Dartmouth College
Dartmouth Digital Commons

Dartmouth Scholarship | Faculty Work

12-7-2010

Information-Preserving Structures: A General Framework for Quantum Zero-Error Information

Robin Blume-Kohout
Perimeter Institute for Theoretical Physics

Hui Khoon Ng
California Institute of Technology

David Poulin
Université de Sherbrooke

Lorenza Viola
Dartmouth College

Follow this and additional works at: https://digitalcommons.dartmouth.edu/facoa

 Part of the Algebra Commons, and the Quantum Physics Commons

Dartmouth Digital Commons Citation

Blume-Kohout, Robin; Ng, Hui Khoon; Poulin, David; and Viola, Lorenza, "Information-Preserving Structures: A General Framework for Quantum Zero-Error Information" (2010). *Dartmouth Scholarship*. 1920.
https://digitalcommons.dartmouth.edu/facoa/1920

This Article is brought to you for free and open access by the Faculty Work at Dartmouth Digital Commons. It has been accepted for inclusion in Dartmouth Scholarship by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.

Information-preserving structures: A general framework for quantum zero-error information

Robin Blume-Kohout,[1,*] Hui Khoon Ng,[2,†] David Poulin,[3,‡] and Lorenza Viola[4,§]

[1]*Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario N2L 2Y5, Canada*

[2]*Institute for Quantum Information, California Institute of Technology, Pasadena, California 91125, USA*

[3]*Département de Physique, Université de Sherbrooke, Québec J1K 2R1, Canada*

[4]*Department of Physics and Astronomy, Dartmouth College, 6127 Wilder Laboratory, Hanover, New Hampshire 03755, USA*

(Received 27 August 2010; published 7 December 2010)

Quantum systems carry information. Quantum theory supports at least two distinct kinds of information (classical and quantum), and a variety of different ways to encode and preserve information in physical systems. A system's ability to carry information is constrained and defined by the noise in its dynamics. This paper introduces an operational framework, using *information-preserving structures*, to classify all the kinds of information that can be perfectly (i.e., with zero error) preserved by quantum dynamics. We prove that every perfectly preserved code has the same structure as a matrix algebra, and that preserved information can always be corrected. We also classify distinct operational criteria for preservation (e.g., "noiseless," "unitarily correctible," etc.) and introduce two natural criteria for *measurement-stabilized* and *unconditionally preserved* codes. Finally, for several of these operational criteria, we present efficient (polynomial in the state-space dimension) algorithms to find all of a channel's information-preserving structures.

DOI: 10.1103/PhysRevA.82.062306 PACS number(s): 03.67.Pp, 03.67.Lx, 03.65.Yz, 89.70.−a

I. INTRODUCTION

Physical systems can be used to store, transmit, and transform information. Different systems can carry different kinds of information; classical systems can carry classical information, while quantum mechanical systems can carry quantum information. The system's dynamics also affect the kind of information that it carries. For example, decoherence [1] can restrict a quantum system to carry only classical information (or none at all). This suggests that perhaps a quantum system's dynamics can select other kinds of information, neither quantum nor classical, but something in between. The central result of this paper is an exhaustive classification of exactly what kinds of information can be selected in this way.

Preservation of information in physical systems is important in several contexts. In communication theory, information originates with a sender ("Alice") who actively conspires with a receiver ("Bob") to transfer it over a communication channel. Computational devices require memory registers that can store information in the face of repeated noise. Experimental and observational sciences require, in a more or less explicit way, the transmission of information from a passive system of interest (perhaps a distant galaxy, or a nanoscale device), through a chain of ancillary systems, to an observer. In each case, achieving the desired transformation requires first that the information be *preserved* by a noisy dynamical process or "channel"—yet, each operational scenario poses a subtly different notion of "preserved."

In this paper we develop a theory that covers all these situations in a unified framework. We start by establishing a general setting for information (and its preservation), using *codes* (Sec. II). We state a minimal necessary condition for information preservation, then prove that it is also sufficient (in a particular strong sense), deriving a powerful structure theorem for preserved codes (Sec. III). On this foundation, we build a hierarchy of different *operational* criteria for preservation (Sec. IV). Stricter criteria correspond to additional operational constraints—for example, that information persists for more than one application of the noise. Some of these criteria encompass previously studied approaches to information preservation, including pointer states [1], decoherence-free subspaces [2] and noiseless subsystems [3–5], and quantum error-correcting codes [6]. Others—notably "measurement-stabilized" and "unconditionally preserved" codes—have not, as far as we know, been explored previously. Our main contribution is to gather them all into a single framework using *information-preserving structures* (IPSs). IPSs classify the kinds of information that dynamical processes can preserve. In particular, we focus here on *perfect IPS*, corresponding to *zero-error* information. Finally, we consider how to find these structures for a given noisy process (Sec. V). It is NP-hard to find a channel's largest correctible IPS, but for stricter preservation criteria it can be much easier. We provide efficient and exhaustive algorithms to find noiseless, unitarily noiseless, and unconditionally preserved IPSs.

Our IPS framework establishes an explicit and rigorous connection between perfectly preserved information and fixed points of channels. By focusing on fixed points (see also [7]), rather than on the noise commutant, it provides a first step toward understanding *approximate IPS*, making contact with stability results for decoherence-free encodings under symmetry-breaking perturbations [8], and with approximate quantum error correction (QEC) [9–12]. Our structure theorem for the fixed points of completely positive maps extends previous results that apply only to unital processes [13,14], or processes with a full-rank fixed state [15]. Our algorithm

*Present address: Theoretical Division, Mail Stop B258; Los Alamos National Laboratory, Los Alamos, NM 87545; robin@blumekohout.com

†Present address: DSO National Laboratories, and Centre for Quantum Technologies, National University of Singapore, Singapore; cqtnhk@nus.edu.sg

‡david.poulin@usherbrooke.ca

§lorenza.viola@dartmouth.edu

 ©2010 The American Physical Society

for finding noiseless and unitarily noiseless codes improves on algorithms that are inefficient (e.g., Refs. [16,17]), or restricted to purely noiseless information [18] or unital channels [19].

Early aspects of this work appeared in Ref. [20]. Here, we provide more results, full proofs, and detailed discussion.

II. PRESERVED INFORMATION

"What kinds of information can a quantum dynamical process preserve?" is a technical question, but one that requires a firm conceptual foundation. This section aims to provide one. We begin with an operational definition of "information," then apply it to quantum theory. We use well-known results on the accuracy with which quantum states can be distinguished to establish a mathematical framework in which this central question can be answered.

"Information" has a variety of meanings. Any crisp definition will inevitably run afoul of some alternative usage. Throughout *this* paper, we will follow this basic operational definition:

Principle 1. Information is a resource, embodied in a physical system, that can be used to answer a question.

A physical system $\mathcal{S}$ can carry information. If one party (Alice) sends it to another (Bob), then the recipient can use it to answer a question. More precisely, possession of $\mathcal{S}$ gives Bob a higher probability of guessing the correct answer. However, if $\mathcal{S}$ evolves during transmission (i.e., it undergoes a dynamical map $\mathcal{E}$) then some information might be lost. As a result, $\mathcal{E}(\mathcal{S})$ may be less useful than $\mathcal{S}$. It is not yet clear how to determine whether information is "preserved," but two principles seem self-evident:

Principle 2. If nothing happens to a system, then all the information in it is preserved.

Principle 3. If a system evolves as $\mathcal{S} \to \mathcal{E}(\mathcal{S})$, and $\mathcal{E}(\mathcal{S})$ is strictly less useful than $\mathcal{S}$ in answering some question, then some information in $\mathcal{S}$ was not preserved.

These simple criteria bracket the (as-yet undefined) notion of preservation—of *all* the information in a system. But information can be *encoded* into one part of a system. Such information may be preserved even if other parts are damaged or destroyed. To properly represent this notion, we appeal to another self-evident principle:

Principle 4. If some property or parameter of a system is already known to all parties (e.g., Alice and Bob), then it carries no useful information.

For example, if a quantum system $\mathcal{S}$ is *known* to be in the state $|\psi\rangle\langle\psi|$, by all parties, then nothing is gained by transmitting it. Since a known property of $\mathcal{S}$ carries no information, disturbing it has no effect on the information embodied in the system. So, we can represent the encoding of information in a very general way by stating a promise or precondition, which *guarantees* certain properties of $\mathcal{S}$. Those properties, being already known, carry no useful information. Information carried by $\mathcal{S}$ *conditional* on the promise can be preserved, even if other properties (constrained by the promise) are disturbed.

Mathematically, a precondition on $\mathcal{S}$ is a restriction of its state, to some (arbitrary) subset. We call such a set a *code*.

Definition 1. A code $\mathcal{C}$ for a system $\mathcal{S}$ is an arbitrary subset of the system's state space.

A code need not be a finite set, and in fact all the codes that interest us contain uncountably many states (because they comprise convex, dense subregions of state space). However, in many of our examples, we will mention or consider finite codes of the form $\mathcal{C} = \{\rho_k\}$, strictly for simplicity.

Codes carry information. Each system $\mathcal{S}$ has a natural "maximum code" containing all its possible states. Smaller codes for that system carry strictly less information, but may be preserved even when the system's maximum code is not. A code that is a strict subset of another preserved code is uninteresting, so we will focus on *maximal* preserved codes.

Definition 2. A preserved code $\mathcal{C}$ is maximal if and only if there exists no preserved $\mathcal{C}_{\text{big}} \supset \mathcal{C}$. That is, if adding any other state would render $\mathcal{C}$ unpreserved.

We can narrow our focus even more. If $\mathcal{S}$ has two preserved codes, $\mathcal{C}_{\text{big}}$ and $\mathcal{C}_{\text{small}}$, where $\mathcal{C}_{\text{big}}$ is strictly "bigger" than $\mathcal{C}_{\text{small}}$, then we are not interested in $\mathcal{C}_{\text{small}}$. $\mathcal{C}_{\text{big}}$ is "bigger" than $\mathcal{C}_{\text{small}}$ if it has a proper subset that is identical or isomorphic to $\mathcal{C}_{\text{small}}$. We can make this rigorous, but only by borrowing a technical definition from the next section (see Definition 4).

Definition 3. A preserved code $\mathcal{C}$ is maximum if and only if there is no preserved $\mathcal{C}_{\text{big}}$ such that $\mathcal{C}$ is isometric to a strict subset $\mathcal{C}_{\text{small}} \subset \mathcal{C}_{\text{big}}$.

We will generally restrict our attention to maximum codes.[1] We need a precise definition of a "preserved" code. We begin by adapting Principles 2 and 3 to codes.

Principle 5. The information in a code $\mathcal{C}$ is preserved by a dynamical map $\mathcal{E}$ if $\mathcal{E}$ leaves every state in $\mathcal{C}$ unchanged.

Principle 6. The information in a code $\mathcal{C}$ is preserved by a dynamical map $\mathcal{E}$ only if $\mathcal{E}(\mathcal{C})$ is as useful as $\mathcal{C}$ for answering any question.

These are sufficient and necessary (respectively) *operational* conditions for preservation. Principle 6 seems much weaker than 5, but we will show that it is actually not. If Principle 6 is satisfied; then there is a physically implementable *recovery operation* that restores every code state. The ability to perform this recovery is a resource—a reasonable one, but a nontrivial one. We will also consider several weaker resources (e.g., restrictions on what recovery operations can be implemented), and the corresponding stronger notions of preservation in Sec. IV.

This concludes the "philosophical" part of our framework, and in what follows we will build on these foundations to establish technical results. Two final points deserve mention, however.

(i) Identifying "information" with codes (arbitrary sets of states) is intended to be a very general paradigm. A system's state, by definition, specifies everything that can be known about that system. Every question that can be answered using $\mathcal{S}$ boils down to a question about the state of $\mathcal{S}$, and variations in that state (restricted to some particular code) encode information. If there are exceptions to this rule—that

[1]Graph theorists may recognize this terminology. Maximal and maximum codes have the same relationship as maximal and maximum cliques, or independent sets. Note, however, that unlike a graph, a channel need *not* have a unique maximum code. If a channel preserves *either* a quantum bit *or* a classical trit, they are incomparable—neither is bigger than the other.

is, notions of information, consistent with Principle 1, that cannot be represented using codes—then we are not aware of them.[2] An extended discussion can be found in Appendix A2.

(ii) Our definition of "information" may not appear congruent with Shannon's theory of communication [21,22]. In fact, it is quite compatible. There are, however, some subtle differences: As mentioned, we focus on zero-error information; furthermore, we consider a *single* use of a communication channel, rather than N uses with $N \to \infty$. An extended discussion can be found in Appendix A1.

A. Systems, states, codes, and channels in quantum theory

So far, we have used a language consistent with a broad range of physical theories. Let us now specialize to quantum theory. States of quantum systems are represented by density operators ρ, which are positive trace-1 operators on the system's Hilbert space $\mathcal{H}$. Quantum dynamical maps (also known as *channels*) are described by completely positive (CP), trace-preserving (TP) linear maps on density operators. A CPTP map $\mathcal{E}$ can be represented in two equivalent ways. In one formulation, the initial system $\mathcal{S}_A$ comes into contact with an *uncorrelated* environment E_0, they evolve unitarily, and then some part E_f of this joint system is discarded[3], yielding a reduced state for the final system $\mathcal{S}_B$:

$$\rho_B = \mathcal{E}(\rho_A) = \mathrm{Tr}_{E_f}\left[U\left(\rho_A \otimes \rho_{E_0}\right)U^\dagger\right]. \tag{1}$$

The other representation of a CP-map is called the operator-sum representation:

$$\rho_B = \mathcal{E}(\rho_A) = \sum_i K_i \rho_A K_i^\dagger, \tag{2}$$

where the *Kraus operators* $\{K_i\}$ satisfy $\sum_i K_i^\dagger K_i = \mathbb{1}$. This representation is mathematically simpler but less physically intuitive (for a complete treatment of CP maps, see Refs. [24,25]). Note that in either representation, $\mathcal{S}_A$ and $\mathcal{S}_B$ may be different systems, with different Hilbert spaces. However, the special case where they are the same is very important—for instance, all continuous-time processes are described by such maps—and we will often implicitly assume it, dropping A and B subscripts and relying on context to illustrate whether "$\mathcal{S}$" refers to the channel's input or its output.

Codes for quantum systems are sets of quantum states. The code $\mathcal{C}$ represents a promise that the system will be prepared in some $\rho \in \mathcal{C}$. Each distinct code represents a potentially distinct kind of information. Note, however, that we are not introducing an infinite proliferation of fundamentally different "kinds" of information, nor are we suggesting that a qubit carries fundamentally different information from a qutrit: Systems with isomorphic state spaces carry the same kind of information. N qutrits equal $N \log_2 3$ qubits, so they carry the same kind of information, but more of it. The important dividing line is between systems that have no asymptotic equivalence, like a qubit and a classical bit.[4]

Now that we have a well-defined mathematical theory, we need a mathematical definition of preservation. Principle 6 uses the very general idea of "questions." A simple and well-defined set of questions turns out to be sufficient: "Was the system prepared in state ρ or state σ?" Here, ρ and σ are states in the code $\mathcal{C}$. In general, these questions cannot be answered with certainty, for most pairs of states are not perfectly distinguishable. But if Bob cannot distinguish them as well as Alice, then information has been lost. Of course, there may well be many other questions that *could* be asked, but it turns out that if *these* well-defined questions are all preserved, then the code can be corrected (and therefore *every* question must be preserved!)

Example 1. Suppose that $\mathcal{S}$ is a quantum bit. If its dynamics are noiseless, then every state passes unchanged through the channel. We can describe the preserved information in terms of a code $\mathcal{C}_{\text{qubit}}$ that contains all the possible states for a qubit. Now, suppose $\mathcal{S}$ experiences a dephasing channel, which transforms an arbitrary superposition of the computational states $|0\rangle$ and $|1\rangle$ into a mixture,

$$\mathcal{E} : \alpha|0\rangle + \beta|1\rangle \longrightarrow |\alpha|^2|0\rangle\langle 0| + |\beta|^2|1\rangle\langle 1|,$$

and which maps the Bloch sphere into itself as in Fig. 1.

The code $\mathcal{C}_{\text{qubit}}$ is no longer preserved. Because the two states $|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}}$ are both mapped to $\rho_B = \frac{1}{2}\mathbb{1}$, Bob cannot answer the question "Was $\mathcal{S}$ prepared in $|+\rangle$ or $|-\rangle$?" However, the more restricted code $\mathcal{C}_{\text{cbit}} = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ is preserved, for Bob can distinguish between these states just as well as Alice. The preserved code describes a different kind of information: one classical bit.

Here are some familiar examples of preserved information, represented as codes.

Example 2. A pointer basis comprises a set of mutually orthogonal "pointer states" $\{|\psi_1\rangle, \ldots, |\psi_N\rangle\}$ that are unaffected

[2] A simple and important example is entanglement between $\mathcal{S}$ and a reference system $\mathcal{R}$. Though not *explicitly* mentioned, entanglement is easy to characterize in our setting. If $\mathcal{S}$ and $\mathcal{R}$ are maximally entangled, then $\mathcal{S}$ can be postselectively prepared in any pure state $|\psi\rangle\langle\psi|$ by projecting $\mathcal{R}$ into some $|\psi'\rangle\langle\psi'|$. Entanglement is preserved if and only if the code containing *all* of these conditional states is preserved.

[3] A technical note is in order here. If the environment E_0 is initially correlated with the input system $\mathcal{S}_A$, then the resulting dynamics is generally *not* CP, and so initial decorrelation is a common assumption in the theory of open quantum systems. For our purposes, it is more than just an assumption. If $\mathcal{S}_A$ is initially correlated with its environment, then the latter contains information about $\mathcal{S}_A$. The system and its environment *together* may contain more information about $\mathcal{S}_A$ than does $\mathcal{S}_A$ itself! In the course of the ensuing interaction, that information may flow back into the system. It is impossible (ill-defined, even) to say whether information in $\mathcal{S}_A$ has been preserved in such a case, for it may have been replaced with information initially residing in E_0. Such an interaction is not, in any sense, "noise." A contrary viewpoint is put forth in Ref. [23], however, which proposes and analyzes error correction for non-CP maps.

[4] Two systems $\mathcal{S}_A$ and $\mathcal{S}_B$ have an asymptotic equivalence if there is a constant R such that for all $\epsilon > 0$ and $N \to \infty$, (i) $N(R-\epsilon)$ copies of $\mathcal{S}_A$ are strictly less powerful than N copies of $\mathcal{S}_B$, and (ii) $N(R+\epsilon)$ copies of $\mathcal{S}_A$ are strictly more powerful that N copies of $\mathcal{S}_B$. Thus, any two finite nontrivial quantum systems have an asymptotic equivalence in this sense.

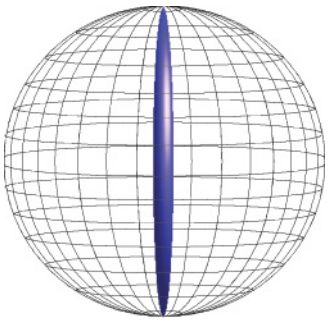


FIG. 1. (Color online) Image of the Bloch sphere under a dephasing channel.

(or "least affected") by noise—as originally introduced in the study of quantum measurement and decoherence [1]. A pointer basis can be described by the code containing all the pointer states (PSs) $|\psi_k\rangle\langle\psi_k|$ and their convex combinations. Classical information is stored in the index k, but not quantum information, because superpositions are not preserved, and thus cannot be included in the code. PSs are preserved in the strongest possible sense: Every state in the code is a fixed point of $\mathcal{E}$.

Example 3. A decoherence-free subspace (DFS) is an entire subspace of the system's Hilbert space, $\mathcal{P} \subseteq \mathcal{H}$, which is invariant under the noise [2] (see also Zurek's prior discussion of "pointer subspaces" [26]). The corresponding code $\mathcal{C}$ contains every density operator supported on $\mathcal{P}$. Since $\mathcal{C}$ includes superpositions of any given basis for $\mathcal{P}$, a DFS preserves quantum information, and can in principle support encoded quantum computation. Like pointer bases, DFSs are preserved in the strongest sense (although the definition is commonly relaxed to allow unitary evolution, especially in the case of Markovian dynamics; see also [27,28]).

Example 4. A noiseless subsystem (NS) is like a DFS in that it can store quantum information. Unlike a DFS, an NS can exist even if no pure state in $\mathcal{H}$ is invariant. According to the original definition [3,4], it suffices that the noise has a trivial action on a "factor" of $\mathcal{H}$. That is, $\mathcal{S}$ supports an NS if there exists a subspace $\mathcal{H}_{AB} \subseteq \mathcal{H}$ that can be factored as $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, so that for every pair of states ρ_A, ρ_B supported on $\mathcal{H}_A$, $\mathcal{H}_B$, respectively,

$$\mathcal{E}(\rho_A \otimes \rho_B) = \rho_A \otimes \rho'_B, \tag{3}$$

for some state ρ'_B on $\mathcal{H}_B$. Thus, the restriction of $\mathcal{E}$ to $\mathcal{H}_{AB}$ obeys

$$\mathcal{E} = \mathbb{1}_A \otimes \mathcal{E}_B, \tag{4}$$

for some CPTP map $\mathcal{E}_B$ on $\mathcal{H}_B$. Since, for every state ρ_{AB} supported on $\mathcal{H}_{AB}$,

$$\mathrm{Tr}_B \mathcal{E}(\rho_{AB}) = \mathrm{Tr}_B \rho_{AB}, \tag{5}$$

it is clear that quantum information is preserved in the reduced state of subsystem A. However, it is not immediately obvious that (as in Examples 2 and 3) there is a corresponding fixed code for $\mathcal{S}$. In fact, the existence of such a code follows from Eq. (4) and the fact that every channel $\mathcal{E}_B$ has at least one fixed state τ_B [29]. Thus, the code $\mathcal{C}_{\mathrm{NS}} = \{\rho_A \otimes \tau_B, \forall\ \rho_A\}$, where ρ_A is arbitrary on $\mathcal{H}_A$, but τ_B is a fixed point of $\mathcal{E}_B$, is invariant under $\mathcal{E}$.

Example 5. Information in a quantum error correcting code (QECC) [6,30] is also preserved, but in a weaker sense than the information in a NS, DFS, or pointer basis. A QECC is a subspace $\mathcal{P}$ for which there exists a physical recovery operation $\mathcal{R}$ so that $(\mathcal{R} \circ \mathcal{E})(|\psi\rangle\langle\psi|) = |\psi\rangle\langle\psi|$ for all $|\psi\rangle \in \mathcal{P}$. As with a DFS, the corresponding "correctable code" contains all states supported on $\mathcal{P}$. Unlike the previous examples, this code is not fixed under $\mathcal{E}$. However, it is clearly preserved, because $\mathcal{P}$ can be turned into a DFS by applying $\mathcal{R}$. An "operator QECC" [31] is an NS for $\mathcal{R} \circ \mathcal{E}$. Another variant stipulates active intervention before the noise occurs [3], in which case the code is "protectable" rather than correctable [18]. While protectable codes will not be further discussed in the present work, the notions of protectability and correctability are not fundamentally different and may, to a large extent, be viewed as "dual" to one another, as elucidated in [11].

