frac{\text{Tr}(f_{\mathcal{A}\mathcal{B}}(t))}{N^2} \left[ S\left(\frac{1}{d} g_{\sigma\nu}^{ij}(t)\right) - S\left(\frac{1}{d} \sum_i g_{\sigma\nu}^{ii}(t)\right) \right], = \sum_n \frac{\lambda_n t}{N^2} \left[ S\left(\frac{1}{d} g_{\sigma\nu}^{ij}\right) - S\left(\frac{1}{d} \sum_i g_{\sigma\nu}^{ii}\right) \right] \Big|_{\sigma, \nu \in \mathcal{B}_n}, \quad (\text{A35})$$

where  $S(\frac{1}{d} g_{\sigma\nu}^{ij}(t) f_{\mathcal{A}\mathcal{B}}(t))$  denotes the entropy of the matrix whose element is  $\frac{1}{d} g_{\sigma_{\mathcal{A}}, \nu_{\mathcal{B}}}^{ij}(t) = \frac{1}{d} g_{\sigma\nu}^{ij}(t) f_{\mathcal{A}\mathcal{B}}(t)$ . We have absorbed the factor  $\lambda_{\nu} t$  into the matrix  $g_{\sigma\nu}^{ij}(t) := \sqrt{\lambda_{\nu}} \lambda_{\sigma} t g_{\sigma\nu}^{ij}$ , and the weight function  $w_{\mathcal{A}}(t)$  into the function  $f_{\mathcal{A}\mathcal{B}}(t) := \sqrt{w_{\mathcal{A}}(t) w_{\mathcal{B}}(t)} f_{\mathcal{A}\mathcal{B}}$ . By definition,  $\text{Tr}(f_{\mathcal{A}\mathcal{B}}) = \sum_{\mathcal{A}} w_{\mathcal{A}}(t) \langle 0 | E_{\mathcal{A}}^{\dagger} E_{\mathcal{A}} | 0 \rangle = \text{Tr}(\rho^{(1)}(t)) + O(\frac{1}{N^2}) = 1 + O(\frac{1}{N^2})$ . In the last step, we again utilized the block diagonal form of the matrix  $g_{\sigma\nu}^{ij}$  to write the result as a sum of each block's contribution, with blocks being labeled by  $\mathcal{B}'_n$ , such that the matrix elements are nonzero only for  $\sigma$  and  $\nu$  in the same block. Note that the entropy difference is the same as in

Eq. (A16). We can bound the decay rate induced by the second class of errors with the following function at leading order of  $\frac{\lambda t}{N^2}$ ,

$$K^{(2)}(t) \leq \sum_n \frac{\lambda_n t}{N^2} f(n) r(n) \log d, \quad (\text{A36})$$

where  $f(n)$  is an upper bound for each matrix element  $g_{\sigma\nu}^{ij}$  in the block  $\mathcal{B}_n$ , and  $r(n)$  is the rank of the block.

At higher order, we can continue to expand the second class of Lindbladian operator and one can repeat these calculations following the same procedure. Errors of the form  $E_\sigma^m$  then induce information decay at  $O(\frac{1}{N^{2m}})$ , with magnitude proportional to  $(\lambda t)^m$ .

In summary, we find that both the first class of errors and the second class of errors contribute to the information decay with some rate function  $K_m(t) \propto \frac{t^m}{N^{2m}}$  as long as the assumptions of Theorem 1 are satisfied. For  $N^2 \gg \lambda t \gg 1$ ,

$$I(S:R) = 2 \ln d - \sum_{m=1}^{\infty} K_m(t). \quad (\text{A37})$$

## APPENDIX B: aKL CONDITION FOR SINGLE TRACE ERRORS

### 1. Nonlocal model

In this section, we present the aKL equations for the most general single-trace error operators. Since the logical states [Eq. (35)] are symmetric in  $a_1^\dagger$  and  $a_2^\dagger$ , we only need to consider the single-trace operators involving  $a_1$  and  $a_1^\dagger$ . In the following, we will only keep track of the terms up to  $O(\frac{1}{N^2})$ .

