arXiv:1003.4994v2 [quant-ph] 26 Apr 2010

The Fidelity Alternative and Quantum Measurement Simulation*

Patrick Hayden[1] and Andreas Winter[2, 3]
[1]*School of Computer Science, McGill University, Montreal, Canada*
[2]*Department of Mathematics, University of Bristol, Bristol BS8 1TW, U.K.*
[3]*Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, Singapore 117542*
(Dated: 23 April 2010)

If a quantum system is subject to noise, it is possible to perform quantum error correction reversing the action of the noise if and only if no information about the system's quantum state leaks to the environment. In this article, we develop an analogous duality in the case that the environment approximately forgets the identity of the quantum state, a weaker condition satisfied by ϵ-randomizing maps. Specifically, we show that the environment approximately forgets quantum states if and only if the original channel approximately preserves pairwise fidelities of pure inputs, an observation we call the fidelity alternative. Using this tool, we then go on to study the task of using the output of a channel to simulate restricted classes of measurements on a space of input states. The case of simulating measurements that test whether the input state is an arbitrary pure state is known as equality testing or quantum identification. We establish that the optimal amortized rate at which quantum states can be identified through a noisy quantum channel is equal to the entanglement-assisted classical capacity of the channel, despite the fact that the task is quantum, not classical, and entanglement-assistance is not allowed. In particular, this rate is strictly positive for every quantum channel, including classical channels, despite the fact that the ability to identify cannot be cloned.

I. INTRODUCTION

Quantum channels in modern quantum information theory [1] are modeled as completely positive and trace-preserving maps $\mathcal{N} : \mathcal{S}(A) \to \mathcal{S}(B)$ between the state spaces of quantum systems with Hilbert spaces A and B. The requirement of *complete* positivity means that $\mathcal{N}$ is not just *positive*, mapping positive semidefinite operators to positive semidefinite operators, but that $\mathrm{id} \otimes \mathcal{N}$ is positive for the identity map id on any $\mathcal{S}(R)$. This distinction plays a central role in the geometry of entanglement because positive but not completely positive maps can be used to identify entangled quantum states [2]. This paper will take as its starting point a similar observation about channel norms.

The Stinespring dilation theorem establishes a fundamental property of quantum channels: for every channel $\mathcal{N}$ there exists an ancilla space E and an isometry $V : A \hookrightarrow B \otimes E$ such that $\mathcal{N}(\rho) = \mathrm{tr}_E\, V\rho V^\dagger$ [3]. This means that quantum noise can always be interpreted as information loss in an otherwise deterministic evolution. Since E and V are essentially unique (up to unitary equivalence), each channel $\mathcal{N}$ also has an associated *complementary channel* $\mathcal{N}^c : \mathcal{S}(A) \to \mathcal{S}(E)$, with $\mathcal{N}^c(\rho) = \mathrm{tr}_B\, V\rho V^\dagger$, which is uniquely defined up to coordinate changes of E.

In quantum Shannon theoretic error correction we try to find two channels $\mathcal{E}$ and $\mathcal{D}$ (an encoder and decoder) such that $\mathcal{D} \circ \mathcal{N} \circ \mathcal{E} \approx \mathrm{id}$. For now we shall consider the encoding $\mathcal{E}$ fixed, so that $\mathcal{N} \circ \mathcal{E}$ can be treated as a single channel. The central insight of quantum error correction [4–7] is that the existence of a decoding operation $\mathcal{D}$ for a channel $\mathcal{N}$, i.e.

$$\forall \rho \in \mathcal{S}(RA) \quad \left\| (\mathrm{id} \otimes \mathcal{D} \circ \mathcal{N})\rho^{RA} - \rho^{RA} \right\|_1 \leq \epsilon, \tag{1}$$

is equivalent to the complementary channel being *completely forgetful*: for all Hilbert spaces R,

$$\forall \rho, \sigma \in \mathcal{S}(RA) \quad \left\| (\mathrm{id} \otimes \mathcal{N}^c)\rho^{RA} - (\mathrm{id} \otimes \mathcal{N}^c)\sigma^{RA} \right\|_1 \leq \delta, \tag{2}$$

with a universal relation between ϵ and δ.

Here we determine a matching duality for the weaker property of the complementary channel being only (approximately) *forgetful*:

$$\forall \rho, \sigma \in \mathcal{S}(A) \quad \left\| \mathcal{N}^c(\rho^A) - \mathcal{N}^c(\sigma^A) \right\|_1 \leq \delta. \tag{3}$$

That this is a much weaker property was noticed in the contexts of approximate encryption and remote state preparation [8, 9]. The difference between Eqs. (2) and (3) is precisely the difference between two norms on superoperators, the naïve one inherited from the trace norm, and the so-called completely bounded norm [7, 10, 11]. Not surprisingly, Eq. (3) will hold provided the main channel approximately preserves the pairwise fidelities between input pure states, a property we call *geometry preservation*:

$$\forall |\psi\rangle, |\varphi\rangle \in A \quad \left| \|\varphi - \psi\|_1 - \|\mathcal{N}(\varphi) - \mathcal{N}(\psi)\|_1 \right| \leq \epsilon. \tag{4}$$

In fact, the reverse is also true. Our investigations will revolve around the *fidelity alternative*, which states that a channel $\mathcal{N}$ is geometry-preserving if and only if its complement $\mathcal{N}^c$ is approximately forgetful, with dimension-independent functions relating δ and ϵ. Thus, an isometry with two outputs can preserve geometry to at most one of them. Symmetrically, the isometry can be forgetful to at most one output. A single choice determines which output preserves geometry and which is forgetful; this choice gives the principle its name.

The geometry preservation property, though much weaker than transmission of quantum information, must nonetheless be considered a way of preserving coherence: by virtue of the fidelity alternative, geometry preservation cannot be cloned. Indeed, if a channel has multiple outputs, one of which is geometry-preserving, then the rest must be forgetful.

---

*A preliminary version of this paper was presented as a contributed talk at the 12th QIP workshop, Santa Fe (NM), 12-16 January 2009.

Via the fidelity alternative, the many known examples of approximately forgetful channels that are not completely forgetful also provide examples of geometry-preserving channels that are not correctable. Most strikingly, it is possible to preserve geometry while almost halving the number of qubits from input to output [12]. In that case, the geometry of the unit sphere in A is necessarily encoded into the eigenvectors *and* eigenvalues of the much smaller output state on B. In contrast to quantum error correction, dimension counting reveals the mixedness of the output state to be crucial to preserving the geometry. Some of the geometry of the input state space of pure quantum states is thus faithfully encoded as noise in the output state.

Moreover, the analogy with the quantum error correction duality can be made much stronger. There is a channel communication task very similar to quantum state transmission which is intimately related to geometry preservation: *quantum identification* [12, 13].

Quantum identification is a cooperative communication game between two parties – conventionally called Alice and Bob – where Alice has a given quantum state that she encodes in some way into the channel, and Bob only wants to simulate measurements consisting of an arbitrary pure state projector and its complement, which can interpreted as performing the experiment asking "Is this the state?" [12]. The idea is that Alice has an encoding channel $\mathcal{E}$ and Bob has, for every pure state φ, a POVM $(D_\varphi, \mathbb{1} - D_\varphi)$ such that

$$\forall|\psi\rangle, |\varphi\rangle \quad \left|\operatorname{tr}\big((\mathcal{N}\circ\mathcal{E})\psi\big)D_\varphi - \operatorname{tr}\psi\varphi\right| \leq \epsilon. \tag{5}$$

Such an object is called an ϵ*-quantum-ID code*. (The name is adapted from the classical case [14–16].)

Note that Bob measures the output of the channel, but the quality of the code is measured by how well the statistics of this measurement approximate the statistics of the ideal measurement he wants to perform on the message state. While it may seem that this is an odd way of defining a quantum communication task, normal quantum error correction can also be described this way; namely, Bob wants to be able to simulate *all* measurements on the message state. Clearly, if he can perform quantum error correction in the usual sense, then he can perform the simulation. But conversely, it follows from the methods of [17–19] that if he only has two measurements approximating generalized X and Z observables sufficiently well, he can build a quantum error correction procedure $\mathcal{D}$. Moreover, a quantum-ID code with $\epsilon = 0$ is itself a quantum error correcting code; there is no difference between error correction and identification if both tasks are to be performed perfectly. Even the task of transmitting classical information is conveniently reflected in this framework. In that case, Bob only wants to simulate the measurement of the generalized Z observable.

With this, one can define in the usual way a *quantum-ID capacity* $Q_{\rm ID}(\mathcal{N})$ of many uses of the channel as the largest rate of qubits that can be encoded and decoded as in Eq. (5) with vanishing error – see Section III for details. Previously it was only known that for the noiseless qubit channel id_2, $Q_{\rm ID}(\mathrm{id}_2) = 2$, double the value of both the the quantum and classical transmission capacities [12].

While reasoning directly about quantum identification codes has proved challenging, the duality between geometry preservation and approximate forgetfulness provides a new approach to studying them. Up to some technical conditions, geometry preservation is equivalent to the existence of a quantum identification code. It is therefore possible to construct quantum identification codes by finding approximately forgetful maps. This approach is fruitful because destroying information is a comparatively indiscriminate task. Indeed, the analogous strategy has led to a number of straightforward proofs of the hashing bound on the quantum capacity of a quantum channel [18, 20–22]. Classical data is not immune to analysis by purification either. The duality between privacy amplification and data compression with quantum side information has recently led to a proof in this spirit [23] of the Holevo-Schumacher-Westmoreland theorem on the classical capacity of a quantum channel [5, 24] .

With the fidelity alternative in hand, it is even possible to calculate a simple formula for an amortized version of the quantum identification capacity; it is exactly equal to the entanglement-assisted classical capacity of a quantum channel.

A. Structure of the paper

Section II contains the formal statement and proof of the fidelity alternative. The duality is studied in more detail in Section III, where forgetfulness is shown to be nearly equivalent to quantum identification. In that section we provide a simple statement whose proof eliminates many technical difficulties, as well as a more flexible version that we prove from first principles. Section IV uses the flexible version of the equivalence to construct quantum identification codes for memoryless quantum channels.

B. Notation

We will restrict our attention throughout to finite dimensional Hilbert spaces. If A is a Hilbert space, we write $\mathcal{S}(A)$ for the set of density operators acting on A. Also, if A and B are two finite dimensional Hilbert spaces, we write $AB \equiv A \otimes B$ for their tensor product. The Hilbert spaces on which linear operators act will be denoted by a superscript. For instance, we write φ^{AB} for a density operator on AB. Partial traces will be abbreviated by omitting superscripts, such as $\varphi^A \equiv \operatorname{tr}_B \varphi^{AB}$. We use a similar notation for pure states, e.g. $|\psi\rangle^{AB} \in AB$, while abbreviating $\psi^{AB} \equiv |\psi\rangle\langle\psi|^{AB}$. We will write id_A for the identity map on $\mathcal{S}(A)$ and id_2 for the identity qubit channel. The symbol $\mathbb{1}^A$ will be reserved for the identity matrix acting on the Hilbert space A and $\pi^A = \mathbb{1}^A/\dim A$ for the maximally mixed state on A.

The trace norm of an operator, $\|X\|_1$ is defined to be $\operatorname{tr}|X| = \operatorname{tr}\sqrt{X^\dagger X}$. The similarity of two density operators φ and ψ can be measured by *trace distance* $\frac{1}{2}\|\varphi - \psi\|_1$, which is equal to the maximum over all possible measurements of the variational distance between the outcome probabilities for the two states. The trace distance is zero for identical states and one for perfectly distinguishable states.

