

## MIT Open Access Articles

### *Self-correcting quantum computers*

The MIT Faculty has made this article openly available. *Please share* how this access benefits you. Your story matters.

**Citation:** Bombin, H, R W Chhajlany, M Horodecki, and M A Martin-Delgado. "Self-correcting quantum computers." *New Journal of Physics* 15, no. 5 (May 1, 2013): 055023. © 2013 IOP Publishing

**As Published:** <http://dx.doi.org/10.1088/1367-2630/15/5/055023>

**Publisher:** IOP Publishing

**Persistent URL:** <http://hdl.handle.net/1721.1/80347>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike 3.0



## Self-correcting quantum computers

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2013 New J. Phys. 15 055023

(<http://iopscience.iop.org/1367-2630/15/5/055023>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 18.51.3.76

The article was downloaded on 07/08/2013 at 19:40

Please note that [terms and conditions apply](#).

## Self-correcting quantum computers

H Bombin<sup>1,5,6</sup>, R W Chhajlany<sup>2</sup>, M Horodecki<sup>3</sup>  
and M A Martin-Delgado<sup>4</sup>

<sup>1</sup> Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>2</sup> Faculty of Physics, Adam Mickiewicz University, Poznań, Poland

<sup>3</sup> Institute of Theoretical Physics and Astrophysics, University of Gdańsk, Poland

<sup>4</sup> Departamento de Física Teórica I, Universidad Complutense, E-28040 Madrid, Spain  
E-mail: [hector.bombin@gmail.com](mailto:hector.bombin@gmail.com)

*New Journal of Physics* **15** (2013) 055023 (42pp)

Received 20 February 2013

Published 29 May 2013

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/15/5/055023

**Abstract.** Is the notion of a quantum computer (QC) resilient to thermal noise unphysical? We address this question from a constructive perspective and show that local quantum Hamiltonian models provide self-correcting QCs. To this end, we first give a sufficient condition on the connectedness of excitations for a stabilizer code model to be a self-correcting quantum memory. We then study the two main examples of topological stabilizer codes in arbitrary dimensions and establish their self-correcting capabilities. Also, we address the transversality properties of topological color codes, showing that six-dimensional color codes provide a self-correcting model that allows the transversal and local implementation of a universal set of operations in seven spatial dimensions. Finally, we give a procedure for initializing such quantum memories at finite temperature.

<sup>5</sup> Current address: Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada.

<sup>6</sup> Author to whom any correspondence should be addressed.



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](http://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

**Contents**

<b>1. Introduction</b>	<b>3</b>
<b>2. Previous results</b>	<b>5</b>
<b>3. Subsystem approach to fault-tolerant quantum computing</b>	<b>5</b>
<b>4. Stabilizer codes and transversal gates</b>	<b>7</b>
4.1. Stabilizer codes . . . . .	7
4.2. Error correction . . . . .	7
4.3. Dressed observables and critical syndromes . . . . .	8
4.4. Transversal gates . . . . .	9
4.5. CSS-like codes . . . . .	10
<b>5. Modeling interaction with the environment</b>	<b>11</b>
5.1. Evolution in the weak coupling limit . . . . .	11
5.2. Fidelity and autocorrelation functions . . . . .	12
5.3. Upper bound for the decay rate . . . . .	13
<b>6. Conditions for self-correcting quantum memories</b>	<b>14</b>
6.1. Stabilizer code models . . . . .	14
6.2. A bound in terms of critical syndromes . . . . .	14
6.3. Sketch of the proof of stability of the Kitaev model . . . . .	15
6.4. A criterion based on connectedness . . . . .	17
6.5. Resilience to noise . . . . .	19
<b>7. Topological stabilizer codes</b>	<b>20</b>
7.1. Local stabilizer codes . . . . .	21
7.2. Homology groups . . . . .	21
7.3. Codes from homology groups . . . . .	22
7.4. Generalized toric codes . . . . .	24
7.5. Generalized color codes . . . . .	26
7.6. Transversality properties of color codes . . . . .	31
<b>8. Self-correcting quantum computers</b>	<b>34</b>
8.1. Initialization from thermal states . . . . .	34
8.2. Measurements . . . . .	35
8.3. Transversal gates . . . . .	36
<b>9. Conclusions</b>	<b>37</b>
<b>10. Later results</b>	<b>38</b>
<b>Acknowledgments</b>	<b>39</b>
<b>Appendix A. Relationship between fidelity and the autocorrelation function: the proof of proposition 1</b>	<b>39</b>
<b>Appendix B. A bound for connected collections</b>	<b>39</b>
<b>Appendix C. Proof of theorem 1</b>	<b>40</b>
<b>References</b>	<b>40</b>

## 1. Introduction

The quest for a robust and stable quantum computer (QC) is a challenge in quantum information science and its construction will imply the discovery of new physics. We may classify QCs according to their protection against different decoherence sources, as follows: (i) bare QC, (ii) externally protected QC and (iii) internally protected QC. This sequence ranges in increasing degree of complexity. Case (i) corresponds to the ideal functioning of a QC without errors. This was the original formulation of the quantum circuit model prior to error correction [1]. The bare QC has already produced new physics by means of small-sized devices with ion traps, optical lattices and others [1, 2]. Case (ii) is a big step forward and corresponds to stabilizing the processing of quantum information by means of acting externally on the QC in order to repair the damage produced by errors and environment interactions. Fault-tolerant quantum computing is the paradigm of this method and the so-called error-threshold theorem is its successful outcome, meaning that quantum computation is not just analogue computation [3–10]. However, this approach has not yet met full experimental success. Case (iii) is a more demanding approach in which we ask the quantum device to have the ability to correct itself whenever an error occurs. Producing such a self-correcting QC will amount to finding a new quantum state of matter, and topological orders [11] constitute promising candidates for this or variants thereof.

In this paper, we take up the challenge of constructing a fully fledged self-correcting QC including all basic operations: (a) initialization, (b) quantum gates and (c) measurements. We address the formulation of this problem from a fundamental viewpoint, i.e. is it, in principle, possible to satisfy operations (a)–(c) in a self-correcting way? We answer positively this question and give an explicit construction of such an internally protected QC. Specifically, our solution is constructed by means of a special type of topological quantum codes, called topological color codes (TCC). To this aim, we construct a new lattice with  $D = 6$  spatial dimensions and the desired properties in section 7.5.1. In order to appreciate the implications of this result, we may draw a comprehensive parallelism between the current achievements for externally and internally protected QCs.

This comparison states that, as long as the error rate  $p$  of the externally protected QC is low enough with respect to the threshold, arbitrary long-term reliable computations can be carried out. The bad news is that typical error thresholds  $p_c$  are very small in practice and no quantum device of this type yet exists. Nevertheless, this result is considered a landmark of quantum computation, a proof of principle that quantum processing is possible. Interestingly enough, as for internally protected QCs, we have a very similar situation, with pros and cons. We shall prove that a complete self-correcting QC can be formulated as long as the spatial dimension  $D$  of the local array of qubits is high enough. This is a remarkable result from a fundamental point of view and sets the critical dimension of our constructions to  $D_c = 6$ . It opens up the way to improve these constructions in the direction of decreasing  $D_c$ , much like methods for increasing  $p_c$  are actively sought after. An important remark is in order here: our models are solely constructed based upon qubits, i.e. spin-1/2 systems and simple interactions. It is known, however, that by changing the content and type of degrees of freedom, a quantum system can lower its critical dimension. Thus, it is conceivable that more elaborate models than those proposed here will be able to perform better as far as  $D_c$  is concerned. If this were not possible, then we would arrive at the remarkable result that a self-correcting QC is not realizable in a three-dimensional (3D) space lattice.

The issue of stability of self-correcting quantum memories [12, 13] at finite temperature has been attracting much attention. One reason for this is that they are more physical than the models of fault-tolerant quantum computation based on the quantum circuit model [14].<sup>7</sup> While the latter is based solely on the naked quantum formalism with no dynamics, the topological memory relies on a suitable Hamiltonian that protects quantum information as classical information is protected in, e.g., a two-dimensional (2D) ferromagnet. The original topological memories are placed on a torus, and it is known that they offer stable qubits only starting from four spatial dimensions [13, 15–18].

Both the toric models and others [19] have some similar drawbacks. They do not support a set of universal gates, and no scheme is known for the initialization of these self-correcting memories in the Hamiltonian spirit. Namely, one would like to have something similar to an Ising model, where the state of definite magnetization is prepared by simply switching on a global magnetic field and then switching it off adiabatically. Thus, disregarding for a while the dimension of the model, an interesting question arises: does there exist a self-correcting QC, allowing universal computation as well as a natural initialization scheme?

In this paper, we propose a first such model. To this end, we use color codes, a class of topological codes with very special transversality properties [20, 21]. Color codes are obtained from colexes, a class of lattices introduced in [22], and their underlying mathematical structure is a simplicial homology that differs from the usual one because it is ‘colored’. In the color code Hamiltonian models excitations take the form of branyons [22], point-like or extended objects subject to topological interactions. These models are local but require interactions between many spin-1/2 systems, a difficulty that can be overcome. For example, Weimer *et al* [23] introduce an experimental proposal to implement the many-body interactions present both in toric and color codes. Alternatively, a local two-body Hamiltonian model that effectively yields a 2D color code has been also proposed [24] (see also [25] in this context).

The main idea of the present paper is that in a suitable spatial dimension it is possible to find color codes that support universal quantum computing and, at the same time, provide self-correction (which we prove giving a Peierls argument in the style of [13, 16]). Namely, we construct a six-dimensional (6D) code that supports the transversal controlled not (CNot) gate and  $\pi/4$  phase gate, and transversal measurement in the  $X$  and  $Z$  basis, as well as initialization in those bases. This allows for universal quantum computation [50]. One more dimension enables local CNot gates.

We also present a scheme of initialization, analogous to that of a ferromagnetic memory. We believe that this is the first proposal of such an initialization, as opposed to standard methods in fault tolerance (usually based on measuring syndrome on some initial product state). There is a remarkable difference between our initialization scheme and that of a ferromagnetic memory. Namely, we need to switch off at some point part of the protecting Hamiltonian, which is not needed in the latter case. Alternatively, we have proposed another initialization method based on code deformation that does not require switching on/off the Hamiltonian (see section 8.1.2). Let us emphasize that we are always using the most realistic scenario; namely we assume finite temperature  $T > 0$ , implying that our notion of self-correction is very strong.

This paper is organized as follows. We first recall the previous results regarding thermal stability of self-correcting quantum memories (section 2). In section 3, we discuss the subsystem picture, which is natural for our analysis of thermal stability. In section 4, we describe the

<sup>7</sup> For critical analysis of fault-tolerant quantum computing from the physical point of view, see [14].

general properties of stabilizer codes. We also introduce ‘dressed’ observables—the ones that define the protected qubit. In section 5, we present our modeling of the interaction with the environment. We recall an upper bound on the decay rate of the autocorrelation functions, and its relation to fidelity. We then formulate sufficient conditions for self-correction, related to the Peierls argument in section 6. In addition, we prove that the given conditions also ensure that the quantum memory is resilient to a certain amount of noise applied suddenly. Next, we turn to the construction of the models, which must satisfy the conditions and at the same time allow a universal set of transversal operations (section 7). We give a unified homological picture of toric codes and color codes, and prove in detail that some color codes possess both properties at the same time. Finally, in section 8, we propose a scheme for initialization, measurement and gate application.

## 2. Previous results

The idea of a thermally stable self-correcting quantum memory goes back to [13]. The authors propose a four-dimensional (4D) topological model and use a Peierls argument to provide heuristic basis in favor of its thermal stability. In [16], this was proved rigorously using the formalism of quantum dynamical semigroups. In this paper, we generalize the latter proof. We give stability conditions that apply to general stabilizer codes and then apply them to toric and color codes.

Most likely, there is no stable quantum memory in 2D. A full proof has been given in [15] in the case of the 2D Kitaev model (see also [27] in this context). The no-go results for stabilizer codes have been provided in [17, 18]. Even though these results do not lead to a full proof of instability in the weak coupling limit, they point out quite clearly that there is no hope of stability in such models.

Finally, let us note that many works have studied the existence of topological order in finite temperature (see e.g. [28–30]) and the main conclusions were compatible with the results on stability of quantum memories.

## 3. Subsystem approach to fault-tolerant quantum computing

The most basic approach to quantum error correction is the following. One considers a subspace, the code, chosen in such a way that if the initial state belongs to the subspace and if not too many errors hit the system, there exists a quantum operation, the recovery map, that restores the initial state.

Such a scheme assumes ideal encoding and decoding, but in reality all operations are exposed to decoherence. The prepared state is already erroneous to a certain extent. Then, in fault-tolerant computing schemes, it is subjected to repetitive error correction that keeps the amount of errors stable. At the end of computation, if the ideal recovery existed, one could map the state back to the code subspace, but this is not really available or needed. Rather, one wants to measure some observable on the state. The measurement is nonideal, but if the state did not suffer too many errors and the measurement does not produce too many either, then the statistics will be arbitrarily close to those expected in the ideal scenario, where encoding and recovery are perfect operations.

To describe the above situation it is more convenient to consider a virtual subsystem [31] (see also [32] in this context). Then the fact that we cannot apply perfect operations does not

exclude that we can prepare the subsystem in a given state with an arbitrarily high fidelity. For example, in standard hard drives, the sign of any single spin varies in time. However, the majority of signs of all spins (i.e. magnetization) are preserved with tremendous perfection.

The virtual subsystem is actually strictly related to the recovery procedure (see [33] in this context). Namely, the observables acting on the code are lifted to the full space by subjecting them to an operation dual to the recovery map [16] (we shall present this construction in section 4.3).

In the virtual subsystem picture, if we were able to start with the state within the code the initial state would be  $\psi_V \otimes \phi_{\text{Anc}}^0$ , where  $V$  denotes the virtual subsystem and the second subsystem is  $\mathcal{H}_{\text{Anc}}$ , where  $\psi$  is the wanted state of the subsystem, then the errors (if not too many of them occur) affect only the second subsystem, whose state turns into a mixed state.

In reality the encoding is not error-free, so that the initial state is of the form  $|\psi\rangle\langle\psi|_V \otimes \rho_{\text{Anc}}^0$ . If the state  $\rho^0$  meets some conditions, then it is capable of accumulating the errors, which then do not affect the state  $\psi_V$ . However, in the absence of some additional mechanism, we will face some saturation, and the errors will affect the system  $V$ . In fault-tolerant schemes, there are two contradictory forces: the environment adds errors (which propagate), while the user tries to correct errors. When these forces are balanced we get a dynamical equilibrium: the system is always in a state of the form  $|\psi\rangle\langle\psi|_V \otimes \rho_{\text{Anc}}$ , i.e. the saturation will never occur.

The above formulation is natural not only for traditional fault-tolerant error correction schemes, but also for self-correcting models. In these models, there is a self-Hamiltonian that causes the environment to do both jobs: adding and removing errors (the latter due to the Boltzmann factor). In such a situation, the system is (after initial equilibration) for a long time in the state  $|\psi\rangle\langle\psi|_V \otimes \rho_\beta$ , where  $\rho_\beta$  is the thermal equilibrium state of the second subsystem. The Gibbs state of the whole system would be proportional to  $I_V \otimes \rho_\beta$ , and the whole system will eventually relax to this state. However, there may exist a critical temperature below which the state of  $V$  will be unaffected for a time that scales exponentially with the size of the total system [13].

In standard fault-tolerant schemes (where there is active error correction), entropy is produced by the environment and reduced by the procedure of error correction. In the self-correcting case, both processes are performed by the environment. The fight (as usual in phase transition phenomena) is between entropy and energy—entropy causes errors and energy forces correction via Boltzmann factors.

Let us emphasize that we do not deal here with the so-called ‘noiseless subsystem’ [31]. The noiseless subsystem is protected against some class of errors for the *arbitrary* state of the rest of the system, while here we have protection only when the rest is in some particular state, e.g. Gibbs state, in the case of self-correcting models. Also an error-correcting code usually does not form a noiseless subsystem: assuming that the code can correct up to  $k$  errors, only if we start with the initial pure state  $\psi_V \otimes \phi_{\text{Anc}}$  the subsystem can indeed tolerate  $k$  errors. Indeed to tolerate these errors we need to start from the code, which in subsystem picture is the span of  $\psi_V \otimes \phi_{\text{Anc}}$  for all  $\psi$  from the subsystem, but for a fixed  $\phi_{\text{Anc}}$ .

The sketched subsystem picture is especially useful if we want to analyze the interaction of the quantum memory with the environment. Then there are known tools to work with autocorrelation functions of observables. To estimate the stability of the memory it suffices to show that these autocorrelations do not decay in time. The suitable observables are those from the algebra of the virtual subsystem, and their autocorrelation functions can be related to the fidelity of the state of the virtual subsystem.



In section 4, we shall describe, first in the standard picture, the error correction based on stabilizer formalism. Then we shall construct suitable ‘dressed’ observables that define the virtual subsystem.

## 4. Stabilizer codes and transversal gates

### 4.1. Stabilizer codes

An error-correcting code [34, 35] is a subspace  $\mathcal{C}$  of the Hilbert space  $\mathcal{H}$  that represents a quantum system. The code subspace  $\mathcal{C}$  is such that any quantum information stored on it can be recovered after it has suffered errors from a set  $\mathcal{E}$  of correctable errors. Usually, error-correcting codes are subspaces of systems of  $n$  qubits, that is,  $\mathcal{H} = \mathcal{H}_2^{\otimes n}$  with  $\mathcal{H}_2$  a 2D Hilbert space with orthonormal basis  $\{|0\rangle, |1\rangle\}$ . If  $\mathcal{C}$  encodes  $k$  qubits, that is, has dimension  $2^k$ , and corrects any error with support on less than  $d/2$  qubits, we say that the code is a  $[n, k, d]$  code. The support of an operator is composed of those qubits in which it acts nontrivially. In this context, the size of the support of an operator  $O$  is usually called its weight and denoted as  $|O|$ .

