

# Limitations of Passive Protection of Quantum Information

Fernando Pastawski,<sup>1</sup> Alastair Kay,<sup>2,1</sup> Norbert Schuch,<sup>1</sup> and Ignacio Cirac<sup>1</sup>

<sup>1</sup>*Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Str. 1, D-85748 Garching, Germany.*

<sup>2</sup>*Centre for Quantum Computation, DAMTP, Centre for Mathematical Sciences,  
University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK*

(Dated: November 19, 2009)

The ability to protect quantum information from the effect of noise is one of the major goals of quantum information processing. In this article, we study limitations on the asymptotic stability of quantum information stored in passive  $N$ -qubit systems. We consider the effect of small imperfections in the implementation of the protecting Hamiltonian in the form of perturbations or weak coupling to a ground state environment. We prove that, regardless of the protecting Hamiltonian, there exists a perturbed evolution that necessitates a final error correcting step when the state of the memory is read. Such an error correction step is shown to require a finite error threshold, the lack thereof being exemplified by the 3D compass model. We go on to present explicit weak Hamiltonian perturbations which destroy the logical information stored in the 2D toric code in a time  $O(\log(N))$ .

PACS numbers: 03.67.Lx, 03.67.Pp

Contents		B. The toric code	19
I. Introduction	1	C. Full Depolarization of the Toric Code's Protected Subspace	20
A. Noise model motivation	2	References	22
B. Outline of results	3		
II. Necessity of error correction on read-out	4		
A. Eigenstate susceptibility to perturbations	4		
B. State evolution in coupled Hamiltonians	5		
C. Discussion	6		
III. Limitations of the 3D compass model	7		
IV. Limitations of the 2D toric code	8		
A. Probabilistic introduction of distant anyons	8		
B. Simple error loops in $O(N)$ time	10		
C. Localization in 2D stabilizer codes	10		
D. Logical errors in $O(\log(N))$ time	11		
Anyon matching that minimizes $L_\infty$	11		
Anyon matching that minimizes $L_1$	12		
E. Discussion	12		
V. Limitations of the 2D Ising model	12		
A. Hamiltonian perturbation proposal	13		
B. Discussion	13		
VI. Aggressive noise models	14		
A. Time-varying Perturbations	14		
1. Adiabatically varying perturbations	14		
2. Rapidly oscillating perturbations	15		
B. Stabilizer Hamiltonians and energetic environment	15		
C. Non-stabilizer Hamiltonians	16		
VII. Further applications	16		
VIII. Conclusions	17		
A. State evolution in perturbed gapped Hamiltonians	18		

## I. INTRODUCTION

Quantum information processing promises exciting new capabilities for a host of computational [1, 2, 3] and cryptographic [4, 5] tasks, if only we can fabricate devices that take advantage of the subtle and very fragile effects of quantum mechanics. The theory of quantum error-correcting codes (QECCs) and fault-tolerance [6, 7, 8, 9] assure that this fragility can be overcome at a logical level once an error rate per element below a certain threshold is achieved. However, providing a scalable physical implementation of computational elements with the required degree of precision and control has proven to be a task of extreme difficulty. Thus, one might hope to design superior fault-tolerant components whose robustness is enforced in a more natural way at a physical level.

A first step in this daunting task is to concentrate not on universal quantum computation, but on one sub-protocol within this; the storage of quantum information. Thus, the aim is to find systems naturally assuring the stability of quantum information, just like magnetic domains in a hard disk provide stable storage of classical information. The quest for such a passive quantum memory was pioneered by Kitaev [10], who introduced the toric code as the first many body *protecting Hamiltonian*. The promising conjunction of properties shown by his proposal has fueled a search, which is yet to provide a definitive result.

For families of protecting Hamiltonians, such as Kitaev's toric code [10, 11], a constant energy gap  $\gamma$  separates the degenerate ground space, used for encoding,

from low energy excited states. Furthermore, the stabilizer representation of these Hamiltonians naturally associates it with a QECC, which permits an error threshold without the use of concatenation [11]. A perturbation theoretic expansion of local errors  $V$  in the Hamiltonian must then cancel to orders increasing with the distance of the associated QECC. Thus, within the region of validity of degenerate perturbation theory, the degeneracy of the ground state space is only split by an amount that is exponentially small in the distance of the code. In turn, it takes an exponentially long time for the splitting to implement logical rotations on the perturbed ground space (e.g. a phase gate).

However, such perturbation theoretic results must be applied with caution. The most important limitation probably arises from the fact that they deal with a closed quantum system whereas actual noise may be better modeled by perturbative coupling to an environment. Furthermore, the rigorous range of validity of perturbation theory [12],  $\|V\| < \gamma/2$ , is extremely restrictive when considering perturbations with an extensive operator norm. Even if local observables can be adapted for to a high degree of accuracy, as shown by Hastings and Wen [13], the global eigenstates of the system may change and become very different. Within our understanding, the possibility of adapting encoding and decoding protocols relies on the perturbation being characterized<sup>1</sup>. However, most of our work will focus on the relevant scenario of an unknown perturbation, where unperturbed (unadapted) encoding and decoding protocols are used.

Recently, Alicki *et al.* have presented results supporting the instability of quantum memories based on Kitaev's 2D toric code [15] and the stability of its 4D version [16] when coupled to a sufficiently cold thermal environment. Chesi *et al.* [17] have made progress in providing a general expression giving a lower bound for the lifetime of encoded information. The approach taken in these articles is thermodynamic in nature and has the advantage of allowing the derivation of positive results. A weak coupling Markovian approximation to an environment at thermal equilibrium is assumed, thus neglecting any memory effects from the environment. In a previous article [18], we considered a Hamiltonian system subject to independent depolarizing noise (corresponding to the high temperature limit of the above approach) and proved that  $O(\log N)$  is the optimal survival time for a logical qubit stored inside  $N$  physical qubits.

Our current approach directly deals with Hamiltonian perturbations and environment couplings without going through a Markovian approximation for the environment. A comparative advantage of our approach is the capability of exactly dealing with certain weak but finite perturbations and couplings, and providing restricted no go

results.

We consider the effect of relatively weak yet unknown perturbations of an  $N$  qubit local protecting Hamiltonian and coupling to an ancillary environment starting out in its ground state. We show that as the number  $N$  of physical subsystems used grows, it is impossible to immunize a quantum subspace against such noise by means of local protecting Hamiltonians only. We further show that if one wishes to recover the quantum state by means of an error correction procedure, the QECC used must have some finite error threshold in order to guarantee a high fidelity; this result is applied to the 3D compass model which is shown not to have such a threshold. In the case of the 2D toric code, we propose Hamiltonian perturbations capable of destroying encoded information after a time proportional to  $\log(N)$ , suggesting that some form of macroscopic energy barrier may be necessary. Weak finite range Hamiltonian perturbations are then presented which destroy classical information encoded into the 2D Ising model; in this case interactions involving a large, yet  $N$  independent, number of qubits are required. Finally, we consider time dependent Hamiltonian perturbations and coupling to an ancillary environment with a high energy density; here we provide constructions illustrating how these more powerful models may easily introduce logical errors in constant time into information protected by all stabilizer Hamiltonians, and certain generalizations.

## A. Noise model motivation

A prerequisite to assess protecting Hamiltonians is a precise definition of the noise model they will be expected to counter. Our aim is to understand the protection lifetime they provide to (quantum) information as well as to identify the properties a good protecting Hamiltonian should have. In order to be able to make such predictions, we will study noise models admitting a mathematically tractable description while striving to keep our choices physically motivated.

To falsify claims of protection against *any* possible noise of a certain class (such as weak local perturbations to the Hamiltonian), it suffices to consider an adversarial choice within such a class. In such a noise model, different perturbations and environments are not assigned probabilities; a perturbation is simply considered possible if it adheres to certain conditions. Finally, we assume that the perturbation has not been characterized. This allows us to derive no-go, or limitation, results from the exact analysis of adversarially engineered noise instances.

The most elementary way in which the Hamiltonian evolution of a closed system can be altered is by including a small perturbation  $V$  to the Hamiltonian  $H$ . A simple physical interpretation for such a perturbation is to associate  $V$  to imperfections in the implementation of the ideal protecting Hamiltonian  $H$ . Furthermore, Hamiltonian perturbations extending beyond the system under

<sup>1</sup> A possible exception to this is given by proposals of adiabatic state preparation [14].

experimental control are modeled by a weak coupling between the system and an environment. We focus on families of protecting Hamiltonians satisfying certain locality and boundedness conditions, and naturally extend similar restrictions on the perturbations and couplings considered.

Let us first introduce some definitions. A family of protecting Hamiltonians  $\{H_N\}$  is parametrized by a number  $N$  which grows with the number of physical subsystems participating in  $H_N$ . A Hamiltonian  $H$  is called “ $k$ -local” when it can be represented as a sum

$$H = \sum_{i=1}^N T_i, \quad (1)$$

with at most  $k$  physical subsystems participating in each interaction term  $T_i$ . The interaction strength of a physical subsystem  $s$  in a  $k$ -local Hamiltonian  $H$  is given by the sum  $\sum_i \|T_i\|$  of operator norms over those interaction terms  $T_i$  in which the physical subsystem  $s$  participates. A family of  $k$ -local Hamiltonians is called “ $J$ -bounded” if, for every Hamiltonian  $H_N$  in the family, the largest interaction strength among the physical subsystems involved is no greater than  $J$ . Finally, a family of Hamiltonians will be  $D$ -dimensional if the physical subsystems involved can be arranged into a  $D$ -dimensional square lattice, such that all interaction terms are kept geometrically local.

We will concentrate on families of  $k$ -local,  $J$ -bounded protecting Hamiltonians, with  $J > 0$ , and  $k, J \sim O(1)$ . Furthermore, the specific Hamiltonians treated in this article admit an embedding into 2, 3 or 4 spatial dimensions and we may assume such embeddings also when dealing with generic protecting Hamiltonians.

The families of Hamiltonian perturbations  $\{V_N\}$  which we will consider will be  $\tilde{J}$ -bounded, with the strength  $\tilde{J}$  small in comparison to  $J$ . The perturbations will be taken to be  $\tilde{k}$ -local, with  $\tilde{k}$  possibly different, and even larger, than  $k$ . This allows, for example, taking into consideration undesired higher order terms which may arise from perturbation theory gadgets [19]. Allowed perturbations should also admit a geometrically local interpretation under the same arrangement of subsystems as the protecting Hamiltonian.

When considering coupling to an environment, an additional set of physical subsystems will be included as the environment state. A family of local environment Hamiltonians  $\{H_N^{(E)}\}$  will be defined on these additional subsystems. The coupling between system and environment will be given by a family of weak local Hamiltonian perturbations  $V_N^{(SE)}$ , acting on both system and environment.

$$\tilde{H}_N = H_N^{(S)} \otimes I_N^{(E)} + I_N^{(S)} \otimes H_N^{(E)} + V_N^{(SE)} \quad (2)$$

Finally, it should be possible to incorporate the additional physical subsystems from the environment while preserving the number of spatial dimensions required for the Hamiltonian. To simplify notation, the sub-index  $N$  shall in general be dropped.

The engineering of  $k$ -body interactions is increasingly difficult as  $k$  grows [19, 20]. This is why we limit our study to families of  $k$ -local Hamiltonians (i.e.  $k$  independent of  $N$ ). It is under such criteria that we exclude proposals such as quantum concatenated-code Hamiltonians [21], for which the required degree of interactions would grow algebraically with the number of qubits.

The  $J$ -bounded condition guarantees that the rate of change for local observables remain bounded. Furthermore, this condition is strictly weaker than the rigorous requirements stated for the simulation through the use of perturbation theory gadgets [19]. There, constant bounds are imposed both on the norm of each interaction as well as on the number of interactions in which each subsystem participates. The  $J$ -bounded condition also leaves out systems with long range interactions, as, for those systems, the total interaction strength of individual physical subsystems diverges as the system size grows. Such long range interacting systems are physically relevant, and may lead to protecting Hamiltonian proposals [22, 23]. However, we abstain from treating such models for which our notion of weak perturbation seems inappropriate.

Each physical subsystem may independently be subject to control imprecision. Such is the case for weak unaccounted “magnetic field” acting on every component of the system or a weak coupling of each component to an independent environment. Thus, relevant physical scenarios involve perturbations with extensive operator norm (i.e. scaling with the number of subsystems). The  $\tilde{J}$ -bounded condition encapsulates these scenarios and seems to better describe what we understand by a weak perturbation.

Finally, it is expected that scalable physical implementations should be mapped to at most three spatial dimensions. This would rule out the 4D toric code Hamiltonian [11], a proposal which was otherwise shown to provide increasing protection against weak local coupling to a sufficiently cold thermal bath [16]. As would occur with an actual physical embedding, we expect that the perturbations considered may be included into the same geometrical picture as the protecting Hamiltonian they affect.

## B. Outline of results

In the following sections, we analyze the problem of obtaining increased protection for quantum information by means of an encoding and a protecting Hamiltonian acting on an increasing number of subsystems. We consider the effect of adversarial noise models consisting of local Hamiltonian perturbations and/or a weakly coupled environment. The aim is to examine the assumptions and limitations of memory schemes based on Hamiltonian protection with a growing number of physical subsystems as quantified by the survival time of stored information.

We prove in complete generality that if the read-out process does not incorporate a recovery procedure then

information retrieval will be unreliable. The figure of merit considered here is  $S(t) = \text{tr}(|\psi(0)\rangle\langle\psi(0)|\rho(t))$ , the overlap between initial and evolved state after a constant time  $t$ . For arbitrary protecting Hamiltonians we provide a completely general construction involving a weakly coupled environment starting in its ground state (Sec. IIB) which yields an exponentially small (in  $N$ ) upper bound on  $S(t)$  after a constant time. For gapped Hamiltonians, a proof proceeding without reference to an environment (Appendix A) can provide an upper bound to the time averaged overlap which is close to  $\frac{1}{2}$ .

From this point on, protecting Hamiltonians are considered together with a recovery operation  $\mathcal{R}$ , applied on read-out, thus providing a more robust figure of merit  $S_{\mathcal{R}}(t) = \text{tr}(\rho(0)\mathcal{R}(\rho(t)))$ . A similar weak coupling construction shows that information content of the 3D compass model [24] can be destroyed in constant time by a zero temperature environment (Sec. III), despite of its error correcting mechanism  $\mathcal{R}$ . This provides a direct, negative, answer to the open question of whether the 3D compass model is a self-correcting quantum memory. From a broader perspective, the structure of our proof strongly suggests that the underlying QECC defining the recovery operation  $\mathcal{R}$  must have a strictly positive error threshold.

We continue by considering the effect of Hamiltonian perturbations on the 2D toric code [10]. The recovery mechanism  $\mathcal{R}$  is then taken as the composition of a fixed syndrome measurement followed by a correction operation pairing the detected anyons. It is shown (Sec. IV) that, although the underlying QECC has an error threshold, it is not protected against combinations of *unknown* weak local Hamiltonian perturbations, even after a final round of error correction is considered. Our claim is based on adversarial weak local perturbations that are capable of destroying the stored information in a time logarithmic in  $N$ . This is stronger than previous results [25] in that, the noise model requires no interaction with the environment and the information is destroyed exponentially faster.

