Channel kets, entangled states, and the location of quantum information

Robert B. Griffiths*
Department of Physics, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213, USA
(Received 20 December 2004; published 28 April 2005)

The well-known duality relating entangled states and noisy quantum channels is expressed in terms of a *channel ket*, a pure state on a suitable tripartite system, which functions as a pre-probability allowing the calculation of statistical correlations between, for example, the entrance and exit of a channel, once a framework has been chosen so as to allow a consistent set of probabilities. In each framework the standard notions of ordinary (classical) information theory apply, and it makes sense to ask whether information of a particular sort about one system is or is not present in another system. Quantum effects arise when a single pre-probability is used to compute statistical correlations in different incompatible frameworks, and various constraints on the presence and absence of different kinds of information are expressed in a set of all-or-nothing theorems which generalize or give a precise meaning to the concept of "no-cloning." These theorems are used to discuss the location of information in quantum channels modeled using a mixed-state environment, the classical-quantum channels introduced by Holevo, and the location of information in the physical carriers of a quantum code. It is proposed that both channel and entanglement problems be classified in terms of pure states (functioning as pre-probabilities) on systems of $p \geqslant 2$ parts, with mixed bipartite entanglement and simple noisy channels belonging to the category $p=3$, a five-qubit code to the category $p=6$, etc., then by the dimensions of the Hilbert spaces of the component parts, along with other criteria yet to be determined.

DOI: 10.1103/PhysRevA.71.042337 PACS number(s): 03.67.−a, 03.67.Hk, 03.67.Pp, 03.67.Mn

I. INTRODUCTION

Understanding entangled states and understanding the properties of quantum channels are two central issues in quantum information theory. At least in a formal sense they are the same problem: the duality mapping one onto the other has been discussed explicitly in recent work [1–4] and employed for various purposes in a much larger collection of papers; see [5–10] for a few examples in addition to those in the extensive bibliography in [3]. The early work most often cited is [11,12], though the basic idea is not complicated and has undoubtedly been rediscovered many times. Nonetheless, one has the impression that this duality has yet to be fully exploited, and much more could be done to relate the concepts used in discussing entanglement and the large number of proposed measures of entanglement to the ideas employed for thinking about quantum channels and the definitions of many different sorts of channel capacity. Perhaps a barrier to its full utilization is the fact that this duality remains something of a mathematical abstraction whose connection with more physical ideas has not been totally clear. One aim of the present paper is to relate this duality to concepts of quantum *information*. To be sure, "information" as it applies to the quantum domain is not at present a very precise concept; the appropriate definitions remain the subject of current research and occasional controversy [13–18]. The term is used here in the very broad sense of *statistical correlation*, an idea familiar in classical physics and classical information theory, which deserves to be better understood and more widely applied in the quantum domain.

The duality under discussion can be formulated in various ways. One which seems particularly helpful characterizes a noisy quantum channel using a *channel ket*, an entangled pure state on a suitable tripartite system; see Sec. II C for the precise definition. While this idea is (at least) implicit in previous work, the main emphasis has been on the duality between a density operator describing a mixed state of a bipartite system and what we here call a *dynamical operator* (following [3], where the term *dynamical matrix* is used), closely connected to the superoperator describing the action of a quantum channel. The channel ket is obtained by "purifying" the dynamical operator using a (possibly fictitious) reference system; in turn, the dynamical operator is a partial trace over the projector corresponding to the channel ket. This relationship is well known and frequently exploited in the case of mixed entangled states (see, e.g., p. 110 of [19]). What is less well known is that there are certain advantages, both formal and conceptual, in using pure states rather than (or at least in addition to) mixed states when discussing the *location* of quantum information—see Sec. IV—and thus occasions when a channel ket provides insights not directly available from a dynamical operator. It should be noted that the principal role of a channel ket is the same as that of a dynamical operator or a density operator: it allows one to calculate probabilities for various properties of a quantum system. These probabilities determine the statistical correlations between events at different times that provide a physical description of a quantum channel, just as the statistical correlations between separate quantum systems at a given moment of time provide a physical description of entanglement.

The remainder of this paper is structured in the following way. After introducing some conventions on notation in Sec. II A, the basic map-ket duality is reviewed in Sec. II B; our treatment differs from previous ones mainly in maintaining what we think is a helpful distinction between operators and their matrices. Channel kets are defined in Sec. II C, with

*Electronic address: rgrif@cmu.edu

some simple examples in Sec. II D. Brief remarks on the inverse problem of turning entangled states into channels are found in Sec. II E.

Quantum information in the sense of statistical correlations is the topic of Sec. III. Sample spaces and probabilities for quantum systems are discussed in Sec. III A and applied to correlated systems in III B. The notion of *particular types* of information about certain subsystems being present or absent in other subsystems, which is central to our later discussions, is introduced in Sec. III C for entangled states and extended to quantum channels, where the ideas are very similar modulo a partial transpose, in Sec. III D. These definitions are *qualitative* and do not depend upon any quantitative measures of information. We believe, however, that once correlations have been defined in a consistent manner, there is no barrier to using quantitative information measures, such as Shannon's mutual information; this should take care of the objections raised in [14]. The point of view adopted here is consistent with and an extension of that in [16].

Following this, Sec. IV contains a set of "all-or-nothing" theorems that apply to qualitative aspects of information. These theorems have a number of interesting consequences, some of which are discussed in Sec. V, where they are applied to two special types of quantum channels—mixed-state environment and "CQ" channels—and to the problem of the location of information in quantum codes. In Sec. VI we propose a scheme, at present rather tentative, for classifying both entanglement and channel problems in terms of pure-state entanglement on p-part systems.

The conclusion, Sec. VII has both a summary and a list of open problems. Appendix A contains the proofs of the theorems of Sec. IV and Appendix B a particular result on bipartite entangled kets used in Appendix A.

II. MAP-KET DUALITY AND CHANNEL KETS

A. Notation

We shall use subscripts a,b,c, etc., and sometimes numbers, to label different subsystems of a system with several parts. The Hilbert space $\mathcal{H}_a$ is associated with system $\mathcal{S}_a$, the tensor product

$$\mathcal{H}_{ab} = \mathcal{H}_a \otimes \mathcal{H}_b, \tag{1}$$

with the combined system $\mathcal{S}_{ab}$ consisting of $\mathcal{S}_a$ and $\mathcal{S}_b$, and so forth. For a ket $|\psi\rangle \in \mathcal{H}_{abc}$ we use the notation

$$\psi = [\psi] = |\psi\rangle\langle\psi|, \tag{2}$$

where the square brackets distinguish a dyad from other types of operator. Partial traces are denoted by

$$\psi_{ab} = \mathrm{Tr}_c(\psi), \quad \psi_a = \mathrm{Tr}_b(\psi_{ab}) = \mathrm{Tr}_{bc}(\psi), \tag{3}$$

and so forth, both for dyads and other operators. Operators on the Hilbert space $\mathcal{H}_a$ themselves form a Hilbert space $\hat{\mathcal{H}}_a$, with inner product $\langle A, A'\rangle = \mathrm{Tr}(A^\dagger A')$.

Because the subscript position is used to label the (sub-)system, indices are often written as superscripts in circumstances in which they are not likely to be confused with exponents. Thus $\mathcal{P}=\{|p^j\rangle\}$ denotes an *orthonormal basis* for the Hilbert space $\mathcal{H}_p$ of dimension d_p, with j taking values between 0 and d_p-1. Two such bases $\mathcal{P}$ and $\overline{\mathcal{P}}=\{|\overline{p^j}\rangle\}$ are called *mutually unbiased* if

$$|\langle \overline{p^j}|p^k\rangle| = 1/\sqrt{d_p}, \tag{4}$$

independent of j and k.

More generally, we shall be interested in a *projective decomposition of the identity* of $\mathcal{H}_p$, hereafter called a "decomposition," a collection $\{P^k\}$ of projectors summing to the identity I_p and mutually orthogonal to each other,

$$I_p = \sum_k P^k, \quad P^k P^l = \delta_{kl} P^k. \tag{5}$$

(Recall that a projector is a Hermitian operator equal to its square, so its eigenvalues are 0 and 1.) No confusion arises if the same symbol $\mathcal{P}$ is used to denote an orthonormal basis $\{|p^j\rangle\}$ or the collection $\{[p^j]\}$ of the corresponding projectors.

Given an orthonormal basis $\{|a^j\rangle\}$ of $\mathcal{H}_a$, any ket $|\psi\rangle$ in $\mathcal{H}_{ab}$ can be expanded in the form

$$|\psi\rangle = \sum_j |a^j\rangle \otimes |\beta^j\rangle, \tag{6}$$

where $|\beta^j\rangle = \langle a^j|\psi\rangle$ is uniquely determined by $|\psi\rangle$ and $|a^j\rangle$. If the $\{|\beta^j\rangle\}$ are mutually orthogonal, we shall call (6) a *Schmidt* expansion and sometimes write it in the alternative form

$$|\psi\rangle = \sum_j \sqrt{p_j}|a^j\rangle \otimes |b^j\rangle, \tag{7}$$

with the $\{|b^j\rangle\}$ an orthonormal basis of $\mathcal{H}_b$, and the p_j summing to 1 when $|\psi\rangle$ is normalized, $\langle\psi|\psi\rangle = 1$. By the *support* of an operator A we shall mean the smallest projector P such that

$$PAP = A, \tag{8}$$

or the subspace $\mathcal{P}$ onto which this P projects. The *rank* of A is the trace of P, or the dimension of $\mathcal{P}$, or the number of nonzero (positive) eigenvalues of $A^\dagger A$, or the rank of the matrix representing A.

B. Maps and kets

Given any linear map $M:\mathcal{H}_a \to \mathcal{H}_b$ and an orthonormal basis $\mathcal{A}=\{|a^j\rangle\}$ of $\mathcal{H}_a$, one can define a corresponding ket

$$|\psi\rangle = \sum_j |a_j\rangle \otimes M|a_j\rangle \tag{9}$$

on the tensor product $\mathcal{H}_{ab}$. Conversely, given such a ket, one can always expand it in the form (6) using the basis $\mathcal{A}$ and define a map M by

$$M|a^j\rangle = |\beta^j\rangle, \tag{10}$$

and its extension to all of $\mathcal{H}_a$ by linearity. These two formulas define the *map-ket duality* used throughout the rest of this paper.

The duality depends, obviously, on the choice of orthonormal basis $\mathcal{A}$; given a different choice $\overline{\mathcal{A}}=\{|\overline{a^j}\rangle\}$, a given

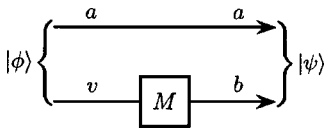

FIG. 1. Circuit illustrating map M - ket $|\psi\rangle$ duality.

map will lead to a different ket and vice versa. For those who (like the author) prefer to write formulas whenever possible in basis-independent form, this dependence is somewhat annoying. One can get around it, as in [2,3], by always using a single basis. We prefer to maintain the usual distinction between operators and matrices. The price for doing this is not exorbitant, because the basis dependence can always be expressed in terms of a suitable unitary transformation on $\mathcal{H}_a$. And if one is primarily concerned with concepts which are *invariant under local unitaries*, meaning unitary operations which are tensor products of unitaries on individual subsystems, such basis dependence is not intolerable.

A way of visualizing the relationship between $|\psi\rangle$ and M, and for understanding the ambiguity associated with the choice of basis, is indicated by the circuit in Fig. 1, where $|\phi\rangle$ is a fully entangled state

$$|\phi\rangle = \sum_j |a^j\rangle \otimes |v^j\rangle \tag{11}$$

on the system $\mathcal{H}_a \otimes \mathcal{H}_v$, $\mathcal{H}_v$ is an auxiliary Hilbert space of the same dimension of $\mathcal{H}_a$, and $M|v^j\rangle = |\beta^j\rangle$, as in (10). Choosing a different fully-entangled state in place of (11) would result in a different relationship between M and $|\psi\rangle$; this is precisely the ambiguity previously discussed, and provides a good way of analyzing it.