A complementary measure is the mixed state fidelity

$$F(\varphi,\psi) = \left\|\sqrt{\varphi}\sqrt{\psi}\right\|_1^2 = \left(\operatorname{tr}\sqrt{\sqrt{\varphi}\psi\sqrt{\varphi}}\right)^2, \tag{6}$$

defined such that when one of the states is pure, $F(\varphi,\psi) = \operatorname{tr}\varphi\psi$. More generally, the fidelity is equal to one for identical states and zero for perfectly distinguishable states. We will make frequent use of the following fundamental inequality between fidelity and trace distance of states [25, Prop. 5]:

$$1-\sqrt{F(\varphi,\psi)} \leq \frac{1}{2}\|\varphi-\psi\|_1 \leq \sqrt{1-F(\varphi,\psi)}. \tag{7}$$

Both measures can be extended to unnormalized states, but Eq. (7) need not hold in that case. Further properties of the distance measures are collected in Appendix A.

II. THE FIDELITY ALTERNATIVE

Our investigations will revolve around the duality between geometry preservation and approximate forgetfulness, which we call the fidelity alternative. The rigorous statement is as follows:

Theorem 1 (Fidelity alternative) *Let* $\mathcal{N}:\mathcal{S}(A)\to\mathcal{S}(B)$ *be a quantum channel with complementary channel* $\mathcal{N}^c:\mathcal{S}(A)\to\mathcal{S}(E)$. *Approximate geometry preservation on* B *implies approximate forgetfulness for* E. *That is,*

$$\begin{aligned}\forall|\psi\rangle,|\varphi\rangle\in A \quad & \|\varphi-\psi\|_1 - \|\mathcal{N}(\varphi)-\mathcal{N}(\psi)\|_1 \leq \delta \\ \textit{implies}\ \forall|\psi\rangle,|\varphi\rangle\in A \quad & \|\mathcal{N}^c(\varphi)-\mathcal{N}^c(\psi)\|_1 \leq 4\sqrt{2}\delta^{1/4}.\end{aligned}$$

Conversely, approximate forgetfulness for E *implies approximate geometry preservation on* B*:*

$$\begin{aligned}\forall|\psi\rangle,|\varphi\rangle\in A \quad & \|\mathcal{N}^c(\varphi)-\mathcal{N}^c(\psi)\|_1 \leq \epsilon \ \textit{implies} \\ \forall|\psi\rangle,|\varphi\rangle\in A \quad & \|\varphi-\psi\|_1 - \|\mathcal{N}(\varphi)-\mathcal{N}(\psi)\|_1 \leq 4\sqrt{2\epsilon}.\end{aligned}$$

Note that we have dropped an absolute value sign as compared to Eq. (4) since $\|\varphi-\psi\|_1 \geq \|\mathcal{N}(\varphi)-\mathcal{N}(\psi)\|_1$ holds automatically for all quantum channels $\mathcal{N}$. (See, for example, [26].)

The duality is a straightforward consequence of two basic results in quantum information theory. The first is that the ability to transmit classical data in two conjugate bases is equivalent to the ability to transmit entanglement. That observation is the basis for the stabilizer approach to quantum error correcting codes [27]. Here we will use a clean approximate formulation due to Renes [19]. The second result is the continuity of the Stinespring dilation of a quantum channel, established by Kretschmann *et al.* [7]. Here we only need a corollary, which can be interpreted as a bound on the information-disturbance trade-off. The theorem is stated in terms of the following norms:

Definition 2 *For a linear superoperator* $\Gamma:\mathcal{S}(A)\to\mathcal{S}(B)$, *let*

$$\|\Gamma\|_\diamond^{(k)} = \max_{\|X\|_1\leq 1}\left\|(\mathrm{id}_k\otimes\Gamma)X\right\|_1,$$

where X *is an operator on* $\mathbb{C}^k\otimes A$. *Define* $\|\Gamma\|_\diamond = \sup_k\|\Gamma\|_\diamond^{(k)}$, *the* completely bounded trace norm *[10] (also known as* diamond norm *[11]).*

Note that the convexity of the trace norm implies that the supremum is achieved on a rank-one operator (if Γ is Hermitian-preserving, in fact on a pure quantum state). Since any operator on A can be "purified" by a system of dimension $\dim A$, it follows that the supremum is achieved when $k=\dim A$.

Theorem 3 (Information-disturbance [7]) *Let* $V:A\to B\otimes E$ *be an isometric extension of the channel* $\mathcal{N}:\mathcal{S}(A)\to\mathcal{S}(B)$ *and let* $\mathcal{N}^c:\mathcal{S}(A)\to\mathcal{S}(E)$ *be the complementary channel. Fix a state* $\rho\in\mathcal{S}(A)$ *and let* $\mathcal{R}:\mathcal{S}(A)\to\mathcal{S}(E)$ *be the channel taking all inputs to* $\mathcal{N}^c(\rho)$. *Then*

$$\frac{1}{4}\inf_{\mathcal{D}}\|\mathcal{D}\circ\mathcal{N}-\mathrm{id}\|_\diamond^2 \leq \|\mathcal{N}^c-\mathcal{R}\|_\diamond \leq 2\inf_{\mathcal{D}}\|\mathcal{D}\circ\mathcal{N}-\mathrm{id}\|_\diamond^{1/2}.$$

Both infimums are over all quantum channels. □

The proof of the fidelity alternative is now a fairly routine matter of combining these results:

Proof of Theorem 1 We begin by assuming approximate geometry preservation. Fix $|\varphi\rangle\perp|\psi\rangle$ in A then set $T=\operatorname{span}(|\varphi\rangle,|\psi\rangle)$. Suppose that

$$\|\mathcal{N}(\omega)-\mathcal{N}(\xi)\|_1 \geq \|\omega-\xi\|_1 - \delta$$

for all $|\omega\rangle,|\xi\rangle\in A$. Then if $|\chi_\pm\rangle = \frac{1}{\sqrt{2}}(|\varphi\rangle\pm|\psi\rangle)$, we have

$$\begin{aligned}\|\mathcal{N}(\varphi)-\mathcal{N}(\psi)\|_1 &\geq 2-\delta \quad\text{and}\\ \|\mathcal{N}(\chi_+)-\mathcal{N}(\chi_-)\|_1 &\geq 2-\delta.\end{aligned}$$

We can therefore transmit data in two conjugate bases through $\mathcal{N}$, which implies that entanglement is also faithfully transmitted. In particular [19, Thm. 1] (with "guessing probability" $1-\delta/2$) implies that there exists a channel $\mathcal{D}:\mathcal{S}(B)\to\mathcal{S}(T)$ such that

$$\left\|(\mathrm{id}_2\otimes\mathcal{D}\circ\mathcal{N})\Phi-\Phi\right\|_1 \leq 2\sqrt{\delta},$$

where $|\Phi\rangle = \frac{1}{\sqrt{2}}(|0\rangle|\varphi\rangle+|1\rangle|\psi\rangle)$. But trace norm monotonicity with respect to dephasing the first system then gives

$$\begin{aligned}&\left\|(\mathrm{id}_2\otimes\mathcal{D}\circ\mathcal{N})\Phi-\Phi\right\|_1\\ &\geq \frac{1}{2}\Big\||0\rangle\langle 0|\otimes[(\mathcal{D}\circ\mathcal{N})\varphi-\varphi]\\ &\qquad + |1\rangle\langle 1|\otimes[(\mathcal{D}\circ\mathcal{N})\psi-\psi]\Big\|_1\\ &= \frac{1}{2}\|(\mathcal{D}\circ\mathcal{N})\varphi-\varphi\|_1 + \frac{1}{2}\|(\mathcal{D}\circ\mathcal{N})\psi-\psi\|_1.\end{aligned}$$

Combining this with Lemma 19 in Appendix A implies that $\|\mathcal{D}\circ\mathcal{N}-\mathrm{id}_2\|_\diamond \le 8\sqrt{\delta}$. The information-disturbance theorem (Theorem 3) then implies that for all $|\omega\rangle \in T$

$$\|\mathcal{N}^c(\varphi) - \mathcal{N}^c(\omega)\|_1 \le 2(8\sqrt{\delta})^{1/2} = 4\sqrt{2}\delta^{1/4}.$$

Since T is an arbitrary two-dimensional subspace of A, however, the inequality must hold for all $|\varphi\rangle$ and $|\omega\rangle$ in A.

For the converse, suppose that, for all states $|\varphi\rangle, |\psi\rangle \in A$, the inequality $\|\mathcal{N}^c(\varphi) - \mathcal{N}^c(\psi)\|_1 \le \epsilon$ holds. Fix $|\varphi\rangle$ and $|\psi\rangle$ then let $\tilde{\mathcal{N}}^c$ be the restriction of $\mathcal{N}^c$ to states on $T = \mathrm{span}(|\varphi\rangle, |\psi\rangle)$. Let $\mathcal{R}$ be the channel on $\mathcal{S}(T)$ that always outputs $\tilde{\mathcal{N}}^c(\psi)$. Then once more by Lemma 19 in Appendix A, $\|\tilde{\mathcal{N}}^c - \mathcal{R}\|_\diamond \le 2\epsilon$. By Theorem 3, there exists a channel $\mathcal{D} : \mathcal{S}(B) \to \mathcal{S}(T)$ such that for all $|\omega\rangle \in T$,

$$\frac{1}{4}\,\|(\mathcal{D}\circ\mathcal{N})\omega - \omega\|_1^2 \le 2\epsilon.$$

Applying the triangle inequality several more times gives:

$$\begin{aligned} 4\sqrt{2\epsilon} &\ge \|(\mathcal{D}\circ\mathcal{N})\varphi - \varphi\|_1 + \|(\mathcal{D}\circ\mathcal{N})\psi - \psi\|_1 \\ &\ge \|\varphi - \psi\|_1 - \|(\mathcal{D}\circ\mathcal{N})(\varphi - \psi)\|_1 \\ &\ge \|\varphi - \psi\|_1 - \|\mathcal{N}(\varphi - \psi)\|_1\,, \end{aligned}$$

where the final inequality used that the quantum channel $\mathcal{D}$ cannot increase the trace norm. Rearranging the final expression gives the desired inequality. □

III. QUANTUM IDENTIFICATION

Quantum identification allows a sender to transmit arbitrary quantum states but only allows the receiver to perform a restricted set of measurements, namely tests to determine whether the transmitted state consists of an arbitrary target state. The receiver gets to choose the target state *after* the sender has transmitted, so the code must work for all targets. If the test can be performed perfectly, then quantum identification is easily seen to be equivalent to quantum state transmission, but in the approximate setting, the tasks are not equivalent.

Definition 4 [12] *An ϵ-quantum-ID code for the channel $\mathcal{N} : \mathcal{S}(A) \to \mathcal{S}(B)$ consists of an encoding map $\mathcal{E} : \mathcal{S}(S) \to \mathcal{S}(A)$ and, for every pure state $|\varphi\rangle \in S$, a POVM $(D_\varphi, \mathbb{1} - D_\varphi)$ acting on $\mathcal{S}(B)$ such that*

$$\forall |\psi\rangle, |\varphi\rangle \in S \quad \Big|\mathrm{tr}\,\big((\mathcal{N}\circ\mathcal{E})\psi\big)D_\varphi - |\langle\varphi|\psi\rangle|^2\Big| \le \epsilon.$$

If the receiver had been able to perform the measurement $(|\varphi\rangle\langle\varphi|, \mathbb{1} - |\varphi\rangle\langle\varphi|)$ on the input state $|\psi\rangle$, then he would have observed outcome $|\varphi\rangle\langle\varphi|$ with probability $|\langle\varphi|\psi\rangle|^2$. The definition therefore ensures that the receiver can simulate the measurement for all input and target states.