In a similar manner, we consider perturbations on the 2D Ising Hamiltonian (Sec. V), which is often used as an example of self-correcting classical memory. Here, Hamiltonian perturbations may transform (classical) code states into an ambiguous state in constant time. While the number  $\tilde{k}$  of bodies in perturbation terms is required to grow as the overall perturbation strength decreases, it shows no dependence on the size  $N$  of the system. In this model, any sequence of local errors connecting the two classical code states must go through states with a macroscopic amount of extra energy, showing that this property alone is not sufficient to give protection.

Beyond the Hamiltonian perturbation model and coupling to a ground state environment, we consider more aggressive noise models (Sec. VI) in which the environment can introduce large amounts of energy. The models considered are time dependent Hamiltonian perturbations and weak Hamiltonian coupling to an environment

starting in a high energy state. For such noise models, even the information storage capabilities of the 4D toric code, a proposal shown to be thermodynamically stable, are completely destroyed.

One might expect that the results presented here are not limited to the task of designing a quantum memory. Rather, they tell us about the difficulty of keeping a state and its time evolution confined within a specific subspace of the system, under the effect of Hamiltonian noise. Such considerations arise in other settings, such as in the models of adiabatic and topological quantum computation. We will outline some of these connections in Sec. VII.

## II. NECESSITY OF ERROR CORRECTION ON READ-OUT

We start our examination of passive quantum memories by proving that a final step of error correction is necessary, regardless of the choice of encoding and protecting Hamiltonian. What we mean by this, is that if such a recovery step is not included on read-out, there is no way of guaranteeing a high fidelity (close to 1) between initial and final states for more than a constant amount of time.

First, we propose weak local perturbations showing an exponentially decreasing overlap between perturbed and unperturbed eigenstates. For general local Hamiltonians and states, we consider a local coupling  $V$  of the system with a  $\gamma$ -bounded environment initialized in its ground state. Averaging over such couplings  $V$ , we are able to derive an exponentially small upper bound  $\langle S(t_f) \rangle_V \leq [1 - \sin^2(2\varepsilon)]^N$  for the overlap between initial and evolved system states at a time  $t_f = \frac{\pi}{\gamma}$ .

### A. Eigenstate susceptibility to perturbations

Consider a  $k$ -local Hamiltonian  $H$ , decomposable into  $k$ -body interaction terms  $T_i$  as described in (1). We choose a perturbation  $V$  such that the initial and final Hamiltonians are related by a composition of local unitary transformations,

$$\tilde{H} = H + V = \mathcal{U}H\mathcal{U}^\dagger, \text{ with } \mathcal{U} = \bigotimes_{l=1}^N e^{i\varepsilon P_l}, \quad (3)$$

where  $P_l$  are normalized local Hermitian operators. Taking this definition,  $V$  can be written as

$$V = \sum_{i=1}^N \mathcal{U}T_i\mathcal{U}^\dagger - T_i, \quad (4)$$

and thus is also  $k$ -local. Furthermore, if  $2k\varepsilon \ll 1$ , it is justified to call  $V$  a perturbation with respect to  $H$ , since all terms are small with respect to those of  $H$ .

Degeneracies in  $H$  are assumed to be infinitesimally lifted to ensure uniquely defined eigenvectors. The overlap between eigenvectors  $|\psi_i\rangle$  of  $H$  and the perturbed eigenvectors  $\mathcal{U}|\psi_i\rangle$  is then given by  $F_{\mathcal{U}} = |\langle\psi_i|\mathcal{U}|\psi_i\rangle|^2$ .

By averaging over all possible directions  $P_i$ , we effectively obtain an independent qubit depolarization.

$$\int \mathcal{U}^\dagger |\psi_i\rangle \langle \psi_i| \mathcal{U} dP_1 \dots dP_N = \Delta_{\lambda(\varepsilon)}^{\otimes N} (|\psi_i\rangle \langle \psi_i|) \quad (5)$$

Here,  $\Delta_\lambda(\rho) = \lambda\rho + (1-\lambda)\frac{I}{2}$  is the qubit depolarizing channel and  $\lambda(\varepsilon) = 1 - \frac{3}{2}\sin^2(\varepsilon)$ . We may then denote  $\langle F \rangle_{\mathcal{U}}$  as average of the overlap  $F_{\mathcal{U}}$  over all local rotations  $\mathcal{U}$  having a given strength  $\varepsilon$ . This average is expressed in terms of the depolarizing channel as

$$\langle F \rangle_{\mathcal{U}} = \langle \psi_i | \Delta_{\lambda(\varepsilon)}^{\otimes N} (|\psi_i\rangle \langle \psi_i|) | \psi_i \rangle. \quad (6)$$

A result of King [26], known as multiplicativity of the maximum output  $p$ -norm for the depolarizing channels, states that

$$\max_{|\phi\rangle} \|\Delta_{\lambda}^{\otimes N} (|\phi\rangle \langle \phi|)\|_p \leq \left( \max_{|\phi\rangle} \|\Delta_{\lambda} (|\phi\rangle \langle \phi|)\|_p \right)^N. \quad (7)$$

For qubit subsystems and for  $p = \infty$ , Eq. (7) bounds the overlap of  $\Delta_{\lambda}^{\otimes N} (|\phi\rangle \langle \phi|)$  with any single pure state, leading to

$$\langle F \rangle_{\mathcal{U}} \leq \left( \frac{1+\lambda}{2} \right)^N = \left( 1 - \frac{3}{4} \sin^2(\varepsilon) \right)^N. \quad (8)$$

Not only does this imply the existence of specific rotations such that  $F$  becomes exponentially small as the number of subsystems  $N$  grows, but that this is true for most rotations  $\mathcal{U}$ . While this is already known under the name of Anderson's orthogonality catastrophe (see, for example [27, 28]), we re-derive it for completeness and as an opportunity to introduce techniques needed throughout the rest of the paper.

## B. State evolution in coupled Hamiltonians

In this section, we consider a weak Hamiltonian perturbation coupling the system to a "cold" environment. The environment is assumed to start in its ground state, corresponding to a cold environment assumption. Averaging over a specific family of such perturbations instances  $V$ , an exponentially small bound on the overlap between the initial state and the evolved state is obtained. This bound,  $\langle S(t_f) \rangle_V \leq [1 - \sin^2(2\varepsilon)/3]^N$ , is obtained after a constant evolution time  $t_f = \frac{\pi}{\gamma}$ , inversely proportional to the strength of the environment Hamiltonian.

Suppose that we start with a state  $|\psi_0\rangle$  "protected" by an  $N$  qubit system Hamiltonian  $H_S$ . We can introduce a simple environment, composed of  $2N$  qubits, each of

which starts in its ground state,  $|0\rangle$ , and which is defined by its Hamiltonian

$$H = H^{(S)} \otimes I^{(E)} + I^{(S)} \otimes H^{(E)} \quad (9)$$

$$H^{(E)} = \gamma \sum_{i=1}^N |1+\rangle \langle 1+|_i^{(E)} - |00\rangle \langle 00|_i^{(E)}. \quad (10)$$

When necessary, we take the supraindices  $(S)$ ,  $(E1)$  and  $(E2)$  to denote the system, the first, and second components of the environment respectively. While both environment components will interact with the system, it is the presence of both which will allow a simple interpretation of the induced decoherence as a probabilistic application of local errors.

We again use the trick of considering a perturbed Hamiltonian  $\tilde{H} = \mathcal{U}H\mathcal{U}^\dagger$  which results from the weak local rotations  $\mathcal{U} = \bigotimes_{j=1}^N U_j$  of the decoupled Hamiltonian  $H$ . The rotation elements will involve both system and environment components,  $U_j = e^{i\varepsilon P_j^{(S)} \otimes X_j^{(E1)}}$ , where the operators  $P_j^{(S)}$  are taken to be Pauli-like operators on site  $j$  of the system.

The perturbation  $V = \mathcal{U}H\mathcal{U}^\dagger - H$  must be decomposable into small local terms. Such a decomposition for  $V$  is given in terms of the decomposition  $H^{(S)} = \sum_i T_i$  into at most  $k$ -body terms. Each perturbation term  $V_i = \mathcal{U}T_i\mathcal{U}^\dagger - T_i$  has an operator norm no greater than  $2\varepsilon k \|T_i\|$  and involves up to  $2k$ -body interactions<sup>2</sup>. The perturbation required to rotate the environment Hamiltonian terms involve at most 3-body terms and a total norm bounded by  $2\varepsilon\gamma$ .

The initial state,  $|\psi_0\rangle|00\rangle^{\otimes N}$  will thus evolve into  $e^{-it\mathcal{U}H\mathcal{U}^\dagger}|\psi_0\rangle|00\rangle^{\otimes N}$ . The survival probability is then  $S(t) = \langle \psi_0 | \rho_S(t) | \psi_0 \rangle$ , where  $\rho_S(t) = \text{tr}_E(\rho(t))$ , and

$$\rho(t) = \mathcal{U}e^{-itH}\mathcal{U}^\dagger |\psi_0\rangle \langle \psi_0| \otimes |00\rangle \langle 00|^{\otimes N} \mathcal{U}e^{itH}\mathcal{U}^\dagger. \quad (11)$$

Here,  $\mathcal{U}$  may be explicitly decomposed as

$$\begin{aligned} \mathcal{U} &= \exp(i\varepsilon \sum_j P_j^{(S)} \otimes X_j^{(E1)}) \\ &= \sum_{\mathbf{p}} \cos(\varepsilon)^{N-w(\mathbf{p})} (i \sin(\varepsilon))^{w(\mathbf{p})} P_{\mathbf{p}}^{(S)} \otimes X_{\mathbf{p}}^{(E1)}, \end{aligned} \quad (12)$$

where  $\mathbf{p}$  denotes a binary vector indicating the sites on which rotations are applied in  $P_{\mathbf{p}}^{(S)}$  and  $w(\mathbf{p})$  is the weight of the bit string  $\mathbf{p}$  (number of non identity factors in  $P_{\mathbf{p}}^{(S)}$ ).

Now consider a time  $t_f = \frac{\pi}{\gamma}$  such that  $e^{-it_f H}$  transforms components of the environment from  $|10\rangle$  to  $|11\rangle$ , while leaving components in state  $|00\rangle$  unaltered. At such

<sup>2</sup> For  $\varepsilon k \ll 1$ , a further decomposition of such terms can be provided in which subterms involving  $k+b$  bodies are of strength  $O(\varepsilon^b)$ , guaranteeing that the strength of terms decays exponentially with the number of bodies involved.

a time  $t_f$ , substituting  $\mathcal{U}$  into expressions (11) allows explicitly tracing over the environment to yield

$$\rho_S(t_f) = \sum_{\mathbf{p}, \mathbf{q}} \cos^2(\varepsilon)^{2N-w(\mathbf{p})-w(\mathbf{q})} \sin^2(\varepsilon)^{w(\mathbf{p})+w(\mathbf{q})} \times \\ P_{\mathbf{p}} e^{-it_f H^{(S)}} P_{\mathbf{q}} |\psi_0\rangle\langle\psi_0| P_{\mathbf{q}} e^{it_f H^{(S)}} P_{\mathbf{p}}. \quad (13)$$

Thus,  $\rho_S(t_f)$  may be considered as the density matrix resulting from the independent probabilistic application of the local unitary rotations prescribed by  $\mathcal{U}$  on  $|\psi_0\rangle\langle\psi_0|$ , followed by the evolution under the unperturbed system Hamiltonian, followed by a second round of random application of the local rotations prescribed by  $\mathcal{U}$ . Defining

$$\mathcal{E}_{R,p}(\rho) = pR\rho R + (1-p)\rho, \quad (14)$$

and the Hamiltonian evolution

$$\mathcal{H}_t(\rho) = e^{-iH^{(S)}t} \rho e^{iH^{(S)}t}, \quad (15)$$

we can take  $p = \sin^2(\varepsilon)$  and define  $\rho_{virt} := \left(\bigotimes_{i=1}^N \mathcal{E}_{P_i,p}\right) |\psi_0\rangle\langle\psi_0|$ , so that we can express  $S(t_f)$  as

$$S(t_f) = \text{tr}(\rho_{virt} e^{-it_f H_S} \rho_{virt} e^{it_f H_S}). \quad (16)$$

This is the overlap between a density matrix and its own unitary evolution, and can be upper bounded by

$$S(t_f) \leq \text{tr}(\rho_{virt}^2). \quad (17)$$

In turn, using the fact that  $P_i$  are Pauli-like operators we may rewrite it as

$$S(t_f) \leq \langle\psi_0| \left(\bigotimes_{i=1}^N \mathcal{E}_{P_i, 2p(1-p)}\right) (|\psi_0\rangle\langle\psi_0|) |\psi_0\rangle. \quad (18)$$

Averaging over the Pauli-like operators, we obtain

$$\langle S(t_f) \rangle_V \leq \langle\psi_0| \Delta_{\lambda(\varepsilon)}^{\otimes N} (|\psi_0\rangle\langle\psi_0|) |\psi_0\rangle, \quad (19)$$

which is the overlap at time  $t_f$  averaged over the proposed family of weak perturbative couplings. Here  $\Delta_{\lambda}(\rho)$  is again the depolarizing channel, and  $\lambda(\varepsilon) = 1 - \frac{4}{3}2p(1-p)$ . Using Eq. (8), we obtain

$$\langle S(t_f) \rangle_V \leq [1 - \frac{4}{3}p(1-p)]^N, \quad (20)$$

which by substituting  $p$  for  $\sin^2(\varepsilon)$  yields

$$\langle S(t_f) \rangle_V \leq [1 - \sin^2(2\varepsilon)/3]^N. \quad (21)$$

By averaging over different possible weak couplings, we obtain an overlap between initial and evolved states which is exponentially decreasing in  $N$ .

The norm  $\gamma$  of Hamiltonian terms in the environment should be bounded, since it is in part these terms which are rotated by  $\mathcal{U}$  to introduce a weak coupling between system and environment. Thus, the proposed evolution

time  $t_f = \frac{\pi}{\gamma}$  is constant. Furthermore, if one considers an environment of  $N$  semi-infinite chains of coupled two level systems (such as Heisenberg chains), it is possible to ensure that the overlap with the initial state is small for all times larger than  $t \sim \frac{\pi}{\gamma}$ , rather than have the recurrences that arise from the discrete spectra of the described model.

In appendix (A), the evolution of an unperturbed eigenstate is considered under the effect of pure Hamiltonian perturbations (no environment). In this case, a constant rate of change in the system state is guaranteed by an energy gap  $\gamma$  in the system Hamiltonian. If an initial state belongs to an energy band separated from the rest of the Hilbert space by such an energy gap  $\gamma$ , we provide an upper bound on the time averaged overlap  $\langle S(t') \rangle_{t' \in [0,t]} \leq \frac{1}{2} + \frac{1}{2\gamma t}$  between initial and evolved state.

### C. Discussion

We were able to show after a constant time  $t_f$ , an exponentially large degradation of the overlap  $S(t_f)$  in terms of  $N$ . If we are to find a benchmark by which to evaluate a memory scheme, we expect that the memory improves as more resources are dedicated to its implementation. It is now clear that many-body quantum states are inherently unstable with respect to the uncorrected overlap, i.e.  $S(t)$  is not the appropriate benchmark.

The fact that we may count on  $N$  physical subsystems to implement a quantum memory should not exclude using only one of them and ignoring whatever noisy evolution is affecting the others. This corresponds to considering an overlap reduced to the relevant subsystem and not on the whole state. Already such a simple idea guarantees that information storage quality is non-decreasing with  $N$ .