C. Channel kets and superoperators

We adopt the following by now fairly standard model for a noisy quantum channel. A unitary time transformation T maps the tensor product $\mathcal{H}_{ae}$ of the Hilbert space $\mathcal{H}_a$ of the *channel entrance* $\mathcal{S}_a$ and the space $\mathcal{H}_e$ of the (initial) *environment* $\mathcal{S}_e$, at some initial time to $\mathcal{H}_{bf} = \mathcal{H}_b \otimes \mathcal{H}_f$, corresponding to the *channel exit* or *output* $\mathcal{S}_b$ and *environment* $\mathcal{S}_f$, at some later time, Fig. 2. Initially the environment is in a *fixed* pure state $|e^0\rangle$, whereas the initial state of the channel is arbitrary, not fixed in advance. Because $|e^0\rangle$ is fixed, the only relevant effect of the unitary operator T is that embodied in the isometry $V: \mathcal{H}_a \to \mathcal{H}_{bf}$ defined by

$$V|a\rangle = T(|a\rangle \otimes |e^0\rangle) \tag{12}$$

and shown schematically in the second part of Fig. 2. Often $\mathcal{H}_a$ and $\mathcal{H}_b$ are identified with each other, and $\mathcal{H}_e$ with $\mathcal{H}_f$. Maintaining the distinction both allows for the possibility, sometimes useful, that the dimensions of $\mathcal{H}_a$ and $\mathcal{H}_b$ may be

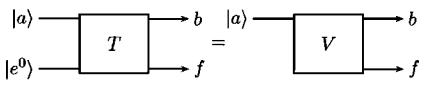

FIG. 2. Quantum channel using a unitary transformation T or isometry V.

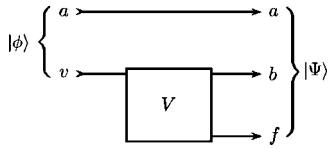

FIG. 3. Circuit for visualizing the channel ket $|\Psi\rangle$.

different, but equally important permits a distinct label. It is sometimes useful to assume that the environment is initially in a mixed, rather than a pure state (see Sec. V A), but there is no loss in generality in assuming a pure state $|e^0\rangle$, since a mixed state can always be purified by introducing an auxiliary system, which can then be thought of as part of $\mathcal{S}_e$.

The *channel ket* $|\Psi\rangle \in \mathcal{H}_{abf}$ is defined as the ket dual to V in the sense of Sec. II B,

$$\sqrt{d_a}|\Psi\rangle = \sum_j |a^j\rangle \otimes V|a^j\rangle \in \mathcal{H}_a \otimes \mathcal{H}_{bf}, \tag{13}$$

using an orthonormal basis $\mathcal{A} = \{|a^j\rangle\}$ of $\mathcal{H}_a$. The normalization $\||\Psi\rangle\| = 1$ is of no great importance—which is why $\sqrt{d_a}$ is placed on the left side of this equation—but does simplify certain formulas. Notice that $|\Psi\rangle$ is a pure state on a tripartite system.

The channel ket can be visualized using Fig. 3, the obvious analog of Fig. 1, as obtained by transmitting the $\mathcal{S}_v$ part of the fully entangled state (11) through the channel, while preserving the $\mathcal{S}_a$ part unchanged. It is important to distinguish the *definition* of the channel ket, given in (13), from this visualization, in that $|\Psi\rangle$ is a mathematical object which functions as a pre-probability, used to calculate probabilities of various events or processes associated with the channel, as discussed in Sec. III, quite apart from whether the channel is being used in the manner just described.

Following the notation of Sec. II A, the symbol Ψ denotes the dyad $|\Psi\rangle\langle\Psi|$, and subscripts are used to indicate its partial traces. Of particular importance is the *dynamical operator*

$$R := \Psi_{ab} = \mathrm{Tr}_f(\Psi) \in \hat{\mathcal{H}}_{ab}, \tag{14}$$

which corresponds to the dynamical matrix defined in [3] (apart from the order ab as against ba); the latter is R with a particular choice of basis. Since Ψ is a positive operator, so is R, and given the normalization in (13), R has unit trace. In addition, because V is an isometry,

$$R_a = \mathrm{Tr}_b(R) = \Psi_a = I_a/d_a. \tag{15}$$

Thus R is a density operator for the bipartite system $\mathcal{H}_{ab}$, with the special property that R_a is proportional to the identity. Hence whatever intuition one possesses for mixed states on bipartite systems can at once be applied to R; e.g., one can ask if it is separable and, if not, how entangled it is according to any of the numerous measures of mixed-state entanglement, etc.

But in addition, R completely determines the properties of the noisy quantum channel—that is, the channel superoperator—up to a unitary transformation of the channel

input $\mathcal{H}_a$ corresponding to different choices for the basis used in the definition (13). The channel superoperator $\mathcal{V}$ is the map from $\hat{\mathcal{H}}_a$ to $\hat{\mathcal{H}}_b$ defined by

$$\mathcal{V}(A) = \mathrm{Tr}_f(VAV^\dagger) \tag{16}$$

for any operator A in $\hat{\mathcal{H}}_a$. To explore how $\mathcal{V}$ is related to R, it is helpful to choose an orthonormal basis $\{|f^l\rangle\}$ for $\mathcal{H}_f$ and expand $|\Psi\rangle$ as

$$|\Psi\rangle = \sum_l |\kappa^l\rangle \otimes |f^l\rangle \in \mathcal{H}_{ab} \otimes \mathcal{H}_f. \tag{17}$$

We shall refer to the expansion coefficients $\{|\kappa^l\rangle\}$ as *Kraus kets*, in that they can, using the duality introduced in Sec. II B, be turned into maps

$$\sqrt{d_a}|\kappa^l\rangle = \sum_j |a^j\rangle \otimes K_l|a^j\rangle, \tag{18}$$

where the K_l are the usual Kraus operators, labeled by subscripts as is the usual convention. They can be used to express the channel superoperator in the familiar form

$$\mathcal{V}(A) = \sum_l K_l A K_l^\dagger. \tag{19}$$

The usual normalization $\Sigma_l K_l^\dagger K_l = I_a$ is the counterpart of (15).

The K_l no longer depend upon the arbitrary choice of basis $\{|a^j\rangle\}$ used in defining $|\Psi\rangle$, as this dependence is undone when kets are changed to maps (using the same basis) in (18), but they do depend upon the choice of basis $\{|f^l\rangle\}$. One can eliminate, or (in degenerate cases) at least mitigate this arbitrariness by making (17) a Schmidt expansion, so that the $\{|\kappa^l\rangle\}$ are orthogonal to one another or, equivalently,

$$\mathrm{Tr}_a(K_l^\dagger K_m) = 0 \quad \text{for } l \neq m. \tag{20}$$

In that case the number of nonzero terms in (17), which could be called the *Kraus rank* of the channel superoperator, is the rank (in the ordinary sense) of the dynamical operator $R = \Psi_{ab}$.

Combining (17) and (18), one obtains the expression

$$R = \Psi_{ab} = \sum_l [\kappa^l] = \frac{1}{d_a}\sum_{j,k} |a^j\rangle\langle a^k| \otimes \sum_l K_l|a^j\rangle\langle a^k|K_l^\dagger \tag{21}$$

for R, and from it another formula

$$\mathcal{V}(A) = \mathrm{Tr}_a[(A \otimes I)Q] \tag{22}$$

for the channel superoperator in terms of the *transition operator* Q, the partial transpose

$$Q = R^{T_A} = \frac{1}{d_a}\sum_{j,k} |a^k\rangle\langle a^j| \otimes \sum_l K_l|a^j\rangle\langle a^k|K_l^\dagger \in \hat{\mathcal{H}}_{ab} \tag{23}$$

of the dynamical operator with respect to the basis $\mathcal{A} = \{|a^j\rangle\}$. Once again, by using this same basis a second time, its effect in defining $|\Psi\rangle$ has been undone, and Q is independent of the basis, consistent with the fact that the superoperator $\mathcal{V}$ in (22) also does not depend upon the choice of basis. Despite their close relationship, Q and R are very different types of operators; the latter is positive, and the former, while it is Hermitian, will typically have negative as well as positive eigenvalues.

The superoperator $\mathcal{V}$ is a map from $\hat{\mathcal{H}}_a$ to $\hat{\mathcal{H}}_b$, so it can be represented as a matrix once orthonormal operator bases have been defined for these two spaces. There are many ways of choosing such bases, but one that is particularly convenient when $\hat{\mathcal{H}}_a$ and $\hat{\mathcal{H}}_b$ are qubits is the *Pauli representation* using $\{\sigma_a^j\}$, with $j=0$ the identity and $j=1,2,3$ the x, y, and z Pauli matrices in the standard basis of $\mathcal{H}_a$, and similarly $\{\sigma_b^j\}$. Expanding the transition operator Q in the Pauli form—see the examples in Sec. II D—often provides a clearer notion of what a noisy channel "does" than is evident by looking at the Kraus operators. There are various ways of generalizing this representation to higher-dimensional spaces. For the case of a channel superoperator there is some advantage to using a basis of Hermitian operators, rather than unitaries as in [20], because the resulting matrix is real. If the basis is again denoted by $\{\sigma^j\}$, with $0 \leqslant j \leqslant d^2-1$ for a d-dimensional Hilbert space, one can again let σ^0 be the identity, so that the orthogonality condition

$$\mathrm{Tr}(\sigma^j\sigma^k) = \delta_{jk}d \tag{24}$$

implies that σ^j for $j>0$ has zero trace—this makes it easy to take partial traces of operators written in Pauli form.

D. Examples of one qubit channels

We use the names for one qubit channels employed in Sec. 8.3 of [19], but employ p in a way which identifies it as the probability of an error. The channel kets are sums of terms of the form $|ab\rangle \otimes |f\rangle$, where a and b are either 0 or 1, but f sometimes takes larger values.

The bit flip channel is described by

$$\sqrt{2}|\Psi\rangle = \sqrt{1-p}(|00\rangle + |11\rangle) \otimes |0\rangle + \sqrt{p}(|01\rangle + |10\rangle) \otimes |1\rangle, \tag{25}$$

leading to a transition operator

$$4Q = I + \sigma_a^1\sigma_b^1 + (1-2p)[\sigma_a^2\sigma_b^2 + \sigma_a^3\sigma_b^3] \tag{26}$$

in the Pauli representation. The dynamical operator R is the same except for a minus sign multiplying the term $\sigma_a^2\sigma_b^2$, reflecting the fact that σ^y changes sign when transposed.

For the amplitude damping channel the corresponding expressions are

$$\sqrt{2}|\Psi\rangle = \sqrt{1-p}(|00\rangle + |11\rangle) \otimes |0\rangle + \sqrt{p}|10\rangle \otimes |1\rangle, \tag{27}$$

$$4Q = I + p\sigma_b^3 + \sqrt{1-p}(\sigma_a^1\sigma_b^1 + \sigma_a^2\sigma_b^2) + (1-p)\sigma_a^3\sigma_b^3. \tag{28}$$

A depolarizing channel requires a larger environment:

$$2|\Psi\rangle = \sqrt{2-3p}(|00\rangle + |11\rangle) \otimes |0\rangle + \sqrt{p}(|00\rangle - |11\rangle) \otimes |1\rangle + \sqrt{2p}(|01\rangle \otimes |2\rangle + |10\rangle \otimes |3\rangle), \tag{29}$$

$$4Q = I + (1-2p)(\sigma_a^1\sigma_b^1 + \sigma_a^2\sigma_b^2 + \sigma_a^3\sigma_b^3). \quad (30)$$

Again, the dynamical operator R is obtained by changing the sign of $\sigma_a^2\sigma_b^2$.

E. From entangled states to channels

As shown in Sec. II C, any noisy channel modeled as in Fig. 2 can be mapped onto an equivalent entangled ket $|\Psi\rangle$ on a tripartite system and thence onto a density operator whose partial transpose determines the channel superoperator. Can one do the reverse, starting with a tripartite ket $|\Psi\rangle$ or a bipartite density operator R. Yes, aside from the condition that Ψ_a (or R_a) be proportional to the identity operator I_a. But if this is not true, can one still turn an entanglement problem into a channel problem.There are at least two approaches, each with advantages and disadvantages.