The previous examples are not exhaustive, but they illustrate the diversity of criteria for "preserved" information. Each example is specified by a different algebraic condition, dictated either by operational constraints or by its relevance to the task at hand. We hope that unifying them will bring clarity to experimental implementations of these ideas [32–34].

The key point of our framework, though, is to explore *beyond* these well-known examples. In particular, all the situations illustrated above can be described intuitively as "quantum information" or "classical information." What we would like to know is whether more exotic codes are possible—whether some weird channel can preserve a form of information that is entirely unlike a pointer basis, NS, or QECC. We need a rigorous criterion for preservation of codes, based on Principles 5 and 6. Principle 5 is straightforward, but Principle 6 refers to *any* operational task. Our strategy will be to identify one particular task—distinguishing between code states. Because we focus on just one task, we will obtain a *necessary* condition. Having done so, our next challenge will be to unify the necessary condition implied by Principle 6 with the sufficient condition implied by Principle 5.

B. Single-shot distinguishability, Helstrom's theorem, and the 1-norm

Suppose that Bob has access to a single copy of the system $\mathcal{S}$, and he wishes to guess whether it was prepared in state ρ or state σ (both of which are in $\mathcal{C}$). He seeks to maximize the probability that his guess is correct, and he knows that the prior probabilities of ρ and σ are (respectively) p and $(1-p)$. He can measure $\mathcal{S}$ to help him decide, and his optimal course of action is determined by Helstrom's theorem [35].

Helstrom's theorem. Suppose a quantum system $\mathcal{S}$ was prepared either in state ρ or in state σ, with respective probabilities p and $(1-p)$. The highest probability of guessing correctly which was prepared is obtained by measuring the Hermitian operator $\Delta_p = p\rho - (1-p)\sigma$, then guessing "$\rho$" upon obtaining a result corresponding to a positive eigenvalue and "σ" in the case of a negative eigenvalue. If a zero eigenvalue is obtained, either guess is equally good. The success probability is given by $P_H(\rho,\sigma;p) = \frac{1}{2}(1 + \|\Delta_p\|_1)$, where $\|\cdot\|_1$ refers to the 1-norm, $\|A\|_1 \equiv \mathrm{Tr}|A| = \mathrm{Tr}\sqrt{A^\dagger A}$.

The success probability P_H is a measure of the *distinguishability* between ρ and σ. It is nonincreasing under any CPTP

map, because the 1-norm is contractive under CPTP maps [36]. So, in order for $\{\mathcal{E}(\rho),\mathcal{E}(\sigma)\}$ to be as distinguishable as $\{\rho,\sigma\}$, we require that for every prior probability p, the Helstrom strategy yields the same success probability for distinguishing ρ from σ as for distinguishing $\mathcal{E}(\rho)$ from $\mathcal{E}(\sigma)$:

$$P_H[\mathcal{E}(\rho),\mathcal{E}(\sigma);p] = P_H(\rho,\sigma;p).$$

If Bob needs to distinguish between two *sets* of states, $\{\rho_k\}$ and $\{\sigma_k\}$, he assigns prior probabilities $\{p_k\}$ and $\{s_k\}$ to the $\{\rho_k\}$ and $\{\sigma_k\}$, respectively. Then his task is to distinguish

$$\rho = \frac{1}{\sum_k p_k}\sum_k p_k\rho_k$$

from

$$\sigma = \frac{1}{\sum_k s_k}\sum_k s_k\sigma_k,$$

where the prior probabilities of ρ and σ are, respectively, $p = \sum_k p_k$ and $1-p = \sum_k s_k$.

This measure of distinguishability is, in fact, a metric on the space of linear operators. Its preservation implies a kind of rigid equivalence, which we make precise with the following definition.

Definition 4. Two codes $\mathcal{C}_1$ and $\mathcal{C}_2$ are 1-isometric (or just "isometric") to each other if and only if there exists a linear 1:1 mapping $f:\mathcal{C}_1\to\mathcal{C}_2$ such that, for all ρ,σ in the convex closure of $\mathcal{C}_1$ and all $p\in[0,1]$,

$$\|pf(\rho)-(1-p)f(\sigma)\|_1 = \|p\rho-(1-p)\sigma\|_1.$$

Definition 5. A code $\mathcal{C}$ is 1-isometric (or just "isometric") for a CPTP process $\mathcal{E}$ only if $\mathcal{C}$ is isometric to $\mathcal{E}(\mathcal{C})$.

So, if $\mathcal{C}$ is isometric for a given map $\mathcal{E}$, then $\|p\mathcal{E}(\rho)-(1-p)\mathcal{E}(\sigma)\|_1 = \|p\rho-(1-p)\sigma\|_1$ for all ρ,σ in the convex closure of $\mathcal{C}$ and $p\in[0,1]$. A stronger characterization is given by the following:

Definition 6. A code $\mathcal{C}$ is fixed by a CPTP channel $\mathcal{E}$ if and only if $\mathcal{E}(\rho)=\rho$ for all $\rho\in\mathcal{C}$.

C. Criteria for preservation

We are now in a position to state Principle 5 more precisely.

Strong condition for preservation. A sufficient condition for $\mathcal{C}$ to be preserved by $\mathcal{E}$ is that $\mathcal{C}$ be fixed by $\mathcal{E}$.

The strong condition is obviously sufficient, but (as demonstrated by error-correcting codes) it is not *necessary* for preservation. Principle 6 implies a host of necessary conditions—one for every operational task. We choose one in particular: We demand that $\mathcal{E}(\rho)$ and $\mathcal{E}(\sigma)$ be just as distinguishable[5] as ρ and σ. We also require that questions like "Was $\mathcal{S}$ prepared in one of the states $\{\rho_1,\rho_2,\rho_3\ldots\}$, or in one of the states $\{\sigma_1,\sigma_2,\sigma_3\ldots\}$?" should be preserved as well, so convex combinations of code states should maintain their pairwise distinguishability. There is nothing inherently special about this particular operational task, except that it produces a useful and convenient mathematical condition:

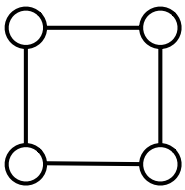
Weak condition for preservation. A necessary condition for $\mathcal{C}$ to be preserved by $\mathcal{E}$ is that $\mathcal{C}$ be isometric for $\mathcal{E}$.

These two criteria form the foundation of our framework. To illustrate their application, here are some examples both simple and subtle.

Example 6. Suppose $\mathcal{S}$ is a classical system with four states labeled $\{0,1,2,3\}$, each perfectly distinguishable from the others. $\mathcal{S}$ passes through a channel that maps state k randomly to k or $k+1$ (mod 4), represented as a stochastic map,

$$\mathcal{E} = \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

A stochastic map's information-preserving properties can conveniently be represented by an adjacency graph for the input states, where state j is connected to state k if $\mathcal{E}(j)$ overlaps with $\mathcal{E}(k)$. This map's adjacency graph is as follows:



The code $\mathcal{C}_4=\{0,1,2,3\}$ representing all information about $\mathcal{S}$ is not preserved, because 0 and 1 are perfectly distinguishable, but $\mathcal{E}(0)$ and $\mathcal{E}(1)$ overlap. A smaller code $\mathcal{C}_2=\{0,2\}$ is preserved, even though neither 0 nor 2 is a fixed point. The code $\mathcal{C}_2'=\{1,3\}$ is also preserved, but the union of $\mathcal{C}_2$ and $\mathcal{C}_2'$ is not preserved. This demonstrates that the set of preserved codes is not convex; distinct preserved codes may rely on mutually contradictory preconditions on $\mathcal{S}$ (e.g., "$\mathcal{S}$ was prepared in 0 or 2" and "$\mathcal{S}$ was prepared in 1 or 3").

Example 7. Why must distinguishability be preserved, not just between code states, but between convex combinations of them?

Let $\mathcal{E}$ be a classical stochastic map on three states $\{0,1,2\}$, which fixes states 0 and 1, but maps $2\to1$. This map "squashes" the classical 3-simplex onto one of its sides, as in Fig. 2. Now, consider a code $\mathcal{C}$ comprising the states on the thick (red) line in Fig. 2.

This code is not preserved by $\mathcal{E}$, because the original code has structure that is missing in its image: States not on the line between "0" and "1" can be unambiguously discriminated (with $p>0$) from states lying on the line. However, there is no way to recover this structure by applying another linear map afterward! Still, if we ignore convex combinations, then all the 1-norm distances $\|p\rho-(1-p)\sigma\|_1$, for $\rho,\sigma\in\mathcal{C}$ are in fact preserved by $\mathcal{E}$. This is because the best way to distinguish any two states in $\mathcal{C}$ is to measure 0 versus $\{1,2\}$, and because the channel maps $2\to1$, it does not actually affect this measurement. If we consider convex combinations, however, we see that $\mathcal{C}$ is not isometric to $\mathcal{E}(\mathcal{C})$, resolving the problem.

[5]Note that ρ and σ need not be perfectly distinguishable to start with. A QECC contains nonorthogonal states that cannot be perfectly distinguished, but they can be distinguished just as well after $\mathcal{E}$ as before.

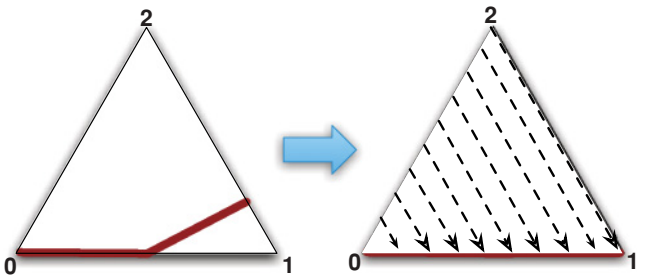


FIG. 2. (Color online) Action of a "squashing" channel on a classical trit, and a code (thick red line) that is not preserved even though all 1-norm distances between code states remain fixed.

Example 8. Why must all the weighted 1-norm distances be preserved, rather than just $\|\rho-\sigma\|_1$?

Consider a channel $\mathcal{E}$ acting on a qutrit Hilbert space $\mathbb{C}^3$, which does nothing to the $\{|0\rangle,|1\rangle\}$ subspace, but maps $|2\rangle\langle 2| \to \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$. Now, consider a code $\mathcal{C}$ comprising all the states of the form,

$$\rho = \tfrac{1}{2}(|\psi\rangle\langle\psi|_{\mathrm{Span}(|0\rangle,|1\rangle)} + |2\rangle\langle 2|).$$

We can think of this code as the set of states that would be prepared by a machine that is supposed to produce qubit states in the $\{|0\rangle,|1\rangle\}$ subspace, but fails 50% of the time and produces $|2\rangle\langle 2|$ instead.

As in Example 7, this code is not preserved by $\mathcal{E}$. In this case, the problem is that Alice can check to see whether the preparation failed or not, but Bob cannot. As before, this intuition is borne out by the fact that no recovery operation exists. However, if we compute the unweighted 1-norm distances $\|\rho-\sigma\|_1$, both before and after $\mathcal{E}$ is applied, then we find that they are unchanged. Only when we require preservation of the weighted 1-norm distances (corresponding to distinguishing states with the aid of prior information), do we correctly derive that $\mathcal{C}$ is not preserved.

As Example 7 demonstrates, it is important that $\mathcal{E}$ preserves distinguishability not just between states in $\mathcal{C}$, but between convex combinations of them. This means that we can (without loss of generality) *extend $\mathcal{C}$ to include all states in its convex closure*. From now on, we will simply assume that any preserved code is convex in this sense, as in Ref. [20]. The weak condition then has a simple geometric interpretation: $\mathcal{E}$ must preserve the 1-norm distance between any two *unnormalized* states $p\rho$ and $(1-p)\sigma$. This means that the entire convex cone of $\mathcal{C}$—that is, the set $\mathcal{C}_+$ containing $x\rho$ for all $x \geqslant 0$ and $\rho \in \mathcal{C}$—must be *isometric* to its image $\mathcal{E}(\mathcal{C}_+)$. Two sets are isometric if there is a distance-preserving mapping (an isometry) between them. Here, the relevant metric is the 1-norm distance,

$$D(A,B) \equiv \|A-B\|_1,$$

and $\mathcal{E}$ is the isometry that preserves it. Thus, preservation requires that the convex cone $\mathcal{C}_+$ evolves *rigidly*, with respect to the 1-norm distance, under $\mathcal{E}$.

Our necessary and sufficient conditions bracket the as-yet-vague notion of a code being preserved by a channel. Fixedness seems too strong, isometry perhaps too weak. One of our main goals in this paper is to derive a single, rigorously stated condition for information to be "preserved" by a channel. We will eventually do so by squeezing the strong and weak conditions together as follows:

Proposition 1. If $\mathcal{C}$ is a maximum isometric code for $\mathcal{E}$ (i.e., it satisfies the weak condition, and there is no larger $\mathcal{C}$ that satisfies the weak condition), then there exists a CPTP map $\mathcal{R}$ such that $\mathcal{R}\circ\mathcal{E}(\rho)=\rho$ for all states $\rho \in \mathcal{C}$.

By proving this proposition, we will demonstrate that the strong and weak conditions for preservation are equivalent—*given* the ability to apply a recovery operation. The proof is somewhat involved. In the next section, we will derive a structure theorem for preserved codes, explore its consequences, and finally derive Proposition 1 as a corollary (Corollary 7) of Lemma 6, which follows from Theorem 1. Anticipating this sequence of derivations, we proffer the following definition of "preserved" now, with the understanding that it will only be justified by what follows:

Definition 7. A code is preserved by a CPTP $\mathcal{E}$ if and only if it satisfies the weak condition, that is,

$$\|\mathcal{E}[p\rho-(1-p)\sigma]\|_1 = \|p\rho-(1-p)\sigma\|_1,$$

for all $\rho,\sigma \in \mathcal{C}$ and $p \in [0,1]$.

III. THE STRUCTURE OF PRESERVED INFORMATION

In Sec. II, we stated plausible necessary and sufficient conditions for a code to be "preserved," and suggested a formal definition of preservation (conditional on some technical results to be proved in what follows). Next, we shall build upon this foundation, elucidating the structures that follow from it. First, we will prove a series of theorems about preserved codes, culminating in a structure theorem showing that preserved codes have the same "shape" as matrix algebras. This indicates that preserved codes are related to algebras, but provides no real context for *how* they are related, nor what role the algebra is playing. So, our second task is to analyze the underlying IPS.

Except where explicitly noted, all the proofs of theorems and lemmas in this section have been deferred to Appendix B.

A. The shape of a preserved code

Suppose that $\mathcal{C}$ is a preserved code for $\mathcal{E}$. Starting from Definition 7, what can we derive about $\mathcal{C}$? Quite a lot, as it turns out. The following two definitions from Ref. [20] will be needed.

Definition 8. A code $\mathcal{C}$ is noiseless for a CPTP map $\mathcal{E}$ if and only if it is preserved by any convex combination of powers of $\mathcal{E}$, $\sum_n q_n\mathcal{E}^n$, with $q_n \geqslant 0$ and $\sum_n q_n = 1$.

Noiselessness is stricter than preservation (every noiseless code is preserved, but many preserved codes are not noiseless), but weaker than fixedness (every fixed code is noiseless, but some noiseless codes are not fixed). Noiseless codes are special because their states remain distinguishable no matter how many times $\mathcal{E}$ is applied (note that only channels whose output space is the same as their input space can have noiseless codes). This captures the operational significance of fixedness—and as we will show below (Lemma 2), there is a close mathematical connection between noiseless and fixed codes.

Definition 9. A code $\mathcal{C}$ is correctable for $\mathcal{E}$ if and only if there exists a CPTP $\mathcal{R}$ such that $\mathcal{C}$ is noiseless for $\mathcal{R}\circ\mathcal{E}$.

Correctable codes can be *made* noiseless, by applying a suitable correction operation every time $\mathcal{E}$ happens. Readers familiar with QEC may worry that our definition is slightly different from the usual one, which requires that $\mathcal{C}$ be *fixed* by $\mathcal{R}\circ\mathcal{E}$, rather than just noiseless. It will turn out that our (apparently weaker) condition implies the usual one, so we obtain the same result with a weaker assumption.[6] We are now in a position to state a key theorem:

Theorem 1. A [convex] code $\mathcal{C}$ is correctable for $\mathcal{E}$ if and only if it is preserved by $\mathcal{E}$.

Although the full proof is rather technical (see Appendix B), one aspect is especially useful and interesting. We prove the theorem by *explicitly* constructing a correction operation for an arbitrary code $\mathcal{C}$. Moreover, the correction operation is independent of $\mathcal{C}$'s structure, and depends only on $\mathcal{C}$'s support. A code's support is the subspace $\mathcal{P}\subseteq\mathcal{H}$, comprising the union of the supports of all $\rho\in\mathcal{C}$. Since the correction only depends on the code's support, every code with the same support will be corrected by the same operation. Remarkably, this operation coincides with the *transpose channel* introduced in Ref. [37], defined as

$$\hat{\mathcal{E}}_{\mathcal{P}}=\Pi\circ\mathcal{E}^{\dagger}\circ\mathcal{N}, \tag{6}$$

where P is the projector onto $\mathcal{P}$, $\Pi(\cdot)=P\cdot P$ is the projection onto $\mathcal{P}$, $\mathcal{E}^{\dagger}$ is the adjoint map of $\mathcal{E}$, and $\mathcal{N}$ is a normalization map $\mathcal{N}(\cdot)=\mathcal{E}(P)^{-1/2}(\cdot)\mathcal{E}(P)^{-1/2}$.

Theorem 1 has two consequences. First, it strongly suggests that Definition 7 captures the critical notions of information preservation. Second, it implies a simple corollary: Every preserved code for $\mathcal{E}$ is noiseless for some other map $\mathcal{R}\circ\mathcal{E}$. This connection from preserved to noiseless codes is a step toward proving Proposition 1. Even more importantly, it will let us derive a structure theorem for preserved codes. To do so, we need another result.

Lemma 2. Every noiseless code $\mathcal{C}$ for $\mathcal{E}$ is isometric to a set of states that are fixed points of $\mathcal{E}$.

This means that noiseless and fixed codes are geometrically equivalent. A noiseless code does not have to be precisely fixed, but it will always be isometric to a fixed code, that is, it will have the same shape. A simple example may be in order.

Example 9. Let $\mathcal{E}$ be a channel on two qubits, labeled A and B, that does nothing to A but depolarizes B:

$$\mathcal{E}(\rho_{AB})=\mathrm{Tr}_B(\rho_{AB})\otimes\frac{\mathbb{1}_B}{2}.$$

Qubit A clearly is an NS under $\mathcal{E}$, whose fixed states are of the form $\mathcal{C}_{\mathrm{NS}}=\rho_A\otimes\frac{\mathbb{1}_B}{2}$. However, there are other noiseless codes. For instance, let $\mathcal{C}$ comprise all states of the form $\rho_A\otimes|0\rangle\langle 0|_B$. Qubit B carries no information, so $\mathcal{E}$'s action on it is irrelevant. None of $\mathcal{C}$'s distinguishability properties are affected by $\mathcal{E}$, even though $\mathcal{C}$ is not actually fixed. Note, however, that $\mathcal{C}$'s image $\mathcal{E}(\mathcal{C})$ is a fixed code. Repeated applications of $\mathcal{E}$ map its noiseless codes to fixed codes.

[6] In the terminology of Ref. [11], a code $\mathcal{C}$ which is fixed by $\mathcal{R}\circ\mathcal{E}$ is referred to as "completely correctable." That complete correctability is, in fact, equivalent to correctability can be alternatively established by exploiting the explicit form of 1-isometric encodings; see Theorem 4 therein.

Lemma 2 implies that a channel has a unique maximum (largest) noiseless code, which is isometric to the set of *all* fixed states:

Corollary 3. Every maximum noiseless code for a channel $\mathcal{E}$ is isometric to the full fixed-point set of $\mathcal{E}$.

A channel can have smaller noiseless codes—even maximal ones. Consider the following example.

Example 10. Let $\mathcal{E}$ be a channel on two qubits, labeled A and B, which acts as follows: It measures B in the $\{|0\rangle,|1\rangle\}$ basis; conditional on $|0\rangle\langle 0|$ it does nothing; conditional on $|1\rangle\langle 1|$, it dephases A and flips B to the $|0\rangle\langle 0|$ state. Every state of the form $\rho_A\otimes|0\rangle\langle 0|_B$ is a fixed point, and so the largest noiseless code encodes a single qubit in A, like in Example 9. However, there is another maximal noiseless code comprising all states of the form $[p|0\rangle\langle 0|+(1-p)|1\rangle\langle 1|]_A\otimes|1\rangle\langle 1|_B$. It is isometric to a strict subset of the fixed points, so it is not a maximum code.

Recall that any preserved code can be made noiseless, by applying a suitable recovery map (Theorem 1). By combining this theorem with Corollary 3, we establish a direct connection between arbitrary preserved codes and fixed states of CPTP maps.

Theorem 4. Every maximum preserved code for a CPTP map $\mathcal{E}$ is 1-isometric to the full set of fixed states for some other CPTP map $\mathcal{R}\circ\mathcal{E}$.

Proof. This follows from combining Lemma 2 with Theorem 1 and Definition 9. ∎

This points the way to the structure theorem we are looking for, provided that we can say something about the fixed points of the unknown CPTP map $\mathcal{R}\circ\mathcal{E}$. Quite a bit is known about fixed points of CPTP maps. In particular, if $\mathcal{H}$ is finite dimensional, and the map is *unital* (meaning that it preserves the identity operator), then its fixed points form a matrix algebra [13,14].

A matrix algebra (a.k.a. finite-dimensional C^* algebra) is a vector space of complex matrices, closed under multiplication and Hermitian conjugation. It follows that

(1) The matrices must be square (otherwise they cannot be multiplied);

(2) The set of *all* $d\times d$ complex matrices (i.e., operators on a d-dimensional Hilbert space $\mathcal{H}$) is an algebra, denoted $\mathcal{M}_d$ or $\mathcal{M}_{\mathcal{H}}$ henceforth;

(3) The set containing only the $d\times d$ identity matrix is an algebra, denoted $\mathbb{1}_d$ or $\mathbb{1}_{\mathcal{H}}$.

Happily, these three simple facts are sufficient to describe *any* matrix algebra. The structure theorem [38] for matrix algebras states that any such matrix algebra $\mathcal{A}$ is unitarily equivalent to the canonical form:

$$\mathcal{A}\simeq\bigoplus_k\mathcal{M}_{A_k}\otimes\mathbb{1}_{B_k}, \tag{7}$$

where A_k and B_k are complex vector spaces of dimension d_k and n_k, respectively. We will refer to each of the subspaces $A_k\otimes B_k$ in the direct sum labeled by k as a "k sector." Each k sector factors into a *noiseless subsystem* (with Hilbert space A_k) and a *noise-full subsystem* (with Hilbert space B_k).[7] Thus, every matrix algebra is built up out of the two simple

[7] Note that in the original definition of [3], a decomposition of the form given in Eq. (7) is applied to the (associative) *error algebra* as

components described in points 2 and 3 in the previous list (the algebra of all $d \times d$ matrices, and the trivial algebra).