Let us define a class of error operators

$$E_{n,m}^P = \frac{:\text{Tr}(P\{a_1^{\dagger n} a_1^m\}):}{N^{\frac{n+m}{2}}}, \quad (\text{B1})$$

where  $P$  denotes a particular permutation of the operators inside the curly brackets. Acting it on the logical states, one obtains

$$E_{n,m}^P |\tilde{i}\rangle = f |\mu^i\rangle + \frac{g}{N^2} |\mu\rangle + O\left(\frac{1}{N^4}\right). \quad (\text{B2})$$

To extract these coefficients, we calculate the overlap

$$\langle \tilde{i} | E_{n',m'}^{P'\dagger} E_{n,m}^P | \tilde{j} \rangle, \quad (\text{B3})$$

which is essentially a four-point function that factorizes into

$$\langle \tilde{O}_i^\dagger E_{n',m'}^{P'\dagger} E_{n,m}^P \tilde{O}_j \rangle = \langle \tilde{O}_i^\dagger \tilde{O}_j \rangle \langle E_{n',m'}^{P'\dagger} E_{n,m}^P \rangle + O\left(\frac{1}{N}\right). \quad (\text{B4})$$

This is the only way to factorize into disconnected correlators, because  $O_i$  and  $O_j$  contain  $a_2^\dagger$ , while  $E_{n,m}^P$ s do not. The two point function  $\langle E_{n',m'}^{P'\dagger} E_{n,m}^P \rangle$  vanishes, unless  $m, m' = 0$ , which is the case that we discussed in the main text. Moreover, the subleading term can be  $O(\frac{1}{N})$  only if  $m' = n' = 0$ , hence the only nontrivial four-point function contribution is when  $m = n$ . In this case, we define  $Z_n^P = \frac{:\text{Tr}(P\{a_1^{\dagger n} a_1^n\}):}{N^n}$  and find that

$$\langle \tilde{j} | Z_n^P | \tilde{i} \rangle = \frac{c_n^P}{N} + O\left(\frac{1}{N^3}\right), \quad (\text{B5})$$

for some constant  $c_n^P$ , which is nonzero only when  $n = 2$ . This is consistent with the aKL condition [Eq. (18)].

So far, we have discussed the order one and  $\frac{1}{N}$  term in Eq. (B5). All the other error operators can only produce order  $\frac{1}{N^2}$  terms when acting on the logical states. This completes the proof of aKL condition.

As a special example, consider the decay mode,  $E_{0,2}$ , which lowers the energy when acting on the logical states. It also satisfies the aKL condition,

$$\begin{aligned} \langle \tilde{1} | E_{0,2}^\dagger E_{0,2} | \tilde{1} \rangle - \langle \tilde{2} | E_{0,2}^\dagger E_{0,2} | \tilde{2} \rangle &= \frac{8}{N^2} + O\left(\frac{1}{N^4}\right), \\ \langle \tilde{2} | E_{0,2}^\dagger E_{0,2} | \tilde{1} \rangle &= \frac{4\sqrt{2}}{N^3} + O\left(\frac{1}{N^5}\right). \end{aligned} \quad (\text{B6})$$

So the decay mode belongs to the second class of errors classified in Appendix A, and lead to a spontaneous decay of quantum information at zero temperature with a rate proportional to  $\frac{1}{N^2}$ .

### 2. Local model

In this section, we prove the aKL condition for the single-trace errors of the form

$$E_{\{n_i\}} = \frac{:\text{Tr}(P\{a_1^{\dagger n_1} a_2^{\dagger n_2} a_1^{n_3} a_2^{n_4}\}):}{N^{\sum_{k=1}^4 \frac{n_k}{2}}}, \quad (\text{B7})$$

with logical states defined as

$$\begin{aligned} |\tilde{\uparrow}\rangle &= \frac{\text{Tr}(a_1^\dagger)}{\sqrt{N}} |0\rangle_{12}, \\ |\tilde{\downarrow}\rangle &= \frac{\text{Tr}(a_2^\dagger)}{\sqrt{N}} |0\rangle_{12}. \end{aligned} \quad (\text{B8})$$