An important class of error-correcting codes is that of stabilizer codes [36, 37], which are defined in terms of an abelian subgroup  $\mathcal{S} \subset \mathcal{P}$  of the Pauli group  $\mathcal{P}$  of  $n$  qubits. This is the group

$$\mathcal{P} = \langle iI, X_1, Z_1, \dots, X_n, Z_n \rangle, \quad (1)$$

which has as generators the usual Pauli operators  $X_i, Z_i$  on the  $i$ th qubit,

$$X = |+\rangle\langle+| - |-\rangle\langle-|, \quad Z = |0\rangle\langle 0| - |1\rangle\langle 1|, \quad (2)$$

where  $|\pm\rangle := (|0\rangle \pm |1\rangle)/\sqrt{2}$ . The subgroup  $\mathcal{S}$  is called the stabilizer of the code  $\mathcal{C}$  because its elements  $|\psi\rangle \in \mathcal{C}$  are defined by

$$\forall s \in \mathcal{S} \quad s|\psi\rangle = |\psi\rangle. \quad (3)$$

Note that  $\mathcal{S}$  cannot contain the element  $-I$ . The number of encoded qubits in a stabilizer code is

$$k = n - g, \quad (4)$$

with  $g$  being the number of independent generators of  $\mathcal{S}$ . To give an expression for the distance of the code, we have to introduce the normalizer group  $\mathcal{N}$ . This is the normalizer of  $\mathcal{S}$  in  $\mathcal{P}$ , which coincides with the centralizer of  $\mathcal{S}$  in  $\mathcal{P}$ : its elements  $n \in \mathcal{N} \subset \mathcal{P}$  commute with all the stabilizers  $s \in \mathcal{S}$ . Then the distance of the code  $\mathcal{C}$  is given by the minimum weight of the elements of  $\mathcal{N} - \mathcal{S}$ , which indeed implement nontrivial unitary operations on encoded states. In fact, it is possible to choose among them the  $X$  and  $Z$  Pauli operators for the  $k$ -encoded qubits:  $\bar{X}_1, \bar{Z}_1, \dots, \bar{X}_k, \bar{Z}_k \in \mathcal{N} - \mathcal{S}$ . These operators must satisfy the usual commutation relations, so that we have  $\bar{X}_i^2 = \bar{Z}_i^2 = 1$ ,  $\{\bar{X}_i, \bar{Z}_i\} = 0$  and  $[\bar{X}_i, \bar{Z}_j] = [\bar{X}_i, \bar{X}_j] = [\bar{Z}_i, \bar{Z}_j] = 0$  for  $i \neq j$ .

### 4.2. Error correction

If we put some quantum information into a given code subspace, we know that as long as the system does not suffer too many errors it is in principle possible to recover the information without any losses. But of course we are interested in recovering it in practice, which is not necessarily just as easy.

Stabilizer codes have a particularly simple structure that makes recovery from errors relatively easy. The first step in error correction is the measurement of a set of generators  $s_i$  of the stabilizer

$$\mathcal{S} = \langle s_1, \dots, s_g \rangle. \quad (5)$$

Let us denote the eigenvalue of the  $i$ th generator after the measurement as  $(-1)^{b_i}$ ,  $b_i = 0, 1 \in \mathbf{Z}_2$ . The list of eigenvalues  $\mathbf{b} = (b_i) \in \mathbf{Z}_2^g$  is called the error syndrome. The second step is then to choose, using the syndrome  $\mathbf{s}$ , a Pauli operator  $K \in \mathcal{P}$  such that

$$K s_i = (-1)^{b_i} s_i K. \quad (6)$$

We set  $\text{synd}(K) := \mathbf{b}$  when (6) holds, so that  $\text{synd}$  defines a group homomorphism

$$\text{synd} : \mathcal{P} \longrightarrow \mathbf{Z}_2^g, \quad (7)$$

$$K \longmapsto \mathbf{b}, \quad (8)$$

with kernel  $\mathcal{N}$ .

If the correction operator  $K$  is applied to the system it will return to the code subspace  $\mathcal{C}$ . Whether it returns to the original state that we had encoded will depend on the errors that occurred and the choice of  $K$ . To clarify this point, suppose for simplicity that the action of errors amounts just to the application of a Pauli operator  $E \in \mathcal{P}$  to the system. Then we have by assumption  $KE \in \mathcal{N}$ , and the error-correction procedure will succeed if in addition  $KE \in \mathcal{S}$ .

In summary, the whole error-correction procedure is encoded in a function

$$\text{corr} : \mathbf{Z}_2^g \longrightarrow \mathcal{P}, \quad (9)$$

$$\mathbf{b} \longmapsto K \quad (10)$$

that takes syndromes  $\mathbf{b}$  to correction operators  $K$ . This  $\text{corr}$  is such that

$$\text{synd}(\text{corr}(\mathbf{b})) = \mathbf{b}.$$

One should choose  $\text{corr}$  in such a way that the probability of success for a given source of errors is maximized. However, in practice it is also important that  $\text{corr}$  can be computed efficiently (say, in polynomial time in the number  $n$  of physical qubits), a requirement that may reduce the success probability with respect to the ideal case.

#### 4.3. Dressed observables and critical syndromes

Take any nontrivial self-adjoint encoded operator  $N \in \mathcal{N} - \mathcal{S}$ . Suppose that we initialize the system in an encoded state and we want to keep track of the value of these encoded operators as errors occur, from a theoretical perspective. In particular, we would like to keep track of the value of  $N$  were we to correct the system using the above procedure. We may for this purpose introduce the operator

$$N_{\text{dr}} := \sum_{\mathbf{b} \in \mathbf{Z}_2^g} \text{corr}(\mathbf{b})^\dagger N \text{corr}(\mathbf{b}) P_{\mathbf{b}}, \quad (11)$$

where  $P_{\mathbf{b}}$  is the projector onto the subspace with syndrome  $\mathbf{b}$ , that is,

$$P_{\mathbf{b}} := 2^{-g} \prod_{i=1}^g (1 + (-1)^{b_i} s_i). \quad (12)$$

Note that for encoded states  $N_{\text{dr}} = N$ . The operator  $N_{\text{dr}}$  is a more stable version of  $N$ , because as long as the system only suffers correctable errors its expectation value does not change. This is not the case for  $N$ , which may change even with an error on a single qubit.

Given an encoded operator  $N$  it is useful to introduce the notion of critical syndromes as those which are just a single-qubit error apart from changing the value of  $N_{\text{dr}}$ . Let  $|\psi_{\mathbf{b}}\rangle$  denote a normalized state with  $P_{\mathbf{b}}|\psi_{\mathbf{b}}\rangle = |\psi_{\mathbf{b}}\rangle$ .

For any Pauli operator  $E \in \mathcal{P}$ , we have

$$N_{\text{dr}} E P_{\mathbf{b}} = S(N, E, \mathbf{b}) E N_{\text{dr}} P_{\mathbf{b}}, \quad (13)$$

where

$$S(N, E, \mathbf{b}) = \pm 1 \quad (14)$$

is negative if and only if the encoded operator

$$\text{corr}(\mathbf{b}) E \text{corr}(\mathbf{b} + \text{synd}(E)) \quad (15)$$

anticommutes with  $N$ . Note that for  $E = E_1 E_2$  we have

$$S(N, E, \mathbf{b}) = S(N, E_1, \mathbf{b} + \text{synd}(E_2)) S(N, E_2, \mathbf{b}). \quad (16)$$

When there exists a single-qubit Pauli operator  $\sigma$  such that  $S(N, \sigma, \mathbf{b}) = -1$ , we say that  $\mathbf{b}$  is  $N$ -critical. For each  $N$ , we denote by  $\text{crit}_N \subset \mathbf{Z}_2^g$  the set of  $N$ -critical syndromes  $\mathbf{b}$ . Similarly, we denote by  $\text{crit}'_N \subset \mathcal{P} \times \mathbf{Z}_2^g$  the set of pairs  $(E, \mathbf{b})$  with  $S(N, E, \mathbf{b}) = -1$ . Note that for any  $(E, \mathbf{b}) \in \text{crit}'_N$  there exists a decomposition  $E = E_1 E_2$  such that  $E_1$  and  $E_2$  have disjoint supports and  $\mathbf{b} + \text{synd}(E_2) \in \text{crit}_N$ .

Let us make a connection with the general discussion from section 3. The dressed observables  $N_{\text{dr}}$  define the virtual subsystem. They are built by means of observables that act solely on the code. Indeed, even though in expression (11) we use an observable  $N$  that acts on the whole Hilbert space, we use only its restriction to the code. Thus, we obtained  $N_{\text{dr}}$  by lifting observables acting on the code to those acting on the full system by using the correction procedure.

The notion of critical syndrome refers to the ‘capacity’ of error storage on the ancilla system: when we are able to stay away from states with a critical syndrome, our virtual system is protected. If the ancilla state  $a$  has critical syndrome, then one more added error affects the virtual subsystem. The system is self-correcting, if, for a long time, the ancilla system is far from a critical syndrome.

#### 4.4. Transversal gates

An important feature of several classes of stabilizers codes is that they allow the implementation of logical gates transversally, that is, performing unitary gates on individual qubits or in small disjoint sets of them. Usually, we are interested in stabilizer codes  $\mathcal{C}$  that encode a single qubit. Then a one-qubit logical gate  $G$  that can be implemented in  $\mathcal{C}$  through a unitary operation of the form

$$U = \bigotimes_i U_i \quad (17)$$

is transversal. Here the index  $i$  runs over physical qubits. Frequently, one has  $U_i = U_j$ , which means that no specific addressing of the physical qubits is necessary; they are all treated on an equal footing.

A particularly simple case is that of  $U$  belonging to the normalizer of  $\mathcal{P}$ , usually called the Clifford group. Then  $U$  is a transversal gate if for any  $s \in \mathcal{S}$

$$U^\dagger s U \in \mathcal{S}, \quad (18)$$

because for any  $|\psi\rangle \in \mathcal{C}$  we must have  $sU|\psi\rangle = U|\psi\rangle$ . An advantage of such transversal gates is that when applied with perfect accuracy, they affect the error syndrome of the code in a definite way. More generally, it is enough for  $U$  to satisfy

$$U^\dagger P_0 U = P_0. \quad (19)$$

In section 7.6.3, we will consider such gates, which are essential for universal quantum computation. It is clear that from condition (19) we get no information about how  $U$  affects the syndrome. However, (17) is a big constraint in this regard. In particular, errors cannot spread. That is, if  $|\psi\rangle$  is an encoded state and  $E$  an error, then  $UE|\psi\rangle = E'U|\psi\rangle$  for some  $E'$  with the same support as  $E$ . A less trivial observation is that, in a certain sense, negative error syndromes (i.e. syndromes with  $b_i = -1$ ) also do not spread. Suppose that we choose a possibly overcomplete set of stabilizer generators which are local in a given lattice of a certain dimension where the qubits live. Take a region  $R$  of this lattice such that no encoded operator has a support completely contained on it. Then if  $O$  is a Pauli operator that commutes with all the local generators that have part of their support in  $R$ , it follows as a slight generalization of the cleaning lemma of [18] that  $O|\psi\rangle = O'|\psi\rangle$  for some  $O'$  with support only outside  $R$ . This means that if  $E|\psi\rangle$  is free of negative error syndromes in a region, then the same will be true for  $UE|\psi\rangle$  because  $E|\psi\rangle = E''|\psi\rangle$  for some  $E''$  with support outside  $R$ .

Similarly, we can consider transversal two-qubit logical gates, which are typically implemented on a pair of equal stabilizer codes through unitaries of the form (17), where now  $i$  runs over pairs of equivalent physical qubits on the two codes, so that the  $U_i$  are two-qubit gates. Finally, it is also possible to consider transversal measurements, where each physical qubit is measured on a given basis and from the net result a measurement for the encoded qubit is obtained.

#### 4.5. CSS-like codes

Calderbank–Shor–Steane (CSS)-like codes are those stabilizer codes for which the generators of the stabilizer can be chosen so that  $\mathcal{S} = \langle s_1^X, \dots, s_{g_1}^X, s_1^Z, \dots, s_{g_2}^Z \rangle$ ,  $g_1 + g_2 = g$ ,

$$s_i^X \in \mathcal{P}_X, \quad s_i^Z \in \mathcal{P}_Z,$$

with  $\mathcal{P}_X := \langle X_1, \dots, X_n \rangle$ ,  $\mathcal{P}_Z := \langle Z_1, \dots, Z_n \rangle$ . In such codes,  $X$ -type ( $Z$ -type) logical operators may be chosen from  $\mathcal{P}_X$  ( $\mathcal{P}_Z$ ), that is, so that

$$\bar{X}_i \in \mathcal{P}_X, \quad \bar{Z}_i \in \mathcal{P}_Z.$$

Then  $X$ -type ( $Z$ -type) errors can only change the value of  $Z$ -type ( $X$ -type) encoded operators, and their correction only involves  $Z$ -type ( $X$ -type) stabilizers. That is, the error-correction procedure can be divided into two subsystems, and the relevant functions are

$$\begin{aligned} \text{synd}_Z : \mathcal{P}_Z &\longrightarrow \mathbf{Z}_2^{g_1}, & \text{corr}_Z : \mathbf{Z}_2^{g_1} &\longrightarrow \mathcal{P}_Z, \\ \text{synd}_X : \mathcal{P}_X &\longrightarrow \mathbf{Z}_2^{g_2}, & \text{corr}_X : \mathbf{Z}_2^{g_2} &\longrightarrow \mathcal{P}_X. \end{aligned} \quad (20)$$

Then for any  $Z$ -type encoded operator  $N \in \mathcal{N} \cap \mathcal{P}_Z$  the corresponding dressed operator takes the form

$$N_{\text{dr}} := \sum_{\mathbf{b} \in \mathbb{Z}_2^{g_2}} \text{corr}_X(\mathbf{b})^\dagger N \text{corr}_X(\mathbf{b}) P_{\mathbf{b}}^Z, \quad (21)$$

where  $P_{\mathbf{b}}^Z$  is the projector onto the subspace with  $Z$ -syndrome  $\mathbf{b}$ , that is,

$$P_{\mathbf{b}}^Z := 2^{-g_2} \prod_{i=1}^{g_2} (1 + (-1)^{b_i} s_i^Z). \quad (22)$$

The dressed  $X$ -type Pauli observables are analogous.

An important consequence of the form of the dressed observables (11) is that transversal measurements are possible. Namely, if we measure all physical qubits in the  $X$  basis ( $Z$  basis), we can recover the value of all encoded  $X$ -type ( $Z$ -type) operators because error correction commutes with measurements. Indeed, after the measurement, correction is entirely classical.

Finally, an analogous transversal initialization is possible. If we initialize a product state of the form  $|0\rangle^{\otimes n} (|+\rangle^{\otimes n})$  and measure all  $X$ -type ( $Z$ -type) stabilizer operators, the resulting state has well-defined  $Z$ -type ( $X$ -type) encoded operators. Namely,  $\bar{Z}_i = 1$  (respectively,  $\bar{X}_i = 1$ ). The state may have many  $X$ -type ( $Z$ -type) stabilizers with the wrong sign, but indeed the correction of  $Z$ -type ( $X$ -type) errors is immaterial because such errors cannot change the encoded state. Of course this is the case when the environment is absent. We shall discuss initialization in the presence of noise in section 8.1.

## 5. Modeling interaction with the environment

In this section, we shall describe the evolution of an open system in the weak coupling approximation. Next, we shall recall the relationship between the fidelity criterion and the decay rate of observables, and a useful upper bound for the latter [16] (in this context see also [38]).

### 5.1. Evolution in the weak coupling limit

We consider a quantum system with a discrete energy spectrum that is coupled to a collection of heat baths leading to the global Hamiltonian

$$H = H^{\text{sys}} + H^{\text{bath}} + H^{\text{int}} \quad \text{with} \quad H^{\text{int}} = \sum_{\alpha} S_{\alpha} \otimes f_{\alpha}, \quad (23)$$

where  $S_{\alpha}$  are system operators and  $f_{\alpha}$  are bath operators.

We assume that the baths are independent, which excludes such phenomena as decoherence free subspaces (see e.g. [31, 39, 40]). The evolution of the system in the Heisenberg picture in the weak coupling limit is the following:

$$\frac{dX}{dt} = i[H^{\text{sys}}, X] + \mathcal{L}_{\text{dis}}(X) =: \mathcal{L}(X), \quad (24)$$

$$\mathcal{L}_{\text{dis}}(X) = \frac{1}{2} \sum_{\alpha} \sum_{\omega} \hat{h}_{\alpha}(\omega) (S_{\alpha}^{\dagger}(\omega) [X, S_{\alpha}(\omega)] \quad (25)$$

$$+ [S_{\alpha}^{\dagger}(\omega), X] S_{\alpha}(\omega)). \quad (26)$$

Here  $S_\alpha(\omega)$  are Fourier transforms of  $S_\alpha$

$$e^{iH_{\text{sys}}t} S_\alpha e^{-iH_{\text{sys}}t} = \sum_{\omega} S_\alpha(\omega) e^{-i\omega t}. \quad (27)$$

The coefficients  $\hat{h}_\alpha$  are Fourier transforms of the autocorrelation functions of  $f_\alpha$ . They describe the rate at which the coupling is able to transfer an energy  $\omega$  from the bath to the system. In the case of thermal baths, they satisfy

$$\hat{h}_\alpha(-\omega) = e^{-\beta\omega} \hat{h}_\alpha(\omega), \quad (28)$$

which is a consequence of the Kubo–Martin–Schwinger condition [41]. The dissipative part commutes with the Hamiltonian part, i.e.  $[\hat{H}, \mathcal{L}] = 0$  (where  $\hat{H}(X) = i[H^{\text{sys}}, X]$ ). Therefore, in the analysis of thermal stability, it is enough to consider solely the dissipative part. An important property of the dissipative generator is that  $-\mathcal{L}$  is positive definite in the scalar product

$$\langle X, Y \rangle_\beta := \text{Tr}(\rho_\beta X^\dagger Y). \quad (29)$$

Moreover, we have  $\mathcal{L}(\mathbf{1}) = 0$ , and if the commutant of the system operators  $S_\alpha$  and  $H^{\text{sys}}$  is trivial, then the eigenvalue 0 is nondegenerate. This means that there are no constants of motion—all other observables eventually decay to identity.

### 5.2. Fidelity and autocorrelation functions

Since the generator is Hermitian in scalar product (29), a strong tool is spectral analysis of  $\mathcal{L}$ . Namely, for any observable satisfying  $\langle A, A \rangle_\beta = 1$  and  $\langle A, I \rangle_\beta = 0$ , one easily finds that

$$\langle A, A(t) \rangle_\beta \geq e^{-\epsilon t} \quad (30)$$

provided that

$$-\langle A, \mathcal{L}(A) \rangle_\beta \leq \epsilon. \quad (31)$$

Thus, to show that an observable  $A$  is stable, it is enough to estimate  $-\langle A, \mathcal{L}(A) \rangle$ , which can be therefore called the *decay rate* for the observable  $A$ . If this quantity decreases exponentially with the size of the system, we obtain stability.