As remarked earlier, the fixed points of a unital map form an algebra. Prior to this work (and the results anticipated in [20]), no such result was known for *arbitrary* nonunital maps. Before stating our main structure theorem, we need to define a couple of terms.

Definition 10. Consider a matrix algebra $\mathcal{A} = \bigoplus_k \mathcal{M}_{A_k} \otimes \mathbb{1}_{B_k}$, which induces a Hilbert space decomposition $\mathcal{H} = \bigoplus_k A_k \otimes B_k$. A *distortion map* for $\mathcal{A}$ is a CPTP map $\mathcal{D}$ such that, for every $X = \sum_k M_{A_k} \otimes \mathbb{1}_{B_k}$ in $\mathcal{A}$,

$$\mathcal{D}(X) = \sum_k M_{A_k} \otimes \tau_k,$$

where τ_k is a positive semidefinite matrix on B_k that does not depend on M_{A_k}. $\mathcal{D}(\mathcal{A})$ is a *distortion* of $\mathcal{A}$. A vector space of matrices $\tilde{A}$ is a *distorted algebra* if it is a distortion of some matrix algebra $\mathcal{A}$.

A distorted algebra is simply an algebra in which each identity factor has been replaced with an arbitrary (but fixed) matrix τ_k in each k sector. A distorted algebra is not an algebra under standard matrix multiplication (because $\tau_k^2 \neq \tau_k$), although it is under a suitably redefined matrix multiplication. More importantly, there exist CP distortion maps that reversibly transform $\tilde{\mathcal{A}} \leftrightarrow \mathcal{A}$, simply by changing the τ_k factors. Thus, $\tilde{\mathcal{A}}$ and $\mathcal{A}$ are isometric.

We can now characterize the fixed points of an arbitrary CPTP map *and* its adjoint (that is, fixed states *and* observables).

Theorem 5. Let $\mathcal{E}$ be a CPTP map on $\mathcal{B}(\mathcal{H})$, and $\mathcal{E}^\dagger$ its adjoint. Let $\mathrm{Fix}(\mathcal{E})$ be the fixed points of $\mathcal{E}$, and $\mathrm{Fix}(\mathcal{E}^\dagger)$ the fixed points of $\mathcal{E}^\dagger$. Then,

(i) Let $\mathcal{P}_0 \subseteq \mathcal{H}$ be the support of $\mathrm{Fix}(\mathcal{E})$. Then $\mathcal{P}_0$ is an invariant subspace under $\mathcal{E}$.

(ii) Let $\mathcal{E}_{\mathcal{P}_0}$ be the restriction of $\mathcal{E}$ to $\mathcal{P}_0$, so $\mathcal{E}_{\mathcal{P}_0} \equiv \Pi_0 \circ \mathcal{E} \circ \Pi_0$, where Π_0 projects onto $\mathcal{P}_0$. Then the fixed points of $\mathcal{E}^\dagger_{\mathcal{P}_0}$ form a matrix algebra $\mathcal{A}$.

(iii) $\mathrm{Fix}(\mathcal{E})$ is a distortion of $\mathcal{A}$.

(iv) $\mathrm{Fix}(\mathcal{E}^\dagger)$ is a 1:1 extension of $\mathcal{A}$ from $\mathcal{P}_0$ to $\mathcal{H}$. That is, for each $X \in \mathcal{A}$, there exists precisely one $X' \in \mathrm{Fix}(\mathcal{E}^\dagger)$ so that $X = \Pi(X') = P_0 X' P_0$.

While Theorem 5 is somewhat intimidating (we shall use all of its pieces in Sec. V), the payoff for its complexity is that it consistently unifies the Schrödinger and Heisenberg pictures of information preservation (see also Refs. [3,4,7]). The Schrödinger approach involves looking at the fixed states in $\mathrm{Fix}(\mathcal{E})$. The Heisenberg approach, on the other hand, emphasizes *observables* of the system, which evolve according to $\mathcal{E}^\dagger$ (since expectation values evolve as $\mathrm{Tr}\{X\mathcal{E}(\rho)\} = \mathrm{Tr}\{\mathcal{E}^\dagger(X)\rho\}$). Fixed states of $\mathcal{E}$ in the Schrödinger picture translate to fixed observables of $\mathcal{E}^\dagger$ in the Heisenberg picture. Theorem 5 shows that *both* such fixed sets are isometric to the *same* matrix algebra $\mathcal{A}$. This algebra determines the structure of preserved codes, so the two pictures (interpreted correctly) yield equivalent characterizations of preserved information.

Some of the results in Theorem 5 were proved previously, in different (though related) contexts. Our characterization of $\mathrm{Fix}(\mathcal{E}^\dagger)$ [parts (ii) and (iv)] follows, in particular, from a classic operator algebra paper by Choi and Effros [39]. Their results are substantially more abstract and less constructive, but Kuperberg subsequently applied them to quantum information (see Ref. [40], Theorems 2.2 and 2.3). The proofs given here are self-contained (and perhaps more accessible to physicists).

The fact that an arbitrary CPTP map's fixed points are isometric to a matrix algebra, together with Theorem 4, nails down the structure of *every* preserved code. If $\mathcal{C}$ is a preserved code for a channel $\mathcal{E}$, then it is isometric (i.e., rigidly equivalent) to a matrix algebra. Furthermore, $\mathcal{E}$'s fixed points are a subspace of matrices that looks very much like an algebra—*except* that each of the identity factors $\mathbb{1}_{B_k}$ has been replaced by some fixed matrix τ_k.

While the domain of $\mathcal{E}$ contains all operators on $\mathcal{H}$, its physical significance comes from its action on positive semidefinite states. Given any algebra $\mathcal{A}$ in the canonical form of Eq. (7), we can easily identify the set $\mathcal{A}_+$ of positive states in $\mathcal{A}$: $\mathcal{A}_+$ contains states of the form $\sum_k p_k \rho_k \otimes \frac{\mathbb{1}_{B_k}}{n_k}$, where the $\{p_k\}$ form a probability distribution, and the $\{\rho_k\}$ are arbitrary states on the noiseless factors.

$\mathcal{E}$'s fixed states $[\mathrm{Fix}(\mathcal{E})_+]$ form a very similar set, comprising states of the form $\sum_k p_k \rho_k \otimes \tau_k$, where the $\{p_k\}$ and $\{\rho_k\}$ are probabilities and arbitrary states as above, and the τ_k are *fixed* density matrices determined by $\mathcal{E}$. Any set of fixed states is a fixed code for $\mathcal{E}$, and $\mathrm{Fix}(\mathcal{E})_+$ is the unique largest fixed code. Lemma 2 implies a relationship between noiseless and fixed codes, which in turn implies

Lemma 6. Let $\mathcal{E} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ be a CP map with a full-rank fixed point, whose fixed points induce (see Theorem 5) the decomposition

$$\mathcal{H} = \bigoplus_k A_k \otimes B_k.$$

Then $\mathcal{C}$ is a [convex] maximum noiseless code for $\mathcal{E}$ if and only if $\mathcal{C}$ comprises all states of the following form

$$\rho = \sum_k p_k \rho_{A_k} \otimes \mu_k, \tag{8}$$

where the ρ_{A_k} are arbitrary states on A_k and each μ_k is a fixed (i.e., the same for all ρ) state on B_k.

Note that the lemma is only proved for channels with a *full-rank fixed point*. We believe that a similar result can be proved for arbitrary channels, but there are some tricky details that obscure the main point. We only need to apply this result to channels of the form $\hat{\mathcal{E}}_{\mathcal{P}} \circ \mathcal{E}$, with $\hat{\mathcal{E}}_{\mathcal{P}}$ defined in Eq. (6). Each such channel, from $\mathcal{B}(\mathcal{P}) \to \mathcal{B}(\mathcal{P})$, is actually unital [since $\hat{\mathcal{E}}_{\mathcal{P}} \circ \mathcal{E}(P) = P$], so it has a full-rank fixed point, and Lemma 6 is sufficient to characterize its noiseless codes: They are isometric to the channel's fixed points, which form a distorted algebra.

So while a channel $\mathcal{E}$ typically has a lot of noiseless codes, they turn out to be trivial variations on a constant theme. The variation is a *gauge*—a particular state μ_k for each of the noise-full subsystems. The actual information is carried by the variation in the code states, which differ only on the

opposed to states, which is why the noiseless factors are identified with B_k.

noiseless factors A_k, and in the weights p_k assigned to the different k sectors. This suggests an obvious way to turn noiseless codes into fixed codes, simply by adjusting the state of the noise-full subsystems. Doing so enables us to finally justify Proposition 1 with the following corollary to Lemma 6.

Corollary 7. For every maximum preserved code $\mathcal{C}$, there exists a CPTP map $\mathcal{R}$ such that $\mathcal{R} \circ \mathcal{E}(\rho) = \rho$ for all states $\rho \in \mathcal{C}$.

We have finally proved the central proposition of the previous section, justifying our definition of "preserved." If and only if a code satisfies Definition 7, there exists a recovery operation that makes it into a fixed code, which is clearly preserved in the strongest possible sense. However, this depends on Bob's ability to apply the necessary recovery immediately after $\mathcal{E}$ happens! Section IV considers the effect of placing operational restrictions on what Bob can do, and how this can change the criteria for preservation.

We note in passing that the framework presented by Kuperberg in [40] is similar and uses much of the same mathematics. However, it only addressed noiseless and unitarily noiseless information (a.k.a. *infinite-distance* codes), not correctable information, or the relationship between preservation and correctability.

B. IPSs: The structures that underlie preserved codes

Taken together, the results we have presented thus far indicate a *rigid algebraic structure* lurking within each CPTP map $\mathcal{E}$, which constrains the shape of its preserved and noiseless codes. The codes themselves are not the structure, however. There are many noiseless codes, all distortions of the same algebra. What matters is their shared structure. In fact, all these noiseless codes are manifestations of a unique noiseless IPS underlying the channel, which we turn to explore next. We begin with an example.

Example 11. Consider the two-qubit channel of Example 9, which depolarizes qubit B. There is an infinite family of maximum noiseless codes for this channel: If τ_B is a valid state for B, then $\mathcal{C}_\tau \equiv \{\rho_A \otimes \tau_B \ \forall \ \rho_A\}$ is a noiseless code. While distinct, these noiseless codes are all equivalent, and share the same recovery operation, $\mathcal{R} = \mathbb{1}$. Thus, they are all manifestations of the same noiseless IPS.

This example demonstrates a noiseless IPS, but a channel can also have correctable codes that are not noiseless. However, these codes are noiseless for the appropriate $\mathcal{R} \circ \mathcal{E}$, so the preserved codes with a common recovery $\mathcal{R}$ also share a common structure. A channel can have multiple preserved IPSs. In a way, each IPS is akin to a hole in the wall of noise, through which information can (if properly aimed) pass unscathed. The preserved codes reflect this structure, but their diversity can also obscure it. If we can concisely describe a channel's IPSs, we have (for all practical purposes) completely classified its preserved codes.

Let us define "information-preserving structure" more precisely. Every maximum preserved code is isometric to an algebra, and preserved codes isometric to the same algebra are essentially trivial variations on a theme. They are manifestations of the same underlying IPS.

Definition 11. An *information-preserving structure* (IPS) for a CPTP map $\mathcal{E}$ is an equivalence class of maximum preserved codes for $\mathcal{E}$. Two codes are equivalent if they are isometric to the same algebra, and are preserved according to the same operational criterion (e.g., Definition 7, Definition 8, or one of the other operational criteria in Sec. IV) with the same recovery operation.

The IPS is *not* itself an algebra. Rather, an IPS is an abstract structure (an equivalence class of codes), whose properties are defined by an associated algebra. It is possible for a channel to have two distinct IPS with the same (isomorphic) algebra.

By looking at the structure theorem for matrix algebras [Eq. (7)], we can interpret any given IPS. It consists of one or more k sectors, each of which contains a noiseless subsystem supported on A_k and a noise-full subsystem supported on B_k. Any information encoded into the A_k factors will be preserved by $\mathcal{E}$, whereas any information encoded into the B_k factors is irreparably damaged. The information-carrying capability of a code is determined entirely by its underlying IPS; distinct codes that share an IPS are equivalent, carrying the same kind and amount of information.

Example 12. Consider a classical stochastic map on four symbols, {0,1,2,3}, which maps each input symbol to a mixture of output symbols as follows:

$$0 \to \{0,1\},\ 1 \to \{2,3\},\ 2 \to \{0,2\},\ 3 \to \{1,3\}.$$

There are exactly two maximal preserved codes for this channel, both of which are actually noiseless: {0,1} and {2,3}. They are equivalent, and both described by the same (commutative) algebra—but this is merely a coincidence. The two codes occupy disjoint subspaces of the input, they both get mapped to output states which span the entire output space in different ways, they have entirely different recovery maps, and by changing the channel slightly, we can easily eliminate either code without affecting the other. They are thus not manifestations of the same IPS.

To make use of an IPS, Alice and Bob use any of the equivalent codes associated with that IPS. Each of these codes is isometric to the IPS's algebra, so the structure of that algebra tells us everything about its information-carrying capability. Since the algebra can be decomposed according to Eq. (7),

$$\mathcal{A} \simeq \bigoplus_k \mathcal{M}_{A_k} \otimes \mathbb{1}_{B_k},$$

we can represent it concisely by its *shape*: the vector $\{d_1, d_2, \ldots, d_n\}$ listing the dimensions of the information-carrying factors $\mathcal{H}_{A_k}$ (the noise-full factors are irrelevant). This is shown pictorially in Fig. 3.

The IPS shape characterizes the type and amount of information an IPS can carry. A k sector with an $\mathcal{H}_{A_k}$ factor of dimension $d_k > 1$ can carry quantum information. Classical information is carried by the choice between the different k sectors. Kuperberg, in Ref. [40], described such a noiseless IPS as a *hybrid quantum memory*, capable of simultaneously storing or transmitting a certain amount of quantum information and a certain amount of classical information. The IPS shape provides a very concise way of describing the noise-free degrees of freedom within a given system's Hilbert space—much more convenient than listing the d^4 real parameters required to specify a quantum process on a d-dimensional Hilbert space!

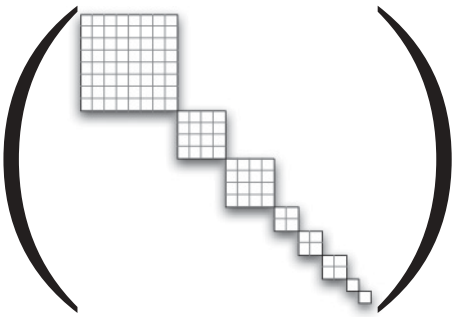


FIG. 3. The block-diagonal shape of the matrices (states) comprising an IPS.

From a physical standpoint, algebraic structure imposes a very strong constraint on the types of information that a quantum process can preserve. *A priori*, we might suppose that any subspace of $\mathcal{B}(\mathcal{H})$ could be "superselected" by some process, however, the theorems proved previously rule out most such possibilities.

Example 13. Consider a single qubit, with $\mathcal{H} = \mathbb{C}^2$. Its dynamics will be described by some CPTP map (or family of them). These dynamics destroy some information while preserving other information, a.k.a. dynamical superselection. Although there are infinitely many different kinds of dynamics, there are only three possible IPSs. The dynamics can preserve the full qubit algebra $\mathcal{M}_2$; or a classical bit, represented (up to unitaries) by the algebra $\mathcal{M}_1 \oplus \mathcal{M}_1 \simeq \text{span}(\{\mathbb{1}, \sigma_z\})$; or nothing, represented by the trivial algebra $\{\mathbb{1}\}$. In particular, there are no CP maps that single out a rebit (a mythical physical system described by a two-dimensional real Hilbert space). This would correspond to preserving information on some equatorial plane of the Bloch sphere, spanned by σ_x and σ_y, while annihilating information about σ_z, as shown in Fig. 4.

But span $(\{\sigma_x, \sigma_y\})$ is not a closed algebra, for σ_x and σ_y generate the full qubit algebra. The fact that no CPTP map can annihilate σ_z while preserving σ_x and σ_y is known, in quantum-information folklore, as the "no-pancake theorem." Our central result might be thought of as a fully general no-pancake theorem, since it rules out the dynamical superselection of all such nonalgebraic IPS.

We can safely talk about "qu*d*its" of information within the code, specified by the IPS shape. Each qudit corresponds to a logical subsystem—a d-dimensional Hilbert space within the full Hilbert space, which need not correspond to a physical subsystem but is nonetheless an independent quantum degree of freedom. Multiple qudits in a direct sum represent a *classical* degree of freedom, for while the different terms in the direct sum correspond to perfectly distinguishable states, superpositions across them are not preserved. We can use these rules to exhaustively catalog all the possible degrees of freedom (up to unitary rotations) within any given quantum system.

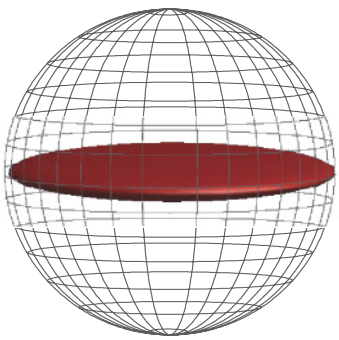


FIG. 4. (Color online) Image of the Bloch sphere under the nonphysical, non-CP "pancake map."

C. Different kinds of IPS

We identified the weak condition as the weakest reasonable condition for information to be preserved. It ensures that Bob can *in principle* restore the system's initial state—but, if Bob has limited resources, then he may be unable to do so *in practice*. Still, Bob's resources may be sufficient to correct a code that satisfies some stronger condition. Each operational constraint on Bob defines some condition on $\mathcal{C}$ that is necessary and sufficient for it to be "preserved" in this situation.

One important example has already appeared, *noiseless* information (Definition 8). Noiseless codes require no correction at all, so noiselessness is a very strong condition. In Sec. IV, we will consider several other conditions. Each such condition defines a distinct class of IPSs. So amongst one or more preserved IPSs a channel may support, one may also be noiseless. A channel's noiseless IPS is unique, because of its relationship to the channel's fixed points (see also Sec. V for further discussion of this point).

Most of the commonly studied techniques for information preservation correspond either to a noiseless IPS, or to a preserved or correctable IPS. Three of the "canonical" structures that we mentioned in Sec. II—pointer bases, DFSs, and NSs—correspond to noiseless IPS. Pointer bases have the shape $\{1,1,1,\ldots\}$, describing a complete set of one-dimensional k sectors (both A_k and B_k are trivial for all k). A DFS has the shape $\{d\}$, describing a single k sector with a trivial $\mathcal{H}_{B_k}$. An NS has the same shape $\{d\}$, but it corresponds to the A_k factor of a single k sector with a nontrivial cofactor B_k.

The relationship between an NS defined in the traditional way as discussed in Example 4 and a noiseless IPS as defined in [20] and in this paper, has some subtleties. A noiseless IPS rests upon a family of noiseless codes, or sets of states, whereas the traditional definition of an NS makes no direct reference to sets of states. The correspondence between the two frameworks arises because Eq. (5) is satisfied if and only if there exist noiseless codes. This does *not* imply that Eq. (5) has anything directly to do with noiseless codes! In particular, a set of states $\{\rho_{AB}\}$ satisfying Eq. (5) need *not* be a noiseless code.

Example 14. Consider a bipartite system AB with Hilbert space $\mathcal{H}_{AB} = \mathcal{H} \otimes \mathcal{H}$, a channel $\mathcal{E}$ that depolarizes system B but leaves A untouched, and the set of states given by $\mathcal{C} = \{|\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi|\}$ for all $|\psi\rangle \in \mathcal{H}$. Since

$$\mathcal{E}(|\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi|) = |\psi\rangle\langle\psi| \otimes \frac{\mathbb{1}}{\dim(\mathcal{H})},$$

$\mathcal{C}$ satisfies Eq. (5). [In fact, every state ρ_{AB} satisfies Eq. (5).] Nonetheless, $\mathcal{C}$ is not noiseless. Equation (5) merely guarantees that a noiseless code will exist.

Error-correcting codes are built upon preserved IPSs. Most QECCs are subspace codes, so a code with a recovery operation $\mathcal{R}$ is a DFS of $\mathcal{R} \circ \mathcal{E}$. While *every* subspace code is associated to an NS of $\mathcal{R} \circ \mathcal{E}$ (as implied by Theorem 6 in [3]), an operator code (OQECC) is also an NS of $\mathcal{R} \circ \mathcal{E}$, for the same $\mathcal{R}$. In each case, the code is built upon the noiseless IPS of $\mathcal{R} \circ \mathcal{E}$, *not* of $\mathcal{E}$ itself. In fact, $\mathcal{E}$ may have no noiseless IPS at all. However, since these codes are correctable for $\mathcal{E}$, they are preserved by it, and so they are associated with preserved IPSs of $\mathcal{E}$.

Example 15. Consider a system of five qubits, and a channel $\mathcal{E}$ that picks one qubit at random and depolarizes it. This is precisely the error model for which the five-qubit QECC was developed [41,42], so $\mathcal{E}$ has a one-qubit preserved IPS. However, it has no noiseless codes at all, because repeatedly applying $\mathcal{E}$ will eventually depolarize all five qubits with high probability.

Example 12 demonstrates that a channel can have more than one preserved IPS. Each is a noiseless IPS for some $\mathcal{R} \circ \mathcal{E}$ (a consequence of Theorem 1), and may be associated with many preserved codes, all of which are corrected by the same $\mathcal{R}$. We would like to have a procedure for listing, or at least counting, all the IPS for a given channel—but unfortunately we do not know how to do this.

What we *can* say (from Theorem 1) is that $\mathcal{E}$'s IPSs comprise all the noiseless IPSs of $\mathcal{R} \circ \mathcal{E}$ for *all* CPTP maps $\mathcal{R}$. A simpler and stronger characterization follows from the structure of the proof. The correction operation for a code depends only on the code's support, so every code with the same support will be corrected by the same operation. This yields a simpler description: $\mathcal{E}$'s IPSs comprise all the noiseless IPSs of $\hat{\mathcal{E}}_{\mathcal{P}} \circ \mathcal{E}$ for all subspaces $\mathcal{P} \subseteq \mathcal{H}$.

While this suggests a way of searching for IPSs (just try every subspace, one at a time), there are uncountably many subspaces to search (see [17]). It may be possible to reduce this problem to searching a countable, even finite set. Unfortunately, it is *not* possible to do so efficiently. Just finding the largest classical code for an arbitrary channel is NP hard, so listing all its preserved IPS is at least this hard. More precisely, let the size of an IPS be measured by the total number of perfectly distinguishable states in one of its preserved codes. Then we have the following:

Lemma 8. The problem of finding the largest preserved IPS for an arbitrary channel $\mathcal{E} : \mathcal{B}(\mathcal{H}_d) \to \mathcal{B}(\mathcal{H}_{d^2})$ that maps a d-dimensional system to a d^2-dimensional system is at least as hard as the NP-complete problem **MAX-CLIQUE**.

