A framework for a quantum mechanical information theory is introduced that is based entirely on density operators, and gives rise to a unified description of classical correlation and quantum entanglement. Unlike in classical (Shannon) information theory, quantum (von Neumann) conditional entropies can be negative when considering quantum entangled systems, a fact related to quantum non-separability. The possibility that negative (virtual) information can be carried by entangled particles suggests a consistent interpretation of quantum informational processes.

Quantum information theory is a new field with potential implications for the conceptual foundations of quantum mechanics. It appears to be the basis for a proper understanding of the emerging fields of quantum computation \[1\], quantum communication, and quantum cryptography \[2\]. Although fundamental results on the quantum noiseless coding theorem \[3\] and the capacity of quantum noisy channels \[4\] have been obtained recently, quantum information is still puzzling in many respects. This is especially true for quantum teleportation and superdense coding, two purely quantum communication schemes devised recently \[5\]. Indeed, these processes can be understood in a consistent way by exploiting a fundamental difference between Shannon theory \[6\] and the capacity on density operators, and gives rise to a unified description of classical correlation and quantum entanglement. Unlike in classical (Shannon) information theory, quantum (von Neumann) conditional entropies can be negative when considering quantum entangled systems, a fact related to quantum non-separability. The possibility that negative (virtual) information can be carried by entangled particles suggests a consistent interpretation of quantum informational processes.

Quantum information theory can include Shannon theory as a special case but describes quantum entanglement as well, thereby providing a unified treatment of classical and quantum information. To be specific, let us consider a composite system consisting of two (classical or quantum) variables, \(A\) and \(B\), and outline in parallel the classical and our quantum information-theoretic treatment of it. In classical information theory, we define the Shannon entropy of a quantum source \(A\) as

\[
H(A) = - \sum_a p(a) \log_2 p(a) ,
\]

where the variable \(A\) takes on value \(a\) with probability \(p(a)\). It is interpreted as the uncertainty about \(A\) [an analogous definition holds for \(H(B)\)]. The quantum analog is the von Neumann entropy \(S(\rho_A)\) of a quantum source \(A\) described by the density operator \(\rho_A\),

\[
S(A) = - \text{Tr}[\rho_A \log_2 \rho_A] ,
\]

where \(\text{Tr}_A\) denotes the trace over the degrees of freedom associated with \(A\). The von Neumann entropy reduces to a Shannon entropy if \(\rho_A\) is a mixed state composed of orthogonal quantum states. The combined classical system \(AB\) is characterized by a joint probability \(p(a,b)\), and therefore by a joint entropy \(H(AB) = - \sum_{a,b} p(a,b) \log_2 p(a,b)\) which is the uncertainty about the entire system. The quantum definition \(S(AB) = - \text{Tr}[\rho_{AB} \log_2 \rho_{AB}]\), a function of the density operator of the combined system \(\rho_{AB}\), is immediate. The classical probabilities observe \(p(a) = \sum_b p(a,b)\) and \(p(b) = \sum_a p(a,b)\); analogously, \(\rho_A = \text{Tr}_B[\rho_{AB}]\) and \(\rho_B = \text{Tr}_A[\rho_{AB}]\). Note that \(\text{Tr}_A\) and \(\text{Tr}_B\) stand for partial traces, while \(\text{Tr}\) is the trace over the joint Hilbert space. The classical conditional entropy is defined as \(H(A|B) = H(AB) - H(B) = - \sum_{a,b} p(a,b) \log_2 p(a|b)\), where \(p(a|b) = p(a,b)/p(b)\) is the probability of \(a\) conditional on \(b\); \(H(A|B)\) characterizes the remaining uncertainty about the source \(A\) when \(B\) is known.
Here, we propose that the correspondence between classical and quantum constructions can be extended by defining the von Neumann conditional entropy

$$S(A|B) = -\text{Tr}[\rho_{AB} \log_2 \rho_{A|B}]$$  \hspace{1cm} (3)$$

based on a conditional “amplitude” operator

$$\rho_{A|B} = \exp_2(-\sigma_{AB}) = \lim_{n \to \infty} \left[ \rho_{AB}^{1/n} (1_A \otimes \rho_B)^{-1/n} \right]^n,$$  \hspace{1cm} (4)$$