However, the criterion of stability of quantum information is measured in the most direct way by fidelity. However, in some natural situations, the fidelity criterion is equivalent to the stability of two complementary observables in terms of the autocorrelation function. The needed relation was obtained in [16]. Let us recall it here, in a slightly more general setting.

As discussed in section 3, we divide the total system into two:  $\mathcal{H}_V \otimes \mathcal{H}_{\text{Anc}}$ , where the first system is the qubit to be protected. We then consider an initial state of the form  $|\psi_V\rangle\langle\psi_V| \otimes \rho_{\text{anc}}$ . The state is subjected to some completely positive trace preserving map  $\Lambda_{V,\text{anc}}^*$ , and we are interested in the fidelity between the initial and final states of the system  $V$ :

$$F(\psi_V; \Lambda_{V,\text{anc}}^*, \rho_{\text{anc}}) = \langle \psi_V | \rho_V^{\text{out}} | \psi_V \rangle, \quad (32)$$

where  $\rho_V^{\text{out}} = \text{Tr}_{\text{anc}}(\Lambda_{V,\text{anc}}^*(|\psi_V\rangle\langle\psi_V| \otimes \rho_{\text{anc}}))$ . In appendix A, we prove the following proposition:

**Proposition 1.** *Consider an arbitrary dichotomic observable  $A_{V,\text{anc}} = A_V \otimes I_{\text{anc}}$  acting on a Hilbert space  $\mathcal{H}_V \otimes \mathcal{H}_{\text{anc}}$  (where  $\mathcal{H}_V = C^2$ ) with  $\psi_\pm$  being eigenvectors of  $A_V$  with*

eigenvalues  $\pm 1$ . Then, for an arbitrary state  $\rho_{\text{anc}}$  and a completely positive trace preserving map  $\Lambda_{V,\text{anc}}^*$ , we have

$$F(\psi_+; \Lambda_{V,\text{anc}}^*, \rho_{\text{anc}}) + F(\psi_-; \Lambda_{V,\text{anc}}^*, \rho_{\text{anc}}) = 1 + \langle A, \Lambda(A) \rangle_\eta, \quad (33)$$

where  $\eta = \frac{1}{2} I_V \otimes \rho_{\text{anc}}$ ,

$$\langle A, B \rangle_\eta = \text{Tr}(\eta A^\dagger B), \quad (34)$$

and the map  $\Lambda$  is dual to  $\Lambda^*$ .

We shall also need the following lemma [42]:

**Lemma 1.** For an arbitrary trace preserving map  $\Gamma$  on a qubit, define  $F(\psi) = \langle \psi | \Gamma(|\psi\rangle\langle\psi|) | \psi \rangle$ , as well as the entanglement fidelity (that quantifies the quality of preservation of a quantum memory)

$$F_e(\Gamma) = \langle \phi_+ | (\Gamma \otimes I) (|\phi_+\rangle\langle\phi_+|) | \phi_+ \rangle, \quad (35)$$

where  $|\phi_+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ . Then  $F_e$  is bounded as follows:

$$F_e(\Gamma) \geq F_{\sigma_x} + F_{\sigma_z} - 1, \quad (36)$$

where  $F_{\sigma_x} = \frac{1}{2}(F(|+\rangle) + F(|-\rangle))$ ,  $F_{\sigma_z} = \frac{1}{2}(F(|0\rangle) + F(|1\rangle))$ , with  $|\pm\rangle$  being eigenvectors of  $\sigma_x$ ,  $|0\rangle, |1\rangle$  being eigenvectors of  $\sigma_z$ .

We shall now apply these facts to the present case. Let  $X$  and  $Z$  be two Pauli observables acting on the memory, i.e. they are of the form  $X_V \otimes \mathbf{1}_{\text{Anc}}$  and  $Z_V \otimes \mathbf{1}_{\text{Anc}}$ . As the state  $\eta$  in the proposition we take the Gibbs state (which is indeed proportional to the identity on the system  $V$  due to the form of the Hamiltonian). Finally, the map  $\Lambda$  is given by  $e^{\mathcal{L}t}$ . Applying the proposition and the lemma, we obtain

$$F_e \geq \frac{1}{2} (\langle X, e^{\mathcal{L}t} X \rangle_\beta + \langle Z, e^{\mathcal{L}t} Z \rangle_\beta) \geq e^{-\epsilon t}, \quad (37)$$

where  $\epsilon$  is an upper bound for the rates  $-\langle X, \mathcal{L}(X) \rangle_\beta$  and  $-\langle Z, \mathcal{L}(Z) \rangle_\beta$ . Therefore, to prove the stability of a quantum memory it is enough to bound the above decay rates. We have actually proved a more general result, namely that the above inequality holds for whatever physical operation (identity preserving completely positive map  $\Lambda$ ) unrelated to the Hamiltonian of the system, i.e. we have

$$F_e \geq \frac{1}{2} (\langle X, \Lambda(X) \rangle_\beta + \langle Z, \Lambda(Z) \rangle_\beta). \quad (38)$$

### 5.3. Upper bound for the decay rate

In [16], the following useful bound for the decay rate was obtained:

$$-\langle A, \mathcal{L}(A) \rangle_\beta \leq 2\hat{h}_{\max} \sum_{\alpha} \langle [S_{\alpha}, A], [S_{\alpha}, A] \rangle_\beta, \quad (39)$$

where

$$\hat{h}_{\max} = \sup_{\alpha, \omega \geq 0} \hat{h}_{\alpha}(\omega). \quad (40)$$

The advantage of this estimate is that it refers directly to the system operators  $S_{\alpha}$  rather than to their Fourier transforms.

## 6. Conditions for self-correcting quantum memories

### 6.1. Stabilizer code models

Given a stabilizer code, we may consider a Hamiltonian model of the form

$$H = - \sum_{i=1}^r t_i s'_i, \quad (41)$$

where  $t_i > 0$  are coupling constants and  $\{s'_i\}$  is a possibly overcomplete set of  $r$  generators of the stabilizer  $\mathcal{S}$ . The encoded Pauli operators  $\bar{X}_1, \bar{Z}_1, \dots, \bar{X}_k, \bar{Z}_k$  commute with  $H$ , showing that each energy level is at least  $2^k$ -fold degenerate. For a particular choice of the encoded operators, one may split the Hilbert space

$$\mathcal{H} = \mathcal{H}_{\text{enc}} \otimes \mathcal{H}_{\text{exc}} = \mathbf{C}^{2^k} \otimes \mathbf{C}^{2^g} \quad (42)$$

in such a way that  $\mathcal{H}_{\text{enc}}$  represents the encoded degrees of freedom and  $\mathcal{H}_{\text{exc}}$  characterizes excitations. Let us emphasize that the above division is not yet the one we considered in section 3. Indeed, the subsystem  $\mathcal{H}_{\text{enc}}$  is defined by the bare observables, and it is not protected. In particular, we may choose a basis  $|\mathbf{b}\rangle$  in  $\mathcal{H}_{\text{exc}}$  with the elements labeled by  $\mathbf{b} \in \mathbf{Z}_2^g$  so that setting  $|\psi, \mathbf{b}\rangle := |\psi\rangle \otimes |\mathbf{b}\rangle$  for any  $|\psi\rangle \in \mathcal{H}_{\text{enc}}$  we have

$$s_i |\psi, \mathbf{b}\rangle = (-1)^{b_i} |\psi, \mathbf{b}\rangle. \quad (43)$$

Recall that  $s_i$  are generators from the complete set (5). Then the states of the form  $|\psi, \mathbf{b}\rangle$  are eigenvectors of the Hamiltonian (41), namely

$$H |\psi, \mathbf{b}\rangle = - \sum_{i=1}^r t_i^{c_i(\mathbf{b})} |\psi, \mathbf{b}\rangle =: E_{\mathbf{b}} |\psi, \mathbf{b}\rangle, \quad (44)$$

where  $c_i(\mathbf{b}) = 0, 1$  are obtained from

$$s'_i = \prod_{j=1}^g s_j^{c_{ij}}, \quad c_{ij} = 0, 1, \quad (45)$$

as  $c_i(\mathbf{b}) = \sum_j c_{ij} b_j$ , with addition modulo 2. Thus, the Gibbs state for the Hamiltonian (41) can be written as

$$\rho(\beta) \propto \sum_{\mathbf{b}} e^{-\beta E_{\mathbf{b}}} P_{\mathbf{b}}. \quad (46)$$

### 6.2. A bound in terms of critical syndromes

In this section, we shall show that the decay rate for a dressed observable  $N_{\text{dr}}$  is bounded by the probability of a critical syndrome multiplied by some polynomial in the size of the system. Thus, if the probability of the critical syndrome is exponentially small in the size of the system, the observable is stable.

Suppose that we have a Hamiltonian model (41) that interacts with a bath through a coupling of the form

$$H_{\text{int}} = \sum_i X_i \otimes f_i + \sum_i Y_i \otimes g_i + \sum_i Z_i \otimes h_i. \quad (47)$$



For simplicity, we will write

$$H_{\text{int}} = \sum_{\sigma} \sigma \otimes f_{\sigma} \quad (48)$$

with  $\sigma$  running over the three types of Pauli operators and over all sites. We will suppose that our model enjoys the following two properties. Firstly, let the  $t_i$  couplings satisfy

$$0 < t_{\min} \leq t_i \leq t_{\max}. \quad (49)$$

Secondly, the number of stabilizers  $s'_i$  (from (41)) that contain in their support a given qubit is bounded. When these conditions hold, it follows that the Bohr frequencies  $\omega$  are bounded; hence, assuming that the spectral density  $h(\omega)$  is finite for all  $\omega$ , we obtain that  $\hat{h}_{\max}$  is finite and (39) applies.

Let us go back to the dressed operators introduced in section 4.3. It is our aim to check how stable they are under thermal fluctuations. From (39) and (16) we have that for any encoded operator  $N$  the decay rate of  $N_{\text{dr}}$  is bounded as follows:

$$-\langle N_{\text{dr}}, \mathcal{L}(N_{\text{dr}}) \rangle_{\beta} \leq 8\hat{h}_{\max} \frac{\sum_{\mathbf{b}} e^{-\beta E_{\mathbf{b}}} \sum_{\sigma} (1 - S(N, \sigma, \mathbf{b}))}{\sum_{\mathbf{b}} e^{-\beta E_{\mathbf{b}}}} \leq 16\hat{h}_{\max} n P_{\text{crit}_N}. \quad (50)$$

Here  $n$  is the total number of sites, which we will suppose to grow polynomially with the size of the system  $L$ , and  $P_{\text{crit}_N}$  is the total probability at a given temperature of finding a syndrome that is critical for  $N$ , that is,

$$P_{\text{crit}_N} = \frac{\sum_{\mathbf{b} \in \text{crit}_N} e^{-\beta E_{\mathbf{b}}}}{\sum_{\mathbf{b}} e^{-\beta E_{\mathbf{b}}}}. \quad (51)$$

When, for a given temperature,  $P_{\text{crit}_N}$  is exponentially small on the size of the system  $L$ , we say that  $N_{\text{dr}}$  is stable: it is self-protected by a free energy barrier produced by the Hamiltonian (41). In the following section, we will derive for a wide class of systems a set of sufficient conditions for this stability to happen at some finite temperature.

### 6.3. Sketch of the proof of stability of the Kitaev model

Before we present conditions ensuring stability, let us first recall the essential features of the proof of stability of the Kitaev model [16]. We shall restrict ourselves to a classical submodel of the Kitaev model, where errors are only bit flips, and there is therefore only one type of excitation. We consider configurations of spins, i.e. patterns of spins-up or spins-down (or in bit notation, this is a pattern of zeros and ones). These spins reside on the plaquettes of some cubic lattice, while excitations reside on the edges. Any such configuration leads to some pattern of excitations, which are closed loops of edges. The energy associated with a given loop (excitation) is proportional to its length.

The proof contains two main steps.

*Step 1.* Show that, with high probability, only configurations leading to small loops appear in the Gibbs state.

This is obtained as follows. Consider a fixed loop, and assume that it is homologically trivial. Then by flipping some spins, we can erase it from the configuration (without affecting any other loop). This means that we have a new spin configuration, which differs from the old one *only* by the loop that has been removed. The energy of the new configuration differs from the old configuration only by the length of the erased loop. This feature, i.e. possibility

of *individually* erasing loops, and the resulting energy difference between the respective spin configurations, allows us to prove that the probability that a given loop will appear is bounded by  $e^{-\beta l}$ , where  $l$  is the size of the loop. This is described in [16], following [44]. Then the probability that arbitrary loops of length  $l$  appear is bounded by

$$p(l) \leq \#(\text{loops}) \times e^{-\beta l}, \quad (52)$$

where  $\#(\text{loops})$  is the number of all loops of length  $l$ . So, we have to count all possible loops of length  $l$ . We first bound the number of loops containing a given spin. Since loops are connected objects, it turns out that this number is at most exponential in  $l$ , i.e. is  $\leq e^{\mu l}$  where  $\mu$  is a constant dependent only on the type of lattice involved. So the entropic factor  $\#(\text{loops})$  is obtained by multiplying this number by the volume  $V$ —i.e. the number of spins in the lattice, which is polynomial in  $L$ —the linear size of the model.

Now this implies that the probability of appearance of loops of size greater than  $cL$  (for some constant  $c$ ) decays exponentially with the system size  $L$  for sufficiently low temperatures, as can be readily seen by summing up the probabilities  $p(l)$  from  $l = cL$  upwards. Thus with high probability spin configurations will contain only loops shorter than  $cL$ .

Apart from trivial loops, typical spin configurations may also contain nontrivial loops. These cannot be individually removed by flipping spins, so the above argument and conclusion is not directly applicable to these excitations (the discussion of this case was actually missing in [16]). However, the following device allows us to generalize the result for trivial loops to the case of nontrivial loops. For the purpose of applying the *erasability* feature, note that *groups* of nontrivial loops can be erased by flipping spins. In the case of the 3D Kitaev model, loops may be grouped (nonuniquely) into subsets of at most four loops (the number  $4 = 3 + 1$  comes from the  $\mathbb{Z}_2$  first homology group of the 3D torus, which can be regarded as a vector space over  $\mathbb{Z}_2$  of dimension 3), such that each such subset can be erased as a whole by flipping spins.

Now, we can rerun the previous argument. Fixing some arbitrary *erasable* set of not more than four loops, the probability of it appearing is then again bounded by  $e^{-\beta l}$ , where  $l$  now is the size of the whole set. The difference now is that the subset is not connected, which influences the counting of possible configurations giving rise to a set of size  $l$ . However, a loose upper bound that is sufficient for our purposes can be obtained by observing that the set can have at most four components, each of which is a loop, i.e. a connected object. Indeed, the number of such sets is trivially bounded from above by the number of ways four loops of lengths  $l_1, l_2, l_3, l_4$  can be chosen to form a subset of length  $l = l_1 + l_2 + l_3 + l_4$ , i.e.  $e^{\mu l} V^4 \text{poly}(l)$  with  $\text{poly}(l)$  the number of ways in which the individual lengths can be chosen to sum up to  $l$ . Again summing probabilities, one obtains that configurations that can be decomposed into erasable subsets of length  $l \geq cL$ , for some constant  $c$ , appear with negligible probability for sufficiently low temperatures.

*Step 2.* Show that configurations with short loops only are not critical, i.e. a single spin flip will not change the sign of the dressed observable.

This proof requires choosing a good error-correcting algorithm (as the latter defines the dressed observable). Note that, due to the previous discussion, only the correction of ‘short’ loops is relevant, e.g. of size smaller than, say,  $L/2$ . But, all of these are homologically trivial and can be individually erased by flipping qubits living in the smallest parallelogram containing the relevant loop, which is unequivocally defined. The function  $\text{corr}$  should be chosen accordingly.

Consider a configuration with short loops, such that each individual loop has a size not greater than, say,  $L/4$ . A single flip affects only a few edges (four in the 3D Kitaev model) and

thus can only affect a very small number of loops (at most two in the 3D Kitaev model) in a given configuration. In the 3D Kitaev model the flip can lead to the following changes:

- (i) ‘single loop  $\leftrightarrow$  single loop’ transition: alter the shape of a single loop along the flipped plaquette;
- (ii) ‘two loops  $\leftrightarrow$  one loop’ transition: divide the loop into two smaller loops or join two loops into a single one;
- (iii) ‘no loop  $\leftrightarrow$  single loop’ transitions: creation of a new single plaquette loop or annihilating such an existing small loop. Thus, if we take the product of the single qubit bit-flip and the correcting operators before and after the change, the resulting product operator will have support on a cube of length  $L/2 + 1$ . Hence, it is a stabilizer and the configuration is not critical.

#### 6.4. A criterion based on connectedness

Here, we present a set of conditions that ensure that the probability of a critical syndrome is exponentially small in the size of the system. These conditions require the excitations to be connected in a suitable way forming clusters, and also that critical configurations always contain clusters of size comparable with the system size. This is essentially an application of Peierls’ argument [43–45] as first done in [13] in the context of quantum computation. These conditions are obtained by extracting the essential properties of the Kitaev 4D model (or the 3D model for one kind of errors) used in the proof of its stability as recalled in the previous section.

For a given Hamiltonian  $H$ , we shall consider an error graph  $\Gamma$  whose nodes represent the set  $\{s'_i\}$  of generators of the stabilizer  $\mathcal{S}$  appearing in  $H$  (41). The nodes represent excitations. Then any syndrome  $\mathbf{b} \in \mathbf{Z}_2^g$  corresponds to a subset  $\text{nds}(\mathbf{b})$ . Namely, the  $i$ th node belongs to the set  $\text{nds}(\mathbf{b})$  if  $c_i(\mathbf{b}) = 1$  (see equation (45)). To be more explicit, an error causes a pattern of excitations (it changes the sign of some generators from the Hamiltonian) that depends on the error only through the syndrome  $\mathbf{b}$ , so that we can unambiguously write  $\text{nds}(\mathbf{b})$ . While  $\mathbf{b}$  tells us which independent generators  $s_i$  were affected by an error,  $\text{nds}(\mathbf{b})$  tells us which Hamiltonian terms  $s'_i$  were affected, which usually are not independent anymore. The structure of links or edges in the graph will depend on the particular model. It provides a notion of connectedness: a set of nodes is connected if between any two of its nodes there exists a path such that all its nodes belong to the set. We say that  $\mathbf{b}$  and  $\mathbf{b}'$  contact if  $\text{nds}(\mathbf{b})$  and  $\text{nds}(\mathbf{b}')$  are not disjoint or if any of the elements of  $\text{nds}(\mathbf{b})$  is linked to an element of  $\text{nds}(\mathbf{b}')$ . Also, given a Pauli operator  $E$  we define  $\text{nds}(E) := \text{nds}(\text{synd}(E))$ .