In the main text, we derived a general formula for the case when  $n_3 = n_4 = 0$ . Moreover, since the logical states are annihilated by more than two  $a_1$  or  $a_2$ , and the error operators are normal ordered, we are left to discuss the case when  $n_3 + n_4 = 1$ . Without loss of generality, we will only

consider the case with  $n_3 = 1$ ,  $n_4 = 0$ . Then the relevant overlaps are

$$\begin{aligned}\langle \tilde{\uparrow} | E_{(n-1,m,0,0)}^\dagger E_{(n,m,1,0)} | \tilde{\uparrow} \rangle &= \frac{n}{N^2} + O\left(\frac{1}{N^4}\right), \\ \langle \tilde{\uparrow} | E_{(n,0,1,0)}^\dagger E_{(n,0,1,0)} | \tilde{\uparrow} \rangle &= \frac{n}{N^2} + O\left(\frac{1}{N^4}\right), \\ \langle \tilde{\downarrow} | E_{(n,m,0,0)}^\dagger E_{(n,m,1,0)} | \tilde{\downarrow} \rangle &= \langle \tilde{\downarrow} | E_{(n,0,1,0)}^\dagger E_{(n,0,1,0)} | \tilde{\downarrow} \rangle = 0, \\ \langle \tilde{\downarrow} | E_{(n,m-1,0,0)}^\dagger E_{(n,m,1,0)} | \tilde{\downarrow} \rangle &= \frac{m}{N^2} + O\left(\frac{1}{N^4}\right).\end{aligned}\quad (\text{B9})$$

Exceptions of these formulas consist of special cases when  $n = 1$ ,  $m = 0$  and  $n = 0$ ,  $m = 1$ , where one can derive that

$$\begin{aligned}\langle \tilde{\uparrow} | \frac{\text{Tr}(a_1^\dagger a_1)}{N} | \tilde{\uparrow} \rangle &= \frac{1}{N}, \quad \langle \tilde{\downarrow} | \frac{\text{Tr}(a_1^\dagger a_1)}{N} | \tilde{\downarrow} \rangle = 0, \\ \langle \tilde{\downarrow} | \frac{\text{Tr}(a_2^\dagger a_1)}{N} | \tilde{\downarrow} \rangle &= \frac{1}{N}.\end{aligned}\quad (\text{B10})$$

These are  $O(\frac{1}{N})$  overlap, which are consistent with Eq. (18). So we complete the proof of aKL (18) for general single-trace errors in this local model.

### 3. Global symmetry

In this section, we prove the aKL condition for the model in Sec. IV. The logical space is spanned by the two states of unit norm

$$\begin{aligned}|1\rangle &= \frac{\text{Tr}(a^{\dagger L})^2}{N^L \xi(L, L)}, \\ |2\rangle &= \frac{\text{Tr}(a^{\dagger L+1}) \text{Tr}(a^{\dagger L-1})}{N^L \xi(L+1, L-1)},\end{aligned}\quad (\text{B11})$$

where the normalization constant are  $\xi(L, L) = \sqrt{2L}(1 + O(\frac{1}{N^2}))$  and  $\xi(L+1, L-1) = \sqrt{L^2-1}(1 + O(\frac{1}{N^2}))$ . Note that these states are not orthogonal but have overlap

$$\langle 1|2\rangle = \frac{L^2 \sqrt{L^2-1}}{\sqrt{2} N^2}.\quad (\text{B12})$$

To check the aKL condition, we insert the error operators  $E_n = \frac{\text{Tr}(a^{\dagger n})}{N^{\frac{n}{2} \xi(n)}}$  with  $\xi(n)$  being the normalization constant, and derive the following equations:

$$\begin{aligned}\langle 1|E_m^\dagger E_n|1\rangle - \langle 2|E_m^\dagger E_n|2\rangle &= \delta_{mn} \left[ \frac{2n}{N^2} + O\left(\frac{1}{N^4}\right) \right], \\ \langle 1|E_m^\dagger E_n|2\rangle &= \delta_{mn} \left[ \frac{L^2 \sqrt{L^2-1}}{\sqrt{2} N^2} + O\left(\frac{1}{N^4}\right) \right].\end{aligned}\quad (\text{B13})$$

Note that the leading term in the second equation is a consequence of the nonorthogonality between our definition of  $|1\rangle$  and  $|2\rangle$ ; it can be easily removed via a change of basis [cf. Eq. (65)]. The above equation also holds for  $E_n = \frac{\text{Tr}(a^n)}{N^{\frac{n}{2} \xi(n)}}$ .