Many variants of the definition have been proposed. In particular, one could imagine drawing a distinction between oblivious ID codes, in which the sender is only given a physical quantum state to send, and visible ID codes, in which the sender knows the identity of the state she is trying to transmit [12]. Entanglement assistance is also interesting and exceptionally powerful in the visible setting [28]. A different task that is nonetheless similar in spirit is to use quantum states as "fingerprints" for identifying classical messages in a model where pairs of messages are to be compared by a referee [29]. For comparing quantum states, however, the simple definition considered here is probably the most natural.

If we integrate the encoding $\mathcal{E}$ and noisy channel $\mathcal{N}$ from Definition 4 into a single map with output B and environment E, we may think of the code Hilbert space S as a subspace of $B\otimes E$. More formally, if we let V be the Stinespring dilation of $\mathcal{N}\circ\mathcal{E}$, then $V : S \hookrightarrow B\otimes E$ and we can identify the code with a subspace of $B\otimes E$. This identification simplifies the notation and we will use it for the remainder of the paper.

The main result of this section is a demonstration that a subspace of $B\otimes E$ is a quantum-ID code for B iff it is approximately forgetful for E. (There is a small technical caveat to the statement: the reduced states on E must also obey a regularity condition for the reverse implication to hold, but we will defer discussion of the details.) For the moment, let us begin by considering the relationship between quantum identification and geometry preservation.

Lemma 5 *Let $S \subseteq B\otimes E$ be a subspace of a tensor product Hilbert space that is an ϵ-quantum-ID code for B. In other words, suppose that, for each pure state $|\varphi\rangle \in S$, there exists an operator $0 \le D_\varphi \le \mathbb{1}$ on B such that for all pure states $|\varphi\rangle, |\psi\rangle \in S$,*

$$\left|\mathrm{tr}\,\psi^B D_\varphi - \mathrm{tr}\,\psi\varphi\right| \le \epsilon.$$

Then, for all $|\varphi\rangle, |\psi\rangle \in S$,

$$F(\varphi,\psi) \le F(\varphi^B,\psi^B) \le F(\varphi,\psi) + 4\sqrt{\epsilon}.$$

Proof Consider the measurement $(D_\varphi, \mathbb{1} - D_\varphi)$ and associated channel $M : \rho \mapsto (\mathrm{tr}\,\rho D_\varphi, 1 - \mathrm{tr}\,\rho D_\varphi)$ which acts on $\mathcal{S}(B)$. By applying the monotonicity of the fidelity under quantum channels to tr_E and M, we get

$$\begin{aligned} F(\psi,\varphi) \le F(\psi^B,\varphi^B) &\le F\big(M(\psi^B), M(\varphi^B)\big) \\ &\le \left(\sqrt{\mathrm{tr}\,\psi^B D_\varphi} + \sqrt{\epsilon}\right)^2 \\ &\le F(\psi,\varphi) + 2\sqrt{\epsilon} + \epsilon + \epsilon, \end{aligned}$$

which proves the lemma. □

The fidelity is therefore approximately preserved by quantum identification codes. Geometry preservation is defined in terms of the trace distance, however, not the fidelity. While it is indeed the case that quantum identification codes preserve geometry, the argument is somewhat more delicate because applying the measurement $(D_\varphi, \mathbb{1} - D_\varphi)$ causes a significant drop in the trace distance even as it leaves the fidelity nearly unchanged. Instead, Theorem 7 will allow us to infer that quantum identification codes preserve geometry by virtue of the fact that their complementary channels are forgetful.

In order to succeed at quantum identification, the following lemma demonstrates that it is sufficient to be able to identify orthogonal states:

Lemma 6 *Let $S \subseteq B \otimes E$ be a subspace of a tensor product Hilbert space such that for $|\varphi\rangle \in S$ there exists $0 \leq D_\varphi \leq \mathbb{1}$ acting on B satisfying*

$$\operatorname{tr}\varphi^B D_\varphi \geq 1-\delta \quad \text{and} \quad \operatorname{tr}\psi^B D_\varphi < \delta$$

whenever $|\psi\rangle \in S$ is orthogonal to $|\varphi\rangle$. Then S is a quantum identification code with error probability $\delta + 2\sqrt{\delta}$.

Proof Let $|\varphi\rangle, |\psi\rangle \in S$ be arbitrary and let $|\varphi'\rangle$ be orthogonal to $|\varphi\rangle$ in $\operatorname{span}(|\varphi\rangle, |\psi\rangle)$. Write

$$|\psi\rangle = \alpha|\varphi\rangle + \beta|\varphi'\rangle.$$

Expanding gives

$$\begin{aligned}\operatorname{tr}\psi^B P &= |\alpha|^2 \operatorname{tr}\varphi^B P + |\beta|^2 \operatorname{tr}\varphi'^B P \\ &\quad + \alpha\overline{\beta}\operatorname{tr}|\varphi\rangle\langle\varphi'|(P\otimes\mathbb{1}) + \overline{\alpha}\beta\operatorname{tr}|\varphi'\rangle\langle\varphi|(P\otimes\mathbb{1}),\end{aligned}$$

which results in

$$\begin{aligned}\left|\operatorname{tr}\psi^B P - |\alpha|^2\right| &\leq |\alpha|^2(1-\operatorname{tr}\varphi^B P) \\ &\quad + |\beta|^2\operatorname{tr}\varphi'^B P + 2|\alpha\beta||\langle\varphi|(P\otimes\mathbb{1})|\varphi'\rangle| \\ &\leq |\alpha|^2(1-\operatorname{tr}\varphi^B P) \\ &\quad + |\beta|^2\operatorname{tr}\varphi'^B P + 2|\alpha\beta|\sqrt{\langle\varphi'|(P\otimes\mathbb{1})|\varphi'\rangle} \\ &\leq \delta + 2\sqrt{\delta},\end{aligned}$$

where we have used the Cauchy-Schwarz inequality and the assumption that orthogonal states in S can be well discriminated. □

Now we are ready to state and prove our main result on the duality between quantum identification and approximate forgetfulness:

Theorem 7 (Identification and forgetfulness) *Quantum-ID codes and forgetfulness are dual in the following quantitative sense. If a subspace $S \subseteq B \otimes E$ is an ϵ-quantum-ID code for B, then E is approximately δ-forgetful:*

$$\forall|\varphi\rangle,|\psi\rangle \in S \quad \frac{1}{2}\left\|\varphi^E - \psi^E\right\|_1 \leq \delta := 7\sqrt[4]{\epsilon}.$$

Conversely, if E is approximately δ-forgetful, then geometry is approximately preserved on B:

$$\forall|\varphi\rangle,|\psi\rangle \in S \quad \left\|\varphi-\psi\right\|_1 - \left\|\varphi^B - \psi^B\right\|_1 \leq \epsilon := 4\sqrt{2\delta}$$

If, in addition, the nonzero eigenvalues of φ^B lie in the interval $[\mu, \lambda]$ for all $|\varphi\rangle \in S$, then S is an η-quantum-ID code for $\eta := 3\delta^{1/4}\sqrt{\lambda/\mu}$.

Remark While it would be desirable to eliminate the eigenvalue condition at the end of the theorem, the condition is fairly natural in this context. If the reduced states φ^E are very close to a single state σ^E for all $|\varphi\rangle \in S$, then all the $|\varphi\rangle$ are very close to being purifications of σ^E, meaning that they differ from one another only by a unitary plus a small perturbation. If σ^E is the maximally mixed state or close to it, then the assumption will be satisfied. □

Proof For the first part, recall that if S is a quantum-ID code with error probability ϵ, then for each pure state $|\varphi\rangle \in S$ there exists an operator $0 \leq D_\varphi \leq \mathbb{1}$ on B such that for all pure states $|\varphi\rangle, |\psi\rangle \in S$,

$$\left|\operatorname{tr}\psi^B D_\varphi - \operatorname{tr}\psi\varphi\right| \leq \epsilon.$$

Just as in the proof of Theorem 1, the hypothesis implies that data can be transmitted in two conjugate bases with guessing probability $1-\epsilon$. Running exactly the same argument as was made in that proof gives that for all $|\varphi\rangle, |\psi\rangle \in S$,

$$\frac{1}{2}\left\|\varphi^E - \psi^E\right\|_1 \leq 4\sqrt{2}(2\epsilon)^{1/4} \leq 7\epsilon^{1/4}. \tag{8}$$

The second part is just a restatement of one direction of the fidelity alternative, but it is a useful step on the way to the third part, which is more challenging since it requires the construction of the decoder, that is, the operators D_φ.

Indeed, given $|\varphi\rangle \in S$, and arbitrary $|\psi\rangle \perp |\varphi\rangle$ in S, we learn from the second part that

$$\|\varphi^B - \psi^B\|_1 \geq 2 - 4\sqrt{2\delta}. \tag{9}$$

By Helstrom's theorem on the optimal discrimination of φ^B and ψ^B [30], there exists a projector $P_{\varphi,\psi}$ on B such that

$$\operatorname{tr}\varphi^B P_{\varphi,\psi} \geq 1 - 2\sqrt{2\delta}, \quad \operatorname{tr}\psi^B P_{\varphi,\psi} \leq 2\sqrt{2\delta}. \tag{10}$$

The problem with using $P_{\varphi,\psi}$ as the decoding is that this projector may indeed depend not only on φ, but also on ψ. Still, let us confirm first that if we manage to find one effect operator D_φ that can deal with all ψ at once, then by Lemma 6 we'll be done. Our strategy for doing so will be to first extend Eq. (10) to all mixed states orthogonal to $|\varphi\rangle$ and supported on S, and then use a minimax argument to extract a single operator independent of ψ.

Lemma 17 in Appendix A can be used directly to see that for all mixed states σ supported on S and orthogonal to φ,

$$F(\varphi^B, \sigma^B) \leq \frac{\lambda^2}{\mu^2}\max F(\varphi^B, \psi^B) \leq 2\delta\frac{\lambda^2}{\mu^2},$$

where the maximization is over all $|\psi\rangle \in S$ orthogonal to $|\varphi\rangle$ and the second inequality is an application of Eq. (7) to Eq. (9). Applying Eq. (7) a second time gives

$$\frac{1}{2}\left\|\varphi^B - \psi^B\right\|_1 \geq 1 - \sqrt{2\delta}\frac{\lambda}{\mu}.$$

Applying Helstrom's theorem to φ^B and σ^B yields a projector P_σ with

$$\operatorname{tr}\varphi^B P_\sigma - \operatorname{tr}\sigma^B P_\sigma \geq 1 - \sqrt{2\delta}\frac{\lambda}{\mu}.$$

Ky Fan's minimax theorem then ensures the existence of a saddle point in the following two-player game [31]. One player selects $0 \leq P \leq \mathbb{1}$ while the other player selects a state σ supported on S and orthogonal to φ. The strategy spaces are therefore closed and convex. The payoff function is $1 - \operatorname{tr} \varphi^B P + \operatorname{tr} \sigma^B P$, which is linear in each argument. Thus, the minimax theorem guarantees that there exists an operator $0 \leq D_\varphi \leq \mathbb{1}$ such that for all σ supported on S and orthogonal to φ,

$$\begin{aligned} \operatorname{tr} \varphi^B D_\varphi &\geq 1 - \sqrt{2\delta}\frac{\lambda}{\mu}, \\ \operatorname{tr} \sigma^B D_\varphi &\leq \sqrt{2\delta}\frac{\lambda}{\mu}, \end{aligned}$$

and applying Lemma 6 finishes the proof. □

Unfortunately, Theorem 7 is not quite strong enough to prove our main result on the quantum identification capacity. To control that ratio of the largest to smallest eigenvalues of the coding states, we need to act on them by typical projectors that cause a slight distortion. To accomodate this complication, we will instead use the following slightly more flexible version of the converse that behaves better with respect to the distortion. In particular, the amount of distortion enters the bound on the quality of the quantum-ID code in a term independent of the eigenvalue constraint. That separation proves to be crucial because the eigenvalues cannot be controlled independent of the distortion.

