Information-theoretic treatment of tripartite systems and quantum channels

Patrick J. Coles, Li Yu, Vlad Gheorghiu, and Robert B. Griffiths

Phys. Rev. A **83**, 062338 — Published 30 June 2011

DOI: 10.1103/PhysRevA.83.062338

Information-theoretic treatment of tripartite systems and quantum channels

Patrick J. Coles,* Li Yu,† Vlad Gheorghiu,‡ and Robert B. Griffiths§
Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

A Holevo measure is used to discuss how much information about a given POVM on system a is present in another system b, and how this influences the presence or absence of information about a different POVM on a in a third system c. The main goal is to extend information theorems for mutually unbiased bases or general bases to arbitrary POVMs, and especially to generalize "all-or-nothing" theorems about information located in tripartite systems to the case of *partial information*, in the form of quantitative inequalities. Some of the inequalities can be viewed as entropic uncertainty relations that apply in the presence of quantum side information, as in recent work by Berta et al. [Nature Physics 6, 659 (2010)]. All of the results also apply to quantum channels: e.g., if $\mathcal{E}$ accurately transmits certain POVMs, the complementary channel $\mathcal{F}$ will necessarily be noisy for certain other POVMs. While the inequalities are valid for mixed states of tripartite systems, restricting to pure states leads to the basis-invariance of the difference between the information about a contained in b and c.

PACS numbers: 03.67.-a, 03.67.Hk

I. INTRODUCTION

A significant part of current quantum information research can be understood as an attempt to find answers to the following question: How much of what kind of information about what is located where? In this paper we provide specific answers to these questions in the case of a general tripartite quantum system: subsystems a, b, and c are described by some sort of quantum state (pre-probability) that induces a joint probability distribution on different properties of these systems. Appropriate statistical correlations can then be thought of in terms of, for example, system b containing information of some sort about certain physical properties of system a. To discuss how much information of this kind is contained in or can be found in b requires some sort of quantitative measure, and it is natural to look for something resembling the well-known Shannon measures in classical information; see [1] for a modern introduction to this subject.

Although it is rather natural to treat systems a, b and c on an equal footing—and that is the perspective of this paper—one can also think of the properties as existing at different times. For example, a might be the entrance to a quantum channel with b, possibly but not necessarily the same physical system, the output of the channel and c the "environment" at this later point in time. Such a *dynamical* perspective is well-known in classical information theory as it applies to a noisy channel, where it can be discussed using the same information measures, e.g., the mutual information $H(X\!:\!Y)$, that apply to statistically-correlated systems (think of a shared key used for cryptographic purposes) at the same time, a *static* perspective. Both perspectives are also possible for problems in quantum information theory, though this has not received as much attention as we think it deserves, and viewing expressions which are formally the same (or closely related) from distinct points of view can make a valuable contribution to one's intuitive understanding of a situation.

Of course, quantum information theory is more general than classical information theory, so the conceptual ideas provided by the latter are insufficient for discussing the quantum world. In this paper we take the perspective that a valuable way to think about the quantum case is to distinguish different *types* or *species* of quantum information [2]. For example, if a is a single qubit the distinction between $|0\rangle$ and $|1\rangle$ constitutes the "z" type of information, whereas the distinction between $|+\rangle$ and $|-\rangle$, with $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$, is the "$x$ type." Each type by itself, even when it refers to microscopic (thus "quantum") properties, follows the usual rules of classical information theory. This allows one to immediately transfer a large body of mathematical formalism and associated physical intuition from the classical to the quantum domain without risk of falling prey to inconsistencies and paradoxes. The quantum nature of the microscopic world then manifests itself through the fact that incompatible types of information, corresponding to non-commuting projective decompositions of the identity, cannot be combined: this is the single framework rule (see, e.g., Ch. 16 of [3]) that allows a fully-consistent use of probabilities in the quantum domain.[1]

*Electronic address: pcoles@andrew.cmu.edu
†Electronic address: liy@andrew.cmu.edu
‡Electronic address: vgheorgh@andrew.cmu.edu
§Electronic address: rgrif@andrew.cmu.edu

[1] It is important to note that a type of information as defined here refers primarily to a *microscopic* quantum property rather than the outcome of a measurement. A correctly constructed measurement apparatus can reveal the property of a microscopic system, so that, for example a Stern-Gerlach apparatus followed by detectors can determine if the spin-half particle entering the apparatus had $S_z = +\hbar/2$ or $-\hbar/2$, corresponding to the qubit

In this paper we generalize the notion of a type of quantum information so that it includes not only a projective decomposition of the identity, a set of projectors that sum to the identity, but also a general POVM, a collection of positive operators that sum to the identity. The idea, discussed in Sec. II A, is that while the operators in a POVM are in general not orthogonal, each corresponds to a projector on a larger Hilbert space, the Naimark extension (which is not unique), and the collection of such projectors sums to the identity on the larger space thus constituting a particular type of quantum information in the sense previously discussed.

The question "How much?" has motivated an ongoing search for measures that extend the very useful idea of entanglement beyond bipartite pure states where it was first introduced. Despite a great deal of effort and a large number of intriguing results [5], it seems fair to say that there remain a large number of unanswered questions even for bipartite mixed states, not to mention the multipartite case. It is not obvious that a single number representing the entanglement, or even a small collection of numbers, will suffice to embody the physical insights needed for a better understanding of such systems. In this paper we introduce measures of information that depend explicitly on the (quantum) type of information one is considering, so we can address the question of, for example, how well a noisy quantum channel performs for different types of input. The definitions and a detailed discussion of these measures will be found in Sec. III; at this point it suffices to note that they are of the Holevo form using the quantum von Neumann entropy, though in some cases they can be generalized using other types of entropy.

As well as direct quantitative measures of information certain *differences* in information measures, e.g., the amount of information of a given type that is in b minus how much is in c, are of interest. We refer to these as entropy or information *biases*. It is not without interest that the *coherent information* [6] when expressed in the language of tripartite systems is (or least can be thought of as) such a bias; see Sec. III C. In Sec. IV we show that under appropriate circumstances an information bias will be independent of the type of information under consideration.

One of the most striking features of quantum information is that if information of a particular type corresponding to some orthonormal basis w of system a is perfectly present (perfect correlation, no noise) in system b for the quantum state under discussion, this prevents or excludes a type of information v corresponding to a basis mutually unbiased (MU) with respect to w—that is, v and w are mutually-unbiased bases (MUBs)—from being present in a third system c. In Sec. V of this paper we present quantitative generalizations of this and some other "all-or-nothing" theorems to situations in which, for example, almost all information of the w type of information about a is in b and one wants to bound how much v information, where v is only approximately MU with respect to w, can be present in c.

In particular, Theorem 5 in Sec. V B presents a bound of this form. It extends to POVMs an important inequality proved in [7], earlier conjectured in [8], using a somewhat simpler proof. This extension was also recently proven in [9] using smooth entropies; in contrast our proof approach is based on the relative entropy. Various consequences, including the application to a channel and its complementary channel, are worked out in various corollaries. As well as thinking of this result as a bound on the amounts of two strongly incompatible (in the sense of almost MUB) types of information about a present in different locations, Theorem 5 constitutes a generalized entropic uncertainty relation for system a when the coupling to another system or systems is taken into account ("quantum side information" in the sense discussed in [8, 10]).

Several additional quantitative generalizations of all-or-nothing results are given in Secs. V A, V C, and V D. The all-or-nothing results can be succinctly stated as follows for orthonormal bases u, v, and w of a, where u and v are MU relative to w (but not necessarily to each other). If the w type of information is perfectly present in b, then (1) ρ_{ac} is block diagonal in the w basis (Lemma 4), (2) the amount of u information in b is equal to the amount of v information in b (Theorem 8), (3) if the v information is perfectly present in b then there is a perfect quantum channel from a to b (Theorem 10), (4) if the w information is completely absent from c, then no information about a is in c: the two are decoupled (Theorem 11).

The remainder of this paper is organized as follows. Section II is an introduction to tripartite systems, including the connection with quantum channels and their complements, and provides details of what we mean by different types of quantum information. Various quantitative measures of information are introduced, and some of their properties discussed, in Sec. III. Our main results, which, as indicated above, provide quantitative bounds on the location of various types of information in different systems, occupy Secs. IV and V. Section VI relates our work to various other approaches and publications. A summary, which provides an overview of how the different theorems are related to each other, is in Sec. VII A, followed by an indication of issues worth further exploration in Sec. VII B. To make the main presentation com-

states $|0\rangle$ or $|1\rangle$, and in this case the z information initially possessed by the particle is translated into distinct macroscopic apparatus states, making the z information "visible" or "classical." (For an important application to quantum information theory of the idea that a macroscopic quantum outcome reveals a prior microscopic state, see [4]; for a detailed discussion of the measurement process in fully quantum terms, see Chs. 17 and 18 of [3].) However, the concept of z information can also be used in situations, such as when a qubit is just entering a quantum channel, where trying to relate it to a measurement, at least as a physical process occurring at that point in time, is not very helpful.

pact and easier to follow, all but the very shortest proofs have been relegated to appendices.

II. SYSTEMS WITH THREE PARTS

A. POVMs and types of information

Much work in contemporary quantum information theory is devoted to particular instances of what may be called the *tripartite system problem* defined in the following way. Let $\mathcal{H}_{abc} = \mathcal{H}_a \otimes \mathcal{H}_b \otimes \mathcal{H}_c$ be a tensor product of Hilbert spaces of dimensions d_a, d_b, d_c, all assumed to be finite, and let

$$I_a = \sum_j P_{aj}, \quad I_b = \sum_k Q_{bk}, \quad I_c = \sum_l R_{cl} \tag{1}$$

be three POVMs, decompositions of their respective identities into *finite* sets of positive operators, hereafter referred to as P_a, etc.[2] [Note that we use the symbols a, b, and c as subscripts (but occasionally on line) to label subsystems, and indices j, k, l, etc. to label the POVM elements.] What can be said about the joint probability distribution

$$\Pr(P_{aj}, Q_{bk}, R_{cl}) = \mathrm{Tr}(P_{aj} Q_{bk} R_{cl} \rho_{abc}), \tag{2}$$

where ρ_{abc} is a density operator acting as a pre-probability (generator of probabilities in the terminology of Sec. 9.4 of [3]), perhaps but not necessarily a projector $|\Omega\rangle\langle\Omega|$ on the pure state $|\Omega\rangle$? In particular, what is its information-theoretic significance? One is, of course, interested in how these probabilities, and the corresponding marginal distributions such as

$$\Pr(P_{aj}, Q_{bk}) = \sum_l \Pr(P_{aj}, Q_{bk}, R_{cl}) = \mathrm{Tr}_{ab}(P_{aj} Q_{bk} \rho_{ab}), \tag{3}$$

with ρ_{ab} the partial trace over $\mathcal{H}_c$ of ρ_{abc}, depend upon the indices j, k, and l. But of equal, or even greater interest is their dependence upon the *choice of POVMs* in (1). Here quantum theory, in contrast to classical physics, allows an enormous number of possibilities.

In what follows we shall want to distinguish various different types of POVM. A *rank-1* POVM is one in which all the positive operators are of rank 1, which is to say proportional to projectors on one-dimensional spaces; we will employ symbols L, M, N to denote such POVMs. When all the POVM elements are projectors (orthogonal projection operators) we have a *projective decomposition* of the identity. A rank-1 projective decomposition is associated with an *orthonormal basis*; e.g., the orthonormal basis $w = \{|w_j\rangle\}$ of $\mathcal{H}_a$ gives rise to the decomposition

$$P_{aj} = |w_j\rangle\langle w_j|. \tag{4}$$

In what follows we use the lower case letters u, v, and w to denote orthonormal bases, and where useful add a subscript, e.g., w_a, to indicate the corresponding system or Hilbert space. A second basis $v = \{|v_j\rangle\}$ is *mutually unbiased* (MU) relative to w—the terms *complementary* or *conjugate* are also in use—thus v and w are mutually unbiased bases (MUBs), when $|\langle v_j|w_k\rangle| = 1/\sqrt{d_a}$ is independent of j and k.

Unlike a general POVM, a projective decomposition can be given a simple physical interpretation: the projectors, or the subspaces onto which they project, form a quantum sample space: a collection of mutually exclusive physical properties one and only one of which is true; see Ch. 5 of [3]. In previous work [2] such a projective decomposition was called a *type of information*: e.g., $\Pi_a = \{\Pi_{aj}\}$ is a type of information about the system a. Two types of information Π_a and Φ_a about the same system are *compatible* provided every projector in one set commutes with every projector in the other set: $\Pi_{aj}\Phi_{ak} = \Phi_{ak}\Pi_{aj}$ for every j and k; otherwise they are *incompatible*. Two distinct rank-1 projective decompositions, or the corresponding orthonormal bases, are necessarily incompatible if they differ by more than simply relabeling the projectors, and two MUBs are incompatible to the maximum extent possible. Probabilistic arguments in quantum mechanics cannot combine results from incompatible decompositions—the *single framework* rule, see Ch. 16 of [3]—without risk of generating contradictions and paradoxes.

However, in the present paper we generalize the notion of a type of information about (say) system a to include any POVM P_a when interpreted using a *Naimark extension*; see [11, 12] or Sec. 9-6 of [13]. Assume that the Hilbert space $\mathcal{H}_a$ is a subspace of a larger Hilbert space $\mathcal{H}_A$, with E_a the operator on $\mathcal{H}_A$ that projects onto $\mathcal{H}_a$. If $\mathcal{H}_A$ has been appropriately chosen there is a projective decomposition $\{\Pi_{Aj}\}$ of its identity I_A such that

$$P_{aj} = E_a \Pi_{Aj} E_a. \tag{5}$$

In addition, one can always arrange that for each j the rank of Π_{Aj} is the same as the rank of P_{aj}, though one may need an additional projector, call it Π_{A0}, which is orthogonal to E_a, so the corresponding P_{a0} is the zero operator. (It is possible to set things up so that the rank of Π_{Aj} exceeds that of P_{aj}, but in light of (5) the reverse is impossible.) An important special case used in proving later results is that any rank-1 POVM N on a is equivalent to some rank-1 projective decomposition (orthonormal basis) on A [12]. One can if one wishes think of $\mathcal{H}_A$ as a tensor product $\mathcal{H}_a \otimes \mathcal{H}_e$, where $\mathcal{H}_e$ is the Hilbert space of some reference system, and $\mathcal{H}_a$ is itself (isomorphic to) the subspace of kets of the form $|\psi\rangle \otimes |e_0\rangle$, with

[2] It is sometimes helpful to imagine the three parts as residing in three different places, say three different laboratories where Alice, Bob, and Carol can carry out separate preparations and measurements on them.