IV. OPERATIONAL CONSTRAINTS AND PRESERVED CODES

Our focus thus far has been on a single notion of preservation. We assumed that Alice and Bob were unlimited in their actions (within the laws of physics), and ended up with a preservation condition that depended only on whether $\mathcal{E}$ actually destroyed some of the information. In this section, we will relax this focus, and consider the effect of restrictions on the sender and receiver. Bob may not want to correct the channel constantly, or he may not know how many times $\mathcal{E}$ has been applied. Alice may have a faulty encoder—or perhaps she is not even cooperative. Operational constraints of this sort lead to alternative conditions for preservation. We shall discuss some of the most useful and interesting operational constraints, and the corresponding types of IPS.

A. Infinite-distance IPSs

Suppose we want to store information in a physical system for a time $T > 0$, during which $\mathcal{E}$ will be applied n times. Further, we cannot perform any active operations on the system during this period. Then the information carried by a code $\mathcal{C}$ remains intact only if $\mathcal{C}$ is preserved by the channel $\mathcal{E}^n$. If T (or n) is unknown in advance, $\mathcal{C}$ has to be preserved by all possible powers of $\mathcal{E}$. One example of a channel for which this holds is a unitary channel:

$$\mathcal{E}(\cdot) = U(\cdot)U^\dagger, \tag{9}$$

for some unitary U. A unitary channel adds no noise at all; it just rotates the code around, and the actual rotation depends on how many times it is applied. As long as we know how many times U has been applied, we can recover *any* initial state by applying U^{-n}.

This kind of behavior can be found even in channels that are not purely unitary.

Example 16. Consider a channel on two qubits, labeled A and B, which applies a unitary U to qubit A and depolarizes qubit B. The channel is not unitary, for it adds entropy to any pure state—but nonetheless, it acts unitarily on qubit A. The code $\mathcal{C} = \{\rho_A \otimes \frac{\mathbb{1}_B}{2} \ \forall \ \rho_A\}$ is preserved by any number of applications of $\mathcal{E}$.

We shall refer to a code that remains preserved no matter how many times $\mathcal{E}$ is applied as *unitarily noiseless* under $\mathcal{E}$. Formally, we define a unitarily noiseless code as in Ref. [20].

Definition 12. A code $\mathcal{C}$ is unitarily noiseless under a CPTP map $\mathcal{E}$ if and only if it is preserved by $\mathcal{E}^n$ for any $n \in \mathbb{N}$.

Notice that to retrieve the information stored in a unitarily noiseless code, we need to know the value of n (or, equivalently, the length of time T), in order to construct the appropriate Helstrom measurement. In the previous example, if we lose track of n, then qubit A will get dephased in the diagonal basis of U. Ensuring that unitarily noiseless codes are preserved indefinitely requires a good clock.

Are there codes for which we do not even need a clock? Certainly—for instance, a code containing fixed states of $\mathcal{E}$. Such a code is fixed not only by $\mathcal{E}$, but also by $\mathcal{E}^n$ for any n, and by any convex combination $\sum_n q_n \mathcal{E}^n$ (where $\{q_n\}$ is a probability distribution). So someone ignorant of n can describe the process by a mixture of different $\mathcal{E}^n$, and information in a fixed code is still preserved! Moreover, only the information-carrying part of the code needs to be invariant under repeated applications, which is the operational motivation for noiseless codes (Definition 8).

Noiseless and unitarily noiseless information are preserved indefinitely. No matter how many times $\mathcal{E}$ is applied, we can still distinguish code states. In classical information theory, the number of errors (i.e., bit flips) required to transform one code word into another is called the *distance* of the code. Under the more general definition of distance introduced by Knill *et al.* [3] (based on defining a single application of $\mathcal{E}$ as an "error"), noiseless and unitarily noiseless codes are *infinite-distance codes*, with respect to the noise model defined by $\mathcal{E}$.

Each infinite-distance code is a manifestation of an underlying noiseless or unitarily noiseless IPS. Infinite-distance IPSs may be viewed as degrees of freedom into which $\mathcal{E}$ introduces no entropy at all, transforming them reversibly (if at all). We do not have to pump entropy out of infinite-distance IPS, and so no active error correction is required. For this reason, these have also been called *passive* error-correcting codes.

B. Constraints on the recovery operation

Suppose that we *can* do something to the system in between applications of $\mathcal{E}$. This is crucial whenever the channel preserves information, but maps it to a part of the Hilbert space that is unprotected against further applications of $\mathcal{E}$. Now we must intervene, applying active correction to move our precious information back into protected sectors, and ensure its continued survival. If we can do absolutely anything, then we can correct any preserved code (thanks to Theorem 1). In practice, however, we may only be able to do certain operations. Any CPTP map can be decomposed into (i) a POVM measurement, followed by (ii) a conditional unitary that depends on the outcome of the POVM. This decomposition suggests two natural restrictions on $\mathcal{R}$: It can consist only of a measurement, or it can be completely unitary.

1. Measurement-stabilized codes

If unitary operations are costly or noisy, but measurements can be performed relatively quickly, the only "corrections" that we can perform effectively are pure measurements. For our purposes,[8] a measurement is a POVM defined by a set of effects,

$$\mathcal{M} = \{E_m\}, \text{ where } \sum_m E_m = \mathbb{1} \quad \text{and} \quad E_m \geqslant 0.$$

The outcome of such measurement is a particular value of m, with probability $Pr(m) = \mathrm{Tr}(E_m \rho)$, and a postmeasurement state,

$$\rho \to E_m^{\frac{1}{2}} \rho E_m^{\frac{1}{2}},$$

where $E_m^{\frac{1}{2}}$ is the unique positive semidefinite square root of E_m.

Can we use measurements to correct noise? At first, it seems implausible—after all, while a measurement provides information, it actually does not *do* anything. However, the existence of unitarily noiseless codes shows that passive information gain, such as knowing how many times $\mathcal{E}$ has been applied, can be useful. This motivates a definition of *measurement-stabilized codes*, whose information is preserved indefinitely provided that a measurement is performed after every application of the channel.

Definition 13. A code $\mathcal{C}$ is *measurement stabilized* for a CPTP map $\mathcal{E}$ if there exists a measurement $\mathcal{M} = \{E_m\}$ such that, conditional on any outcome m, $\mathcal{C}$ is unitarily noiseless for $\mathcal{M} \circ \mathcal{E}$.

Stabilizer codes for Pauli channels [25] are an example of measurement-stabilized codes. Stabilizer codes divide the system into two degrees of freedom, the code and the syndrome. Measuring the syndrome "collapses" the error, revealing which Pauli unitary transformed the information-carrying subsystem. In the usual paradigm, we would undo this unitary—but this is not actually necessary, as long as we keep track of the current "Pauli frame" [43] by recording the results of each syndrome measurement as the system evolves.

The key to reconciling the behavior of stabilizer codes with Definition 13 is conditioning on the syndrome measurements. Since each syndrome measurement collapses the syndrome subsystem into a particular basis state, we can see the overall system's dynamics, conditional on the measurement record, as a rather strange time-dependent unitary evolution: At each time step, the code subspace gets transformed by some Pauli operator P_l, and the syndrome state jumps from $|k\rangle \to |k + l\rangle$. Since the code evolves unitarily at every step, it is unitarily noiseless, and the information in it can be recovered at any time.

At first glance, this may seem trivial, for as we observed above, *any* correction operation $\mathcal{R}$ can be written as a measurement followed by a conditional unitary. So, given a generic correctable code, couldn't we just do the measurement, skip the conditional unitary, and keep track of which unitary we did not do? This does not work, in general, because $\mathcal{E}$ may have moved the code to a different subspace which is not, itself, a preserved code. Stabilizer codes for Pauli channels can be measurement stabilized because they actually comprise a large set of preserved codes, and (conditional on the syndrome measurement) the channel merely permutes the codes while transforming them unitarily. It is an open question whether all measurement-stabilized codes are of this form (that is, a large set of isomorphic codes, indexed by a syndrome), or if the previous definition permits other structures.

2. Unitarily correctable codes

In some systems, we have the opposite situation: Measurements are slow and/or hard, while unitary evolution is fast and relatively easy (liquid-state NMR quantum computation is an extreme example; most solid-state architectures also fall into this category). Now we can only apply unitary gates after each application of $\mathcal{E}$. The authors of Ref. [31] considered this situation, and demanded that there exist a unitary matrix U on $\mathcal{H} = (\mathcal{H}_A \otimes \mathcal{H}_B) \oplus \mathcal{H}_C$ such that $\mathrm{Tr}_B\{U\mathcal{E}(\rho_{AB})U^\dagger\} = \mathrm{Tr}_B \rho_{AB}$ for all $\rho_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$. The A subsystem is a *unitarily correctable*[9] subsystem (see also [19]).

Definition 14. A code $\mathcal{C}$ is unitarily correctable for a channel $\mathcal{E}$ if there exists a unitary correction map $\mathcal{U}(\cdot) = U \cdot U^\dagger$, for some unitary operator U, so that $\mathcal{C}$ is noiseless for $\mathcal{U} \circ \mathcal{E}$.

[8] This careful definition may seem pedantic. However, "measurements" are sometimes defined very generally, with an update rule involving *any* square root of the effect E_m. This trivializes our distinction between measurements and arbitrary CP maps. The convention we adopt here is known as *Lüder's Rule*, and defines the unique minimally disturbing (and maximally repeatable) implementation of a given measurement.

[9] The authors of [31] called this "unitarily noiseless," but we believe the term "unitarily correctable" is more appropriate.

Unitarily correctable codes are interesting in part because $\mathcal{E}$ does not inject entropy into the code states.[10] If it did, the error could not be corrected by a unitary operation. Kribs and Spekkens considered unitarily correctable codes in some detail in Ref. [19], and noted that while any preserved code is "unitarily recoverable"—that is, there is a unitary that puts the information back into the subsystem where it originated—this need not suffice to correct the errors, and cooling may be required to protect the information against subsequent iterations of the noise. Sufficient conditions for unitary correctability have likewise been directly derived from the structure of 1-isometric encodings [11] (see, in particular, Proposition 1 therein).

Example 17. Consider two qubits labeled A and B, and let $\mathcal{E}$ act as follows: B is measured in the $\{|0\rangle,|1\rangle\}$ basis; if the result was "1", then A is depolarized. Finally, B is depolarized. The code $\mathcal{C} = \{\rho_A \otimes |0\rangle\langle 0|_B \ \forall\ \rho_A\}$ is preserved. It is unitarily recoverable—in fact, no recovery is necessary because the information remains in the A subsystem. It is *not* unitarily correctable, however, because unless B is cooled to the $|0\rangle$ state, $\mathcal{E}$'s next iteration may damage the information.

Kribs and Spekkens also pointed out that under certain circumstances, unitarily correctable codes can be found efficiently. This observation is closely related to our next topic.

C. Unconditionally preserved information

If a code $\mathcal{C}$ is preserved, then Bob can distinguish between states in $\mathcal{C}$ (and their convex combinations) just as well as Alice. So if we want to know "Was the system prepared in $|\psi\rangle \in \mathcal{C}$?," Bob can answer just as well as Alice could have, by discriminating $|\psi\rangle\langle\psi|$ from a convex combination of all states orthogonal to $|\psi\rangle$. What he cannot do is determine whether the initial state was in $\mathcal{C}$. Information is preserved *conditional* on the system being prepared in $\mathcal{C}$, as illustrated by the following example.

Example 18. Let $\mathcal{E}$ be the following [effectively classical] channel from a d-dimensional system to itself. On the subspace $\mathcal{H}_{d-1}$ spanned by $\{|0\rangle \ldots |d-2\rangle\}$, $\mathcal{E}$ acts as the identity channel. However, $|d-1\rangle$ is decohered and mapped to the maximally mixed state $\frac{1}{d}\mathbb{1}$.

The code comprising all states on $\mathcal{H}_{d-1}$ is preserved, so Bob can distinguish between $|0\rangle$ and any convex combination of $|1\rangle \ldots |d-2\rangle$. If the input state was supported on $\mathcal{H}_{d-1}$, Bob can determine whether $|0\rangle$ was prepared. Without this promise, however, *any* measurement result on the output is consistent with the input state $|d-1\rangle$.

Sometimes, a channel preserves some properties of the input state *irrespective* of what it is. For instance, if $\mathcal{E}$ is the identity channel, then Bob can make any measurement that Alice can. His conclusions from those measurements do not depend on any prior information about the input. The following example is less trivial.

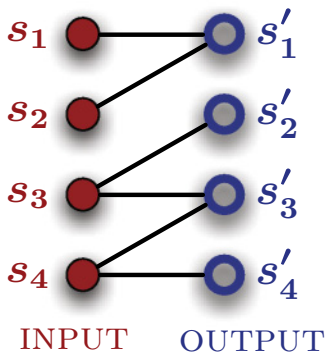


FIG. 5. (Color online) An input-output graph for a classical channel on four states. Vertices on the left represent input states, those on the right represent output states, and edges represent possible mappings.

Example 19. Consider the classical channel whose action is shown in Fig. 5, which corresponds to a stochastic map of the form,

$$\mathcal{E} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

Bob can measure $\{1'\}$ versus $\{2',3',4'\}$, and from the result infer exactly what Alice would have gotten had she measured $\{1,2\}$ versus $\{3,4\}$. So this property of the input state is unconditionally preserved: No matter what the input state was, Bob can determine whether it was in $\{1,2\}$ or not. Note that unconditional preservation need *not* be related to noiselessness—applying this channel twice ruins the information.

This illustrates *unconditionally preserved information*. The most natural way to define unconditional preservation is not in terms of states or codes, however, but rather in terms of measurements (see also Ref. [7], in which closely related ideas are developed).

Definition 15. Let $\mathcal{E} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ be a channel, and $\mathcal{M} = \{P_1, \ldots, P_n\}$ a projective measurement on Hilbert space $\mathcal{H}$ (so $\sum_k P_k = \mathbb{1}$). Then $\mathcal{M}$ is unconditionally preserved by $\mathcal{E}$ if and only if there exists another measurement $\mathcal{M}' = \{Q_1, \ldots, Q_n\}$ on $\mathcal{H}'$ such that $\mathcal{M}'$ simulates $\mathcal{M}$: that is, $\mathrm{Tr}[P_k\rho] = \mathrm{Tr}[Q_k\mathcal{E}(\rho)]$ for all density matrices ρ on $\mathcal{H}$.

This condition on measurements is based in the Heisenberg picture of quantum mechanics, in which states stay fixed, but measurements evolve according to $\mathcal{E}^\dagger$. In order for $\mathcal{M}$ to be unconditionally preserved, there must be some measurement $\mathcal{M}'$ that evolves into $\mathcal{M}$. We can also define an equivalent condition on states.

Definition 16. A code $\mathcal{C}$ is unconditionally preserved by a channel $\mathcal{E}$ if and only if the Helstrom measurement for every weighted pair of states $\{p\rho, (1-p)\sigma\}$ in the convex closure of $\mathcal{C}$ is unconditionally preserved.

[10] Actually, it is slightly more technical than this: Given any unitary correctable code, there is another code associated with the same unitarily correctable IPS, into which $\mathcal{E}$ does not inject any entropy. This is directly related to the fact that a code can be noiseless without being fixed—in both cases, repeated application of $\mathcal{E}$, or $\mathcal{U} \circ \mathcal{E}$, causes the code to converge toward a fixed code, whose entropy does not increase thereafter.

The second definition is strictly more general: Every unconditionally preserved measurement $\mathcal{M} = \{P_1, \ldots, P_2\}$ can be identified uniquely with a code,

$$\mathcal{C} = \left\{ \frac{P_1}{\mathrm{Tr}(P_1)}, \ldots, \frac{P_n}{\mathrm{Tr}(P_n)} \right\},$$

which is unconditionally preserved if and only if $\mathcal{M}$ is. Every classical code whose support is all of $\mathcal{H}$ defines a single unconditionally preserved measurement. Quantum (or hybrid) codes whose support is all of $\mathcal{H}$ define entire algebras of unconditionally preserved measurements. Codes restricted to a subspace do not generally correspond to unconditionally preserved measurements.

The code associated with a given unconditionally preserved measurement spans the entire Hilbert space. Therefore, following the proof of Theorem 1, it can be corrected using a transpose map $\hat{\mathcal{E}}_{\mathcal{P}}$—where $\mathcal{P}$ is the entire Hilbert space! Since this statement holds for every unconditionally preserved measurement, we can correct *every* unconditionally preserved code using a single unique recovery, which we denote $\hat{\mathcal{E}}$:

$$\hat{\mathcal{E}}(\cdot) = \mathcal{E}^\dagger\left(\mathcal{E}(\mathbb{1})^{-\frac{1}{2}}(\cdot)\mathcal{E}(\mathbb{1})^{-\frac{1}{2}}\right). \tag{10}$$

It follows that every unconditionally preserved measurement consists of projectors P_k that are fixed points of $\hat{\mathcal{E}} \circ \mathcal{E}$. There exists a *unique* unconditionally preserved IPS, which contains all the unconditionally preserved codes. Moreover, we can find its structure quite easily by constructing and diagonalizing $\hat{\mathcal{E}} \circ \mathcal{E}$. Other codes are hard to find, precisely because we need to know their support $\mathcal{P}$.

Kribs and Spekkens observed that if $\mathcal{E}$ is unital [that is, $\mathcal{E}(\mathbb{1}) = \mathbb{1}$], then its unitarily correctable codes are fixed points of $\mathcal{E}^\dagger\mathcal{E}$. This is an interesting special case of unconditional preservation. If $\mathcal{C}$ is unitarily correctable, then the channel does not add any entropy to it—thus, every pure state in the code remains pure. But if $\mathcal{E}$ is unital, it cannot map two orthogonal subspaces to overlapping subspaces of the same size, because this would cause a pile-up of probability on the overlapping portion. So every unitarily correctable code for a unital channel must be unconditionally preserved, because no other subspace can be piled on top of it in the output space. Finally, for a unital channel, $\hat{\mathcal{E}} = \mathcal{E}^\dagger$, so $\mathcal{E}^\dagger$ corrects every unconditionally preserved code.

V. APPLICATIONS

In this section, we present three applications of the IPS framework that we have derived. First, we state a very simple algorithm that efficiently finds all noiseless and unitarily noiseless codes for a given map $\mathcal{E}$. We then present a similar algorithm to find all the unconditionally preserved codes. Finally, we show how to address so-called "initialization-free" DFSs and NSs within our framework.

A. Finding infinite-distance codes

Our discussion suggests a natural strategy for finding all the preserved codes of a channel $\mathcal{E}$: First, find all its preserved IPSs; then build codes from the IPSs. Unfortunately, there is a potential IPS for each and every subspace $\mathcal{P} \subseteq \mathcal{H}$. So, searching for IPSs seems to require an exhaustive search over all subspaces of $\mathcal{H}$ (see [17]). We can find *some* preserved codes by picking particular subspaces, but we may not find the largest IPS (or any of them). Since the problem is NP hard, an efficient algorithm seems unlikely (though it should be noted that we have only proven that finding the best *classical* code is NP hard—other special cases, for instance, the largest quantum code, might conceivably be easier).

Let us focus instead on *noiseless* codes. The noiseless IPS of $\mathcal{E}$ is unique because all the maximum noiseless codes are isometric to $\mathcal{E}$'s fixed points. So, to find the unique noiseless IPS, we need only determine the structure of $\mathcal{E}$'s fixed points. Theorem 5 defines this structure, and suggests an efficient algorithm to find it.

Algorithm for finding noiseless IPS

(1) Write $\mathcal{E}$ as a $d^2 \times d^2$ matrix, where d is the dimension of the Hilbert space.

(2) Diagonalize the matrix, and extract its eigenvalue-1 right and left eigenspaces [corresponding to $\mathrm{Fix}(\mathcal{E})$ and $\mathrm{Fix}(\mathcal{E}^\dagger)$, respectively].

(3) Compute $\mathcal{P}_0$, the support of $\mathrm{Fix}(\mathcal{E})$, and project $\mathrm{Fix}(\mathcal{E}^\dagger)$ onto $\mathcal{P}_0$ to obtain a basis for $\mathcal{A}$.

(4) Find the shape of $\mathcal{A}$.

In the last step, we need to find the canonical decomposition, Eq. (7), of a finite-dimensional matrix algebra specified as a linear span. This can be done efficiently using, for example, the algorithm presented in Ref. [44]. This canonical decomposition step is also present in existing algorithms for finding NSs [17,18]. Our algorithm improves on previous algorithms by providing a straightforward method of finding $\mathcal{A}$ as a linear span. Its hardest step is diagonalizing a $d^2 \times d^2$ matrix, which runs in time $O(d^6)$. As such, it is more efficient than algorithms (such as [16,17]) that require exhaustive search over states or subspaces in $\mathcal{H}$, for these sets grow exponentially in volume with d.

We can generalize this algorithm to find an arbitrary channel's unitarily noiseless IPS. Whereas the noiseless IPS consists of $\mathcal{E}$'s fixed points—operators X such that $\mathcal{E}(X) = X$—the unitarily noiseless IPS consists of *rotating points*, operators X such that $\mathcal{E}(X) = e^{i\phi_X} X$.

Definition 17. Let $\mathcal{E} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ be a CPTP map. An operator $X \in \mathcal{B}(\mathcal{H})$ is a unitary eigenoperator of $\mathcal{E}$ if and only if $\mathcal{E}(X) = e^{i\phi} X$ for some $\phi \in \mathbb{R}$. The rotating points of $\mathcal{E}$ comprise all operators in the span of its unitary eigenoperators.

Note that a rotating point need not be an eigenoperator—for instance, a linear combination of two unitary eigenoperators with different phases is a rotating point, but not itself an eigenoperator. As an example, consider the unitary qubit channel $\mathcal{E}(\rho) = e^{-i\phi\sigma_z}\rho e^{i\phi\sigma_z}$. The Pauli operators σ_x and σ_y are not eigenoperators, but they are rotating points.

Lemma 9. If $\mathcal{C}$ is a maximum unitarily noiseless code for a CP map $\mathcal{E}$, then $\mathcal{C}$ is isometric to the set of all (positive trace-1) states in the span of the rotating points of $\mathcal{E}$. In other words, there exists a map $\mathcal{E}_{\mathrm{inf}}$ such that $\|p\mathcal{E}_{\mathrm{inf}}(\rho) - (1-p)\mathcal{E}_{\mathrm{inf}}(\sigma)\|_1 = \|p\rho - (1-p)\sigma\|_1$ for any $\rho, \sigma \in \mathcal{C}$, $p \in [0,1]$, and $\mathcal{E}_{\mathrm{inf}}(\rho)$ and $\mathcal{E}_{\mathrm{inf}}(\sigma)$ are in the span of the rotating points of $\mathcal{E}$.