For more general operators  $E_{(n,m)} := \frac{\text{Tr}(a^{\dagger n} a^m)}{N^{\frac{m+n}{2}}}$ , with  $n \neq m$ , one can show that

$$\begin{aligned}\langle 1|E_{n',m'}^\dagger E_{n,m}|1\rangle - \langle 2|E_{n',m'}^\dagger E_{n,m}|2\rangle &= \delta_{n'-m', n-m} \left[ \frac{2}{N^2} + O\left(\frac{1}{N^4}\right) \right], \\ \langle 1|E_{n',m'}^\dagger E_{n,m}|2\rangle &= \delta_{n'-m', n-m} \left[ \frac{L^2 \sqrt{L^2-1}}{\sqrt{2} N^2} + O\left(\frac{1}{N^4}\right) \right],\end{aligned}\quad (\text{B14})$$

which indicates the aKL condition holds.

For the special case of  $n = m$ , the operator also incurs no phase error, because

$$\langle 1| \frac{\text{Tr}(a^{\dagger n} a^n)}{N^n} |1\rangle = \langle 2| \frac{\text{Tr}(a^{\dagger n} a^n)}{N^n} |2\rangle = \frac{2L}{N}.\quad (\text{B15})$$

We thereby conclude that the aKL conditions hold for all the single trace operators.

### 4. Spin model

The approximate Knill-Laflamme condition for the spin model can be calculated using the spin algebra. Recall that we can construct two states that span the code subspace in a way analogous to our definition in the oscillator model,

$$\begin{aligned}|1\rangle &= \frac{\text{Tr}(S^{+L})^2}{N^L \eta(L, L)}, \\ |2\rangle &= \frac{\text{Tr}(S^{+L-1}) \text{Tr}(S^{+L+1})}{N^L \eta(L+1, L-1)},\end{aligned}\quad (\text{B16})$$

where  $\eta(L, L')$  is the normalization constant. They have the same leading order behavior as  $\xi(L, L')$  in the harmonic oscillator model, but have different forms in subleading terms. The overlap of these two states is

$$\langle 1|2\rangle = \frac{f(L) L \sqrt{L^2-1}}{\sqrt{2} N^2},\quad (\text{B17})$$

where  $f(L)$  is some complicated function of  $L$ . Again, we insert error operators of the form  $E_n = \frac{\text{Tr}(S^{\pm n})}{N^{\frac{n}{2}\eta(n)}}$  and  $E_{n,m} = \frac{\text{Tr}(S^{+n}S^{-m})}{N^{\frac{m+n}{2}}}$  into the correlator, and derive the aKL equations for the spin model,

$$\begin{aligned} \langle 1|E_m^\dagger E_n|1\rangle - \langle 2|E_m^\dagger E_n|2\rangle &= \delta_{mn} \left[ \frac{2n}{N^2} + O\left(\frac{1}{N^4}\right) \right], \\ \langle 1|E_{n',m'}^\dagger E_{n,m}|1\rangle - \langle 2|E_{n',m'}^\dagger E_{n,m}|2\rangle &= \delta_{n'-m',n-m} \\ &\times \left[ \frac{2}{N^2} + O\left(\frac{1}{N^4}\right) \right]. \end{aligned} \quad (\text{B18})$$

Again, the bit flip error only appears at  $O(\frac{1}{N^4})$

$$\langle 1|E_m^\dagger E_n|2\rangle = \delta_{mn} \left[ \frac{f(L)L\sqrt{L^2-1}}{\sqrt{2}N^2} + O\left(\frac{1}{N^4}\right) \right], \quad (\text{B19})$$

where we recall the first term can be removed by choosing an orthonormal basis. This completes our proof for the aKL condition.