$|e_0\rangle$ a fixed, normalized ket in $\mathcal{H}_e$. In this case the density operator on A is $\rho_A = \rho_{ae} = \rho_a \otimes |e_0\rangle\langle e_0|$, and E_a in (5) is simply $I_a \otimes |e_0\rangle\langle e_0|$. Starting with the projective decomposition Π_A for the larger system A, one can think of the corresponding positive operators defined in (5) as convenient mathematical tools for computing probabilities in those cases in which the density operator ρ_A has support in the subspace $\mathcal{H}_a$ onto which E_a projects. From this perspective, and using the corresponding Naimark extensions for b and c, one could reformulate the results given in later sections of this paper in terms of projective decompositions on the larger Hilbert spaces. However, the use of POVMs provides in many cases a simpler mathematical form, and a theorem that refers to an arbitrary POVM obviously includes projective decompositions as particular cases. Nonetheless, when thinking in physical or operational terms about the type of information represented by a general POVM P_a it is helpful to employ its Naimark counterpart.

The information about a POVM $P_a = \{P_{aj}\}$ is said to be *completely* or *perfectly present* in system b provided the conditional density operators

$$p_j \rho_{bj} = \mathrm{Tr}_a(P_{aj}\rho_{ab}); \quad p_j := \Pr(P_{aj}) = \mathrm{Tr}(P_{aj}\rho_a), \quad (6)$$

on $\mathcal{H}_b$ are mutually orthogonal: $\rho_{bj}\rho_{bj'} = 0$ for $j \neq j'$. Conversely, this type of information is (completely) *absent* from b when the conditional density operators ρ_{bj} are identical. One can visualize this in terms of measurements as follows. Suppose a POVM $\{P_{aj}\}$ measurement is carried out on system a. Can the value of j be deduced by carrying out an appropriate sort of measurement on system b? If the ρ_{bj} are orthogonal to each other this is clearly possible using (projective) measurements corresponding to a suitable decomposition $\{Q_{bk}\}$ of I_b. But in the other extreme in which the ρ_{bj} are identical it is clear that no measurement on b will provide any information about j. It is worth noting that one obtains the same values in (6) by replacing the formulas with $p_j\rho_{bj} = \mathrm{Tr}_A(\Pi_{Aj}\rho_{Ab})$ and $p_j = \mathrm{Tr}(\Pi_{Aj}\rho_A)$, so all of our measures in Sec. III B quantifying the presence of the P_a information in b, depending only on $\{p_j, \rho_{bj}\}$, will be unaffected by replacing P_a with its Naimark extension Π_A.

B. Quantum channels

In some sense the most natural way to state the various results given below in Secs. IV and V is in terms of correlations in which all three parts a, b, and c are treated, at least formally, in a symmetrical fashion. But some of the more interesting applications are to quantum channels and complementary channels, in which the channel entrance is not treated in the same way, either formally or intuitively, as the channel output. Hence in order to facilitate application of our results to the case of channels, we provide a brief explanation, using ideas in [14, 15], of why the "tripartite" and the "channel" problem are not only closely related to each other, but in some sense identical problems in the case where one restricts attention to a pure-state pre-probability $|\Omega\rangle \in \mathcal{H}_{abc}$.

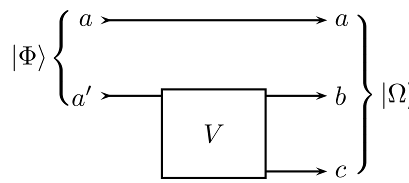

FIG. 1: How $|\Omega\rangle$ is produced by applying the isometry V to an entangled state $|\Phi\rangle$.

Consider the situation shown in Fig. 1 where

$$|\Omega\rangle = (I_a \otimes V)|\Phi\rangle. \quad (7)$$

is the result of applying an isometry

$$V = \sum_j |s_j\rangle\langle a'_j| \quad (8)$$

to the a' part of an entangled state $|\Phi\rangle \in \mathcal{H}_a \otimes \mathcal{H}_{a'}$, with $\mathcal{H}_{a'}$ a copy (i.e., the same dimension) of $\mathcal{H}_a$. Here $\{|a'_j\rangle\}$ is some orthonormal basis of $\mathcal{H}_{a'}$ held fixed during the following discussion, we are assuming that $d_a \leqslant d_b d_c$, and the requirement that V be an isometry, which is to say $V^\dagger V = I_a$ is equivalent to the assumption that the kets $\{|s_j\rangle\}$ form an orthonormal collection spanning the subspace $\mathcal{H}_s = V\mathcal{H}_{a'}$ of $\mathcal{H}_{bc}$.

If, in particular, $|\Phi\rangle$ is the fully-entangled state

$$|\Phi\rangle = (1/\sqrt{d_a}) \sum_j |a_j\rangle \otimes |a'_j\rangle, \quad (9)$$

then

$$|\Omega\rangle = (1/\sqrt{d_a}) \sum_j |a_j\rangle \otimes |s_j\rangle, \quad (10)$$

is an example of what in [14] is called a *channel ket*, characterized by the property that

$$\rho_a = \mathrm{Tr}_{bc}(|\Omega\rangle\langle\Omega|) = I_a/d_a. \quad (11)$$

Indeed, given a pre-probability $|\Omega\rangle$ such that (11) holds, it is necessarily a fully-entangled state on $\mathcal{H}_a \otimes \mathcal{H}_{bc}$, so it will have a Schmidt form (10) for $\{|a_j\rangle\}$ a given orthonormal basis of $\mathcal{H}_a$, and using the orthonormal collection of states $\{|s_j\rangle\}$ corresponding to this Schmidt decomposition one can define a corresponding isometry V by means of (8). Thus by employing map-state duality (see e.g. [14] or Ch. 11 of [15]) one can move from a channel ket $|\Omega\rangle$ satisfying (11) to an isometry V or the reverse.

From an information-theoretic perspective the isometry V corresponds to saying that all information about the system a is in the system bc, and in fact in the subspace $\mathcal{H}_s$ of $\mathcal{H}_{bc}$ spanned by the $|s_j\rangle$ in (10). The partial traces onto $\mathcal{H}_b$ and $\mathcal{H}_c$ in a sense "project down" parts of this information onto these subsystems. Thus, not surprisingly, the projector Υ onto $\mathcal{H}_s$ along with its partial traces down to $\mathcal{H}_b$ and $\mathcal{H}_c$,

$$\Upsilon = VV^\dagger = \sum_j |s_j\rangle\langle s_j|,$$
$$\Upsilon_b = \mathrm{Tr}_c(\Upsilon), \quad \Upsilon_c = \mathrm{Tr}_b(\Upsilon), \tag{12}$$

play useful roles in our thinking about these problems, as they in a sense describe, in a basis-independent way, how the subspace $\mathcal{H}_s$ is "oriented" relative to the factor spaces $\mathcal{H}_b$ and $\mathcal{H}_c$. Note that while Υ is a projector, Υ_b and Υ_c are positive operators but (in general) not projectors.

The isometry V in (8) can be used to define a *quantum channel* from a to b through the superoperator

$$\mathcal{E}(A) = \mathrm{Tr}_c(VAV^\dagger) = \sum_l K_l A K_l^\dagger \tag{13}$$

that maps the space $\mathcal{L}(\mathcal{H}_a)$ of operators on $\mathcal{H}_a$ to the corresponding space $\mathcal{L}(\mathcal{H}_b)$ of operators on $\mathcal{H}_b$. Here the Kraus operators are maps from $\mathcal{H}_a$ to $\mathcal{H}_b$ of the form

$$K_l = \langle c_l|V = \sum_j \langle c_l|s_j\rangle\langle a_j|, \tag{14}$$

where $\{|c_l\rangle\}$ is an orthonormal basis of $\mathcal{H}_c$, and $\langle c_l|s_j\rangle$ is a ket in $\mathcal{H}_b$, defined in an obvious way, not just a complex number. Because V is an isometry the Kraus operators satisfy the usual closure condition

$$\sum K_l^\dagger K_l = I_a. \tag{15}$$

The *complementary* channel from a to c,

$$\mathcal{F}(A) = \mathrm{Tr}_b(VAV^\dagger) = \sum_m L_m A L_m^\dagger, \tag{16}$$

is defined in a similar way with Kraus operators

$$L_m = \langle b_m|V = \sum_j \langle b_m|s_j\rangle\langle a_j|, \tag{17}$$

for $\{|b_m\rangle\}$ some orthonormal basis of $\mathcal{H}_b$, and these again satisfy the closure condition analogous to (15).

The superoperators $\mathcal{E}$ and $\mathcal{F}$, and their adjoints $\mathcal{E}^\dagger$ and $\mathcal{F}^\dagger$ relative to the usual Frobenius inner product $\langle P, Q\rangle = \mathrm{Tr}(P^\dagger Q)$, can be expressed directly in terms of $\rho_{abc} = |\Omega\rangle\langle\Omega|$, or its partial traces such as ρ_{ab}, using formulas such as

$$\mathcal{E}(A) = d_a \mathrm{Tr}_a[(A^\mathrm{T} \otimes I_b)\rho_{ab}],$$
$$[\mathcal{E}^\dagger(B)]^\mathrm{T} = d_a \mathrm{Tr}_b[(I_a \otimes B)\rho_{ab}], \tag{18}$$

where $^\mathrm{T}$ denotes the transpose relative to the basis $\{|a_j\rangle\}$ employed in (9) and (10). The complete positivity of $\mathcal{E}$ is equivalent to the requirement that ρ_{ab} be a positive operator; in some respects this is simpler and more compact than the traditional definition. For it to be trace preserving it is necessary that ρ_a be I_a/d_a, (11). Since in general neither ρ_b nor ρ_c is proportional to the corresponding identity, the adjoints $\mathcal{E}^\dagger$ and $\mathcal{F}^\dagger$ are not (in general) trace preserving, and in this sense are not quantum channels. This is one respect in which "tripartite" language is more flexible than "channel" language.

It is also worth observing that the superoperator $\mathcal{E}$ uniquely determines $|\Omega\rangle$ up to local unitaries on $\mathcal{H}_a$ and $\mathcal{H}_c$ for a fixed d_c. This is because a set of Kraus operators is generated, (14), using an orthonormal basis $\{|c_l\rangle\}$ of $\mathcal{H}_c$, and one can invert the process by writing $V = \sum_l |c_l\rangle K_l$, where of course the result depends on the choice of basis $\{|c_l\rangle\}$. Different orthonormal bases on $\mathcal{H}_c$, as is well-known, simply give rise to different collections of Kraus operators which represent the same quantum channel or operation. In this sense a channel completely determines its complementary channel for a fixed d_c, and vice versa, up to local unitaries. However, different insights may emerge by considering one rather than the other, or by thinking about the two together.

We say there exists a *perfect quantum channel* from a to b when all types of information about a are perfectly present in b. This by itself implies that $\rho_a = I_a/d_a$ (see Theorem 3 in [14]), and thus $\mathcal{E}$ in (18) is trace-preserving. It obviously suffices to check that the information associated with every orthonormal basis is present in b, but there are also weaker conditions that ensure the presence of a perfect quantum channel; e.g. see [2, 16] and the discussion in Sec. V D.

A more general relationship is possible between an isometry V and a tripartite pure state, by starting with (7), the circuit in Fig. 1, but assuming that $|\Phi\rangle$, while no longer fully entangled, has full Schmidt rank:

$$|\Phi\rangle = \sum_k \sqrt{\pi_k}\, |a_k\rangle \otimes |a'_k\rangle, \tag{19}$$

with $\pi_k > 0$ for every k. With V an isometry of the form (8), j replaced by k, and

$$\rho_a = \sum_k \pi_k |a_k\rangle\langle a_k| \tag{20}$$

the partial trace of $|\Phi\rangle\langle\Phi|$ down to $\mathcal{H}_a$, one has

$$\rho_b = \mathrm{Tr}_{ac}(|\Omega\rangle\langle\Omega|) = \mathcal{E}(\rho_a), \tag{21}$$

where $\mathcal{E}$ is the superoperator corresponding to V through (13). A similar result holds for the complementary a to c channel. The ket $|\Omega\rangle$ determines the projector $\Upsilon = VV^\dagger$ uniquely, but V itself only up to a unitary transformation on $\mathcal{H}_a$. Conversely, two isometries V and $\widetilde{V}$ giving rise to the same Υ can be used to generate the same $|\Omega\rangle$ by using two different entangled states $|\Phi\rangle$ and $|\widetilde{\Phi}\rangle$.

The partially entangled $|\Phi\rangle$ (19) is useful when addressing the following question: Suppose an ensemble $\{p_j, \rho_j\}$ of states is sent through the quantum channel $\mathcal{E}$. How can one relate the outputs $\mathcal{E}(\rho_j)$ of the channel to corresponding outcomes of a POVM measurement P_a

on the tripartite state $|\Omega\rangle$? Suppose the density operator $\rho_a = \sum_j p_j\rho_j$ for the ensemble is of the form (20), i.e., choose $|\Phi\rangle$ in (19) such that this is the case. Then define P_a through

$$P_{aj}^{\mathrm{T}} = p_j W \rho_j W^\dagger, \tag{22}$$

where $^{\mathrm{T}}$ denotes the transpose in the basis $\{|a_k\rangle\}$, and

$$W = \sum_k (1/\sqrt{\pi_k})|a_k\rangle\langle a'_k|. \tag{23}$$

It is straightforward to show that P_{aj} is a positive operator with the same rank as ρ_j (since W is nonsingular), and $\sum_j P_{aj} = I_a$. The probability of outcome j for the POVM is p_j, and the corresponding conditional density operator is

$$\rho_{bcj} = \mathrm{Tr}_a(P_{aj}\rho_{abc}) = V\rho_j V^\dagger, \tag{24}$$

with $\rho_{abc} = |\Omega\rangle\langle\Omega|$. Tracing this down to b yields $\mathcal{E}(\rho_j)$, the outcome when ρ_j is sent through the channel.

The preceding discussion requires some fairly obvious changes if some of the π_k in (19) are zero. First, $\Upsilon = VV^\dagger$ is not uniquely determined by $|\Omega\rangle$, since the $|s_k\rangle$ in (8) corresponding to zero π_k are unknown. Second, the sum in (23) must be restricted to the k with $\pi_k > 0$, whereas (22) remains the same.