One may further generalize this by realizing that the relevant subsystem need not correspond to an actual physical subsystem. This corresponds to preparing code states of a QECC and performing a single round of error correction  $\mathcal{R}$  on read-out. The resulting benchmark is  $S_{\mathcal{R}}(t) = \text{tr}(\rho(0)\mathcal{R}(\rho(t)))$ , quantifying the quality of the information extracted from the evolved state and not of the state itself. Robust error corrected logical observables<sup>3</sup> can analogously be defined via the extension provided from the code space to the whole Hilbert space by the recovery operation  $\mathcal{R}$ . In the following sections, we shall consider protecting Hamiltonians associated with such error correcting codes and robust logical observables.

<sup>3</sup> Alicki et al. [16] use the name dressed observables, whereas Chesi et al. [17] use the self-explanatory name of error corrected logical operators, which we will adopt.

### III. LIMITATIONS OF THE 3D COMPASS MODEL

A desirable property for a quantum memory is that any sequence of local operators mapping between different logical code-states should have energy penalties which grow with the system size. It has been shown that this happens for schemes in four dimensions, such as the 4D toric code [11]. Seeking to provide such an example in three spatial dimensions, Bacon proposed the 3D compass model [24], a scheme based on subsystem error correcting codes [29] and requiring only 2-body nearest neighbor interactions. Furthermore, mean field arguments suggest that this model might show such an increasingly large energetic barrier.

However, we will show that the zero temperature (local, but non-Markovian) environment construction of the previous section is capable of giving a false read-out from the code after constant time. We argue that this is due to the choice of the prescribed error correction protocol by showing that the same flaw is present for the 4D toric code if one uses a similar error correction technique. We show that this failure is a general feature of all quantum memory protocols for which the underlying quantum error correcting code does not have a local error threshold (is unable to handle errors on a constant fraction of the sites).

For the 3D compass model, quantum information is first encoded into states of a 2-local Hamiltonian  $H_S$  defined on a  $N \times N \times N$  arrangement of two level systems (where  $N$  is an odd number).

$$H_S = -\lambda \sum_{i,j=1}^N \sum_{l=1}^{N-1} (X_{l,i,j} X_{l+1,i,j} + X_{i,l,j} X_{i,l+1,j} + Z_{i,l,j} Z_{i,l+1,j} + Z_{i,j,l} Z_{i,j,l+1}). \quad (22)$$

This is not a stabilizer code but a subsystem code. This means that a recovery operation need not correct certain errors which have no effect on the logical observables, and information may be preserved even if the recovered state is different.

First, note that pairs of planes of operators  $\hat{Z}_l = \prod_{i,j} Z_{i,j,l} Z_{i,j,l+1}$  and  $\hat{X}_l = \prod_{i,j} X_{l,i,j} X_{l+1,i,j}$  commute with the Hamiltonian  $H$  for all  $l$ . This also holds for logical operators, which consist of products along a single plane  $\bar{Z} \equiv \bar{Z}_l = \prod_{i,j} Z_{i,j,l}$  and  $\bar{X} \equiv \bar{X}_l = \prod_{i,j} X_{l,i,j}$  operators respectively, for an arbitrarily chosen  $l$ . In the ground space, the choice of  $l$  is irrelevant, since the operators  $\bar{Z}_l, \bar{X}_l$  all have  $+1$  eigenvalues. Provided  $N$  is odd,  $\bar{X}$  and  $\bar{Z}$  anti-commute, giving a qubit algebra. In the presence of errors (outside of the ground space), the error corrected logical observables will be defined as the majority vote among plane observables

$$\bar{Z}_{ec} = \text{maj}_l \bar{Z}_l \quad \bar{X}_{ec} = \text{maj}_l \bar{X}_l \quad (23)$$

where maj stands for a majority vote among the  $\pm 1$  eigenvalued commuting operators. Measuring all pairs of adjacent planes  $\bar{Z}_l$  allows a majority vote error correction

scheme to be performed on the value of the  $\bar{Z}_l$  plane observables without extracting whether the corrected state yields  $+1$  or  $-1$  values for all such planes.

Considering a perturbation on the system plus an environment, as presented in section II B. By explicitly developing the final expectation values for the observables of interest ( $\text{tr}(\bar{X}_{ec} \rho_S(t_f))$  and  $\text{tr}(\bar{X}_{ec} \rho_S(t_f))$ ), it can be seen that the information stored in the code will not be reliable after a time  $t_f$ . For this, we can pick up from the evolved state of the system in Eq. (16)

$$\rho_S(t_f) = \left( \bigotimes_{i=1}^N \mathcal{E}_{P_i,p} \right) \circ \mathcal{H}_{t_f} \circ \left( \bigotimes_{i=1}^N \mathcal{E}_{P_i,p} \right) |\psi_0\rangle\langle\psi_0|. \quad (24)$$

Since all plane observables ( $\bar{X}_l$  and  $\bar{Z}_l$ ) of a given type mutually commute, and also do so with the Hamiltonian, we can independently consider the probability of each plane observable having suffered a flip. If the  $P_i$  in Eq. (14) are taken to be single  $X$  or  $Z$  rotations, they will anticommute with overlapping  $\bar{Z}_l$  or  $\bar{X}_l$  plane observables respectively, changing their value upon an odd number of applications. Taking the  $P_i$  to be simple  $Z$  operators, the probability of flipping the value of an  $\bar{X}_l$  plane observables by applying  $\bigotimes_{i=1}^{N^3} \mathcal{E}_{P_i,p}$  once is given by

$$\begin{aligned} p_{\text{plane}*} &= \sum_{i \in \text{odd}}^{i \leq N} \cos^{2N^2-2i}(\varepsilon) \sin^{2i}(\varepsilon) \binom{N^2}{i} \\ &= \frac{1 - \cos^{N^2}(2\varepsilon)}{2}, \end{aligned} \quad (25)$$

which is exponentially close to  $1/2$ . Since all observables involved commute with the system Hamiltonian, the probability of observing any result configuration will be preserved by  $\mathcal{H}_{t_f}$ . Finally, a second round of errors  $\bigotimes_{i=1}^{N^3} \mathcal{E}_{P_i,p}$  will again flip the observed value for each plane with a probability  $p_{\text{plane}*}$ . The final independent probability of flipping the value of each plane is

$$p_{\text{plane}} = 2p_{\text{plane}*}(1 - p_{\text{plane}*}) = \frac{1 - \cos^{2N^2}(2\varepsilon)}{2}. \quad (26)$$

The proposed correction scheme is equivalent to a majority voting among such planes. Thus, if more than half the planes suffer such an error, the majority vote will fail. The probability for incorrectly measuring the error corrected logical observable  $\bar{X}_{ec}$  on read-out is then

$$p_{\text{logic}} = \sum_{i=(N+1)/2}^N p_{\text{plane}}^i (1 - p_{\text{plane}})^{N-i} \binom{N}{i}. \quad (27)$$

Given that  $\frac{1}{2}[1 - \cos^{2N^2}(2\varepsilon)] \leq p_{\text{plane}} \leq \frac{1}{2}$ , we have that

$$\frac{1}{2}[1 - N \cos^{2N^2}(2\varepsilon)] \leq p_{\text{logic}} \leq \frac{1}{2}. \quad (28)$$

Assuming  $\varepsilon$  to be a small constant independent of  $N$ , the probability  $p_{\text{logical}}$  will exponentially approach  $1/2$

for large  $N$ . We conclude that the encoding is not robust against the error model posed by local coupling to a cold adversarial environment.

The problem lies in the error correction mechanism rather than the protecting Hamiltonian itself. This becomes apparent if one applies a similar analysis to the 4D Toric code. There, the suggested noise model does not present a problem, since the usual error correction [11] of the 4D toric code has an error threshold. That is, provided the probability of per-site error,  $p = \sin^2 \varepsilon$  is below this threshold, there exist error correction criteria which succeed with a probability approaching 1 exponentially with  $N$ . On the other hand, we could consider a majority voting version of error correction in this setting, where we measure hyperplanes of  $X$  operators, and apply a majority vote to choose the correct result. In this case, an analysis completely analogous to that of Bacon's 3D compass code would hold proving such a read-out technique unreliable. We conclude that the error correction procedure of the compass code does not allow sufficient resolution to use any potentially topological properties of the encoded quantum information.

The errors introduced by  $\bigotimes_{i=1}^N \mathcal{E}_{P_i, p}$  are sufficiently general to suggest a necessary criterion for Hamiltonian protection of information from weak coupling to a cold environment. Even if we consider the first round of errors and the Hamiltonian evolution as part of the encoding procedure, information should still be able to withstand the probabilistic application of arbitrary local errors. This means that the information, either quantum or classical, should be encoded in such a way as to provide a finite error threshold in the thermodynamic (large  $N$ ) limit. Of course, in practice, it will be desirable to have a code with a fault-tolerance threshold such that when we try to implement the round of error correction, faulty operations can be compensated for.

#### IV. LIMITATIONS OF THE 2D TORIC CODE

We will now show how local Hamiltonian perturbations are capable of introducing uncorrectable errors in Kitaev's 2D toric code Hamiltonian. The introduction of such errors will strongly rely on the lack of string tension on the toric code, suggesting a macroscopic energy barrier may be a necessary requirement. A brief introduction to the toric code Hamiltonian is provided in appendix B, and is recommended to the unfamiliar reader.

Logical operations in the 2D toric code can be realized by creating a pair of anyons, propagating them so as to complete a non-trivial loop, and finally annihilating them. It is roughly such a scheme that will be followed by the perturbations we develop here. Repeating techniques from section II, we may consider the initial state as containing a superposition of local errors which are interpreted as neighboring anyon pairs. A perturbation construction due to Kay [25] allows the deterministic propagation of such anyon pairs along predefined adversar-

ial paths on the lattice. Syndrome measurement allows restricting to a probabilistic picture where error strings corresponding to anyon propagation paths are present with a predefined probability. It is finally the recovery procedure which may possibly complete these errors into logical operations by selecting an incorrect anyon matching.

A family of weak local perturbations capable of probabilistically introducing distant anyon pairs will first be presented. As before, the initial state  $|\psi(0)\rangle$  is assumed to be a ground state of the unperturbed Hamiltonian  $H$ , in this case an  $N \times N$  toric code as in Eq. (B1). After a time  $t_f$  proportional to the maximum desired anyon propagation distance  $D$ , unperturbed syndrome read-out on  $|\psi(t_f)\rangle$  will probabilistically detect distant (as well as local) anyon pairs. Our construction will then be applied to produce a simple set of  $O(N)$  distance anyons such that no syndrome based error correction may be reliably applied. Later, shorter yet more elaborate anyon propagation paths will require explicit analysis of the error correcting probability of different anyon pairing protocols. In this context, we find weak Hamiltonian perturbations are capable of introducing logical errors with a large probability ( $\approx \frac{1}{2}$ ) in a time  $t_f$  logarithmic in the system size  $N$ .

##### A. Probabilistic introduction of distant anyons

Kay [25] showed that local errors (anyons) in the 2D toric code, and other local stabilizer Hamiltonians lacking string tension, can be propagated into logical errors corresponding to almost complete loop operators by a local Hamiltonian perturbation  $P$ . While in his work the initial presence of the anyons was assumed, here, anyons will be introduced with a certain amplitude by a generalization of the Hamiltonian perturbation  $P$ .

Consider introducing perturbations of the form  $V = \mathcal{U}(H+P)\mathcal{U}^\dagger - H$ , where  $\mathcal{U} = \bigotimes_i U_i$  decomposes into weak local unitary rotations, and  $P$  is, as in [25], a weak local perturbation capable of deterministically propagating anyons in a given time  $t_f$ . The perturbed Hamiltonian

$$\tilde{H} = \mathcal{U}(H + P)\mathcal{U}^\dagger \quad (29)$$

induces a time evolution which can be written as

$$|\psi(t)\rangle = e^{-it\tilde{H}}|\psi(0)\rangle = \mathcal{U}e^{-it(H+P)}\mathcal{U}^\dagger|\psi(0)\rangle. \quad (30)$$

In this context,  $\mathcal{U}$  and  $P$  are chosen such that:

1. Neighboring vertex anyon pairs are created by  $\mathcal{U}^\dagger$ , with a certain small amplitude  $O(\varepsilon)$ , by applying weak  $Z$  rotations on connecting edges.
2. Each of the anyons is deterministically propagated by  $P$  along a predefined path. Thus, local excitation pairs become strings of errors defining new positions for the anyon pair.



3. Finally,  $\mathcal{U}$  is unable to remove both the anyons created by  $\mathcal{U}^\dagger$  after at least one of them has been propagated. Moreover, if none were present,  $\mathcal{U}$  creates an anyon pair with amplitude  $O(\varepsilon)$ .

Propagation paths for each anyon are not allowed to overlap but are otherwise completely independent. The propagation of the  $i$ -th anyon along its path  $\ell^{(i)}$  may be attributed to a specific component  $P_i$  of  $P = \sum_i P_i$ . In turn, each component  $P_i$  admits a decomposition

$$P_i = \sum_{j=1}^{|\ell^{(i)}|} J_j^{(i)} T_{\ell_{j-1}^{(i)}, \ell_j^{(i)}}, \quad (31)$$

in terms of local interaction terms  $T_{p,q}$ , where  $\ell_j^{(i)}$  are the anyon locations along the path  $\ell$ . As in [25], the scalar coefficients  $J_j^{(i)}$  are chosen to implement a perfect state transfer [30, 31, 32] and each term  $T_{p,q}$  implement a swap among vertex anyons on  $p$  and  $q$ . If  $p$  and  $q$  are neighboring vertices,  $T_{p,q}$  is defined as

$$T_{p,q} = Z_s \text{frac}(\mathbb{1} - A_p A_q) 2, \quad (32)$$

where  $A_p$  and  $A_q$  are the vertex stabilizer operators corresponding to  $p$  and  $q$  respectively (appendix B) and  $Z_s$  is a  $Z$  rotation on physical site  $s$  corresponding to the edge connecting  $p$  and  $q$ . Furthermore, by allowing  $p$  and  $q$  to be next nearest neighbors, it is possible to have crossing anyon paths  $\ell^{(j)}$ ,  $\ell^{(i)}$  without having them overlap in the anyon locations used. If vertices  $p$  and  $q$  are not neighbors, the same effect is obtained by substituting  $Z_s$  in (32) for a tensor product of  $Z$  operators along an edge path  $\overline{pq}$  from  $p$  to  $q$ ,

$$T_{p,q} = \bigotimes_{s \in \overline{pq}} Z_s \frac{(\mathbb{1} - A_p A_q)}{2}. \quad (33)$$

The distance  $D$  is the maximum number of steps among the different anyon propagation paths  $D = \max_i |\ell^{(i)}|$ . It will be taken as  $D = N/2 - 1$  in section (IV B) and as  $D = O(\log N)$  in section (IV D). Fixing the strength of perturbation terms in  $P$  as  $J_j^{(i)} = \frac{\varepsilon}{D} \sqrt{j(|\ell^{(i)}| + 1 - j)}$  allows the perturbation  $P$  to remain  $\varepsilon$ -bounded while allowing simultaneous perfect anyon transfer in a time  $t_f = \frac{D\pi}{2\varepsilon}$ . Similarly to previous sections, by taking the rotations  $U_j = e^{i\varepsilon Z_j}$  as  $\varepsilon$  weak, the final perturbation required  $V$  will also be composed of  $O(\varepsilon)$  strength interactions involving at most 8 bodies each.