The first is to begin with a unitary operator or isometry as in Fig. 2, but then instead of "throwing away" the environment $\mathcal{H}_f$, apply a projector F to this part of the output, and condition on the resulting state. One can think of this as carrying out a measurement on $\mathcal{H}_f$ that determines whether F is true or false and throwing away the results of all experiments in which it is false. The consequence of an appropriately chosen "post-selection" of this type will be a set of (conditional) probabilities that correspond to those of the original ket or density operator; in other words, one obtains the same pre-probability—see Sec. III below. The second approach is based on Fig. 3, and the idea is to replace the fully entangled $|\phi\rangle$ with a different entangled state, chosen so ϕ_a is no longer proportional to I_a, but to Ψ_a (or R_a).

The question remains as to whether either of these procedures is worthwhile, and that depends on one's goals. Rather than turning entanglement problems into channel problems, it may be simpler to do the reverse, as in the classification scheme proposed in Sec. VI. This allows the mathematical structure of the two types of problem to be compared. If, on the other hand, there is quite a bit of useful mathematical and physical intuition to be wrung from contemplating how quantum systems develop in time, the approaches mentioned in the previous paragraph may be worthwhile. Until the channel-entanglement duality has been more thoroughly explored, it is hard to say which approach is best. In any case, there are significant entanglement problems that map in a simple way onto channel problems, and a study of what entanglement does and does not mean in such cases might be very helpful.

III. QUANTUM INFORMATION

A. Sample spaces and probabilities

The basic concept of "information" used in the following discussion is that of a *statistical correlation*. This morning's newspaper contains information because the symbols are correlated in an appropriate way with yesterday's events. Information is contained in a photon traveling down an optical fiber because its properties are correlated with whatever produced it, and with the effects produced by the further processes it will undergo. An encrypted message contains information in that its symbols are correlated with those in the key used to encrypt or decrypt it. Shannon's information theory provides numerical measures for these statistical correlations, which apply to quantum as well as to classical systems (which, of course, are in fact quantum mechanical), when probabilities have been properly defined.

Standard probability theory [21–23] is based on the idea of a *sample space* of mutually exclusive properties. A quantum sample space or *framework* can be constructed using the mutually exclusive properties associated with a decomposition of the identity (Sec. II A) of the Hilbert space used to describe the system. Given such a sample space one can assign probabilities using the standard formula

$$p_k = \langle P^k \rangle = \langle \psi | P^k | \psi \rangle = \mathrm{Tr}(P^k \rho), \quad (31)$$

where the quantum system is assumed to be described by a ket $|\psi\rangle$ or density operator ρ functioning as a *pre-probability*—i.e., as a device for generating probabilities [24]. Probabilities in quantum mechanics are often discussed in terms of *measurements*, which provide a good approach to understanding them in operational terms, even though it is rather unsatisfactory from a fundamental perspective (the infamous "measurement problem;" see, e.g., [25]). For present purposes such measurements should be thought of as *ideal projective measurements* which reveal the (microscopic) properties they are designed to measure; see the discussion in Chaps. 17 and 18 of [24]. We shall have no need of more complicated concepts such as POVMs (see, e.g., p. 90 of [19], or Chap 7 of [26]). From time to time there have been proposals to introduce nonstandard notions of probability into quantum mechanics, but these have not proven very successful, and we shall not use them.

In quantum mechanics, in contrast to classical physics, one is typically interested in a variety of sample spaces that are incompatible with each other, but whose probabilities can all be generated from a single pre-probability. For example, what is the probability that $S_x = +1/2$, or that $S_z = -1/2$, for a spin-half particle¿ The same ket or density operator may be used to answer these questions by inserting different projectors in (31), but there is no way of combining the answers to make them refer to a single physical system, as it makes no sense to talk about $S_x = +1/2$ AND $S_z = -1/2$ or any other logical combination of propositions associated with incompatible decompositions of the identity whose projectors do not commute with each other. Traditional textbooks state that S_x and S_z cannot be simultaneously *measured*, which is correct. But the reason such joint measurements are impossible in a quantum world is that the combined properties *do not exist*: such a combination is incompatible with the mathematical structure of the quantum Hilbert space (see Chap. 4 of [24]). Treating incompatible sample spaces as if they were compatible and combining the probabilities of one with the other is the same sort of mistake as ignoring the difference between xp and px when these symbols refer to quantum operators.

Consequently, one must be careful when giving a *physical* interpretation to the various *mathematical* constraints, such as those in Sec. IV, relating probabilities on different incompatible sample spaces generated by a single pre-probability.

They cannot refer to a single quantum system, as it cannot be simultaneously described by incompatible frameworks. Instead, one must take a counterfactual approach: "This is what happens when a qubit initially in state $|0\rangle$ is sent through the channel, *but if instead it had been* in the state $(|0\rangle+|1\rangle)/2$, then…." To be sure, counterfactuals can themselves produce headaches in quantum theory if improperly used; for a consistent approach, see Chap. 19 of [24]. Alternatively, one can imagine different experiments carried out on an array of nominally identical systems.

In comparison with classical physics, the new and unfamiliar element in quantum information theory is the multiplicity of incompatible sample spaces and probability distributions associated with them, even when one is using a single pre-probability. Finding good ways to think about this is a fundamental problem, perhaps the fundamental problem, of quantum information and thus a major challenge to our understanding the world in quantum terms.

B. Correlations

Consider two systems $\mathcal{S}_a$ and $\mathcal{S}_b$, with Hilbert spaces $\mathcal{H}_a$ and $\mathcal{H}_b$, and let $\{A^j\}$ and $\{B^k\}$ be decompositions of the respective identities I_a and I_b. On the tensor product $\mathcal{H}_{ab}=\mathcal{H}_a\otimes\mathcal{H}_b$ used to describe the combined systems the projectors $\{A^jB^k\}$ form a decomposition of I_{ab}, and thus a sample space, to which probabilities may be assigned as in (31):

$$\Pr(A^j,B^k)=\langle A^jB^k\rangle=\mathrm{Tr}[(A^j\otimes B^k)\rho], \tag{32}$$

with $\rho=|\psi\rangle\langle\psi|$ for a pure state $|\psi\rangle$. The marginal distributions

$$\Pr(A^j)=\sum_k \Pr(A^j,B^k)=\langle A^j\rangle,$$

$$\Pr(B^k)=\sum_j \Pr(A^j,B^k)=\langle B^k\rangle \tag{33}$$

are obtained by summing or by inserting $A^j\otimes I$ (i.e., A^j) or $I\otimes B^k$ (i.e., B^k) on the right side of (32). One can think of $\Pr(A^j,B^k)$ as the joint probability distribution of two random variables which take on integer values j and k, and apply to it any standard measure of correlation including, if one wants, the Shannon mutual information $I(A:B)$. Note, in particular, the condition for *statistical independence*:

$$\Pr(A^j,B^k)=\Pr(A^j)\Pr(B^k),\ \text{or}\ \langle A^jB^k\rangle=\langle A^j\rangle\langle B^k\rangle. \tag{34}$$

If one thinks of $\mathcal{S}_a$ and $\mathcal{S}_b$ as physically separated systems, then the joint probability distribution (32) will be the same as that of the outcomes of ideal measurements of $\{A^j\}$ and $\{B^k\}$ carried out on the separate systems. Consequently, the measurement outcomes will be correlated in precisely the same way as the quantum properties that have been measured, and one can use either the language of properties (our approach) or of measurement outcomes to discuss these statistical correlations. Discussions of measurements in textbooks often refer to "observables" rather than decompositions. Given a decomposition $\{A^j\}$, one can always construct a corresponding observable $O=\Sigma_j a_jA^j$ with distinct (real) eigenvalues: $a_j\neq a_k$ for $j\neq k$. But for our purposes these eigenvalues play no role, so the language of decompositions tends to be clearer than that referring to observables.

C. Information present and absent

Because of the multiplicity of incompatible quantum sample spaces, one needs to identify different *types* or *varieties* of information potentially available about a particular system. Given a decomposition $\mathcal{A}=\{A^j\}$ of I_a, we shall say that the $\mathcal{A}$ information about $\mathcal{S}_a$ is *present*, or *perfectly present*, in another system $\mathcal{S}_b$ for a given pre-probability provided there exists a decomposition $\mathcal{B}=\{B^k\}$ of I_b such that

$$\langle A^jB^k\rangle=\delta_{jk}\langle A^j\rangle=\delta_{jk}\langle B^k\rangle, \tag{35}$$

where one may have to renumber the projectors in one of the collections to satisfy this condition. A little thought will show that the first equality implies the second. The symmetry of the definition implies that when some type of information about $\mathcal{S}_a$ is available in $\mathcal{S}_b$, there is also some type of information about $\mathcal{S}_b$ available in $\mathcal{S}_a$. Although we shall not make use of it in this paper, it is worth mentioning that the Shannon mutual information $I(\mathcal{A}:\mathcal{B})$ in this case is $(-\Sigma_j p_j\log p_j)$ with $p_j=\langle A^j\rangle$.

If the $\mathcal{A}$ information about $\mathcal{S}_a$ is present in $\mathcal{S}_b$ (in the sense just defined) for *every* decomposition of I_a, we shall say that *all* the (quantum) information about $\mathcal{S}_a$ is in $\mathcal{S}_b$. Clearly it suffices to check this for every orthonormal basis $\{|a^j\rangle\}$. Less obvious (theorem 4 in Sec. IV) is the fact that one need not check them all: two properly chosen incompatible bases suffice. We shall say that $\mathcal{S}_a$ and $\mathcal{S}_b$ are *informationally equivalent* when all information about $\mathcal{S}_a$ is in $\mathcal{S}_b$ and all information about $\mathcal{S}_b$ is in $\mathcal{S}_a$.

The $\mathcal{A}=\{A^j\}$ information about $\mathcal{S}_a$ is (completely) *absent* from $\mathcal{S}_b$ provided *any* choice of a decomposition $\{B^k\}$ of I_b is statistically independent, (34). A little thought shows that this is equivalent to the requirement that

$$\mathrm{Tr}_a(A^j\rho)=\langle A^j\rangle\rho_b=p_j\rho_b \tag{36}$$

for every j, where $\rho_b=\mathrm{Tr}_a(\rho)$ is the reduced density operator for ρ_b. [Note that it suffices to require that the operators defined by the left side of (36) be proportional to one another; when that is so, summing them shows they are all proportional to ρ_b.] In other words, for every j such that p_j is not zero, the density operator *conditional* on A^j,

$$\overrightarrow{\rho_b^j}=\mathrm{Tr}_a(A^j\rho)/p_j, \tag{37}$$

is the same as ρ_b.

If for every decomposition $\mathcal{A}$ of I_a—it suffices to check all orthonormal bases—the corresponding information about $\mathcal{S}_a$ is absent from $\mathcal{S}_b$, one can show [theorem 1 (iii) in Sec. IV] that

$$\rho=\rho_a\otimes\rho_b, \tag{38}$$

from which it follows that all information of any sort about $\mathcal{S}_b$ is also absent from $\mathcal{S}_a$. In this case we shall say that $\mathcal{S}_a$ and $\mathcal{S}_b$ are (completely) *uncorrelated*. No conceivable measurement on one of these systems will provide any information about the other.

In the case of three or more systems, the presence or absence of particular types of information about $\mathcal{S}_a$ satisfies some intuitively obvious rules. If $\mathcal{A}$ information about $\mathcal{S}_a$ is present in $\mathcal{S}_b$, it is also present in the combined system $\mathcal{S}_b$ and $\mathcal{S}_c$, denoted by $\mathcal{S}_{bc}$. If it is absent from $\mathcal{S}_{bc}$, it is absent from both $\mathcal{S}_b$ and $\mathcal{S}_c$. The same is true when "$\mathcal{A}$ information" is replaced by "all information."

These definitions of information perfectly present or completely absent make no reference to any sort of numerical measure of correlation and thus are useful for a *qualitative* rather than a quantitative discussion of quantum information. This is not to say that quantitative measures are unimportant—far from it—but they lie outside the scope of this paper. It is hoped that the qualitative approach developed here will help organize and motivate quantitative discussions (see Sec. VII B).