III. INFORMATION MEASURES

A. Entropies

All the information measures that we will introduce are based on some sort of entropy. In classical information theory [1] the usual starting point is the Shannon entropy

$$H(P) = H(\{p_j\}) = -\sum_j p_j \log p_j, \tag{25}$$

where P denotes a random variable or its corresponding probability distribution. Given two random variables P and Q the entropy $H(P,Q)$ is obtained by replacing p_j in (25) by the joint probability distribution $p_{jk} = \Pr(P_j, Q_k)$ and summing over both j and k. The *conditional entropy* and *mutual information* are then defined by:

$$\begin{aligned} H(P|Q) &= H(P,Q) - H(Q), \\ H(P:Q) &= H(P) + H(Q) - H(P,Q). \end{aligned} \tag{26}$$

The quantum entropy most closely analogous to Shannon's H is the von Neumann entropy

$$S(\rho) = -\mathrm{Tr}(\rho\log\rho), \tag{27}$$

but we have also studied some other possibilities:

$$\begin{aligned} S_R(\rho) &= \frac{1}{1-q}\log\mathrm{Tr}(\rho^q), \quad 0 < q \leqslant 1, \\ S_T(\rho) &= \frac{1}{1-q}[\mathrm{Tr}(\rho^q) - 1], \quad 0 < q \leqslant \infty, \\ S_Q(\rho) &= 1 - \mathrm{Tr}(\rho^2). \end{aligned} \tag{28}$$

Here S_R, S_T, and S_Q are the *Renyi, Tsallis*, and *quadratic* (often misleadingly called *linear*) entropies, respectively. Some of our results are valid for all these entropies, in which case they are stated for S_K, where K denotes either no subscript (von Neumann) or else one of the three symbols R, T, Q.

All of these entropies are strictly concave, $S_K(\sum p_j\rho_j) \geqslant \sum p_j S_K(\rho_j)$ for $0 < p_j < 1$ and $\sum p_j = 1$, with equality if and only if all ρ_j's are equal, provided the parameter q in the case of S_R and S_T is in range specified in (28). Both S_R and S_T are equal to S in the limit $q = 1$, and S_T interpolates between S and S_Q as q goes from 1 to 2. The entropies S, S_Q, and S_T for $q \geqslant 1$, are subadditive [17] in the sense that $S_K(\rho_a) + S_K(\rho_b) \geqslant S_K(\rho_{ab})$, but only the von Neumann S has the property of strong subadditivity on a tripartite system (p. 519 of [6]):

$$S(\rho_{ab}) + S(\rho_{bc}) \geqslant S(\rho_{abc}) + S(\rho_b). \tag{29}$$

Given a bipartite quantum system with a density operator ρ_{ab}, partial traces ρ_a and ρ_b, the *quantum conditional entropy* and the *quantum mutual information* are defined as (p. 514 of [6])

$$\begin{aligned} S(a|b) &= S(\rho_{ab}) - S(\rho_b), \\ S(a:b) &= S(\rho_a) + S(\rho_b) - S(\rho_{ab}), \end{aligned} \tag{30}$$

which are formally analogous to the quantities in (26). Note that $S(a|b)$ can be negative. On the other hand, $S(a:b)$ is nonnegative and vanishes for a product state $\rho_{ab} = \rho_a \otimes \rho_b$, and thus can be regarded in some sense as a measure of how much information about a is in b or vice versa. Thought of in this way it has the property that for a tripartite system abc,

$$S(a:bc) \geqslant S(a:b), \tag{31}$$

i.e., there is less information about a in b, a subsystem of bc, than there is in bc, which seems a reasonable requirement for a measure of information. Note that (31) is equivalent to (29), a property not shared (in general) by the other entropies defined in (28).

We shall later prove our main result using the relative entropy:

$$S(\rho||\sigma) = \mathrm{Tr}(\rho\log\rho) - \mathrm{Tr}(\rho\log\sigma), \tag{32}$$

which has the useful property [18] that it is nonincreasing under the action of a quantum channel $\mathcal{E}$,

$$S(\rho||\sigma) \geqslant S(\mathcal{E}(\rho)||\mathcal{E}(\sigma)). \tag{33}$$

The extension of (32) to general positive operators is natural, and [19] for any positive operators A, B, and C, if $C \geqslant B$ (i.e. $C - B$ is a positive operator),

$$S(A||B) \geqslant S(A||C). \tag{34}$$

B. Distinguishability measures

While $S(a:b)$ can serve as an overall indication of how much information about a is in b, or vice versa, it is not a measure that depends on the *type* of information, so cannot be used to compare how well different types of information about a are found in, or transmitted to b. For this purpose one could use a *fidelity* measure: how closely a state on $\mathcal{H}_b$ resembles its counterpart on $\mathcal{H}_a$. However, this requires making some identification between the two Hilbert spaces, which is not easy to do if they are of different dimension, or else one needs an additional map or channel to carry $\mathcal{H}_b$ back to $\mathcal{H}_a$. For this and other reasons we prefer to use a *distinguishability* measure. Thus suppose $P_a = \{P_{aj}\}$ is a decomposition of the identity I_a of $\mathcal{H}_a$, (1), and $\{p_j, \rho_{bj}\}$ is the ensemble of conditional states on $\mathcal{H}_b$ defined in (6). Two extreme cases were discussed in Sec. II A: that in which the P_a type of information is perfectly present in b, which means $\rho_{bj}\rho_{bk} = 0$ for $j \neq k$, thus conditional density operators perfectly distinguishable; and the P_a type of information (completely) absent from b, meaning the ρ_{bj} are identical for all j and thus indistinguishable. Our goal is to assign numerical values to situations lying between these extremes.

Ideally one might use some collection of numbers referring to the distinguishability of every pair of conditional density operators ρ_{bj}, see [20] for an overview of distinguishability measures for two density operators. However, we shall employ a much coarser but still useful characterization in which a single number, in some sense an "average" distinguishability, is assigned to each information type, thereby allowing us to focus on our primary goal: elucidating how the amount of information depends upon the type considered, for a given pre-probability (density operator or channel). As is customary in information theory we want a measure that is nonnegative, that is (formally) invariant under local unitary operations, and, naturally, we prefer simple mathematical expressions that have a clear intuitive interpretation. This still leaves many possibilities, but among them we have found that measures based on the Holevo function

$$\chi_K(\{p_j, \rho_j\}) = S_K(\sum_j p_j \rho_j) - \sum_j p_j S_K(\rho_j) \qquad (35)$$

are particularly useful, where $\{p_j, \rho_j\}$ denotes an *ensemble* associated with a particular Hilbert space $\mathcal{H}$: each ρ_j a density operator on this space, and the $\{p_j\}$ a probability distribution. Here S_K could be any of the entropies defined in (27) or (28); S without a subscript refers to the von Neumann entropy, and the corresponding χ has no subscript. Because each of these entropies is a strictly concave function (for q in the appropriate range indicated in (28)), χ_K is nonnegative and equal to zero if and only if the ρ_j are identical.

When (35) is applied to the ensemble $\{p_j, \rho_{bj}\}$ of (6), states in $\mathcal{H}_b$ conditional on the decomposition $P_a = \{P_{aj}\}$ in (1), the result is

$$\chi_K(P_a, b) := S_K(\rho_b) - \sum_j p_j S_K(\rho_{bj}), \qquad (36)$$

a measure of the amount of information of type P_a in b. This is also a numerical measure of what is sometimes called quantum side information [8, 10].

While P_a can refer to a general projective decomposition of I_a or a POVM, we will often be interested in an orthonormal basis $\{|w_j\rangle\}$, projectors $|w_j\rangle\langle w_j|$, of $\mathcal{H}_a$, in which case we will write $\chi_K(w, b)$, omitting the a subscript when it is obvious from the context. One can easily show using the concavity of S_K that

$$\chi_K(P_a, b) \geqslant \chi_K(\widetilde{P}_a, b), \qquad (37)$$

where P_a and $\widetilde{P}_a$ are POVMs, and $\widetilde{P}_a$ is a coarse-graining of P_a formed by summing some of the P_{aj} elements. Also, as a consequence of (29), see [21],

$$\chi(P_a, bc) \geqslant \chi(P_a, b), \qquad (38)$$

so a subsystem b of bc cannot contain more information than bc itself. (This does not hold for χ_K with $K = R$, T or Q.)

In the case of a quantum channel $\mathcal{E}$ (13) from a to b associated with isometry V from a to bc, we define

$$\chi_K(P_a, \mathcal{E}) := S_K[\mathcal{E}(\sum p_j \rho_{aj})] - \sum p_j S_K[\mathcal{E}(\rho_{aj})], \quad (39)$$

where P_a is a POVM, $I_a = \sum P_{aj} = d_a \sum p_j \rho_{aj}$, with

$$\rho_{aj} = P_{aj}/\mathrm{Tr}(P_{aj}), \quad p_j = \mathrm{Tr}(P_{aj})/d_a. \qquad (40)$$

Note that $\mathcal{E}(\sum p_j \rho_{aj}) = \mathrm{Tr}_c(VV^\dagger)/d_a = \Upsilon_b/d_a$ [see (12)] in the first term of (39) is independent of the POVM P_a. Equation (39) is some measure for how well $\mathcal{E}$ preserves the distinguishability of the P_a ensemble; e.g. if $\mathcal{E}$ perfectly preserves the orthogonality of an input orthonormal basis w then $\chi(w, \mathcal{E}) = \log d_a$, otherwise $\chi(w, \mathcal{E}) < \log d_a$ (see Lemma 1 below).

In contrast to $\chi(P_a, b)$, the quantity [10]

$$H(P_a|b) := H(P_a) - \chi(P_a, b) \qquad (41)$$

is a measure of *absence* of the P_a type of information from b, where $H(P_a)$ is the Shannon entropy (25) associated with the probabilities defined in (6).[3] One can also think of $H(P_a|b)$ as the *missing information* about P_a given the quantum system b, and it is a natural quantum analog of $H(P_a|Q_b) = H(P_a) - H(P_a : Q_b)$ [see (26)], where one

[3] Following [6], we use H for classical entropy and S for quantum entropy. For conditional entropy, we use H if the first argument is classical as in (41), and S if the first argument is more general (quantum) as in (30).

identifies $\chi(P_a, b)$ as a quantum analog of $H(P_a\!:\!Q_b)$.[4] In contrast to $S(a|b)$ defined in (30), $H(P_a|b)$ is non-negative (see Lemma 1); it equals the Shannon missing information $H(P_a)$ in the case when b provides no information about P_a, and it equals zero only when b perfectly contains the P_a information.

We remark that an alternative way of defining $H(P_a|b)$, similar to that employed in [7–9], is to introduce the quantum channel $\mathcal{E}_P$ from $ab \to eb$ defined by

$$\mathcal{E}_P(\rho_{ab}) = \sum_j |e_j\rangle\langle e_j| \otimes \mathrm{Tr}_a(P_{aj}\rho_{ab}), \tag{42}$$

where $\{|e_j\rangle\}$ is an orthonormal basis for an auxiliary system e. Then $H(P_a|b)$ is the von Neumann conditional entropy $S(e|b)$ of the state $\mathcal{E}_P(\rho_{ab})$.

Lemma 1. This lemma summarizes some useful properties of the $\chi(P_a, b)$ and $H(P_a|b)$ measures.

(i) For any ensemble $\{p_j, \rho_j\}$

$$\chi(\{p_j, \rho_j\}) = S(\sum_j p_j\rho_j) - \sum_j p_j S(\rho_j) \leqslant H(\{p_j\}), \tag{43}$$

with equality if and only if the ρ_j are mutually orthogonal.

(ii) Let P_a and Q_b be any two POVMs on a and b respectively, and ρ_{ab} any state on $\mathcal{H}_{ab}$. Then

$$H(P_a\!:\!Q_b) \leqslant \chi(P_a, b) \leqslant \min\{S(\rho_a), S(\rho_b), S(a\!:\!b)\}, \tag{44}$$

and hence by (41), (43), and (44),

$$0 \leqslant H(P_a|b) \leqslant H(P_a|Q_b). \tag{45}$$

□

Part (i) is from [6] (Theorem 11.10, p. 518). The left-hand-side of (44) is Holevo's bound (p. 531 of [6]), and the right-hand-side of (44) is similar to Proposition 1 of [22] though we prove it in Appendix A since we have explicitly inserted the bound on χ.

C. Entropy biases and coherent information

In addition to quantitative measures of information about one system present in another it is useful to have measures of *information differences*. In what follows we shall make use of two quantities of this type. When considering two systems b and c,

$$\Delta S_K(b, c) := S_K(\rho_b) - S_K(\rho_c), \tag{46}$$

is the *entropy bias*, while for information type P_a,

$$\Delta\chi_K(P_a; b, c) := \chi_K(P_a, b) - \chi_K(P_a, c) \tag{47}$$

is the *information bias*. Analogous quantities for the complementary channels $\mathcal{E}$ and $\mathcal{F}$ (to b and c respectively) arising from isometry V are:

$$\Delta S_K(\mathcal{E}, \mathcal{F}) := S_K(\Upsilon_b/d_a) - S_K(\Upsilon_c/d_a),$$
$$\Delta\chi_K(P_a; \mathcal{E}, \mathcal{F}) := \chi_K(P_a, \mathcal{E}) - \chi_K(P_a, \mathcal{F}). \tag{48}$$

Unlike our information measures these quantities can (obviously) be negative. When using the von Neumann entropy we omit the subscript K and denote these quantities, e.g., by $\Delta S(b, c)$ and $\Delta\chi(P_a; b, c)$.

The *coherent information* I_{coh} (Sec. 12.4.2 of [6]) is a particular instance of the entropy bias for the tripartite pure state $|\Omega\rangle$:

$$I_{\mathrm{coh}}(\rho_{a'}, \mathcal{E}) = \Delta S(b, c) \tag{49}$$

where, see the discussion in Sec. II B associated with (19), the quantum channel $\mathcal{E}$ corresponds to an isometry V which yields $|\Omega\rangle$ when applied to an entangled state $|\Phi\rangle$ chosen so that the partial trace of $|\Phi\rangle\langle\Phi|$ down to a' yields the density operator $\rho_{a'}$. The density operators ρ_b and ρ_c needed to define the entropy bias, (46), on the right side of (49) are the partial traces of $|\Omega\rangle\langle\Omega|$ down to systems b and c, respectively. It can also be seen more directly, for the maximally-mixed input state, that $I_{\mathrm{coh}}(I_{a'}/d_{a'}, \mathcal{E}) = \Delta S(\mathcal{E}, \mathcal{F})$.

Despite the connection in (49), the entropy bias in (46) seems more natural in the state or static point of view, which lacks the notion of inputs and outputs, than I_{coh}. The latter has always been thought of as a function of a trace-preserving superoperator $\mathcal{E}$ and an input state $\rho_{a'}$ to a channel, whereas the biases in (46) and (47) are simply functions of the tripartite state $|\Omega\rangle$, without making reference to how it may have been generated by the combination of an isometry and a partially-entangled state.

IV. BASIS INVARIANCE

We begin our discussion of how the amount of information about system a in some other system(s) depends on the type of information with two cases in which certain quantities are actually independent of type. In both of them, a pure-state pre-probability is assumed.