Theorem 8 *Let $S \subseteq B \otimes E$ be a subspace and $0 \leq X \leq \mathbb{1}$ an operator acting on $B \otimes E$ such that $\operatorname{tr}(X\varphi X^\dagger) \geq 1 - \epsilon$ for all $|\varphi\rangle \in S$, with $0 \leq \epsilon \leq 1/15$. For any state $|\omega\rangle \in S$, write $\tilde{\omega} = X\omega X^\dagger$. If there exists a state Ω such that*

$$\forall |\varphi\rangle \in S \quad \big\|\tilde{\Omega}^E - \tilde{\varphi}^E\big\|_1 \leq \delta$$

and, in addition, the nonzero eigenvalues of $\tilde{\Omega}^E$ lie in the interval $[\mu, \lambda]$, then S is an η-quantum-ID code for $\eta := 3(30\lambda\delta/\mu + 4\sqrt{\epsilon})^{1/2}$.

Proof Let $|\varphi\rangle$ and $|\psi\rangle$ be orthonormal states in S. We will begin by showing that $\tilde{\varphi}^B$ and $\tilde{\psi}^B$ can be effectively distinguished. To this end, consider the states

$$\begin{aligned} |\vartheta_\pm\rangle &= \frac{1}{\sqrt{2}}|\varphi\rangle \pm \frac{1}{\sqrt{2}}|\psi\rangle, \\ |\chi_\pm\rangle &= \frac{1}{\sqrt{2}}|\varphi\rangle \pm \frac{i}{\sqrt{2}}|\psi\rangle, \end{aligned}$$

which form two orthogonal pairs. Then

$$\begin{aligned} \tilde{\vartheta}^E_\pm &= \frac{1}{2}\tilde{\varphi}^E + \frac{1}{2}\tilde{\psi}^E \pm \frac{1}{2}\big(\operatorname{tr}_B |\tilde{\varphi}\rangle\langle\tilde{\psi}| + \operatorname{tr}_B |\tilde{\psi}\rangle\langle\tilde{\varphi}|\big), \\ \chi^E_\pm &= \frac{1}{2}\tilde{\varphi}^E + \frac{1}{2}\tilde{\psi}^E \mp \frac{i}{2}\big(\operatorname{tr}_B |\tilde{\varphi}\rangle\langle\tilde{\psi}| - \operatorname{tr}_B |\tilde{\psi}\rangle\langle\tilde{\varphi}|\big), \end{aligned}$$

and, by assumption,

$$\frac{1}{2}\|\tilde{\vartheta}^E_+ - \tilde{\vartheta}^E_-\|_1 \leq \delta \quad \text{and} \quad \frac{1}{2}\|\tilde{\chi}^E_+ - \tilde{\chi}^E_-\|_1 \leq \delta.$$

Combining these relations reveals that $\|\operatorname{tr}_B |\tilde{\varphi}\rangle\langle\tilde{\psi}| \pm \operatorname{tr}_B |\tilde{\psi}\rangle\langle\tilde{\varphi}|\|_1 \leq 4\delta$, hence by the triangle inequality, $\|\operatorname{tr}_B |\tilde{\varphi}\rangle\langle\tilde{\psi}|\|_1 \leq 8\delta$. But this gives us, by virtue of Lemma 16,

$$F(\tilde{\varphi}^B, \tilde{\psi}^B) \leq 64\delta^2. \tag{11}$$

To proceed as in the proof of Theorem 7, we need to show that any $|\varphi\rangle \in S$ and mixed state σ supported on the orthogonal complement of $|\varphi\rangle$ in S can also be distinguished. In order to apply Lemma 17 in Appendix A, we will show that the largest and smallest nonzero eigenvalues of φ^B, or equivalently, φ^E, are well-behaved modulo a little bit of truncation. Indeed, let (O) and (p) be the eigenvalues of $\tilde{\Omega}^E$ and $\tilde{\varphi}^E$, respectively, in nonincreasing order. Then

$$\big\|(O) - (p)\big\|_1 \leq \big\|\tilde{\Omega}^E - \tilde{\varphi}^E\big\|_1 \leq \delta.$$

Define the set

$$J = \big\{j \,:\, (1-\gamma)p_j \leq O_j \leq (1+\gamma)p_j\big\}.$$

Then

$$\gamma \sum_{j \notin J} p_j \leq \sum_{j \notin J} |O_j - p_j| \leq \delta,$$

implying that

$$\sum_{j \in J} p_j = \sum_j p_j - \sum_{j \notin J} p_j \geq (1 - \epsilon) - \epsilon/\gamma.$$

Fixing $\gamma = 1/2$ implies that for each $|\varphi\rangle \in S$, there is a positive semidefinite operator $\hat{\varphi}^B \leq \tilde{\varphi}^B$ satisfying $\operatorname{tr} \hat{\varphi}^B \geq 1 - 3\epsilon$ and whose eigenvalues lie in the interval $[\mu/2, 3\lambda/2]$.

Now let $|\varphi\rangle \in S$ and consider any state $\sigma = \sum_i q_i \psi_i$ whose support lies in the the orthogonal complement of $|\varphi\rangle$ in S. Let $\hat{\sigma} = \sum_i q_i \hat{\psi}_i$. Then by Lemma 17,

$$\begin{aligned} F(\hat{\varphi}^B, \hat{\sigma}^B) &\leq \frac{9\lambda^2}{\mu^2} \max F(\hat{\varphi}^B, \hat{\psi}^B) \\ &\leq \frac{9\lambda^2}{\mu^2} \max F(\tilde{\varphi}^B, \tilde{\psi}^B) \\ &\leq \frac{9\lambda^2}{\mu^2} 64\delta^2 = \frac{576\lambda^2\delta^2}{\mu^2}. \end{aligned}$$

Both maximizations are over states $|\psi\rangle \in S$ such that $\langle\varphi|\psi\rangle = 0$. The second inequality follows from the fact that $\hat{\varphi}^B \leq \tilde{\varphi}^B$ (and likewise for ψ) along with Lemma 18 while the third arises by substituting in the result of Eq. (11). Introducing one last decoration for our states, let $\bar{\varphi}^B = \hat{\varphi}^B / \operatorname{tr} \hat{\varphi}^B$ and likewise for σ. Applying Eq. (7) with attention paid to the fact that $\hat{\varphi}^B$ and $\hat{\sigma}^B$ are not normalized gives

$$\frac{1}{2}\big\|\bar{\varphi}^B - \bar{\sigma}^B\big\|_1 \geq 1 - \frac{24\lambda\delta}{\mu}\frac{1}{1 - 3\epsilon} \geq 1 - \frac{30\lambda\delta}{\mu},$$

where the final inequality uses that $\epsilon \leq 1/10$. Applying Helstrom's theorem to $\bar{\varphi}^B$ and $\bar{\sigma}^B$ implies that there exists a projector P_σ such that

$$\operatorname{tr} \bar{\varphi}^B P_\sigma - \operatorname{tr} \bar{\sigma}^B P_\sigma \geq 1 - \frac{30\lambda\delta}{\mu}.$$

Next we invoke Ky Fan's minimax theorem, just as in the proof of Theorem 7, for the payoff function $1 - \operatorname{tr} \bar{\varphi}^B P + \operatorname{tr} \hat{\sigma}^B P$, with the strategy space of the second player the convex hull of states $\hat{\psi}^B$, where $|\psi\rangle \in S$ ranges over states orthogonal to $|\varphi\rangle$. This provides an operator $0 \leq D_\varphi \leq \mathbb{1}$ such that

$$\operatorname{tr} \bar{\varphi}^B D_\varphi \geq 1 - \frac{30\lambda\delta}{\mu} \quad \text{and} \tag{12}$$

$$\operatorname{tr} \hat{\sigma}^B D_\varphi \leq \frac{30\lambda\delta}{\mu}. \tag{13}$$

But

$$\begin{aligned} \left| \operatorname{tr} \varphi^B D_\varphi - \operatorname{tr} \bar{\varphi}^B D_\varphi \right| &\leq \|\varphi^B - \bar{\varphi}^B\|_1 \\ &\leq \|\varphi^B - \hat{\varphi}^B\|_1 + \|\hat{\varphi}^B - \bar{\varphi}^B\|_1 \\ &\leq \sqrt{4 \cdot 3\epsilon} + \left|1 - (1 - 3\epsilon)\right| \\ &\leq 4\sqrt{\epsilon}, \end{aligned}$$

where the last line follows from the gentle measurement lemma (Appendix A, Lemma 20) and the fact that $\hat{\varphi}^B = \bar{\varphi}^B / (\operatorname{tr} \hat{\varphi}^B)$. Similarly, for any $\hat{\sigma}^B = \sum_i q_i \hat{\psi}_i^B$ a convex combination of states arising from $|\psi_i\rangle \in S$ perpendicular to $|\varphi\rangle$,

$$\begin{aligned} \left| \operatorname{tr} \sigma^B D_\varphi - \operatorname{tr} \hat{\sigma}^B D_\varphi \right| &\leq \|\sigma^B - \hat{\sigma}^B\|_1 \\ &\leq \sum_i q_i \|\psi_i^B - \hat{\psi}_i^B\|_1 \\ &\leq \sqrt{4 \cdot 3\epsilon} \leq 4\sqrt{\epsilon}. \end{aligned}$$

Combining these estimates with the outcome of the minimax theorem in Eq. (12) and Lemma 6 completes the proof. □

IV. QUANTUM IDENTIFICATION CAPACITY

While it might not be possible to design low error quantum-ID codes for any given channel, the situation becomes more promising if many uses of the channel are allowed. In analogy with classical and quantum data transmission, we can define asymptotic quantum-ID codes as follows.

Definition 9 (Quantum-ID capacity [12]) *A rate Q is said to be* achievable for quantum identification over $\mathcal{N}$ *if for all $\epsilon > 0$ and sufficiently large n, there are ϵ-quantum-ID codes for $\mathcal{N}^{\otimes n}$ with encoding domain S of dimension at least 2^{nQ}. The* quantum identification capacity $Q_{\rm ID}(\mathcal{N})$ *is defined as the supremum of the achievable rates.*

The capacity should be interpreted as the number of qubits that can be identified per use of the channel $\mathcal{N}$ in the limit of many uses of the channel. The only nontrivial channel for which the quantum identification capacity was known prior to this paper was the identity channel: asymptotically, a noiseless qubit channel can be used to identify two qubits. That is, $Q_{\rm ID}(\mathrm{id}_2) = 2$ [12]. As we will see below, the theory of the quantum identification capacity is considerably simpler when the given channel $\mathcal{N}$ can be used in conjunction with noiseless channels to the receiver. This obviously increases the capacity, so the interesting question is how much the use of $\mathcal{N}$ increases the quantum identification capacity over what would have been achievable with the noiseless channels alone. When defining the achievable amortized rates it is therefore necessary to subtract off two qubits for every noiseless qubit channel used per copy of $\mathcal{N}$.