D. Correlations for channels

The preceding discussion referred to properties of separated systems $\mathcal{S}_a$ and $\mathcal{S}_b$ at the same time. Basically the same ideas apply in the case of quantum channels, where $\mathcal{S}_a$ is the channel input at an earlier time and $\mathcal{S}_b$ its output at a later time (Sec. II C). The only difference is the manner in which one calculates a joint probability distribution; (32) is replaced by

$$\Pr(A^j, B^k) = \langle A^j B^k \rangle = \mathrm{Tr}[(A^j \otimes B^k)Q]. \tag{39}$$

Here the transition operator Q [see (23)] takes the place of the density operator in (32). The marginals are once again given by (33). The fact that Q is the partial transpose of a density operator R guarantees that the probabilities in (39) are well defined; indeed, they behave very much like those of a bipartite system described by R.

One can once again visualize $\{B^k\}$ in terms of idealized measurements of what emerges from the channel, but the corresponding intuitive picture of $\{A^j\}$ is an ideal *preparation*. Of course, it is no more possible to prepare a quantum system in a state of two (or more) incompatible properties than it is to measure such a state, for such states do not exist in the quantum world. And just as an ideal measurement reveals a property possessed by a quantum system at a slightly earlier time, an ideal preparation results in a quantum system having a specific property at a slightly later time. The language of "preparation" and "measurement" is useful both for providing quantum concepts with intuitive content and for relating quantum theory to laboratory experiments, but it should be used to illuminate, not replace, the notion of statistical correlations among microscopic properties, whether at the same or at different times, as this is the more fundamental concept.

The correlations obtained using a transition operator, (39), are not entirely the same as those arising from a density operator, (32), but the differences are rather subtle. Given a pair of decompositions $\mathcal{A}$ and $\mathcal{B}$, there is no way of telling whether the joint probability distribution comes from a density or a transition operator. What can happen with *sets* of correlations for *incompatible* decompositions, when they are generated by a single pre-probability, is best illustrated by means of an example. For a perfect one-qubit channel, $p=0$ in (26), each component of angular momentum of a spin-half particle is identical at the entrance and at the exit,

$$\langle \sigma_a^x \sigma_b^x \rangle = \langle \sigma_a^y \sigma_b^y \rangle = \langle \sigma_a^z \sigma_b^z \rangle = 1. \tag{40}$$

However, this type of correlation is impossible for two separate systems at the same time. What one can, instead, achieve by using an appropriate (pure state) density operator is

$$\langle \sigma_a^x \sigma_b^x \rangle = -\langle \sigma_a^y \sigma_b^y \rangle = \langle \sigma_a^z \sigma_b^z \rangle = 1 \tag{41}$$

or something similar: one of the terms (it need not be $\langle \sigma_a^y \sigma_b^y \rangle$) must have a minus sign, or else there are three minus signs, as in the famous spin-singlet state used in discussions of the Einstein-Podolsky-Rosen paradox. Similarly, (41) is impossible for a quantum channel.

Interesting as these differences, which arise from the partial transpose in (23), may be, they are basically irrelevant to the concerns of this paper. The definitions of information perfectly present or completely absent given in Sec. III C above and the theorems in Sec. IV below apply equally to channels and entangled states. In both cases the fundamental issue is statistical correlations and what quantum theory has to say about them, and that is exactly the same once proper account is taken of the partial transpose.

IV. ALL-OR-NOTHING THEOREMS

It is convenient to organize a number of qualitative "all-or-nothing" results on the location of quantum information in a series of eight theorems. The first four refer to bipartite and the last four to tripartite systems. In several cases there are separate results depending upon whether the pre-probability is a pure state, indicated by a ket $|\Psi\rangle$, or a density operator ρ. The former are stronger than the latter, and the reader should keep in mind that any result that is valid for a density operator applies equally to the case of a pure state, even if that is not explicitly stated.

While the theorems are stated for entangled states, thought of as different systems at a single instant of time, they apply equally to correlations at two different times in a quantum channel, for which $|\Psi\rangle$ is the channel ket. The bipartite systems used in the first four theorems are sometimes designated $\mathcal{S}_{ab}$ and sometimes $\mathcal{S}_{ac}$. This makes the notation consistent with the later theorems for tripartite systems, where information *about* $\mathcal{S}_a$ is *present* in $\mathcal{S}_b$ and/or *absent* from $\mathcal{S}_c$. Note that, in agreement with the definitions in Sec. III C, "present" means perfectly or completely present; "absent" means completely absent. The proofs will be found in Appendix A.

The tripartite theorems have a no-cloning "smell" to them and represent an attempt to give this important, but somewhat elusive, notion a precise information-theoretic content. The absence of theorems for p-part systems with $p \geqslant 4$ reflects our inability to find results of corresponding generality, and we hope our readers will be more successful. But keep in mind that a tripartite theorem might, for example, be usefully applied to $\mathcal{S}_{abcd}$ thought of as consisting of $\mathcal{S}_a$, $\mathcal{S}_b$, and $\mathcal{S}_{cd}$—a strategy employed in discussing quantum codes in Sec. V C.

Theorem 1. Absence of information.

(i) If $\mathcal{A}=\{A^l\}$ is a decomposition of I_a, the $\mathcal{A}$ information about $\mathcal{S}_a$ is absent from $\mathcal{S}_c$ for a pre-probability $|\Psi\rangle \in \mathcal{H}_{ac}$ if and only if

$$PA^lP = a_lP, \tag{42}$$

where P is the projector on the support of Ψ_a and the a_l are (non-negative) constants. The following is equivalent to (42):

$$\langle p^j|A^l|p^k\rangle = a_l\delta_{jk}, \tag{43}$$

where $\{|p^j\rangle\}$ is a collection of orthonormal states which span the support of Ψ_a, so that $P=\Sigma_j|p^j\rangle\langle p^j|$.

(ii) If $\mathcal{A}=\{|a^j\rangle\}$ is an orthonormal basis and all $\mathcal{A}$ information about $\mathcal{S}_a$ is absent from $\mathcal{S}_c$ for $|\Psi\rangle \in \mathcal{H}_{ac}$, then

$$|\Psi\rangle = |\alpha\rangle \otimes |\gamma\rangle \tag{44}$$

is a product state on $\mathcal{H}_a \otimes \mathcal{H}_c$.

(iii) All information about $\mathcal{S}_a$ is absent from $\mathcal{S}_c$ for a pre-probability $\rho \in \hat{\mathcal{H}}_{ac}$ if and only if

$$\rho = \rho_a \otimes \rho_c, \tag{45}$$

which implies that all information about $\mathcal{S}_c$ is absent from $\mathcal{S}_a$ (the two are uncorrelated).

Theorem 2. Presence of particular information.

(i) The $\mathcal{A}=\{A^l\}$ information about $\mathcal{S}_a$ is present in $\mathcal{S}_b$ for $\rho \in \hat{\mathcal{H}}_{ab}$ if and only if

$$\Lambda^l\Lambda^m = 0 \quad \text{for } l \neq m, \tag{46}$$

where

$$\Lambda^l = \mathrm{Tr}_a(A^l\rho). \tag{47}$$

(ii) The $\mathcal{A}=\{A^l\}$ information about $\mathcal{S}_a$ is present in $\mathcal{S}_b$ for $|\Psi\rangle \in \mathcal{H}_{ab}$ if and only if

$$[A^l, \Psi_a] = 0 \tag{48}$$

for all l. In particular, if $\mathcal{A}=\{|a^j\rangle\}$ is an orthonormal basis, (48) is equivalent to the requirement that

$$|\Psi\rangle = \sum_j |a^j\rangle \otimes |\beta^j\rangle \tag{49}$$

be a Schmidt expansion—i.e., $\langle\beta^k|\beta^j\rangle=0$ for $j \neq k$.

(iii) If the $\mathcal{A}=\{A^l\}$ information about $\mathcal{S}_a$ is present in $\mathcal{S}_b$ for $\rho \in \hat{\mathcal{H}}_{ab}$, then for all l

$$[A^l, \rho_a] = 0. \tag{50}$$

Note that if $\mathcal{A}$ is an orthonormal basis, (48) and (50) are equivalent to the assertion that the Ψ_a or ρ_a matrices are diagonal in this basis.

Theorem 3. Presence of all information.

(i) All information about $\mathcal{S}_a$ is in $\mathcal{S}_b$ for $|\Psi\rangle \in \mathcal{H}_{ab}$ if and only if

$$\Psi_a = I_a/d_a; \tag{51}$$

i.e., $|\Psi\rangle$ is maximally entangled.

(ii) All information about $\mathcal{S}_a$ is in $\mathcal{S}_b$ for $\rho \in \hat{\mathcal{H}}_{ab}$ if and only if there are Hilbert spaces $\mathcal{H}_d$ and $\mathcal{H}_e$ whose tensor product is $\mathcal{H}_b$ or a subspace of $\mathcal{H}_b$, and ρ is of the form

$$\rho = \phi \otimes \rho_e \in \hat{\mathcal{H}}_{ad} \otimes \hat{\mathcal{H}}_e, \tag{52}$$

where $\phi=|\phi\rangle\langle\phi|$ projects on a fully entangled state $|\phi\rangle \in \mathcal{H}_{ad}$. This last implies (but is not implied by)

$$\rho_a = I_a/d_a. \tag{53}$$

(iii) All information about $\mathcal{S}_a$ is in $\mathcal{S}_b$ and all information about $\mathcal{S}_b$ is in $\mathcal{S}_a$, i.e., the two systems are informationally equivalent, if and only if the pre-probability is a fully entangled pure state: maximally entangled with $\mathcal{H}_a$ and $\mathcal{H}_b$ of the same dimension.

The utility of theorem 3 increases significantly through the existence of some (seemingly) rather weak conditions which imply that all information about $\mathcal{S}_a$ is in $\mathcal{S}_b$. To this end we need the following definition. Two decompositions $\mathcal{A}=\{A^j\}$ and $\overline{\mathcal{A}}=\{\overline{A}^k\}$ of I_a are *strongly incompatible* if there exists no projector P, apart from $P=0$ and $P=I_a$, that commutes with all the $\{A^j\}$ and all the $\{\overline{A}^k\}$. This is, for example, the case when $\mathcal{A}=\{|a^j\rangle\}$ and $\overline{\mathcal{A}}=\{|\overline{a}^l\rangle\}$ are two orthonormal bases for which

$$\langle a^j|\overline{a}^k\rangle \neq 0 \tag{54}$$

for all j and k, a condition which is fulfilled when the two bases are mutually unbiased, (4), but is obviously much weaker. Strong incompatibility is weaker still; it is possible for a number of the inner products in (54) to vanish provided a sufficient number are nonzero. Indeed, two decompositions can be strongly incompatible without all of the projectors or, in some cases, any of the projectors being onto pure states. We shall not pursue the matter further at this point, but instead state the desired result:

Theorem 4. Strong incompatibility. Let $\mathcal{A}$ and $\overline{\mathcal{A}}$ be two strongly incompatible decompositions of I_a, according to the preceding definition, and suppose that both the $\mathcal{A}$ and the $\overline{\mathcal{A}}$ information about $\mathcal{S}_a$ is in $\mathcal{S}_b$. Then

$$\rho_a = I_a/d_a, \tag{55}$$

and if, in addition, $\rho=\Psi$ is a pure state on $\mathcal{H}_{ab}$, then *all* information about $\mathcal{S}_a$ is in $\mathcal{S}_b$.

The following theorems refer to a tripartite system $\mathcal{S}_{abc}$.