The quantum state before measurement at time  $t_f$  is given in Eq. (30). Expanding  $\mathcal{U}^\dagger$  from  $U_j = e^{i\varepsilon Z_j}$ , we get

$$|\psi(t_f)\rangle = \mathcal{U} e^{-i(P+H)t_f} \bigotimes_j (\cos \varepsilon \mathbb{1}_j - i \sin \varepsilon Z_j) |\psi(0)\rangle, \quad (34)$$

where the index  $j$  ranges over sites of non trivial action for  $\mathcal{U}$ . The state  $|\psi(0)\rangle$  is a ground space eigenstate of

$H$ , and assuming the locations  $j$  on which  $\mathcal{U}$  acts are non neighboring, each  $Z_j$  will increase the energy respect to  $H$  by  $\gamma$ . Furthermore since the energy of a state respects to  $H$  depends only on anyon number and  $P$  is anyon number preserving, we have  $[H, P] = 0$ , allowing us to write

$$|\psi(t_f)\rangle = \mathcal{U} e^{-iPt_f} \bigotimes_j (\cos \varepsilon \mathbb{1}_j - i e^{-i\gamma t_f} \sin \varepsilon Z_j) |\psi(0)\rangle. \quad (35)$$

Since all the propagations in  $P$  commute and correspond to exact transfer of each anyon created by  $\mathcal{U}^\dagger$  precisely at time  $t_f$ , we may write

$$|\psi(t_f)\rangle = \mathcal{U} \prod_j (\cos \varepsilon \mathbb{1} - i e^{-i\gamma t_f} \sin \varepsilon \bigotimes_{i \in \ell^{(j)}} Z_i) |\psi(0)\rangle \quad (36)$$


where  $\ell^{(j)}$  is the path given by the union of  $\{j\}$  and the two propagation paths  $\ell^{(j+)}$  and  $\ell^{(j-)}$  of  $P$  corresponding to the each of the two anyons created by  $Z_j$ . By expanding  $\mathcal{U}$ , we obtain

$$\begin{aligned} |\psi(t_f)\rangle = \prod_j \left[ \cos^2 \varepsilon \mathbb{1} - i e^{-i\gamma t_f} \sin \varepsilon \cos \varepsilon \bigotimes_{i \in \ell^{(j)}} Z_i \right. \\ \left. + i \cos \varepsilon \sin \varepsilon Z_j + \sin^2 \varepsilon e^{-i\gamma t_f} Z_j \bigotimes_{i \in \ell^{(j)}} Z_i \right] |\psi(0)\rangle. \end{aligned} \quad (37)$$

The state  $|\psi(t_f)\rangle$  described by Eq. (37) corresponds to a coherent quantum superposition of applying different error paths. For such unitary evolutions, initially orthogonal states will remain orthogonal and thus fully distinguishable. However, there are at least two mechanisms which lead us to consider a mixed density matrix as the final state. The first, is due to the fact that the actual perturbation applied is not known, and can for instance be taken probabilistically among the family of perturbations described. The second, is unperturbed syndrome measurement  $\mathcal{M}$ , which is the first step of a quantum error correction procedure to recover the initial state.

Syndrome measurement  $\mathcal{M}$  will probabilistically project the state  $|\psi(t_f)\rangle$  into a subspace consistent with a fixed anyon distribution. This is the first step of the recovery operation  $\mathcal{R} = \mathcal{C} \circ \mathcal{M}$ , the sequential application of unperturbed syndrome measurement  $\mathcal{M}$  followed by a syndrome dependent correction operation  $\mathcal{C}$ . Analysis of different correction strategies  $\mathcal{C}$  need only focus on the resulting mixed state  $\mathcal{M}(|\psi(t_f)\rangle\langle\psi(t_f)|)$ . Since for any anyon configuration there is at most one combination of operators yielding it in Eq. (37), the state  $|\psi(t_f)\rangle\langle\psi(t_f)|$  is reduced to a probabilistic application of these operators on  $|\psi(0)\rangle\langle\psi(0)|$ . Again, taking  $\mathcal{E}_{R,p}(\rho) = pR\rho R + (1-p)\rho$ , one may verify that

$$\mathcal{M}(|\psi(t_f)\rangle\langle\psi(t_f)|) = \bigcirc_j \mathcal{E}_{Z_j,p} \mathcal{E}_{\bigotimes_{i \in \ell^{(j)}} Z_i,p} (|\psi(t_f)\rangle\langle\psi(t_f)|) \quad (38)$$



Case	Probability	Errors	Anyons
A	$(1-p)^2$	...III...III...	...0000...0000...
B	$p(1-p)$	...ZZZ...ZIZ...	...0000...0110...
C	$(1-p)p$	...IZI...III...	...0110...0000...
D	$p^2$	...ZIZ...ZIZ...	...0110...0110...

FIG. 1: Two hollow dots indicate positions where a pair of vertex anyons may be created by  $\mathcal{U}^\dagger$  and/or by  $\mathcal{U}$  with probability  $p$ . Anyons created by  $\mathcal{U}^\dagger$  are propagated by  $P$  along the darkening path. A table is provided indicating the probability of possible error configurations and their corresponding syndrome observables (1 (0) representing anyon presence (absence)).

with  $p = \sin^2 \varepsilon$ . Note that the order of application is arbitrary, since the  $\mathcal{E}_{R,p}$  superoperators commute. Thus, one may consider independent probabilities  $p$  for observing each anyon pair created by  $\mathcal{U}^\dagger$  and propagated by  $P$  (or unpropagated anyon pairs created by  $\mathcal{U}$ ). Hence, when instantiating the Hamiltonian perturbation described on a certain set of anyon propagation paths, one need only deal with the independent probabilities of measuring propagated and unpropagated anyon pairs.

### B. Simple error loops in $O(N)$ time

The aim of this subsection is to provide a simple ensemble of perturbations, employing the above construction in such a way that resulting anyon configurations are provably ambiguous, by which we mean that a single anyon configuration could have, with equal likelihood, originated from logically inequivalent errors. This means that for such configurations, the anyon pairing recovery procedure  $\mathcal{C}$  can do no better than guessing, and will complete a logical error with a 50% probability for any possible choice of  $\mathcal{C}$ .

Let us first consider weakly perturbing only in the vicinity of a single row. The joint effect of many such perturbations, will then be shown to produce further degradation of stored information. So,  $\mathcal{U}^\dagger$  introduces a single  $Z$  error (neighboring anyon pair) on a physical start site  $s$  of the row with probability  $p$ . The paths for the perturbation  $P$  are chosen such that both of the produced anyons propagate along the row in opposite directions up to final neighboring locations which are diametrically opposite  $s$  (see figure 1). For  $\varepsilon$  weak perturbations, this requires no more than  $O(N/\varepsilon)$  time. Finally, with probability  $p$ ,  $\mathcal{U}$  may introduce an error at site  $s$  or counter an element of the propagated error chain. As can be seen from the figure, if the anyon introduction site  $s$  is chosen uniformly at random, there are observable anyon configurations which occur with probability  $2p(1-p)$ , which are completely ambiguous (e.g. cases B and C are indistinguishable under exchange of initial site  $s$ ). However, if such a syndrome is measured, the correction proto-

col has a 50% chance of completing a horizontal  $Z$  loop on the lattice, which is equivalent to applying a completely dephasing channel on one of the encoded qubits with probability  $2p - 2p^2$ .

By applying such a perturbation family to  $i \leq N$  rows of the lattice, the probability of not having such a logically dephasing action take place becomes  $(1 - 2p + 2p^2)^i$ , which may be made arbitrarily small for large  $N$  (i.e. an odd number of horizontal  $Z$  loops is completed with a probability exponentially close to  $1/2$ ). Completely analogous string like perturbations exist for any of four logical operators defining the 2-qubit algebra associated to the ground space. Again, by simultaneously considering such perturbations on a sufficiently large set of parallel lines these operators too will be completed with a probability exponentially close to  $1/2$ . Furthermore, by allowing anyons to hop directly to next nearest neighbors (i.e. Eq. (33)), it becomes possible to simultaneously introduce perpendicular yet commuting loop operations as a result of anyon removal.

Simultaneously introducing the four logical operators independently with probability exponentially close to  $\frac{1}{2}$ , would yield a state exponentially close to a maximal mixture over the code space. Our proof requires terms from different perturbation paths to commute, indicating a possible obstacle to achieving this. In practice however, given that different anyons follow roughly ballistic trajectories with a relatively small spread, this does not pose an issue. In appendix C, we show how it is possible to select the set of anyon trajectories in the perturbation such that the order of anyon crossing is well defined (exponentially well in  $N$ ). In turn, this implies an exponentially small deviation from the result of performing such anyon propagations in order, resulting in a state exponentially close to a maximal mixture on the four dimensional code-space.

### C. Localization in 2D stabilizer codes

In the perturbations constructed to introduce logical errors in the toric code, there is a strong use of the energy degeneracy of subspaces with the same number of anyons. The strengths of the different stabilizer terms in the 2D toric code manifest as strengths of local magnetic fields in the effective Hamiltonian of the propagation [25]. However, having exactly the same strength for all local Hamiltonian terms is not an essential feature of the 2D toric code or of stabilizer Hamiltonians in general.

In the unperturbed picture of stabilizer Hamiltonians, excitations are completely localized. However, when different excitations live in a degenerate energy space, perturbations may be very effective at propagating them. With the hope of obtaining some localization with respect to anyon propagation terms, different stabilizer term strengths may be randomly chosen from some range  $\gamma_{\text{upper}} > \gamma_{\text{lower}} > 0$ .

However, a Hamiltonian perturbation may “smooth”

this distribution to take on, at random, only a finite number of discrete energy values, separated by  $\varepsilon$ , the strength of the perturbation. The number of such possible values is given by  $\lceil \frac{\gamma_{\text{upper}} - \gamma_{\text{lower}}}{\varepsilon} \rceil$ , which is therefore also the average spacing between sites of the same energy. Hence, by selecting propagation terms of a similar size, the hopping scheme can route around any defects, and give a logical error. Thus, the argument is unable to guarantee protection against any constant sized perturbation. Nevertheless, it may be that the perturbation terms necessary to break the code should involve a larger number of bodies, which would definitely be an improvement.

In the case of the 2D toric code [25], and all other 2D local stabilizer Hamiltonians [33, 34], there are always logical operations with string like support. This means that, albeit with some possible energetic smoothening, the scheme presented in section IV B can be adapted to introduce logical errors in arbitrary 2D stabilizer Hamiltonians, meaning that the asymptotic lifetime which 2D  $N \times N$  stabilizer codes may guarantee against weak local perturbations cannot be more than  $O(N)$ .

#### D. Logical errors in $O(\log(N))$ time

In the previous section, we gave a rigorous upper bound of  $O(N)$  on the information lifetime of the toric code. This bound coincides with the one provided by Kay in [25], which required initial anyons in the system to be introduced by an unspecified environment. In this subsection, we provide an exponentially tighter bound by concentrating on specific choices for error correction protocols. We argue that it is possible for a Hamiltonian perturbation to introduce ambiguous distributions of anyon configurations in a time logarithmic in  $N$ , i.e. after a time  $t_f \sim O(\log N)$ , error correction succeeds with probability not much higher than  $1/2$ . Figure 2 schematically presents one such perturbation, indicating where anyon pairs should be introduced, and paths  $P_k$  along which they should propagate. The fact that the trajectories have only simple crossings allows them to be implemented by weak local Hamiltonian perturbation terms involving at most 8 bodies, as obtained from Eq. (33), with  $p, q$  being next nearest neighbors. Furthermore, the trajectory length is no more than twice the distance at which the anyon pair is finally separated.

The length of anyon propagation trajectories is  $8S + 4$ , where

$$S = \left\lceil \frac{\ln N}{2p} \right\rceil, \quad (39)$$

and each has  $2S$  simple crossings with other trajectories. The time required to perform such a propagation by fixed strength local perturbations is proportional to  $S$  (i.e. logarithmic in  $N$ ).

A relevant property of such a perturbation is that anyons observed when performing unperturbed error correction after an evolution time  $t_f$  are always collinear.

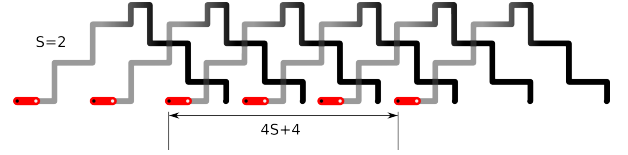


FIG. 2: Anyon pairs corresponding to each thick red edge may be created by  $\mathcal{U}^\dagger$ . After a time  $t_f$ , the right anyon from each pair introduced will be propagated a distance  $4S + 2$  to the right introducing  $Z$  errors along the darkening paths. Finally,  $\mathcal{U}$  acting on the same red segment may move an unpropagated anyon one position to the right or create a neighboring anyon pair on it. The number of big steps (or equivalently of crossings) during the upward propagation is given by  $S$ , which in the case of the figure is 2.

The anyon type and line direction may be chosen to coincide with any of the logical operations, translating to the fact that any logical error may be introduced. This also has the desirable effect of simplifying the analysis of anyon matching criteria. There are only two logically inequivalent anyon matchings on the line, which are the two perfect matchings in which each anyon is paired with one of its two nearest neighbors (i.e. right or left). The point is that one matching will be logically equivalent to the actual trajectories performed by the anyons, canceling any errors introduced, whereas the other will complete the actual paths into a logical error. A simple criterion to determine which case we are dealing with is to count how many times the actual trajectories, together with the anyon matching, cross a vertical line or any homologically equivalent curve. An odd number of crossings means that a logical error has been completed, whereas an even number of crossings means that the proposed pairing has been successful at error correcting.

We study the success probability of two apparently reasonable matching criteria. The first minimizes the furthest distance among paired anyons. The second, for which a polynomial algorithm is known [35], consists of minimizing the sum of distances among paired anyons. Proofs and numerics will be provided for the large  $N$  regime given by  $N \gg 4S + 2$  which convey a high logical error rate.

#### Anyon matching that minimizes $L_\infty$

Let us first consider minimizing the furthest distance among paired anyons. This is the  $L_\infty$  norm of the vector with components given by the individual distances among anyons paired by the matching. We will prove that the probability of introducing a particular logical error is close to  $1/2$  by considering two disjoint scenarios. The first is the very unlikely scenario in which, on syndrome measurement, two consecutive anyons are measured at a distance  $\geq D$  (by consecutive, we mean no additional anyons were measured in the interval between

them). The second is composed of anyon distributions consistent with the measurement of a fixed pair of consecutive anyons at a distance  $\leq D$ . For such distributions, the number of activated anyon paths passing completely over the fixed pair is shown to be odd with probability very close to  $1/2$ .

Let us first bound the probability of observing two consecutive anyons at a distance greater than  $D$  in the syndrome measurement for the evolved state. Given a fixed region of length  $D$ , at least  $\lfloor D/4 \rfloor$  different potential anyon paths start and end in it. Furthermore, assuming  $D < 4S$ , the probabilities for not measuring anyons in this region are independent and are  $1 - p$  for each end of an anyon path and  $(1 - p)^2$  for each start of an anyon path, since both  $\mathcal{U}$  and  $\mathcal{U}^\dagger$  could have created anyons in this case. The anyon-free region can begin in any of  $N$  locations of the full loop. Thus, regardless of correlations, the probability of having  $D$  consecutive anyon-free sites is upper bounded by  $N(1 - p)^{3\lfloor D/4 \rfloor}$ .