where $\sigma_{AB} = 1_A \otimes \log_2 \rho_B - \log_2 \rho_{AB}$. $1_A$ being the unit matrix in the Hilbert space of $A$, and $\otimes$ the tensor product in the joint Hilbert space. Eq. (4) is a quantum generalization of the conditional probability $p(a|b)$ and reduces to it in the classical limit (diagonal $\rho_{AB}$). In general, $\rho_{A|B}$ is a positive Hermitian operator on the joint Hilbert space [just as $\rho_{AB}$ is a function of $a$ and $b$], whose spectrum is invariant under frame changes of the product form $U_A \otimes U_B$ on $\rho_{AB}$. We refer to $\rho_{A|B}$ as an amplitude operator to emphasize that it retains the quantum phases, in contrast to $p(a|b)$. It is defined on the support of $\rho_{AB}$ since the latter is included in the support of $1_A \otimes \rho_B$, so that $S(A|B)$ is well-defined. Indeed, inserting Eq. (4) in Eq. (3) results in $S(A|B) = S(AB) - S(B)$, just as for Shannon entropies. However, $\rho_{A|B}$ is not a density operator as its eigenvalues can exceed 1, in which case the operator $\sigma_{AB}$ is not positive semi-definite, in contrast to its classical counterpart. It is precisely for this reason that the von Neumann conditional entropy can be negative. In Shannon theory, $H(A|B)$ is always non-negative, reflecting that the classical entropy of a composite system $AB$ cannot be lower than the entropy of any subsystem $A$ or $B$, i.e., $\max[H(A), H(B)] \leq H(AB)$. Quantum entropies on the other hand are known to be non-monotonic, i.e., $S(A) > S(AB)$ or $S(B) > S(AB)$ is possible when $A$ and $B$ are quantum entangled subsystems (see, e.g., [8]). The above operator-based formalism naturally accounts for this non-monotonicity.

The appearance of nonclassical ($> 1$) eigenvalues of $\rho_{A|B}$ and nonclassical ($< 0$) conditional entropies can be related to quantum non-separability and the violation of entropic Bell inequalities, as we show in [8]. It can be proven that any separable state $\rho_{AB} = \sum w_i |\psi_i^A\rangle \otimes |\psi_i^B\rangle$ (with $0 \leq w_i \leq 1$ and $\sum w_i = 1$) is such that $\sigma_{AB} \geq 0$, and is therefore associated with $S(A|B) = \text{Tr}[\rho_{AB} \sigma_{AB}] \geq 0$. Thus, all the eigenvalues of $\rho_{A|B}$ (and $\rho_{B|A}$) are $\leq 1$ for any convex mixture of product states, which implies that it is a necessary condition for separability in a Hilbert space of arbitrary dimensions. A negative conditional entropy implies that $\rho_{A|B} \not\leq 1$ while the converse is not true, so that $S(A|B) \geq 0$ is a weaker necessary condition. As an example, this criterion can be applied to two spin-1/2 particles in a Werner state, i.e., a mixture of a singlet fraction $x$ and a random fraction $(1-x)$. A simple calculation shows that $\rho_{A|B}$ admits three eigenvalues equal to $(1-x)/2$, and a fourth equal to $(1 + 3x)/2$. The above separability criterion is thus fulfilled when the latter does not exceed 1, that is for $x \leq 1/3$. Thus, for this particular case (in fact, for any mixture of Bell states), our condition is simply equal to Peres’ [11] and then happens to be also sufficient. Numerical evidence suggests, however, that it is weaker than Peres’ in general. This will be investigated elsewhere.