Theorem 5. All information absent. If for $|\Psi\rangle \in \mathcal{H}_{abc}$ all information about $\mathcal{S}_a$ is absent from $\mathcal{S}_c$, there are Hilbert spaces $\mathcal{H}_d$ and $\mathcal{H}_e$ whose tensor product $\mathcal{H}_{de}$ is either $\mathcal{H}_b$ or a subspace of $\mathcal{H}_b$, and $|\Psi\rangle$ is of the form

$$|\Psi\rangle = |\chi\rangle \otimes |\psi\rangle \in \mathcal{H}_{ad} \otimes \mathcal{H}_{ce}. \tag{56}$$

Only if the support of Ψ_b is a proper subspace of $\mathcal{H}_b$ will $\mathcal{H}_{de}$ differ from $\mathcal{H}_b$, and in that case it can be identified with the subspace. The "hidden product" structure of (56) turns out to be a surprisingly useful tool.

Theorem 6. Particular information present for a pure state. For a pre-probability $|\Psi\rangle \in \mathcal{H}_{abc}$, the following holds:

(i) If $\mathcal{A}=\{|a^j\rangle\}$ is an orthonormal basis of $\mathcal{H}_a$, a necessary and sufficient condition for the $\mathcal{A}$ information to be present in $\mathcal{S}_b$ is that

$$\Psi_{ac} = \sum_j |a^j\rangle\langle a^j| \otimes \Gamma^j, \tag{57}$$

where the $\{\Gamma^j\}$ are (positive) operators on $\mathcal{H}_c$.

(ii) If for some decomposition $\mathcal{A}=\{A^k\}$ of I_a,

$$\Psi_{ac} = \sum_k A^k \otimes \overline{\Gamma}^k, \tag{58}$$

the $\mathcal{A}$ information about $\mathcal{S}_a$ is in $\mathcal{S}_b$, and if $\overline{\mathcal{A}}=\{\overline{A^l}\}$ is a compatible decomposition of I_a in the sense that all the $\{\overline{A^l}\}$ projectors commute with all the $\{A^k\}$ projectors, then the $\overline{\mathcal{A}}$ information is also present in $\mathcal{S}_b$. (In particular, $\overline{\mathcal{A}}$ may be an orthonormal basis in which the $\{A^k\}$ are diagonal.)

Theorem 7. Particular information present for a mixed state. Suppose that the $\mathcal{A}=\{|a^j\rangle\}$ information about $\mathcal{S}_a$ is in $\mathcal{S}_b$ for $\rho \in \hat{\mathcal{H}}_{abc}$. Then the following holds:

(i) The reduced density operator on $\mathcal{H}_{ac}$ is of the form

$$\rho_{ac} = \sum_j |a^j\rangle\langle a^j| \otimes \Gamma^j, \tag{59}$$

where the $\{\Gamma^j\}$ are (positive) operators on $\mathcal{H}_c$.

(ii) If $\overline{\mathcal{A}}=\{|\overline{a^k}\rangle\}$ is another orthonormal basis of $\mathcal{H}_a$, and $\mathcal{A}$ and $\overline{\mathcal{A}}$ are mutually unbiased, then no $\overline{\mathcal{A}}$ information is in $\mathcal{S}_c$, and

$$\mathrm{Tr}(\rho[\overline{a^k}]) = 1/d_a, \tag{60}$$

independent of k.

Theorem 8. No splitting theorem.

(i) If for $\rho \in \hat{\mathcal{H}}_{abc}$ all the information about $\mathcal{S}_a$ is in $\mathcal{S}_b$, then there is no information about $\mathcal{S}_a$ in $\mathcal{S}_c$,

$$\rho_{ac} = \rho_a \otimes \rho_c. \tag{61}$$

(ii) If for $|\Psi\rangle \in \mathcal{H}_{abc}$ all the information about $\mathcal{S}_a$ is in $\mathcal{S}_{bc}$ and none of it is in $\mathcal{S}_c$, then it is all in $\mathcal{S}_b$.

(iii) If for $\rho \in \hat{\mathcal{H}}_{abc}$ all the information about $\mathcal{S}_a$ is in $\mathcal{S}_{bc}$, but none of it is in $\mathcal{S}_c$, then the dimension of $\mathcal{H}_b$ is not less than that of $\mathcal{H}_a$.

Note that (iii) in this last theorem is a weaker result than (ii), for if all the $\mathcal{S}_a$ information is in $\mathcal{S}_b$, then by theorem 3 (ii) the dimension of $\mathcal{H}_b$ cannot be less than that of $\mathcal{H}_a$. The difference between (ii) and (iii) turns out be of some interest for understanding quantum codes, Sec. V C.

V. APPLICATIONS

A. Channels with mixed-state environment

There is no loss in generality in assuming the environment for a quantum channel is initially in a pure state, Fig. 2, provided the dimension d_e of $\mathcal{H}_e$ is at least d_a^2. The question has been raised [27,28] as to what channels can be produced using a smaller d_e—e.g., $d_e=d_a$—if one assumes an initial mixed state for the environment.

Such a channel can be modeled in the manner indicated in Fig. 4, with a "large" environment $\mathcal{S}_{ed}$ initially in a pure state $|\chi\rangle$, which when traced down to $\mathcal{H}_e$ yields the desired mixed-state density operator. The unitary transformation T maps $\mathcal{H}_{ve}$ onto $\mathcal{H}_{bc}$ to produce the analog of Fig. 3, where f has become the pair cd, and $|\phi\rangle$ is again the fully entangled state (11). The channel ket

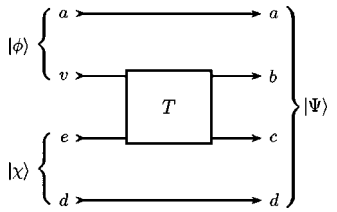

FIG. 4. Channel and channel ket $|\Psi\rangle$ for a mixed-state environment.

$$|\Psi\rangle = (I_a \otimes T \otimes I_d)(|\phi\rangle \otimes |\chi\rangle \tag{62}$$

is a pure state of $\mathcal{H}_{abcd}$.

This channel ket has the interesting property

$$\Psi_{ad} = \Psi_a \otimes \Psi_d, \tag{63}$$

which means that $\mathcal{S}_a$ and $\mathcal{S}_d$ are uncorrelated; no information about one is available in the other. It follows from the fact that the product state on the right side of (62) has this property, which is preserved during time development because the *unitary* operator T does not act on $\mathcal{H}_{ad}$. As a consequence, Ψ_{ad} (and therefore also its partial traces Ψ_a and Ψ_d) is independent of time. Note that this invariance is *not* true (in general) if T is not a unitary operator. The reason, in physical terms, is that a general map from $\mathcal{H}_{ve}$ to $\mathcal{H}_{bc}$ can be thought of as involving post selection, based upon some sort of joint measurement. Since $\mathcal{S}_a$ is correlated with $\mathcal{S}_z$ and $\mathcal{S}_d$ with $\mathcal{S}_e$ through the entangled initial states, the final state of affairs *conditioned on the outcome of such a measurement* may very well contain correlations between $\mathcal{S}_a$ and $\mathcal{S}_d$.

Not only is (63) a consequence of our model of a mixed-state environment, it comes close to being the very essence of the matter in light of theorem 5 applied to the tripartite $\mathcal{H}_a \otimes \mathcal{H}_{bc} \otimes \mathcal{H}_d$, for that tells us that $|\Psi\rangle$ necessarily involves a "hidden product" structure. What is required to bring that structure to light is a suitable unitary transformation, which is T in Fig. 4. To be sure, theorem 5 does not tell us that $|\phi\rangle$ shall be fully entangled—which suggests that the problem of a channel with a mixed-state environment is actually part of a more general information-theoretical question about entangled states on four-part systems, and exploring it from this perspective may be useful. In addition, our analysis suggests a close connection between such channels and properties of unitary transformations on bipartite systems.

B. CQ channels

The notion of a CQ or "classical-quantum" channel was introduced in [29] and has been the subject of some recent studies [7,8] in connection with entanglement-breaking channels, which were introduced in [30]. An entanglement-breaking channel may be defined as one in which the dynamical operator R in (21) is separable, in the standard way in which that term is applied to density operators (see, e.g.,

[31,32], Sec. 2.2.3 of [1]), and a CQ channel is a particular case of an entanglement-breaking channel in which R has the form

$$R = (1/d_a)\sum_j |a^j\rangle\langle a^j| \otimes B^j, \tag{64}$$

using a suitably chosen orthonormal basis $\mathcal{A}=\{|a^j\rangle\}$ for $\mathcal{H}_a$ and positive operators B^j of unit trace [to ensure (15)] on $\mathcal{H}_b$. The remarks which follow apply equally to a QC or "quantum-classical" channel, with the roles of a and b interchanged.

Introducing the channel ket $|\Psi\rangle \in \mathcal{H}_{abf}$ with $R=\Psi_{ab}$, (14), allows one to apply theorem 6 (i) in order to characterize a CQ channel as one in which there is an orthonormal basis for the channel entrance such that the information associated with this basis is *perfectly present in the environment* $\mathcal{S}_f$ at the later time. Note that such a characterization is *not* immediately obvious from considering the dynamical operator or, equivalently, the channel superoperator, for these are obtained by tracing out, thus ignoring, the environment, whereas the property which provides the simplest characterization in information-theoretic terms has very much to do with what information is available *in* the environment.

Using a channel ket in no way reduces the value of the insights provided in the studies cited above, nor does it supply (at least in any obvious sense) alternative tools for arriving at the technical results in those papers. But it does suggest a genuinely quantum-mechanical and information-theoretical description of what is "classical" (the C in CQ) about a CQ channel: namely, the environment provides perfect decoherence in a particular basis, as a consequence of which no information in any "complementary," which is to say mutually unbiased basis, is available at the channel exit, theorem 7 (ii). This is typical of what is generally referred to as "classical communication."

C. Information location in quantum codes

Quantum codes allow quantum information to be preserved against the effects of noise, whether due to interaction with the environment in a quantum communication setting, or imperfect gates in a quantum computer, and thus they have received a great deal of attention; for an introduction, see [33] and Chap. 10 of [19]. Our purpose here is not to contribute to the technical literature, but instead to point out how the basic operation of such a code can be understood in terms of the presence or absence of certain types of information in certain places.

The standard scenario is one in which the quantum information is embedded in a *code* $\mathcal{B}$, a K-dimensional subspace of the Hilbert space

$$\mathcal{H}_d = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \mathcal{H}_n \tag{65}$$

associated with n *carriers* of the coded information. The simplest situation is one in which $K=2=d_m$ for $1 \leqslant m \leqslant n$, but most of what we have to say applies more generally. Define the *security* s of the code to be the largest integer such that the encoded information is entirely absent from any set of s or fewer carriers [in a sense made precise in (68) below]. That is, an eavesdropper could learn nothing at all by carrying out arbitrary measurements on a set of s carriers, but could learn something from a suitable set of $s+1$ carriers. In the literature it is customary to refer to $s+1$ as the "distance" d of the code, using an analogy with classical codes in which d is the minimum Hamming distance between two code words. For a quantum code the notion of "distance" is somewhat obscure, as is the notion of code word, whereas s has a simple intuitive interpretation.

For analyzing the security and the error-correction properties of the code it is convenient to define a channel ket

$$\sqrt{K}|\Psi\rangle = \sum_j |a^j\rangle \otimes |b^j\rangle \in \mathcal{H}_a \otimes \mathcal{H}_d, \tag{66}$$

where the $\{|a^j\rangle\}$ form an orthonormal basis of the channel entrance $\mathcal{H}_a$, with $d_a=K$, and the $\{|b^j\rangle\}$ an orthonormal basis of the code subspace $\mathcal{B}$ with projector

$$B = \sum_j |b^j\rangle\langle b^j|. \tag{67}$$

Thus the encoding operation maps $\mathcal{H}_a$ onto $\mathcal{B}$. One can visualize $|\Psi\rangle$ using Fig. 3, but with $\mathcal{S}_b$ and $\mathcal{S}_f$ combined to form $\mathcal{S}_d$.

The security condition introduced earlier can now be stated as

$$\Psi_{au} = \Psi_a \otimes \Psi_u, \tag{68}$$

where u denotes any subset of s integers drawn from $\{1,2,\ldots,n\}$. Note that if (68) holds for such a set, it also holds for a smaller set; simply, take an appropriate partial trace of both sides. In view of theorem 1 (iii), (68) expresses precisely what we want to say by the security condition: if it is satisfied, no conceivable measurement on $\mathcal{S}_u$ will reveal anything about any sort of information in the channel entrance, whereas if it is *not* satisfied, *some* sort of information will be at least partially available to an eavesdropper.