### APPENDIX C: MASTER EQUATION IN LINDBLAD FORM

Suppose the coupling between system and thermal reservoir is

$$\begin{aligned} H &= H_S + H_B + V_I, \\ V_I &= \sum_a E_a \otimes b_a + \text{H.c.}, \end{aligned} \quad (\text{C1})$$

where  $H_S$  and  $H_B$  are the system and bath Hamiltonian, respectively, and  $b_a$ s are the bath modes.

In the interaction picture, time evolution is described by

$$\begin{aligned} \dot{\rho}(t) &= -i[V_I(t), \rho(t)], \\ &= -i[V_I(t), \rho_0] - \left[ V_I(t), \int_0^t dt' [V_I(t'), \rho(t')] \right]. \end{aligned} \quad (\text{C2})$$

The essential simplification is enabled by the Born-Markov approximation, which allows one to approximate the full density matrix as  $\rho(t) \sim \rho_S(t) \otimes \rho_B$  at all times. After tracing out the thermal state, we arrive at

$$\begin{aligned} \dot{\rho}_{IS}(t) &= \int_0^t dt' \sum_{ab} [E_b(t') \rho_{IS}(t) E_a^\dagger(t) \\ &\quad - E_a^\dagger(t) E_b(t') \rho_{IS}(t)] \text{Tr}(\rho_B b_a^\dagger(t) b_b(t')) + \text{H.c.} \end{aligned} \quad (\text{C3})$$

For sufficiently long time  $t$ , we can treat the initial time as  $-\infty$ . We also assume that the bath modes are uncorrelated. Then we define

$$\int_{-\infty}^t dt' \text{Tr}(\rho_B b_a^\dagger(t) b_b(t')) e^{i\omega(t-t')} := \frac{\gamma(\omega)}{2} \delta_{ab}. \quad (\text{C4})$$

For operators  $E_a(t)$  with the decomposition,

$$E_a(t) = \sum_k E_a^{(k)} e^{-i\epsilon_a^{(k)} t}, \quad (\text{C5})$$

Eq. (C3) simplifies as

$$\begin{aligned} \dot{\rho}_{IS}(t) &= \sum_{a,k} \frac{\gamma(\epsilon_a^{(k)})}{2} [E_a^{(k)} \rho_{IS}(t) E_a^\dagger(t) \\ &\quad - E_a^\dagger(t) E_a^{(k)} \rho_{IS}(t)] e^{-i\epsilon_a^{(k)} t} + \text{H.c.} \end{aligned} \quad (\text{C6})$$

For example, consider the error operator  $E_{ijkl} = a_j^\dagger a_l^{\dagger k}$  in Sec. IV. It can be decomposed as

$$\begin{aligned} E_{ijkl}(t) &= \widetilde{a_j^\dagger a_l^{\dagger k}} e^{-2i\mathcal{J}t} + \left[ \frac{(\widetilde{a^{\dagger 2}})_l^i}{N} \delta_j^k + \frac{(\widetilde{a^{\dagger 2}})_j^k}{N} \delta_l^i \right. \\ &\quad \left. + \frac{\text{Tr}(a^\dagger)}{N} \widetilde{a_j^\dagger} \delta_l^k + \frac{\text{Tr}(a^\dagger)}{N} \widetilde{a_l^{\dagger k}} \delta_j^i \right] e^{-i\mathcal{J}t} \\ &\quad + \frac{\text{Tr}(a^{\dagger 2})}{N^2} \delta_l^i \delta_j^k + \frac{\text{Tr}(a^\dagger)^2}{N^2} \delta_j^i \delta_l^k, \end{aligned} \quad (\text{C7})$$

where the operator under the tilde vanishes when contracting with the Kronecker delta, e.g.  $a_j^\dagger \delta_j^i = 0$ . Consequently, the tensor coefficients are almost orthogonal to each other, i.e.

$$\sum_{ijkl} E_{ijkl}^{\dagger(m)} E_{ijkl}^{(n)} \propto \delta^{mn} + O\left(\frac{1}{N}\right), \quad (\text{C8})$$

where the  $O(\frac{1}{N})$  term denotes operators that has  $O(\frac{1}{N})$  expectation value with respect to any low energy state.