We adapt the algorithm for finding a noiseless IPS by shifting its focus from fixed points to rotating points. It is

useful to note that the support of the rotating points is the same as the support $\mathcal{P}_0$ of the fixed points. Therefore, we just need to replace step 2 previously mentioned by the following:

(2′). Diagonalize the matrix, and extract the right and left eigenoperators with unit modulus eigenvalues. Let $\mathrm{Fix}(\mathcal{E})$ [$\mathrm{Fix}(\mathcal{E}^\dagger)$] be the linear span of the unit-modulus right (left) eigenoperators.

This again runs in time $O(d^6)$ as before. It is (to our knowledge) the first efficient algorithm to find unitarily noiseless codes for arbitrary channels. We note in passing that both algorithms—for finding noiseless and unitarily noiseless IPS—rely on the codes having infinite distance, so they are unlikely to be adaptable to finding other kinds of IPS.

B. Finding the unconditionally preserved IPS

We know that preserved codes are in general hard to find, but in the previous section we saw how to take advantage of infinite-distance codes' structure to find the unique noiseless and unitarily noiseless IPSs. Unconditionally preserved IPSs are another special case. A channel has a unique unconditionally preserved IPS, and we can find it efficiently. The algorithm is extremely simple: Construct

$$\hat{\mathcal{E}}(\cdot) = \mathcal{E}^\dagger\left(\mathcal{E}(\mathbb{1})^{-\frac{1}{2}}(\cdot)\mathcal{E}(\mathbb{1})^{-\frac{1}{2}}\right) \tag{11}$$

as a matrix, diagonalize $\hat{\mathcal{E}} \circ \mathcal{E}$, and extract its fixed points (the eigenspace with eigenvalue $+1$). These will form an algebra, which defines the IPS we are looking for.

However, one might reasonably inquire why the unconditionally preserved IPS is interesting and useful. If we ask "What information is preserved by a given channel $\mathcal{E}$?" then one possible answer consists of an exhaustive list of all the channel's preserved IPSs. This is somewhat unsatisfactory for three reasons. First, we do not know how to find such a list (though we know that it is generally hard). Second, it might be very very long, even for channels on small systems. Third, the preserved codes corresponding to these IPSs represent information that *could* be preserved by the channel, depending on what the sender chooses to do, and conditional upon prior agreement between sender and receiver.

Unconditional preservation provides an alternative answer. Every channel has a unique unconditionally preserved IPS, comprising all the information that is *definitely* preserved by $\mathcal{E}$. In the important case where the "sender" is a natural process, this IPS represents everything that the observer can determine with certainty. Any further conclusions are only valid conditional upon certain prior assertions about the "distant" system (e.g., that its state lay in some subspace $\mathcal{P}$). This interpretation alone is sufficient reason to consider the unconditionally preserved IPS—independent of the happy accident that it is unique and easily calculable.

C. Initialization-free DFS and NS

As discussed in Sec. III C, DFSs and NSs are manifestations of $\mathcal{E}$'s noiseless IPS. Since a DFS is just an NS with a trivial noise-full subsystem, we shall focus on the NS.[11]

We can demand further operational requirements on an NS. One particular criterion is *robustness* against initialization errors—that is, we demand not only that information encoded *in* the NS be preserved indefinitely, but also that if Alice failed to prepare a state within the NS, that this can be detected by Bob. Such "initialization-free" (IF) NS were first studied in Ref. [27], and have been further characterized in Ref. [28] in the context of Markovian dynamics.

If we decompose the system's Hilbert space as

$$\mathcal{H} = (A \otimes B) \oplus C,$$

and A supports an NS, then we can write an arbitrary density operator in the following block form:

$$\rho = \begin{pmatrix} \rho_{AB} & \bar{\rho} \\ \bar{\rho}^\dagger & \rho_C \end{pmatrix}. \tag{12}$$

The NS is said to be perfectly initialized whenever $\bar{\rho}$ and ρ_C are zero. If, in practice, it is not possible to guarantee preparation within $A \otimes B$, then we need a special kind of NS that is insensitive to such initialization errors. The NS is *initialization-free* if the (possibly subnormalized) state ρ_{AB} on $A \otimes B$ satisfies the NS condition of Eq. (5), even when ρ_C is not zero. In other words, an IF-NS is one that is *immune to interference* coming in from orthogonal subspaces of $\mathcal{H}$ (i.e., states that would not have been prepared if the system had been perfectly initialized).

Our framework, as it turns out, provides a simple and elegant condition for initialization-free NSs: *An NS is IF if and only if it is noiseless and unconditionally preserved*. So, we can find a channel's IF noiseless structures by intersecting its noiseless IPS and its unconditionally preserved IPS. In the remainder of this section, we will demonstrate this equivalence.

Given a Hilbert space $\mathcal{H}$ and a channel $\mathcal{E}$, the channel's noiseless IPS defines a subspace decomposition $\mathcal{H} = \mathcal{P}_0 \bigoplus \overline{\mathcal{P}}_0$. Subspace $\mathcal{P}_0$ is the support of the noiseless IPS. The noiseless IPS also defines a canonical decomposition of $\mathcal{P}_0$ into k sectors $(A_k \otimes B_k)$, so we write the Kraus operators of $\mathcal{E}$ accordingly, as

$$K_i = \begin{pmatrix} \sum_k \mathbb{1}_{A_k} \otimes \kappa_{i,B_k} & D_i' \\ 0 & C_i' \end{pmatrix}. \tag{13}$$

Each k sector is an invariant subspace. So each NS (A_k) is automatically resilient to initialization errors that prepare states in the wrong k sector (but still within $\mathcal{P}_0$).

However, if faulty initialization puts support on $\overline{\mathcal{P}}_0$, then this error may spill into the noiseless sector. Specifically, the D_i' blocks in Eq. (13) map $\overline{\mathcal{P}}_0$ into $\mathcal{P}_0$, which can interfere with information stored in noiseless codes. Since every NS is immune to interference from other k sectors within $\mathcal{P}_0$, let us consider interference from $\overline{\mathcal{P}}_0$.

[11] Ref. [27] actually discusses a more general case, allowing unitary evolution of the NS. This is what we call a unitarily noiseless code. To be consistent with the usual definition of NS, we use the "strict" NS condition given in Eq. (5), but everything in this section can easily be generalized by using a channel's unitarily noiseless IPS instead of its noiseless IPS.

Consider, for the sake of simplicity, a noiseless IPS containing a single k sector, so $\mathcal{P}_0 = A \otimes B$ [as in Eq. (12)]. Let P be the projector onto $\mathcal{P}_0$. The Kraus operators are

$$K_i = \begin{pmatrix} A_i & D_i \\ 0 & C_i \end{pmatrix}, \tag{14}$$

and if the initial state is ρ as given in Eq. (12), then the final state on $\mathcal{P}_0 = A \otimes B$ is

$$P\mathcal{E}(\rho)P = \sum_i A_i \rho_{AB} A_i^\dagger + \sum_i D_i \rho_C D_i^\dagger + \sum_i (A_i \bar{\rho} D_i^\dagger + D_i \bar{\rho}^\dagger A_i^\dagger). \tag{15}$$

For perfect initialization, only the first term is present. The remaining terms represent interference from faulty initialization on $\overline{\mathcal{P}_0}$. The NS is IF if and only if they vanish, which requires

$$\sum_i D_i \rho_C D_i^\dagger = -\sum_i (A_i \bar{\rho} D_i^\dagger + D_i \bar{\rho}^\dagger A_i^\dagger). \tag{16}$$

Since ρ_C is positive semidefinite, the left-hand side of Eq. (16) is also positive semidefinite. But the right-hand side of Eq. (16) must be traceless, because in order for $\mathcal{E}$ to be trace preserving, $\sum_i A_i^\dagger D_i = 0$, and so

$$\mathrm{Tr}\left[\sum_i (A_i \bar{\rho} D_i^\dagger + D_i \bar{\rho}^\dagger A_i^\dagger)\right] = 2\mathrm{ReTr}\left(\sum_i A_i^\dagger D_i \bar{\rho}^\dagger\right) = 0. \tag{17}$$

So the left-hand side is positive semidefinite *and* traceless, which means it vanishes—and so Eq. (16) holds if and only if $\sum_i D_i \rho_C D_i^\dagger = 0$ for all ρ_C—which implies $D_i = 0$ for all i.

This means that in order for an NS whose support is $\mathcal{P}_k = A_k \otimes B_k$ to be IF, the channel must not map anything from $\overline{\mathcal{P}_0}$ into $\mathcal{P}_k$. That is, $\mathcal{P}_k$ is orthogonal to $\mathcal{E}(\rho_C)$ for every $\rho_C \geqslant 0$ on $\overline{\mathcal{P}_0}$ (and, by Lemma 1.1 in Appendix B1, it is sufficient to consider just one full-rank ρ_C on $\overline{\mathcal{P}_0}$). But this is precisely the condition for the corresponding code to be unconditionally preserved: Bob must be able to determine whether the system was correctly initialized, which means that the channel must not map any part of $\overline{\mathcal{P}_0}$ back into $\mathcal{P}_k$.

VI. CONCLUSIONS AND OUTLOOK

We have presented a framework characterizing the information preserved by a quantum process, described by an arbitrary CPTP $\mathcal{E}$ map acting on a finite-dimensional quantum system. Information is carried by codes; codes are preserved if their associated information can be extracted after passing through the channel; preservation implies correctability. Preserved codes are built upon the channel's information-preserving structures (IPSs), which in turn inherit matrix algebra structure from fixed point sets of CPTP maps. This allows for a very elegant and concise description of the full information-carrying capability of any code. We also discussed several operational variations on preservation, with particular attention to infinite-distance codes, and applied the theory to find all of a channel's noiseless, unitarily noiseless, and unconditionally preserved codes.

A number of important open problems and directions for further investigation remain. We have not explicitly addressed continuous-time quantum processes. Such a process is described by a one-parameter family $\{\mathcal{E}_t : t \geqslant 0\}$ of CPTP maps. A special subclass with particular physical significance is Markovian noise, where $\mathcal{E}(t) = e^{t\mathcal{L}}$ for some Liouville semigroup generator $\mathcal{L}$ [45]. In principle, our definitions of noiseless and unitarily noiseless codes extend to the Markovian setting, suggesting connections to recent studies of DFSs and NSs under Markovian noise (see in particular Refs. [27,28,46,47]), and to earlier approaches such as "damping bases" developed in the context of quantum optics [48]. However, we believe it will be necessary to extend our notion of correctability to address continuous-time QEC, as developed, for instance, in [49].

Our analysis has focused on information preservation under the uncontrolled ("free") evolution of an open system. The ability to control that system's dynamics *while* it is experiencing noise (rather than correcting the errors after they occur) raises questions that are interesting for practical quantum information processing and from a control-theoretic perspective. It would be valuable to know how to synthesize dynamics that support a given (desired) IPS, using externally applied control, much as DFSs or NSs can be engineered using open-loop unitary manipulations [50] or closed-loop feedback protocols [28,46].

Our current framework does not address "postselective" preservation of information, where the information is preserved conditional on a particular measurement outcome. Another natural direction for generalization is to relax the "zero-error" requirement, looking at imperfectly preserved information under CPTP channels or more general noisy dynamics. Preliminary investigations [10] indicate that partial extensions of some of the structures present in the perfect case carry over to the approximate case, but a variety of interesting complications arise. A final question that deserves further investigation arises when the information-carrying system is not *initially* fully decoupled from its environment. This particular kind of initialization error can produce noise which cannot be described by CP maps, and its analysis must address the influence of (weak) initial correlation with the environment on the information (supposedly) stored within the system.

ACKNOWLEDGMENTS

The authors acknowledge support in part by the Gordon and Betty Moore Foundation (D.P. and H-K.N.); by the Natural Sciences and Engineering Research Council of Canada and Fonds Québécois de la Recherche sur la Nature et les Technologies (D.P.); by the National Science Foundation under Grants No. PHY-0803371, No. PHY-0456720, No. PHY-0555417, and No. PHY-0903727 (L.V.); and by the Government of Canada through Industry Canada and the Province of Ontario through the Ministry of Research & Innovation (R.B.K.). We also gratefully acknowledge extensive conversations with Robert Spekkens, Daniel Gottesman, Cedric Beny, and Wojciech Zurek.

APPENDIX A: OUR FRAMEWORK FOR ANALYZING INFORMATION

1. Our notion of information: Relation to Shannon theory

The most common technical meaning of "information" comes from Shannon's theory of communication [21,22,51]. Here, Alice and Bob are connected by a communication channel $\mathcal{E}$ (a dynamical map between input states and output states), and also have:

(1) A code book that tells Bob which signals Alice might send;

(2) The patience and ability to send signals requiring arbitrarily many uses of the channel;

(3) A willingness to tolerate a very small probability of failure;

(4) A guarantee that $\mathcal{E}$ will be applied exactly once.

Although this paradigm is the backbone of both classical and quantum information theory, it is not unique. Any or all of the above resources may be unavailable:

(1) Sometimes there is no code book restricting the possible signals. In scientific applications, the source of information is generally a natural phenomenon rather than a canny and cooperative sender. This *observational* paradigm restricts the questions whose answers the receiver can learn.

(2) In real-time applications, a signal has to be transmitted within a strictly limited number (N) of channel uses. This eliminates the second resource (encoding over arbitrarily many uses), and motivates *single-shot* capacity: What can we accomplish with a single use of the channel $\mathcal{E}^{\otimes N}$?

(3) Some applications demand perfect reliability. This eliminates the third resource (tolerance of arbitrarily small failure probability), and yields *zero-error information theory* [52,53].

(4) Memory devices, which store information rather than transmitting it, may violate the guarantee that $\mathcal{E}$ is applied exactly once. We may wish our information to be preserved for an arbitrary number of clock cycles, or $\mathcal{E}$ may be a snapshot of a continuous process. When $\mathcal{E}$ may be applied many times, we turn to *error correction*. *Correctible* information requires active correction after each iteration of $\mathcal{E}$; *noiseless* information persists through repeated iterations of $\mathcal{E}$ with no intervention.

In this paper, we are concerned primarily with identifying the kinds of information that can be preserved, rather than the rate at which information can be sent or stored. So, we focus on zero-error information and the single-shot paradigm. This does not really affect the generality of our results: Since they apply to arbitrary channels, we can discuss $\mathcal{E}^{\otimes N}$ for any N. We do not know for certain, however, whether tolerating an asymptotically small amount of error changes the *kinds* of information that can be preserved by $\mathcal{E}^{\otimes N}$.

The other two resources (a pre-existing code book, and exact knowledge of $\mathcal{E}$) are quite important. They yield different preservation criteria, with substantially different consequences, and we consider them separately.

2. On the usefulness and generality of codes

Our framework for analyzing preserved information relies on *codes* to describe different kinds of information. A code is an arbitrary set of preparations (states) for a physical system $\mathcal{S}$, representing the alternatives available to the sender. Essentially, a code describes a very generalized "subsystem," in which information can be encoded. We settled on this formalism after quite a bit of thought and exploration, and expect that some readers may seek a more extensive explanation of why we believe it is useful, general, and powerful. The most efficient way to do so might be to anticipate some potential objections.

(1) *Using "questions" to define information seems inherently classical, and inadequate to describe quantum information.* The idea of a question, with a definite answer, is indeed inherently classical. Human beings are unavoidably classical, and as Bohr famously insisted [54], our descriptions and perceptions of Nature are always classical. As such, we believe that a precise and general definition of "information" must rely on classical concepts. We can nonetheless describe quantum information in this framework. The difference between a classical bit and a quantum bit is that the bit admits just one sharp question, "Is the bit 0 or 1?" whereas the qubit supports an infinite continuum of inequivalent sharp questions, "Is the qubit in state $|\psi\rangle$ or state $|\psi_\perp\rangle$," for every orthogonal basis $\{|\psi\rangle,|\psi_\perp\rangle\}$. By using classical questions as a common denominator to define both classical and quantum information in the same lingua franca, we have a framework that is open to novel forms of information—rather than begging the question of whether they exist.

(2) *This definition does not seem to capture entanglement as a form of information—that is, that $\mathcal{E}$ might preserve entanglement between $\mathcal{S}$ and a reference system $\mathcal{R}$.* Entanglement is a peculiarly quantum form of correlation, wherein the state of $\mathcal{S}$ is conditional upon observations on the reference system. Projecting $\mathcal{R}$ into a state $|\psi\rangle$ *steers* [55] $\mathcal{S}$ into a corresponding ρ_ψ. It is not difficult to show that $\mathcal{E}$ preserves this entanglement if and only if it also preserves the code comprising *all* ρ_ψ into which $\mathcal{S}$ can be steered. Thus, the code paradigm does address entanglement as a form of information.

(3) *Preserved information should be addressed in the Heisenberg picture, by considering preserved observables rather than states.* In fact, our analysis proceeds along these lines; we demand that every measurement for distinguishing between code states be reproducible on Bob's end. However, the code $\mathcal{C}$ is a crucial ingredient in defining a kind of information, because it determines which measurements need to be reproducible! Otherwise, it is easy to identify *all* POVMs that can be reproduced on Bob's end with "preserved information" [7], an approach that we believe is subtly flawed. A preserved measurement $\mathcal{M}$ represents perfectly preserved information *only* if there is some circumstance under which Alice would measure $\mathcal{M}$ in order to answer a question. If $\mathcal{M}$ is inherently noisy and error-laden, then for any question Alice might ask, there is always some $\mathcal{M}'$ that would yield a better answer. The fact that $\mathcal{M}$ can be reproduced by Bob is irrelevant if Alice would never choose to make that measurement.

(4) *The whole idea of a code is appropriate only in the communication-theoretic paradigm, not the observational one.* If the input to the channel is controlled by an oblivious system (e.g., a distant star) rather than a cooperative sender, then the receiver or observer cannot rely on preparation within the code. This is correct—and yet the framework works nonetheless. If any information is perfectly preserved by the channel, then there *must* be at least two input states that

remain distinguishable at the output. Conversely, if the channel mixes up *every* pair of input states, then there is absolutely no question that Bob can answer as well as Alice. It is true that the semantic meaning of a "code" is inappropriate to the observational paradigm, since an oblivious "sender" is unlikely to cooperate by carefully preparing within a code. Ultimately, this is why we focus not on codes, but on the underlying IPS. The existence of a preserved code is merely a symptom of the underlying structure; if a code exists, then there is potentially an entire equivalence class of codes. This is especially true in the case of unconditionally preserved information (the only kind relevant to observation), where the recovery map $\hat{\mathcal{E}}$ [recall Eq. (11)] does not depend on any prior information about the code (e.g., a subspace projector P). An unconditionally preserved IPS is isometric to a subalgebra that spans the system's entire Hilbert space (rather than a subspace $\mathcal{P}$). Every observable in this algebra can be observed faithfully by the observer at the channel's output. Thus, in this situation, the code framework is ancillary to the real question—but it works nonetheless.

APPENDIX B: PROOFS

In this section, we present complete proofs of the technical results stated in the main text.

1. Preserved information is correctable

Theorem 1. A [convex] code $\mathcal{C}$ is correctable for $\mathcal{E}$ if and only if it is preserved by $\mathcal{E}$.

Proof 2. The "only if" direction is straightforward. For any $\rho,\sigma \in \mathcal{C}$, any $p \in [0,1]$, define the weighted difference $\Delta = p\rho - (1-p)\sigma$. If $\mathcal{C}$ is correctable, then there exists a CPTP $\mathcal{R}$ such that, for every such Δ, $\|\Delta\|_1 = \|(\mathcal{R}\circ\mathcal{E})(\Delta)\|_1$. The 1-norm is contractive under CPTP maps [36], so

$$\|(\mathcal{R}\circ\mathcal{E})(\Delta)\|_1 \leqslant \|\mathcal{E}(\Delta)\|_1 \leqslant \|\Delta\|_1.$$

Combining these two expressions yields $\|\mathcal{E}(\Delta)\|_1 = \|\Delta\|_1$, which means that $\mathcal{C}$ is preserved by $\mathcal{E}$.

To prove that preservation implies correctability, we give an explicit correction operation. This operation is known as the *transpose channel* [37], defined as

$$\hat{\mathcal{E}}_{\mathcal{P}} = \Pi \circ \mathcal{E}^{\dagger} \circ \mathcal{N},$$

where $\mathcal{P}$ is the joint support of all $\rho \in \mathcal{C}$, Π is the projection onto $\mathcal{P}$, P is the projector onto $\mathcal{P}$, $\mathcal{E}^{\dagger}$ is the adjoint map of $\mathcal{E}$, and $\mathcal{N}$ is a normalization map given below. If the operator sum representation (OSR) of $\mathcal{E}$ is

$$\mathcal{E}(\rho) = \sum_i E_i \rho E_i^{\dagger},$$

then the OSRs for these maps are

$$\Pi(\rho) = P\rho P,$$
$$\mathcal{E}^{\dagger}(\rho) = \sum_i E_i^{\dagger}\rho E_i,$$
$$\mathcal{N}(\rho) = \mathcal{E}(P)^{-\frac{1}{2}}\rho\mathcal{E}(P)^{-\frac{1}{2}},$$
$$\hat{\mathcal{E}}_{\mathcal{P}}(\rho) = \sum_i \left(PE_i^{\dagger}\mathcal{E}(P)^{-\frac{1}{2}}\right)\rho\left(\mathcal{E}(P)^{-\frac{1}{2}}E_iP\right).$$

Note that the inverse in $\mathcal{E}(P)^{-\frac{1}{2}}$ is taken on the support of $\mathcal{E}(P)$. It is simple to verify that $\hat{\mathcal{E}}_{\mathcal{P}}$ is a trace-preserving CP map.

To prove that $\hat{\mathcal{E}}_{\mathcal{P}}$ corrects the code $\mathcal{C}$, we need a couple of technical lemmas. The first makes rigorous the notion of a channel's action on a subspace.

Lemma 1.1. Let $\mathcal{E} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ be a CP map, and X_0 be a positive semidefinite operator on $\mathcal{H}$. If X is an operator on the support of X_0, then $\mathcal{E}(X)$ is an operator on the support of $\mathcal{E}(X_0)$.

Proof. Both X_0 and X are diagonalizable, so X_0 has a smallest eigenvalue, and X has a largest eigenvalue. Thus, for some $\epsilon > 0$, $X_0 > \epsilon X$, which means that $X_0 - \epsilon X > 0$. Since $\mathcal{E}$ is CP, $\mathcal{E}(X_0 - \epsilon X) \geqslant 0$. Because it is linear, $\mathcal{E}(X_0) \geqslant \epsilon\mathcal{E}(X)$. This implies that X is supported on the support of X_0. ∎

Now, recall that discriminating between two code states involves a binary (Helstrom) measurement that projects onto one of two orthogonal subspaces. Our second lemma states that if a channel $\mathcal{E}$ preserves a code $\mathcal{C}$, it also preserves the orthogonality of these subspaces.

Lemma 1.2. Let $\mathcal{E}$ be a CP map, ρ and σ be states in a code $\mathcal{C}$ that is preserved by $\mathcal{E}$, and $p \in [0,1]$. Let us write $\Delta = p\rho - (1-p)\sigma$ in terms of its positive and negative parts, as $\Delta = \Delta_+ - \Delta_-$, where $\Delta_\pm$ are positive operators with disjoint supports. Then $\mathcal{E}(\Delta_+)$ and $\mathcal{E}(\Delta_-)$ have disjoint supports.