Definition 10 (Amortized quantum-ID capacity) *A rate Q is said to be* achievable for amortized quantum identification over $\mathcal{N}$ *if for all $\epsilon > 0$ and sufficiently large n, there are ϵ-quantum-ID codes for $\mathrm{id}_C \otimes \mathcal{N}^{\otimes n}$ with encoding domain S such that $Q \leq \frac{1}{n}(\log_2 \dim S - 2\log_2 \dim C)$. The* amortized quantum identification capacity $Q_{\rm ID}^{\rm am}(\mathcal{N})$ *is defined as the supremum of the achievable rates.*

The fidelity alternative is a very powerful tool for studying the quantum-ID capacities. As a warm-up, the fact that the complements of quantum-ID codes are forgetful supplies a quick answer to an open question from [12]:

Theorem 11 *If $\mathcal{N}$ is an antidegradable channel, that is, if there exists channel $\mathcal{T}$ such that $\mathcal{N} = \mathcal{T} \circ \mathcal{N}^c$, then $Q_{\rm ID}(\mathcal{N}) = 0$. This is true in particular for the noiseless cbit channel $\overline{\mathrm{id}}_2$.*

Proof If $\mathcal{N}$ is antidegradable, then so is $\mathcal{N} \circ \mathcal{E}$ for any encoding map. Given a quantum-ID code for the channel $\mathcal{N}$ that encodes as little as one qubit, the channel $\mathcal{N} \circ \mathcal{E}$ will be geometry-preserving. But if $\mathcal{N}$ is antidegradable, then the channel $\mathcal{N} \circ \mathcal{E}$ will be also. Hence, by the fidelity alternative, the channel complementary to $\mathcal{N} \circ \mathcal{E}$ and, by antidegradability, $\mathcal{N} \circ \mathcal{E}$ itself would be approximately forgetful, contradicting the assumption. □

As usual, quantitative statements about asymptotically achievable rates and upper bounds on the identification capacities are naturally expressed in terms of entropies. For a bipartite density matrix φ^{AB}, we write

$$H(A)_\varphi \equiv H(\varphi^A) \equiv -\operatorname{tr} \varphi^A \log_2 \varphi^A$$

for the *von Neumann entropy* of φ^A. The mutual information of the state φ^{AB} is defined to be

$$I(A:B)_\varphi = H(A)_\varphi + H(B)_\varphi - H(AB)_\varphi$$

while the coherent information and the conditional entropy are, respectively,

$$\begin{aligned} I(A\rangle B)_\varphi &= H(B)_\varphi - H(AB)_\varphi \\ H(A|B)_\varphi &= H(AB)_\varphi - H(B)_\varphi. \end{aligned}$$

Our main theorem on the quantum identification capacities includes a concise formula for $Q_{\rm ID}^{\rm am}$ that eliminates the optimization over multiple channel uses.

Theorem 12 (Quantum identification capacity) *For any quantum channel $\mathcal{N}$, its quantum-ID capacity is given by $Q_{\rm ID}(\mathcal{N}) = \sup_n \frac{1}{n} Q_{\rm ID}^{(1)}(\mathcal{N}^{\otimes n})$, where*

$$Q_{\rm ID}^{(1)}(\mathcal{N}) = \sup_{|\varphi\rangle} \bigl\{ I(A:B)_\rho \text{ s.t. } I(A\rangle B)_\rho > 0 \bigr\},$$

where $|\varphi\rangle$ *is the purification of any input state to* $\mathcal{N}$ *and* $\rho^{AB} = (\mathrm{id} \otimes \mathcal{N})\varphi$, *and where we declare the* sup *to be* 0 *if the set above is empty.*

Furthermore, the amortized quantum-ID capacity equals

$$Q_{\mathrm{ID}}^{\mathrm{am}}(\mathcal{N}) = \sup_{|\varphi\rangle} I(A:B)_\rho = C_E(\mathcal{N}),$$

the entanglement-assisted classical capacity of $\mathcal{N}$ *[32].*

Remark It follows from Theorem 12 that the amortized quantum-ID capacity of a noiseless cbit channel is one. Reconciling this observation with Theorem 11, which asserts this channel's unamortized quantum-ID capacity is zero, reveals that *some* amortized noiseless quantum communication is necessary to achieve $Q_{\mathrm{ID}}^{\mathrm{am}}$ without determining how much. In fact, inspection of the proof of Theorem 12 reveals that, for the noiseless cbit channel $\overline{\mathrm{id}}_2$, a zero rate of noiseless side qubits is sufficient to achieve the maximum value of one. These observations extend to cq-channels, so named because they consist of a destructive measurement resulting in classical information, followed by the preparation of a state conditioned on the measurement outcome. For these channels, the entanglement-assisted capacity C_E is equal to the unassisted classical capacity C, also known as the Holevo capacity [24, 33]. As a result, $Q_{\mathrm{ID}}(\mathcal{N}) = 0$ for all such channels even as $Q_{\mathrm{ID}}^{\mathrm{am}}(\mathcal{N}) = C(\mathcal{N})$, the latter strictly positive for all nontrivial channels. The difference in all cases can be traced to a sublinear amount of free quantum communication in the amortized setting.

This effect can be viewed as an instance of *(un-)locking* since the quantum-ID rate increases from strictly 0 to an arbitrarily large amount by the addition of any positive rate of quantum communication, cf. [17, 34, 35]. Unlike the previously known examples where a certain finite rate is always required, however, here an arbitrarily small rate of extra quantum communication is sufficient to bring about an unbounded increase in the capacity. □

The intuition behind the achievability of the rates in Theorem 12 is quite simple. The structure of an amortized code is illustrated in Figure 1. Fix a state $|\varphi\rangle$ purifying any input to the channel $\mathcal{N}$ and let $|\rho\rangle^{ABE}$ be $(\mathbb{1} \otimes U_{\mathcal{N}})|\varphi\rangle$, where $U_{\mathcal{N}}$ is the Stinespring extension of $\mathcal{N}$. The encoding will embed the input into a random subspace of a typical subspace of A^n, producing states highly entangled between B^nC and E^nF. By arranging for E^nF to be slightly smaller than B^nC in the appropriate sense, one ensures that the states are indistinguishable on the environment. By the fidelity alternative, they can therefore be identified by Bob. Letting $R = \frac{1}{n}\log\dim C$ and $f = \frac{1}{n}\log\dim F$, the condition for forgetfulness to the environment is roughly

$$H(B)_\rho + R > H(E)_\rho + f,$$

so $f - R$ is chosen to be very slightly less than $H(B)_\rho - H(E)_\rho$. Moreover, measure concentration for the choice of random subspace will make it possible to choose the coding subspace almost as large as the ambient space, which in qubit terms has effective size

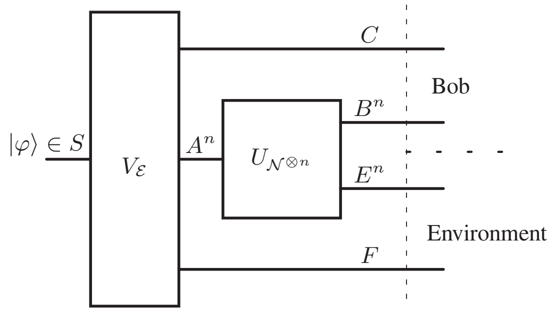

FIG. 1: Structure of a quantum identification code. $U_{\mathcal{N}^{\otimes n}}$ and $V_{\mathcal{E}}$ are the Stinespring extensions of the noisy channel $\mathcal{N}^{\otimes n}$ and the encoding operation $\mathcal{E}$. The receiver, Bob, has access to the channel output B^n as well as C, which consists of nR qubits transmitted noiselessly from the receiver. (In the non-amortized setting, there is no C.) The encoding map $\mathcal{E}$ is generally noisy, so part of its output is transmitted to the environment.

$$nH(A)_\rho + nR + nf.$$

The rate of the amortized code will therefore be

$$\begin{aligned} H(A)_\rho + R + f - 2R &= H(A)_\rho + f - R \\ &\approx H(A)_\rho + H(B)_\rho - H(E)_\rho. \end{aligned}$$

Since ρ is pure, $H(E)_\rho = H(AB)_\rho$ and the rate is precisely the mutual information.

The detailed proof of the achievability of the rates in Theorem 12 builds on the techniques developed in Refs. [36] and [37] analyzing the properties of generic quantum states. The proof will combine the following theorem, originally motivated by the foundations of statistical mechanics, with the duality between quantum identification and approximate forgetfulness formulated in Theorem 7 or, more precisely, its technical variant Theorem 8.

Theorem 13 (Random versus average states [37]) *Let* S *be a subspace of* $B \otimes E$, Ω *be the maximally mixed state on* S, *and* X *any operator acting on* $B \otimes E$ *with* $\|X\|_\infty \leq 1$. *If* $|\varphi\rangle \in S$ *is chosen according to the unitarily invariant measure, then for all* $\epsilon > 0$

$$\Pr\left\{ \left\| \mathrm{tr}_B X\Omega X^\dagger - \mathrm{tr}_B X\varphi X^\dagger \right\|_1 \geq \eta \right\} \leq \eta'$$

where

$$\begin{aligned} \eta &= \epsilon + \sqrt{\tilde{d}_E/\tilde{d}_B} \quad \text{and} \\ \eta' &= 2\exp(-C\epsilon^2 \dim S). \end{aligned}$$

Here $C > 0$ *is a constant,* $\tilde{d}_E = \dim\operatorname{supp}\mathrm{tr}_B XX^\dagger$ *is an upper bound on the dimension of the support of* $\mathrm{tr}_B X\Omega X^\dagger$ *and* $\tilde{d}_B = 1/\operatorname{tr}[(\mathrm{tr}_E X\Omega X^\dagger)^2]$ *can be thought of as the effective dimension of* B.