Assume now that a pair of consecutive anyons is measured at a distance no greater than  $D$ . There are at least  $\lfloor (4S - D)/4 \rfloor$  potential anyon paths going over this region, each with independent probability  $p$  of being observed. On syndrome read-out, the number of such paths that is activated is odd with a probability approaching  $1/2$  at least as fast as  $\frac{1}{2}(1 \pm (1 - 2p)^{\lfloor (4S - D)/4 \rfloor})$ . Since the  $L_\infty$  norm correction completes a logical error if the most distant consecutive anyon pair is covered by an odd number of activated anyon paths, then by inserting  $S = \lceil \frac{\ln(N)}{2p} \rceil$  and  $D = \lceil \frac{8 \ln(N)}{5p} \rceil$ , we get a probability lower bound for logical errors which approaches  $1/2$  as  $1/2(1 - N^{-1/5})$ .

#### Anyon matching that minimizes $L_1$

Let us now consider the anyon pairing criterion that minimizes the total sum of distances among paired anyons. Since all anyons are found on a loop of length  $N$ , this criterion will always choose a pairing with total distance no greater than  $N/2$ . Thus it will successfully error correct if and only if the total distance of regions of the loop covered an odd number of times by observed anyons is no greater than  $N/2$ . By taking  $S$  to be  $\lceil \frac{\ln N}{2p} \rceil$ , we expect to find roughly half of the sites flipped. To see this, note that, on average, each site is covered approximately  $S\varepsilon$  times. Moreover, the probability of each site being covered an odd number of times is  $\frac{1}{2}[1 - (1 - 2p)^S]$ . For small  $p$  and the chosen value of  $S$ , the average number of sites covered an odd number of times is approximated by  $\frac{N}{2} - \frac{1}{2N}$ . Furthermore, we expect the actual number of such sites to approximately follow a normal distribution around this value, which would imply that logical errors are completed with a probability of close to  $\frac{1}{2}$ . However, since the flipping of different nearby sites are highly correlated events, it is not clear how to go about proving this. Instead, computer simulations (Fig. 3) provide very strong numerical evidence.

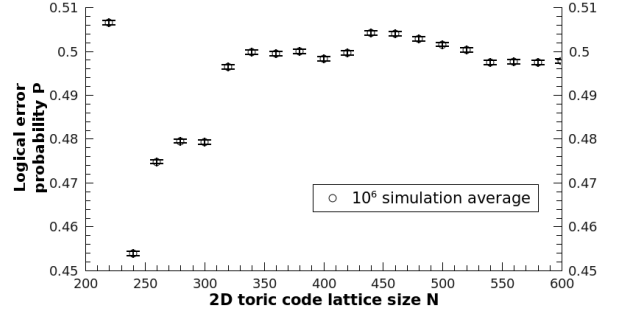


FIG. 3: The average probability of error for  $L_1$  correction after the system evolves for a time  $t_f$  under the described Hamiltonian perturbation. Here, anyon pairs arise, and evolve to distances of  $\lceil 20 \ln(N) \rceil$ , with a probability of 10%, all collinear on a line of length  $N$ . Each point represents an average over  $10^6$  random samples, with error bars representing the magnitude of estimated statistical errors.

## E. Discussion

We have proven that Hamiltonian perturbations can completely destroy the information stored in the 2D toric code in a time proportional to  $N$ . The only assumptions are that the precise Hamiltonian perturbation is unknown, and that recovery begins by performing unperturbed syndrome measurements. A simple family of Hamiltonian perturbations with associated probabilities, was used to justify that the introduction of logical errors in  $O(N)$  time is fully independent from the correction protocol used. This approach remains applicable for arbitrary 2D stabilizer codes, even when the stabilizer terms are of uneven strength.

Furthermore, we have argued that logical errors may be introduced by weak local perturbations in a time logarithmic with the system size. In particular, two apparently reasonable anyon pairing schemes were shown to provide an unreliable recovery mechanism against weak local perturbations acting for  $O(\log N)$  time.

A fully general proof, including all possible error correction strategies based on syndrome measurement is currently lacking for the  $O(\log N)$  error introduction. The generality of the  $O(N)$  construction is obtained by considering a family of different perturbations which could produce the same syndrome outcomes through topologically inequivalent error paths. It may be fruitful to apply such an approach for a general proof of logical errors produced in  $O(\log N)$  time.

## V. LIMITATIONS OF THE 2D ISING MODEL

In present day classical computers, magnetic domains are widely used to provide passive safekeeping of classical information. The Ising model is usually used to elucidate the origin of such long lived magnetized states as a collective effect arising from microscopic local 2-body

interactions

$$H_{\text{Ising}} = -J \sum_{\langle i,j \rangle} Z_i Z_j. \quad (40)$$

In two and higher spatial dimensions, the nearest neighbor Ising model presents a finite temperature phase transition between a disordered phase and an ordered magnetized phase. However, it has long been known that such a system loses its asymptotic bistability under the bias produced by even the weakest of magnetic fields [36, 37, 38]

$$\tilde{H}_{\text{Ising}} = -J \sum_{\langle i,j \rangle} Z_i Z_j + \varepsilon \sum_j Z_j. \quad (41)$$

Such studies consider the dynamics of minority droplets in a 2D Ising model as given by phenomenological equations or the Metropolis algorithm. For a Metropolis algorithm in which such a systematic magnetic field  $\varepsilon$  is included, there is only one stable phase parallel to the field. The anti-parallel phase becomes metastable, with a lifetime exponential in  $J/\varepsilon$ . Dependence of the information lifetime on lattice size  $N$  appears only for the small  $N \lesssim J/\varepsilon$  and will thus not appear if one first takes the limit for large  $N$ .

In this section we consider storing one bit of classical information subject to a quantum evolution of a perturbed 2D Ising Hamiltonian. The observable on which the classical bit is encoded is assumed to be the overall direction for magnetization  $\bar{Z} = \text{maj}_j Z_j$ . The perturbations may conceptually be split into two parts,  $Z$  parallel magnetic fields which introduce additional degeneracies to the Hamiltonian, and transverse magnetic fields or many body terms which couple the new ground states, introducing a hopping between them. The perturbation terms considered will not show support or intensity growing with  $N$ , and they will be capable of introducing logical errors to the unperturbed logical observable  $\bar{Z}$  in a time also independent of  $N$ .

#### A. Hamiltonian perturbation proposal

Consider dividing the  $N \times N$  2D periodic lattice with a chessboard pattern of squares of  $M \times M$  spins where  $M > 4J/\varepsilon_{\text{max}}$  and  $J$  is, again, the nearest neighbor Ising coupling constant and  $\varepsilon_{\text{max}}$  is the greatest local perturbation strength one expects the Hamiltonian to protect information against. For simplicity, we assume  $N = 2nM$ , where  $n$  is an integer. Consider alternately introducing  $\pm \varepsilon Z_j$  magnetic fields in the lattice site  $j$  belonging to white/black squares of the chessboard pattern respectively. The value  $\varepsilon$  is chosen homogeneously for each square such that the energy difference,  $2\varepsilon M^2$ , from fully field parallel and anti-parallel configurations of each square exactly matches the maximum energy difference for border Ising terms  $8MJ$ . For  $N \gg J/\varepsilon$  such a perturbation is always possible.

The point is that now, the ground space of the system acquires a much higher degeneracy, i.e. between  $2^{2n^2+1} - 1$  and  $2^{4n^2}$ . Each black square could be fully magnetized parallel to its preferred field direction or parallel to the direction of its four neighboring squares if it is opposite. By taking one spin variable for each square, the ground states may be identified with those of a 2D  $n \times n$  anti-ferromagnetic Ising model with magnetic field. Three important ground states are the two fully magnetized lattice configurations and the checkerboard configuration in which all spins are fully aligned to their local magnetic field. The gap of these ground states with respect to low lying excited states is  $2\varepsilon$ , which is the energy penalty of flipping a corner lattice site of a square that is fully oriented in the direction of the field but anti-parallel to the two neighboring squares adjacent to the stated corner.

A flipping term for each square of the chessboard should be of the form  $\alpha \bigotimes_j X_j$ , where the  $j$  is taken over all the  $M^2$  sites in the square. Such terms can be introduced either on all black squares or all white squares. This would respectively couple one of the two fully magnetized configurations with the checkerboard configuration, achieving a full swap of state in a time  $t_{\text{flip}} = \frac{\pi}{\alpha}$ . This evolution is exact when such  $M^2$ -body terms of norm  $\alpha$  are allowed, which implies that a proof of Hamiltonian stability will not only require assuming sufficiently weak perturbation terms but also a specific bound for the number of bodies on which such terms act.

Let us now focus on the magnitude of  $\alpha$ . This is the coefficient for a many body term, in which the size of the support scales like  $M^2 = (4J/\varepsilon)^2$ , independent of  $N$ . One may consider obtaining such a term from the  $M^2$ -th order degenerate perturbation theory expansion of fields of the form  $\varepsilon_2 \sum_j X_j$ . For perturbation theory to be strictly valid, one needs  $M^2 \varepsilon_2 < \varepsilon$ . Even then, this small magnitude must be taken to the  $M^2$ -th power to obtain the first non vanishing expansion term. The time required to flip all spins in a plaquette is then proportional to:

$$\alpha \approx \varepsilon_2 \left( \frac{\varepsilon_2}{\varepsilon} \right)^{M^2} \approx \varepsilon M^{-2(M^2-1)} = \varepsilon \left( \frac{\varepsilon}{4J} \right)^{\frac{32J^2}{\varepsilon^2} + 2}. \quad (42)$$

This expression has no dependence on  $N$  and the same perturbation can be introduced in all squares of a given color to yield a fixed flip time. Furthermore, we note that the state of those chessboard squares which are not perturbed is fixed and may be traced out exactly. Hence, the second set of perturbations applied are fully independent and degenerate perturbation theory may be rigorously applied.

#### B. Discussion

Although the flip time shows no dependence on  $N$ , it grows faster than exponentially in terms of  $4J/\varepsilon$ . It may

well be that for magnetic domains, describable by such a 2D Ising Hamiltonian as Eq. (40), the ratio  $4J/\varepsilon$  is sufficiently large to provide a lifetime longer than would be experimentally verifiable. More importantly, we are dealing with an extremely simplified model, with the particularity of neglecting any long range interactions of actual physical systems.

The fact that the perturbation is unknown means that if such a checkerboard state is observed on read-out, information is not recoverable. Such schemes may clearly be generalized to higher dimensions and to deformations of the checkerboard pattern. The existence of such perturbations elucidates important limitations for statements one may formally prove about the classical memory reliability of the Ising model, and therefore what conclusions one might draw about the presence of a macroscopic energy barrier (string tension) which the 2D Ising model certainly possesses. However, it is not clear that these arguments can be applied, for instance, to the 4D toric code since it is a feature of classical memories, but not quantum ones, that local fields can split degeneracies.

## VI. AGGRESSIVE NOISE MODELS

In previous sections, we explored the effects of Hamiltonian perturbations on quantum memories, and particularly on the 2D toric code. We also considered examples of local cold environments perturbatively coupled to a system, as illustrated by Secs. II B and III. The only energy available in these scenarios was due to local perturbations on the system plus environment. Intuitively, a small but constant energy density proportional to  $\varepsilon$  was allowed. While this energy is potentially  $O(N^d)$  for a  $d$  spatial dimension lattice of  $N^d$  qubits, it is difficult to concentrate it in specific regions in order to generate logical errors. In comparison, stabilizer codes only require  $O(N^{d-1})$  energy to implement a logical gate through local rotations.

More aggressive noise models may locally introduce large amounts of energy into the system while keeping perturbation magnitudes weak. Such an example is provided by weak yet time dependent Hamiltonian perturbations. These are relevant when one considers effective protecting Hamiltonians in the interaction picture [39]. Another possibility is to consider the weak coupling of the system to an environment which starts in a high energy state. Noise constructions for these models shall be presented in this section.

In calling such noise models aggressive, we convey the fact that we do not expect “reasonable” Hamiltonian protection schemes to guarantee a long lifetime against such models. Thus, their study may help identify required restrictions on the noise model in order to allow for provably robust Hamiltonian protected memory models. Furthermore, it may provide insight regarding potentially fruitful proof techniques.

### A. Time-varying Perturbations

When considering Hamiltonian perturbations, we assumed that we were unable to determine the new ground space due to the perturbation, and thus encoded in the original ground state space. One might consider an intermediate setting where encoding can be achieved in the perturbed code-space, perhaps due to an adiabatic evolution such as proposed by [14], or by a precise determination of, and compensation for, any perturbations present at the start of the storage time. However, in real experiments, stray fields etc. responsible for perturbations may fluctuate in time. Again, if one can track the changes in perturbations, the proof of Hastings and Wen [13] continues to hold because Lieb-Robinson bounds apply to time-varying local Hamiltonians, and we can therefore adapt the final error correction step as well. Instead, we proceed assuming it is impossible to precisely learn this time variation.

#### 1. Adiabatically varying perturbations

One extreme case to consider is that the perturbation varies adiabatically, so that the system remains in its ground state space. If we do not apply error correction, then we are concerned with how long it takes before the initial and final ground states have a small overlap. We shall assume that the original Hamiltonian  $H$  of  $N$  qubits has an energy gap  $\gamma$ , and we will consider the time-varying perturbation

$$V = \mathcal{U}(t)H\mathcal{U}^\dagger(t) - H$$

where, as before,

$$\mathcal{U}(t) = \prod_{j=1}^N e^{-it\varepsilon X_j/T}$$

and  $T$  is the total time of the evolution, i.e. small local rotations are gradually introduced. At any time  $0 \leq t \leq T$ , the effective Hamiltonian  $\mathcal{U}(t)H\mathcal{U}^\dagger(t)$  has the same energy gap as  $H$ , which means that the adiabatic condition is satisfied for  $T \sim 1/\text{poly}(\gamma)$ . From previous considerations, Eqn. (A9), we know that the overlap of the initial state  $|\psi(0)\rangle$  and the evolved state, the ground state of the adiabatically perturbed Hamiltonian, have an average overlap of no more than  $\text{tr}(P_0)(1 - \frac{3}{4}\sin^2(t\varepsilon/T))^N$ . For large  $N$  and small  $\varepsilon$ , this means that the final overlap is of the order  $\text{tr}(P_0)\exp(-\frac{3\varepsilon^2 N}{4})$  if a phase of error correction is not involved.

When error correction is introduced to this scenario, this maps into the situation where our quantum memory is initially encoded in the perturbed subspace, but decoding is using the original, unperturbed, error correction strategy. In the specific instance of the perturbation  $\mathcal{U}(T)$ , we find that  $X$  rotations are applied probabilistically on each site, and hence our QECC must have a superior error threshold.

Hastings and Wen [13] reveal a similar interpretation holds for all possible perturbations (assuming the gap remains open as the perturbation is introduced) since all local terms are converted into quasi-local rotations.

## 2. Rapidly oscillating perturbations

Another extreme scenario is when perturbations are allowed to oscillate with arbitrary frequencies. A simple construction shows that for stabilizer Hamiltonians, this allows the introduction of arbitrary errors in constant time. We expect that optimal control theory may provide the tools to generalize such results to arbitrary Hamiltonians.