From the definition (66) it is obvious that $\Psi_a=I_a/d_a$, so by theorem 3 (i) all information about $\mathcal{S}_a$ is in $\mathcal{S}_d$. Thus by theorem 8 (ii), if none of this information is in $\mathcal{S}_u$, it must be in the complement of this system in $\mathcal{S}_d$. That is, all the information about $\mathcal{S}_a$ is available in any collection of $n-s$ carriers; given any such a set, there will be a means of extracting or recovering the information from it even if the other carriers are ignored. This provides a preliminary understanding in information-theoretic terms of how a quantum error-correcting code functions, though some additional points remain to be dealt with.

In order to relate the security of the code to the discussion of error correction found in [34,35], it is helpful to introduce the following definition. An operator F on $\mathcal{H}_d$ will be said to have a *base* $\mathcal{S}_w$, where w is some subset of, and $\tilde{w}$ its complement in, $\{1,2,\ldots,n\}$, provided

$$F = F_w \otimes I_{\tilde{w}} \in \hat{\mathcal{H}}_w \otimes \hat{\mathcal{H}}_{\tilde{w}}, \tag{69}$$

and w is the *smallest* set for which F can be written in this form. The *size* of the base of F is the number of carriers in $\mathcal{S}_w$, the number of integers in w.

A code has security s when for every operator F with a base whose size does not exceed s it is the case that

$$BFB = b(F)B, \text{ or } \langle b^j|F|b^k\rangle = b(F)\,\delta_{jk}, \tag{70}$$

and s is the largest integer for which this is the case. Here $b(F)$ is a (complex) number that depends upon F, but not on j or k, and B is the projector in (67). The two equalities in (70) are equivalent because the second is simply the first expressed as a matrix when one extends $\{|b^j\rangle\}$ to an orthonormal basis of $\mathcal{H}_d$. To see that (70) is correct, first apply it in the case where F is a projector in $\hat{\mathcal{H}}_u$ for some u for which (68) holds, and use theorem 1 (i), with $\mathcal{H}_a$ in the theorem replaced by $\mathcal{H}_d$, and $\mathcal{H}_c$ by $\mathcal{H}_a$, to the decomposition $\{F, I_d - F\}$ of I_d. Any operator on $\mathcal{H}_u$ can be expressed as a linear combination of projectors, and hence by linearity, and the "if and only if" of theorem 1 (i), we arrive at the equivalence of (68) and (70) as statements that $\mathcal{H}_a$ and $\mathcal{H}_u$ are uncorrelated.

Now (70) is very similar to the necessary and sufficient condition

$$\langle b^j|K_l^\dagger K_m|b^k\rangle = b_{lm}\delta_{jk} \tag{71}$$

of [35] (in a slightly different notation) for a code to be able to correct a class of errors corresponding to the Kraus operators $\{K_l\}$ acting on the space $\mathcal{H}_d$. If these Kraus operators have a base no larger than t, then $F = K_l^\dagger K_m$ has a base that is no larger than $2t$, and we arrive at the condition

$$s = 2t \tag{72}$$

relating the security s to the maximum number of errors t which can be corrected. That is, a code which allows full recovery of information when t carriers are tampered with in any way, and *one does not know which* carriers have been affected, must allow full recovery when any *known* set of $s = 2t$ carriers have been tampered with; in the latter case the information will be recovered from the $n-2t$ remaining carriers. Thus the well-known five qubit code—see [34,36] and p. 469 of [19]—allows error recovery in the case of tampering with any one of the five carriers, but also if any two are stolen, since recovery is then carried out on the three that remain. (For a helpful discussion of this somewhat confusing point, see [37].)

The foregoing considerations make it possible to understand in information-theoretical terms the quantum Singleton lower bound

$$n \geqslant 4t + \log K/\log D \tag{73}$$

on the number of carriers, each assumed to have a Hilbert space of dimension D, in a quantum code [38]; also see p. 568 of [19]. One argues as follows. In order to correct up to t errors on unknown carriers the code must have a security of $s = 2t$: there is no information about $\mathcal{S}_a$ in any collection of $2t$ carriers, so by theorem 8 (ii) all the information about $\mathcal{S}_a$ is in any set of $n-2t$ carriers, as we noted earlier. But in a set of $n-2t$ carriers, no information can be present in a subset of $2t$ carriers, and thus by theorem 8 (iii), the Hilbert space of $n-2t-2t = n-4t$ carriers must have a dimension greater than or equal to $d_a = K$. This last assertion is equivalent to (73).

Note how in carrying out this argument it is essential to distinguish between a pure state pre-probability $|\Psi\rangle$ and a mixed state pre-probability ρ. The former is needed when using theorem 8 (ii) to infer the presence of all the information about $\mathcal{S}_a$ in any collection of $n-2t$ carriers, given that it is absent from any collection of size $2t$. However, these $n-2t$ carriers along with $\mathcal{S}_a$ form a system whose pre-probability is a density operator, and as a consequence we cannot use the fact that no $2t$ of *these* carriers contain information to infer that it must be present in a set of $n-4t$ carriers, something that is (at least in general) not true. By using theorem 8 (iii) instead of theorem 8 (ii), we correctly infer that the leftover collection of $n-4t$ carriers has a certain minimal size, *not* that it contains all the information.

The foregoing discussion focused on codes for which arbitrary errors in t or fewer carriers can be corrected. What of codes designed for the correction of errors of a more specific sort? Once again (71) applies, but only to a more specialized class of operators. Consider, for example, the three-qubit code which is adequate for bit-flip errors ([33] or p. 430 of [19]). Such errors can be represented by a Pauli σ^x on a single qubit, and what (71) is telling us is that *no $\mathcal{X}$ information about any pair of qubit code carriers can be present in $\mathcal{S}_a$*, where the sample space $\mathcal{X}$ is the orthonormal basis $\{|x_1^\pm x_2^\pm\rangle\}$ if the carriers are 1 and 2; here $|x^\pm\rangle$ are the eigenstates of σ^x.

The statement about absence at the channel input $\mathcal{S}_a$ of certain types of information about some of the carriers can be misinterpreted if thought of in terms of some backwards-in-time "influence" which the carriers exert on the channel input. Instead, keep in mind that the real issues have to do with statistical correlations between states of affairs at different times as represented in appropriate sample spaces or frameworks. Error recovery depends, of course, on information being present in appropriate locations, and quantum no-cloning (loosely speaking) allows us to connect the presence of information in one place with its absence someplace else. Presence and absence should always be thought of in terms of statistical correlations.

VI. CLASSIFICATION OF CHANNEL AND ENTANGLEMENT PROBLEMS

The fact that the properties of a quantum channel can be deduced from those of a channel ket, and likewise the properties of an entangled mixed state from those of a suitable purification, suggests the possibility of classifying these two types of quantum information problem in a single scheme based on pure states of a p-part system. Of course, for each p one should then introduce additional categories with some information-theoretical significance. The dimensions of the p subsystems are meaningful parameters, and other features, such as the "all" or "nothing" character of certain types of information, could assist in classifying particular cases. The motivation behind such a classification scheme is to have a useful way of comparing different types of experimental phenomena or theoretical models, one that may suggest analogies in instances where these are not immediately evident. Seeing how it relates to other problems does not, of course,

automatically provide a solution or even a better way of thinking about a particular question, but could in some cases suggest an alternative approach or allow the application of a different set of ideas.

There are two reasons for preferring a classification using entangled states to one based on channels. First, every channel problem (of the sort under discussion) maps in a simple and natural way to an entanglement problem, while the reverse is subject to some qualifications, as discussed in Sec. II E. Second, entangled states have a higher "conceptual symmetry;" for example, it is more natural to ask what happens if two subsystems of a bipartite system are interchanged than what will occur if the channel is, so to speak, operated in a time-reverse mode. The utility of pure states as against mixed states is less obvious, but the results in Sec. IV suggest that this may lead to a simpler classification using the location of quantum information, assuming that is a useful way to proceed.

Now let us consider some preliminary results. The Schmidt expansion for bipartite pure states provides a complete classification, up to local unitaries, for $p=2$, and the by now standard pure-state entanglement measure has proven itself a remarkably useful tool for their study. Noiseless quantum channels described by unitary time development fall in this category and correspond to fully entangled states.

The difficult problems start with $p=3$, which includes both mixed-state entanglement and the standard model for noisy quantum channels. Classifying the two together immediately raises the question of how various mixed-state entanglement measures [1,39] may be related to the many different types of quantum channel capacity that have been defined [1,40]. There is a brief discussion in Sec. 6.3.3 of [1], which notes that the equivalence of the (simple) quantum capacity and a one-way distillation entanglement measure was demonstrated in [34]. But we know of no systematic attempt to relate objects which ought to have a close connection. Or if they do not have a close connection, why is that?

If one further classifies $p=3$ problems according to the sizes of the subsystems, the obvious starting point is pure states of three qubits. Some one-qubit noisy channel problems fall in this category, as does the simplest cloning problem [41]. Leaving aside cases of a product state of one qubit with an entangled state of the other two, which in some sense belong to the $p=2$ class, the remaining states fall into two classes, "W" and "GHZ," under the equivalence generated by

$$|\Psi'\rangle = (A \otimes B \otimes C)|\Psi\rangle, \tag{74}$$

where A, B, and C are nonsingular operators [42]. This is a very interesting result which does not seem to have been generalized to larger subsystems. However, even for qubits it may not represent a complete classification scheme, for operations of the form (74) do not, in general, preserve all the properties that are of interest from an information-theoretic perspective (in which $|\Psi\rangle$ functions as a pre-probability).

The general one-qubit noisy quantum channel falls in the $p=3$ category, with two subsystems (entrance and exit of the channel) of dimension 2 and one (the environment) of dimension 4. A quite general description of such channels has been worked out in [43], and this work can and should be regarded as a significant step in classifying a large and important set of tripartite pure states. There are, on the other hand, entangled states which escape this classification (for the reasons explained in Sec. II E), and it would be interesting if the methods used in [43] could be extended to these as well.

A unitary transformation mapping a bipartite system to itself can be thought of as a $p=4$ problem, equivalent to a fully entangled state between two bipartite systems. In the case of two qubits such unitaries can be written down explicitly in terms of three real parameters [44], up to local unitaries on the individual qubits, and this provides a convenient description of an important class of $p=4$ pure states in which each subsystem has dimension 2. Beyond this very little seems to be known at present about the four qubit problem. A one-qubit channel with a mixed-state environment falls in this category, as explained in Sec. V A. The *entanglement of purification* introduced in [45] is an example of a $p=4$ problem not limited to qubits, as is the general problem of a channel corresponding to a mixed-state environment.

As noted in Sec. V C, a quantum code with n carriers falls in the $p=n+1$ category of states for which there is an absence of correlations between one particular subsystem (the channel entrance) and various collections of other subsystems. Relating quantum codes to more general problems of multipartite entanglement is an interesting and challenging problem [20].

VII. CONCLUSION

A. Summary

The fundamental idea underlying the duality discussed in Sec. II is that the correlation of events at different times that characterize a quantum channel are "the same thing" as the correlation of properties of an entangled quantum system at different points in space. At the mathematical level the correspondence is expressed by a simple partial transpose (23) that carries the dynamical density operator R, into the transition operator Q representing the channel superoperator. In physical terms the duality says that the correlations which express the location of information about one quantum system in another are of basically the same nature, whether they refer to properties of a single system at two different times or to two different systems at the same time. This is well established in classical information theory, where the same tools are used for both circumstances, and it works equally well in quantum systems given appropriate sample spaces or frameworks, as explained in Sec. III.