For example, in the 3D Kitaev model the qubits live in plaquettes and  $X$ -type excitations are associated with links. Such excitations are linked in the graph if the corresponding edges share a site of the original lattice. Connected sets take the form of closed paths.

Rather than a single Hamiltonian model (41), consider a family of models  $H_L$  with increasing system size  $L$ , each with a corresponding graph  $\Gamma_L$ . The number of nodes  $|\Gamma_L|$  is assumed polynomial in  $L$ . We show that the following conditions, if satisfied by the graphs  $\Gamma_L$ , imply the stability of the encoded information. Here  $\lambda, \mu, \nu$  are integers and  $\xi$  a positive number.

- (i) For any single-qubit Pauli  $\sigma$ , we have  $|\text{nds}(\sigma)| \leq \lambda$ .
- (ii) The number of links meeting at a node is at most  $\mu$ .

(iii) For every  $\mathbf{b} \in \mathbf{Z}_2^g$ , there exists a decomposition

$$\mathbf{b} = \sum_i \mathbf{b}_i, \quad (53)$$

such that the  $\mathbf{b}_i$ -s do not contact and have at most  $\nu$  connected components each.

(iv) If the total length of a syndrome satisfies  $|\text{nds}(\mathbf{b})| < \xi L$ , for some number  $\xi$ , then the syndrome  $\mathbf{b}$  is not critical for any encoded operator  $N$ , i.e.  $\mathbf{b} \notin \text{crit}_N$ . Furthermore, for such  $\mathbf{b}$ , given any  $\mathbf{b}'$  not contacting  $\mathbf{b}$

$$\text{corr}(\mathbf{b} + \mathbf{b}') = \text{corr}(\mathbf{b}) \text{corr}(\mathbf{b}'). \quad (54)$$

Under these conditions one can characterize the type of syndromes that are critical w.r.t. encoded information. Consider a general syndrome that can be decomposed into components (53) of various lengths. We show that such a syndrome can be critical only if it contains a large component  $\mathbf{b}_0$  with  $|\text{nds}(\mathbf{b}_0)| \geq (\xi L - \lambda)/\lambda\mu$ . Suppose, to the contrary, that  $\mathbf{b} = \sum_{i=1}^r \mathbf{b}_i$  consists of small components, i.e.  $|\text{nds}(\mathbf{b}_i)| < (\xi L - \lambda)/\lambda\mu$ . If  $\sigma$  is a single-qubit Pauli operator, then (i) and (ii) imply that at most  $\lambda\mu$  of the  $\mathbf{b}_i$ -s make contact with  $\text{nds}(\sigma)$ , say  $\mathbf{b}_1, \dots, \mathbf{b}_k$ . Let us denote  $\tilde{\mathbf{b}} = \sum_{i=1}^k \mathbf{b}_i$ . Clearly, we have  $|\tilde{\mathbf{b}}| \leq \xi L - \lambda$  and  $|\tilde{\mathbf{b}} + \text{synd}(\sigma)| \leq \xi L$ ; hence using (iv) we obtain

$$\begin{aligned} \text{corr}(\mathbf{b}) &= \text{corr}(\tilde{\mathbf{b}}) \prod_{i=k+1}^r \text{corr}(\mathbf{b}_i), \\ \text{corr}(\mathbf{b} + \text{synd}(\sigma)) &= \text{corr}(\tilde{\mathbf{b}} + \text{synd}(\sigma)) \prod_{i=k+1}^r \text{corr}(\mathbf{b}_i). \end{aligned} \quad (55)$$

It follows that

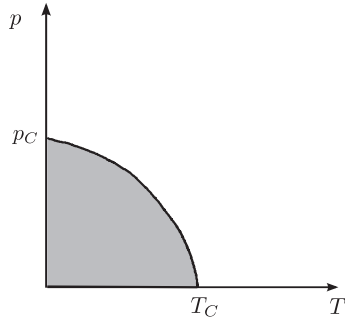
$$\text{corr}(\mathbf{b})\sigma \text{corr}(\mathbf{b} + \text{synd}(\sigma)) = \text{corr}(\tilde{\mathbf{b}})\sigma \text{corr}(\tilde{\mathbf{b}} + \text{synd}(\sigma)) \quad (56)$$

and hence due to (15) that  $S(N, \sigma, \mathbf{b}) = S(N, \sigma, \tilde{\mathbf{b}})$ , i.e.  $\mathbf{b}$  is critical with respect to  $\sigma$  if and only if  $\tilde{\mathbf{b}}$  is critical with respect to  $\sigma$ . However, the latter satisfies  $|\tilde{\mathbf{b}}| < \xi L$ ; hence by (iv) it is not critical, so that  $\mathbf{b} \notin \text{crit}_N$ . In particular, this result means that an error is not critical even if it contains a large number of small error components. This result corresponds to step 2 in the example of the 3D Kitaev model above.

We now aim to bound the probability of a critical syndrome  $P_{\text{crit}_N}$  (much like in step 1 in the 3D Kitaev model). For this purpose, we first bound the probability  $P_{\mathbf{b}_0}$  to find an excitation configuration  $\text{nds}(\mathbf{b})$  such that  $\mathbf{b}_0$  is one of the components in the decomposition (53). Denote the set of all configurations containing  $\mathbf{b}_0$  by  $\{\mathbf{b}'_k\}$ . We now apply Peierls' argument to show that the probability  $P_{\mathbf{b}_0}$  is suppressed by the Boltzman factor at temperature  $1/\beta$ . Indeed, erasing the component  $\mathbf{b}_0$  leads to configurations  $\{\mathbf{b}'_k + \mathbf{b}_0\}$  whose energies are related to the energies of the original configurations as  $E_{\mathbf{b}'_k + \mathbf{b}_0} + E_{\mathbf{b}_0} = E_{\mathbf{b}'_k}$ . Thus, we obtain

$$P_{\mathbf{b}_0} = \frac{\sum_k e^{-\beta E_{\mathbf{b}'_k}}}{\sum_{\mathbf{b} \in \mathbf{Z}_2^g} e^{-\beta E_{\mathbf{b}}}} = \frac{e^{-\beta E_{\mathbf{b}_0}} \sum_k e^{-\beta E_{\mathbf{b}'_k + \mathbf{b}_0}}}{\sum_{\mathbf{b} \in \mathbf{Z}_2^g} e^{-\beta E_{\mathbf{b}}}} \leq e^{-\beta E_{\mathbf{b}_0}} \leq e^{-\beta t_{\min} |\text{nds}(\mathbf{b}_0)|}. \quad (57)$$

The fact that the energy corresponding to a syndrome depends only on the length of the syndrome follows from the form of the Hamiltonian (41). In order to obtain  $P_{\text{crit}_N}$ , in accordance with the previous paragraph, one needs the probability of appearance of any error component of



**Figure 1.** The shaded area describes schematically the protected region.  $T_C$  is the critical temperature in the absence of noise, while  $p_C$  is noise threshold for zero temperature.

a large enough length. The number  $N(l)$  of connected collections of nodes of size  $l$  that contain a given node is bounded as follows for some  $\tau$  (see appendix B):

$$N(l) \leq e^{\tau l}. \quad (58)$$

Thus, the total number of collections of  $l$  nodes with at most  $\nu$  connected components is bounded by  $|\Gamma_L|^\nu l^{\nu-1} e^{\tau l} \leq |\Gamma_L|^{2\nu} e^{\tau l}$ . Putting together these results we obtain

$$P_{\text{crit}_\nu} \leq |\Gamma_L|^{2\nu} \sum_{l \geq \xi L / \lambda \mu} e^{-\delta l} = \text{poly}(L) e^{-\delta \xi L / \lambda \mu} \frac{1}{1 - e^{-\delta}} \quad (59)$$

with  $\delta = \beta t_{\min} - \tau$  positive below a critical temperature. This shows the stability of  $N_{\text{dr}}$  under the given assumptions, as the critical probability decreases exponentially fast with the system size.

### 6.5. Resilience to noise

It is natural to expect that a self-protecting quantum memory will also be resilient, to some extent, to suddenly applied noise. This is interesting since it allows us to expose the memory to errors, which will be an unavoidable byproduct of performing encoded gates. As we shall show, as long as the amount of noise is below a threshold, we are safe. This threshold will reduce as we increase the temperature until it vanishes at the critical temperature, as schematically depicted in figure 1. For zero temperature, this is just the standard error threshold of the error-correcting code.

To fix ideas, let us consider a process  $\Lambda_p$  in which each qubit undergoes a depolarizing channel characterized by an ‘error’ probability  $p$ . That is, any given operator  $N$  is mapped in the Heisenberg picture to

$$\Lambda_p(N) = \sum_E (1-p)^{n-|E|} p^{|E|} E^\dagger N E, \quad (60)$$

where the sum runs over a set of representatives of  $\mathcal{P}/iI$ . We are under the error threshold when in the  $L \rightarrow \infty$  limit for any given encoded operator  $N$  we have

$$\langle N_{\text{dr}}, \Lambda_p(N_{\text{dr}}) \rangle_\beta \longrightarrow 1. \quad (61)$$

However, setting  $\eta := \log((1-p)/p)$ ,

$$\langle N_{\text{dr}}, \Lambda_p(N_{\text{dr}}) \rangle_\beta = \frac{\sum_{E, \mathbf{b}} e^{-\beta E_{\mathbf{b}} - \eta |E|} S(N_{\text{dr}}, E, \mathbf{b})}{\sum_{E, \mathbf{b}} e^{-\beta E_{\mathbf{b}} - \eta |E|}}. \quad (62)$$

That is,  $\langle N_{\text{dr}}, \Lambda_p(N_{\text{dr}}) \rangle_\beta = 1 - 2P_{\text{crit}'_N}$  with

$$P_{\text{crit}'_N} = \frac{\sum_{(E, \mathbf{b}) \in \text{crit}'_N} e^{-\beta E_{\mathbf{b}} - \eta |E|}}{\sum_{E, \mathbf{b}} e^{-\beta E_{\mathbf{b}} - \eta |E|}}. \quad (63)$$

This expression is similar to (51), and we can use a similar reasoning to show that  $P_{\text{crit}'_N} \rightarrow 0$  for low enough temperature and noise. Firstly, we attach to each Pauli operator  $E = \bigotimes_i \sigma_i$  the set

$$\text{nds}'(E) = \bigcup_i \text{nds}(\sigma_i) \subset \Gamma_L,$$

where the  $\sigma_i$  are single-qubit Pauli operators. Note that  $|\text{nds}(E)| \leq |\text{nds}'(E)| \leq \lambda |E|$ . Thus, when  $\lambda |E| + |\text{nds}(\mathbf{b})| < \xi L$ , we have  $(E, \mathbf{b}) \notin \text{crit}'_N$ .

Secondly, we need a substitute of the decomposition (53) to deal with pairs  $(E, \mathbf{b})$ . To this end, note that there exist  $E_j$  such that  $E = \prod_j E_j$  and the sets  $\text{nds}'(E_j)$  form the connected components of  $\text{nds}'(E)$ . Then we arrange the  $\mathbf{b}_i$ -s and  $E_j$ -s in maximal connected collections. That is, if one of these collections is  $\{\mathbf{b}_{a_1}, \dots, \mathbf{b}_{a_r}; E_{a_1}, \dots, E_{a_s}\}$ , then the set  $\text{nds}(\mathbf{b}_{a_1}) \cup \dots \cup \text{nds}(\mathbf{b}_{a_r}) \cup \text{nds}'(E_{a_1}) \cup \dots \cup \text{nds}'(E_{a_s})$  is connected and disconnected from any set  $\text{nds}(\mathbf{b}_i)$  or  $\text{nds}'(E_j)$  that is not part of the collection. We denote each collection by  $(E_a, \mathbf{b}_a)$ , with  $\mathbf{b}_a = \sum_i \mathbf{b}_{a_i}$ ,  $E_a = \sum_i E_{a_i}$ , so that  $\mathbf{b} = \sum_a \mathbf{b}_a$  and  $E = \prod_a E_a$ . This gives the desired decomposition of  $(E, \mathbf{b})$ .

Finally, we observe that when  $\lambda |E_a| + |\text{nds}(\mathbf{b}_a)| < \xi L$  for all pairs  $(E_a, \mathbf{b}_a)$ , we have

$$\begin{aligned} \text{corr}(\mathbf{b}) &= \prod_a \text{corr}(\mathbf{b}_a), \\ \text{corr}(\mathbf{b} + \text{synd}(E)) &= \prod_a \text{corr}(\mathbf{b}_a + \text{synd}(E_a)), \end{aligned} \quad (64)$$

so that  $S(N, E, \mathbf{b}) = \prod_a S(N, E_a, \mathbf{b}_a)$ . We omit the rest of the details.

## 7. Topological stabilizer codes

In this section, we shall construct codes that allow a universal set of transversal gates and, at the same time, give rise to self-correcting local Hamiltonian models. The codes will belong to the important class of local stabilizer codes, in particular, topological codes. These are constructed from manifolds where the physical qubits are placed, and have two main characteristics:

- There exists a set of generators of the stabilizer  $\mathcal{S}$  with a spatially local support.
- The encoded operators in  $\mathcal{N} - \mathcal{S}$  are global, i.e. they have a support which is topologically nontrivial.

We will give below a more precise characterization. The quantum Hamiltonian models (41) obtained from topological codes are topologically ordered: their ground spaces have a degeneracy of topological origin and their excitations are gapped and subject only to topological

interactions. As we will see, excitations can be either particles or higher-dimensional objects. Self-correction will appear in codes with excitations of the latter kind.

The original example of topological stabilizer codes is the toric codes, as introduced by Kitaev *et al* [12, 13]. These codes and their generalizations are based on homology theory [46]. Indeed, the number of encoded qubits in a given manifold is dictated by its Betti numbers, i.e. the dimensions of homology groups, which are topological invariants.

An important drawback of toric codes is that they do not allow the transversal implementation of a sufficiently rich set of gates. There exists, however, a class of topological stabilizer codes that surmount this difficulty: color codes. In 2D [20], they allow the transversal implementation of the so-called Clifford group, the normalizer of the Pauli group in the group of unitary operations on  $n$  qubits. In 3D [21], they allow the transversal implementation of a set of operations that, when measurements are included, gives rise to transversal universal quantum computation. Below we will analyze in detail the transversality properties of general  $D$ -dimensional color codes. In particular, we will exhibit a 6D code that supports transversal CNot gate as well as  $\pi/4$  phase gate. The code will also allow transversal initialization and measurement in  $X$  and  $Z$  basis. Altogether, this enables universal quantum computation [50]. By adding one more dimension we obtain local CNot gates.

### 7.1. Local stabilizer codes

A family of codes  $\{\mathcal{C}_L\}$  is local if there exist integers  $\alpha$  and  $\beta$  such that

- (i) each stabilizer  $\mathcal{S}_L$  has a set of generators  $\{s_1, \dots, s_{g_L}\}$  such that  $|s_j| \leq \alpha$ ,  $j = 1, \dots, g_L$ ,
- (ii) each physical qubit is in the support of at most  $\beta$  such generators  $s_j$  and
- (iii) for any integer  $d$ , there exists a code  $\mathcal{C}_L$  with distance bigger than  $d$ .

Given a code  $\mathcal{C}_L$  in the family, it is always possible to construct a local quantum Hamiltonian model such that its ground state is  $\mathcal{C}_L$ . Namely, we may set

$$H_L = - \sum_{i=1}^{g_L} s_i. \quad (65)$$

Topological stabilizer codes are local codes in which locality has a geometrical meaning. Physical qubits are placed in a lattice in a given manifold, in such a way that the generators of the stabilizer have a support contained in a small neighborhood of the lattice. The support of nontrivial encoded operators, in turn, must be global, in such a way that the distance of the codes grows with the size of the lattice [47].

Below we will study in full generality two important examples of topological stabilizer codes, toric codes and color codes. We will unify the treatment of both types of codes through homology theory.

### 7.2. Homology groups

Recall that a homology group is defined from an exact sequence such as

$$A_2 \xrightarrow{\partial_2} A_1 \xrightarrow{\partial_1} A_0, \quad (66)$$

where  $A_i$  are Abelian groups and  $\partial_i$ , the boundary operators, are homomorphisms with

$$\partial_1 \circ \partial_2 = 0. \quad (67)$$

The elements of the subgroup  $Z_1 := \ker \partial_1 \subset A_1$  are called cycles. The elements of the subgroup  $B_1 := \text{im } \partial_2 \subset A_1$  are called boundaries. Then the condition (67) states that all boundaries are cycles,  $B_1 \subset Z_1$ . Because of this, we can divide cycles in homology classes. Two cycles  $z, z'$  are homologous,  $z \sim z'$ , if  $z - z'$  is a boundary. The homology group  $H$  has as its elements the corresponding homology classes. It is the quotient

$$H := Z_1/B_1. \quad (68)$$

In our case, the elements of the groups  $A_i$  will be chains of suitable geometrical objects from a lattice in a given manifold, and the homology group will be a topological invariant of the manifold, independent of the particular lattice considered. We are only interested in  $\mathbf{Z}_2$  homology, in which the chains in  $A_i$  are binary formal sums of geometrical objects in a set  $S_i$ . That is, if  $a \in A_i$ , then

$$a = \sum_{s \in S_i} b_s s, \quad b_s = 0, 1. \quad (69)$$

When needed, we will regard  $\mathbf{Z}_2$  chains as sets in the usual way. That is, we identify the chain  $a$  above with the set  $\{s \in S_i \mid b_s = 1\}$ .