Consider a stabilizer Hamiltonian  $H_0$  and a logical error to implement  $L = P_M P_{M-1} \dots P_2 P_1$ , which is decomposed into Pauli operators  $P_i$  on different sites. We then consider the time dependent Hamiltonian perturbation

$$V(t) = \varepsilon(t) \sum_i e^{-iH_0 t} P_i e^{iH_0 t}. \quad (43)$$

This perturbation is weak if  $\varepsilon(t)$  is sufficiently small. Furthermore, given that  $H_0$  is a stabilizer Hamiltonian,  $V(t)$  may be written as a sum of local terms (at least as local as the stabilizer operators). Finally, the dependence of  $\varepsilon(t)$  on time, is to allow for  $\varepsilon(0) = 0$  which makes the initial encoding equivalent for both perturbed and unperturbed Hamiltonians.

The point of such a perturbation, is that it is possible to explicitly calculate the evolution of the system state in the interaction picture.

$$|\psi_I(t)\rangle = \Pi_i e^{-iP_i \int_0^t \varepsilon(t') dt'} |\psi_I(0)\rangle \quad (44)$$

This means that after a constant time  $t_f$  such that  $\frac{\pi}{2} = \int_0^{t_f} \varepsilon(t') dt'$ , the target operation  $L$  is perfectly implemented in the interaction picture. If  $|\psi(0)\rangle$  is an eigenstate, then  $L$  is also implemented in the Schrödinger picture, modulo a global phase.

Taking  $L$  to be a logical operator of the stabilizer code used, this means that time dependent perturbations of sufficiently high frequency can destroy stored information in constant time. Here, sufficiently high frequency refers to having perturbation terms which oscillate with frequencies at least as high as those corresponding to localized excitations.

## B. Stabilizer Hamiltonians and energetic environment

In what follows, we consider a model in which an environment starts out in an arbitrarily energetic state. However, the couplings between system and environment are required to remain small and local.

For simplicity, we assume that the system is defined by a stabilizer Hamiltonian  $H_S$  and that it starts out in an

eigenstate  $|\psi_0\rangle$  of all stabilizer operators. We will consider a sequence of  $M$  Pauli operators on different sites  $L = P_M P_{M-1} \dots P_2 P_1$  compounding to a logical operation. In the case of translationally invariant stabilizer codes, explicit constructions for these operators are given in [34]. Finally, we may assume a code state  $|\psi_0\rangle$ , such that  $\langle\psi_0|L|\psi_0\rangle = 0$ .

Motivated by the realization that, in order to introduce logical errors, we need to transfer some energy from the environment to the system, we choose a specific environment Hamiltonian  $H_E = -H_S^*$  (at this point, the complex conjugate is unnecessary, but will become useful later). This means that all steps up in energy in the system correspond to an identical step down in energy in the environment. We start the environment state in  $|\psi_0^*\rangle$ .

In this scenario, the coupling

$$H_{SE} = \varepsilon \sum_{i=1}^M P_{S,i} \otimes P_{E,i}^* \quad (45)$$

is enough to produce the logical error  $L$  in constant time  $\frac{\pi}{2\varepsilon}$ . To see this, consider the two states  $P_{S,\mathbf{i}}|\psi_0\rangle P_{E,\mathbf{i}}^*|\psi_0^*\rangle$  and  $P_{S,\mathbf{i}}P_{S,\mathbf{i}}|\psi_0\rangle P_{E,\mathbf{i}}^*P_{E,\mathbf{i}}^*|\psi_0^*\rangle$ . Here, the subindex  $\mathbf{i}$  is an arbitrary binary vector indicating which values of  $j$  a product of  $P_{S,j}$  (respectively  $P_{E,j}^*$ ) should be taken over. First of all, since we have assumed  $H_S$  is a stabilizer, and the  $P_{S,i}$  are Pauli operators, the aforementioned states are zero eigenstates of  $H_S \otimes \mathbb{1}_E + \mathbb{1}_S \otimes H_E$ . Furthermore, the effective Hamiltonian for the perturbation term  $\varepsilon P_{S,\mathbf{i}} \otimes P_{E,\mathbf{i}}^*$  acting on the pair of states  $P_{S,\mathbf{i}}|\psi_0\rangle P_{E,\mathbf{i}}^*|\psi_0^*\rangle$  and  $P_{S,\mathbf{i}}P_{S,\mathbf{i}}|\psi_0\rangle P_{E,\mathbf{i}}^*P_{E,\mathbf{i}}^*|\psi_0^*\rangle$  is just a matrix

$$\varepsilon \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

independent of  $\mathbf{i}$ . This means that we can consider the action of the different  $P_{S,i} \otimes P_{E,i}^*$  terms independently:

$$e^{-iH_{SE}t}|\psi_0\rangle = \left( \bigotimes_{i=1}^M e^{-i\varepsilon P_{S,i} \otimes P_{E,i}^* t} \right) |\psi_0\rangle.$$

Due to the effective Hamiltonian,  $P_{S,\mathbf{i}}|\psi_0\rangle P_{E,\mathbf{i}}^*|\psi_0^*\rangle$  is mapped to  $P_{S,\mathbf{i}}P_{S,\mathbf{i}}|\psi_0\rangle P_{E,\mathbf{i}}^*P_{E,\mathbf{i}}^*|\psi_0^*\rangle$  in a time  $\pi/(2\varepsilon)$ . Thus, the effect of the entire perturbation is to rotate, in a time  $\pi/(2\varepsilon)$  from  $|\psi_0\rangle$  to  $L|\psi_0\rangle$ .

Of course, this approach requires the environment state to have a very high initial energy, namely to start in one of its highest energy states. A refinement of this argument allows us to only change the sign of stabilizers in  $H_E$  which share support with  $L$ . For a local stabilizer code in  $d$  spatial dimensions consisting of  $N^d$  qubits, it was shown [33, 34] that there are logical operators  $L$  with support on  $k \propto N^{d-1}$  sites. The initial state of the environment  $|\psi_E\rangle = |\psi_0^*\rangle$  is still an eigenvector of these stabilizers, with the same eigenvalues. Thus, coupling to an environment with an energy proportional to  $N^{d-1}$ ,



may also introduce logical errors in the same time. This means that the energy required from the environment per system qubit tends to 0 as  $\frac{1}{N}$  (compare to perturbations, which introduce an energy  $\varepsilon$  per site). The catch however, is that the distribution of the energy in the initial environment is highly specific and is in general very different from distributions that may be provided by low temperature thermal states.

We conclude that no stabilizer Hamiltonian will be capable of providing a guarantee for the logical integrity of stored information under the presence of an adversarial, weakly coupled, local environment. Further statistical assumptions such as energy distribution associated to a low temperature environment state need to be included in addition to the weak local coupling assumptions.

### C. Non-stabilizer Hamiltonians

Stabilizer Hamiltonians are not the only possible candidates for providing information protection, although they are particularly attractive because local errors remain as local errors (not propagating or multiplying in the absence of perturbations). Let us now consider the more general case of distance preserving Hamiltonians i.e. ones which might not leave local errors perfectly localized, but do not increase the distance of the error as defined by an error correcting code<sup>4</sup>. Using the same construction as in the previous section, we will show that weak coupling to an environment can also introduce the relevant logical error into a distance preserving Hamiltonian (i.e. a logical operation converting between the most distant code states) for classical memories, by which we mean that one set of local errors becomes irrelevant, say  $Z$  errors, and the presence of a logical error on the classical bit depends only on the local  $X$  errors present. The distance preserving assumption means that the number of  $X$  errors is preserved,  $[H_S, \sum_i Z_i] = 0$ . The maximum distance between any 2 states is for the eigenstates  $\bigotimes_i |0\rangle_i$  and  $\bigotimes_i |1\rangle_i$ , suggesting we should use these states for encoding.

Similarly to the previous subsection, we introduce an environment, and a perturbative coupling between system and environment,

$$H(\varepsilon) = H_S \otimes \mathbb{1}_E + \mathbb{1}_S \otimes H_E + \varepsilon \sum_i X_{i,S} \otimes X_{i,E}. \quad (46)$$

The system Hamiltonian is weakly coupled to a “mirror” system  $H_E = -H_S^*$ . This perturbative coupling

is responsible for the evolution of a mirrored state  $|0\rangle^{\otimes N} |0\rangle^{\otimes N}$ , eigenstate of the unperturbed Hamiltonian  $H(0)$ .

To analyze the evolution, let us consider the action of the operators  $X_i$  in terms of the eigenstates of  $H_S$ . Due to the commutation relation, there must be  $\binom{N}{m}$  eigenstates  $\{|\psi_{m,j}\rangle\}$  of  $H_S$  which are simultaneous eigenvectors of  $\sum_i Z_i$  with eigenvalue  $2m - N$ . We can thus express the eigenvectors  $|\psi_{m,j}\rangle$  of  $H_S$  in terms of the canonical basis as

$$|\psi_{m,j}\rangle = \sum_{\mathbf{i}: w(\mathbf{i})=m} \alpha(m)_{\mathbf{i},j} X_{\mathbf{i}} |0\rangle^{\otimes N} = \sum_{\mathbf{i}: w(\mathbf{i})=m} \alpha(m)_{\mathbf{i},j} |\mathbf{i}\rangle, \quad (47)$$

where  $\mathbf{i}$  are binary vectors with  $m$  non zero components and  $|\mathbf{i}\rangle$  are the respective states from the canonical basis. The matrix  $\alpha(m)$  is unitary as it relates two orthonormal bases of the same subspace. Define

$$\begin{aligned} |\bar{m}\rangle &= \frac{1}{\sqrt{\binom{N}{m}}} \sum_{\mathbf{i}: w(\mathbf{i})=m} |\mathbf{i}\rangle |\mathbf{i}\rangle \\ &= \frac{1}{\sqrt{\binom{N}{m}}} \sum_{j,k, (i): w(\mathbf{i})=m} \alpha(m)_{\mathbf{i},j}^* \alpha(m)_{\mathbf{i},k} |\psi_{m,j}\rangle |\psi_{m,k}^*\rangle \\ &= \frac{1}{\sqrt{\binom{N}{m}}} \sum_j |\psi_{m,j}\rangle |\psi_{m,j}^*\rangle. \end{aligned} \quad (48)$$

From this, one obtains that

$$(H_S \otimes \mathbb{1}_E - \mathbb{1}_S \otimes H_S^*) |\bar{m}\rangle = 0, \quad (49)$$

implying that any non trivial evolution of  $|\bar{m}\rangle$  arises exclusively from the perturbative coupling and is given by

$$H(\varepsilon) |\bar{m}\rangle = \varepsilon J_m |\bar{m} - 1\rangle + \varepsilon J_{m+1} |\bar{m} + 1\rangle, \quad (50)$$

with  $J_m = \sqrt{m(N+1-m)}$ . These are precisely the coefficients performing perfect state transfer between  $|\bar{0}\rangle = |0\rangle^{\otimes 2N}$  and  $|\bar{N}\rangle = |1\rangle^{\otimes 2N}$  in a constant time  $t = \frac{\pi}{2\varepsilon}$  [30, 32].

These results exclude the possibility of proving robustness against weak adversarial coupling to an arbitrarily initialized environment, even of many classical memories using the repetition code (such as Ising models). We learn that if the environment can provide enough energy, then even weak local couplings may be sufficient to produce logical operations. This also motivates the desire to encode in the ground state space of the Hamiltonian since, were we to encode in a higher energy subspace, the environment needs less energy to cause destructive effects. Alternatively, the mechanism presented here could present a useful way to implement gates on a memory.

## VII. FURTHER APPLICATIONS

Constructed perturbations and results presented in this article have focused on elucidating limitations of passive quantum memories. However, our results may be recast in the following other scenarios.

<sup>4</sup> As an aside, note that Bacon’s 3D compass code [24] is an example of a code where the errors do not remain in fixed positions, but preserve the values of the observables (since the observables commute with the Hamiltonian), which reminds us (Sec. III) that error correction is already likely to become much more problematic for these codes.



**Adiabatic Quantum Computation-** The standard approach to adiabatic quantum computation consists of implementing an adiabatic evolution

$$H(t) = f(t)H_i + (1 - f(t))H_f, \quad (51)$$

between Hamiltonians  $H_i$  and  $H_f$ , where  $f(0) = 1$  and  $f(T) = 0$ . While the ground state of the initial Hamiltonian  $H_i$  is expected to be readily prepared, the ground state of the final Hamiltonian  $H_f$  encodes the result of the desired quantum computation. An energy gap no less than  $\gamma$  between ground state and excited states of  $H(t)$  is required for the duration of the adiabatic evolution.

In this context, it is possible that Hamiltonian perturbations could change the initial or final ground state, and maybe even close the gap during the Hamiltonian trajectory. For example, a time dependent perturbation

$$V(t) = (\mathcal{U}H(t)\mathcal{U}^\dagger - H(t)), \quad (52)$$

with  $\mathcal{U}$  defined as in Eq. (3), can make the perturbed initial and final ground states almost orthogonal to the unperturbed versions (see Eq. (8)), while keeping the same gap as  $H(t)$ . Even assuming the perturbed initial ground state is exactly prepared, only if the final state belongs to a code space with an error threshold, will it be possible to reliably recover the desired result, as in VIA.

Connections between adiabatic quantum computation and passive quantum memories can be expected to continue into the regime where error correction is incorporated, and future studies may better elucidate the issues involved in developing a fault-tolerant theory of adiabatic quantum computation [40].

**Topological Quantum Computation-** Difficulties in implementing quantum memories can also be related to some of the difficulties in implementing a topological quantum computation. In particular, in section IV we illustrated how constant Hamiltonian perturbations can create and propagate anyons in the 2D toric code. In the context of topological quantum computation, where gates are implemented through the braiding of anyons, the existence of perturbations capable of creating and propagating anyon pairs is at least equally disturbing as in the memory scenario.

**Quantum Simulations-** One of the most interesting uses of a quantum computer is likely to be the simulation of other quantum systems. While one could express these simulations in terms of the circuit model of quantum computation, and from there create a circuit-based theory of fault-tolerance for quantum simulation, it would be advantageous to understand how this could be implemented more directly, via the simulation of an encoded Hamiltonian.

A logical first step would be to encode the state of each subsystem to be simulated into a quantum memory. Thus, establishing when quantum memories exist, or when they fail, has implications in this case. One of the most commonly applied techniques in Hamiltonian simulation is that of the Trotter-Suzuki decomposition,

where pulses of non-commuting Hamiltonians are combined into one effective Hamiltonian to some accuracy  $\delta$ . This inaccuracy may be treated as a time dependent Hamiltonian perturbation. Given the power such perturbations were shown to have, it is with great care that one should consider the use of passive quantum memories as elements for such quantum simulators.

## VIII. CONCLUSIONS

In this paper, we have studied several constraints on the extent to which a many body Hamiltonian can be expected to protect quantum information. First of all, we showed that gapped local Hamiltonians have eigenstates which are asymptotically unstable under local Hamiltonian perturbations. This result, commonly referred to as Anderson's orthogonality catastrophe [27] shows that a gap is not sufficient to guarantee protection against errors [23, 41]. We proved that a weakly coupled cold environments, can in constant time, alter the evolution of any quantum state leading to an exponentially small overlap between initial and final states. Taking these results together, we conclude that quantum memory schemes must incorporate at least one final round of error correction.