Proof. The triangle inequality for the 1-norm, together with the fact that $\mathcal{E}$ is TP, gives

$$\begin{aligned}\|\mathcal{E}(\Delta)\|_1 &= \|\mathcal{E}(\Delta_+) - \mathcal{E}(\Delta_-)\|_1 \\ &\leqslant \|\mathcal{E}(\Delta_+)\|_1 + \|\mathcal{E}(\Delta_-)\|_1 \\ &= \mathrm{Tr}(\Delta_+) + \mathrm{Tr}(\Delta_-). \end{aligned} \tag{B1}$$

Because $\mathcal{C}$ is preserved, $\|\mathcal{E}(\Delta)\|_1 = \|\Delta\|_1 = \mathrm{Tr}(\Delta_+) + \mathrm{Tr}(\Delta_-)$. This implies equality throughout Eq. (B1), that is, $\|\mathcal{E}(\Delta_+) - \mathcal{E}(\Delta_-)\|_1 = \|\mathcal{E}(\Delta_+)\|_1 + \|\mathcal{E}(\Delta_-)\|_1$. This is possible if and only if $\mathcal{E}(\Delta_+)$ and $\mathcal{E}(\Delta_-)$ have disjoint supports. ∎

Armed with these results, we wish to prove that $\mathcal{C}$ is noiseless for $\hat{\mathcal{E}}_{\mathcal{P}} \circ \mathcal{E}$. To do so, we will show that for *every* Helstrom measurement $\{\mathcal{P}_+,\mathcal{P}_-\}$ that distinguishes between two states in $\mathcal{C}$, the subspaces $\mathcal{P}_\pm$ are invariant under $\mathcal{E}$. First, we prove this for the special case where the measurement forms a partition of $\mathcal{P}$ (that is, Δ is full rank).

Lemma 1.3. Define $\mathcal{E}$ and Δ as in Lemma 1.2. Define $\mathcal{P}_\pm \equiv \mathrm{supp}(\Delta_\pm)$ and $P_\pm$ as the projector onto $\mathcal{P}_\pm$. Then, if Δ is full rank on $\mathcal{P}$, then $\mathcal{P}_+$ and $\mathcal{P}_-$ are invariant subspaces under $\hat{\mathcal{E}}_{\mathcal{P}} \circ \mathcal{E}$.

Proof. $\hat{\mathcal{E}}_{\mathcal{P}}$ is a composition of three CP maps, so $\hat{\mathcal{E}}_{\mathcal{P}} \circ \mathcal{E}$ can be written as a composition of four maps: $\hat{\mathcal{E}}_{\mathcal{P}} \circ \mathcal{E} = \Pi \circ \mathcal{E}^{\dagger} \circ \mathcal{N} \circ \mathcal{E}$. Let us define the subspaces $\mathcal{Q}_\pm \equiv \mathrm{supp}[\mathcal{E}(\Delta_\pm)]$, and $Q_\pm$ as the projectors onto $\mathcal{Q}_\pm$. We will prove the lemma by following the subspaces $\mathcal{P}_\pm$ through each of the four maps.

By Lemma 1.1, $\mathcal{E}$ maps every operator on $\mathcal{P}_+$ to an operator on $\mathcal{Q}_+$, and every operator on $\mathcal{P}_-$ to one on $\mathcal{Q}_-$. By Lemma 1.2, $\mathcal{Q}_\pm$ are disjoint. Thus, $\mathcal{E}$ maps $\mathcal{P}_\pm$ to disjoint subspaces $\mathcal{Q}_\pm$.

Now we consider $\mathcal{N}$. $P_\pm$ and $\Delta_\pm$ have the same support, so $\mathcal{E}(P_\pm)$ is supported on $\mathcal{Q}_\pm$. Thus, $\mathcal{E}(P_+)$ and $\mathcal{E}(P_-)$ have disjoint supports, and because $P = P_+ + P_-$,

$$\mathcal{E}(P)^{-1/2} = \mathcal{E}(P_+)^{-1/2} + \mathcal{E}(P_-)^{-1/2},$$

and so $\mathcal{N}$ maps $\mathcal{Q}_+ \to \mathcal{Q}_+$ and $\mathcal{Q}_- \to \mathcal{Q}_-$.

Now we consider $\mathcal{E}^\dagger$. Using the cyclic property of the trace, $\mathrm{Tr}[Q_\pm \mathcal{E}(P_\mp)] = 0$ implies $\mathrm{Tr}[P_\mp \mathcal{E}^\dagger(Q_\pm)] = 0$. By Lemma 1.1, $\mathcal{E}^\dagger$ does not map $\mathcal{Q}_\pm$ into $\mathcal{P}_\mp$, which means that $\mathcal{E}^\dagger$ maps $\mathcal{Q}_\pm$ to $\mathcal{P}_\pm$.

Thus, $\mathcal{E}^\dagger \circ \mathcal{N} \circ \mathcal{E}$ maps $\mathcal{P}_\pm \to \mathcal{Q}_\pm \to \mathcal{Q}_\pm \to \mathcal{P}_\pm$. The final projection Π has no effect on any operator in $\mathcal{P}$, so $\hat{\mathcal{E}}_\mathcal{P} \circ \mathcal{E}$ maps $\mathcal{P}_\pm \to \mathcal{P}_\pm$. ∎

Lemma 1.3 is the core of the proof for Theorem 1. To complete the proof, we need to extend it to cases where Δ is not full rank, and therefore $\{\mathcal{P}_+, \mathcal{P}_-\}$ do not form a partition of $\mathcal{P}$.

Lemma 1.4. Lemma 1.3 holds even if Δ is not full rank on $\mathcal{P}$.

Proof. There exists a full-rank (on $\mathcal{P}$) state $\rho_0 \in \mathcal{C}$. This follows because $\mathcal{P}$ is the support of $\mathcal{C}$, and $\mathcal{C}$ is convex. For any $\epsilon \in (0, \ldots, 1)$, $(1-\epsilon)\rho + \epsilon\rho_0$ is full rank. So we consider, in place of ρ, a sequence of full-rank states $\{\rho_n'\}$, where $\rho_n' = (1-\epsilon_n)\rho + \epsilon_n \rho_0$, and $\{\epsilon_n\}$ converges to 0. Lemma 1.3, applied to the sequence of full-rank weighted differences $\Delta'^{(n)} = p\rho_n' - (1-p)\sigma$, implies that the corresponding partitions $\{\mathcal{P}_+'^{(n)}, \mathcal{P}_+'^{(n)}\}$ are invariant subspaces. As $n \to \infty$, $\Delta_-'^{(n)}$ converges to Δ_-, and $\mathcal{P}_-'^{(n)}$ converges to $\mathcal{P}_-$, while $\mathcal{P}_+'^{(n)}$ converges to the orthogonal complement of $\mathcal{P}_-$ in $\mathcal{P}$. Thus $\mathcal{P}_-$ is invariant under $\hat{\mathcal{E}}_\mathcal{P} \circ \mathcal{E}$. The same argument, but with σ replaced by $\sigma_n' = (1-\epsilon_n)\sigma + \epsilon_n\rho_0$, shows that $\mathcal{P}_+$ is invariant under $\hat{\mathcal{E}}_\mathcal{P} \circ \mathcal{E}$. ∎

Armed with Lemmas 1.3 and 1.4, it is now easy to prove that $\mathcal{C}$ is noiseless for $\hat{\mathcal{E}}_\mathcal{P} \circ \mathcal{E}$. Consider an arbitrary convex combination of powers of $\mathcal{E}$,

$$\mathcal{F} \equiv \sum_n p_n (\hat{\mathcal{E}}_\mathcal{P} \circ \mathcal{E})^n,$$

where $\{p_n\}$ is a probability distribution over non-negative integers. Let Δ be a weighted difference of code states. By Lemmas 1.3 and 1.4, the supports of Δ_+ and Δ_- are invariant and disjoint subspaces. Since $\mathcal{F}$ is trace preserving,

$$\begin{aligned} \|\mathcal{F}(\Delta)\|_1 &= \mathrm{Tr}[\mathcal{F}(\Delta_+)] + \mathrm{Tr}[\mathcal{F}(\Delta_-)] \\ &= \mathrm{Tr}(\Delta_+) + \mathrm{Tr}(\Delta_-) = \|\Delta\|_1. \end{aligned} \tag{B2}$$

This condition—satisfied for all Δ—is sufficient for $\mathcal{C}$ to be noiseless. ∎

2. The structure of noiseless codes

Lemma 2. Every noiseless code $\mathcal{C}$ for $\mathcal{E}$ is isometric to a set of states that are fixed points of $\mathcal{E}$.

Proof. Consider the CPTP map,

$$\mathcal{E}_\infty = \lim_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^{N} \mathcal{E}^n.$$

The limit is well defined for any map on a finite-dimensional Hilbert space. Note that $\mathcal{E} \circ \mathcal{E}_\infty = \mathcal{E}_\infty$, so $\mathcal{E}[\mathcal{E}_\infty(\rho)] = \mathcal{E}_\infty(\rho)$ for any $\rho \in \mathcal{C}$. That is, $\mathcal{E}_\infty$ projects onto the fixed points of $\mathcal{E}$. Now, if $\mathcal{C}$ is noiseless for $\mathcal{E}$, then it is preserved by any convex combination of powers of $\mathcal{E}$, and hence by $\mathcal{E}_\infty$. Since $\mathcal{C}$ is preserved by $\mathcal{E}_\infty$, $\mathcal{C}$ is isometric to $\mathcal{E}_\infty(\mathcal{C})$ (see Definition 7). As noted above, $\mathcal{E}_\infty(\mathcal{C})$ consists entirely of fixed states, so $\mathcal{C}$ is isometric to a set of fixed states. ∎

Corollary 3. Every maximum noiseless code for a channel $\mathcal{E}$ is isometric to the full fixed-point set of $\mathcal{E}$.

Proof. Let $\mathcal{C}$ be a noiseless code for $\mathcal{E}$. By Lemma 2, $\mathcal{C}$ is isometric to a subset of the fixed states. The fixed states themselves form a noiseless code $\mathcal{C}_{\max}$. If $\mathcal{C}$ is isometric to a proper subset of the fixed states, then $\mathcal{C}$ is strictly smaller than $\mathcal{C}_{\max}$, and is therefore not maximum. ∎

A similar result for preserved codes follows from the fact that they can be made noiseless (Theorem 1).

Theorem 4. Every maximum preserved code for a CPTP map $\mathcal{E}$ is 1-isometric to the full set of fixed states for some other CPTP map $\mathcal{R} \circ \mathcal{E}$.

Proof. This follows from combining Lemma 1 with Theorem 2 and Definition 9. ∎

These results tell us that maximum preserved codes have the same structure as fixed-state sets—but not what that structure is. The following theorem fills that gap, defining the structure of an arbitrary CPTP map's fixed points. It also characterizes the fixed points of the adjoint map $\mathcal{E}^\dagger$ [defined so that if $\mathcal{E}(\rho) = \sum_i E_i \rho E_i^\dagger$, then $\mathcal{E}^\dagger(\rho) = \sum_i E_i^\dagger \rho E_i$]. This extra result is useful in Sec. V, in the algorithm for finding noiseless codes of $\mathcal{E}$.

Theorem 5. Let $\mathcal{E}$ be a CPTP map on $\mathcal{B}(\mathcal{H})$, and $\mathcal{E}^\dagger$ its adjoint. Let $\mathrm{Fix}(\mathcal{E})$ be the fixed points of $\mathcal{E}$, and $\mathrm{Fix}(\mathcal{E}^\dagger)$ the fixed points of $\mathcal{E}^\dagger$. Then,

(i) Let $\mathcal{P}_0 \subseteq \mathcal{H}$ be the support of $\mathrm{Fix}(\mathcal{E})$. Then $\mathcal{P}_0$ is an invariant subspace under $\mathcal{E}$.

(ii) Let $\mathcal{E}_{\mathcal{P}_0}$ be the restriction of $\mathcal{E}$ to $\mathcal{P}_0$, so $\mathcal{E}_{\mathcal{P}_0} \equiv \Pi_0 \circ \mathcal{E} \circ \Pi_0$, where Π_0 projects onto $\mathcal{P}_0$. Then the fixed points of $\mathcal{E}^\dagger_{\mathcal{P}_0}$ form a matrix algebra $\mathcal{A}$.

(iii) $\mathrm{Fix}(\mathcal{E})$ is a distortion of $\mathcal{A}$.

(iv) $\mathrm{Fix}(\mathcal{E}^\dagger)$ is a 1:1 extension of $\mathcal{A}$ from $\mathcal{P}_0$ to $\mathcal{H}$. That is, for each $X \in \mathcal{A}$, there exists precisely one $X' \in \mathrm{Fix}(\mathcal{E}^\dagger)$ so that $X = \Pi(X') = P_0 X' P_0$.

Proof. First, we will prove that $\mathcal{P}_0$ is an invariant subspace under $\mathcal{E}$, using the following lemma.

Lemma 5.1. $\mathrm{Fix}(\mathcal{E})$ contains a positive, full-rank (on $\mathcal{P}_0$) operator. There exists $\rho_0 \in \mathrm{Fix}(\mathcal{E})$, such that $\langle\psi|\rho_0|\psi\rangle > 0$ for all pure states $|\psi\rangle \in \mathcal{P}_0$.

Proof. Let $\rho_0 \equiv \mathcal{E}_\infty(\mathbb{1})$, where $\mathbb{1}$ is the identity on $\mathcal{H}$. Since $\mathcal{E}_\infty$ is CP and projects onto fixed points of $\mathcal{E}$, ρ_0 must be a non-negative fixed point of $\mathcal{E}$, and hence is in $\mathrm{Fix}(\mathcal{E})$. Let $\mathcal{Q} \subseteq \mathcal{P}_0$ be the support of ρ_0. We want to show that $\mathcal{Q} = \mathcal{P}_0$. Suppose $\mathcal{Q}$ is a proper subspace of $\mathcal{P}_0$. Then, there exists $|\psi\rangle$ in $\mathcal{P}_0 \backslash \mathcal{Q}$ such that $\langle\psi|\rho_0|\psi\rangle = 0$, but there exists $X \in \mathrm{Fix}(\mathcal{E})$ such that $\langle\psi|X|\psi\rangle \neq 0$. Let Y be one of the four possible Hermitian operators: $\pm(X + X^\dagger)$, $\pm i(X - X^\dagger)$, chosen so that $\langle\psi|Y|\psi\rangle < 0$ (this must be true for at least one of the four possibilities). Since $X^\dagger$, $-X$, and iX are all in $\mathrm{Fix}(\mathcal{E})$ if $X \in \mathrm{Fix}(\mathcal{E})$, Y is also in $\mathrm{Fix}(\mathcal{E})$, so $\mathcal{E}_\infty(Y) = Y$. Now consider the operator $\rho = \mathbb{1} + \delta Y$, where $\delta > 0$ is chosen small enough so that ρ is non-negative. Then, $\mathcal{E}_\infty(\rho) = \rho_0 + \delta Y$. However, $\langle\psi|\mathcal{E}_\infty(\rho)|\psi\rangle$, which contradicts the CP property of $\mathcal{E}_\infty$. Therefore, $\mathcal{Q} = \mathcal{P}_0$, and ρ_0 is the desired positive, full-rank fixed operator. ∎

Applying Lemma 1.1 to ρ_0 implies that $\mathcal{P}_0$ is an invariant subspace under $\mathcal{E}$, which proves part (i) of the theorem.

Now, to prove part (ii), we consider $\mathcal{E}_{\mathcal{P}_0} \equiv \Pi_0 \circ \mathcal{E} \circ \Pi_0$, the restriction of $\mathcal{E}$ to $\mathcal{P}_0$. Its Kraus operators are $\{K_i\} = \{P_0 E_i P_0\}$, where P_0 is the projector onto $\mathcal{P}_0$. Since $\mathcal{P}_0$ is an invariant

subspace, $E_i P_0 = P_0 E_i P_0 \, \forall \, i$, which means that $\mathcal{E}_{\mathcal{P}_0}$ is TP (i.e., $\sum_i K_i^\dagger K_i = P_0$). Furthermore, since all of $\mathcal{E}$'s fixed points are supported on $\mathcal{P}_0$, $\mathcal{E}_{\mathcal{P}_0}$ has the same fixed points as $\mathcal{E}$.

We can now show that $\mathcal{E}_{\mathcal{P}_0}^\dagger$'s fixed points must commute with its Kraus operators.

Lemma 5.2. For any $X \in \mathcal{B}(\mathcal{P}_0)$, $\mathcal{E}_{\mathcal{P}_0}^\dagger(X) = X$ if and only if $[X, K_i] = 0$ for all i.

Proof. If $[X, K_i] = 0 \, \forall \, i$, then

$$\mathcal{E}_{\mathcal{P}_0}^\dagger(X) = \sum_i K_i^\dagger X K_i = \Big(\sum_i K_i^\dagger K_i \Big) X = P_0 X = X.$$

Conversely, suppose $\mathcal{E}_{\mathcal{P}_0}^\dagger(X) = X$. Consider the quantity,

$$\sum_i [X, K_i]^\dagger [X, K_i] = \mathcal{E}_{\mathcal{P}_0}^\dagger(X^\dagger X) - X^\dagger X,$$

after some algebra. By construction, this is non-negative. Now, observe that

$$\begin{aligned} \mathrm{Tr}\big\{\rho_0 \big[\mathcal{E}_{\mathcal{P}_0}^\dagger(X^\dagger X) - X^\dagger X\big]\big\} &= \mathrm{Tr}\big\{\mathcal{E}_{\mathcal{P}_0}(\rho_0) X^\dagger X\big\} - \mathrm{Tr}\{\rho_0 X^\dagger X\} \\ &= 0, \end{aligned}$$

since ρ_0 is fixed under $\mathcal{E}$ (and hence $\mathcal{E}_{\mathcal{P}_0}$). Because ρ_0 is full rank and positive, for any positive operator $Y \in \mathcal{B}(\mathcal{P}_0)$, $\mathrm{Tr}(\rho_0 Y) = 0 \Leftrightarrow Y = 0$. Therefore, $\mathcal{E}_{\mathcal{P}_0}^\dagger(X^\dagger X) - X^\dagger X = 0$, and $\sum_i [X, K_i]^\dagger [X, K_i] = 0$. Since every term in the sum is non-negative, we conclude that $[X, K_i] = 0 \; \forall \; i$. (Note: This proof is adapted from a result in [56].) ∎

Lemma 5.2 tells us that the fixed points of $\mathcal{E}_{\mathcal{P}_0}^\dagger$ are precisely the commutant in $\mathcal{B}(\mathcal{P}_0)$ of the Kraus operators $\{K_i\}$. Commutants are closed under addition and multiplication, and the fixed points of $\mathcal{E}_{\mathcal{P}_0}^\dagger$ are closed under Hermitian conjugation. Therefore, the fixed points of $\mathcal{E}_{\mathcal{P}_0}^\dagger$ form a matrix algebra, which completes the proof of part (ii) of the theorem.

Let us denote this matrix algebra $\mathcal{A}$. The structure theorem for matrix algebras (see Eq. (7) and Ref. [38]) states that, in some basis, we can write $\mathcal{A}$ as

$$\mathcal{A} \cong \bigoplus_k \big(\mathcal{M}_{A_k} \otimes \mathbb{1}_{B_k} \big), \tag{B3}$$

which induces a natural Hilbert space decomposition:

$$\mathcal{H} = \mathcal{P}_0 \oplus \overline{\mathcal{P}_0} = \left[\bigoplus_k (A_k \otimes B_k) \right] \oplus \overline{\mathcal{P}_0}. \tag{B4}$$

In this basis, we can say something about the Kraus operators of $\mathcal{E}$.

Lemma 5.3. Given a CPTP map $\mathcal{E}$ on $\mathcal{B}(\mathcal{H})$, let $\mathcal{P}_0$ be the support of its fixed points, and $\mathcal{A}$ the algebra fixed by $\mathcal{E}_{\mathcal{P}_0}^\dagger$ (as in Theorem 5). In the decomposition of $\mathcal{H}$ induced by $\mathcal{A}$ [Eq. (B4)], the Kraus operators of $\mathcal{E}$ have the form:

$$E_i = \begin{pmatrix} \bigoplus_k (\mathbb{1} \otimes K_{i,k}) & D_i \\ 0 & C_i \end{pmatrix}, \tag{B5}$$

for some operators $K_{i,k} \in \mathcal{B}(B_k)$, $C_i \in \mathcal{B}(\overline{\mathcal{P}_0})$, and $D_i \in \mathcal{B}(\overline{\mathcal{P}_0}, \mathcal{P}_0)$.

Proof. The E_i operators can always be written in the 2×2 block form given previously. Since C_i and D_i are arbitrary, we need only show that the upper left block is of the given form, and that the lower left block must vanish. The upper left block of each E_i is a Kraus operator K_i of $\mathcal{E}_{\mathcal{P}_0}$. These are the Hermitian conjugates of the Kraus operators for $\mathcal{E}_{\mathcal{P}_0}^\dagger$, which (by Lemma 5.2) commute with $\mathcal{A}$. Therefore, they must be of the form,

$$K_i = \bigoplus_k \big(\mathbb{1}_{A_k} \otimes K_{i,B_k} \big),$$

which is the desired form for the upper left block. Finally, we observe that the lower left block maps operators on $\mathcal{P}_0$ to operators on $\overline{\mathcal{P}_0}$. Since $\mathcal{P}_0$ is an invariant subspace, this block must vanish. ∎

In light of the above, $\mathcal{E}_{\mathcal{P}_0}$ acts trivially on each of the "noiseless" A_k factors, but does something nontrivial on each of the "noisy" B_k factors. Furthermore, $\mathcal{E}$ acts identically to $\mathcal{E}_{\mathcal{P}_0}$ on the $\mathcal{P}_0$ subspace, but may do anything at all to its complement (including mapping states on $\overline{\mathcal{P}_0}$ onto $\mathcal{P}_0$).

The next step of the proof is to show that $\mathrm{Fix}(\mathcal{E})$ is a distortion of $\mathcal{A}$. Recall that $\mathcal{E}$ and $\mathcal{E}_{\mathcal{P}_0}$ have the same fixed points, so we need only characterize the fixed points of $\mathcal{E}_{\mathcal{P}_0}$. We will do so by constructing a vector space of fixed operators, then showing that this exhausts the fixed points of $\mathcal{E}_{\mathcal{P}_0}$.

Lemma 5.4. Following the notation in Theorem 5, let

$$\mathcal{A} = \bigoplus_k \mathcal{M}_{A_k} \otimes \mathbb{1}_{B_k}$$

be the algebra fixed by $\mathcal{E}_{\mathcal{P}_0}^\dagger$. Then there exist positive semidefinite operators $\tau_k \in \mathcal{B}(\mathcal{H}_{B_k})$ such that the following distortion of $\mathcal{A}$,

$$\tilde{\mathcal{A}} = \bigoplus_k \mathcal{M}_{A_k} \otimes \{\tau_{B_k}\},$$

consists entirely of operators that are fixed by $\mathcal{E}$.