Theorem 2. Consider a bipartite system with a pure-state pre-probability $\rho_{ab} = |\Psi\rangle\langle\Psi|$. Let N be a rank-1 POVM on a, let w be an orthonormal basis (thus also a rank-1 POVM) on a, then

$$\chi_K(w, b) = \chi_K(N, b) = S_K(\rho_a) \tag{50}$$

is *independent* of the basis w or rank-1 POVM N.

[4] It is straightforward to show that $\chi(P_a, b)$ becomes $H(P_a\!:\!Q_b)$ if one replaces the conditional density operators ρ_{bj} in (36) with conditional probability distributions $\Pr(Q_b|P_a = P_{aj})$, and also replaces $S()$ with $H()$.

Proof. Apply (36) to w, setting $S_K(\rho_b) = S_K(\rho_a)$ and the second term in (36) to zero because each ρ_{bj} is a pure state, proving $\chi_K(w,b) = S_K(\rho_a)$. From Sec. II A, N is equivalent to an orthonormal basis v_{ae} on $\mathcal{H}_a \otimes \mathcal{H}_e$ assuming the state on ae is $\rho_{ae} = \rho_a \otimes |e_0\rangle\langle e_0|$, where $|e_0\rangle$ is some pure state on e. Thus, $\chi_K(N,b) = \chi_K(v_{ae},b) = S_K(\rho_{ae})$, but $S_K(\rho_{ae}) = S_K(\rho_a)$ for all entropy functions under consideration. □

This implies that if the w information about a is absent from b, $\chi_K(w,b) = 0$, *all* types are absent and $|\Psi\rangle$ is a product state, which is one form of the Absence theorem of [2]. And it generalizes in that if the w information is *almost* absent from b, then by (37) $\chi(w,b) \geqslant \chi(P,b)$, any other type P is *almost* absent from b. On the other hand, one can read (50) as a statement that all (rank-1) types of information are equally present; the only problem is interpreting the common value of $\chi_K(w,b) = S_K(\rho_a)$. In the case of the von Neumann entropy, $\chi(w,b) = S(\rho_a)$ is the usual entanglement measure of $|\Psi\rangle$, and is an upper bound on the Shannon mutual information (Lemma 1) that can be achieved by performing measurements in the Schmidt bases on a and b. Note that reading (50) in reverse provides a natural interpretation for $S_K(\rho_a)$; it is the amount of information about any rank-1 type N contained in a system b that purifies ρ_a, as measured by $\chi_K(N,b)$.

The following useful result for tripartite pure states and complementary channels (see Sec. VI A) is proved in Appendix B.

Theorem 3. Let M and N be rank-1 POVMs on a, and let v and w be orthonormal bases (thus also rank-1 POVMs) on a.

(i) Consider a tripartite system with a pure-state pre-probability $\rho_{abc} = |\Omega\rangle\langle\Omega|$. Then the information bias defined in (47),

$$\Delta\chi_K(w;b,c) = \Delta\chi_K(N;b,c) = \Delta S_K(b,c), \tag{51}$$

where K denotes any of the entropies defined in (27) or (28), is equal to the corresponding entropy bias, and thus *independent* of the choice of orthonormal basis or rank-1 POVM. It follows that the difference:

$$\chi_K(M,b) - \chi_K(N,b) = \chi_K(M,c) - \chi_K(N,c), \tag{52}$$

is the same for b and c, which obviously holds if M and N are replaced by v and w.

(ii) Likewise, for complementary quantum channels $\mathcal{E}$ and $\mathcal{F}$, the information bias defined in (48),

$$\Delta\chi_K(w;\mathcal{E},\mathcal{F}) = \Delta\chi_K(N;\mathcal{E},\mathcal{F}) = \Delta S_K(\mathcal{E},\mathcal{F}) \tag{53}$$

is invariant to the choice of orthonormal basis w or rank-1 POVM N, and

$$\chi_K(M,\mathcal{E}) - \chi_K(N,\mathcal{E}) = \chi_K(M,\mathcal{F}) - \chi_K(N,\mathcal{F}). \tag{54}$$

□

This theorem provides a natural interpretation for the entropy bias of a tripartite pure state: this is the amount by which more (or less if the bias is negative) w information about a is present in b than it is in c. The theorem tells us that this excess, which we call the information bias, does not depend upon the orthonormal basis w, allowing us to drop the w from $\Delta\chi_K(b,c)$ under these conditions. This theorem is used in proving several of the results that follow, including Theorems 8, 10, and 11.

Example 1. As an illustration, suppose that in the case of a qubit, $d_a = 2$, the z information associated with the standard $|0\rangle, |1\rangle$ basis is perfectly transmitted from a to b, while no information in the conjugate x basis is transmitted; i.e., we have a perfect "classical" channel from a to b. Setting $M = z$ and $N = x$ in (52) and using Lemma 1, $H(z) = \chi(z,c) - \chi(x,c)$, which can only be true if $\chi(x,c) = 0$ and $H(z) = \chi(z,c)$. The z information is thus perfectly transmitted from a to c, saying the "classical" information (in this sense) is always copied to another party, and further by the basis-invariance of $\Delta\chi(b,c) = 0$, that the ab and the ac channels are equally effective in terms of the χ measure.[5] This conclusion can be reached by alternative lines of argument, but it illustrates the nontrivial content of Theorem 3.

V. GENERALIZING ALL-OR-NOTHING THEOREMS

In this section we consider various quantitative generalizations, using the information measures introduced in Sec. III, of some "all-or-nothing" theorems [2], which have the general form that in a multipartite system if a particular type or types of information about a particular subsystem a is perfectly present or absent in some other subsystem, then some other types of information about a will also be perfectly present or absent in other locations. In each subsection below we provide a quantitative generalization of such a theorem to situations of partial presence or absence, indicating the connection with the all-or-nothing theorem if it is not already clear.

A. Truncation

The Truncation theorem of [2] states that if $\Pi = \{\Pi_j\}$ is a projective decomposition of I_a, and if the Π type of information about a is perfectly present in c, then for any third system b, the density operator ρ_{ab} is truncated or block-diagonal (or "pinched", p. 50 of [23]) in the sense that $\rho_{ab} = \sum_j \Pi_j \rho_{ab} \Pi_j$. The following result is a generalization of this theorem to the case of partial information presence in c, and also allows for more general

[5] An explicit example of this is the GHZ state $(|000\rangle + |111\rangle)/\sqrt{2}$.

POVMs P in Part (ii). The all-or-nothing result comes out by setting $H(\Pi|c) = 0$ (perfect information presence) in (55) below, which implies that $\rho_{ab} = \sum_j \Pi_j \rho_{ab} \Pi_j$ since $S(\rho||\sigma) = 0$ only if $\rho = \sigma$. More generally, ρ_{ab} will be "close" (in the relative entropy sense) to the truncated form if $H(\Pi|c)$ is small.

Lemma 4. Let $\Pi = \{\Pi_j\}$ be a projective decomposition of I_a and let $P = \{P_j\}$ be any POVM on a.

(i) Let ρ_{abc} be a pure state, then

$$H(\Pi|c) = S(\rho_{ab}|| \sum_j \Pi_j \rho_{ab} \Pi_j). \tag{55}$$

(ii) Let ρ_{abc} be *any* state, then

$$H(P|c) \geqslant S(\rho_{ab}|| \sum_j P_j \rho_{ab} P_j). \tag{56}$$

□

The proof can be found in Appendix C. This lemma is also used in proving the uncertainty relation in the next section.

B. Information exclusion relations

An exclusion relation refers to incompatible types of information such that the presence of one type in one subsystem "hinders" or to some extent "excludes" the incompatible type from being present in a *different* subsystem. Thus the Exclusion theorem of [2] asserts that if v and w are mutually unbiased bases on a, and the v information about a is perfectly present in b, then the w information about a is (completely) absent from c. A quantitative extension of this to partial presence and absence can be based on the following theorem, where the incompatibility of two POVMs $P = \{P_j\}$ and $Q = \{Q_k\}$ is quantified using:

$$r(P,Q) := \max_{j,k} \left\| \sqrt{P_j}\sqrt{Q_k} \right\|_\infty^2 . \tag{57}$$

Here $\|\cdot\|_\infty$ denotes the supremum norm: the maximum singular value of the operator.

Our main result, with proof in Appendix D, is:

Theorem 5. Let ρ_{abc} be any state on $\mathcal{H}_{abc}$.

(i) Let $P = \{P_j\}$ and $Q = \{Q_k\}$ be any two POVMs on $\mathcal{H}_a$, with $H(\cdot|\cdot)$ defined in (41) and r in (57). Then

$$H(P|b) + H(Q|c) \geqslant \log[1/r(P,Q)], \tag{58}$$

where each $H(\cdot|\cdot)$ term is bounded by, e.g.:

$$H(P|b) \geqslant \log[1/\sqrt{r(P,P)}]. \tag{59}$$

(ii) Specializing (58) to the case of orthonormal bases $v = \{|v_j\rangle\langle v_j|\}$ and $w = \{|w_k\rangle\langle w_k|\}$, we obtain:

$$H(v|b) + H(w|c) \geqslant \log[1/r(v,w)], \tag{60}$$

where in this case (57) reads

$$r(v,w) = \max_{j,k} |\langle v_j|w_k\rangle|^2. \tag{61}$$

(iii) The right-hand-side of (60) is largest when v and w are MUBs, $r(v,w) = 1/d_a$:

$$H(v|b) + H(w|c) \geqslant \log d_a. \tag{62}$$

□

We remark that (60) is equivalent to the main inequality conjectured in [8] and proven in [7], see Sec. VI B, and (58) was also recently proven in [9] using smooth entropies, an approach different from ours. Our proof approach is based on the relative entropy; we will go into more detail about this approach in a subsequent article [24].

It is useful to view the inequalities in Theorem 5 in two different ways, as information exclusion relations *and* as entropic uncertainty relations. The fact that they contain both principles can be seen, for example, in the MUB case by rewriting (62) as:

$$H(v) + H(w) \geqslant \chi(v,b) + \chi(w,c) + \log d_a. \tag{63}$$

Viewed from the left-hand-side it looks like an entropic uncertainty relation: a lower bound on an entropic sum. Viewed from the right-hand-side it looks like an information exclusion relation: an upper bound on an information sum. We note here that setting $H(v|b) = 0$ in (62) implies $H(w|c) = \log d_a$, the maximum value, and thus c contains no information about w, demonstrating that our result implies (and thus generalizes) the Exclusion theorem from [2].

As (60) was proven in [7], consider the following example illustrating how (58) goes beyond (60).

Example 2. Set Q to the w basis, and let P be a POVM composed of n pure states or rank-1 operators each with trace d_a/n and each of which is unbiased with respect to the w basis. Applying (58) gives

$$H(P|b) + H(w|c) \geqslant \log n. \tag{64}$$

Now suppose c contains all the w information, $H(w|c) = 0$. This implies that $H(P|b) = \log n$, which in turn implies *two* conditions, the probabilities of the P_j are equal, so there is maximal missing information about which P_j state system a is in, *and* the P information must be perfectly absent from b: $\chi(P,b) = 0$. The latter means that all states in P get mapped by (6) to the same output density operator ρ_{bj} on b. For example, for $d_a = 2$ consider setting w to the z basis (standard basis); then P could be the four states making up the x and y bases or three states forming an equilateral triangle in the xy plane of the Bloch sphere or any symmetric set of states in the xy plane. Imagining P to be composed of a very large number of states in the xy plane, by continuity *all* states in the xy plane must get mapped to the same output density operator ρ_{bj} on b when $H(z|c) = 0$; a result that does not come out of pairing z with a particular MUB, say x, and using (60). This all-or-nothing result is implied by the Truncation theorem of [2], but (64) also describes the partial information case, saying that the ρ_{bj} associated with P must be fairly indistinguishable if $H(w|c)$ is small.

Inspired by (and strengthening) a result in [25], Eq. (59) is, in some sense, an uncertainty relation for a *single* POVM. Rewriting it as

$$H(P) \geqslant \chi(P,b) + \log[1/\sqrt{r(P,P)}], \tag{65}$$

it strengthens the bound $H(P) \geqslant \chi(P,b)$ from Lemma 1; stating that if the P measurement outcome is fairly certain [$H(P)$ small], this can partially exclude the P information from another system b [$\chi(P,b)$ small]. The idea is that a POVM is generally not a set of mutually-exclusive properties (Sec. II A) so it has some intrinsic incompatibility, as measured by $\log[1/\sqrt{r(P,P)}]$. For example, if P is composed of n rank-1 operators each with trace d_a/n, then $\log[1/\sqrt{r(P,P)}] = \log(n/d_a)$.

Some information exclusion relations below for quantum channels are proven in Appendix E. Although they follow from Theorem 5, they bring to mind a slightly different picture [16], as one imagines Alice sending "incompatible ensembles" P and Q respectively through $\mathcal{E}$ and $\mathcal{F}$, and if the $\mathcal{F}$ channel transmits the Q ensemble well to Carol, then the $\mathcal{E}$ channel must be constructed in such a way that at its output Bob will have difficultly discerning which member of the P ensemble Alice sends.

Corollary 6. For complementary quantum channels $\mathcal{E}$ and $\mathcal{F}$, χ given by (39),

(i) Let P and Q be any two POVMs, with $H(P) = H(\{p_j\})$ where p_j is given by (40) and likewise for $H(Q)$,

$$\chi(P,\mathcal{E}) \leqslant H(P) - \log[1/\sqrt{r(P,P)}], \tag{66}$$

$$\chi(P,\mathcal{E}) + \chi(Q,\mathcal{F}) \leqslant H(P) + H(Q) - \log[1/r(P,Q)]. \tag{67}$$

(ii) For orthonormal bases v and w,

$$\chi(v,\mathcal{E}) + \chi(w,\mathcal{F}) \leqslant \log[d_a^2 r(v,w)]. \tag{68}$$

(iii) For MUBs v and w,

$$\chi(v,\mathcal{E}) + \chi(w,\mathcal{F}) \leqslant \log d_a. \tag{69}$$

□

As another corollary to Theorem 5, some uncertainty relations for a single system [25, 26] (see Sec. VI B) can be strengthened for mixed states, with the proof in Appendix F.

Corollary 7.

(i) For any state ρ, let N be a rank-1 POVM and let P be any POVM, then

$$H(N) \geqslant \log[1/\sqrt{r(N,N)}] + S(\rho), \tag{70}$$

$$H(N) + H(P) \geqslant \log[1/r(N,P)] + S(\rho). \tag{71}$$

(ii) For any state ρ of a qubit (dimension $d = 2$) and any complete set of three MUBs x, y, and z:

$$H(x) + H(y) + H(z) \geqslant 2\log 2 + S(\rho). \tag{72}$$

□

While one might conjecture that (70) or (71) generalizes to the case where N is an arbitrary POVM, it is easy to see that this is false. Imagine a highly mixed state such that $S(\rho)$ is very large, yet N and P are composed of coarse-grained projectors with very high rank, so $H(N)$ and $H(P)$ would be small, violating the inequality.