In Shannon information theory, we define the mutual (or correlation) entropy $H(A:B) = H(A) - H(AB)$ as the decrease of the entropy of $A$ due to the knowledge of $B$, resulting in $H(A:B) = H(A) + H(B) - H(AB) = -\sum_{a,b} p(a,b) \log_2 p(a,b)$, where $p(a:b) = p(a)p(b)/p(a,b)$ is a mutual probability. The mutual entropy corresponds to the information gained about $A$ by measuring $B$. It is symmetric, i.e., $H(A:B) = H(B:A)$, and can be viewed as the amount of entropy shared by $A$ and $B$. Also, $H(AB) \geq 0$, as the entropy of $A$ can only be reduced through the knowledge of $B$. As we now show, a quantum mechanical extension of this concept can be straightforwardly defined, i.e., we can construct a von Neumann mutual entropy $S(AB)$ based on a mutual amplitude operator

$$\rho_{A:B} = \lim_{n \to \infty} \left[ (\rho_A \otimes \rho_B)^{1/n} \rho_{AB}^{-1/n} \right]^n,$$  \hspace{1cm} (5)$$

generalizing the mutual probability $p(a:b)$. Again, the associated quantum mutual entropy

$$S(A:B) = -\text{Tr}[\rho_{AB} \log_2 \rho_{A:B}]$$  \hspace{1cm} (6)$$

is invariant under frame changes of the product form, and can be written as $S(AB) = S(A) + S(B) - S(AB)$. The quantum mutual entropy $S(AB)$ is a natural extension of $H(A:B)$ which measures quantum as well as classical correlations, and reduces to it for diagonal density matrices (probability distributions). It does not discriminate purely quantum entanglement from correlation, however, but rather unifies their information-theoretic description. While $S(A:B) \geq 0$ just as its classical counterpart, it can exceed the entire uncertainty of the source ensemble, namely

$$S(A:B) \leq 2\min[S(A), S(B)],$$  \hspace{1cm} (7)$$

as implied by the Araki-Lieb inequality (see [8]). This precisely occurs for quantum entangled subsystems, and is forbidden classically due to the non-negativity of the Shannon conditional entropy, i.e., $H(A:B) \leq \min[H(A), H(B)]$.

The above unified treatment of classical and quantum correlation is illustrated by considering three limiting cases and their associated entropy Venn diagrams as defined in Fig. [1]. For independent quantum systems (I), one has $\rho_{AB} = \rho_A \otimes \rho_B$, so that $\rho_{A|B} = \rho_A \otimes 1_B$ [analogous to $p(a|b) = p(a)$] and $S(AB) = S(A)$, thereby saturating the lower bound $S(AB) = 0$ as in the classical case. This is illustrated in Fig. [1b] for two quantum bits (qubits), i.e., if systems $A$ and $B$ belong to
a 2-state Hilbert space. For two (maximally) classically correlated systems (II), the classical upper bound 
\[ S(A:B) = \min[S(A), S(B)] \] is saturated. The range between the classical and quantum upper bounds corresponds to a purely quantum (classically forbidden) regime, namely quantum entanglement. The quantum upper bound, Eq. (7), is saturated for maximally entangled systems such as EPR pairs (III): the conditional entropies are negative, while \( S(A:B) \) exceeds the value which defines 100% correlations, so that entanglement might be viewed as supercorrelation.

FIG. 1. (a) Entropy Venn diagram for a bipartite system \( AB \). (b) Diagram for two qubits with \( S(A) = S(B) = 1 \): (I) independent 50/50 mixtures of states \(|0\rangle \) and \(|1\rangle \); (II) maximally classically (anti)correlated qubits, i.e., a 50/50 mixture of \(|01\rangle \) and \(|10\rangle \); (III) fully entangled EPR state with wavefunction \( 2^{-1/2}(|01\rangle - |10\rangle) \), or any Bell state in general.

The information-theoretic formalism discussed here should thus be regarded as an extension of a classical formalism beyond its original range to identify the signature of quantum dynamics. As shown in further work (see [12] and references therein), it can be successfully applied to multipartite entangled systems by extending Shannon construction [e.g., defining the quantum conditional mutual entropy \( S(A:B|C) \)]. Most of the standard concepts of Shannon theory then find an intuitive quantum analog, which results in a convenient framework for analyzing quantum channels or error correction, for example. Here, we restrict ourselves to sketching an information-theoretic description of quantum teleportation and superdense coding relying on this framework. We argue that a conservative picture of information flow in these processes, consistent with unitarity, is only possible if the existence of negative conditional entropies is recognized [13].