Proof This is a slight modification of [37, Thm. 2]. In the original, the theorem bounds $\| \operatorname{tr}_B \Omega - \operatorname{tr}_B \varphi\|_1$ under similar hypotheses but η includes a correction dependent on $\operatorname{tr} X\Omega X^\dagger$. The correction disappears if the argument is applied to $\| \operatorname{tr}_B X\Omega X^\dagger - \operatorname{tr}_B X\varphi X^\dagger\|_1$ instead under the assumption that $\|X\|_\infty \leq 1$, which ensures that the map $\rho \mapsto X\rho X^\dagger$ is 1-Lipschitz. $\square$

In order to use Theorem 13 to make statements about random subspaces, we will use the following lemma

Lemma 14 *Let f be an real-valued function on $\mathbb{C}P^d$ (identified with rank one projectors acting on $\mathbb{C}^d$) and suppose that f is α-Lipschitz with respect to the trace norm. Let μ be the unitarily invariant measure on $\mathbb{C}P^d$ and $\hat{\mu}$ the unitarily invariant measure on the space of k-dimensional subspaces of $\mathbb{C}^d$. If*

$$\mu\left\{|\xi\rangle; f(\xi) > \eta\right\} \leq g(d)$$

then

$$\hat{\mu}\left\{S; \max_{|\xi\rangle\in S, \langle\xi|\xi\rangle=1} f(\xi) > (1+\alpha)\eta\right\} \leq \left(\frac{5}{\eta}\right)^{2k} g(d).$$

Proof This is a standard discretization argument. Fix a k-dimensional subspace $S_0 \subseteq \mathbb{C}^d$. According to Ref. [8], there is a trace norm η-net M for the rank one projectors on S_0 of cardinality no more than $(5/\eta)^{2k}$. If U is distributed according to the Haar measure ν, then US_0 is distributed according to the unitarily invariant measure. So, we find by the triangle inequality that

$$\begin{aligned}
&\hat{\mu}\left\{S; \max_{|\xi\rangle\in S, \langle\xi|\xi\rangle=1} f(\xi) > (1+\alpha)\eta\right\} \\
&= \nu\left\{U; \max_{|\xi\rangle\in S_0, \langle\xi|\xi\rangle=1} f(U\xi U^\dagger) > (1+\alpha)\eta\right\} \\
&\leq \nu\left\{U; \max_{|\xi\rangle\in M, \langle\xi|\xi\rangle=1} f(U\xi U^\dagger) > \eta\right\} \\
&\leq \left(\frac{5}{\eta}\right)^{2k} \mu\{|\xi\rangle; f(\xi) > \eta\},
\end{aligned}$$

where the second inequality is just the union bound over elements of the net. $\square$

The following theorem collects the facts we will need about type and typical projectors. We omit their definitions, which will not be required here and can be found in Ref. [38].

Theorem 15 (Typicality) *Let $|\rho\rangle \in A\otimes B\otimes E$ and set $|\psi\rangle = |\rho\rangle^{\otimes n}$. For any $\delta, \epsilon > 0$ sufficiently small there exist projectors Π^B, Π_1^E and Π_2^E on $B^{\otimes n}$ and $E^{\otimes n}$, respectively, and a projection Π_t^A onto a fixed type subspace of A^n such that the states*

$$|\psi_t\rangle = \frac{\Pi_t^A \otimes \mathbb{1}^B \otimes \mathbb{1}^E |\psi\rangle}{\sqrt{\langle\psi|\Pi_t^A \otimes \mathbb{1}^B \otimes \mathbb{1}^E|\psi\rangle}} \quad \textit{and}$$

$$|\tilde{\psi}_t\rangle = \frac{\Pi_t^A \otimes \Pi^B \otimes \Pi_2^E\Pi_1^E |\psi\rangle}{\sqrt{\langle\psi|\Pi_t^A \otimes \mathbb{1}^B \otimes \mathbb{1}^E|\psi\rangle}}$$

satisfy the following conditions for $X = A^n, B^n, E^n$ and sufficiently large n:

1. *$\psi_t^{A^n} = \Pi_t^A/\operatorname{rank}\Pi_t^A$.*
2. *$\|\psi_t - \tilde{\psi}_t\|_1 \leq \epsilon$.*
3. *$\operatorname{tr}[(\tilde{\psi}_t^X)^2] \leq 3(1-3\epsilon)^{-1}2^{-n[H(X)_\rho - c\delta]}$.*
4. *$2^{n[H(X)_\rho-\delta]} \leq \operatorname{rank}\Pi^X \leq 2^{n[H(X)_\rho+\delta]}$.*
5. *The largest eigenvalue of $\tilde{\psi}_t^{E^n}$ is bounded above by $(1-3\epsilon)^{-1}2^{-n[H(E)_\rho-c\delta]}$.*
6. *The ratio of the largest to the smallest nonzero eigenvalue of $\tilde{\psi}_t^{E^n}$ is at most $2^{2n\delta}$.*

where Π^A and Π^E should respectively be understood to be Π_t^A and $\Pi_2^E\Pi_1^E$ in property 4, and $c > 0$ is a constant.

Proof If Π_2^E is removed and property 6 omitted, then the theorem is precisely a result proved in Ref. [38], with Π_1^E the typical projector for ρ on E^n. Π_2^E will be a projector that removes all eigenvalues of the reduced density operator on E^n below the stated threshold. Let

$$|\xi\rangle = \frac{\Pi_t^A \otimes \Pi^B \otimes \Pi_1^E|\psi\rangle}{\sqrt{\langle\psi|\Pi_t^A\otimes \mathbb{1}^B \otimes \mathbb{1}^E|\psi\rangle}}$$

The largest eigenvalue of ξ^{E^n} is bounded above by $(1-3\epsilon)^{-1}2^{-n[H(E)_\rho-c\delta]}$ according to property 5 as stated above and the state's rank is at most $2^{n[H(E)_\rho+\delta]}$ by property 4. Applying Lemma 21 to the eigenvalues of ξ^{E^n} reveals that the sum of all eigenvalues less than or equal to $2^{-2n\delta}/\operatorname{rank}\xi^{E^n}$ is at most

$$\frac{2^{-2n\delta}}{1-3\epsilon} \leq 2^{-n\delta}$$

for sufficiently large n. We can therefore let Π_2^E be the orthogonal projection onto the direct sum of the eigenspaces of ξ^{E^n} corresponding to eigenvalues larger than $2^{-2n\delta}/\operatorname{rank}\xi^{E^n}$. Let λ be the largest eigenvalue of ξ^{E^n}. The ratio of the largest to the smallest eigenvalue after the application of Π_2^E will be at most

$$\frac{\lambda}{2^{-2n\delta}/\operatorname{rank}\tilde{\psi}_t^{E^n}} \leq \frac{\lambda}{2^{-2n\delta}\cdot\lambda} = 2^{2n\delta}.$$

A redefinition of ϵ completes the proof. $\square$

Proof (Direct coding part of Theorem 12) The regular and amortized cases can be handled simultaneously. Fix an input state φ as in Theorem 12, let $|\rho\rangle^{ABE}$ be a purification of $(\mathrm{id}\otimes\mathcal{N})\varphi$ and let $|\psi\rangle = |\rho\rangle^{\otimes n}$. To construct the code, we will need to project ψ^{A^n} to a type subspace having favorable properties. $\psi_t^{A^nB^n}$ is the Choi-Jamiolkowski state for the channel $\mathcal{N}^{\otimes n}$ restricted to the type subspace A_t defined by the projector Π_t^A. Call this channel $\mathcal{N}_t$, write U_t for its Stinespring dilation, and consider $\mathcal{N}_t \otimes \mathrm{id}^C \otimes \mathrm{id}^F$. C will play the role of the noiseless channel from Alice to Bob in the case of the amortized capacity and F will represent quantum information discarded by Alice at the encoding stage. Our code will consist of a subspace of S' of $A_t \otimes C \otimes F$ selected according to the unitarily invariant measure, which then defines a subspace S of $(B^n \otimes C) \otimes (E^n \otimes F)$. Our aim will be to show that S is likely to be approximately forgetful for $E^n \otimes F$ when C and F are chosen appropriately, allowing for an application of Theorem 8.

Let $\Omega = \psi_t^{B^n E^n} \otimes \pi^C \otimes \pi^F$ be the image under $U_t \otimes \mathbb{1}^C \otimes \mathbb{1}^F$ of the maximally mixed state on $A_t \otimes C \otimes F$. (Recall that π^Z denotes the maximally mixed state on Z.) Define $|\tilde{\psi}_t\rangle$ as in Theorem 15 and let $\tilde{\Omega} = \tilde{\psi}_t^{B^n E^n} \otimes \pi^C \otimes \pi^F$. Then

$$\tilde{\psi}_t^{B^n E^n} = (\Pi^B \otimes \Pi_2 \Pi_1^E) \psi_t^{E^n B^n} (\Pi^B \otimes \Pi_1^E \Pi_2^E)$$

so for $X = \Pi^B \otimes \Pi_2^E \Pi_1^E$, Theorem 13 states that a randomly chosen state $|\omega\rangle$ in $U_t(A_t) \otimes C \otimes F$ will satisfy

$$\Pr\left[\left\|\tilde{\Omega}^{E^n F} - \tilde{\omega}^{E^n F}\right\|_1 \geq \frac{\eta}{2}\right] \leq \eta'$$

for $\tilde{\omega} = X \omega X^\dagger$ and where, for any $\nu > 0$,

$$\begin{aligned}\frac{\eta}{2} &= \nu + \sqrt{\operatorname{rank}[\Pi_2^E \Pi_1^E \otimes \mathbb{1}^F] \cdot \operatorname{tr}[(\tilde{\psi}_t^{B^n} \otimes \pi^C)^2]},\\ \eta' &= 2 \exp\big(-C\nu^2 \dim(A_t \otimes C \otimes F)\big).\end{aligned}$$

We will fix ν to be $\nu = 2^{-3n\delta}$. So by Lemma 14, a random S in $U_t(A_t) \otimes C \otimes F$ chosen according to the unitarily invariant measure will satisfy

$$\begin{aligned}&\Pr_S\left[\max_{|\omega\rangle \in S} \left\|\tilde{\Omega}^{E^n F} - \tilde{\omega}^{E^n F}\right\|_1 \geq \eta\right]\\ &\qquad \leq 2\left(\frac{10}{\eta}\right)^{2|S|} \exp\big(-C\nu^2 \dim(A_t \otimes C \otimes F)\big)\end{aligned}$$

since the function $\omega \mapsto \|\tilde{\Omega}^{E^n F} - \tilde{\omega}^{E^n F}\|_1$ is 1-Lipschitz with respect to the trace norm. For convenience, let $\dim F = 2^{nf}$ and $\dim C = 2^{nR}$. Since $\dim A_t \geq 2^{n[H(A)_\rho - \delta]}$, choosing $\dim S$ to be $2^{n[H(A)_\rho + R + f - 8\delta]}$ will lead to

$$\max_{|\omega\rangle \in S} \left\|\tilde{\Omega}^{E^n F} - \tilde{\omega}^{E^n F}\right\|_1 < \eta \tag{14}$$

with high probability for sufficiently large n provided η decays at most exponentially with n.