Note that (72) is a tight bound, achieved for example when the state is along the z-axis of the Bloch sphere, such that $H(x) = H(y) = \log 2$ and $H(z) = S(\rho_a)$.

C. Suppression of differences

The following is a bipartite result, proved in Appendix G, saying that the presence of some type of information P about a in b *suppresses the difference* in the presence of two other types of information, M and N, about a in b. Note that a similar result holds for quantum channels.

Theorem 8. Let ρ_{ab} be any state,

(i) For any POVM P on a; rank-1 POVMs M and N on a,

$$\begin{aligned}
&|\chi(M,b) - \chi(N,b)| \leqslant H(P|b) + \\
&\quad \max\{H(M) - \log[1/r(P,M)], H(N) - \log[1/r(P,N)]\}, \\
&|H(M|b) - H(N|b)| \leqslant H(P|b) + \\
&\quad \max\{H(M) - \log[1/r(P,N)], H(N) - \log[1/r(P,M)]\}.
\end{aligned} \tag{73}$$

(ii) For orthonormal bases u, v, w on a, with u and v each MU with respect to w:

$$\begin{aligned}
|\chi(u,b) - \chi(v,b)| &\leqslant H(w|b), \\
|H(u|b) - H(v|b)| &\leqslant H(w|b).
\end{aligned} \tag{74}$$

(iii) Let u, v, w be as in (ii), but in addition assume that the w type is perfectly present in b, then

$$\begin{aligned}
\chi_K(u,b) = \chi_K(v,b) &= S_K(\rho_b) - S_K(\rho_{ab}), \\
H(u|b) = H(v|b) &= \log d_a + S(a|b),
\end{aligned} \tag{75}$$

meaning that all types MU to w are present *to the same degree* in b, in this sense. □

The difference suppression effect is most apparent in part (ii) of this theorem, where (74) says that the presence in b of the w information forces types MU to the w type to be equally present in b, in the sense of having the same χ and H quantities. As an illustration, consider $d_a = 2$, let the z information about a be perfectly present in b, then all types in the xy plane of the Bloch sphere are present in b to the same degree, bringing to mind the image of a prolate spheroid (American football), with z being the major axis, for the information about a present in b.

D. Decoupling theorems

The preceding results can be used to generalize of some all-or-nothing decoupling theorems, which provide sufficient conditions about the information content of b and/or c to ensure that c is completely uncorrelated to or decoupled from a. For example, the No Splitting theorem of [2] states that if all types of information about a are perfectly present in b, then all types of information about a are perfectly absent from c. (The name is motivated by the idea that a perfect quantum channel from a to b allows no diversion or split off of information to a third location c.) In our notation this corresponds to the assertion that when $H(w|b) = 0$ for every orthonormal basis w of $\mathcal{H}_a$, then $\chi(w, c) = 0$ for every such basis. What follows is a quantitative generalization, a corollary of Theorem 5, with the No Splitting theorem the special case when $\alpha = 0$.

Corollary 9. Let α be some positive constant.

(i) For any state ρ_{abc}, if $H(w|b) \leqslant \alpha$ for every orthonormal basis w of $\mathcal{H}_a$, then $\chi(w, c) \leqslant \alpha$ for every such basis.

(ii) For complementary quantum channels $\mathcal{E}$ and $\mathcal{F}$, if $\chi(w, \mathcal{E}) \geqslant \log d_a - \alpha$ for every orthonormal basis w of the channel input, then $\chi(w, \mathcal{F}) \leqslant \alpha$ for every such basis.

Proof. For any orthonormal basis w there is a MU basis v, and thus (62) implies that $H(w|c) \geqslant \log d_a - \alpha$ and hence, because $H(w)$ cannot exceed $\log d_a$, $\chi(w, c)$ cannot be greater than α. The channel version follows by the same argument using (69). □

The Presence theorem of [2] states that if two *strongly incompatible* types of information about a are perfectly present in b, then all information about a is perfectly present in b, which is to say there is a perfect quantum channel from a to b. Unfortunately, "strongly incompatible" is a complicated concept, and it is not obvious how to extend it to a quantitative measure in the general case. Instead, we consider two POVMs N and P, and the case where they are MUBs implies they are strongly incompatible types of information. The following theorem, proved in Appendix H, combines the notions of "presence" and "no splitting", and gradually specializes from POVMs to orthonormal bases to MUBs. Note that part (ii) of this theorem is stated for channels to remind the reader that each of our results for states has some analogous formulation for channels.[6]

Theorem 10. For any POVM P on a; rank-1 POVMs M and N on a; orthonormal bases u, v, w on a;

(i) For any bipartite state ρ_{ab}

$$H(N|b) + H(P|b) \geqslant \log[1/r(N, P)] + S(a|b), \quad (76)$$

where, for any tripartite state ρ_{abc},

$$\chi(M, b) \geqslant [-S(a|b)], \text{ and } H(M|c) \geqslant [-S(a|b)]. \quad (77)$$

(ii) For complementary quantum channels $\mathcal{E}$ and $\mathcal{F}$,

$$\chi(u, \mathcal{E}) \geqslant \Delta\chi(\mathcal{E}, \mathcal{F}) \geqslant \chi(v, \mathcal{E}) + \chi(w, \mathcal{E}) - \log[d_a^2 r(v, w)], \quad (78)$$

$$\chi(u, \mathcal{F}) \leqslant \log[d_a^3 r(v, w)] - [\chi(v, \mathcal{E}) + \chi(w, \mathcal{E})]. \quad (79)$$

(iii) For MUBs v and w,

$$S(a:b) \geqslant 2\log d_a - 2[H(v|b) + H(w|b)], \quad (80)$$

$$S(a:c) \leqslant H(v|b) + H(w|b). \quad (81)$$

□

Corollary 9 gave a condition to guarantee that no information is present in c, and it is that all information is present in b. But what part (iii) of Theorem 10 shows is that one need not check that every single type of information is present in b; rather, simply check that b contains *two types* that are MUBs and this will completely decouple c from a [8]. Part (ii) emphasizes that the information in c (transmitted by $\mathcal{F}$) can be upper bounded and that in b (transmitted by $\mathcal{E}$) lower bounded even when the two bases are not MUBs. Part (i) generalizes this notion further to POVMs. By (76) one can lower-bound $[-S(a|b)]$, some measure of the entanglement between a and b, just by knowing that b contains information about a rank-1 POVM on a and an arbitrary POVM on a. By (77) this serves to lower-bound both the M information in b and the M information *missing* from c, for *any* rank-1 POVM M. The application of such a relation, specialized to orthonormal bases, to quantum cryptography was discussed previously in [7], and the generalization to POVMs might turn out to be useful.

There is a seemingly odd restriction in (76) that either N or P must be composed of rank-1 elements. One might conjecture that (76) holds for arbitrary POVMs, but this is false. One can see this by choosing $\rho_{ab} = \rho_a \otimes \rho_b$ in which case (76) reduces to (71). As discussed following Corollary 7, (71) could be violated dramatically if $S(\rho_a)$ was large but both N and P were composed of high-rank projectors.

The following decoupling theorem considers the situation where some type of information w is both perfectly present in b and absent from c. It shows that this simple condition strikingly is enough to *completely decouple* c from a, and furthermore, for pure states, it leads to the suppression of differences between *all* types of information in b. The theorem gradually specializes from all states to pure states to channels, with the proof in Appendix I.

Theorem 11. Let L, M, N be rank-1 POVMs and P be any POVM on a; let v and w be orthonormal bases on a,

(i) Let ρ_{abc} be any state, then

$$S(a:c) \leqslant \chi(N, c) + H(N|b) - \log[1/\sqrt{r(N, N)}]. \quad (82)$$

---

[6] When one obtains an upper bound on a χ quantity for a channel $\mathcal{F}$, as in Theorems 6, 9, 10, and 11, this bound also holds if one composes *any* channel $\mathcal{G}$ with $\mathcal{F}$, i.e. $\chi(w, \mathcal{G} \circ \mathcal{F}) \leqslant \chi(w, \mathcal{F})$ by (38), which is useful if one is interested in bounding information in a *subsystem* of the output of $\mathcal{F}$.

If the N type of information is perfectly present in b, then

$$\chi_K(P, c) \leqslant \chi_K(N, c). \tag{83}$$

If, in addition, the N type of information is absent from c, then *all* types of information about a are absent from c, i.e. a and c are completely uncorrelated: $\rho_{ac} = \rho_a \otimes \rho_c$.

(ii) In the special case of pure states $\rho_{abc} = |\Omega\rangle\langle\Omega|$,

$$|\chi(L, b) - \chi(M, b)| \leqslant \chi(N, c) + H(N|b) - \log[1/\sqrt{r(N, N)}] \tag{84}$$

and thus, in the extreme case where the N type of information is perfectly present in b and absent from c,

$$\chi(L, b) = \chi(M, b) = S(\rho_a) \tag{85}$$

is *independent* of rank-1 POVM (or orthonormal basis).

(iii) For complementary channels $\mathcal{E}$ and $\mathcal{F}$,

$$\begin{aligned} \chi(v, \mathcal{F}) &\leqslant \chi(w, \mathcal{F}) + [\log d_a - \chi(w, \mathcal{E})], \\ \chi(v, \mathcal{E}) &\geqslant \chi(w, \mathcal{E}) - \chi(w, \mathcal{F}). \end{aligned} \tag{86}$$

Thus, if the w type of information is perfectly present in the $\mathcal{E}$ channel and absent from the $\mathcal{F}$ channel, the same is true for *all* types of information. This is a necessary and sufficient condition for $\mathcal{E}$ being a perfect quantum channel and $\mathcal{F}$ being a completely noisy channel. □

VI. CONNECTION WITH OTHER WORK

A. Difference of Holevo quantities

Schumacher and Westmoreland [27] remarked that a difference in χ quantities associated with sending an ensemble of pure states through complementary quantum channels depends only on the average density operator of the input ensemble. This situation is equivalent to the one considered in (51) of Theorem 3, where a rank-1 POVM N acts on system a of a tripartite pure state $|\Omega\rangle$. The equivalence follows from the discussion in Sec. II B; decompose $|\Omega\rangle$ into an isometry V acting on half of a bipartite pure state $|\Phi\rangle$, as in (7) and Fig. 1, then by the construction in (22), any pure-state ensemble at the input of V can be produced by an appropriate choice of N and $|\Phi\rangle$. Despite this equivalence, the notion of basis invariance or invariance to the rank-1 POVM N emerges naturally out of the state view, since the average density operator of the input ensemble to V is unaffected by choice of N. If one is willing to restrict to inputting the maximally-mixed average density operator, then the basis-invariance emerges in the channel view as well, as in (53).

B. Entropic uncertainty relations

Our inequalities are related to several entropic uncertainty relations in the literature (see [28] for a recent review), which are translated below into our notation. Maassen and Uffink [29] proved an entropic uncertainty relation for measurements in orthonormal bases v and w on system a for any state ρ_a:

$$H(v) + H(w) \geqslant \log[1/r(v, w)]. \tag{87}$$

Krishna and Parthasarathy [25] generalized this to POVMs P and Q,

$$H(P) + H(Q) \geqslant \log[1/r(P, Q)], \tag{88}$$

and also stated an uncertainty relation for a single POVM

$$H(P) \geqslant \log[1/\sqrt{r(P, P)}]. \tag{89}$$

Hall [30] incorporated into (87) the idea of "classical" side information, i.e. information about the outcome of a POVM X_e acting on a system e that may be correlated to a:

$$H(v) + H(w) \geqslant \log[1/r(v, w)] + H(v : X_e) + H(w : X_e). \tag{90}$$

Considering e to be a composite system bc and $X_e = Q_b \otimes R_c$ a composite POVM, it follows from (90) that:

$$H(v|Q_b) + H(w|R_c) \geqslant \log[1/r(v, w)], \tag{91}$$

see the discussion in [8] where (91) was termed the weak complementary information tradeoff and was ascribed to Cerf et al. [31].

The inequalities in Theorem 5: (58), (59), and (60) respectively strengthen (88), (89), and (87) by allowing for quantum side information, for example, information about property P contained in another quantum system b, as measured by $\chi(P, b)$. The presence of such χ quantities, reducing the left-hand-sides of the Theorem 5 inequalities, is precisely what strengthens these bounds. Equation (90) follows from (87) [and thus (60)] by an argument that can be found in [8, 30]. Equation (91) follows from (60) using the Holevo bound (45), $H(v|Q_b) \geqslant H(v|b)$ and $H(w|R_c) \geqslant H(w|c)$.

Equation (60) is precisely the "strong complementary information tradeoff" conjectured by Renes and Boileau [8] and later proven by Berta et al. [7]. It is straightforward to show that our definition of $H(v|b)$ in (41) is equivalent the definition employed in [8] and [7], see (42).

The main inequality in Berta et al. [7],

$$H(v|b) + H(w|b) \geqslant \log[1/r(v, w)] + S(a|b) \tag{92}$$

was formulated for orthonormal bases v and w, and we generalized it to POVMs (with at least one POVM being rank-1) in (76). Also, (92) is equivalent to (60) as follows. Apply (60) to a pure state ρ_{abc} and use $H(w|b) = H(w|c) - \Delta\chi(b, c)$ with $S(a|b) = -\Delta\chi(b, c)$ to get (92). Conversely, starting from (92), follow the reverse process to prove (60) for pure states ρ_{abc}, and then (60) for mixed states follows from (38). Thus, since (60) is generalized to *two* arbitrary POVMs by (58), (58) and

(76) provide two alternative generalizations of (92). To prove (92), Berta et al. first proved an uncertainty relation involving smooth minimum and maximum entropies, and then invoked a lemma that these entropies approach the desired von Neumann entropic quantities under an appropriate asymptotic limit. In contrast, our proof does not use smooth entropies, but invokes the monotonicity of the relative entropy under quantum operations, so the approaches are conceptually different.

Christandl and Winter [16] derived an information exclusion relation for quantum channels, which can be rearranged and expressed in our notation to read:

$$\chi(x,\mathcal{E})+\chi(z,\mathcal{F})\leqslant \log d_a, \tag{93}$$

where x and z are orthonormal bases related to each other by the d-dimensional quantum Fourier transform, and $\mathcal{E}$ and $\mathcal{F}$ are complementary quantum channels. Equation (69) of Corollary 6 generalizes this to arbitrary MUBs, and (67) further generalizes to input ensembles associated with POVMs.