Proof. Let $X = \sum_k X_{A_k} \otimes \tau_{B_k}$ be an element of $\tilde{\mathcal{A}}$. By Lemma 5.3,

$$\begin{aligned} \mathcal{E}(X) &= \sum_i K_i X K_i^\dagger \\ &= \sum_k X_{A_k} \otimes \Big(\sum_i K_{i,k} \tau_k K_{i,k}^\dagger \Big) \\ &= \sum_k X_{A_k} \otimes \mathcal{E}_{B_k}(\tau_k), \end{aligned}$$

where for each k, $\mathcal{E}_{B_k} : \mathcal{B}(\mathcal{H}_{B_k}) \to \mathcal{B}(\mathcal{H}_{B_k})$ is a CPTP map with Kraus operators $\{K_{i,k}\}$. Schauder's fixed-point theorem [29] states that every CPTP map has at least one fixed state. If we let τ_k be a fixed state of $\mathcal{E}_{B_k}$, then $\mathcal{E}(X) = X$. ∎

Now we need to show that $\tilde{\mathcal{A}}$ contains *all* the fixed points of $\mathcal{E}$.

Lemma 5.5. Following the notation in Theorem 5, let $\mathcal{A}$ be defined as in Lemma 5.4. Then every fixed point of $\mathcal{E}$ is in $\tilde{\mathcal{A}}$.

Proof. $\tilde{\mathcal{A}}$ is closed under linear combination, so it is a vector subspace of $\mathcal{B}(\mathcal{P}_0)$. Its dimension is easily calculated:

$$\dim(\tilde{\mathcal{A}}) = \dim(\mathcal{A}) \sum_k \dim(A_k)^2.$$

Let us view $\mathcal{E}_{\mathcal{P}_0}$ and $\mathcal{E}_{\mathcal{P}_0}^\dagger$ as matrices (L and $L^\dagger$, respectively) that act on vectors in $\mathcal{B}(\mathcal{P}_0)$. Since each element of $\mathcal{A}$ is fixed

by $\mathcal{E}$, and is therefore an eigenvector of $\mathcal{E}_{\mathcal{P}_0}$ with eigenvalue $+1$, $\mathcal{E}_{\mathcal{P}_0}$ has a $+1$ eigenspace of dimension at least $\dim(\mathcal{A})$. Furthermore, if $\mathcal{E}$ had another fixed point outside of $\mathcal{A}$, then $\mathcal{E}_{\mathcal{P}_0}$'s $+1$ eigenspace would be strictly larger than that.

Let $\{O_i\}$ be an orthonormal basis (in the Hilbert-Schmidt inner product) for $\mathcal{B}(\mathcal{P}_0)$. L has matrix elements $L_{ij} = \mathrm{Tr}\{O_i^\dagger \mathcal{E}_{\mathcal{P}_0}(O_j)\}$, and $L^\dagger$ is its Hermitian conjugate. The eigenvalues of a matrix and its Hermitian conjugate are complex conjugates of each other. Thus, the dimensions of the $+1$ eigenspaces of L and $L^\dagger$ are equal, and $\mathrm{Fix}(\mathcal{E})$ and $\mathrm{Fix}(\mathcal{E}^\dagger_{\mathcal{P}_0}) = \mathcal{A}$ have the same dimension. So $\mathcal{E}$ has no fixed points outside of $\tilde{\mathcal{A}}$. ∎

These two lemmas prove that $\mathrm{Fix}(\mathcal{E}) = \tilde{\mathcal{A}}$ is a distortion of $\mathcal{A}$.

Finally, let us consider the fixed points of $\mathcal{E}^\dagger$. We begin by showing that they are in 1:1 correspondence with the fixed points of $\mathcal{E}^\dagger_{\mathcal{P}_0}$, by showing that $P_0\mathrm{Fix}(\mathcal{E}^\dagger)P_0 = \mathcal{A}$. The first step is relatively straightforward.

Lemma 5.6. Following the notation in Theorem 5, $P_0\mathrm{Fix}(\mathcal{E}^\dagger)P_0 \subseteq \mathcal{A}$.

Proof. The Kraus operators of $\mathcal{E}^\dagger$ are [by Eq. (B5) in Lemma 5.3]

$$E_i^\dagger = \begin{pmatrix} \bigoplus_k (\mathbb{1} \otimes K_{i,k}^\dagger) & 0 \\ D_i^\dagger & C_i^\dagger \end{pmatrix}.$$

Let X be an element of $\mathrm{Fix}(\mathcal{E}^\dagger)$. By writing X in block-diagonal form with respect to the decomposition $\mathcal{H} = \mathcal{P}_0 \oplus \overline{\mathcal{P}_0}$, and noting that $\mathcal{E}^\dagger(X) = \sum_i E_i^\dagger X E_i$, it is straightforward to show that

$$\mathcal{E}^\dagger_{\mathcal{P}_0}(P_0 X P_0) = P_0 \mathcal{E}^\dagger(X) P_0,$$

and since $\mathcal{E}^\dagger(X) = X$, we conclude that $P_0 X P_0$ is a fixed point of $\mathcal{E}^\dagger_{\mathcal{P}_0}$, and therefore is an element of $\mathcal{A}$. So $P_0\mathrm{Fix}(\mathcal{E}^\dagger)P_0 \subseteq \mathcal{A}$. ∎

Now we need to show that $\mathcal{A} \subseteq P_0\mathrm{Fix}(\mathcal{E}^\dagger)P_0$. This is a bit more difficult, and requires a technical lemma. Let us partition the Hilbert-Schmidt space into subspaces as follows:

$$\begin{aligned} \mathcal{K} &\equiv \mathcal{B}(\mathcal{H}), \\ \mathcal{K}_0 &\equiv \mathcal{B}(\mathcal{P}_0), \\ \overline{\mathcal{K}}_0 &\equiv \mathcal{K}/\mathcal{K}_0. \end{aligned}$$

We can write the matrix representing $\mathcal{E}$ in block form as

$$L = \begin{pmatrix} L_{\mathcal{E}_{\mathcal{P}_0}} & L_{\mathcal{G}} \\ 0 & L_{\mathcal{F}} \end{pmatrix}. \tag{B6}$$

Here, L corresponds to the map $\mathcal{E}$, which acts on vectors in $\mathcal{K}$. $L_{\mathcal{E}_{\mathcal{P}_0}}$ corresponds to the map $\mathcal{E}_{\mathcal{P}_0}$ and maps $\mathcal{K}_0$ back into itself. $L_{\mathcal{F}}$ maps $\overline{\mathcal{K}}_0$ back into itself, while $L_{\mathcal{G}}$ maps $\overline{\mathcal{K}}_0$ to $\mathcal{K}_0$. Because $\mathcal{P}_0$ is an invariant subspace, L does not map $\mathcal{K}_0$ to $\overline{\mathcal{K}}_0$. The matrix for $\mathcal{E}^\dagger$ is the Hermitian conjugate $L^\dagger_{\mathcal{E}}$.

Lemma 5.7. $L_{\mathcal{F}}$ has no fixed points.

Proof. Suppose there exists $X \in \overline{\mathcal{K}}_0$ such that $L_{\mathcal{F}}(X) = X$. Define $Y = L_{\mathcal{G}}(X)$. Then

$$L\begin{pmatrix} 0 \\ X \end{pmatrix} = \begin{pmatrix} Y \\ X \end{pmatrix},$$

and the action of $\mathcal{E}^n$ on the operator corresponding to $\binom{0}{X}$ is given by

$$L^n\begin{pmatrix} 0 \\ X \end{pmatrix} = \begin{pmatrix} \sum_{m=0}^{n-1} L^m_{\mathcal{E}_{\mathcal{P}_0}}(Y) \\ X \end{pmatrix}.$$

If Y is orthogonal to the subspace $\mathrm{Fix}(\mathcal{E})$, then as $n \to \infty$, the sum converges to

$$\lim_{n\to\infty} L^n\begin{pmatrix} 0 \\ X \end{pmatrix} = \begin{pmatrix} (\mathbb{1} - \mathcal{E}_{\mathcal{P}_0})^{-1}(Y) \\ X \end{pmatrix}.$$

This is a fixed point of $\mathcal{E}$ not contained in $\mathrm{Fix}(\mathcal{E})$, which contradicts the definition of $\mathrm{Fix}(\mathcal{E})$. On the other hand, if Y is *not* orthogonal to $\mathrm{Fix}(\mathcal{E})$, then the sum diverges as $n \to \infty$. This implies that $\mathcal{E}$ is noncontractive, which violates complete positivity [36]. So, either way, we have a contradiction. ∎

Using Lemma 5.7, we can show that every fixed point of $\mathcal{E}^\dagger_{\mathcal{P}_0}$ has an extension to a fixed point of $\mathcal{E}^\dagger$.

Lemma 5.8. Let $X_0 \in \mathcal{A}$ be a fixed point of $\mathcal{E}^\dagger_{\mathcal{P}_0}$. Then there exists a fixed point $X \in \mathcal{B}(\mathcal{H})$ of $\mathcal{E}^\dagger$ such that $P_0 X P_0 = X_0$.

Proof. Both X_0 and X are vectors in the Hilbert-Schmidt space $\mathcal{K} = \mathcal{B}(\mathcal{H})$. Using the decomposition $\mathcal{K} = \mathcal{K}_0 \oplus \overline{\mathcal{K}_0}$, we can write X_0 in block form:

$$X_0 = \begin{pmatrix} X_0 \\ 0 \end{pmatrix}.$$

In this block form, we choose

$$X = \begin{pmatrix} X_0 \\ (\mathbb{1}_{\overline{\mathcal{K}}_0} - L^\dagger_{\mathcal{F}})^{-1} L^\dagger_{\mathcal{G}} X_{\mathcal{K}_0} \end{pmatrix}.$$

Note that $L_{\mathcal{F}}$ has no fixed points (by Lemma 5.7), so $\mathbb{1}_{\overline{\mathcal{K}}_0} - \mathcal{L}^\dagger_{\mathcal{F}}$ is invertible, which means that X is well defined. Furthermore, $P_0 X P_0 = X_0$ by construction. To show that X is a fixed point of $\mathcal{E}^\dagger$, we simply compute

$$\begin{aligned} L^\dagger(X) &= \begin{pmatrix} L^\dagger_{\mathcal{E}_{\mathcal{P}_0}} & 0 \\ L^\dagger_{\mathcal{G}} & L^\dagger_{\mathcal{F}} \end{pmatrix} \begin{pmatrix} X_0 \\ (\mathbb{1}_{\overline{\mathcal{K}}_0} - L^\dagger_{\mathcal{F}})^{-1} L^\dagger_{\mathcal{G}} X_0 \end{pmatrix} \\ &= \begin{pmatrix} L^\dagger_{\mathcal{E}_{\mathcal{P}_0}}(X_0) \\ (\mathbb{1} + L^\dagger_{\mathcal{F}}(\mathbb{1} - L^\dagger_{\mathcal{F}})^{-1}) L^\dagger_{\mathcal{G}}(X_0) \end{pmatrix} \\ &= \begin{pmatrix} X_0 \\ (\mathbb{1} - L^\dagger_{\mathcal{F}} + L^\dagger_{\mathcal{F}})(\mathbb{1} - L^\dagger_{\mathcal{F}})^{-1} L^\dagger_{\mathcal{G}}(X_0) \end{pmatrix} \\ &= \begin{pmatrix} X_0 \\ (\mathbb{1}_{\overline{\mathcal{K}}_0} - L^\dagger_{\mathcal{F}})^{-1} L^\dagger_{\mathcal{G}} X_0 \end{pmatrix} \\ &= X. \end{aligned}$$ ∎

Lemma 5.8. implies that $\mathcal{A} \subseteq P_0\mathrm{Fix}(\mathcal{E}^\dagger)P_0$. Combining this with Lemma 5.6, we conclude that $\mathcal{A} = P_0\mathrm{Fix}(\mathcal{E}^\dagger)P_0$, which completes the proof of Theorem 5. ∎

Now, we want to show that $\mathcal{E}$'s noiseless codes have a rigid structure dictated by the fixed points.

Lemma 6. Let $\mathcal{E} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ be a CP map with a full-rank fixed point, whose fixed points induce (see Theorem 5) the decomposition

$$\mathcal{H} = \bigoplus_k (A_k \otimes B_k).$$

Then $\mathcal{C}$ is a [convex] maximum noiseless code for $\mathcal{E}$ if and only if $\mathcal{C}$ comprises all states of the following form

$$\rho = \sum_k p_k \rho_{A_k} \otimes \mu_k, \tag{B7}$$

where the ρ_{A_k} are arbitrary states on A_k and each μ_k is a fixed (*i.e.*, the same for all ρ) state on B_k.

Proof. If $\mathcal{C}$ has the given structure, then

(1) It is maximum, since it is isometric to the full set of fixed states of $\mathcal{E}$.

(2) It is noiseless, because $\mathcal{E}$ leaves the states on subsystem A_k intact, and every ρ_k state has the same noise-full state μ_k. So $\mathcal{E}$ preserves all the weighted 1-norm distances between code states.

To show the converse, we must show that if $\mathcal{C}$ is *not* of this form, then it is not maximum noiseless. If $\mathcal{C}$ is not of this form, then either

(1) it contains only a strict subset of the states given above; or,

(2) it contains at least one state with correlations (off-diagonal elements) between different k sectors; or

(3) it contains at least one state with correlations between A_k and B_k; or

(4) it contains states that differ on B_k.

If $\mathcal{C}$ is a strict subset, then it is obviously not maximum.

The key to proving the converse is showing that the condition for noiselessness (Definition 8) forbids correlations between the k sectors as well as between A_k and B_k. The proof relies both on convexity and on the code being maximum. First, recall the map $\mathcal{E}_\infty$ from Lemma 2, which projects onto the fixed-point set $\mathrm{Fix}(\mathcal{E})$. Given the structure of $\mathrm{Fix}(\mathcal{E})$, the CPTP $\mathcal{E}_\infty$ must act on states on $\mathcal{P}_0$ as

$$\mathcal{E}_\infty(\rho) = \bigoplus_k \left(\mathrm{Tr}_{B_k}\{P_k \rho P_k\} \otimes \tau_k\right), \tag{B8}$$

where τ_k is the fixed state on B_k from Theorem 5, and P_k projects onto the kth sector. From Lemma 2, we know that for every fixed state of the form $\rho_f \equiv \bigoplus_k (\sigma_{A_k} \otimes \tau_k)$, there exists exactly one code state $\rho \in \mathcal{C}$ such that $\mathcal{E}_\infty(\rho) = \rho_f$. From Eq. (B8), this demands $\mathrm{Tr}_{B_k}\{P_k \rho P_k\} = \sigma_{A_k}$ for all k.

Now, focus on the case with only two k sectors, labeled 1 and 2. Consider two fixed states in these sectors with block-diagonal form:

$$\rho_{f1} = \begin{pmatrix} \rho'_{f1} & 0 \\ 0 & 0 \end{pmatrix}, \quad \rho_{f2} = \begin{pmatrix} 0 & 0 \\ 0 & \rho'_{f2} \end{pmatrix}.$$

The two code states that are isometric to the fixed points must, respectively, be of the form,

$$\rho_1 = \begin{pmatrix} \rho'_1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \rho_2 = \begin{pmatrix} 0 & 0 \\ 0 & \rho'_2 \end{pmatrix}.$$

By convexity of $\mathcal{C}$, any convex combination of ρ_1 and ρ_2 must also be in $\mathcal{C}$. This excludes from $\mathcal{C}$ any state with on-diagonals equal to this convex combination, but nonzero off-diagonals, since the two different states will have the same image (and hence indistinguishable) under $\mathcal{E}_\infty$. Generalizing this to any number of k sectors, we find that any code state in $\mathcal{C}$ must be block diagonal: $\rho = \bigoplus_k \rho'_k$.

Next, consider the state ρ'_k for the kth sector. We need to show that only product states of $A_k \otimes B_k$ are allowed. We first consider a fixed state ρ'_f on this sector of the form $|\psi\rangle\langle\psi|_{A_k} \otimes \tau_k$. Since the state on A_k is pure, the corresponding code state whose image under $\mathcal{E}_\infty$ is ρ'_f must also be pure on A_k. It is hence a product state of the form $|\psi\rangle\langle\psi|_{A_k} \otimes \mu_k$. Next, suppose $\rho'_f = \sigma_{A_k} \otimes \tau_k$, where σ_{A_k} is in general a mixed state writable as $\sigma_{A_k} = \sum_\alpha q_\alpha |\psi_\alpha\rangle\langle\psi_\alpha|_{A_k}$. Now, each state $|\psi_\alpha\rangle\langle\psi_\alpha|_{A_k} \otimes \tau_k$ is a fixed state, with corresponding code state $\rho'_{k,\alpha} = |\psi_\alpha\rangle\langle\psi_\alpha|_{A_k} \otimes \mu_{k,\alpha}$. By convexity, the state $\sum_\alpha q_\alpha \rho'_{k,\alpha}$ is also in $\mathcal{C}$ and maps to $\rho_f = \sigma_{A_k} \otimes \tau_k$ under $\mathcal{E}_\infty$. This excludes from $\mathcal{C}$ any other state with nonzero correlations between A_k and B_k, but with the reduced state on A_k equal to σ_{A_k}. Furthermore, we must have that $\mu_{k,\alpha} = \mu_k \ \forall\ \alpha$ in order for the (1-norm) distinguishability between the $\rho'_{k,\alpha}$'s to remain unchanged under $\mathcal{E}_\infty$. Therefore, ρ'_k must be of the form $\sigma_{A_k} \otimes \mu_k$ for some μ_k. ∎

We knew already that noiseless codes are isometric to fixed states (Lemma 2) and that fixed states are isometric to algebras (Theorem 5). Now we know explicitly what these codes look like. The isometry is very similar to the one between the fixed states [$\mathrm{Fix}(\mathcal{E})$] and the underlying algebra $\mathcal{A}$: A noiseless code is obtained from $\mathrm{Fix}(\mathcal{E})$ just by changing the state of the noise-full factors.[12]

Finally, it follows from this lemma that not only can we make preserved codes noiseless, but we can also make them fixed.

Corollary 7. For every maximum preserved code $\mathcal{C}$, there exists a CPTP map $\mathcal{R}$ such that $\mathcal{R} \circ \mathcal{E}(\rho) = \rho$ for all states $\rho \in \mathcal{C}$.

Proof. From Theorem 1, we know that every preserved code $\mathcal{C}$ is correctable, so there exists a recovery map $\mathcal{R}_0$ such that $\mathcal{C}$ is noiseless for $\mathcal{R}_0 \circ \mathcal{E}$, and $\mathcal{R}_0 \circ \mathcal{E}$ is unital. By Lemma 6, $\mathcal{C}$ contains states all of the form $\rho = \sum_k p_k \rho_{A_k} \otimes \mu_k$. Now let $\mathcal{R} = \mathcal{T} \circ \mathcal{R}_0$, where $\mathcal{T}$ does nothing to the A_k subsystems, but replaces the state of each B_k subsystem with μ_k. (Constructing such a map is simple, and it is manifestly CPTP.) Now, every $\rho \in \mathcal{C}$ is a fixed state of $\mathcal{R} \circ \mathcal{E}$. ∎

3. Finding preserved IPS is hard

Lemma 8. The problem of finding the largest preserved IPS for an arbitrary channel $\mathcal{E} : \mathcal{B}(\mathcal{H}_d) \to \mathcal{B}(\mathcal{H}_{d^2})$ that maps a d-dimensional system to a d^2-dimensional system is at least as hard as the NP-complete problem **MAX-CLIQUE**.

Proof. The proof is straightforward, and proceeds in three steps. First, we review a known result connecting classical channels with graphs. Second, we show that finding the largest code for a certain set of classical channels is equivalent to **MAX-CLIQUE**. Third, we observe that the classical channels can be embedded in quantum channels.

(1) A classical channel $\mathcal{E}_c$ maps a set of input symbols $\{1, \ldots, N\}$ into mixtures of a set of output symbols $\{1, \ldots, M\}$. For each input symbol n, its *image* $\mathcal{I}(n)$ is the set of output

[12] Since $\mathcal{C}$ only contains *states*, we are really restricting to the positive trace-1 operators in $\mathcal{M}_k$ within $\mathrm{Fix}(\mathcal{E})$ and $\mathcal{A}$. This is what we mean by "$\mathcal{C}$ is isometric to a matrix algebra."

symbols to which $\mathcal{E}$ maps it with nonzero probability. A set of input symbols $\mathcal{C} = \{n_1, \ldots, n_k\}$ is a preserved zero-error code for $\mathcal{E}$ if and only if the images of all the n_j are disjoint—that is, it is possible to unambiguously identify which of the input symbols was sent. We can define the channel's *adjacency graph* G (see Example 6) as follows: The vertices are labeled by input symbols $\{1, \ldots, N\}$, and two vertices $\{n,m\}$ are connected by an edge if and only if the images $\mathcal{I}(n)$ and $\mathcal{I}(m)$ are overlapping. Now, a code $\mathcal{C}$ is a subgraph of G, and it is preserved if and only if no two of its vertices are connected—that is, if it is an *independent set* of G. The largest code is a *maximum independent set* of G. An independent set for G is a clique for its dual graph G', and finding the maximum clique for an arbitrary G' is a well-known NP-complete problem called **MAX-CLIQUE**.

(2) We have not yet shown that finding a classical channel's largest code is NP complete—perhaps all channel's adjacency graphs are easy instances of **MAX-CLIQUE**? This turns out not to be the case; any graph H can be the adjacency graph of a classical channel. Let H be a graph with vertices $\{1, \ldots, d\}$, and let $\mathcal{E}$ be a classical channel from $\{1, \ldots, d\} \to \{1, \ldots, d^2\}$, defined as follows:

(a) The d input symbols are denoted $v \in \{1, \ldots, d\}$, and the d^2 output symbols are denoted by ordered pairs $u \in \{1, \ldots, d\} \times \{1, \ldots, d\}$.

(b) For each input symbol $v \in \{1, \ldots, d\}$, $\mathcal{E}$ maps v (with nonzero probability) to each of the d output symbols $\{(v,x) : x = 1, \ldots, d\}$.

(c) For each input symbol v, $\mathcal{E}$ maps each input symbol v to output symbol (v',v) if and only if H contains the edge (v',v).

Note that each output symbol (a,b) can be produced by at most two input symbols (a and b). So, if two input symbols v and v' are connected in H, then $\mathcal{E}$ maps both of them to the output symbol (v',v), and so they are connected in the adjacency graph G. But, if they are not connected in H, then they are not mapped to the same output symbol, so they are not connected in G. Ergo, $G = H$, and any graph can be produced as the adjacency graph of a channel.