The third diagram in Fig. (I) exhibits how the negative conditional entropy of a member of an EPR pair, \( S(A|B) = -1 \), is balanced by the (unconditional) entropy \( S(B) = 1 \) of its partner. As entropy can be viewed as “latent” information, this suggests characterizing the members of an EPR pair by a virtual information content of ±1 bit [14], a concept that turns out to be very fruitful when analyzing the information flow. To distinguish these maximally entangled qubits, we call them cbits [15]. In this picture inspired by particle physics, the ebit (\( e \)) and anti-ebit (\( \bar{e} \)) can be viewed as virtual conjugate particles, and the preparation of an EPR pair appears as the creation of an “\( e\bar{e} \) pair” from the entropy “vacuum” \( S(e\bar{e}) = 0 \). The central point is that the virtual information content of an ebit can be revealed (i.e., converted to real information) in interactions with qubits and classical bits (cbits), as we shall see. In quantum teleportation (see Fig. (II)), an unknown qubit (\( q \)) is transported with perfect fidelity through the transmission of two cbits (\( 2c \)), after the sender and the receiver have shared an \( e\bar{e} \) pair. The sender performs a joint measurement (\( M \)) of the qubit and the ebit in the 2-particle Bell basis (i.e., 4 orthogonal maximally entangled 2-particle states), thereby generating 2 cbits. The receiver reconstructs the qubit from the 2 cbits by applying to one of 4 possible unitary transforms (\( U \)) in the 1-particle Hilbert space. Only if the (virtual) information content of \( e \) and \( \bar{e} \) is taken into account properly (1 bit and −1 bit, respectively) can the information flow be conserved through the (M) and (U) stages. More specifically, if \( q \) is maximally entangled with an external reference (not represented here), the entropy conservation rule for (M) is

\[ S(2c) = S(qe) = S(q) + S(e) = 1 + 1 = 2, \]  

since \( q \) and \( e \) are initially independent, while the conditional entropy of the remaining particle is \( S(\bar{e}qe) = -1 \). At (U), we have

\[ S(q') = S(qe\bar{e}) = S(qe) + S(\bar{e}qe) = 2 - 1 = 1, \]  

where \( q' \) is the outgoing qubit [16]. In the particle physics language, if the \( \bar{e} \) is replaced by an \( e \) going backwards in time (see diagram in lower half of Fig. (II)), the whole process is formally equivalent to the transmission of 1 qubit via 2 cbits, but with the additional burden of 1 ebit. A violation of causality by the \( \bar{e} \) is prevented by the fact that its information content cannot be revealed without the presence of the 2 cbits, which travel causally.
In superdense coding (see Fig. 2b), 2 cbits are apparently transported via 1 qubit (a 2-state particle!). Our analysis suggests that the negative information content of the τ is exploited by the sender so that 2 cbits can be packed into a single qubit via the unitary transform (U). Subsequently, the receiver performs a joint measurement (M) of the qubit and ebit in the 2-particle Bell basis, thereby recovering the two encoded cbits. The entropy conservation rule for (U) is

\[ S(q|e) = S(2c|\tau|e) = S(2c) + S(\tau|e) = 2 - 1 = 1, \]

(10)
since 2c and τ are initially independent, while the unconditional entropy of the remaining particle is \( S(e) = 1 \). At (M), we have

\[ S(2c') = S(q)e = S(q|e) + S(e) = 1 + 1 = 2, \]

(11)
where 2c' are the outgoing cbits. The cbits 2c left with the sender must be ignored from an informational point of view, as they are correlated with 2c'. The factor 2 that is apparently gained here is directly connected to the factor 2 in Eq. (10), and originates from the fact that \( S(2c|q|e) = 2 \), that is, the additional information (about 2c) conveyed by q when knowing e exceeds the entropy of q by a factor of 2. It is the maximum compression allowed by quantum mechanics (no other quantum communication scheme could be more efficient). However, this compression is only apparent since the information flow is conserved in both (U) and (M) stages if a qubit–antiqubit picture is used. Formally, the 2 cbits are distributed over two particles (see lower half of Fig. 2b) although the ebit does not appear during the considered period of transmission as it is sent backwards in time.