Our results strengthen some uncertainty relations in the case of mixed states. In the special case where N is a rank-1 POVM, (70) and (71) respectively strengthen (89) and (88) with the addition of the $S(\rho_a)$ term. Sánchez-Ruiz [26] proved an entropic uncertainty relation for sets of d_a+1 MUBs, which when applied to qubits ($d_a=2$) gives:

$$H(x)+H(y)+H(z)\geqslant 2\log 2. \tag{94}$$

Likewise this is strengthened for mixed states by (72). Bounds depending on the purity of ρ_a were also given in [26]; in the qubit case these bounds are implied by (72).

C. No Splitting and Decoupling

Kretschmann et al. [32] have studied the degree to which a channel is error-correctable using a diamond-norm measure, and showed that when a channel is nearly perfect (in this sense) its complementary channel transmits very little information, and vice versa. Bény and Oreshkov [33] formulated a similar theorem for complementary channels, but in a general, symmetric fashion, using a fidelity measure. Hayden and Winter [34] have studied the degree to which a channel preserves the distinguishability of input states, and formulated the tradeoff in geometry-preservation between complementary channels using a trace-distance measure. Each of these formulations generalize the No Splitting principle (see Sec. V D), although their information measures are of a different nature from the one we employ, and the connection between our approach and theirs remains to be determined. Intuitively, the No Splitting theorem should also be related to the notion that entanglement is monogamous. Quantitative expressions of entanglement monogamy have been found in terms of the concurrence and the squashed entanglement [5]; as these are "global" measures of correlation, their relation to our information-type-specific measure is not obvious.

Renes and Boileau [8] formulated a decoupling theorem as a corollary to their conjectured uncertainty relation [Eq. (60)], stating that if b contains the information about two sufficiently incompatible orthonormal bases of a, then the coupling of c to a can be upper-bounded. This is quite similar to our Theorem 10, which extends this notion to two sufficiently incompatible POVMs.

VII. CONCLUSIONS

A. Summary

Since our technical results in Secs. IV and V involve a large number of theorems, the following comments are intended to assist the reader in seeing how they are related to one another and to the definitions given earlier in Secs. II and III.

In Sec. II A we generalize an earlier [2] notion of types of quantum information to include general POVMs on a Hilbert space $\mathcal{H}_a$ for system a, by noting that the associated probabilities are the same as those for a projective decomposition of the identity on a larger Hilbert space $\mathcal{H}_A$, the Naimark extension, and a rank-1 POVM corresponds to an orthonormal basis of the extension. Various measures for different types of information are introduced and discussed in Sec. III. For uniformity of notation, Shannon entropies and related quantities are denoted by $H()$; e.g., $H(P_a)$ is the missing information about type P_a, as determined by its probability distribution, when the quantum state is assumed known. For quantum entropies we use $S()$ for the von Neumann entropy, and $S_K()$, where K can be R or T or Q for Renyi, Tsallis, and quadratic entropies, respectively.

We use the Holevo function $\chi(P_a,b)$, or $\chi_K(P_a,b)$ for S_K, (36), as a measure of the amount of information of type P_a about system a which is present in system b, along with the complementary quantity $H(P_a|b)$, (41), as a corresponding measure of the amount of information about P_a that is still missing given system b. While the analogy is not exact, $\chi(P_a,b)$ is similar to Shannon's mutual information $H(P_a:Q_b)$, whereas $H(P_a|b)$ resembles Shannon's conditional entropy $H(P_a|Q_b)$. In particular, $H(P_a|b)$, like $H(P_a|Q_b)$, is nonnegative, so retains some of the intuition of the latter quantity, in contrast to the quantum conditional entropy $S(a|b)$, (30), which can be of either sign. We use the term information bias for the difference between the amount of type P_a information about a in b and the amount in c, $\chi(P_a,b)-\chi(P_a,c)=\Delta\chi(P_a;b,c)$, which can have either sign. Similarly, we refer to $\Delta S(b,c)=S(\rho_b)-S(\rho_c)$ as the entropy bias, and add a subscript K when using an alternative to the von Neumann entropy. Our most extensive results are for the von Neumann entropy and its associated information measures. However, in some cases, see Theorems 2, 3, 8, and 11, these results also hold for a more general χ_K, and stating them in this form seems worthwhile, as for certain purposes these other measures could be useful.

While the most natural and symmetrical, in terms of treating the different parts on the same footing, formulation of our results is in terms of a tripartite system, some of the more interesting and significant applications are to quantum channels and complementary channels. The relationship between the tripartite and the channel perspectives is worked out in some detail in Sec. II B, and in Sec. III C we relate the coherent information for a quantum channel to a corresponding tripartite entropy bias. In several theorems the channel counterparts of tripartite results are stated separately, because while the formal results are in some sense the same, one's intuition about their significance can be different.

Our first set of results are the equalities in Theorems 2 and 3 of Sec. IV, which apply for pure quantum states of bipartite and tripartite systems, respectively. The first says that the amount of information about a in b is independent of the type of information, provided the latter is a rank-1 POVM; this includes an orthonormal basis. The second says that the difference between the amount of information concerning such a rank-1 POVM in b and in c is independent of the type considered, and equal to the corresponding entropy bias. Equivalently, given two rank-1 POVMs M and N, the difference between the amount of M and N information about a found in b is the same as the corresponding difference in c. While these results are limited to pure states, they are important for the proofs of many of the later results. They also extend from von Neumann to other quantum entropies, so they are stated in this more general form.

Perhaps the simplest way of viewing the collection of inequalities that make up Sec. V is that the main theorems are quantitative generalizations of all-or-nothing theorems which can be stated quite concisely for types of information associated with orthonormal bases v and w of system a. A central result of this paper is Theorem 5, and part (iii) of this theorem tells us that if the v information about a is perfectly present in b, which is to say $H(v|b) = 0$, then the mutually unbiased (MU) w type of information must be perfectly absent from c: $H(w|c) = \log d_a$ means that $\chi(w, c) = 0$. Part (ii) allows for bases that are not MU at the cost of a weaker bound on the H measures, while part (i) is not restricted to bases but applies to quite general types of information P and Q. The generalization to POVMs is, in turn, based on Lemma 4, which itself generalizes the Truncation theorem [2]: if the $v = \{v_j\}$ information is perfectly present in c then ρ_{ab} commutes with the v_j projectors.

The connections of Theorem 5 to literature entropic uncertainty relations are given in Sec. VI B. Broadly speaking we think that the addition of quantum side information to uncertainty relations [7, 8] not only strengthens certain bounds but also gives further conceptual insight into the nature of complementarity, in that side information about complementary observables in different locations (Sec. V B) must be constrained as well. We also note a recent experimental study [35]. Further remarks on the significance of Theorem 5 can be found in the discussion that follows it in Sec. V B.

Corollary 6 of Theorem 5 gives the corresponding result for quantum channels, generalizing to partial information and to arbitrary POVMs or orthonormal bases the all-or-nothing theorem: if the v information is perfectly present in (or transmitted by) the $\mathcal{E}$ channel, any MU type of information w must be absent from (or destroyed by) the complementary channel $\mathcal{F}$. In addition, Corollary 7 of Theorem 5 provides strengthened information inequalities for a single system described by a mixed state.

The idea behind Theorem 8 is encapsulated in the observation that if the information about an orthonormal basis w of a is perfectly present in b, so $H(w|b) = 0$, and u and v are bases of a that are MU with respect to w (but not necessarily with respect to each other) then the u and v types are present in b in equal amounts. If, on the other hand the w information is less than perfectly present in b, this theorem provides quantitative bounds on the difference between the u and v types of information in b. Similarly, the requirement that u and v be MU relative to w can be relaxed, and they can even be replaced with rank-1 POVMs, and w with a general POVM, see part (i) of the theorem, at the price of appropriately weakening the bounds that confine the differences.

The results in Sec. V D provide quantitative generalizations of conditions that ensure system c is completely uncorrelated to (or decoupled from) system a, $\rho_{ac} = \rho_a \otimes \rho_c$. Corollary 9 of Theorem 5 says that the correlations between a and c are tightly upper-bounded if system b *almost* perfectly contains all types of information about a, and gives the analogous result for complementary channels $\mathcal{E}$ and $\mathcal{F}$. But Theorem 10 stresses the importance of the presence of *just two* (sufficiently incompatible) types of information. That is, if b perfectly contains the information about two MUBs of a, then b contains *all* types of information about a, and c is completely uncorrelated to a; a generalization of this statement for the partial information case is given in part (iii) of Theorem 10. Parts (ii) and (i) of this theorem respectively illustrate that this idea can be extended, at the price of weakened bounds, to any two orthonormal bases or to two POVMs in which at least one of the POVMs is rank-1. The relevance of inequalities like (76) of Theorem 10, where the presence of two types of information about a in b can be used to upper bound the information about a in c, to quantum cryptography was discussed in [7].

The same sort of decoupling occurs when a single type of information about a associated with an orthonormal basis w is perfectly present in b and completely absent from c. Theorem 11 contains this interesting result together with certain quantitative generalizations, both when the type of information in question is only partially absent from c, and when it is not perfectly present in b.

B. Future outlook

There are various ways in which the results summarized above suggest problems which deserve further attention and research. One has to do with the difference between rank-1 and higher-rank POVMs, or orthonormal bases as against coarser projective decompositions of the identity. In a number of cases the results we have obtained for the former are distinctly stronger than for the latter, but the reason for this is not always clear. Since applications of quantum information theory to macroscopic systems, in particular to problems of decoherence, lead rather naturally to coarse decompositions or POVMs, a good intuitive understanding in addition to formal expressions would be of value. A second item concerns the use of the $r(P,Q)$ overlap measure for POVMs, or its $r(v,w)$ counterpart for orthonormal bases, see (57) and (61). While this provides the basis of significant inequalities in Theorem 5 and later, the fact that $r(P,Q)$ requires one to maximize over all pairs of elements from the two POVMs hints that stronger results might well be possible were one to use a more refined perspective on how the POVMs are related to each other, or the sorts of information that they provide.

While qualitative inequalities are certainly an advance over simple all-or-nothing results, it would be even better if one could express information tradeoffs in terms of equalities of the sort which could conceivably allow one to completely characterize how a quantum channel is related to its complementary channel using a (hopefully small) number of parameters with a clear intuitive significance. The equalities in Theorem 3, as applied either to channels or, more generally, pure-state tripartite systems, hint that something like this might be possible, but thus far we have not found it.

Any advance in understanding tripartite systems raises an obvious question: what about systems with four (or more) parts? It is, of course, possible to study them by thinking of two of the parts as constituting a single object, and then applying results for tripartite systems. But there is probably some "residual" aspect of a system of four parts which cannot be captured in this way, just as there are residual aspects of tripartite systems which cannot be understood simply in terms of combining two of them so as to yield a bipartite system. We think that our results in this paper have helped to clarify some of this tripartite residual, and we hope they provide hints on ways to deal with more complicated cases.

Acknowledgments

We thank Michael Zwolak, Luc Tartar, Edward Gerjuoy, Shiang-Yong Looi, and Danquynh Nguyen for helpful discussions. We are especially grateful to an anonymous referee for pointing out a serious error in an earlier version of this paper. The research described here was supported by the Office of Naval Research and by the National Science Foundation through Grant No. PHY-0757251.

Appendix A: Proof of Lemma 1

Proof. (ii) The inequality $\chi(P_a,b)\leqslant S(\rho_b)$ obviously follows from (36). Now let c be a system that purifies ρ_{ab}. Then by (38), $\chi(P_a,b)\leqslant\chi(P_a,bc)=S(\rho_a)-\sum_j p_j S(\rho_{bcj})\leqslant S(\rho_a)$.

To prove $S(a:b)\geqslant\chi(P_a,b)$, as in Sec. II A think of P_a as a *projective* measurement $\widetilde{w}_{ae}$ on system ae, where $\widetilde{w}_{ae}$ is a coarse graining of some orthonormal basis (rank-1 projectors) w_{ae}. Let c purify ρ_{ab} such that $\rho_{abce}=\rho_{abc}\otimes|e_0\rangle\langle e_0|$ is a pure state. Then, $S(a:b)=S(\rho_{ae})+S(\rho_b)-S(\rho_c)=\chi(w_{ae},bc)+\chi(w_{ae},b)-\chi(w_{ae},c)\geqslant\chi(w_{ae},b)\geqslant\chi(\widetilde{w}_{ae},b)=\chi(P_a,b)$, by the Theorems in Sec. IV, by (38), and by (37). $\square$

Appendix B: Proof of Theorem 3

Proof. (i) For orthonormal basis $w=\{|w_j\rangle\}$, insert (36) into (47) to obtain

$$\begin{aligned}&\chi_K(w,b)-\chi_K(w,c)\\&\quad=S_K(\rho_b)-S_K(\rho_c)-\sum_j p_j[S_K(\rho_{bj})-S_K(\rho_{cj})].\end{aligned}\tag{B1}$$

The final term vanishes, for the following reason. Write $|\Omega\rangle=\sum_j|w_j\rangle\otimes|s_j\rangle$ in the form (10) with $|w_j\rangle$ replacing $|a_j\rangle$, so from (6) the conditional density operators in (B1) are given by

$$p_j\rho_{bj}=\mathrm{Tr}_c\Big(|s_j\rangle\langle s_j|\Big),\quad p_j\rho_{cj}=\mathrm{Tr}_b\Big(|s_j\rangle\langle s_j|\Big).\tag{B2}$$

Since $|s_j\rangle$ is a pure state the partial traces ρ_{bj} and ρ_{cj} have the same eigenvalues (determined by the Schmidt expansion coefficients of $|s_j\rangle$), except one may have more zeros than the other if $d_b\neq d_c$. Since $S_K(\rho)$ is a function only of the nonzero (positive) eigenvalues of ρ, each term in the final sum in (B1) vanishes, and we are left with (51). The generalization to rank-1 POVMs follows by the equivalence of N to an orthonormal basis v_A on $\mathcal{H}_A$, the Naimark extension of $\mathcal{H}_a$ as in Sec. II A. Since ρ_{Abc} is a tripartite pure state, then $\Delta\chi_K(N;b,c)=\Delta\chi_K(v_A;b,c)=S_K(\rho_b)-S_K(\rho_c)$.

(ii) Equation (53) follows from (51) by applying it to a channel ket $|\Omega\rangle$ constructed from V by (7). Alternatively, it can be proven directly from (39) and (48), obtaining an equation similar to (B1),

$$\begin{aligned}\Delta\chi_K(P;\mathcal{E},\mathcal{F})=&S_K(\Upsilon_b/d_a)-S_K(\Upsilon_c/d_a)\\&+\sum_j p_j[S_K(\rho_{bj})-S_K(\rho_{cj})],\end{aligned}\tag{B3}$$

where the final term vanishes again since $\rho_{bj}=\mathrm{Tr}_c[V\rho_{aj}V^\dagger]$ and $\rho_{cj}=\mathrm{Tr}_b[V\rho_{aj}V^\dagger]$ have the same (nonzero) spectrum, as the ρ_{aj} in (40) are rank-1 operators.