Finally, we need to show that for each such graph, we can construct a quantum channel. This is rather easy. Let the input space be $\mathcal{H}_d$ and the output space be $\mathcal{H}_{d^2}$. Let $\{|1\rangle, \ldots, |d\rangle\}$ be a basis for $\mathcal{H}_d$. Then the $\mathcal{E}$ we will consider acts as follows: First, it dephases in the given basis (i.e., measures it); and then it acts as the classical channel above. ∎

4. Unitarily noiseless codes

The analysis of unitarily noiseless codes follows closely that of the noiseless codes. The rotating points of $\mathcal{E}$ replace its fixed points, with a CPTP map that projects onto their span playing the role that $\mathcal{E}_\infty$ does for noiseless codes.

Lemma 9. If $\mathcal{C}$ is a maximum unitarily noiseless code for a CP map $\mathcal{E}$, then $\mathcal{C}$ is isometric to the set of all (positive trace-1) states in the span of the rotating points of $\mathcal{E}$. In other words, there exists a map $\mathcal{E}_{\text{inf}}$ such that $\|p\mathcal{E}_{\text{inf}}(\rho) - (1-p)\mathcal{E}_{\text{inf}}(\sigma)\|_1 = \|p\rho - (1-p)\sigma\|_1$ for any $\rho,\sigma \in \mathcal{C}$, $p \in [0,1]$, and $\mathcal{E}_{\text{inf}}(\rho)$ and $\mathcal{E}_{\text{inf}}(\sigma)$ are in the span of the rotating points of $\mathcal{E}$.

Proof. By Definition 17, a rotating point X of $\mathcal{E}$ is a linear combination of operators X_k such that $\mathcal{E}(X_k) = e^{i\phi_k}X_k$. Let $\text{Rot}(\mathcal{E})$ be the complex span of all rotating points of $\mathcal{E}$. It is convenient to move to the Hilbert-Schmidt space, where $\text{Rot}(\mathcal{E})$ can be viewed as a subspace spanned by the vectors corresponding to the rotating points. Clearly, $\text{Rot}(\mathcal{E})$ is an invariant subspace under the linear map $\mathcal{E}$, in the sense that any vector in $\text{Rot}(\mathcal{E})$ gets mapped under $\mathcal{E}$ to another vector in $\text{Rot}(\mathcal{E})$. Let $\mathcal{E}_R$ denote $\mathcal{E}$ restricted to $\text{Rot}(\mathcal{E})$. We view $\mathcal{E}$ and $\mathcal{E}_R$ as matrices acting on vectors in the Hilbert-Schmidt space.

Even though $\mathcal{E}$ may not be a diagonalizable matrix, we can still write it in the Jordan normal form [57]: There exists an invertible matrix S such that $\mathcal{E} = SJS^{-1}$, where J is the matrix $J = \text{diag}[J_1, J_2, \ldots, J_K]$. Each J_k is called a *Jordan block*, and it is zero except on the diagonal and first-off-diagonal:

$$J_k = \begin{pmatrix} \lambda_k & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_k & 1 \\ & & & \lambda_k \end{pmatrix}. \tag{B9}$$

The Jordan form for $\mathcal{E}$ is unique up to permutation of the Jordan blocks. Note that any vector $|v\rangle$ is an eigenvector of J if and only if $S|v\rangle$ is an eigenvector of $\mathcal{E}$.

Lemma 9.1. For any k, the support of J_k contains exactly one unit eigenvector of $\mathcal{E}$. The corresponding eigenvalue is λ_k.

Proof. Let $\{|v_\alpha^{(k)}\rangle\}_{\alpha=1}^m$ be the ordered basis for the support of J_k in which J_k takes the form Eq. (B9). Clearly, $J_k|v_1^{(k)}\rangle = \lambda_k|v_1^{(k)}\rangle$, so $S|v_1^{(k)}\rangle$ is an eigenvector of $\mathcal{E}$ with eigenvalue λ_k. To show that this is the only eigenvector in this Jordan block, let $|v\rangle \equiv \sum_\alpha \mu_\alpha |v_\alpha^{(k)}\rangle$ be a vector in the support of J_k. From the form of J_k in Eq. (B9), it is easy to see that the coefficients $\{\mu_\alpha\}$ satisfy the equation $J_k|v\rangle = a|v\rangle$ for some constant a only if $\mu_{\alpha+1} = (a-\lambda_k)\mu_\alpha$ for $\alpha = 1, \ldots, m-1$, and $(a-\lambda_k)\mu_m = 0$. The only nontrivial solution is $a = \lambda_k$ and $\mu_1 \neq 0, \mu_{\alpha>1} = 0$. ∎

This lemma tells us that the rotating points of $\mathcal{E}$ are mutually orthogonal, unless there are degenerate eigenspaces of rotating points. In that case, we can still pick an orthonormal basis for each degenerate eigenspace (already done in the Jordan normal form), and these bases, together with the nondegenerate rotating points, form an orthonormal basis of rotating points for $\text{Rot}(\mathcal{E})$. We denote this basis as $\{X_l\}$. $\mathcal{E}_R$ is diagonal in this basis, with entries $e^{i\phi_l}(= \lambda_l)$. Note that, for any CPTP map $\mathcal{E}$, the following lemma from [57] holds.

Lemma 9.2. Any eigenvalue λ of $\mathcal{E}$ must satisfy $|\lambda| \leqslant 1$.

This, together with Lemma 9.1, implies that $|\lambda_k| \leqslant 1 \; \forall \, k$.

Next, consider powers of $\mathcal{E}$. $\mathcal{E}^n$ can be written using the Jordan normal form as SJ^nS^{-1} where $J^n = \text{diag}[J_1^n, J_2^n, \ldots, J_K^n]$ with each J_k^n being an upper-triangular matrix:

$$J_k^n = \begin{pmatrix} \lambda_k^n & \binom{n}{1}\lambda_k^{n-1} & \binom{n}{2}\lambda_k^{n-2} & \cdots \\ 0 & \lambda_k^n & \binom{n}{1}\lambda_k^{n-1} & \cdots \\ 0 & 0 & \lambda_k^n & \cdots \\ & & & \ddots \end{pmatrix}. \tag{B10}$$

Using the form of J_k^n in Eq. (B10), we can show the following fact about the rotating points of $\mathcal{E}$.

Lemma 9.3 Any (nondegenerate) rotating point of $\mathcal{E}$ must occur in a one-dimensional Jordan block.

Proof. (This proof follows ideas from [58] for the proof of Lemma 9.2.) Suppose there exists a rotating point X such that it belongs to some $m \times m$ Jordan block J_k with $m > 1$. Let $\{X_\alpha^{(k)}\}_{\alpha=1}^m$ be an operator basis for the operators in the support (as vectors) of J_k, with $X_1^{(k)} \equiv X$. Consider the completely mixed state $\rho_{\mathbb{1}} \equiv \mathbb{1}/d$ (d is the dimension of the Hilbert space). Let σ be some operator in the span of $\{X_\alpha^{(k)}\}_{\alpha=2}^m$ and consider the operator $\rho \equiv \rho_{\mathbb{1}} + \eta\sigma$ where η is a positive number chosen small enough so that ρ is positive. Applying $\mathcal{E}^n$ to ρ gives $\mathcal{E}^n(\rho) = \mathcal{E}^n(\rho_{\mathbb{1}}) + \eta\mathcal{E}^n(\sigma)$. Since $\mathcal{E}$ is TP, $\mathcal{E}^n(\rho_{\mathbb{1}})$ remains finite. However, since X is a rotating point, we know that $|\lambda_k| = 1$, and the entries of J_k^n grows in amplitude as n increases, and hence the entries of $\mathcal{E}^n(\sigma)$ (viewed as a vector) grow in amplitude. For large enough n (η fixed), there will be a choice of σ such that $\mathcal{E}^n(\rho)$ is no longer positive semidefinite. But this violates the assumption that $\mathcal{E}$ is a CPTP map. Hence, we must have that $m = 1$. ∎

Lemma 9.3 tells us that any Jordan block J_k with $m > 1$ must have $|\lambda_k| < 1$.

Now, let $\{Y_\beta\}$ be an operator basis for operators outside of $\mathrm{Rot}(\mathcal{E})$. Y_β's are the operators occurring in Jordan blocks with $|\lambda_k| < 1$, and hence $\lim_{n\to\infty} \mathcal{E}^n(Y_\beta) = 0$ since Eq. (B10) tells us that $\lim_{n\to\infty} J_k^n = 0$ if $|\lambda_k| < 1$. We can use $\{X_l\} \bigcup \{Y_\beta\}$ as an operator basis for $\mathcal{B}(\mathcal{H})$, and write any operator $A \in \mathcal{B}(\mathcal{H})$ as $A = \sum_l a_l X_l + \sum_\beta b_\beta Y_\beta$. Then,

$$\lim_{n\to\infty} \mathcal{E}^n(A) = \lim_{n\to\infty} \left(\sum_l a_l (\mathcal{E}_R)^n(X_l) + \sum_\beta b_\beta \mathcal{E}^n(Y_\beta) \right) = \sum_l a_l \lim_{n\to\infty} (\mathcal{E}_R)^n(X_l), \tag{B11}$$

assuming the limit $\lim_{n\to\infty}(\mathcal{E}_R)^n(X_l)$ exists for all l.

To work out what $\lim_{n\to\infty}(\mathcal{E}_R)^n(X_l)$ is, we need the following lemma.

Lemma 9.4. For every $\epsilon > 0$, there exists some $N_\epsilon \in \mathbb{N}$ such that $\|(\mathcal{E}_R)^{N_\epsilon} - \mathbb{1}_R\| < \epsilon$, where $\mathbb{1}_R$ is the identity operator on $\mathrm{Rot}(\mathcal{E})$.

Proof. Recall that $\mathcal{E}_R$ is a diagonal matrix, with entries $e^{i\phi_l}$, $l = 1, \ldots, M$, where $M = \dim[\mathrm{Rot}(\mathcal{E})]$. Therefore, $(\mathcal{E}_R)^n$ is also diagonal, with entries $e^{in\phi_l}$, and in particular $(\mathcal{E}_R)^0 = \mathbb{1}$. The set of *all* such matrices forms an n torus with a finite volume $(2\pi)^M$. Each $(\mathcal{E}_R)^n$ is surrounded by an ϵ neighborhood $\mathcal{N}_n$, containing all matrices X on the torus such that $\|(\mathcal{E}_R)^n - X\| < \epsilon$. Each such neighborhood has volume at least ϵ^M, and so if we consider the neighborhoods of $(\mathcal{E}_R)^n$ for $n = 0, \ldots, (2\pi/\epsilon)^M$, then at least one pair must overlap. Denote the pair with overlapping neighborhoods

If ϕ_l's are all rational multiples of 2π (that is, $\phi_l = \frac{2\pi p_l}{q_l}$, $p_l, q_l \in \mathbb{N}$), then choosing N_ϵ to be the lowest common multiple of all q_l works.

Otherwise, a more complicated analysis is required. To have

$$\|(\mathcal{E}_R)^{N_\epsilon} - \mathbb{1}_R\| = \max_l |\exp(iN_\epsilon\phi_l) - 1| = 2\max_l |\sin(N_\epsilon\phi_l/2)| < \epsilon,$$

it suffices to demand $N_\epsilon\phi_l (\mathrm{mod}\, 2\pi) < \epsilon$ for all l. Consider the point $[n\phi_1 (\mathrm{mod}\, 2\pi), \ldots, n\phi_M (\mathrm{mod}\, 2\pi)]$, where we always take the smallest non-negative value of $n\phi_l (\mathrm{mod}\, 2\pi)$. As n increases from 0, this point traces out a trajectory on the surface of an M-dimensional torus. If there is at least one ϕ_l that is a rational multiple of 2π, this trajectory will eventually close upon itself, and the path length of the trajectory is finite. If there is no such ϕ_l, the trajectory will cover the surface of the torus, which has finite area (since it is finite dimensional). Consider hyperspheres of (Euclidean) diameter ϵ centered at $[n\phi_1 (\mathrm{mod}\, 2\pi), \ldots, n\phi_M (\mathrm{mod}\, 2\pi)]$ for each $n \in \mathbb{N}$. Because the trajectory either has finite length or traverses a space of finite area, some of these hyperspheres will eventually overlap, that is, there exists finite r and $s > r$ such that the hyperspheres centered at points with $n = r$ and $n = s$ overlap. The distance between the centers of the overlapping hyperspheres is $\sqrt{\sum_l [(s-r)\phi_l (\mathrm{mod}\, 2\pi)]^2} < \epsilon$, which implies that $(s-r)\phi_l (\mathrm{mod}\, 2\pi) < \epsilon$ for all l. Therefore, we can choose $N_\epsilon = s - r$. ∎

We can view the limit $\lim_{n\to\infty}(\mathcal{E}_R)^n$ equivalently as the limit $\lim_{n\to\infty}(\mathcal{E}_R)^{N_\epsilon n}$. Intuitively, provided we choose ϵ to decrease fast enough, this should converge to $\mathbb{1}_R$. More precisely, we can write $(\mathcal{E}_R)^{N_\epsilon} = \mathbb{1}_R + \mathcal{G}_\epsilon$, where $\mathcal{G}_\epsilon$ is some map (need not be CP) on $\mathrm{Rot}(\mathcal{E})$ such that $\|\mathcal{G}_\epsilon\| < \epsilon$. Now consider the map $(\mathcal{E}_R)^{N_\epsilon n} = (\mathbb{1}_R + \mathcal{G}_\epsilon)^n = \sum_{m=0}^n \binom{n}{m} \mathcal{G}_\epsilon^m$, for $n \in \mathbb{N}$, which gives

$$\|(\mathcal{E}_R)^{N_\epsilon n} - \mathbb{1}_R\| \leqslant \sum_{m=1}^n \binom{n}{m} \left\|\mathcal{G}_\epsilon^m\right\| \leqslant \epsilon(2^n - 1). \tag{B12}$$

Let us choose $\epsilon = 3^{-n}$ (actually, $\epsilon = C_0^{-n}$ for any choice of $C_0 > 2$ works). Then taking the limit $n \to \infty$ of Eq. (B12), we conclude that $\lim_{n\to\infty}(\mathcal{E}_R)^{N_\epsilon n} = \mathbb{1}_R$.

From this, we see that Eq. (B11) can be rewritten as

$$\lim_{n\to\infty} \mathcal{E}^n(A) = \sum_l a_l X_l \quad \in \mathrm{Rot}(\mathcal{E}). \tag{B13}$$

Therefore, $\mathcal{E}_{\mathrm{inf}} \equiv \lim_{n\to\infty} \mathcal{E}^{nN_\epsilon}$ (with ϵ depending on n as above) is the projection onto $\mathrm{Rot}(\mathcal{E})$. Since a unitarily noiseless code is preserved under any power of $\mathcal{E}$, it must be preserved under $\mathcal{E}_{\mathrm{inf}}$, which gives the desired isometry condition. ∎

Note that $\mathcal{E}_{\mathrm{inf}}$ is CPTP simply because $\mathcal{E}$ is CPTP, and the set of CPTP maps on a finite-dimensional Hilbert space is closed under composition. Furthermore, it projects every operator onto the span of the rotating points of $\mathcal{E}$. Observe that $\mathrm{Rot}(\mathcal{E})$ is precisely the set of fixed points of $\mathcal{E}_{\mathrm{inf}}$.

[1] W. H. Zurek, Phys. Rev. D **24**, 1516 (1981); Rev. Mod. Phys. **75**, 715 (2003).

[2] G. M. Palma, K.-A. Suominen, and A. K. Ekert, Proc. R. Soc. London A **452**, 567 (1996); P. Zanardi and M. Rasetti, Phys. Rev. Lett. **79**, 3306 (1997); L.-M. Duan and G.-C. Guo, *ibid.* **79**, 1953 (1997); D. A. Lidar, I. L. Chuang, and K. B. Whaley, *ibid.* **81**, 2594 (1998).

[3] E. Knill, R. Laflamme, and L. Viola, Phys. Rev. Lett. **84**, 2525 (2000).

[4] L. Viola, E. Knill, and R. Laflamme, J. Phys. A **34**, 7067 (2001); L. Viola and E. Knill, Phys. Rev. A **68**, 032311 (2003).

[5] P. Zanardi, Phys. Rev. A **63**, 012301 (2000); J. Kempe, D. Bacon, D. A. Lidar, and K. B. Whaley, *ibid.* **63**, 042307 (2001).

[6] P. W. Shor, Phys. Rev. A **52**, R2493 (1995); A. M. Steane, Phys. Rev. Lett. **77**, 793 (1996); C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996); E. Knill and R. Laflamme, *ibid.* **55**, 900 (1997).

[7] C. Beny, A. Kempf, and D. W. Kribs, Phys. Rev. Lett. **98**, 100502 (2007); C. Beny, e-print arXiv:0802.0685; Ph.D. thesis, e-print arXiv:0901.3629.

[8] D. Bacon, D. A. Lidar, and K. B Whaley, Phys. Rev. A **60**, 1944 (1999).

[9] B. Schumacher and M. D. Westmoreland, Quant. Info. Proc. **1**, 5 (2002); R. Klesse, Phys. Rev. A **75**, 062315 (2007); F. Buscemi, *ibid.* **77**, 012309 (2008); H.-K. Ng and P. Mandayam, e-print arXiv:0909.0931.

[10] H. K. Ng and R. Blume-Kohout (in preparation).

[11] F. Ticozzi and L. Viola, Phys. Rev. A **81**, 032313 (2010).

[12] C. Beny and O. Oreshkov, Phys. Rev. Lett. **104**, 120501 (2010).

[13] A. Arias, A. Gheondea, and S. Gudder, J. Math. Phys. **43**, 5872 (2002).

[14] D. W. Kribs, Proc. Edinb. Math. Soc. **46**, 421 (2003).

[15] A. Frigerio, Lett. Math. Phys. **2**, 79 (1977); Commun. Math. Phys. **63**, 269 (1978).

[16] W. H. Zurek, Prog. Theor. Phys. **89**, 281 (1993).

[17] M.-D. Choi and D. W. Kribs, Phys. Rev. Lett. **96**, 050501 (2006).

[18] E. Knill, Phys. Rev. A **74**, 042301 (2006).

[19] D. W. Kribs and R. W. Spekkens, Phys. Rev. A **74**, 042329 (2006).

[20] R. Blume-Kohout, H. K. Ng, D. Poulin, and L. Viola, Phys. Rev. Lett. **100**, 030501 (2008).

[21] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).

[22] C. E. Shannon and W. W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).

[23] A. Shabani and D. A. Lidar, Phys. Rev. A **80**, 012309 (2009).

[24] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory* (Springer, Berlin, 1983).

[25] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[26] W. H. Zurek, Phys. Rev. D **26**, 1862 (1982).

[27] A. Shabani and D. A. Lidar, Phys. Rev. A **72**, 042303 (2005).

[28] F. Ticozzi and L. Viola, IEEE Trans. Autom. Control **53**, 2048 (2008).

[29] A. Granas and J. Dugundji, *Fixed Point Theory* (Springer, New York, 2003).

[30] D. Kribs, R. Laflamme, and D. Poulin, Phys. Rev. Lett. **94**, 180501 (2005).

[31] D. Kribs, R. Laflamme, D. Poulin, and M. Lesosky, Quantum Inf. Comput. **6**, 382 (2006).

[32] P. G. Kwiat, A. J. Berglund, J. B. Altepeter, and A. G. White, Science **290**, 498 (2000); D. Kielpinski *et al.*, *ibid.* **291**, 1013 (2001); E. M. Fortunato *et al.*, New J. Phys. **4**, 5 (2002); J. B. Altepeter *et al.*, Phys. Rev. Lett. **92**, 147901 (2004); M. Carravetta, O. G. Johannessen, and M. H. Levitt, *ibid.* **92**, 153003 (2004).

[33] L. Viola *et al.*, Science **293**, 2059 (2001); E. M. Fortunato, L. Viola, M. A. Pravia, E. Knill, R. Laflamme, T. F. Havel, and D. G. Cory, Phys. Rev. A **67**, 062303 (2003).

[34] D. G. Cory, M. D. Price, W. Maas, E. Knill, R. Laflamme, W. H. Zurek, T. F. Havel, and S. S. Somaroo, Phys. Rev. Lett. **81**, 2152 (1998); E. Knill, R. Laflamme, R. Martinez, and C. Negrevergne, *ibid.* **86**, 5811 (2001); N. Boulant, L. Viola, E. M. Fortunato, and D. G. Cory, *ibid.* **94**, 130501 (2005); J. Chiaverini *et al.*, Nature (London) **432**, 602 (2004).

[35] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, London, 1976).

[36] D. Pérez-García, M. M. Wolf, D. Petz, and M. B. Ruskai, J. Math. Phys. **47**, 083506 (2006).

[37] H. Barnum and E. Knill, J. Math. Phys. **43**, 2097 (2002).

[38] K. Davidson, C*-*Algebras by Example, Fields Institute Monographs* (Amer. Math. Soc., Providence, 1996).

[39] M. D. Choi and E. G. Effros, J. Funct. Anal. **24**, 156 (1977).

[40] G. Kuperberg, IEEE Trans. Inf. Theory **49**, 1465 (2003).

[41] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[42] R. Laflamme, C. Miquel, J.-P. Paz, and W. H. Zurek, Phys. Rev. Lett. **77**, 198 (1996).

[43] E. Knill, Nature (London) **434**, 39 (2005).

[44] J. Holbrook, D. Kribs, and R. Laflamme, Quant. Info. Proc. **2**, 381 (2004).

[45] R. Alicki and K. Lendi, *Quantum Dynamical Semigroups and Applications* (Springer, Berlin, 1987).

[46] F. Ticozzi and L. Viola, Automatica **45**, 2002 (2009).

[47] O. Oreshkov and J. Calsamiglia, e-print arXiv:1002.2219.

[48] H.-J. Briegel and B.-G. Englert, Phys. Rev. A **47**, 3311 (1993).

[49] C. Ahn, H. M. Wiseman, and G. J. Milburn, Phys. Rev. A **67**, 052310 (2003); B. A. Chase, A. J. Landahl, and J. M. Geremia, *ibid.* **77**, 032304 (2008); H. Mabuchi, New J. Phys. **11**, 105044 (2009).

[50] L. Viola, E. Knill, and S. Lloyd, Phys. Rev. Lett. **85**, 3520 (2000).

[51] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).

[52] C. E. Shannon, IRE Trans. Inf. Theory **IT-2**, 8 (1956).

[53] J. Körner and A. Orlitsky, IEEE Trans. Inf. Theory **44**, 2207 (1998).

[54] J. A. Wheeler and W. H. Zurek, *Quantum Theory and Measurement* (Princeton University Press, Princeton, 1983).

[55] E. Schrödinger, Proc. Cambridge Philos. Soc. **31**, 555 (1935); **32**, 446 (1936); F. Verstraete, Ph.D. thesis, Katholieke University Leuven, 2002; H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. **98**, 140402 (2007).

[56] G. Lindblad, Lett. Math. Phys. **47**, 189 (1999).

[57] K. Hoffman and R. Kunze, *Linear Algebra*, 2nd Ed. (Prentice Hall, New Jersey, 1971).

[58] B. M. Terhal and D. P. DiVincenzo, Phys. Rev. A **61**, 022301 (2000).