We have proposed a quantum mechanical extension of Shannon information theory which incorporates entanglement via negative conditional entropies. This analysis suggests the possibility that a qubit (the fundamental quantum of information) could have an analogously defined anti-qubit (a quantum of negative information), formally equivalent to a qubit traveling backwards in time as anticipated in [4]. The concept that negative virtual information can be carried by entangled particles provides interesting insight into the information flow in quantum communication processes such as quantum teleportation and superdense coding. Furthermore, this leads us to conjecture that these processes can be recast into reactions involving information quanta (e, q, c, τ) described by diagrams, much like particle physics reactions. The (M) and (U) operations then correspond to vertices (see lower half of Fig. 2b) which describe the dual information-conserving processes (U) \( 2c + \tau \rightarrow q \) and (M) \( q + e \rightarrow 2c \), summarizing Eqs. (8–11). As the cτ contain no readable information, they cannot appear in the external lines of diagrams, just like the virtual particles of quantum field theory.

We thank Steve Koonin, Asher Peres, Barry Simon, and Armin Uhlmann for enlightening discussions, as well as an anonymous referee. This work has been funded by NSF grants PHY 94-12818 and PHY 94-20470, and by DARPA/ARO through the QUIC Program (#DAAH04-96-1-3086).

[1] D.P. DiVincenzo, Science 270, 255 (1995); S. Lloyd, ibid. 261, 1569 (1993); ibid. 273, 1073 (1996)
[2] A. Ekert, Nature 358, 14 (1992); C.H. Bennett, G. Brassard, N.D. Mermin, Phys. Rev. Lett. 68, 557 (1992).
[3] B. Schumacher, Phys. Rev. A 51, 2738 (1995); R. Jozsa and B. Schumacher, J. Mod. Opt. 41, 2343 (1994).
[4] S. Lloyd, Phys. Rev. A 55, 1613 (1997).
[5] C.H. Bennett and S.J. Wiesner, Phys. Rev. Lett. 69, 2881 (1992); C.H. Bennett et al., Phys. Rev. Lett. 70, 1895 (1993).
[6] C. E. Shannon and W. Weaver, The mathematical theory of communication (University of Illinois Press, 1949).
[7] This is more obvious if \( \rho_{AB} \) and \( 1_A \otimes \rho_B \) commute, in which case the Trotter decomposition in Eq. (1) reduces to \( \rho_{AB} = \rho_{AB}(1_A \otimes \rho_B)^{-1} \).
[8] A. Wehrl, Rev. Mod. Phys. 50, 221 (1978).
[9] N. J. Cerf and C. Adami, Phys. Rev. A 55, 3371 (1997).
[10] The concavity of \( S(AB) \) in \( \rho_{AB} \) also implies that \( S(A|B) \geq \sum_i w_i S(\rho_{iA}) \geq 0 \), so that \( S(A|B) \geq 0 \) is a necessary condition for separability.
[11] A. Peres, Phys. Rev. Lett. 77, 1413 (1996).
[12] N. J. Cerf and C. Adami, in Proc. of 4th Workshop on Physics and Computation; to appear in Physica D (1997).
[13] The details of this analysis will be reported elsewhere.
[14] We define information as virtual if its extraction would violate causality. In other words, virtual information cannot give rise to superluminal communication between the two particles of an EPR pair. Whether positive or negative virtual information is carried by a member of an EPR pair is not an intrinsic property of the particle, but rather can only be assigned after the information dynamics is completed. In teleportation or superdense coding, positive (virtual) information is carried by that particle which is measured. If both particles are measured (e.g., in Bell experiments), no information transfer can take place and no virtual information content can be assigned to either particle.
[15] The term ebit was coined in C.H. Bennett et al., Phys. Rev. A 53, 2046 (1996) to describe the “shared entanglement” of two particles, as “an undirected informational resource” giving the ability to transmit information (although it is not information in itself). Here, an ebit (antiebit) is that member of an EPR pair which carries positive (negative) virtual information.
[16] In this description, the cbits remain unamplified. The original teleportation scheme can be understood in a similar manner. Then, the leftover cbits after (U) are in a product state with \( q' \), so that the remaining entropy \( S(2c) = 2 \) should not be included in the entropy balance. It simply originates from the measurement of two \( |0 \rangle + |1 \rangle \) states in the standard basis, as revealed by a more detailed analysis.