The nonclassical "peculiarities" of quantum information emerge when one uses a single pre-probability, either a pure state or a density operator, or their counterparts for a quantum channel, to generate probability distributions and thus correlations for a variety of different, incompatible frameworks (sample spaces). It is here that "no-cloning" plays a central role, and the eight all-or-nothing theorems of Sec. IV are intended to make that idea more precise and more widely applicable. While the theorems are expressed in entanglement language, the duality allows their immediate applica-

tion to quantum channels. In many cases the results are more precise (and in others their derivation is easier) when the pre-probability is a pure rather than a mixed state, which in the case of a quantum channel means a channel ket rather than a dynamical operator. This suggests that channel kets are a useful tool for analyzing the properties of noisy quantum channels, and the applications in Sec. V bear this out. Whether pure states are equally advantageous for classifying entangled states and quantum channels in a single scheme remains to be demonstrated, but the preliminary results in Sec. VI are encouraging.

B. Open questions

The eight all-or-nothing theorems of Sec. IV provide a useful first step in describing in a systematic way how information can be divided up or spread out over an entangled quantum system. But one suspects there remains much more to be said, both about bipartite and tripartite systems, and also about systems with $p \geqslant 4$ parts. In addition, every qualitative theorem of the type found in Sec. IV ought to be the limiting case of one or perhaps several *quantitative* theorems in which the complete presence or absence of information is replaced by quantitative measures—Shannon entropies are an obvious, but not the unique possibility—and constraints are provided in the form or rigorous inequalities, or perhaps even equalities, if one is lucky. While some ideas of this sort have been put forward—e.g., [46,47]—a great deal more could be done.

To be sure, several entanglement measures have been proposed for bipartite mixed states [1,39], and to a lesser extent for systems with $p \geqslant 3$ parts; see [20] and the references given there. But rarely do these have a specific information-theoretical content or basis, and it is an open question whether, and if so how, they can be understood in such terms—i.e., related to statistical correlations forming part of a consistent probabilistic description of a quantum system. To be sure, entanglement measures can be useful even if they have no connection to information theory, but if there is such a connection, understanding what it is could be a useful contribution to the subject.

Discussions of quantum channel capacities seem better anchored in an information-theoretic framework than those concerning entanglement measures, though perhaps more thought should be given as to how to translate "classical," which occurs rather frequently in such discussions, into appropriate quantum mechanical terms; we no longer live in a classical world. Relating these capacities to entanglement measures seems at present a largely open question, and answering it could make a valuable contribution understanding both entanglement and noisy channels.

The task of classifying entangled pure states of p-part systems in the manner suggested in Sec. VI can be regarded as complete for $p=2$, but for $p=3$ it has just begun, and very little is known about $p \geqslant 4$ systems apart from work on quantum codes. Extending the latter to more general entangled states could make a significant contribution to our understanding of multipartite entanglement, which at present is quite limited.

ACKNOWLEDGMENTS

I thank L. Yu for providing some of the references and for a critical reading of the text. The research described here received support from the National Science Foundation through Grant No. PHY-0139974.

APPENDIX A: PROOFS OF THEOREMS IN SEC. IV

Theorem 1 (i). Expand $|\Psi\rangle$ in Schmidt form,

$$|\Psi\rangle = \sum_j \sqrt{q_j}|a^j\rangle \otimes |c^j\rangle, \tag{A1}$$

and let J be the collection of j values for which $q_j > 0$. For the $\{A^l\}$ information to be absent from $\mathcal{S}_c$, it must be the case [see (36)] that

$$\mathrm{Tr}_a(\Psi A^l) = \sum_{j,k} \sqrt{p_j p_k}\langle a^j|A^l|a^k\rangle(|c^j\rangle\langle c^k|) \tag{A2}$$

is proportional to

$$\Psi_c = \sum_{j\in J} p_j[c^j], \tag{A3}$$

which means that

$$\langle a^j|A^l|a^k\rangle = a_l\delta_{jk} \tag{A4}$$

for all j and k in J. This is the same as (43), which is the same as (42).

*Theorem 1 (ii).*Expand $|\Psi\rangle$ in the orthonormal basis $\{|a^j\rangle\}$ [see (6)]:

$$|\Psi\rangle = \sum_j |a^j\rangle \otimes |\gamma^j\rangle. \tag{A5}$$

The requirement that no information about $\{|a^j\rangle\}$ be in $\mathcal{S}_c$ means that all the $|\gamma^j\rangle$ must be proportional to each other, and thus to a single ket $|\gamma\rangle$, which means that $|\Psi\rangle$ is of the form (44).

Theorem 1 (iii). The "if" part is obvious. To prove that (45) holds if all information about $\mathcal{S}_a$ is absent from $\mathcal{S}_c$, let $\{|a^j\rangle\}$ and $\{|c^l\rangle\}$ be bases in which ρ_a and ρ_c are diagonal,

$$\rho_a = \sum_j p_j[a^j], \quad \rho_c = \sum_l q_l[c^l], \tag{A6}$$

and write

$$\rho = \sum_{jk}\sum_{lm} \langle a^j c^l|\rho|a^k c^m\rangle(|a^j\rangle\langle a^k| \otimes |c^l\rangle\langle c^m|). \tag{A7}$$

The absence of all information implies that

$$\mathrm{Tr}_a(A\rho) = \langle A\rangle\rho_c \tag{A8}$$

for *any* operator $A \in \hat{\mathcal{H}}_a$—see (36), and note that the collection of all projectors is an operator basis for $\hat{\mathcal{H}}_a$. Insert $A = |a^k\rangle\langle a^j|$ in (A8), and use (A7) to evaluate the left side and (A6) the right. The conclusion is that

$$\langle a^j c^l|\rho|a^k c^m\rangle = p_j q_l \delta_{jk}\delta_{lm}, \tag{A9}$$

which is (45).

Theorem 2 (i). If the $\mathcal{A}$ information is present in $\mathcal{S}_b$ (35) implies that

$$\mathrm{Tr}(A^l B^m) = \mathrm{Tr}_b(\Lambda^l B^m) = \delta_{lm}\,\mathrm{Tr}_b(\Lambda^l), \tag{A10}$$

since $\langle A^l\rangle = \mathrm{Tr}_b(\Lambda^l)$. If P and Q are positive operators such that $\mathrm{Tr}(PQ)=0$, then $PQ=0$. Using this and the fact that the B^m are projectors, so that $\Lambda^l = \Lambda^l B^m + \Lambda^l(I_b - B^m)$, one sees that (A10) implies that

$$B^m \Lambda^l B^m = \delta_{lm}\Lambda^l, \tag{A11}$$

and (46) is a consequence of $B^l B^m = \delta_{lm} B^l$. Conversely, (46) implies that one can simultaneously diagonalize the collection $\{\Lambda^l\}$ and choose the B^l projecting onto appropriate blocks in such a way that (A11), and therefore (A10) and (35) are satisfied.

Theorem 2 (ii). Choose an orthonormal basis $\{|a^j\rangle\}$ in which the $\{A^l\}$ are diagonal, and expand $|\Psi\rangle$ in this basis, (49), without assuming it is in Schmidt form. Then

$$\Psi_a = \sum_{jk} \langle\beta^k|\beta^j\rangle |a^j\rangle\langle a^k| \tag{A12}$$

and

$$\Lambda^l = \sum_{j\in J_l} |\beta^j\rangle\langle\beta^j|, \tag{A13}$$

where J_l is the collection of j values for which $A^l|a^j\rangle = |a^j\rangle$. One can show that Ψ_a commutes with all the A^l if and only if $\langle\beta^k|\beta^j\rangle = 0$ whenever $j\in J_l$ and $k\in J_m$ with $m\neq l$. But this last is equivalent to (46). If the A^l project onto one-dimensional states, then $\langle\beta^k|\beta^j\rangle = 0$ for $j\neq k$, so (49) is in Schmidt form.

Theorem 2 (iii). Purify ρ to a ket $|\Psi\rangle \in \mathcal{H}_{abc}$. Use the fact that the $\mathcal{A}$ information is present in $\mathcal{S}_{bc}$, and apply part (ii) of the theorem with $\mathcal{S}_{bc}$ in place of $\mathcal{S}_b$ to infer that $\Psi_a = \rho_a$ commutes with all the A^l.

Theorem 3. Part (i) is an immediate consequence of 2(ii), for it is only multiples of the identity that commute with all projectors. The proofs of (ii) and (iii) are given below, following that of theorem 8.

Theorem 4. By theorem 2, ρ_a or Ψ_a must commute with all the $\{A^j\}$ and all the $\{\overline{A^k}\}$, and must therefore, by the definition of strong incompatibility, be multiples of I_a. The final statement is a consequence of theorem 3.

Theorem 5. Let $\{|a^j\rangle\}$ and $\{|c^k\rangle\}$ be orthonormal bases of $\mathcal{H}_a$ and $\mathcal{H}_c$ which diagonalize Ψ_a and Ψ_c,

$$\Psi_a = \sum_j p_j[a^j], \quad \Psi_b = \sum_k q_k[c^k], \tag{A14}$$

and expand $|\Psi\rangle$ in these bases:

$$|\Psi\rangle = \sum_{jk} |a^j\rangle \otimes |\beta^{jk}\rangle \otimes |c^k\rangle. \tag{A15}$$

The condition $\Psi = \Psi_a \otimes \Psi_c$ expressing the absence of all $\mathcal{S}_a$ information from $\mathcal{S}_b$, theorem 1 (iii), implies that

$$\langle\beta^{j'k'}|\beta^{jk}\rangle = p_j q_k \delta_{jj'}\delta_{kk'}. \tag{A16}$$

Therefore if we restrict our attention to the $j\in J$ and $k\in K$ for which $p_j > 0$ and $q_k > 0$, we can construct an orthonormal set

$$|b^{jk}\rangle = |\beta^{jk}\rangle / \sqrt{p_j q_k} \tag{A17}$$

of kets in $\mathcal{H}_b$, and rewrite (A15) in the form

$$|\Psi\rangle = \sum_{jk} \sqrt{p_j q_k}\,|a^j\rangle \otimes |b^{jk}\rangle \otimes |c^k\rangle. \tag{A18}$$

The spaces $\mathcal{H}_d$ and $\mathcal{H}_e$ are then *defined* as having orthonormal bases $\{|d^j\rangle\}$ and $\{|e^k\rangle\}$ such that

$$|b^{jk}\rangle = |d^j\rangle \otimes |e^k\rangle, \tag{A19}$$

so that $|\Psi\rangle$ is of the form (56) with

$$|\chi\rangle = \sum_j \sqrt{p_j}|a^j\rangle \otimes |d^j\rangle, \quad |\psi\rangle = \sum_k \sqrt{q_k}|c^k\rangle \otimes |e^k\rangle. \tag{A20}$$

Theorem 6 (i). Expand $|\Psi\rangle$ in the orthonormal basis $\{|a^j\rangle\}$

$$|\Psi\rangle = \sum_j |a^j\rangle \otimes |\zeta^j\rangle, \tag{A21}$$

with $|\zeta^j\rangle \in \mathcal{H}_{bc}$, and write

$$\Psi = \sum_{jk} |a^j\rangle\langle a^k| \otimes \zeta^{jk}, \quad \zeta^{jk} := |\zeta^j\rangle\langle\zeta^k|. \tag{A22}$$

If the $\{|a^j\rangle\}$ information is in $\mathcal{S}_b$, then, by theorem 2 (i),

$$\zeta_b^{jj}\zeta_b^{kk} = 0 \quad \text{for } j\neq k, \tag{A23}$$

where, following our usual notation, $\zeta_b^{jj} = \mathrm{Tr}_c(\zeta^{jj})$. Now apply (B3) in Appendix B, with a replaced by b, b replaced by c, $|e\rangle = |\zeta^j\rangle$, and $|g\rangle = |\zeta^k\rangle$, to conclude that (A23) holds if and only if

$$\zeta_c^{jk} = 0 \quad \text{for } j\neq k. \tag{A24}$$

[Note that $\mathrm{Tr}_c(CC^\dagger) = 0$ implies that $C=0$.] But (A24) inserted in (A22) implies (57) with $\Gamma^j = \zeta_c^{jj}$. Conversely, (57) implies (A24), which implies (A23), which, using theorem 2 (i), implies that the $\{|a^j\rangle\}$ information is in $\mathcal{S}_b$.