When working with such  $\mathbf{Z}_2$  homology groups we may consider the dual exact sequence

$$A_0 \xrightarrow{\partial_1^*} A_1 \xrightarrow{\partial_2^*} A_2. \quad (70)$$

The dual boundary operators  $\partial_i^*$  are defined so that for  $a \in S_{i-1}$ ,  $b \in S_i$ , we have  $a \in \partial_i^* b$  if and only if  $b \in \partial_i^* a$ . The dual exact sequence (70) gives rise to a dual homology group  $H^* := Z_1^*/B_1^*$  which is isomorphic to  $H$ . Note that for  $a \in A_1$ ,  $s_2 \in S_2$  and  $s_0 \in S_0$  we have

$$\begin{aligned} s_0 \in \partial_1 a &\iff |\partial_1^* s_0 \cap a| \equiv 1 \pmod{2}, \\ s_2 \in \partial_2^* a &\iff |\partial_2 s_2 \cap a| \equiv 1 \pmod{2}. \end{aligned} \quad (71)$$

Let us illustrate the introduced notions using 2D toric code. Here,  $A_2$  is a set of plaquettes,  $A_1$ —a set of links and  $A_0$  a set of vertices. A boundary is a set of closed loops of links which is contractible, and cycles are arbitrary sets of closed loops. The dual boundary is a set of closed loops in a dual lattice which is contractible (e.g. a dual boundary of a single vertex is the set of links that contain this vertex). Dual cycles are arbitrary sets of closed loops in a dual lattice.

### 7.3. Codes from homology groups

Given a  $\mathbf{Z}_2$  homology group  $H$  obtained from a lattice in a manifold, as those discussed in the previous section, we can introduce in a natural way a CSS-like stabilizer code  $\mathcal{C}$  with a basis labeled by the elements of  $H$ . The procedure can be summarized as follows. Recall that the abelian groups  $A_i$  in the exact sequence (66) are constructed from sets of geometrical objects  $S_i$ . To define the code, we attach physical qubits to the elements of  $S_1$ ,  $X$ -type stabilizers to the elements of  $S_2$  and  $Z$ -type stabilizers to the elements of  $S_0$ . In what follows, we detail the procedure.

First, we attach a physical qubit to each of the elements of  $S_1 = \{r_j\}$ , so that the number of qubits  $n$  is given by  $n = |S_1|$ . For each  $a \in A_1$  with

$$a = \sum_j c_j r_j, \quad c_j = 0, 1, \quad (72)$$



we define an  $X$ -type and a  $Z$ -type Pauli operator setting

$$X_a := \bigotimes_j X_j^{c_j}, \quad Z_a := \bigotimes_j Z_j^{c_j}. \quad (73)$$

Note that the definitions (73) describe two group isomorphisms. Moreover,

$$X_a Z_{a'} = (-1)^{|a \cap a'|} Z_{a'} X_a. \quad (74)$$

Then using (71),

$$a \in Z_1 \iff \forall b \in B_1^*, \quad [X_a, Z_b] = 0, \quad (75)$$

$$a \in Z_1^* \iff \forall b \in B_1, \quad [Z_a, X_b] = 0. \quad (76)$$

As a consequence of (67),  $[X_b, Z_{b'}] = 0$  for  $b \in B_1, b' \in B_1^*$  and we can define the stabilizer group  $\mathcal{S}$  as the subgroup of  $\mathcal{P}$  with elements of the form

$$X_b Z_{b'}, \quad b \in B_1, b' \in B_1^*. \quad (77)$$

It follows then that the operators of the form

$$X_z Z_{z'}, \quad z \in Z_1, z' \in Z_1^*, \quad (78)$$

form, up to phases  $\pm 1, \pm i$ , the normalizer group  $\mathcal{N}$ . If  $\bar{X}_1, \dots, \bar{X}_k \in \mathcal{P}_X$  and  $\bar{Z}_1, \dots, \bar{Z}_k \in \mathcal{P}_Z$  form a Pauli basis for encoded operators, then  $X_i = X_{z_i}$  and  $Z_i = Z_{z'_i}$  for suitable cycles  $z_i \in Z_1, z'_i \in Z_1^*$ . The cycles  $z_i (z'_i)$  are representatives of classes  $\bar{z}_i \in H (\bar{z}'_i \in H^*)$  that form an independent set of generators of  $H (H^*)$ , and thus we have the desired result.

It is instructive to work out explicitly a basis for the resulting code  $\mathcal{C}$ . To this end, for any  $a \in A_1$  we set

$$|a\rangle_0 := X_a |0\rangle^{\otimes n}, \quad |a\rangle_+ := Z_a |+\rangle^{\otimes n}. \quad (79)$$

Then  $\{|a\rangle_0\}$  and  $\{|a\rangle_+\}$  are bases of the Hilbert space, with elements labeled by the chains  $a \in A_1$ . For any cycles  $z \in Z_1, z' \in Z_1^*$ , we define the states

$$|\bar{z}\rangle_0 := \sum_{c \sim z} |c\rangle_0, \quad |\bar{z}'\rangle_+ := \sum_{c \sim z'} |c\rangle_+. \quad (80)$$

Then, as can be easily checked,  $\{|\bar{z}\rangle_0\}$  and  $\{|\bar{z}'\rangle_+\}$  are alternative bases for the topological stabilizer code  $\mathcal{C}$  with elements labeled by homology classes of  $H$  and  $H^*$ , respectively.

**7.3.1. Error correction.** We analyze now error correction in the code  $\mathcal{C}$  just described. First, we need a generating set for the stabilizer group, which can be obtained by attaching an operator to each of the elements of  $\mathcal{S}_2$  and  $\mathcal{S}_0$ . Set for any  $a_2 \in A_2, a_0 \in A_0$ ,

$$X_{a_2} := X_{\partial_2 a_2}, \quad Z_{a_0} := Z_{\partial_1^* a_0}. \quad (81)$$

Then, if  $\mathcal{S}_2 = \{p_1, \dots, p_l\}, \mathcal{S}_0 = \{q_1, \dots, q_m\}$ , we may write

$$\mathcal{S} := \langle X_{p_1}, \dots, X_{p_l}, Z_{q_1}, \dots, Z_{q_m} \rangle. \quad (82)$$

Indeed, the above set may be overcomplete. We can choose in that case suitable subsets  $S'_2 = \{p_1, \dots, p_{g_1}\} \subset S_2$ ,  $S'_0 = \{q_1, \dots, q_{g_2}\} \subset S_0$ . We may naturally regard any  $s_2 \subset S'_2$  as some  $\mathbf{b} \in \mathbf{Z}_2^{g_1}$ , setting  $b_i = 1$  when  $p_i \in s_2$ , and similarly for  $S'_0$  and  $\mathbf{Z}_2^{g_2}$ . Then, for  $a \in A_1$ ,

$$\begin{aligned} \text{synd}_X(Z_a) &= S'_2 \cap \partial_2^* a, \\ \text{synd}_Z(X_a) &= S'_0 \cap \partial_1 a. \end{aligned} \quad (83)$$

That is, for a given Pauli error  $X_a Z_{a'}$  the error syndrome determines the boundaries of  $a$  and  $a'$ . Error correction amounts to choosing an operator  $X_c Z_{c'} = \text{corr}(\text{synd}(X_a Z_{a'}))$  with  $\partial_1 c = \partial_1 a$  and  $\partial_2^* c' = \partial_2^* a'$ . It will succeed if and only if  $a + c \in B_1$  and  $a' + c' \in B_1^*$ , that is, if and only if the correction operator and the error are equivalent up to homology.

*7.3.2. Hamiltonian and excitations.* Given the generators of the stabilizer in (82), we can consider the quantum Hamiltonian model

$$H = - \sum_{i=1}^l X_{p_i} - \sum_{i=1}^m Z_{q_i}. \quad (84)$$

The excitations of this model are gapped, with two units of energy per stabilizer violation. We say that the excitations related to  $X_{p_i}$  ( $Z_{q_i}$ ) generators are  $X$ -type ( $Z$ -type) excitations. Then, as it follows from the previous analysis of error correction,  $X$ -type excitations are arranged in the form of  $B_1^*$  boundaries and  $Z$ -type excitations are arranged in the form of  $B_1$  boundaries.

#### 7.4. Generalized toric codes

Consider a lattice in a  $D$ -manifold. We call the  $n$ -dimensional elements of the lattice  $n$ -cells: 0-cells are vertices, 1-cells are links, 2-cells are faces and so on. For  $n = 0, \dots, D$ , let  $S_n$  be the set of all  $n$ -cells, and  $C_n$  the corresponding abelian group of chains. We in addition set  $C_{D+1} = C_{-1} = 0$ , with 0 the trivial group. For  $n = 1, \dots, D$  we introduce the boundary operators

$$C_n \xrightarrow{\partial_n} C_{n-1} \quad (85)$$

in the usual way. That is, they are homomorphisms such that if the  $(n-1)$ -cells  $s_i \in S_{n-1}$  form the geometrical boundary of the  $n$ -cell  $s$ , then  $\partial s = \{s_i\}$ . We also define the homomorphisms  $\partial_{D+1}$  and  $\partial_0$  in the only possible way.

We can obtain a generalized toric code for each  $d$  with  $d = 0, \dots, D$ . This is the code that corresponds to the exact sequence

$$C_{d+1} \xrightarrow{\partial_{d+1}} C_d \xrightarrow{\partial_d} C_{d-1}, \quad (86)$$

as described in the previous section. The corresponding homology group is the standard  $d$ th homology group of the manifold,

$$H_d := \frac{Z_d}{B_d}. \quad (87)$$

Thus the number of encoded qubits  $k$  is  $h_d$ , the  $d$ th Betti number of the manifold. For example, for  $d = 1$  we have the homology group of curves, so that the number of encoded qubits is equal to the number of independent nonboundary loops in the manifold. Note that the 0th and the  $D$ th

code are just classical repetition codes, with one encoded qubit per connected component of the manifold.

The dual sequence can be visualized more easily in the dual lattice, in which  $r$ -cells become  $(D - r)$ -cells. Indeed, in the dual lattice the dual exact sequence corresponds to the exact sequence of the  $\bar{d}$ th homology group, where

$$\bar{d} := D - d. \quad (88)$$

Thus, in a  $d$ th toric code the relevant homology groups are  $H_{\bar{d}}$  and  $H_d$ . Cycles take the form thus of closed  $\bar{d}$ - and  $d$ -branes, which are mapped to operators using (73). The commutation rules of closed brane operators are purely topological. Namely, given a closed  $d$ -brane  $z \in Z_d$  and a closed  $\bar{d}$ -brane  $z' \in Z_{\bar{d}}$ , the operators  $X_z$  and  $Z_{z'}$  anticommute if  $z$  and  $z'$  intersect an odd number of times. Otherwise they commute.

*7.4.1. Branyons.* For  $d = 1, \dots, D - 1$ , the Hamiltonian model (84) of a  $d$ th toric code contains two kinds of brane-like excitations, which are created by open  $d$  and  $\bar{d}$ -brane operators.  $Z$ -type excitations are closed  $(d - 1)$ -branes and  $X$ -type excitations are closed  $(\bar{d} - 1)$ -branes in the dual lattice. Indeed, both kinds of excitations are not only closed but also boundaries. For example, for  $D = 2$  and  $d = 1$  the excitations are 0-branes, that is, particles. They can only be created or destroyed in pairs since they have to be boundaries of strings. These two kinds of particles interact topologically and are thus called anyons. More generally,  $Z$ -type  $(d - 1)$ -branes and  $X$ -type  $(\bar{d} - 1)$ -branes interact topologically and thus deserve the name of branyons [22].

Since  $(n - 1)$ -branyons appear at the boundaries of  $n$ -brane operators,  $n = d, D - d$ , it follows that such operators represent branyon processes. In particular, processes that do not create or destroy excitations are related to closed branes. Then one can derive from the commutation rules for closed brane operators the topological  $\pm 1$  phases appearing in branyon processes.

*7.4.2. Self-correcting models.* To fix ideas, let us consider a family of hypercubic lattices in a  $D$ -dimensional torus. Each lattice has the same length  $L$  in the  $D$  directions. We will show that, for  $d > 1$ , the models with Hamiltonian (84) for the  $d$ th code protect  $Z$ -type encoded operators in the sense of section 6. Then it follows by duality that for  $d < D - 1$   $X$ -type operators are protected, and thus that for  $1 < d < D - 1$ , we have self-correcting quantum memories, which thus exist for  $D \geq 4$ .

Recall that we have to define a connection scheme for excitations and a correction procedure such that conditions (i)–(iv) of section 6.4 are satisfied. The notion of connectedness of excitations, that is, of  $(d - 1)$ -cells, is naturally given by the lattice. To translate this into the language of section 6.4, we just construct a graph with  $(d - 1)$ -cells as its nodes in which nodes are connected by a link when the corresponding cells meet at a  $(d - 2)$ -cell (note that  $d \geq 2$ ). Conditions (i) and (ii) are clearly satisfied. As for condition (iii), we first note that any connected set of excitations is closed. Some connected sets are boundaries, so that they can be switched off independently and give no further problems (i.e. they constitute  $\text{nds}(E)$  for some Pauli operator  $E$ ). But those which have nontrivial homology cannot be switched off independently. Instead, we have to gather several of them together to get a boundary, but only in a finite amount. Indeed, setting  $v = h_{d-1} + 1$ , condition (iii) is satisfied because  $h_{d-1}$  is the dimension of  $H_{d-1}$  as a  $\mathbf{Z}_2$  vector space.

Finally, we need a correction procedure, that is, the function  $\text{corr}_X$  in (20). We define it in two steps. In the first step, we divide the syndrome  $\mathbf{b}$  in the components  $\mathbf{b}_i$  of (53). In the second step, for each  $\mathbf{b}_i$  we find a suitable collection  $c'$  of  $d$ -cells such that  $\partial c' = c = \text{nds}(\mathbf{b}_i)$ . The choice of the collection of  $d$ -cells  $c'$  must be such that it ensures stability. In fact, we only consider those connected sets of  $(d-1)$ -cells that can be put inside a hypercube of length  $l$  with  $l \leq L$ , and choose any  $c'$  that is contained in the hypercube. This task is always possible because  $c$  has trivial homology, and it can be easily checked that it can be accomplished in polynomial time in  $l$ . We leave the correction of other sets of  $(d-1)$ -cells undefined, because it is not important.

Now that we have a correction procedure, we can check condition (iv). It suffices to show that no  $Z$ -syndrome  $\mathbf{b} \in \mathbf{Z}_2^{g_2}$  that only contains connected sets in  $\text{nds}(\mathbf{b})$  of size at most  $L/2 - 1$  can be critical. To this end, take any  $d$ -cell  $t$  and let us check whether  $\mathbf{b}$  is  $X_t$ -critical. Let  $s_i$  be the connected components of the set of  $(d-1)$ -cells in  $\text{nds}(\mathbf{b}) \cap \partial_1 t$ . Both  $t$  and these sets  $s_i$  are contained in a hypercube of length  $L-1$ , because each  $s_i$  is contained in a hypercube of length  $L/2 - 1$ . Now recall that a syndrome  $\mathbf{b}$  is critical with respect to a given error  $E$  and encoded operator  $N$ , iff the operator (15) anticommutes with  $N$ . Using the notation of (83) we have that the operator of (15)  $\text{corr}_X(\mathbf{b})X_t\text{corr}_X(\mathbf{b} + S'_0 \cap \partial_1 t)$  has support in the hypercube of length  $L-1$  and thus has trivial homology; hence, due to (75) it commutes with any encoded  $Z$ -type operators  $N \subset \mathcal{P}_Z$ . This implies condition (iv) and thus the stability of  $N_{\text{dr}}$  for any given  $N \subset \mathcal{P}_Z$ .

### 7.5. Generalized color codes

Color codes are local codes that are constructed in a special kind of lattices called  $D$ -colexes that live in  $D$ -manifolds. We will first describe color codes without referring to homology and then give a homological picture to fit them in the general framework given above. As in the case of toric codes, several color codes labeled with  $d = 0, \dots, D$  can be obtained from a given lattice, which again will display  $d-1$  dimensional and  $D-d-1$  dimensional excitations. However, unlike in toric codes these excitations can branch and form nets.

**7.5.1.  $D$ -colexes.** A  $D$ -colex  $\Lambda$  is a  $D$ -dimensional lattice with certain colorability properties. Such lattices  $\Lambda$  can be described [22] in terms of their 1-skeleton or graph, which is  $(D+1)$ -valent and has  $D+1$  colorable edges. Here, instead, we choose to define a  $D$ -colex  $\Lambda$  in terms of its dual lattice  $\Delta$ . From now on,  $D$  is an integer,  $D > 0$ .

**Definition 1.** A lattice  $\Lambda$  is a  $D$ -colex if its dual lattice  $\Delta$  is a simplicial lattice with  $(D+1)$ -colored vertices that form a closed  $D$ -manifold.

Recall that an  $n$ -simplex is the convex hull of a set of  $(n+1)$  affinely independent points. In a simplicial  $D$ -dimensional lattice all the  $n$ -dimensional cells are homeomorphic to  $n$ -simplices. A closed manifold is a compact manifold without boundaries. For example, for  $D=1$  it must be a circle, and for  $D=2$  [20] it could be a sphere, a projective plain, a torus, etc. That the vertices of  $\Delta$  are  $(D+1)$ -colored means that we have attached labels to vertices from a set  $Q$  with  $D+1$  elements or colors, in such a way that two vertices connected by a link never share the same color. At first sight it may seem unclear whether a  $D$ -colex can be constructed in arbitrary manifolds, due to the colorability constraints. However, in [22] a construction is given to obtain  $D$ -colexes from arbitrary lattices. In the case of hypercubic lattices, the construction

is particularly simple. The resulting dual lattice  $\Delta$  has vertices also forming a  $D$ -dimensional hypercubic lattice, which we can label by their coordinates  $(x_j) \in \mathbf{Z}^D$ . The  $D$ -simplices are as follows. Let  $\mathbf{e}_k =: (\delta_{kj})$  be the standard basis. For each permutation of  $D$  elements  $\sigma_j$ ,  $\mathbf{z} = (z_j) \in \mathbf{Z}^D$  and  $(d_j) \in \{1, -1\}^D$ , there is a simplex with vertex coordinates

$$\{2z, 2z + d_1 \mathbf{e}_1, 2z + d_1 \mathbf{e}_1 + d_2 \mathbf{e}_2, \dots, 2z + \sum_{i=1}^D d_i \mathbf{e}_i\}.$$

This lattice is  $(D+1)$ -colorable: the coloring  $c \in \{0, \dots, D\}$  of a vertex  $(x_i)$  is given by  $c = \sum_i (x_i \bmod 2)$ .