Equations (52) and (54) follow immediately from (51) and (53), respectively. □

Appendix C: Proof of Lemma 4

Proof. (i)

$$S(\rho_{ab}||\sum_j \Pi_j\rho_{ab}\Pi_j)$$
$$= -S(\rho_{ab}) - \mathrm{Tr}[\rho_{ab}\log(\sum_j \Pi_j\rho_{ab}\Pi_j)] \quad \text{(C1)}$$
$$= -S(\rho_c) - \mathrm{Tr}[\rho_{ab}\sum_k \Pi_k \log(\sum_j \Pi_j\rho_{ab}\Pi_j)\sum_l \Pi_l] \quad \text{(C2)}$$
$$= -S(\rho_c) - \mathrm{Tr}[\sum_k \Pi_k\rho_{ab}\Pi_k \log(\sum_j \Pi_j\rho_{ab}\Pi_j)]$$
$$- \sum_{k,l\neq k} \mathrm{Tr}[\rho_{ab}\Pi_k \log(\sum_j \Pi_j\rho_{ab}\Pi_j)\Pi_l] \quad \text{(C3)}$$
$$= -S(\rho_c) + S(\sum_j \Pi_j\rho_{ab}\Pi_j) \quad \text{(C4)}$$
$$= -S(\rho_c) + H(\Pi) + \sum_j p_j S(\rho_{abj}) \quad \text{(C5)}$$
$$= H(\Pi) - \chi(\Pi, c) = H(\Pi|c), \quad \text{(C6)}$$

where $p_j = \mathrm{Tr}(\Pi_j\rho_{ab})$ and $p_j\rho_{abj} = \Pi_j\rho_{ab}\Pi_j$. The last term in (C3) disappears because $\log(\sum_j \Pi_j\rho_{ab}\Pi_j)$ is block diagonal with respect to the Π_j projectors, and then one takes an off-diagonal element of it. Step (C5) follows from Lemma 1, part (i).

(ii) For clarity, we include the subscript a on the POVM P_a. Think of $P_a = \{P_{aj}\}$ as a projective measurement $\Pi_A = \{\Pi_{Aj}\}$ on an extended Hilbert space $\mathcal{H}_A$ (Naimark extension), with $\mathcal{H}_a$ a subspace and E_a the projector onto this subspace, and $P_{aj} = E_a\Pi_{Aj}E_a$. The state ρ_{Ab} is the same as ρ_{ab} but now just embedded in a larger space, that is: $\rho_{Ab} = E_a\rho_{Ab}E_a = \rho_{ab}$. Let $E_a^\perp$ be the projector onto the orthogonal complement of $\mathcal{H}_a$, note $E_a^\perp\rho_{Ab}E_a^\perp = 0$, and let the channel $\mathcal{F}$ be defined by $\mathcal{F}(\rho) = E_a\rho E_a + E_a^\perp\rho E_a^\perp$. Then if d is a system that purifies ρ_{abc}, we have:

$$H(P_a|c) = H(\Pi_A|c) \quad \text{(C7)}$$
$$\geqslant H(\Pi_A|cd) = S(\rho_{Ab}||\sum_j \Pi_{Aj}\rho_{Ab}\Pi_{Aj}) \quad \text{(C8)}$$
$$\geqslant S(\mathcal{F}(\rho_{Ab})||\mathcal{F}(\sum_j \Pi_{Aj}\rho_{Ab}\Pi_{Aj})) \quad \text{(C9)}$$
$$= S(E_a\rho_{Ab}E_a||E_a(\sum_j \Pi_{Aj}\rho_{Ab}\Pi_{Aj})E_a +$$
$$E_a^\perp(\sum_j \Pi_{Aj}\rho_{Ab}\Pi_{Aj})E_a^\perp) \quad \text{(C10)}$$
$$= S(E_a\rho_{Ab}E_a||\sum_j E_a\Pi_{Aj}E_a\rho_{Ab}E_a\Pi_{Aj}E_a) \quad \text{(C11)}$$
$$= S(\rho_{ab}||\sum_j P_{aj}\rho_{ab}P_{aj}). \quad \text{(C12)}$$

Note that the term with $E_a^\perp$ in (C10) disappeared because it lies outside of the support of $E_a\rho_{Ab}E_a$. □

Appendix D: Proof of Theorem 5

Proof. First let us prove the single-POVM uncertainty relation as follows, defining $\lambda_{\max}(\cdot)$ to be the maximum eigenvalue. From Lemma 4,

$$H(P|b) \geqslant S(\rho_{ac}||\sum_j P_j\rho_{ac}P_j) \quad \text{(D1)}$$
$$\geqslant S(\rho_c||\sum_j \mathrm{Tr}_a[P_j\rho_{ac}P_j]) \quad \text{(D2)}$$
$$\geqslant S(\rho_c||\sum_j \lambda_{\max}(P_j)\mathrm{Tr}_a[P_j\rho_{ac}]) \quad \text{(D3)}$$
$$\geqslant S(\rho_c||\max_j \lambda_{\max}(P_j)\sum_j \mathrm{Tr}_a[P_j\rho_{ac}]) \quad \text{(D4)}$$
$$= S(\rho_c||\max_j \lambda_{\max}(P_j)\rho_c) \quad \text{(D5)}$$
$$= -\log\max_j \lambda_{\max}(P_j) = -\log\max_j \|P_j\|_\infty \quad \text{(D6)}$$
$$\geqslant -\log\max_{j,k} \|\sqrt{P_j}\sqrt{P_k}\|_\infty. \quad \text{(D7)}$$

We invoked (33) for step (D2). We used (34) for step (D3), $\lambda_{\max}(P_j)I_a \geqslant P_j$ which implies $\mathrm{Tr}_a[\lambda_{\max}(P_j)I_aT_{ac}] \geqslant \mathrm{Tr}_a[P_jT_{ac}]$, where $T_{ac} = \sqrt{P_j}\rho_{ac}\sqrt{P_j}$ is a positive operator. We also used (34) for step (D4), $\max_j \lambda_{\max}(P_j)\sum_j A_j \geqslant \sum_j \lambda_{\max}(P_j)A_j$ where the A_j are positive operators.

Now for the two-POVM uncertainty relation, consider the quantum channel [as in (42)] $\mathcal{E}_Q(\rho_{ab}) = \sum_k |e_k\rangle\langle e_k| \otimes \mathrm{Tr}_a(Q_k\rho_{ab})$ associated with the Q measurement, where $\{|e_k\rangle\}$ is an orthonormal basis of an auxiliary system e. One can verify that $\mathcal{E}_Q$ is trace-preserving, and its complete positivity follows from the fact that $(\mathcal{E}_Q \otimes \mathcal{I}_c)(\rho_{abc}) = \sum_k |e_k\rangle\langle e_k| \otimes \mathrm{Tr}_a(Q_k\rho_{abc})$ is a positive operator for any system c, where $\mathcal{I}_c$ is the identity channel for c. Also, define $G_{jk} = \sqrt{P_j}Q_k\sqrt{P_j}$, and note

$G_{jk} \leqslant \lambda_{\max}(G_{jk}) I_a$, and $r(P,Q) = \max_{j,k} \lambda_{\max}(G_{jk})$. Then from Lemma 4,

$$
\begin{aligned}
H(P|c) &\geqslant S(\rho_{ab}|| \sum_j P_j \rho_{ab} P_j) \geqslant S(\mathcal{E}_Q(\rho_{ab}) || \sum_j \mathcal{E}_Q(P_j \rho_{ab} P_j)) && \text{(D8)}\\
&= S(\sum_l |e_l\rangle\langle e_l| \otimes \mathrm{Tr}_a\{Q_l \rho_{ab}\} || \sum_{j,k} |e_k\rangle\langle e_k| \otimes \mathrm{Tr}_a\{G_{jk}\sqrt{P_j}\rho_{ab}\sqrt{P_j}\}) && \text{(D9)}\\
&\geqslant S(\sum_l |e_l\rangle\langle e_l| \otimes \mathrm{Tr}_a\{Q_l \rho_{ab}\} || \sum_{j,k} \lambda_{\max}(G_{jk}) |e_k\rangle\langle e_k| \otimes \mathrm{Tr}_a\{P_j \rho_{ab}\}) && \text{(D10)}\\
&\geqslant S(\sum_l |e_l\rangle\langle e_l| \otimes \mathrm{Tr}_a\{Q_l \rho_{ab}\} || r(P,Q) I_e \otimes \rho_b) && \text{(D11)}\\
&= -\log r(P,Q) - S(\sum_l |e_l\rangle\langle e_l| \otimes \mathrm{Tr}_a\{Q_l \rho_{ab}\}) - \mathrm{Tr}[(\sum_l |e_l\rangle\langle e_l| \otimes \mathrm{Tr}_a\{Q_l \rho_{ab}\}) \log(I_e \otimes \rho_b)] && \text{(D12)}\\
&= -\log r(P,Q) - H(Q) - \sum_l q_l S(\rho_{bl}^Q) + S(\rho_b) = -\log r(P,Q) - H(Q|b), && \text{(D13)}
\end{aligned}
$$

where $q_l = \mathrm{Tr}(Q_l \rho_{ab})$ and $q_l \rho_{bl}^Q = \mathrm{Tr}_a(Q_l \rho_{ab})$. We invoked (33) for step (D8), and we invoked (34) for steps (D10) and (D11). [For (D11), $\lambda_{\max}(G_{jk}) \leqslant r(P,Q)$ for each j,k, so replacing each $\lambda_{\max}(G_{jk})$ with $r(P,Q)$ makes the overall operator larger.] Step (D13) involves Lemma 1, part (i). □

Appendix E: Proof of Corollary 6

Proof. (i) Consider a channel ket $|\Omega\rangle$ on $\mathcal{H}_{abc}$ with $P = \{P_j\}$ and $Q = \{Q_k\}$ two POVMs on a, and apply (58) and (59) to $|\Omega\rangle$:

$$
\begin{aligned}
\chi(P,b) &\leqslant H(P) - \log[1/\sqrt{r(P,P)}]\\
\chi(P,b) + \chi(Q,c) &\leqslant H(P) + H(Q) - \log[1/r(P,Q)].
\end{aligned} \tag{E1}
$$

Now decompose $|\Omega\rangle = (I_a \otimes V)|\Phi\rangle$ as in Sec. II B and Fig. 1, where system a' (of the same dimension as a) is introduced and fed into isometry V, and the state $|\Phi\rangle = (1/\sqrt{d_a}) \sum_j |j\rangle_a \otimes |j\rangle_{a'}$ is maximally entangled, expanded here in the computational bases on a and a'. By map-state duality [14], think of $|\Phi\rangle$ as an isometry $\hat{V}$ from $\mathcal{H}_a$ to $\mathcal{H}_{a'}$, with $\hat{V}^\dagger \hat{V} = I_a$ *and* $\hat{V}\hat{V}^\dagger = I_{a'}$ since $d_a = d_{a'}$. This means that $\widetilde{P} = \{\widetilde{P}_j\} = \{\hat{V} P_j \hat{V}^\dagger\}$ and $\widetilde{Q} = \{\widetilde{Q}_k\} = \{\hat{V} Q_k \hat{V}^\dagger\}$ are POVMs on a'. If outcome P_j of P occurs on a, then element $\widetilde{P}_j$ will get fed into the isometry V, so $\chi(P,b) = \chi(\widetilde{P},\mathcal{E})$ and likewise $\chi(Q,c) = \chi(\widetilde{Q},\mathcal{F})$, where $\mathcal{E}$ and $\mathcal{F}$ are the (complementary) channels to b and c, respectively, associated with isometry V. Also, since $\rho_a = I_a/d_a$ for a channel ket, the probability for P_j in (6) given by $p_j = \mathrm{Tr}(P_j \rho_a) = \mathrm{Tr}(P_j)/d_a$ reduces to the corresponding formula in (40), so $H(P) = H(\widetilde{P})$ and likewise $H(Q) = H(\widetilde{Q})$. Finally, show that $r(\widetilde{P},\widetilde{Q}) = r(P,Q)$ as follows:

$$
\begin{aligned}
&\|(\widetilde{P}_j)^{1/2}(\widetilde{Q}_k)^{1/2}\|_\infty^2 = \lambda_{\max}[\hat{V}(Q_k)^{1/2} P_j (Q_k)^{1/2} \hat{V}^\dagger]\\
&= \lambda_{\max}[(Q_k)^{1/2} P_j (Q_k)^{1/2}] = \|(P_j)^{1/2}(Q_k)^{1/2}\|_\infty^2,
\end{aligned} \tag{E2}
$$

where $\lambda_{\max}[\cdot]$ denotes the maximum eigenvalue and we used the fact that $(\widetilde{Q}_k)^{1/2} = \hat{V}(Q_k)^{1/2}\hat{V}^\dagger$, which follows from $[\hat{V}(Q_k)^{1/2}\hat{V}^\dagger]^2 = \hat{V} Q_k \hat{V}^\dagger$ since $(Q_k)^{1/2}$ and $\hat{V}(Q_k)^{1/2}\hat{V}^\dagger$ are positive operators. Thus from (E1),

$$
\begin{aligned}
\chi(\widetilde{P},\mathcal{E}) &\leqslant H(\widetilde{P}) - \log[1/\sqrt{r(\widetilde{P},\widetilde{P})}]\\
\chi(\widetilde{P},\mathcal{E}) + \chi(\widetilde{Q},\mathcal{F}) &\leqslant H(\widetilde{P}) + H(\widetilde{Q}) - \log[1/r(\widetilde{P},\widetilde{Q})].
\end{aligned} \tag{E3}
$$

Since $\hat{V}$ is a one-to-one mapping of the set of POVMs on a to the set of POVMs on a', then (E3) must be true for *all* POVMs on a', and one can replace $\widetilde{P}$ and $\widetilde{Q}$ with P and Q in (E3) for simplicity.

(ii) Equation (68) follows from (67) since $H(v) = H(w) = \log d_a$ from (40). □

Appendix F: Proof of Corollary 7

Proof. (i) For (70), let b be a system that purifies ρ_a, apply (59), and by Theorem 2, $\chi(N,b) = S(\rho_a)$. For (71), again let b purify ρ_a, and apply (58). System c is completely uncorrelated to a, so $H(P|c) = H(P)$, and by Theorem 2, $\chi(N,b) = S(\rho_a)$.