Similar perturbations show why the 3D compass model [24], a self-correcting quantum memory proposal, is not reliable, revealing that the code and error correcting process must possess an error correcting threshold. Similar conclusions may be drawn in scenarios where information encoding and evolution is according to a perturbed Hamiltonian but read-out and decoding are not.

Further explicit counterexamples illustrate that while we may, in some cases, be able to adapt for known perturbations, arbitrary unknown perturbations can destroy the storage properties of codes such as the 2D Toric code in a time  $O(\log N)$ . In this case, the proposed adversarial Hamiltonian perturbation heavily relies on the absence of a macroscopic energy barrier (it is possible to transform orthogonal encoded states via a sequence of local operations while keeping intermediate states in a low energy subspace). By considering the 2D Ising model, we have argued that, in and of itself, a macroscopic energy barrier is not sufficient to protect against perturbations.

Finally, we have considered strong noise models such as time varying Hamiltonian perturbations and weak coupling to an arbitrarily initialized environment. We showed that these noise models could apply logical transformations on information protected by stabilizer Hamiltonians or distance preserving classical memories in constant time. We expect these result to provide insight into how one may prove properties of passive quantum memories and under which assumptions. For instance, since such time-varying Hamiltonian perturbations can destroy the 4D toric code, then when trying to prove robustness against static perturbations, Lieb-Robinson bounds are unlikely to be beneficial.

Having proven a variety of limitations for quantum

memory models and elucidated some required conditions, the next step is to incorporate this deeper understanding into new designs for quantum memories. One major route is to establish a set of necessary and sufficient conditions under which a quantum memory is protected against unknown weak static perturbations. Under such a model, we may once again raise the question of whether good protecting Hamiltonians in two or three spatial dimensions exist. Furthermore, one would hope to find similar conditions under an extended perturbation model allowing a perturbatively coupled local environment. Here, a central problem is to determine which physically realistic assumption may be made on the environment such that positive results are still attainable (i.e. conditions on the initial state of the environment, such as it being prepared in its ground state). Finally, one may study the possibility of engineering an out of equilibrium environment to provide additional protection to quantum information.

We acknowledge the DFG for support (projects Munich Advanced Photonics and FOR635), and the EU (QUOVADIS and SCALA). We would like to thank the hospitality of the Pedro Pascual center for sciences at Benasque while part of this work was being done. ASK is also supported by Clare College, Cambridge. FP thanks Miguel Aguado for helpful discussions.

## APPENDIX A: STATE EVOLUTION IN PERTURBED GAPPED HAMILTONIANS

Energy gaps are considered as a positive feature for a protecting Hamiltonian, since they are expected to provide an energetic barrier which an error process is required to overcome. However, it will be shown that for sufficiently large  $N$ , the fidelity of the unperturbed eigenstates acquires an upper bound close to  $1/2$  after being evolved under the effect of a perturbed Hamiltonian for a time inversely proportional to the gap energy  $\gamma$ .

If a system is perturbed, but we do not know the nature of the perturbation, the best strategy is, arguably, to continue using the unperturbed encoding (i.e. the eigenstates of the unperturbed Hamiltonian). The survival probability for an unperturbed eigenstate  $|\psi_0\rangle$  of  $H$  after evolution under a perturbed Hamiltonian  $\tilde{H}$  for a given time  $t$  (Eq. (3)) is, without error correction

$$S(t) = \left| \langle \psi_0 | e^{-it\tilde{H}} | \psi_0 \rangle \right|^2 = \left| \langle \psi_0 | \mathcal{U} e^{-itH} \mathcal{U}^\dagger | \psi_0 \rangle \right|^2, \quad (\text{A1})$$

i.e. we can express  $S(t)$  as the overlap of  $\mathcal{U}^\dagger |\psi_0\rangle$  with itself under the evolution of the unperturbed Hamiltonian  $H$ . Furthermore, in terms of the eigenstate decomposition

$$\mathcal{U}^\dagger |\psi_0\rangle = \sum_j \alpha_j |\psi_j\rangle \quad \text{where} \quad H |\psi_j\rangle = E_j |\psi_j\rangle, \quad (\text{A2})$$

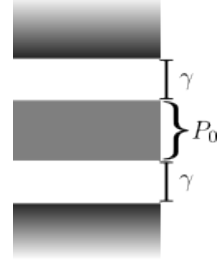


FIG. 4: There is an energy gap  $\gamma$  separating the eigenenergies corresponding to an exponentially small subspace  $P_0$  from the energies of the Hamiltonian eigenstates giving rise to the rest of the Hilbert space.

$S(t)$  may be expanded as

$$S(t) = \sum_{i,j} |\alpha_i|^2 |\alpha_j|^2 \cos[(E_i - E_j)t]. \quad (\text{A3})$$

Assume the initial state  $|\psi_0\rangle$  belongs to an energy subspace  $P_0$  of  $H$  (i.e.  $\langle \psi_0 | P_0 | \psi_0 \rangle = 1$ ), and that  $H$  imposes an energetic gap  $\gamma$  between the subspace  $P_0$  and its orthogonal subspace (see figure 4). This allows the sum in Eq. (A3) to be split as

$$\begin{aligned} S(t) = & \sum_{|\psi_i\rangle, |\psi_j\rangle \in P_0} |\alpha_i|^2 |\alpha_j|^2 \cos((E_i - E_j)t) \\ & + \sum_{|\psi_i\rangle, |\psi_j\rangle \notin P_0} |\alpha_i|^2 |\alpha_j|^2 \cos((E_i - E_j)t) \\ & + 2 \sum_{|\psi_i\rangle \in P_0, |\psi_j\rangle \notin P_0} |\alpha_i|^2 |\alpha_j|^2 \cos((E_i - E_j)t) \end{aligned} \quad (\text{A4})$$

We then define  $R$ , the  $\mathcal{U}P_0\mathcal{U}^\dagger$  subspace overlap of  $|\psi_0\rangle$  as

$$R = \langle \psi_0 | \mathcal{U} P_0 \mathcal{U}^\dagger | \psi_0 \rangle = \sum_{|\psi_i\rangle \in P_0} |\alpha_i|^2. \quad (\text{A5})$$

Taking the time average  $\langle S(t') \rangle_{t' \in [0,t]} = \frac{1}{t} \int_0^t S(t') dt'$ , and noting that

$$|E_i - E_j| \geq \gamma \Rightarrow \left| \int_0^t \cos((E_i - E_j)t') dt' \right| \leq \frac{1}{\gamma}, \quad (\text{A6})$$

Poincaré recurrences are averaged out, providing a bound

$$\langle S(t') \rangle_{t' \in [0,t]} \leq R^2 + (1 - R)^2 + \frac{2}{\gamma t} R(1 - R). \quad (\text{A7})$$

Although the bound in Eq. (A7) is minimized for  $R = 1/2$ , this does not imply that the smallest values for  $\langle S(t') \rangle_{t' \in [0,t]}$  are actually obtained for  $R = 1/2$ .

A sufficient condition for the existence of a weak perturbation yielding  $R = \frac{1}{2}$  may now be obtained by means of continuity arguments. First, note that  $R$  depends continuously on the parameter  $\varepsilon$  appearing in the definition of the rotation  $\mathcal{U}$ , and  $R = 1$  for  $\varepsilon = 0$ . This means that

if, for some  $\varepsilon_0 > 0$ , we find that  $R < 1/2$ , then  $R$  must be equal to  $1/2$  for some smaller positive value  $0 < \varepsilon < \varepsilon_0$ .

As in the previous subsection, we may take  $\langle R \rangle_{\mathcal{U}}$  as an average of the overlap  $R$  over different directions of the rotation  $\mathcal{U}$ . An expression for  $\langle R \rangle_{\mathcal{U}}$ , in terms of the depolarizing channel is given by

$$\langle R \rangle_{\mathcal{U}} = \text{tr} \left( P_0 \Delta_{\lambda(\varepsilon)}^{\otimes N} (|\psi_i\rangle\langle\psi_i|) \right). \quad (\text{A8})$$

Including the dimension of the subspace  $P_0$ , the same bound as in Eq. (8) may be used, leading to

$$\langle R \rangle_{\mathcal{U}} \leq \text{tr}(P_0) \left( 1 - \frac{3}{4} \sin^2(\varepsilon) \right)^N. \quad (\text{A9})$$

If the asymptotic growth of  $\text{tr}(P_0)$  is slower than  $(1 - \frac{3}{4} \sin^2(\varepsilon))^{-N}$ , the bound (A9) will be exponentially decreasing with  $N$ . This means that for sufficiently large  $N$ , and for most directions of rotation, there is some small rotation parameter  $\varepsilon$  yielding  $R = 1/2$ . For the important case of small  $\varepsilon$  and a constant dimension  $\text{tr}(P_0)$ , large  $N$  refers to  $N \sim \mathcal{O}(\varepsilon^{-2})$ .

For those  $\mathcal{U}$  leading to  $R = \frac{1}{2}$ , the time averaged survival probability  $\langle S(t') \rangle_{t' \in [0, t]}$  for the corresponding perturbation may be bounded as

$$\langle S(t') \rangle_{t' \in [0, t]} \leq \frac{1}{2} + \frac{1}{2\gamma t}. \quad (\text{A10})$$

We thus obtain that the overlap of initial encoded states and uncorrected evolved states will drop to values not much larger than  $\frac{1}{2}$  in a time inversely proportional to the gap  $\gamma$ .

## APPENDIX B: THE TORIC CODE

Kitaev introduced the toric code [10] with the intention of achieving reliable storage of quantum information at the physical level, as in classical stable storage, rather than by periodically performing explicit error correction procedures. He proposed that the Hamiltonian of the physical system being used to store the quantum information could, by its nature, make the information stable. His proposal consisted of a 2D system with non trivial topology (such as the surface of a torus) with a stabilizer Hamiltonian composed of local terms. Qubits could then be stored in the ground subspace with a degeneracy of  $4^g$ , with  $g$  being the genus of the surface on which the physical qubits are located.

In the toric code Hamiltonian, the physical qubits are located on the edges of a planar grid covering the 2D surface. For concreteness and simplicity, we shall restrict to the case where the surface is a torus and the grid is an  $N \times N$  square lattice (i.e.  $2N^2$  physical qubits). The Hamiltonian is composed of commuting terms which are products of Pauli operators on different sites (it is a stabilizer Hamiltonian). For each vertex  $s$  of the grid, there

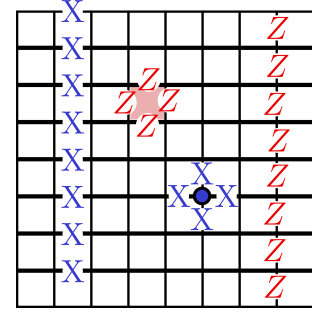


FIG. 5: Each edge in the grid represents a physical qubit and opposite sides of the grid are identified by toric periodic boundary conditions. Typical plaquette and vertex operators are depicted near the center. Two vertical loop operators,  $\bar{X}_1$  and  $\bar{Z}_2$ , which allow breaking the degeneracy are also presented. One can take these to be the  $X$  and  $Z$  operators for the first and second logically encoded qubits respectively. The complementary (anticommuting) operators are given by analogous horizontal loops.

is a star (or vertex) term  $A_s = \prod_{j \in \text{star}(s)} X_j$  which is the product of  $X$  operators over all the qubits of edges reaching  $s$ . Analogously, for each face  $p$  of the grid, there is a plaquette (or face) term  $B_p = \prod_{j \in \text{boundary}(p)} Z_j$  which is the product of  $Z$  operators over all the qubits of edges surrounding the face  $p$ . Since each vertex and face have either 0 or 2 common edges, the terms  $A_s$  and  $B_p$  always commute. Hence all terms of the toric code Hamiltonian

$$H = - \sum_s A_s - \sum_p B_p \quad (\text{B1})$$

commute, and may be simultaneously diagonalized. Since  $\prod_s A_s = I$  and  $\prod_p B_p = I$ , there are only  $2N^2 - 2$  independent binary quantum numbers associated to these terms (stabilizer operators) and each valid configuration determines a subspace of dimension 4. Due to this, violations of plaquette (vertex) conditions  $A_s|\psi\rangle = |\psi\rangle$  ( $B_p|\psi\rangle = |\psi\rangle$ ) always come in respective pairs. Following usual nomenclature, virtual particles called vertex (plaquette) anyons are respectively associated to these excitations. The set of stabilizers may be completed with a pair of logical observables consisting of the product of  $Z$  ( $X$ ) operators along non contractible loops on the lattice (dual lattice), which may not be expressed as a product of plaquette (star) terms as illustrated in figure 5. Together with the set of Hamiltonian stabilizers any commuting pair of these four logical operators ( $\bar{X}_1, \bar{X}_2, \bar{Z}_1$  and  $\bar{Z}_2$ ) uniquely determine the state.

A stated prerequisite for using the toric code as a protecting Hamiltonian is that the energy splitting of the ground space due to Hamiltonian perturbations should be small. This is argued through the use of degenerate perturbation theory and the fact that it only gives non-zero splitting when the order taken is at least the lattice width/height, claiming an exponential suppression of perturbations in the ground space.

The interaction terms in this Hamiltonian may be used as the syndrome measurements of an error correcting code with the desirable property that they are all geometrically local. Such codes provide a way of obtaining a fault tolerance threshold without requiring the use of concatenated quantum error correction. In this case, increasing the lattice size allows periodic measurements to suppress the effect of errors up to any desired accuracy [11] provided the accumulated error probability between measurements is below a certain threshold.

We will briefly review how error syndromes are interpreted and corrected, making the simplifying assumption that the error syndromes are measured perfectly. These syndromes, i.e. measurements of the stabilizers, reveal the presence of any anyons on the lattice, but do not distinguish between them, so it is up to us to determine how these anyons should be paired up in order to annihilate them. For each of the two kinds of anyon, the error correcting procedure will pair up the anyons and annihilate them by applying a connecting string of operators on them. If the connections performed and the actual origin of the anyons form topologically trivial loops (contractible loops), the error correction will have been successful. If however, the actual error pathways, together with the connections performed by the error correction procedure complete one, or an odd number of, non-trivial loops, then a logical error will have been implemented.

Different criteria for pairing anyons may lead to logically different results. This is illustrated in figure 6, where two different criteria are used to pair up six anyons. In particular, if one of the criteria compensates the actual error path, allowing recovery of the initial state, the other will complete the error path into an undesired logical operation.

There are two correction protocols which we will consider, as they are expected to perform adequately when correcting a small proportion of randomly located errors. The first, which we refer to as  $l_1$ -EC, consists of minimizing the sum of distances among paired anyons, for which there is a polynomial time algorithm [35]. The second,  $l_\infty$ -EC minimizes the furthest distance among paired anyons.