Now let us determine how to choose f and R in order to ensure a small value for η. Observe that by properties 3 and 4 in Theorem 15,

$$\begin{aligned}&\operatorname{rank} \Pi_2^E \Pi_1^E \Pi_1^E \otimes \mathbb{1}^F \leq 2^{n[H(E)_\rho + \delta + f]} \quad \text{and}\\ &\operatorname{tr}[(\tilde{\psi}_t^{B^n} \otimes \tfrac{1}{|C|}\mathbb{1}^C)^2] \leq 3(1-3\epsilon)^{-1} \cdot 2^{-n[H(B)_\rho - c\delta - R]}.\end{aligned}$$

Therefore,

$$\eta \leq \nu + 3 \cdot 2^{n[H(E)_\rho - H(B)_\rho + f - R + (1+c)\delta]/2}$$

provided ϵ is chosen smaller than $1/10$. There are two cases to consider:

Case 1. First suppose that $I(A\rangle B)_\rho > 0$ or, equivalently, that $H(E)_\rho < H(B)_\rho$. Under these circumstances, amortization is not required. Choosing $R = 0$ and $f = H(B)_\rho - H(E)_\rho - (7+c)\delta$ leads to $\eta \leq \nu + 3 \cdot 2^{-3n\delta} \leq 4 \cdot 2^{-3n\delta}$. The rate of the associated code will be

$$\begin{aligned}Q &= \frac{1}{n}\log_2 \dim S\\ &= H(A)_\rho + R + f - 8\delta\\ &= H(A)_\rho + H(B)_\rho - H(E)_\rho - (7+c)\delta - 8\delta\\ &= I(A:B)_\rho - (15+c)\delta.\end{aligned}$$

Case 2. Now suppose that $I(A\rangle B)_\rho \leq 0$ so that $H(E)_\rho \geq H(B)_\rho$. In this case we set $R = H(E)_\rho - H(B)_\rho + (7+c)\delta$ and $f = 0$ to again achieve $\eta \leq 4 \cdot 2^{-3n\delta}$. This time, the rate of the code will be

$$\begin{aligned}Q &= \frac{1}{n}\log_2 \dim S - 2R\\ &= H(A)_\rho + R + f - 8\delta - 2R\\ &= H(A)_\rho + H(B)_\rho - H(E)_\rho - (15+c)\delta\\ &= I(A:B)_\rho - (15+c)\delta.\end{aligned}$$

We have established that the subspace S corresponds to a code of the correct rate. Applying Theorem 8 to $\tilde{\Omega}$ and the states in S with $X = \Pi^B \otimes \mathbb{1}^C \otimes \Pi_2^E \Pi_1^E \otimes \mathbb{1}^F$ will complete the proof. Recalling that the ratio of the largest to the smallest nonzero eigenvalues of $\tilde{\Omega}^{E^n F}$ is at most $2^{2n\delta}$, the theorem asserts that S is a quantum-ID code with error probability at most

$$3\left(30 \cdot 2^{2n\delta} \cdot (4 \cdot 2^{-3n\delta}) + 4\sqrt{\epsilon}\right)^{1/2},$$

which can be made arbitrarily small for sufficiently large n. □

Proof (Converse for Theorem 12) We will address both the regular and amortized capacities at the same time. Consider an amortized quantum-ID code for n copies of $\mathcal{N}$ as illustrated in Figure 1. The Stinespring dilations of $\mathcal{N}^{\otimes n}$ and $\mathcal{E}$ together have three output registers: one for the channel input, one for the transmission to Bob and one going to the environment. Calling $\widehat{B} := B^n C$ and $\widehat{E} = E^n F$ in the figure, the quantum-ID code is equivalent to a subspace $S \subseteq \widehat{B} \otimes \widehat{E}$, and we can apply our lemmas.

A key observation is that for any orthogonal pure state ensemble $\{p_x, \varphi_x\}$ on S decomposing the maximally mixed state,

$$H(\widehat{B}) \geq H(\widehat{B}|X) = H(\widehat{E}|X) = H(\widehat{E}) - o(n). \tag{15}$$

The first inequality is just the concavity of the entropy function while the first equality follows from the fact that φ_x is pure on $\widehat{B}\widehat{E}$. The final relation is a consequence of Theorem 7: the fidelity alternative implies that if states can be identified on $\widehat{B}$ then they must be indistinguishable on $\widehat{E}$. Continuity of the von Neumann entropy in the form of the Fannes inequality [39] shows the correction to be $o(n)$. Thus, sending one half of a maximally entangled state between A and an auxiliary space also named A into the above circuit, we obtain a multipartite pure state with respect to which

$$\begin{aligned}\log_2 |A| = H(A) &\leq H(A) + H(\widehat{B}) - H(\widehat{E}) + o(n) \\ &= I(A : \widehat{B}) + o(n) \\ &= I(A : B^n) + I(A : C|B^n) + o(n) \\ &\leq I(A : B^n) + 2\log_2 \dim C + o(n).\end{aligned}$$

Therefore, the amortized quantum identification capacity is bounded above by $\lim_{n\to\infty} \frac{1}{n} g(\mathcal{N}^{\otimes n})$ where $g(\mathcal{N}) = \max_{|\varphi\rangle} I(A : B)_\rho$ for $\rho = (\mathrm{id} \otimes \mathcal{N})\varphi$. It is well-known, however, that $g(\mathcal{N}^{\otimes n}) = n g(\mathcal{N})$ so the limit is not necessary [32].

On the other hand, in the non-amortized case, $\dim C = 1$, and the above Eq. (15) yields

$$I(A\rangle B^n) = I(A\rangle \widehat{B}) = H(\widehat{B}) - H(\widehat{E}) \geq -o(n). \qquad (16)$$

To obtain the claimed formula, we need to make two observations. First, if $Q_{\rm ID}(\mathcal{N}) > 0$, then also $Q(\mathcal{N}) > 0$; this is obtained by restricting to a two-dimensional subspace of S for large n (and small ϵ), improving Eqs. (15) and (16) to

$$\begin{aligned}H(\widehat{B}) &\geq 1 - o(1) + H(\widehat{B}|X) \\ &= 1 - o(1) + H(\widehat{E}|X) = 1 - o(1) + H(\widehat{E}),\end{aligned}$$

showing that the coherent information is indeed positive, hence implying positive quantum capacity. Second, in that case the normalized mutual information $\frac{1}{n} I(A : B^n)$ is upper bounded by $\frac{1}{n+o(n)} I(AA' : B^n B'^{o(n)}) + o(1)$, where the $A'B'^{o(n)}$ system is chosen as above with positive coherent information rate, such that $I(AA'\rangle B^n B'^{o(n)}) > 0$. This shows that $\sup_n \frac{1}{n} Q_{\rm ID}^{(1)}(\mathcal{N}^{\otimes n})$ is indeed an upper bound on all achievable rates. □

V. CONCLUSION AND OPEN QUESTIONS

The fidelity alternative states that geometry preservation and approximate forgetfulness are complementary properties, much like quantum data transmission and complete forgetfulness. Subject to some technical conditions, geometry preservation is itself equivalent to quantum identification, an operational task very much in the spirit of quantum data transmission but strictly weaker. Just as analyzing complete forgetfulness has proved a versatile and effective tool for studying asymptotic quantum error correction, approximate forgetfulness provides a new approach to asymptotic quantum identification. Indeed, by focusing on approximate forgetfulness of the complementary channel, we have established that the amortized quantum identification capacity is exactly equal to the entanglement-assisted capacity.

The fidelity alternative suggests a number of possible extensions, such as asking what happens if geometry is preserved not only for pure states but for higher rank mixed states. Would such a property have an operational interpretation and corresponding interpretation in terms of a form of forgetfulness intermediate between the weak form studied here and complete forgetfulness? It would also be interesting to understand geometry preservation as a type of pseudo-isometry [40] from projective space to the Grassmannian of subspaces corresponding to the supports of the mixed output states.

Meanwhile, Theorem 12 poses an entertaining and potentially deep puzzle: why do amortized quantum identification and entanglement-assisted classical communication result in the same capacity in the absence of any known operational relationship between these tasks? The theorem also leaves open the important problem of evaluating the quantum identification capacity formula in the unamortized case. Similarly, the theorem fails to determine how much extra quantum communication is necessary to achieve the amortized capacity. In particular, does there exist a channel where the required rate is strictly positive? We do expect this to be true, but have been unable to establish it rigorously.

Acknowledgments

We thank Marco Piani for sharing his proof of Lemma 19 with us and Mark Wilde for helpful comments. PH was supported by the Canada Research Chairs program, the Perimeter Institute, CIFAR, FQRNT's INTRIQ, MITACS, NSERC, ONR through grant N000140811249 and QuantumWorks. AW was supported through an Advanced Research Fellowship of the U.K. EPSRC, the EPSRC's "QIP IRC", the European Commission IP "QAP", by a Wolfson Research Merit Award of the Royal Society and a Philip Leverhulme Prize. The Centre for Quantum Technologies is funded by the Singapore Ministry of Education and the National Research Foundation as part of the Research Centres of Excellence programme.

Appendix A: Miscellaneous Facts

The following results were used in various proofs but have been collected here so as not to distract from the main line of argument in the paper. This first pleasing little relation provides a convenient way to calculate mixed state fidelity:

Lemma 16 *For pure states φ, ψ on a bipartite system $B \otimes E$,*

$$F(\varphi^B, \psi^B) = \big\| \mathrm{tr}_B\, |\varphi\rangle\langle\psi| \big\|_1^2. \qquad \text{(A1)}$$

Proof This is a straightforward calculation:

$$\begin{aligned}\big\|\mathrm{tr}_B\, |\varphi\rangle\langle\psi|\big\|_1 &= \max_{\|X\|_\infty \leq 1} \left|\mathrm{tr}\left(\mathrm{tr}_B\, |\varphi\rangle\langle\psi|\right) X\right| \\ &= \max_{U \text{ unitary}} \left|\mathrm{tr}\left(\mathrm{tr}_B\, |\varphi\rangle\langle\psi|\right) U\right| \\ &= \max_{U \text{ unitary}} \left|\mathrm{tr}\, |\varphi\rangle\langle\psi| (\mathbb{1} \otimes U)\right| \\ &= \max_{U \text{ unitary}} \sqrt{F\big((\mathbb{1} \otimes U)\varphi(\mathbb{1} \otimes U^\dagger), \psi\big)} \\ &= \sqrt{F(\varphi^B, \psi^B)},\end{aligned}$$

invoking, successively, the duality between trace and sup norm, the fact that the maximum is always attained at a unitary, the defining property of the partial trace, and in the last line Uhlmann's relation [41, 42]. ☐

The following lemma provides conditions under which mixing preserves near-orthogonality.

Lemma 17 *Let ρ and σ_i, for all i, be states on the same Hilbert space such that there exist projectors P and Q_i of rank $\leq r$, and $\mu P \leq \rho \leq \lambda P$, $\mu Q_i \leq \sigma_i \leq \lambda Q_i$ such that $\mu r \leq 1$. If furthermore for all i, $F(\rho, \sigma_i) \leq \epsilon$, then*

$$F\big(\rho, \overline{\sigma}\big) \leq \delta := \epsilon \frac{\lambda^2}{\mu^2}$$

for every $\overline{\sigma} = \sum_i p_i \sigma_i$ in the convex hull of the σ_i.

Proof We use the definition of the fidelity to first obtain

$$\epsilon \geq \left(\operatorname{tr} \sqrt{\sqrt{\rho} \sigma_i \sqrt{\rho}} \right)^2 \geq \mu^2 \left(\operatorname{tr} P Q_i P \right)^2 .$$

Invoking the definition again, we now get from this

$$\begin{aligned} \sqrt{F\big(\rho, \overline{\sigma}\big)} = \left\| \sqrt{\rho} \sqrt{\overline{\sigma}} \right\|_1 &\leq \lambda \operatorname{tr} \sqrt{\sum_i p_i P Q_i P} \\ &\leq \lambda r \sqrt{\sum_i p_i \frac{1}{\mu r} \mu \operatorname{tr} P Q_i P} \\ &\leq \lambda r \sqrt{\frac{\epsilon}{\mu r}} \leq \sqrt{\epsilon} \frac{\lambda}{\mu}, \end{aligned}$$

using the concavity of the square root twice in turn [43]. ☐

Lemma 18 *Let $0 \leq \tilde{\rho} \leq \rho$ and $0 \leq \tilde{\sigma} \leq \sigma$. Then $F(\tilde{\rho}, \tilde{\sigma}) \leq F(\rho, \sigma)$.*

Proof Denoting unitary congruence of matrices (in particular having the same spectrum) by $\sim$, we have

$$\sqrt{\tilde{\rho}} \tilde{\sigma} \sqrt{\tilde{\rho}} \leq \sqrt{\tilde{\rho}} \sigma \sqrt{\tilde{\rho}} \sim \sqrt{\sigma} \tilde{\rho} \sqrt{\sigma} \leq \sqrt{\sigma} \rho \sqrt{\sigma} \sim \sqrt{\rho} \sigma \sqrt{\rho}.$$