We will denote the set of all  $n$ -simplices in  $\Delta$  as  $\Delta_n, n = 0, \dots, D$ . The coloring of vertices induces a coloring of  $n$ -simplices. Namely, given a subset of colors  $q \subset Q$  with  $n+1$  elements, we define the subset  $\Delta_q \subset \Delta_n$  of those  $n$ -simplices such that their  $n+1$  vertices are labeled with the elements of  $q$ . A coloring for the cells of the  $D$ -colex  $\Lambda$  is also induced. Let us write  $\lambda^*$  for the dual  $(D-n)$ -simplex of an  $n$ -cell  $\lambda$  and  $\delta^*$  for the dual  $(D-n)$ -cell of an  $n$ -simplex  $\delta$ . Then if  $\delta$  is a  $q$ -cell we say that  $\delta^*$  is a  $\bar{q}$  cell, with  $\bar{q}$  the complement of  $q$  in  $Q$ . As with simplices, the set of all  $n$ -cells of  $\Lambda$  is  $\Lambda_n$  and the set of  $q$ -cells is  $\Lambda_q \subset \Lambda_n, q \subset Q, |q| = n$ . Note, in particular, that vertices are colorless. Indeed, the colorability of vertices in  $\Lambda$  has its own content: vertices are bicolorable—that is, the graph of  $\Lambda$  is bipartite—if and only if the manifold is orientable. In general,  $\Lambda$  may be composed of  $m$  connected components  $\Lambda^\alpha, \alpha = 1, \dots, m$ . We will regard each of these components as a ‘ $(D+1)$ -cell’, setting  $\Lambda_Q := \Lambda_{D+1} := \{\Lambda^\alpha\}$ .

The previous notation for  $\Delta$  and  $\Lambda$  can be extended to individual simplices and cells. Namely,  $\delta_n \subset \Delta_n$  will denote the set of  $n$ -simplices that are part of an  $m$ -simplex  $\delta \in \Delta_m, D \geq m \geq n \geq 0$ . For example, if  $\delta \in \Delta_1$  is a link, then  $\delta_0 = \{\epsilon, \eta\}$  with  $\epsilon, \eta$  the vertices in the ends of  $\delta$ . Similarly,  $\lambda_n \subset \Lambda_n$  denotes the set of  $n$ -cells that are part of an  $m$ -cell  $\lambda \in \Lambda_m, m \geq n$ , and  $\lambda_q \subset \Lambda_q$  denotes the set of  $q$ -cells that are part of a  $q'$ -cell  $\lambda \in \Lambda_{q'}, q \subset q' \subset Q$ .

An important feature of  $D$ -colexes is that all their  $d$ -cells,  $0 < d \leq D$ , are  $(d-1)$ -colexes themselves. This can be easily checked in the dual simplicial lattice  $\Delta$ . To this end, we first choose a vertex  $\delta \in \Delta_0$ . If we keep only those simplices that meet at  $\delta$ , and reduce their dimension in one by erasing the vertex  $\delta$ , it is easy to check that the resulting simplicial lattice  $\Delta'$  is dual to a  $(D-1)$ -colex  $\Lambda'$  with the topology of a  $(D-1)$ -sphere. This is the colex that correspond to the  $D$ -cell  $\delta^*$ . The process can then be continued recursively to get the colexes attached to cells of smaller dimension. Thus, there is a one-to-one correspondence between subcolexes and cells. Each  $d$ -cell  $\lambda \in \Lambda_d$  has as its boundary a  $(d-1)$ -subcolex. If  $\lambda$  is  $q$ -colored,  $q = \{q_i\} \subset Q$ , then its  $(d-1)$ -colex is composed of  $q_i$ -links,  $i = 1, \dots, d$ .

Given a coloring  $r$  and a cell  $\lambda \in \Lambda_q, q \subset r \subset Q$ , there exists a unique  $r$ -cell, denoted as  $\lambda^r \in \Lambda_r$ , such that  $\lambda \in (\lambda^r)_q$ . It is natural to define the intersection of two cells  $\mu \in \Lambda_q, \nu \in \Lambda_r, q, r \subset Q$ , as the collection of disjoint  $(q \cap r)$ -cells  $\mu \cap \nu := \mu_{q \cap r} \cap \nu_{q \cap r}$ . When  $\mu \cap \nu$  is empty or has a single element for any cells  $\mu$  and  $\nu$ , we will say that  $\Lambda$  is nice. In a nice  $D$ -colex, if  $\mu \in \Lambda_q$  and  $\lambda \in \mu_s$  with  $s \subset q \subset Q$ , there exists for any  $r \subset Q$  with  $s = q \cap r$  a unique  $\nu \in \Lambda_r$  with  $\mu \cap \nu = \{\lambda\}$ , namely  $\nu = \lambda^r$ .

**7.5.2. Color codes.** From a  $D$ -colex  $\Lambda$  we can obtain several color codes, labeled with an integer  $d$  such that  $0 \leq d \leq D$ . As we will see, the  $d$ th code is dual to the  $\bar{d}$ th code. The codes with  $d = \bar{d}$  are self-dual, something that turns out to have interesting computational consequences.

To construct the  $d$ th color code from a  $D$ -colex  $\Lambda$ , we first attach a qubit to each vertex  $v \in \Lambda_0$ . As generators of the stabilizer  $\mathcal{S}$  we choose certain  $(d+1)$ -cell and  $(\bar{d}+1)$ -cell operators, as we explain next. For any  $n$ -cell  $\lambda \in \Lambda_n$ ,  $0 \leq n \leq D+1$ , we introduce the  $X$  and  $Z$  cell operators

$$X_\lambda := \prod_{v \in \lambda_0} X_v, \quad Z_\lambda := \prod_{v \in \lambda_0} Z_v. \quad (89)$$

An important property of cell operators is that for any integers  $1 \leq n, m \leq D+1$  with  $n+m > D+1$ , and for any cells  $\mu \in \Lambda_n$ ,  $\nu \in \Lambda_m$  we have [22]

$$[X_\mu, Z_\nu] = 0. \quad (90)$$

Also, for any  $q$ -cell  $\lambda$ ,  $q \subset Q$ , and for any coloring  $q' \subset q$ , we have the relations [22]

$$X_\lambda = \prod_{\mu \in \lambda_{q'}} X_\mu, \quad Z_\lambda = \prod_{\mu \in \lambda_{q'}} Z_\mu. \quad (91)$$

**Definition 2.** The  $d$ th color code of  $\Lambda$ ,  $0 \leq d \leq D$ , is the stabilizer code  $\mathcal{C}$  with stabilizer  $\mathcal{S}$  generated by the operators of the form  $X_\lambda$ ,  $\lambda \in \Lambda_{d+1}$ , and  $Z_{\lambda'}$ ,  $\lambda' \in \Lambda_{\bar{d}+1}$ .

From this definition it is clear that dual codes are related through the exchange of  $X$  and  $Z$  operators, and thus are equal up to a change of basis.

From the relations (91) it follows [22] that the number of encoded qubits is

$$k = \binom{D}{d} h_d, \quad (92)$$

where  $h_d = h_{\bar{d}}$  is the  $d$ th Betti number of the manifold. The origin of this particular value will be clarified in the following sections.

Note that color codes with  $d=0$  or  $d=D$  are not interesting quantum codes, as they are repetition ‘classical’ codes, which only correct either  $X$  or  $Z$  errors and encode a qubit per connected component of  $\Lambda$ . Because of this, below we will mainly focus on color codes with  $0 < d < D$ .

**7.5.3. A simplicial homology.** In this section, we will work only in the dual lattice  $\Delta$ , since it is more suited to explain the underlying homological structure of color codes. We attach to each set of  $n$ -simplices  $\Delta_n$ ,  $0 \leq n \leq D$ , the abelian group  $C_n$  of  $\mathbf{Z}_2$  chains, together with a trivial group  $C_{-1} = \{0\}$ , and to each  $n = 0, \dots, D$  a boundary operator

$$C_D \xrightarrow{\hat{\partial}_n} C_{n-1} \quad (93)$$

defined for  $n > 0$  by

$$\hat{\partial}_n \delta := \sum_{\epsilon \in \delta_{n-1}} \epsilon, \quad (94)$$

where  $\delta \in \Delta_D$ . The exact sequence from which the  $d$ th color code is obtained is

$$C_{\bar{d}-1} \xrightarrow{\hat{\partial}_{\bar{d}}^*} C_D \xrightarrow{\hat{\partial}_d} C_{d-1}, \quad (95)$$

where  $0 \leq d \leq D$ . It gives rise to the homology group

$$\hat{H}_d := \hat{Z}_d / \hat{B}_d. \quad (96)$$

It follows from (92) that the homology groups  $\hat{H}_{\bar{d}}$ ,  $\hat{H}_d$  are isomorphic to the sum of  $\binom{D}{d}$   $h_d$  copies of  $\mathbf{Z}_2$ . In other words, we have

$$\hat{H}_d \sim \binom{D}{d} H_d. \quad (97)$$

The origin of the binomial coefficient is related to color and will be explained in the next section.

**7.5.4. Brane-nets.** We will now give a geometric interpretation to the  $D$ -simplex cycles discussed in the previous section in terms of branes. For any  $n$ -cell  $\lambda \in \Lambda_n$  we define  $c_\lambda \in C_D$  as

$$c_\lambda := \sum_{v \in \lambda_0} v^*. \quad (98)$$

For  $0 < n < D$ , we will now introduce boundary operators that take  $n$ -cells to  $(n - 1)$ -chains in a ‘natural’ way. For  $\lambda \in \Lambda_n$ , we set

$$\partial\lambda := \hat{\partial}_n c_\lambda. \quad (99)$$

We say that these definitions are natural because  $\partial\lambda$ , as a set of simplices glued together, is homeomorphic to the boundary of  $\lambda$  in the usual sense. Moreover, if  $\lambda$  is a  $q$ -cell,  $q \subset Q$ , then  $\partial\lambda$  only contains  $q$ -simplices. Thus, we can regard  $n$ -cells as pieces of  $n$ -branes. Indeed, coloring is important here, and we should regard a  $q$ -cell,  $q \subset Q$ , as a piece of a  $q$ -brane. As we will see next, such pieces can be put together to form closed  $q$ -branes, that is, branes without boundaries in the sense of the  $\partial$  just defined, which is just a reinterpretation of  $\hat{\partial}_n$ .

Let us fix a particular color subset  $q$  with  $|q| = n$  and introduce the  $q$ -colored chain subgroup  $C_D^q$ , the subgroup of  $C_D$  generated by all  $c_\lambda$  chains with  $\lambda \in \Lambda_q$ . That is, the chains in  $C_D^q$  represent the  $q$ -branes just discussed. We may then set  $Z_q := \hat{Z}_n \cap C_D^q$ ,  $B_q := \hat{B}_n \cap C_D^q$  and  $H_q := Z_q/B_q$  with the net result

$$H_q \sim H_n. \quad (100)$$

Thus, the homology of  $q$ -branes for a fixed  $q$  is just the usual brane homology. This result is shown in [22] by constructing from the colex  $\Lambda$  a  $q$ -reduced lattice in which all  $n$ -cells correspond to  $q$ -cells of  $\Lambda$ .

In summary, the cycles in  $Z_q$  can be regarded as closed  $q$ -branes, which are also cycles in  $\hat{Z}_n$ . In general, cycles in  $\hat{Z}_n$  will combine several pieces of such  $q$ -branes of different colors, forming  $n$ -brane nets [22]. In this sense,  $\hat{H}_n$  is an  $n$ -brane net homology group, as opposed to the usual  $H_n$  homology group of  $n$ -branes. This discussion already suggests that  $\hat{H}_n$  contains several copies of  $H_n$ , but how many of them? To answer this, let us consider color subsets  $q_i \subset Q$ ,  $i = 1, \dots, D + 2 - n$ , with  $|q_i| = n$  and  $|q_i \cap q_j| = n - 1$  for  $i \neq j$ . It turns out [22] that if  $z_i$ ,  $i = 1, \dots, D + 1 - n$ , are  $q_i$ -branes, all of them homologous in the usual sense of  $H_n$  homology, then

$$\sum_i z_i \sim 0 \quad (101)$$

up to  $\hat{H}_n$  homology, which shows that only  $\binom{D}{n}$  color combinations are independent, which finally explains (96).

In a  $d$ th color code the relevant homology groups are  $\hat{H}_{\bar{d}}$  and  $\hat{H}_d$ , with the corresponding  $\bar{d}$ - and  $d$ -brane nets, which are mapped to operators using (73). The commutation rules of closed

brane-net operators are purely topological. Namely, given a closed  $q$ -brane net  $z \in \hat{Z}_q$ ,  $|q| = d$ , and a closed  $q'$ -brane net  $z' \in \hat{Z}_{q'}$ ,  $|q'| = \bar{d}$ , the operators  $X_z$  and  $Z_{z'}$  anticommute if  $z$  and  $z'$  intersect an odd number of times and  $q \cap q' = \emptyset$ . Otherwise they commute.

**7.5.5. Branyon nets.** For  $d = 1, \dots, D$ , the Hamiltonian model (84) of a  $d$ th color code contains as in toric codes two kinds of brane-like excitations,  $(\bar{d} - 1)$ -branyons and  $(d - 1)$ -branyons. These are created with the brane-net operators discussed in the previous sections, in particular with open brane nets since the excitations appear at their boundary. Then, unlike in toric codes, branyons can form nets and we must talk about branyon nets. As branes in the previous sections, branyons can be labeled with colors. Namely, a  $q$ -branyon is the boundary of a  $q$ -brane.

We will only discuss  $(d - 1)$ -branyons, which are related to violations of the stabilizer conditions at  $(\bar{d} + 1)$ -cells. The case of  $(\bar{d} - 1)$ -branyons is analogous. If  $\lambda \in \Lambda_{\bar{d}+1}$  is a  $q$ -cell, we say that the corresponding excitation is a  $(Q - q)$ -branyon excitation. We know that branyon excitation must be arranged in closed structures. This fact is reflected [22] in a ‘conservation law’ at  $(\bar{d} + 2)$  cells. Namely, for  $\lambda$  a  $q$ -cell,  $|q| = \bar{d} + 2$ , and for any  $r, s \subset q$ ,  $|r| = |s| = \bar{d} + 1$  we have

$$\prod_{\mu \in \lambda_r} Z_\mu = \prod_{\mu \in \lambda_s} Z_\mu, \quad (102)$$

which is just a special case of (91). This implies, for example, that at a given  $(d + 2)$  cell there cannot exist a single excited  $(d + 1)$ -cell.

In the previous section we have seen that only  $\binom{D}{d}$  color combinations are enough to label an independent set of  $d$ -brane operators. We will now show that we can locally transform a set of branyon excitations into a new set that only contains branyons from  $\binom{D}{d}$  independent colors. By ‘locally’ we mean that it is enough to act in the neighborhood of the undesired excitations. We start by choosing a color  $q_0 \in Q$ . We want to show that we can eliminate all  $q$ -branyon excitations with  $q_0 \in q$  or, to put it in a different way, all excitations located at cells  $\lambda \in \Lambda_{\bar{d}+1}$  of color  $q$  with  $q_0 \notin q$ . The total set of excitations corresponds to the boundary of some  $c \in C_D$ ; that is, they can be switched off with  $X_c$ . Note that all the  $q$ -cells with  $q_0 \notin q$  that touch a  $(Q - q_0)$ -subcolex are part of it. Then the restriction of  $X_c$  to those  $(Q - q_0)$ -subcolexes that contain the undesired excitations will do the job.

Once we have reduce the branyonic excitations to the  $\binom{D}{d}$  color combinations, each color may be treated independently as in the previous section. That is, the branyons of each color form boundary  $(d - 1)$ -branes.

Finally, we note that, as in the case of generalized toric codes, one can derive [22] from the commutation rules for closed brane operators the topological  $\pm 1$  phases appearing in branyon processes.

**7.5.6. Self-correcting models.** Regarding self-correction, color code models behave very much like toric code models. That is, for  $d > 1$ , the models with Hamiltonian (84) for the  $d$ th code protect  $Z$ -type encoded operators, and for  $d < D - 1$  they protect  $X$ -type operators. Instead of repeating the analysis already done for toric codes, we will just mention the new features that appear in color codes. In particular, we consider the correction of  $X$ -type errors.

The first difference is the lattice, but any periodic  $D$ -colex will do, and the existence of periodic lattices follows from the procedure given in [22] to construct  $D$ -colexes from arbitrary



lattices. The second difference is that now we deal with branyon nets instead of branyons, but this does not make such a big difference due to the discussion in the previous sections. In this regard, we first need a notion of connectedness of  $(d+1)$ -cells. This is already implicit, for example, in (102): we construct a graph with  $(\bar{d}+1)$ -cells as its nodes in which nodes are connected by a link when the corresponding cells are part of the same  $(\bar{d}+2)$ -cell (note that  $d \geq 2$ ). As for the corr function, we transform each connected set of excitations using the procedure of the previous section into a collection of  $(d-1)$  boundaries of  $\binom{D}{d}$  colors, so that each color can be independently treated exactly as in a toric code.

### 7.6. Transversality properties of color codes

Color codes can have very special transversality properties when the lattice and its topology are suitably tuned. In this section, we will treat this issue in detail. To this end, we start explaining a simple way to consider  $D$ -colexes in manifolds with boundary. We then describe a particular topology for the manifold: in the shape of a  $D$ -simplex in which each face has different properties. Finally, we analyze the transversality properties of the resulting simplicial codes.

*7.6.1. Boundaries.* Up to now we have only considered colexes  $\Lambda$  in closed  $D$ -manifolds. However, richer color codes can be obtained by introducing suitable boundaries in the manifold [20, 21]. The easiest way to describe such boundaries is as collections of missing cells in the lattice. For a missing cell in  $\Lambda$ , we really mean a missing generator of the stabilizer. When the removed generator is independent from the remaining ones, the number of encoded qubits  $k$  gets increased (4). This can be understood also in homological terms. Consider a  $q$ -brane  $c \in C_D$  with a boundary,  $|q| = d$ . Given an encoded state  $|\psi\rangle$ , the state  $X_c|\psi\rangle$  will have stabilizer violations at those  $(\bar{d}+1)$ -cells that form the boundary of  $c$ , where as above  $\bar{d} = D - d$ . But, if we eliminate the generators that correspond to these  $(\bar{d}+1)$ -cells, then  $X_c$  will belong to the stabilizer and thus we should consider  $c$  a closed  $q$ -brane. This way, as cells are eliminated we are introducing in the manifold areas in which suitable branes can have a boundary and still be closed. It is as if we had removed a piece of manifold in these areas, thus creating a boundary. Such an introduction of a boundary does not change homology only in the way just discussed. In addition, when we remove a  $q$ -cell the corresponding generator, which can be regarded as a  $d$ -brane, transforms into an encoded operator in  $\mathcal{N} - \mathcal{S}$ . Thus, this  $d$ -brane should no longer be considered a boundary, but rather a nontrivial cycle.