### APPENDIX C: FULL DEPOLARIZATION OF THE TORIC CODE'S PROTECTED SUBSPACE

In the main body of the paper (Sec. IV B), we gave a construction for a single logical error  $X_1$ ,  $X_2$ ,  $Z_1$  or  $Z_2$  to be applied with a probability exponentially close to 50%, independent of the model used for error correction. This is not sufficient to show that we get full depolarization of the two-qubit subspace because it is not automatically clear that all 4 logical errors can be introduced simultaneously in the same model; the problem being that crossing paths for anti-commuting operations do not necessarily have a well-defined phase, and the perfect state transfer operations can fail. Indeed, if two non commuting

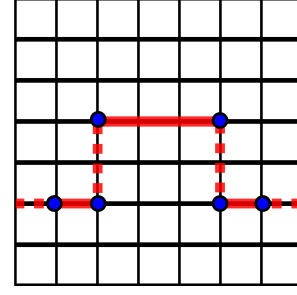


FIG. 6: Illustration of a possible configuration of three vertex anyon pairs (small circles). Segments indicate possible qubits where  $Z$  rotations could be introduced in order to remove the anyons. Solid and dotted segments illustrate the anyon matching arising from  $l_1$ -EC and  $l_\infty$ -EC respectively. Since together they complete a non-trivial loop, the matchings are logically inequivalent.

anyon propagation paths of equal length cross at their midpoints, the amplitude corresponding to full propagation on both paths can be seen to be 0 at times! It is the aim of this section to extend the setting of (Sec. IV B) to multiple logical errors while ensuring that the failure probability remain exponentially small with system size, thereby allowing a fully depolarizing map on the code-space with probability exponentially close to 1.

The basic idea behind this construction is that, for large systems, the propagation of the anyons is essentially ballistic. Hence, we can divide our lattice into sections, and ensure that the paths for anyons of different types only cross in regions where we can be (almost) guaranteed of the order in which the anyons pass through. It is then our task to bound the error probability.

Let us first consider the probability  $p_s(t)$  of finding a propagated anyon at site  $s$  after a propagation time  $t$  which is given in [42] as

$$p_s(t) = f(s; D, \sin^2 t) = \sin^{2s} t \cos^{2(D-s)} t \binom{D}{s}, \quad (C1)$$

where  $f$  is the binomial distribution function and  $D$  is the propagation length (i.e. there are  $D + 1$  possible anyon sites in the path). Here time has been normalized such that perfect transfer occurs at  $t = \frac{\pi}{2}$ . Correspondingly, if  $P$  is the perfect transfer Hamiltonian for vertex anyons and  $\Pi_0$  is the projector onto the subspace with a unique anyon at the transfer start site, then

$$e^{-itP} \Pi_0 = \sum_s \alpha_s(t) Z^{\otimes s} \Pi_0 \quad \text{where } |\alpha_s(t)| = \sqrt{p_s(t)} \quad (C2)$$

and  $Z^{\otimes s}$  is the tensor product of  $s$  consecutive  $Z$  operators along the anyon propagation path.

We are now in condition to compare the actual evolution imposed by two non commuting anyon propagations  $|\psi(t)\rangle$  and an ordered idealization of it  $|\psi_{b;a}(t)\rangle$

$$\begin{aligned} |\psi(t)\rangle &= \mathcal{U} e^{-it(P_a + P_b)} \mathcal{U}^\dagger |\psi_0\rangle \\ |\psi_{b;a}(t)\rangle &= \mathcal{U} e^{-itP_a} e^{-itP_b} \mathcal{U}^\dagger |\psi_0\rangle. \end{aligned} \quad (C3)$$

We will assume that the physical qubit corresponding to the crossing of both paths is between anyon sites  $s_a - 1$  and  $s_a$  of the anyon path associated to  $P_a$  and between anyon sites  $s_b - 1$  and  $s_b$  of the anyon path associated to  $P_b$ . Furthermore, we will assume  $s_a \gg s_b$ , where what is meant by ( $\gg$ ) will soon be made clear. Under these conditions, we will see that  $|\psi(t)\rangle$  and  $|\psi_{b;a}(t)\rangle$  are almost equal (at least during the time period corresponding to perfect state transfer).

By definition, we have that  $\langle\psi(0)|\psi_{b;a}(0)\rangle = 1$ . Let us now bound how fast this overlap can actually decay

$$\frac{d}{dt} \langle\psi(t)|\psi_{b;a}(t)\rangle = i \langle\psi(t)|[P_b, e^{-itP_a}]e^{-itP_b}\mathcal{U}^\dagger|\psi_0\rangle. \quad (\text{C4})$$

This allows bounding

$$\left| \frac{d}{dt} \langle\psi(t)|\psi_{b;a}(t)\rangle \right| \leq \| [P_b, e^{-itP_a}]e^{-itP_b}\mathcal{U}^\dagger|\psi_0\rangle \|. \quad (\text{C5})$$

Now let  $\Pi_\emptyset^{(a)}$  and  $\Pi_0^{(a)}$  be projectors onto the subspace with no anyons in the path of  $P_a$  and the subspace where a single anyon is located at the initial site and define  $\Pi_\emptyset^{(b)}$  and  $\Pi_0^{(b)}$  analogously. Recalling that  $|\psi_0\rangle$  is a code state and our choice of rotation  $\mathcal{U}$ , we have

$$\begin{aligned} (\Pi_\emptyset^{(a)} + \Pi_0^{(a)})\mathcal{U}^\dagger|\psi_0\rangle &= \mathcal{U}^\dagger|\psi_0\rangle \\ (\Pi_\emptyset^{(b)} + \Pi_0^{(b)})\mathcal{U}^\dagger|\psi_0\rangle &= \mathcal{U}^\dagger|\psi_0\rangle. \end{aligned} \quad (\text{C6})$$

Commuting these projectors and using the expansion (C2) of the perfect transfer we may express the RHS of equation (C5) by

$$\left\| [P_b, \sum_s \alpha_s(t) Z^{\otimes s} \Pi_0^{(a)}] \sum_r \alpha_r(t) X^{\otimes r} \Pi_0^{(b)} \mathcal{U}^\dagger|\psi_0\rangle \right\| \quad (\text{C7})$$

There is only one possible non commuting term in  $P_b$  and this only for  $s \geq s_a$ . Furthermore, this term cancels for all but two terms in the sum over  $s'$ . We may then rewrite (C7) as

$$\left\| 2J_{s_b} \sum_{s \geq s_a} \alpha_s(t) Z^{\otimes s} \Pi_0^{(a)} \times (\alpha_{s_b}(t) X^{\otimes s_b-1} + \alpha_{s_b-1}(t) X^{\otimes s_b}) \Pi_0^{(b)} \mathcal{U}^\dagger|\psi_0\rangle \right\| \quad (\text{C8})$$

Where  $J_{s_b}$  is the strength of the term performing an anyon swap between sites  $s_b$  and  $s_b - 1$ . Since each coefficient accompanies an orthogonal component of the state, we may recall the definition in (C2) and rewrite (C8) as

$$2J_{s_b} \sqrt{[p_{s_b-1}(t) + p_{s_b}(t)] \sum_{s \geq s_a} p_s(t) \sin^2 \varepsilon}, \quad (\text{C9})$$

where  $\sin^2 \varepsilon$  is the amplitude of  $\Pi_0^{(a)} \Pi_0^{(b)} \mathcal{U}^\dagger|\psi_0\rangle$ . An exponentially small upper bound will now be given for the expression inside the square root.

$$\begin{aligned} & [p_{s_b-1}(t) + p_{s_b}(t)] \sum_{s \geq s_a} p_s(t) \\ & \leq \sum_{r \leq s_b} f(r, D, \sin^2 t) \sum_{s \geq s_a} f(s, D, \sin^2 t) \quad (\text{C10}) \\ & = F(s_b, D, \sin^2 t) F(D - s_a, D, \cos^2 t), \end{aligned}$$

where  $F(k, N, p) = \sum_{i=0}^k f(i, N, p)$  is the cumulative binomial distribution function. Assuming  $\frac{s_b}{D} \leq \sin^2 t \leq \frac{s_a}{D}$  we may use Hoeffding's inequality [43] to bound (C10) as

$$e^{-2 \frac{(D \sin^2 t - s_b)^2}{D}} e^{-2 \frac{(D \sin^2 t - s_a)^2}{D}} \leq e^{-\frac{(s_a - s_b)^2}{D}}, \quad (\text{C11})$$

with equality holding for  $\sin^2 t = \frac{s_a + s_b}{2D}$ . In turn, a tighter bound can be obtained by using Hoeffding's inequality on a single factor of (C10) when  $\sin^2 t \geq \frac{s_a}{D}$  or  $\frac{s_b}{D} \geq \sin^2 t$ .

Taking  $D = N/2 - 1$  as in Sec. IV B and  $s - r \geq s_a - s_b \geq D/6$  for instance, the obtained upper bound becomes exponentially small in  $N$ . In turn, the derivative (C5) is exponentially small, meaning that the actual evolution is approximated by the ordered evolution with exponentially good precision in  $N$ .

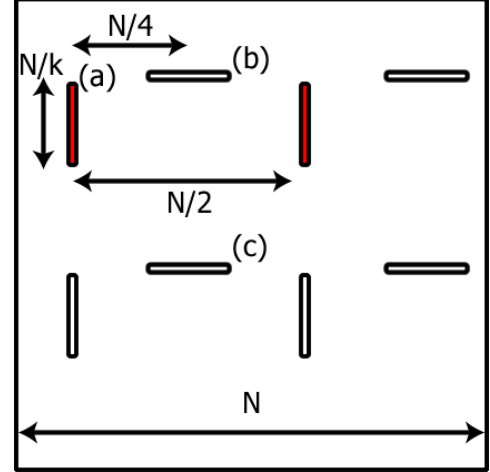


FIG. 7: In an  $N \times N$  lattice, there are two sets of  $N/k$  rows ( $k \sim O(1)$ ) and two sets of columns and rows, each of which corresponds to the construction of (Sec. IV B) for a different error type ( $\bar{X}_1, \bar{Z}_2$  are introduced by columns starting at horizontal stripes and  $\bar{Z}_1$  and  $\bar{X}_2$  are introduced by rows starting from vertical stripes).

We have formally proven that for two non commuting anyon propagation paths which intersect with a sufficiently large offset (i.e.  $\geq D/6$ ) the evolution can be accurately approximated by ordered anyon propagation. There is no obstacle in generalizing this result to many such anyon paths, as required to introduce logical errors with high probability. Some leading factors of order  $N^2$  appear but crucial factors remain exponentially decreasing in  $N$ .

In Fig. 7, we illustrate a configuration allowing the simultaneous introduction of all possible logical errors by anyon propagation within the lattice. The marked stripes of width  $N/k$  indicate locations where perpendicular anyon propagations begin or end. The perturbation to be introduced is chosen randomly as in Sec. IV B, such that each propagation row/column starts with equal probability in either of each pair of opposing stripes. Taking  $k$  fixed allows sufficiently many repetitions of the single row/column construction that the probability of introducing each type of logical error approaches  $\frac{1}{2}$  exponentially fast with  $N$ . Thus, after a perturbed evolution

for time  $t_f = O(N)$ , and a final application of an arbitrary error correcting protocol based on unperturbed syndrome measurement, the resulting state is exponen-

tially close to the maximally mixed state  $\mathbf{1}/4$  of the code-space.

- 
- [1] P. W. Shor, SIAM J. Comput. **26**, 1484 (1997).
  - [2] A. M. Childs and W. van Dam, e-print arXiv:0812.0380 (2008).
  - [3] A. W. Harrow, A. Hassidim, and S. Lloyd, Phys. Rev. Lett. **103**, 150502 (2009).
  - [4] C. H. Bennet and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems & Signal Processing, Bangalore, India* (IEEE, New York, 1984), pp. 175–179.
  - [5] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
  - [6] P. W. Shor, Phys. Rev. A **52**, R2493 (1995).
  - [7] D. Aharonov and M. Ben-Or, quant-ph/9611025 (1996).
  - [8] D. Gottesman, Phys. Rev. A **57**, 127 (1998).
  - [9] D. Gottesman, e-print arXiv:0904.2557 (2009).
  - [10] A. Y. Kitaev, Annals of Physics **303**, 2 (2003).
  - [11] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Jour. Math. Phys. **43**, 4452 (2002).
  - [12] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966).
  - [13] M. B. Hastings and X. G. Wen, Phys. Rev. B **72**, 045141 (2005).
  - [14] A. Hamma and D. A. Lidar, Phys. Rev. Lett. **100**, 030502 (2008).
  - [15] R. Alicki, M. Fannes, and M. Horodecki, e-print arXiv:0810.4584 (2008).
  - [16] R. Alicki, M. Horodecki, P. Horodecki, and R. Horodecki, e-print arXiv:0811.0033 (2008).
  - [17] S. Chesi, D. Loss, S. Bravyi, and B. M. Terhal, e-print arXiv:0907.2807 (2009).
  - [18] F. Pastawski, A. Kay, N. Schuch, and I. Cirac, Phys. Rev. Lett. **103**, 080501 (2009).
  - [19] S. Bravyi, D. P. DiVincenzo, D. Loss, and B. M. Terhal, Phys. Rev. Lett. **101**, 070503 (2008).
  - [20] M. M. Wolf, Nat. Phys. **4**, 834 (2008), ISSN 1745-2473.
  - [21] D. Bacon, Phys. Rev. A **78**, 042324 (2008).
  - [22] S. Chesi, B. Röthlisberger, and D. Loss, e-print arXiv:0908.4264 (2009).
  - [23] A. Hamma, C. Castelnovo, and C. Chamon, Phys. Rev. B **79**, 245122 (2009).
  - [24] D. Bacon, Phys. Rev. A **73**, 012340 (2006).
  - [25] A. Kay, Phys. Rev. Lett. **102**, 070503 (2009).
  - [26] C. King, Trans. Inform. Theory **49**, 221 (2003), ISSN 0018-9448.
  - [27] P. W. Anderson, Phys. Rev. Lett. **18**, 1049 (1967).
  - [28] L. C. Venuti and P. Zanardi, Phys. Rev. Lett. **99**, 095701 (2007).
  - [29] D. Kribs, R. Laflamme, and D. Poulin, Phys. Rev. Lett. **94**, 180501 (2005).
  - [30] M. Christandl, N. Datta, A. Ekert, and A. J. Landahl, Phys. Rev. Lett. **92**, 187902 (2004).
  - [31] P. Karbach, Phys. Rev. A **72**, 030301 (2005).
  - [32] A. Kay, e-print arXiv:0903.4274 (2009).
  - [33] S. Bravyi and B. Terhal, New J. Phys. **11**, 043029 (2009).
  - [34] A. Kay and R. Colbeck, e-print arXiv:0810.3557 (2008).
  - [35] P. R. A. Vaidya, *Geometry helps in matching* (ACM, 1988), pp. 422–425, ISBN 0-89791-264-0.
  - [36] H. L. Richards, S. W. Sides, M. A. Novotny, and P. A. Rikvold, Journal of Magnetism and Magnetic Materials **150**, 37 (1995), ISSN 0304-8853.
  - [37] E. Cirillo and J. Lebowitz, Journal of Statistical Physics **90**, 211 (1998).
  - [38] G. Grinstein, IBM J. Res. & Dev. **48** (2004).
  - [39] K. R. Brown, Phys. Rev. A **76**, 022327 (2007).
  - [40] D. A. Lidar, Phys. Rev. Lett. **100**, 160506 (2008).
  - [41] S. P. Jordan, E. Farhi, and P. W. Shor, Phys. Rev. A **74**, 052322 (2006).
  - [42] A. Kay and M. Ericsson, New Journal of Physics **7**, 143 (2005), ISSN 1367-2630.
  - [43] W. Hoeffding, Journal of the American Statistical Association **58**, 13 (1963), ISSN 01621459.