More generally, a tensor product of individual creation and annihilation operators can always be decomposed into different components obtained by index contractions, which are orthogonal to each other up to  $O(\frac{1}{N})$ , in the sense of Eq. (C8). As a result, the Lindbladian operator can always be written as a diagonal sum of components that correspond to different contractions,

$$\sum_m \gamma(\epsilon^{(m)}) \sum_a E_a^{(m)} \rho E_a^{\dagger(m)} = \sum_{ma} \gamma(\epsilon^{(m)}) E_a^{(m)} \rho E_a^{\dagger(m)}. \quad (\text{C9})$$

### APPENDIX D: EXPECTATION VALUE OF THE CASIMIR

In this section, we evaluate the expectation value of the penalty Hamiltonian  $H_G$  [in Eq. (59)] for a thermal state  $\rho_T$  at temperature  $T < \frac{\mathcal{J}}{2 \log N}$ ,

$$\text{Tr}(\rho_T H_G) = \sum_{m=0}^{\infty} d_m m \mathcal{J} e^{-\frac{m(\mathcal{J}+\omega)}{T}}, \quad (\text{D1})$$

where  $m$  is the highest weight of a particular representation, with degeneracy  $d_m$  and energy  $m\mathcal{J}$ . Examples of nonsinglet states are the  $N^2$  number of  $a_j^{\dagger i}|0\rangle$ s in the adjoint representation with energy  $\mathcal{J}$ , and the  $N^4$  number of  $a_j^{\dagger i} a_l^{\dagger k}|0\rangle$ s in the tensor product representation with energy  $2\mathcal{J}$ .

In the oscillator model, we expect  $d_m$  to scale as  $d_m \sim N^{2m}$ . This factor cancels with the factor  $e^{-\frac{m\mathcal{J}}{T}}$ , since

$\mathcal{J} > 2T \log N$ . Therefore, we estimate the energy expectation value to be

$$\text{Tr}(\rho_T H_G) < \sum_m m \mathcal{J} e^{-\frac{m\omega}{T}}. \quad (\text{D2})$$

This is an  $O(1)$  number times the energy scale  $\mathcal{J}$ . So we conclude that, for a thermal state with temperature  $T < \frac{\mathcal{J}}{2 \log N}$ , the energy expectation value of  $H_G$  per oscillator is  $O(1)$  times  $\frac{\mathcal{J}}{N^2}$ .