Although for our purposes it will be much more convenient to describe boundaries as big missing cells in the lattice, this would not be convenient at all in an implementation of the model. Fortunately it is possible to adopt a different approach, as discussed in [48] in the context of 2D toric code models and also in [49] in a more general framework. In this alternative approach boundaries appear at interfaces between the topologically ordered phase and other phases in which some of the branyons are condensated.

*7.6.2. Simplicial color codes.* There exists a very special class of color codes which live in simplex-shaped lattices. For  $D = 2$  we have triangular codes [20], which bear this name because their lattice is a triangle: each side of the triangle corresponds to a different kind of boundary. For  $D = 3$  we have tetrahedral codes [21], and more generally for arbitrary  $D$  we may consider simplicial codes. Instead of giving a detailed description of the properties of each

of the boundaries in a simplicial code, it is just easier to give a constructive procedure to obtain such codes, as we do next.

Suppose that we have at our disposal a  $D$ -colex  $\Lambda$  with the topology of a  $D$ -sphere,  $D \geq 2$ . Since spheres are homologically trivial for  $D \geq 2$ ,  $\Lambda$  yields trivial color codes with  $k = 0$ , no encoded qubits. Let us now choose a vertex of  $\Lambda$  and proceed to remove the corresponding physical qubit and all the  $(d + 1)$ - and  $(\bar{d} + 1)$ -cells that meet at this vertex. The resulting lattice looks like a  $D$ -simplex, because it is a  $D$ -ball with its boundary  $(D - 1)$ -sphere divided in the  $D + 1$  regions that were in contact with each of the  $D + 1$  removed  $D$ -cells. It can be checked using (91) that only two independent generators of  $\mathcal{S}$  are eliminated in the process, one of  $X$  type and the other of  $Z$  type. Thus (4), all color codes derived from the new lattice  $\Lambda'$  encode exactly one qubit. Since the original lattice has an even number of qubits [21], it follows that the operators

$$\bar{X} := X^{\otimes n}, \quad \bar{Z} := Z^{\otimes n}, \quad (103)$$

with  $n$  the total number of remaining qubits, are the encoded  $X$  and  $Z$  operators: they belong to the normalizer and

$$\{\bar{X}, \bar{Z}\} = 0. \quad (104)$$

It is easy to construct the smallest example of a simplicial color code for each dimension  $D \geq 2$ . We start with a  $(D + 1)$ -hypercube, in which we color parallel links with the same color to get [22] a  $D$ -colex with the topology of a  $D$ -sphere. Then as explained above we eliminate one of its vertices and all neighboring cells. As a result we get a color code with  $2^{D+1} - 1$  physical qubits in which all  $n$ -cells have  $2^n$  vertices.

**7.6.3. Transversal gates.** We will now consider the transversality properties of color codes. As it turns out, simplicial colexes give the most suitable color codes for the transversal implementation of quantum gates. Thus, we restrict our discussion to them.

First, it is clear that the transversal implementation of the logical  $X$  and  $Z$  operators is always possible, since they take the form (103). For the same reason, transversal measurements of  $X$  and  $Z$  are always possible. Similarly, the two-qubit CNot gate

$$C_X := |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X \quad (105)$$

can always be implemented transversally on pairs of equal simplicial codes by applying it to each pair of equivalent qubits. The one-qubit Hadamard gate

$$H := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (106)$$

is more restrictive. As it exchanges  $X$  and  $Z$  operators, it can only be applied to dual codes, those with  $\bar{d} = d$ . Again, it can be obtained by applying a Hadamard to each of the physical qubits independently.

We want to consider in general the family of one-qubit phase gates

$$R_k := |0\rangle\langle 0| + \exp\left(\frac{\pi i}{2^k}\right) |1\rangle\langle 1|, \quad (107)$$

where  $k$  is a non-negative integer. The case  $k = 0$  is just the  $X$  gate. The case  $k = 1$  is important since the gates  $\{C_X, H, R_1\}$  generate the Clifford group. Still more interesting is the case  $k = 2$ , because  $C_X$  and  $R_2$ , together with  $X$  and  $Z$  initializations and measurements, are enough for

universal quantum computation [50]. To study in detail when these gates can be implemented transversally on simplicial color codes, we will introduce next the notion of  $(j, k)$ -good colexes.

From now on,  $j, k$  are integers with  $0 \leq j \leq D, k \geq 0$ .

**Definition 3.** A  $D$ -colex  $\Lambda$  is  $(j, k)$ -good if

$$\forall c \in \hat{B}_j, \quad |c| \equiv 0 \pmod{2^{k+1}}. \quad (108)$$

Note that all colexes are  $(j, 0)$ -good. From the first expression for encoded states in (80) and the fact that simplicial colexes have an odd number of physical qubits, it follows that simplicial color codes obtained from  $(d, k)$ -good colexes allow the transversal implementation of  $R_k$  gates. More exactly, if we apply  $R_k$  to each physical qubit independently we will get a logical  $R_k^s$  gate, with an odd power  $s$ . But since  $s$  and  $2^{k+2}$  are coprime, there exists an integer  $r$  such that  $R_k^{sr} = R_k^{2^{k+2}+1} = R_k$ .

The problem with definition 3 is that it is not really operative. But we can give a simpler characterization of  $(j, k)$ -goodness. For its proof, see appendix C

**Theorem 1.** A nice  $D$ -colex  $\Lambda$  is  $(j, k)$ -good for some  $k > 0$  if and only if  $j' := 2j - D \geq 0$ , all  $j$ -subcolexes are  $(j', k - 1)$ -good and

$$\forall \lambda \in \Lambda_{j+1}, \quad |\lambda_0| \equiv 0 \pmod{2^{k+1}}. \quad (109)$$

Moreover, the  $\Leftarrow$  implication holds even if  $\Lambda$  is not nice.

Since  $(0, k)$ -good colexes only exist for  $k = 0$ , by repeatedly applying theorem 1 it follows that a  $(j, k)$ -good and nice  $D$ -colex may exist only if

$$j \geq \frac{k}{k+1}D. \quad (110)$$

Several conclusions may be derived from this result. For example, dual codes only allow the transversal implementation of the  $R_1$  gate. Indeed, if this were not the case, i.e. a dual code allowed transversal implementation of  $R_2$ , then this would imply existence of a stabilizer code supporting the application of a universal set of operations, in contrast to the result present in [51]. For fixed  $\bar{d}$ , there is a minimal dimension  $D$  for which  $(d, k)$ -good nice  $D$ -colexes exist, namely

$$D \geq (k+1)\bar{d}. \quad (111)$$

In particular, setting  $\bar{d} = 1$  it follows that the minimal dimension  $D$  for which  $R_k$  may be transversally implementable in a color code is  $D = k + 1$ .

Once the bound (111) is established, it only remains to be shown that it is saturated by some lattice. But, the simplest examples of simplicial lattices given in the previous section indeed saturate the bound. The next question is whether periodic lattices that saturate it exist. This is indeed the case for  $D = 2, 3$ .

As for higher dimensions, we are particularly interested in the case  $D = 6, k = 2, \bar{d} = 2$ . Indeed, we need  $\bar{d} > 1$  if the system is to be self-correcting and  $k > 1$  to obtain universal computation with transversal operations. Thus,  $D = 6$  is the lowest dimensionality that puts the two features together. As can be readily checked, the construction of section 7.5.1 provides a suitable periodic 6-colex. Finally, by adding one more dimension we can obtain local CNot gates.

## 8. Self-correcting quantum computers

### 8.1. Initialization from thermal states

*8.1.1. Initialization with an external field.* In the previous sections, we have studied how quantum memories resilient to thermal noise can be obtained from local quantum models in 4D and above, from both toric codes and color codes. However, a quantum memory is not of much use if we cannot initialize it. We will now sketch a possible procedure to initialize the self-correcting topological quantum memories as long as we can control the couplings of the local quantum Hamiltonian—without any sort of physical qubit addressing. In particular, we will consider toric and color code models with  $d > 1$ , so that the dressed version of  $Z$ -type encoded operators  $N$  are protected, and give a procedure to initialize all such operators to  $N = 1$  with a probability of failure exponentially small in the system size. For  $D = d = 2$ , this is just the usual initialization of a 2D ferromagnet with an external magnetic field.

Our starting point is a modification of the Hamiltonian (84). We introduce single-qubit Zeeman terms and some couplings to reflect the fact that now the strength of the different terms can be controlled:

$$H = -g_X \sum_{i=1}^l X_{p_i} - g_Z \sum_{i=1}^m Z_{q_i} - g_0 \sum_s Z_s, \quad (112)$$

where  $s$  runs over all qubits and  $g_X, g_Z, g_0 \geq 0$ . Before considering the details, let us sketch the initialization procedure. We start from a thermalized state with  $g_Z = g_0 = 1$  and  $g_X = 0$ . Then we switch off the Zeeman terms. Finally, we ‘carefully’ switch on the  $X$ -type stabilizers until  $g_1 = 1$ , as discussed below. Note that the Hamiltonian is exactly solvable at all times, since all the terms with nonzero coupling at a given time commute. Moreover, all the terms commute with  $Z$ -type encoded operators, precisely the ones that we want to initialize with eigenvalue 1.

We first want to show that, for temperatures  $1/\beta$  below a critical value, the initial thermalized state satisfies, for any encoded operator  $N \in \mathcal{N} \cap \mathcal{P}_Z$  and some  $t > 0$ ,

$$1 - \langle N_{\text{dr}} \rangle_\beta \leq e^{-\delta L}. \quad (113)$$

Recall that  $g_X = 0$  and  $g_Z = g_0 = 1$ . Both for toric and color codes we already have a graph of the excitations that correspond to  $g_Z$  terms, where nodes are labeled by the  $q_i$  elements of the lattice. We enlarge this graph by adding a node per physical qubit, to represent the excitations that correspond to the Zeeman  $g_0$  terms. We also add a link from the node of the physical qubit  $s$  to the node of  $q_i$  if  $\{X_s, Z_{q_i}\} = 0$ , and a link between the nodes of two physical qubits  $s$  and  $s'$  if they are both linked to some  $q_i$ . For example, in a toric self-correcting code with  $d = 4$  the  $q_i$  nodes correspond to edges and the  $s$  nodes to plaquettes, so that the links of the graph just represent the usual notion of connectedness for plaquettes and edges in a lattice. The Zeeman term is responsible for the tension of membranes built of plaquettes, while the Kitaev term for the tension of the loops, i.e. boundaries of the membranes.

It is then clear that any connected set of excitations can be independently switched off with a suitable operator in  $\mathcal{P}_X$ . We can then repeat the reasoning of section 6.4 to show that those excitation configurations which contain connected sets of size bigger than, say,  $L/2$  are exponentially suppressed with the size of  $L$ . Therefore, it suffices to observe that the rest of the configurations belong to the sector with  $N_{\text{dr}} = 1$  for any  $N \in \mathcal{N} \cap \mathcal{P}_Z$ , as we detail next using

the language of section 7.3. First, the eigenstates of the Hamiltonian may be written as  $|a\rangle_0$  for some  $a \in A_1$ . The corresponding set of excitations corresponds both to the chain  $a$  and its boundary  $\partial_1 a \in A_0$ . But,  $\text{corr}(S'_2 \cap \partial a) = X_{a'}$  for some  $a' \in A_1$  such that both  $a$  and  $a'$  lie in the same hypercube of length  $L/2$ , and thus  $X_a X_{a'} \in \mathcal{S}$ . The result follows.

Once we have shown that the initial thermal state is in the desired sector, it is enough to check that this continues to be true as we switch off and on the interactions as indicated. This follows from the fact that  $g_Z = 1$  at all times, as long as we keep the temperature of the system stable. In this regard, the process of switching on  $g_X$  could be a problem since at the start we have arbitrarily many violations of the  $X_{p_i}$  stabilizers and if  $g_X$  increases too fast we will put a lot of energy in the system. To avoid this, we require that the switching on is done slowly enough for the system to stay close to the thermal state. In this regard, we assume that the thermalization time is independent of the system size, so that the time required to perform a gate does not scale either. This is a natural assumption for the systems considered here given the nature of their excitations, at least as long as the coupling with the environment is uniform.

*8.1.2. Initialization by code deformation.* An alternative way to initialize a topological quantum code is to use the code deformation approach as discussed in [48, 53]. Without going into details, the idea is to start from a code with no encoded qubits, and to change its geometry progressively to introduce an encoded qubit and at the same time initialize it in a well-defined manner. Note that the geometry is dictated by the stabilizer generators, and these in turn are the terms in the Hamiltonian, so that we are required to control the couplings of the protecting Hamiltonian at least with a certain degree of spatial resolution, even if no detailed control is required.

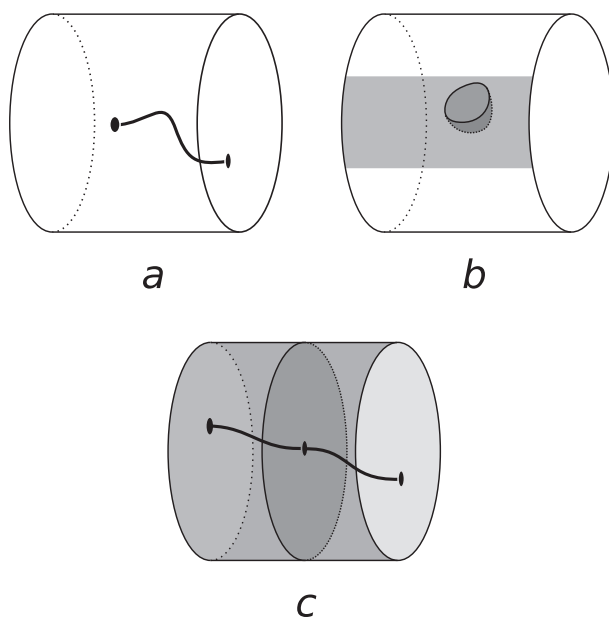
If the final encoded qubit should be initialized to  $Z = 1$ , the logical operator  $Z$  should correspond (up to the deforming process) to a stabilizer in the original trivial code. Also, along the whole process the geometry must be such that the  $X$  logical operator, when it exists, has a global support, ensuring that the encoded qubit is topologically protected from bit-flip errors. On the other hand, it will not be protected at all times from phase-flip errors, but just as in the initialization with Zeeman terms such errors are immaterial. As an example, figure 2 shows how this could be done in a 3D toric code by changing boundary conditions (stabilizers) over time.

There is thus a parallelism in these two ways to initialize the quantum memory, in the sense that they rely on ‘unprotecting’ at some point the logical operator that provides the initialization basis. It is unclear whether this is a general feature of the initialization of a self-correcting (or maybe just self-protecting) quantum memory or a particularity of our approach.

## 8.2. Measurements

The fact that we are dealing with CSS-like codes implies that we may make transversal  $X$  and  $Z$  measurements. Typically, measuring a physical qubit in the  $Z$  basis requires an interaction  $Z \otimes H_{\text{app}}$  with a measuring apparatus. In this regard, during the measurement we can switch off the  $X$ -type stabilizers, which do not commute with the measuring Hamiltonian. Switching off this part of the protecting Hamiltonian is not a source of problems since the corresponding errors do not affect the measurement.

Alternatively, code deformation could also be used to make the measurements.



**Figure 2.** Initialization of a 3D toric code by changing the boundary conditions over time. (a) At the start, open string operators with their endpoints on any boundary surface (as the one displayed) are stabilizers. There are no encoded qubits. (b) Boundary conditions change over a region (shaded) that extends as time passes. On this region string operators cannot end. Instead, membrane operators with their boundary in the region (as the one displayed) are stabilizers. (c) When the region with the alternate boundary conditions becomes nonsimply connected, an encoded qubit appears. The  $X$  logical operator is the membrane shown, whereas the  $Z$  logical operator corresponds to the string. The latter is initialized to 1 by continuity, as such a string operator was before a stabilizer.

### 8.3. Transversal gates

In section 7.6.3, we showed that 6D color codes are enough to have a set of universal transversal operations. Local CNOTs require an additional dimension to remain local, so that we need a total of seven spatial dimensions.

In order to perform transversal gates we have to temporarily modify the Hamiltonian of the system. This means adding new terms that will not commute with the protecting ones and thus make the analysis difficult. In any event, when a gate is halfway completed the system will, in general, be in a state with a high energy with respect to the protecting Hamiltonian. This could easily be a source of problems. To circumvent them, we propose to apply gates that are fast in comparison with the protective Hamiltonian. Thus, during the application of gates, the protective Hamiltonian can be neglected.

Note that we can represent the faulty gate as an ideal gate preceded by some noise operation, so that it is enough to consider a situation with a trivial gate (identity), i.e. qubits are subject solely to interaction with the environment. Thus, the process can be seen as a noisy channel, so that we can use the results of section 6.5, which tell us that the quantum memory can resist a small amount of noise. Indeed, the results in that section were obtained for a depolarizing

channel acting independently on each qubit, but they can be generalized to more complicated noises as long as correlations are local.

Because fast gates may result in a large amount of noise [14] (see footnote 7), we must create separation between the time scales of the gate and the protective Hamiltonian by decreasing the coupling of the protective Hamiltonian rather than making the gate faster. Then, however, the ratio  $T/\Delta$  (where  $\Delta$  is the minimal separation between the energy levels of the protective Hamiltonian) will become larger, and the threshold for resilience against white noise will decrease. Therefore, in order to keep the noise produced by the gate below that threshold, we shall tune down temperature. In this way, if we are given good enough gates, we are able to choose the coupling constant of the protective Hamiltonian and the temperature in such a way that the gate is applied to the logical qubit with arbitrary good fidelity.

However, this is not really the end of the story. Even if we could perform transversal gates perfectly and in an instant, the gates themselves will increase the number of excitations. In the case of single-qubit gates we know that excitations cannot spread much, so this is not a big problem. The case of CNot gates is worse, because they copy errors (and thus excitations) from one code to another. It is intuitively clear, however, that as long as excitations are scarce enough this will not affect the encoded qubits. A rigorous approach to this problem requires using similar arguments to those in sections 6.4 and 6.5, but considering two copies of the  $\Gamma_L$  lattices together.

In summary, the application of a transversal gate will put some amount of energy per volume unit in the system. As long as the temperature is low enough and the energy is not too much, this will not produce an error. Of course, we have to wait for some time between gates, so that the injected energy can be dissipated.