Hence, since the square root is operator monotone [43] and the trace is invariant under unitary basis change, $\operatorname{tr} \sqrt{\sqrt{\tilde{\rho}} \tilde{\sigma} \sqrt{\tilde{\rho}}} \leq \operatorname{tr} \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}}$, completing the proof. ☐

The next lemma constrains the increase of the maximal output trace norm when tensoring with a fixed-size identity transformation:

Lemma 19 *Let $\Gamma : \mathcal{S}(A) \rightarrow \mathcal{S}(B)$ be a linear superoperator. Then for any t any positive integer,*

$$\|\Gamma\|_\diamond^{(t)} \leq t \, \|\Gamma\|_\diamond^{(1)} .$$

Proof Write X, an operator on $\mathbb{C}^t \otimes A$ such that $\|X\|_1 \leq 1$, in its singular value decomposition as $\sum_j s_j |v_j\rangle\langle w_j|$, with $0 \leq s_j \leq 1$ and $\langle v_j | v_k \rangle = \langle w_j | w_k \rangle = \delta_{jk}$. By convexity (triangle inequality), $\|\Gamma\|_\diamond^{(t)}$ is attained with a rank-one $X = |v\rangle\langle w|$, and for the following fix Schmidt decompositions $|v\rangle = \sum_k \alpha_k |e_k\rangle |f_k\rangle$ and $|w\rangle = \sum_\ell \beta_\ell |g_\ell\rangle |h_\ell\rangle$. Then,

$$\begin{aligned} \|(\mathrm{id}_t \otimes \Gamma) X\|_1 &= \big\|(\mathrm{id}_t \otimes \Gamma) |v\rangle\langle w|\big\|_1 \\ &= \left\| (\mathrm{id}_t \otimes \Gamma) \left(\sum_{k\ell} \alpha_k \beta_\ell |e_k\rangle\langle g_\ell| \otimes |f_k\rangle\langle h_\ell| \right) \right\|_1 \\ &\leq \sum_{k\ell} \alpha_k \beta_\ell \big\| (\mathrm{id}_t \otimes \Gamma) \left(|e_k\rangle\langle g_\ell| \otimes |f_k\rangle\langle h_\ell| \right) \big\|_1 \\ &= \sum_{k\ell} \alpha_k \beta_\ell \left\| \Gamma \left(|f_k\rangle\langle h_\ell| \right) \right\|_1 \leq t \|\Gamma\|_1^{(1)}. \end{aligned}$$

where the first step is just the triangle inequality and the next follows from the fact that $\|X\|_1 = \sum_j s_j \leq 1$. The final inequality uses the fact that $\sum_{k=1}^t \alpha_{jk}$ and $\sum_{l=1}^t \beta_{jl}$ are both bounded above by $\sqrt{t}$ since $\|\alpha_j\|_2 = \|\beta_j\|_2 = 1$. ☐

Remark The factor t is optimal, as the example of the matrix transposition shows where the bound of the lemma becomes an equality. ☐

Lemma 20 (Gentle measurement [44, 45]) *Let ρ be a state, and $0 \leq X \leq \mathbb{1}$ be an operator on some Hilbert space, such that $\operatorname{tr} \rho X \geq 1 - \epsilon$. Then, $\big\| \rho - \sqrt{X} \rho \sqrt{X} \big\|_1 \leq 2\sqrt{\epsilon}$.* ☐

The following, final, lemma is used to argue that the small eigenvalues of a density operator can be discarded without causing much disturbance.

Lemma 21 *Let $(p_1, p_2, \ldots, p_r)$ be a probability density with $p_i \geq p_{i+1}$ for all i and let $\chi = \{i; p_i \leq D/r\}$ for some $0 \leq D \leq 1$. Then, $\sum_{i \in \chi} p_i \leq D$.*

Proof Since evidently $|\chi| \leq r$,

$$\sum_{i \in \chi} p_i \leq |\chi| \frac{D}{r} \leq r \frac{D}{r} = D,$$

and that's it. ☐

[1] C. H. Bennett and P. W. Shor. Quantum information theory. *IEEE Trans. Inf. Theory*, 44(6):2724–2742, 1998.

[2] M. Lewenstein, B. Kraus, J. I. Cirac, and P. Horodecki. Optimization of entanglement witnesses. *Phys. Rev. A*, 62:052310, 2000.

[3] W. F. Stinespring. Positive functions on C^*-algebras. *Proc. Amer. Math. Soc.*, 6:211–216, 1955.

[4] M. A. Nielsen, C. M. Caves, B. Schumacher, and H. Barnum. Information-theoretic approach to quantum error correction and reversible measurement. *Proc. R. Soc. A*, 454:277–+, 1998.

arXiv:quant-ph/9706064.

[5] B. Schumacher and M. D. Westmoreland. Approximate quantum error correction. *Quantum Inf. Proc.*, 1(1-2):5–12, 2002. arXiv:quant-ph/0112106.

[6] I. Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inf. Theory*, 51(1):44–55, 2005. arXiv:quant-ph/0304127. See also Open Sys. Inf. Dyn. 15(1), special issue on quantum capacity (2008).

[7] D. Kretschmann, D. Schlingemann, and R. F. Werner. The Information-Disturbance Tradeoff and the Continuity of Stinespring's Representation. *IEEE Trans. Inf. Theory*, 54(4):1708–1717, 2008. arXiv:quant-ph/0605009.

[8] P. Hayden, D. Leung, P. W. Shor, and A. Winter. Randomizing quantum states: Constructions and applications. *Comm. Math. Phys.*, 250:371–391, 2004. arXiv:quant-ph/0307104.

[9] C. H. Bennett, P. Hayden, D. Leung, P. W. Shor, and A. Winter. Remote preparation of quantum states. *IEEE Trans. Inf. Theory*, 51(1):56–74, 2005. arXiv:quant-ph/0307100.

[10] V. Paulsen. *Completely Bounded Maps and Operator Algebras*, volume 78 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2003.

[11] A. Kitaev. Quantum computations: algorithms and error correction. *Russian Math. Surveys*, 52:1191–1249, 1997.

[12] A. Winter. Quantum and classical message identification via quantum channels. In O. Hirota, editor, *Festschrift "A. S. Holevo 60"*, pages 171–188. Rinton Press, 2004. Reprinted in Quantum Inf. Comput. 4(6&7):563-578 (2004); arXiv:quant-ph/0401060.

[13] A. Winter. Identification via quantum channels in the presence of prior correlation and feedback. In R. Ahlswede *et al.*, editor, *Information Transfer and Combinatorics*, volume 4123 of *Lecture Notes in Computer Science*, pages 486–504. Springer, 2006. arXiv:quant-ph/0403203.

[14] R. Ahlswede and G. Dueck. Identification via channels. *IEEE Trans. Inf. Theory*, 35(1):15–29, 1989.

[15] R. Ahlswede and G. Dueck. Identification in the presence of feedback – a discovery of new capacity formulas. *IEEE Trans. Inf. Theory*, 35(1):30–36, 1989.

[16] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Trans. Inf. Theory*, 48(3):569–579, 2002. arXiv:quant-ph/0012127.

[17] M. Christandl and A. Winter. Uncertainty, Monogamy, and Locking of Quantum Correlations. *IEEE Trans. Inf. Theory*, 51(9):3159–3165, 2005. arXiv:quant-ph/0501090.

[18] P. Hayden, P. W. Shor, and A. Winter. Random quantum codes from Gaussian ensembles and an uncertainty relation. *Open Sys. Inf. Dyn.*, 15(1):71–89, 2008. arXiv:0712.0975.

[19] J. M. Renes. Approximate quantum error correction via complementary observables. arXiv:1003.1150, 2010.

[20] P. Hayden, M. Horodecki, J. T. Yard, and A. Winter. A decoupling approach to the quantum capacity. *Open Sys. Inf. Dyn.*, 15:7–19, 2008. arXiv:quant-ph/0702005.

[21] R. Klesse. A random-coding based proof for the quantum coding theorem. *Open Sys. Inf. Dyn.*, 15(1):21–45, 2008. arXiv:0712.2558.

[22] M. Horodecki, J. Oppenheim, and A. Winter. Quantum State Merging and Negative Information. *Comm. Math. Phys.*, 269:107–136, 2007. arXiv:quant-ph/0512247.

[23] J. M. Renes. Duality of privacy amplification against quantum adversaries and data compression with quantum side information. arXiv:1003.0703, 2010.

[24] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Theory*, 44:269–273, 1998.

[25] C. A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum mechanical states. *IEEE Trans. Inf. Theory*, 45:1216–1227, 1999.

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

[27] D. Gottesman. *Stabilizer codes and quantum error correction*. PhD thesis, Caltech, 1997. arXiv:quant-ph/9705052.

[28] A. Abeyesinghe, P. Hayden, G. Smith, and A. Winter. Optimal superdense coding of entangled states. *IEEE Trans. Inf. Theory*, 52(8):3635–3641, 2006. arXiv:quant-ph/0407061.

[29] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf. Quantum fingerprinting. *Phys. Rev. Lett.*, 87(16):167902, 2001. arXiv:quant-ph/0102001.

[30] C. W. Helstrom. Quantum detection and estimation theory. *J. Stat. Phys.*, 1(2):231–252, 1969.

[31] Ky Fan. Minimax theorems. *Proc. Nat. Acad. Sci. USA*, 39:42–47, 1953.

[32] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Trans. Inf. Theory*, 48(10):2637–2655, 2002. arXiv:quant-ph/0106052.

[33] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56:131–138, 1997.

[34] D. P. DiVincenzo, M. Horodecki, D. W. Leung, J. A. Smolin, and B. M. Terhal. Locking classical correlations in quantum states. *Phys. Rev. Lett.*, 92:067902, 2004. arXiv:quant-ph/0303088.

[35] K. Horodecki, M. Horodecki, P. Horodecki, and J. Oppenheim. Locking entanglement measures with a single qubit. *Phys. Rev. Lett.*, 94:200501, 2005. arXiv:quant-ph/0404096.

[36] P. Hayden, D. W. Leung, and A. Winter. Aspects of Generic Entanglement. *Comm. Math. Phys.*, 265:95–117, 2006. arXiv:quant-ph/0407049.

[37] S. Popescu, A. J. Short, and A. Winter. Entanglement and the foundations of statistical mechanics. *Nature Phys.*, 2(11):754–758, 2006. arXiv:quant-ph/0511225.

[38] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. The mother of all protocols: restructuring quantum information's family tree. *Proc. R. Soc. A*, 465(2108):2537–2563, 2009. arXiv:quant-ph/0606225.

[39] M. Fannes. A continuity property of the entropy density for spin lattice systems. *Comm. Math. Phys.*, 31:291–294, 1973.

[40] G. D. Mostow. *Strong rigidity of locally symmetric spaces*. Princeton University Press, 1973.

[41] R. Jozsa. Fidelity for mixed quantum states. *J. Mod. Opt.*, 41:2315–2323, 1994.

[42] A. Uhlmann. The 'transition probability' in the state space of a $*$-algebra. *Rep. Math. Phys.*, 9:273, 1976.

[43] R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics. Springer, 1996.

[44] A. Winter. Coding theorem and strong converse for quantum channels. *IEEE Trans. Inf. Theory*, 45(7):2481–2485, 1999.

[45] T. Ogawa and H. Nagaoka. A new proof of the channel coding via hypothesis testing in quantum information theory. In *Proc. 2002 IEEE ISIT*, page 73, 2002. arXiv:quant-ph/0208139.