- 
- [1] A. Y. Kitaev, *Proceedings of the 3rd International Conference of Quantum Communication and Measurement*, edited by O. Hirota, A. S. Holevo, and C. M. Caves (Plenum, New York, 1997).
- [2] A. Yu. Kitaev, Fault-tolerant quantum computation by anyons, *Ann. Phys. (Amsterdam)* **303**, 2 (2003).
- [3] Alexei Kitaev and John Preskill, Topological Entanglement Entropy, *Phys. Rev. Lett.* **96**, 110404 (2006).
- [4] R. Alicki, M. Fannes, and M. Horodecki, On thermalization in Kitaev's 2D model, *J. Phys. A* **42**, 065303 (2009).
- [5] R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki, On thermal stability of topological qubit in Kitaev's 4D model, [arXiv:0811.0033](https://arxiv.org/abs/0811.0033).
- [6] Alexei Kitaev, Anyons in an exactly solved model and beyond, *Ann. Phys. (Amsterdam)* **321**, 2 (2006).
- [7] Ahmed Almheiri, Xi Dong, and Daniel Harlow, Bulk locality and quantum error correction in AdS/CFT, *J. High Energy Phys.* **04** (2015) 163.
- [8] S. W. Hawking and Don N. Page, Thermodynamics of black holes in anti-de Sitter space, *Commun. Math. Phys.* **87**, 577 (1982).
- [9] Edward Witten, Anti-de Sitter space, thermal phase transition, and confinement in gauge theories, *Adv. Theor. Math. Phys.* **2**, 505 (1998).
- [10] Ning Bao, ChunJun Cao, and Guanyu Zhu, Deconfinement and error thresholds in holography, *Phys. Rev. D* **106**, 046009 (2022).
- [11] Fernando Pastawski, Beni Yoshida, Daniel Harlow, and John Preskill, Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence, *J. High Energy Phys.* **06** (2015) 149.
- [12] Gong Cheng, ChunJun Cao, and Brian Swingle, companion paper (to be published).
- [13] Emanuel Knill, Raymond Laflamme, and Lorenza Viola, Theory of Quantum Error Correction for General Noise, *Phys. Rev. Lett.* **84**, 2525 (2000).
- [14] Cédric Bény and Ognian Oreshkov, General Conditions for Approximate Quantum Error Correction and Near-Optimal Recovery Channels, *Phys. Rev. Lett.* **104**, 120501 (2010).
- [15] Eric Mintun, Joseph Polchinski, and Vladimir Rosenhaus, Bulk-Boundary Duality, Gauge Invariance, and Quantum Error Corrections, *Phys. Rev. Lett.* **115**, 151601 (2015).
- [16] Alexey Milekhin, Quantum error correction and large  $N$ , *SciPost Phys.* **11**, 094 (2021).
- [17] Eric Dennis, Alexei Kitaev, Andrew Landahl, and John Preskill, Topological quantum memory, *J. Math. Phys. (N.Y.)* **43**, 4452 (2002).
- [18] Dave Bacon, Operator quantum error-correcting subsystems for self-correcting quantum memories, *Phys. Rev. A* **73**, 012340 (2006).
- [19] Benjamin J. Brown, Daniel Loss, Jiannis K. Pachos, Chris N. Self, and James R. Wootton, Quantum memories at finite temperature, *Rev. Mod. Phys.* **88**, 045005 (2016).
- [20] Olivier Landon-Cardinal, Beni Yoshida, David Poulin, and John Preskill, Perturbative instability of quantum memory based on effective long-range interactions, *Phys. Rev. A* **91**, 032303 (2015).
- [21] Courtney G. Brell, A proposal for self-correcting stabilizer quantum memories in 3 dimensions (or slightly less), *New J. Phys.* **18**, 013050 (2016).
- [22] Keiichiro Furuya, Nima Lashkari, and Mudassir Moosa, Renormalization group and approximate error correction, *Phys. Rev. D* **106**, 105007 (2022).
- [23] Robert J. Harris, Nathan A. McMahon, Gavin K. Brennen, and Thomas M. Stace, Calderbank-Shor-Steane holographic quantum error-correcting codes, *Phys. Rev. A* **98**, 052301 (2018).
- [24] ChunJun Cao and Brad Lackey, Approximate Bacon-Shor code and holography, *J. High Energy Phys.* **05** (2021) 127.
- [25] Kfir Dolev, Vladimir Calvera, Samuel S. Cree, and Dominic J. Williamson, Gauging the bulk: Generalized gauging maps and holographic codes, *J. High Energy Phys.* **05** (2022) 158.
- [26] Ofer Aharony, Joseph Marsano, Shiraz Minwalla, Kyriakos Papadodimas, and Mark Van Raamsdonk, The hagedorn - deconfinement phase transition in weakly coupled large  $N$  gauge theories, *Adv. Theor. Math. Phys.* **8**, 603 (2004).
- [27] Nissan Itzhaki and John McGreevy, The large  $N$  harmonic oscillator as a string theory, *Phys. Rev. D* **71**, 025003 (2005).
- [28] David Berenstein, A toy model for the AdS/CFT correspondence, *J. High Energy Phys.* **07** (2004) 018.

- 
- [29] Igor R. Klebanov, String theory in two-dimensions, in *Spring School on String Theory and Quantum Gravity (to be followed by Workshop)* (1991), [arXiv:hep-th/9108019](https://arxiv.org/abs/hep-th/9108019).
- [30] Clifford V. Johnson, Consistency Conditions for Non-Perturbative Completions of JT Gravity, [arXiv:2112.00766](https://arxiv.org/abs/2112.00766).
- [31] Stefano Chesi, Beat Röthlisberger, and Daniel Loss, Self-correcting quantum memory in a thermal environment, *Phys. Rev. A* **82**, 022305 (2010).
- [32] Daniel Gottesman, Alexei Kitaev, and John Preskill, Encoding a qubit in an oscillator, *Phys. Rev. A* **64**, 012310 (2001).