## 9. Conclusions

We have provided a quantum-topology-based solution to the problem of constructing a self-correcting QC. One of the main problems we have to face is the existence of tensionless string excitations (errors) that propagate throughout the quantum device causing the QC to collapse. Thus, a mechanism for reabsorption of errors is necessary and a natural way to achieve this in a topological scenario is by increasing the spatial dimension, but this has to be done with an appropriate class of topological codes such as TCCs that provide a complete set of tools for a self-correcting QC. This situation is reminiscent of another fundamental problem in gauge theories, i.e. color confinement in QCD. In fact, there exists the string picture of confinement based on string tension on quarks [54]. We may conjecture that a possible solution to the critical dimension of the self-correcting QC is to find appropriate confinement mechanisms, i.e. we arrive at the following relationship:

$$\text{Self-correcting QC} \rightleftharpoons \text{Confinement problem.}$$

The solution of this confinement problem is an open question in topological quantum computation.

Related to this central problem, we have constructed a model for a self-correcting QC based on topological stabilizer codes. The model, to be local, requires seven spatial dimensions—the codes require six dimensions and an extra dimension is needed to keep CNot gates local. Our model consists of (1) a lattice of qubits exhibiting proper topological and geometrical properties so that (2) a code on this lattice allows a universal set of transversal gates (including

measurements) and (3) a protecting Hamiltonian, with controllable coupling constants. We have described the procedures for initialization and measurement of the QC and how gates should be performed. For the latter, we have also proposed an initialization method based on code deformations that avoids switching on/off Hamiltonians<sup>8</sup>.

We have then obtained two rigorous results concerning the self-protection of such a QC. Firstly, we have shown within the weak-coupling limit regime that the memory is stable, generalizing thereby the results of [16]. Secondly, we have shown that if the memory is exposed to a local noise for a short time the quantum information is preserved. Our further results are physically well motivated and established. These comprise the initialization scheme and a reasoning showing that we can perform gates on protected qubits with high fidelity.

We would like to emphasize that as the formulation of a self-correcting QC is a challenge, we decided to relax the number of spatial dimensions  $D$  in which our proposals are formulated. The point is to address this problem as a fundamental issue regardless of the dimensionality. Once we have seen that this is possible under the circumstances considered in this paper, we may reflect on possible ways to lower the dimensionality  $D$  where the lattices are embedded. To this end, we should consider our models as simplified models as far as the types of degrees of freedom are concerned. We have only used spin-1/2 or qubits. It is an open question whether by including additional degrees of freedom like fermions or bosons, we could induce interactions that may lower the dimension in which self-correcting QCs are possible [13]. A similar possibility is to consider models with generalized spins and discrete non-abelian gauge symmetry groups [52] which are found to possess string tension and confining properties. Generally speaking, there are instances of quantum theories whose critical dimensions for being stable can be lowered by introducing additional types of degrees of freedom or a combination thereof. This could also be the case here and this possibility we leave open, but with the new insight gained here, we see that at least with simpler models it is possible to achieve the property of self-correctedness.

Finally, let us note that our results are in contradiction with [26], where it is argued that a QC could be used to draw work from a heat bath. It would be worthwhile to get more insight into this problem, especially regarding the points which are more physically established in our approach, particularly the limitations in the access to the quantum memory.

## 10. Later results

We have already commented on previous results on self-correcting quantum devices in section 2. After the completion of this work, a series of very interesting articles have also appeared addressing the possibility of QCs with self-correcting capabilities. In particular, some of them introduce new topological codes in more physically accessible dimensions like 3D codes possessing some self-correcting properties [55–57]. We would like to emphasize a basic difference between those proposals and ours for the sake of clarity in the quest for a self-correcting QC. Namely, all those novel schemes do not work at a fixed temperature, i.e. in order to increase fidelity, the temperature must be decreased. In contrast, in this paper we consider a strong notion of self-correction involving a phase transition at a certain critical temperature. This

<sup>8</sup> Let us note in passing that we have also proposed to temporarily switch off the protecting Hamiltonian, so that we can avoid problems due to the combined effect of the protecting Hamiltonian and the one needed to perform the gates.



means that we are using the most demanding or realistic scenario: we assume finite temperature  $T > 0$ , implying that our notion of self-correction is very strong. We prefer this strong notion rather than introducing a weaker notion of self-correction in which memories are initialized at  $T = 0$  and demanding a size-dependent thermalization time.

## Acknowledgments

We are grateful to Robert Alicki for asking questions that led to this paper and for numerous discussions. MH thanks Jonathan Oppenheim for discussions. HB and MAM-D acknowledge financial support from a PFI grant of EJ-GV, DGS grants under contract, FIS2006-04885 and the ESF INSTANS 2005-10. RWC and MH are supported by EC IP SCALA and by the Polish Ministry of Science and Higher Education through grant no. NN20223193. RWC also acknowledges support from the Foundation of Polish Science (FNP). The support from the Polish research network LFPPI is also acknowledged. Part of this work was done in the National Quantum Information Centre of Gdansk. Part of this work was initiated during the Madrid 2008 conference on the Mathematical Foundations of Quantum Control and Quantum Information Theory (Fundacion Areces).

## Appendix A. Relationship between fidelity and the autocorrelation function: the proof of proposition 1

Let us denote  $X = P_+ - P_-$ , with  $P_+ + P_- = I$ . From the definition of  $F(\psi; \Lambda_{V,\text{anc}}^*, \rho_{\text{anc}})$ , we obtain

$$F(\psi_+; \Lambda_{V,\text{anc}}^*, \rho_{\text{anc}}) = \text{Tr}(I_V \otimes \rho_{\text{anc}} P_+ \Lambda(P_+)) \quad (\text{A.1})$$

and similarly for  $F(\psi_-; \Lambda_{V,\text{anc}}^*, \rho_{\text{anc}})$ . On the other hand, using that  $\Lambda$  preserves the identity, we obtain

$$X \Lambda(X) = 2(P_+ \Lambda(P_+) + P_- \Lambda(P_-)) - I. \quad (\text{A.2})$$

This, together with equation (A.1), implies the required formula.

## Appendix B. A bound for connected collections

We want to show that, in a graph with at most  $\mu$  links meeting at each node, the number of connected collections of nodes of size  $k$  that contain a given node  $n$  is bounded by  $e^{\tau k}$  for a suitable  $\tau$  that only depends on  $\mu$ . The idea of the proof is to express a collection as a word  $W$  in a suitable alphabet with  $\mu + 1$  letters. For a given node, each neighboring node is identified by one of the first  $\mu$  letters, and there is an extra ‘back’ letter. A valid word can have at most  $l/2$  ‘back’ letters among its first  $l$  letters, and we construct a collection from a valid word  $W$  as follows. At each step we keep an ordered list of nodes  $(n_1, \dots, n_l)$ . At the start the list is  $(n)$ , so that it has a single element. At the  $j$ th step if the  $j$ th letter of  $W$  is not a ‘back’ letter, we add the corresponding neighboring node of  $n_l$ , the last element of the list. If instead it is a ‘back’ letter, we remove the last node in the list. The collection is formed by those nodes that were in the list at some point in the process, so that clearly at most  $2k - 3$  letters are necessary to encode a collection of size  $k$  and thus we can take  $\tau = 2\log(\mu + 1)$ .

### Appendix C. Proof of theorem 1

First we need some extra notions for those cases in which  $j' \geq 0$ . Recall that each  $\lambda \in \Lambda_{j+1}$  is a  $j$ -colex  $\Lambda'$ . We introduce a group homomorphism  $h_\lambda : C_D \rightarrow C'_j$  that restricts  $c \in C_D$  to  $C'_j$ , the  $j$ -simplex chain group of  $\lambda$ . In particular for  $c = \sum_{v \in V} v^*$  with  $V \subset \Lambda_0$  we set  $h_\lambda(c) := \sum_{v \in V'} v^*$  with  $V' = V \cap \lambda_0$ . This homomorphism maps  $\hat{B}_j$  to  $\hat{B}'_j$ . To check this, it suffices to show that for any  $\mu \in \Lambda_{j+1}$  we have  $h_\lambda(c_\mu) = \sum_{v \in N} c_v$  for some  $N \subset \lambda_{j'+1}$ . But, if  $\mu \cap \lambda = \{v^i\}$  we just set  $N = \bigcup_i (v^i)_{j'+1}$ .

When  $\Lambda$  is nice, we can in addition define a group homomorphism  $h^\lambda : \hat{B}'_j \rightarrow \hat{B}_j$  such that for every  $c' \in \hat{B}'_j$  we have  $h^\lambda(h^\lambda(c')) = c'$ . For each  $v \in \Lambda_{j'}$  we set  $h^\lambda(c'_v) := c_{v^q}$ , where  $q = Q - (r - s)$  with  $s, r \subset Q$  such that  $v \in \Lambda_s, \lambda \in \Lambda_r$ . Here,  $c'_v$  is meant to denote  $h_\lambda(c_v)$ .

We can now prove each of the implication directions:

$\Rightarrow$  / First, for any  $\lambda \in \Lambda_{j+1}$  we have  $c_\lambda \in \hat{B}_j$ , so that  $|\lambda_0| = |c_\lambda| \equiv 0 \pmod{2^{k+1}}$ . Suppose that  $j' \leq 0$ . Then, for any two colorings  $q, r \subset Q$  with  $|q| = |r| = j + 1$ ,  $|q \cap r| = \emptyset$  and for any  $v \in \Lambda_0$  we get, using the niceness of  $\Lambda$ , that  $|c_{v^q} + c_{v^r}| = |c_{v^q}| + |c_{v^r}| - 2 \equiv -2 \pmod{2^{k+2}}$ , a contradiction. Finally, suppose that  $\lambda \in \Lambda_{j+1}$  is not  $(j', k - 1)$ -good as a  $j$ -subcolex. Then, with the same notation as above, there exists a chain  $c' \in \hat{B}'_j$  with  $|c'| \not\equiv 0 \pmod{2^k}$ . For  $c = h^\lambda(c')$ , we have by assumption  $|c| \equiv 0 \pmod{2^{k+1}}$ , but then  $|c + c_\lambda| = |c| + |c_\lambda| - 2|c'| \not\equiv 0 \pmod{2^{k+1}}$ . Since  $c, c_\lambda \in \hat{B}_j$ , this is a contradiction.

$\Leftarrow$  / It suffices to show that for any  $c \in \hat{B}_j$  with  $|c| \equiv 2 \pmod{2^{k+1}}$  and for any  $\lambda \in \Lambda_{j+1}$  we have  $|c + c_\lambda| \equiv 0 \pmod{2^k}$ . But, for  $c' = h_\lambda(c)$  we get  $|c + c_\lambda| = |c| + |c_\lambda| - 2|c'|$  and since  $|c'| \equiv 0 \pmod{2^k}$  the result follows.

### References

- [1] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
- [2] Cirac J I and Zoller P 1995 Quantum computations with cold trapped ions *Phys. Rev. Lett.* **74** 4091–4
- [3] Shor P 1996 Fault-tolerant quantum computation *Proc. Symp. on the Foundations of Computer Science* (Los Alamitos: CA) (arXiv:quant-ph/9605011)
- [4] Kitaev Yu A 1997 Quantum error correction with imperfect gates *Quantum Communication, Computing and Measurement* ed O Hirota, A S Holevo and C M Caves pp 181–8
- [5] Knill E and Laflamme R 1996 Concatenated quantum codes arXiv:quant-ph/9608012
- [6] Knill E, Laflamme R and Zurek W 1996 Accuracy threshold for quantum computation arXiv:quant-ph/9610011
- [7] Knill E, Laflamme R and Zurek W 1998 Resilient quantum computation: error models and thresholds *Science* **279** 342
- [8] Aharonov D and Ben-Or M 1999 Fault-tolerant quantum computation with constant error rate arXiv:quant-ph/9906129
- [9] Kitaev Yu A 1997 Quantum computing: algorithms and error correction *Russ. Math. Surveys* **52** 1191–249
- [10] Preskill J 1998 reliable quantum computers *Proc. R. Soc. Lond. A* **454** 385–410
- [11] Wen X-G 2004 *Quantum Field Theory of Many-Body Systems: From the Origin of Sound to an Origin of Light and Electrons* (Oxford: Oxford University Press)
- [12] Kitaev Yu A 2003 Fault-tolerant quantum computation by anyons *Ann. Phys.* **303** 2
- [13] Dennis E, Kitaev A, Landahl A and Preskill J 2002 Topological quantum memory *J. Math. Phys.* **43** 4452–505
- [14] Alicki R, Lidar D A and Zanardi P 2006 Internal consistency of fault-tolerant quantum error correction in light of rigorous derivations of the quantum Markovian limit *Phys. Rev. A* **73** 052311

- [15] Alicki R, Fannes M and Horodecki M 2008 On thermalization in Kitaev's 2D model arXiv:0810.4584
- [16] Alicki R, Horodecki M, Horodecki P and Horodecki R 2008 On thermal stability of topological qubit in Kitaev's 4D model arXiv:0811.0033
- [17] Kay A and Colbeck R 2008 Quantum self-correcting stabilizer codes arXiv:0810.3557
- [18] Bravyi S and Terhal B 2008 A no-go theorem for a two-dimensional self-correcting quantum memory based on stabilizer codes arXiv:0810.1983
- [19] Bacon D 2006 Operator Quantum error correcting subsystems for self-correcting quantum memories *Phys. Rev. A* **73** 012340
- [20] Bombin H and Martin-Delgado M A 2006 Topological quantum distillation *Phys. Rev. Lett.* **97** 180501
- [21] Bombin H and Martin-Delgado M A 2007 Topological computation without Braiding *Phys. Rev. Lett.* **98** 160502
- [22] Bombin H and Martin-Delgado M A 2007 Exact topological quantum order in  $D = 3$  and beyond: branyons and Brane-Net condensates *Phys. Rev. B* **75** 075103
- [23] Weimer H, Muller M, Lesanovsky I, Zoller P and Buchler H P 2009 Digital coherent and dissipative quantum simulations with Rydberg atoms arXiv:0907.1657
- [24] Bombin H, Kargarian M and Martin-Delgado M A 2008 arXiv:0811.0911
- [25] Herdman C M, Kevin C Young, Scarola V W, Sarovar M and Whaley K B 2010 Stroboscopic generation of topological protection *Phys. Rev. Lett.* **104** 230501
- Kempe J, Kitaev A and Regev O 2004 The complexity of the local Hamiltonian problem *SIAM J. Comput.* **35** 1070–97
- Jordan S P and Farhi E 2008 Perturbative gadgets at arbitrary orders *Phys. Rev. A* **77** 062329
- [26] Alicki R 2009 Quantum memory as a perpetuum mobile of the second kind arXiv:0901.0811
- [27] Kay A 2008 The non-equilibrium reliability of quantum memories arXiv:0807.0287
- [28] Castelnovo C and Chamon C 2008 Topological order in a 3D toric code at finite temperature *Phys. Rev. B* **78** 155120
- [29] Iblisdir S, Perez-Garcia D, Aguado M and Pachos J 2008 Scaling law for topologically ordered systems at finite temperature arXiv:0806.1853
- [30] Kargarian M 2009 Finite temperature topological order in 2D topological color codes *Phys. Rev. A* **80** 012321
- [31] Knill E, Laflamme R and Viola L 2000 Theory of quantum error correction for general noise *Phys. Rev. Lett.* **84** 2525
- [32] Alicki R 2006 False qubits? Polarization of light and Josephson junction arXiv:quant-ph/0606242v2
- [33] Knill E 2006 On protected realizations of quantum information arXiv:quant-ph/0603252
- [34] Shor P W 1996 Scheme for reducing decoherence in quantum computer memory *Phys. Rev. A* **52** R2493
- [35] Steane A M 1996 Error correcting codes in quantum theory *Phys. Rev. Lett.* **77** 793
- [36] Gottesman D 1996 Class of quantum error-correcting codes saturating the quantum Hamming bound *Phys. Rev. A* **54** 1862
- [37] Calderbank A R, Rains E M, Shor P W and Sloane N J A 1997 Quantum error correction and orthogonal geometry *Phys. Rev. Lett.* **78** 405
- [38] Alicki R and Fannes M 2008 Decay of fidelity in terms of correlation functions arXiv:0809.4180
- [39] Alicki R 1988 *Physica A* **150** 455
- [40] Zanardi P and Rasetti M 1997 *Phys. Rev. Lett.* **79** 3306
- [41] Alicki R and Lendi L 2008 *Quantum Dynamical Semigroups and Applications* 2nd edn (Berlin: Springer)
- [42] Hofmann H 2005 *Phys. Rev. Lett.* **94** 160504 (arXiv:quant-ph/0411011)
- [43] Peierls R 1936 *Proc. Camb. Phil. Soc.* **32** 477
- [44] Griffiths R 1964 Peierls proof of spontaneous magnetization in a two-dimensional Ising ferromagnet *Phys. Rev. A* **136** 437
- [45] Dobrushin R L 1965 *Teoriya Veroyatnostei i ee Primeneniya* **10** 209  
Dobrushin R L 1965 *Theory Probab. Appl.* **10** 193 (Engl. transl.)
- [46] Bombin H and Martin-Delgado M A 2007 Homological error correction: classical and quantum codes *J. Math. Phys.* **48** 052105

- [47] Bombin H and Martin-Delgado M A 2007 Optimal resources for topological 2D stabilizer codes: comparative study *Phys. Rev. A* **76** 012305
- [48] Bombin H and Martin-Delgado M A 2008 An interferometry-free protocol for demonstrating topological order *Phys. Rev. B* **78** 165128
- [49] Bombin H and Martin-Delgado M A 2011 Nested topological order *New J. Phys.* **13** 125001
- [50] Knill E, Laamme R and Zurek W 1996 Accuracy threshold for quantum computation arXiv:quant-ph/9610011
- [51] Zeng B, Cross A and Chuang I L 2007 Transversality versus universality for additive quantum codes arXiv:0706.1382v3 [quant-ph]
- [52] Bombin H and Martin-Delgado M A A 2008 A family of non-Abelian Kitaev models on a lattice: topological confinement and condensation *Phys. Rev. B* **78** 115421
- [53] Bombin H and Martin-Delgado M A 2009 Quantum measurements and gates by code deformation *J. Phys. A: Math. Theor.* **42** 095302
- [54] 't Hooft G 1976 *High Energy Physics* ed M Zichichi (Bologna: Editrice Compositori)
- [55] Haah J 2011 Local stabilizer codes in three dimensions without string logical operators *Phys. Rev. A* **83** 042330
- [56] Bravyi S and Haah J 2011 Analytic and numerical demonstration of quantum self-correction in the 3D cubic code arXiv:1112.3252
- [57] Michnicki K 2012 3-d quantum stabilizer codes with a power law energy barrier arXiv:1208.3496