Theorem 6 (ii). Let $\{|a^j\rangle\}$ be any basis in which the A^k in (58) are diagonal. Then (57) is a consequence of (58): simply write each A^k as a sum of a suitable collection of $[a^j]$. Thus by (i), the $\{|a^j\rangle\}$ and *a fortiori* the $\{A^k\}$ information is in $\mathcal{S}_b$. For a compatible decomposition $\overline{\mathcal{A}} = \{\overline{A^l}\}$, use a basis $\{|a^j\rangle\}$ in which both these and the $\{A^k\}$ are diagonal.

Theorem 7 (i). Purify ρ to $|\Psi\rangle \in \mathcal{H}_{abcd}$, and apply theorem 6 (i) with c replaced by cd to conclude that Ψ_{acd} is of the form (57) with operators Γ^j on $\mathcal{H}_{cd}$. Now trace both sides over $\mathcal{H}_d$ to get the equivalent of (59).

Theorem 7 (ii). Multiply both sides of (59) by $[\overline{a^k}]$. First trace over $\mathcal{H}_a$ and use the definition of mutually unbiased bases in (4) to conclude that the resulting operator (on $\mathcal{H}_c$) does not depend on k, so the $\{|\overline{a^k}\rangle\}$ information is absent from $\mathcal{S}_c$ according to the definition in Sec. III C; see the

comment following (36). Next, trace over $\mathcal{H}_c$ to get (60).

Theorem 8 (i). Given an arbitrary orthonormal basis $\overline{\mathcal{A}}$ of $\mathcal{H}_a$, one can always find another basis $\mathcal{A}$ with $\overline{\mathcal{A}}$ and $\mathcal{A}$ mutually unbiased. As the $\mathcal{A}$ information is, by assumption, in $\mathcal{S}_b$, the $\overline{\mathcal{A}}$ information cannot be in $\mathcal{S}_c$, by theorem 7 (ii).

Theorem 8 (ii). All the information about $\mathcal{S}_a$ is in $\mathcal{S}_{bc}$, so $\Psi_a = I_a/d_a$ by theorem 3 (i). But as there is no information about $\mathcal{S}_a$ in $\mathcal{S}_c$, theorem 5 tells us $|\Psi\rangle$ is of the form (56), with $\chi_a = \Psi_a = I_a/d_a$, and therefore, once again invoking theorem 3 (i), all the information about $\mathcal{S}_a$ is in $\mathcal{S}_b$.

Theorem 8 (iii). [The following argument is from p. 569 of [19], where it is ascribed to [48], and it makes use of some well-known properties of the von Neumann entropy

$$S(\rho) = -\operatorname{Tr}(\rho \log \rho); \tag{A25}$$

see, e.g., pp. 513 and 515 of [19].] Upon purifying ρ to $|\Psi\rangle \in \mathcal{H}_{abcd}$ one finds that

$$S(\Psi_a) + S(\Psi_c) = S(\Psi_{ac}) = S(\Psi_{bd}) \leqslant S(\Psi_b) + S(\Psi_d). \tag{A26}$$

The first equality is a consequence of the absence of information about $\mathcal{S}_a$ in $\mathcal{S}_c$, thus $\Psi_{ac} = \Psi_a \otimes \Psi_c$ by theorem 1 (iii). The second equality reflects the fact that $|\Psi\rangle$ is a pure state on $\mathcal{H}_{ac} \otimes \mathcal{H}_{bd}$, and the final inequality is a standard result for a density operator on a tensor product. Since all information about $\mathcal{S}_a$ is in $\mathcal{S}_{bc}$, it must be absent from $\mathcal{S}_d$ by part (i) of this theorem, so we can interchange the roles of $\mathcal{S}_c$ and $\mathcal{S}_d$ in (A26) to obtain

$$S(\Psi_a) + S(\Psi_d) \leqslant S(\Psi_b) + S(\Psi_c), \tag{A27}$$

and by adding this to (A26) arrive at

$$S(\Psi_a) \leqslant S(\Psi_b). \tag{A28}$$

By theorem 3 (i) (replace b by bcd) we know that $\Psi_a = I_a/d_a$, so the left side of (A28) is $\log d_a$ and as the right side cannot exceed $\log d_b$, therefore $d_b \geqslant d_a$.

Theorem 3 (ii) and *Theorem 3 (iii).* Purify ρ to $|\Psi\rangle \in \mathcal{H}_{abc}$. If all information about $\mathcal{S}_a$ is in $\mathcal{S}_b$ (for ρ and for $|\Psi\rangle$), then by theorem 8 (i) there is none in $\mathcal{S}_c$, so by theorem 5 $|\Psi\rangle$ has the product structure of (56), where in addition $|\chi\rangle$ must be maximally (fully) entangled, so we arrive at (52). If, on the other hand, (52) is correct, then $\phi_a = I_a/d_a$, and all the information about $\mathcal{S}_a$ is in $\mathcal{S}_d$ and, therefore, in $\mathcal{S}_b$. To prove theorem 3 (iii), note that if $|\Psi\rangle$ is given by (56) and d_e is 2 or more, I_e has a nontrivial decomposition, and the corresponding information obviously cannot be in $\mathcal{S}_a$. Thus if all the information about $\mathcal{S}_b$ is in $\mathcal{S}_a$, it is the case that $d_e = 1$ and $\mathcal{H}_{de}$ is the same as $\mathcal{H}_d$, and the latter is the same as $\mathcal{H}_b$, for were it a proper subspace, Ψ_b would not be proportional to I_b.

APPENDIX B: FOUR ENTANGLED KETS

Let

$$D^{ef} = |e\rangle\langle f|, \quad D_a^{ef} = \operatorname{Tr}_b(D^{ef}), \quad D_b^{ef} = \operatorname{Tr}_a(D^{ef}) \tag{B1}$$

denote the dyad and its partial traces for two kets $|e\rangle$ and $|f\rangle$ on $\mathcal{H}_{ab} = \mathcal{H}_a \otimes \mathcal{H}_b$.

Theorem. Let $|e\rangle$, $|f\rangle$, $|g\rangle$, $|h\rangle$ be any four kets on $\mathcal{H}_{ab}$. Then

$$\operatorname{Tr}_a(D_a^{ef} D_a^{gh}) = \operatorname{Tr}_b(D_b^{eh} D_b^{gf}). \tag{B2}$$

In particular, if $|f\rangle = |e\rangle$ and $|h\rangle = |g\rangle$, then

$$\operatorname{Tr}_a(D_a^{ee} D_a^{gg}) = \operatorname{Tr}_b(D_b^{eg} D_b^{ge}). \tag{B3}$$

Proof. Let $\{|a^j\rangle\}$ be a fixed orthonormal basis of $\mathcal{H}_a$, and expand each ket in the form

$$|w\rangle = \sum_j |a^j\rangle \otimes |w^j\rangle. \tag{B4}$$

Direct calculation shows that the left and right sides of (2) are both equal to

$$\sum_{jk} \langle f^k|e^j\rangle\langle h^j|g^k\rangle. \tag{B5}$$

[1] M. Keyl, Phys. Rep. **369**, 431 (2002).
[2] F. Verstraete and H. Verschelde, e-print quant-ph/0202124.
[3] K. Życzkowski and I. Bengtsson, Open Syst. Inf. Dyn. **3**, 42 (2004).
[4] P. Arrighi and C. Patricot, Ann. Phys. (N.Y.) **311**, 26 (2004).
[5] J. I. Cirac, W. Dür, B. Kraus, and M. Lewenstein, Phys. Rev. Lett. **86**, 544 (2001).
[6] M. Reimpell and R. F. Werner, e-print quant-ph/0307138.
[7] M. Horodecki, P. W. Shor, and M. B. Ruskai, Rev. Math. Phys. **15**, 629 (2003).
[8] M. B. Ruskai, Rev. Math. Phys. **15**, 643 (2003).
[9] M. Hamada, Int. J. Quantum Inf. **1**, 443 (2003).
[10] W. Dür, J. I. Cirac, and P. Horodecki, Phys. Rev. Lett. **93**, 020503 (2004).
[11] A. Jamiołkowski, Rep. Math. Phys. **3**, 275 (1972).
[12] M.-D. Choi, Linear Algebr. Appl. **10**, 285 (1975).
[13] Č. Brukner and A. Zeilinger, Phys. Rev. Lett. **83**, 3354 (1999).
[14] Č. Brukner and A. Zeilinger, Phys. Rev. A **63**, 022113 (2001).
[15] D. Deutsch and P. Hayden, Proc. R. Soc. London, Ser. A **456**, 1759 (2000).
[16] R. B. Griffiths, Phys. Rev. A **66**, 012311 (2002).
[17] C. G. Timpson, Stud. Hist. Philos. Mod. Phys. **34**, 441 (2003).
[18] A. Duwell, Stud. Hist. Philos. Mod. Phys. **34**, 479 (2003).
[19] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, England, 2000).
[20] A. J. Scott, Phys. Rev. A **69**, 052330 (2004).
[21] W. Feller, *An Introduction to Probability Theory and its Applications*, 3rd ed. (Wiley, New York, 1968), Vol. 1.
[22] S. M. Ross, *Introduction to Probability Models*, 7th ed. (Academic Press, San Diego, 2000).
[23] M. H. DeGroot and M. J. Schervish, *Probability and Statistics*,

3rd ed. (Addison-Wesley, Boston, 2002).
[24] R. B. Griffiths, *Consistent Quantum Theory* (Cambridge University Press, Cambridge, UK, 2002).
[25] P. Mittelstaedt, *The Interpretation of Quantum Mechanics and the Measurement Process* (Cambridge University Press, Cambridge, UK, 1998).
[26] W. M. de Muynck, *Foundations of Quantum Mechanics, an Empiricist Approach* (Kluwer Academic, Dordrecht, 2002).
[27] B. M. Terhal, I. L. Chuang, D. P. DiVincenzo, M. Grassl, and J. A. Smolin, Phys. Rev. A **60**, 881 (1999).
[28] C. Zalka and E. Rieffel, J. Math. Phys. **43**, 4376 (2002).
[29] A. S. Holevo, Russ. Math. Surveys **53**, 1295 (1999).
[30] P. W. Shor, J. Math. Phys. **43**, 4334 (2002).
[31] M. Lewenstein, D. Bruß, J. I. Cirac, B. Kraus, M. Kuś, J. Samsonowicz, A. Sanpera, and R. Tarrach, J. Mod. Opt. **47**, 2841 (2000).
[32] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Lett. A **283**, 1 (2001).
[33] A. W. Steane, in *Introduction to Quantum Computation and Information*, edited by H.-K. Lo, S. Popescu, and T. Spiller (World Scientific, Singapore, 1998), pp. 184–212.
[34] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).
[35] E. Knill and R. Laflamme, Phys. Rev. A **55**, 900 (1997).
[36] R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, Phys. Rev. Lett. **77**, 198 (1996).
[37] M. Grassl, T. Beth, and T. Pellizzari, Phys. Rev. A **56**, 33 (1997).
[38] E. M. Rains, IEEE Trans. Inf. Theory **45**, 1827 (1999).
[39] M. J. Donald, M. Horodecki, and O. Rudolph, J. Math. Phys. **43**, 4252 (2002).
[40] D. Kretschmann and R. F. Werner, New J. Phys. **6**, 26 (2004).
[41] C.-S. Niu and R. B. Griffiths, Phys. Rev. A **60**, 2764 (1999).
[42] W. Dür, G. Vidal, and J. I. Cirac, Phys. Rev. A **62**, 062314 (2000).
[43] M. B. Ruskai, S. Szarek, and E. Werner, Linear Algebr. Appl. **347**, 159 (2002).
[44] B. Kraus and J. I. Cirac, Phys. Rev. A **63**, 062309 (2001).
[45] B. M. Terhal, M. Horodecki, D. W. Leung, and D. P. DiVincenzo, J. Math. Phys. **43**, 4286 (2002).
[46] M. J. W. Hall, Phys. Rev. Lett. **74**, 3307 (1995).
[47] M. J. W. Hall, Phys. Rev. A **55**, 100 (1997).
[48] J. Preskill, www.theory.caltech.edu/people/preskill/ph229/ (1998).