(ii) Equation (72) follows from (71) applied to MUBs x and y:

$$
H(x) + H(y) \geqslant \log 2 + S(\rho_a). \tag{F1}
$$

Denote X, Y, and Z as the Pauli operators whose eigenvectors are the x, y, and z bases. Consider a (possibly mixed) state in the xy plane of the Bloch sphere:

$$\rho_a = (I_a + \alpha X + \beta Y)/2, \tag{F2}$$

for which $H(z) = \log 2$, so (72) clearly holds for states of this form using (F1). Now consider varying ρ_a along a vertical path within the Bloch sphere, from the state ρ_a (in the xy plane) to a state ρ_a' with some z component but with the same x and y components:

$$\rho_a' = (I_a + \alpha X + \beta Y + \gamma Z)/2, \tag{F3}$$

Denoting the relevant state with a subscript, note that $H(x)_{\rho_a} = H(x)_{\rho_a'}$ and $H(y)_{\rho_a} = H(y)_{\rho_a'}$ remain constant, so to prove (72) for general states ρ_a', we just need to show that $H(z)$ decreases more slowly than $S(\rho_a)$ along this path. This would be true if:

$$H(z)_{\rho_a'} - S(\rho_a') \geqslant H(z)_{\rho_a} - S(\rho_a) = \log 2 - S(\rho_a). \tag{F4}$$

Due to the isotropic nature of the Bloch sphere, it is sufficient to check that (F4) holds for an initial state along the x-axis: $\rho_a = (I_a + \alpha X)/2$ and $\rho_a' = (I_a + \alpha X + \gamma Z)/2$, since $H(z)$ and $S(\rho_a)$ will vary in the same way along a vertical path regardless of an initial unitary rotation about z. But for such a state, $S(\rho_a) = H(x)_{\rho_a} = H(x)_{\rho_a'}$, and (F4) reduces to $H(z)_{\rho_a'} + H(x)_{\rho_a'} \geqslant \log 2 + S(\rho_a')$, which is (F1) applied to MUBs z and x. Thus, varying along a vertical path from a state in the xy plane to a state with some z-component keeps the values of $H(x)$ and $H(y)$ constant, while not decreasing the value of $H(z) - S(\rho_a)$, proving the result in general. $\square$

Appendix G: Proof of Theorem 8

Proof. Let c be a system that purifies ρ_{ab}. Re-write (58) as

$$\begin{aligned} \chi(M,c) &\leqslant H(P|b) + H(M) + \log r(P,M), \\ \chi(N,c) &\leqslant H(P|b) + H(N) + \log r(P,N). \end{aligned} \tag{G1}$$

Taken together, these two inequalities give an upper bound on the difference $|\chi(M,c) - \chi(N,c)|$. The difference is at most the one computed by allowing the χ quantity with the highest upper bound in (G1) to reach its bound, and allowing the other χ quantity to be zero. Thus,

$$\begin{aligned} &|\chi(M,c) - \chi(N,c)| \leqslant H(P|b) + \\ &\max\{H(M) + \log r(P,M), H(N) + \log r(P,N)\}. \end{aligned} \tag{G2}$$

By (52), substitute b for c on the left-hand-side.

Rearranging (G1) to lower bound $H(M|c)$ and $H(N|c)$, and upper-bounding each respectively by $H(M)$ and $H(N)$, we can upper-bound their difference by the (maximum) difference between the upper bound of one and the lower bound of the other:

$$\begin{aligned} &|H(M|c) - H(N|c)| \leqslant H(P|b) + \\ &\max\{H(M) + \log r(P,N), H(N) + \log r(P,M)\}. \end{aligned} \tag{G3}$$

Again invoke (52) to switch from c to b and obtain (73).

Now assuming u and v are MU with respect to w, (74) follows from (73) by setting $r(u,w) = r(v,w) = 1/d_a$, and by noting that $H(u) \leqslant \log(d_a)$ and likewise for $H(v)$, so that the max{} term in (73) is non-positive.

Further specializing to the case of $H(w|b) = 0$ and v MU to w, then (62) implies $H(v|c) = \log d_a$ and $\chi(v,c) = 0$, and in turn that $\chi_K(v,c) = 0$, because all χ_K measures are zero under the same conditions. Then by Theorem 3, $H(v|b) = H(v|c) - \Delta\chi(b,c) = \log d_a + S(a|b)$, and $\chi_K(v,b) = \Delta\chi_K(b,c) = S_K(\rho_b) - S_K(\rho_{ab})$. $\square$

Appendix H: Proof of Theorem 10

Proof. (i) First let c purify ρ_{ab}, and by Theorem 3, add the basis-invariant quantity $\Delta\chi(c,b) = H(N|b) - H(N|c) = S(\rho_c) - S(\rho_b) = S(a|b)$ to both sides of (58), setting $Q = N$, to obtain (76). Now to prove (77), let cd purify ρ_{ab} so that $\rho_{abc} = \mathrm{Tr}_d(\rho_{abcd})$ is a general (possibly mixed) state. Again by Theorem 3, $[-S(a|b)] = \chi(M,b) - \chi(M,cd) \leqslant \chi(M,b)$ and $[-S(a|b)] = H(M|cd) - H(M|b) \leqslant H(M|cd) \leqslant H(M|c)$ by (38).

(ii) The argument for complementary quantum channels is the same. Add the basis-invariant quantity $\Delta\chi(\mathcal{E},\mathcal{F})$ to (68) to obtain (78), and obtain (79) using $\chi(u,\mathcal{F}) \leqslant \log d_a - \Delta\chi(\mathcal{E},\mathcal{F})$.

(iii) Equation (80) follows from $S(a:b)/2 \geqslant -S(a|b) \geqslant \log d_a - [H(v|b) + H(w|b)]$. For (81), let cd purify ρ_{ab}, then $H(v|b) + H(w|b) \geqslant \log d_a + S(a|b) \geqslant S(\rho_a) + S(a|b) = S(a:cd) \geqslant S(a:c)$. $\square$

Appendix I: Proof of Theorem 11

Proof. (i) First let us prove (82) for pure $\rho_{abc} = |\Omega\rangle\langle\Omega|$.

$$\begin{aligned} S(a:c) &= S(\rho_a) - \Delta\chi(b,c) \\ &\leqslant H(N) - \log[1/\sqrt{r(N,N)}] - \Delta\chi(b,c) \\ &= H(N|b) + \chi(N,c) - \log[1/\sqrt{r(N,N)}], \end{aligned} \tag{I1}$$

where the first line follows from Theorem 3, and the second line is from (70). Now consider any ρ_{abc}. Apply the just-proven result (I1) to ρ_{abcd}:

$$S(a:c) \leqslant H(N|bd) + \chi(N,c) - \log[1/\sqrt{r(N,N)}]. \tag{I2}$$

where ρ_{abcd} is a purification of ρ_{abc}. Then, (82) is obtained by noting that $H(N|bd) \leqslant H(N|b)$ from (38).

If information about a rank-1 POVM N is perfectly present in b, this implies that the elements of N must be orthogonal and hence normalized, i.e. N is some orthonormal basis $w = \{|w_j\rangle\}$. (The outputs ρ_{bj} cannot all be orthogonal if the inputs are not orthogonal.) By the Truncation theorem of [2], $\rho_{ac} = \sum_j p_j |w_j\rangle\langle w_j| \otimes \rho_{cj}$, i.e. c is at-most classically correlated to the w basis on a. Then the conditional density operators on c (σ_{ck}

occurring with probability q_k) associated with POVM $P = \{P_k\}$ are related to those associated with the w basis by $q_k\sigma_{ck} = \mathrm{Tr}_a(P_k\rho_{ac}) = \sum_j M_{kj}p_j\rho_{cj}$, where $M_{kj} = \langle w_j|P_k|w_j\rangle$. Now use the concavity of the entropy S_K (all of our entropy functions have this property, see Sec. III A) and $\sum_k M_{kj} = 1$ to show that:

$$\begin{aligned}\chi_K(P,c) &= S_K(\rho_c) - \sum_k q_k S_K(\sigma_{ck})\\ &\leqslant S_K(\rho_c) - \sum_{k,j} M_{kj}p_j S_K(\rho_{cj})\\ &= S_K(\rho_c) - \sum_j p_j S_K(\rho_{cj}) = \chi_K(w,c). \end{aligned} \tag{I3}$$

The remark that $\rho_{ac} = \rho_a \otimes \rho_c$ when all types are absent from c seems obvious, although it is rigorously proven in Theorem 1 of [14].

(ii) To prove (84) for pure states, note that the right-hand-side of (82) is an upper bound on $\chi(L,c)$ and $\chi(M,c)$ by (44), so it must also upper-bound their difference:

$$|\chi(L,c)-\chi(M,c)| \leqslant \chi(N,c)+H(N|b)-\log[1/\sqrt{r(N,N)}]. \tag{I4}$$

By (52), b can replace c on the left-hand-side.

In the case where information about N is perfectly present in b and absent from c, $\rho_{ac} = \rho_a\otimes\rho_c$ by part (i) of this theorem, and $S(\rho_b) = S(\rho_{ac}) = S(\rho_a)+S(\rho_c)$ by the additivity of S for product states. Thus by Theorem 3, for any rank-1 POVM L, $\chi(L,b) = \chi(L,b) - \chi(L,c) = S(\rho_b) - S(\rho_c) = S(\rho_a)$.

(iii) Equation (86) follows immediately from $\chi(v,\mathcal{F}) \leqslant \log d_a - \Delta\chi(\mathcal{E},\mathcal{F})$ and $\chi(v,\mathcal{E}) \geqslant \Delta\chi(\mathcal{E},\mathcal{F})$, where $\Delta\chi(\mathcal{E},\mathcal{F}) = \chi(w,\mathcal{E}) - \chi(w,\mathcal{F})$ is basis-invariant by Theorem 3. In the extreme case where the w type of information is perfectly present in $\mathcal{E}$ and absent from $\mathcal{F}$, $\Delta\chi(\mathcal{E},\mathcal{F}) = \log d_a$, hence $\chi(v,\mathcal{F}) = 0$ and $\chi(v,\mathcal{E}) = \log d_a$ for *all* v. □

[1] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 2nd edition, 2005.

[2] Robert B. Griffiths. Types of quantum information. *Phys. Rev. A*, 76:062320, 2007.

[3] Robert B. Griffiths. *Consistent Quantum Theory*. Cambridge University Press, Cambridge, 2002.

[4] Robert B. Griffiths and Chi-Sheng Niu. Semiclassical fourier transform for quantum computation. *Phys. Rev. Lett.*, 76:3228–3231, 1996.

[5] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, 2009.

[6] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 5th edition, 2000.

[7] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner. The uncertainty principle in the presence of quantum memory. *Nature Physics*, 6:659, 2010.

[8] Joseph M. Renes and Jean-Christian Boileau. Conjectured strong complementary information tradeoff. *Phys. Rev. Lett.*, 103:020402, 2009.

[9] M. Tomamichel and R. Renner. The Uncertainty Relation for Smooth Entropies, 2010. e-print arXiv:1009.2015 [quant-ph].

[10] I. Devetak and A. Winter. Classical data compression with quantum side information. *Phys. Rev. A*, 68:042301, 2003.

[11] Asher Peres. Neumark's theorem and quantum inseparability. *Found. Phys.*, 20:1441–1453, 1990.

[12] R. Jozsa, M. Koashi, N. Linden, S. Popescu, S. Presnell, D. Shepherd, and A. Winter. Entanglement cost of generalised measurements. e-print arXiv:quant-ph/0303167.

[13] Asher Peres. *Quantum Theory: Concepts and Methods*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993.

[14] Robert B. Griffiths. Channel kets, entangled states, and the location of quantum information. *Phys. Rev. A*, 71:042337, 2005.

[15] Ingemar Bengtsson and Karol Życzkowski. *Geometry of Quantum States*. Cambridge University Press, Cambridge, 2006.

[16] M. Christandl and A. Winter. Uncertainty, monogamy, and locking of quantum correlations. *Information Theory, IEEE Transactions on*, 51:3159–3165, 2005.

[17] Koenraad M. R. Audenaert. Subadditivity of q-entropies for $q > 1$. *Journal of Mathematical Physics*, 48:083507, 2007.

[18] V. Vedral. The role of relative entropy in quantum information theory. *Rev. Mod. Phys.*, 74:197–234, 2002.

[19] M. Ohya and D. Petz. *Quantum Entropy and Its Use*. Springer, 1st edition, 1993.

[20] C. A. Fuchs. Distinguishability and Accessible Information in Quantum Theory, 1996. e-print arXiv:quant-ph/9601020.

[21] Benjamin Schumacher, Michael Westmoreland, and William K. Wootters. Limitation on the amount of accessible information in a quantum channel. *Phys. Rev. Lett.*, 76:3452–3455, 1996.

[22] Shengjun Wu, Uffe V. Poulsen, and Klaus Mølmer. Correlations in local measurements on a quantum state, and complementarity as an explanation of nonclassicality. *Phys. Rev. A*, 80:032319, Sep 2009.

[23] Rajendra Bhatia. *Matrix Analysis*. Springer-Verlag, New York, 1997.

[24] P. J. Coles, M. Zwolak, L. Yu, and R. B. Griffiths. (Unpublished).

[25] M. Krishna and K. Parthasarathy. An Entropic Uncertainty Principle for Quantum Measurements. *Indian J. of Statistics*, Ser. A 64:842, 2002.

[26] Jorge Sánchez-Ruiz. Improved bounds in the entropic uncertainty and certainty relations for complementary observables. *Physics Letters A*, 201:125 – 131, 1995.

[27] Benjamin Schumacher and Michael D. Westmoreland. Quantum privacy and quantum coherence. *Phys. Rev. Lett.*, 80:5695–5697, 1998.

[28] S. Wehner and A. Winter. Entropic uncertainty relations

– a survey. *New Journal of Physics*, 12:025009, 2010.
[29] Hans Maassen and J. B. M. Uffink. Generalized entropic uncertainty relations. *Phys. Rev. Lett.*, 60:1103–1106, 1988.
[30] Michael J. W. Hall. Information exclusion principle for complementary observables. *Phys. Rev. Lett.*, 74:3307–3311, 1995.
[31] Nicolas J. Cerf, Mohamed Bourennane, Anders Karlsson, and Nicolas Gisin. Security of quantum key distribution using d-level systems. *Phys. Rev. Lett.*, 88:127902, 2002.
[32] Dennis Kretschmann, David W. Kribs, and Robert W. Spekkens. Complementarity of private and correctable subsystems in quantum cryptography and error correction. *Phys. Rev. A*, 78:032330, 2008.
[33] Cédric Bény and Ognyan Oreshkov. General conditions for approximate quantum error correction and near-optimal recovery channels. *Phys. Rev. Lett.*, 104:120501, 2010.
[34] P. Hayden and A. Winter. The Fidelity Alternative and Quantum Measurement Simulation. e-print arXiv:1003.4994 [quant-ph].
[35] R. Prevedel, D. R. Hamel, R. Colbeck, K. Fisher, and K. J. Resch. Experimental investigation of the uncertainty principle in the presence of quantum memory, 2010. e-print arXiv:1012.0332 [quant-ph].