Quantum Information Theory Tutorial

Mark M. Wilde

Hearne Institute for Theoretical Physics,
Department of Physics and Astronomy,
Center for Computation and Technology,
Louisiana State University,
Baton Rouge, Louisiana, USA

mwilde@lsu.edu

Reference: *Quantum Information Theory*
published by Cambridge University Press (2nd edition forthcoming)

July 10, 2016, ISIT 2016, Barcelona, Spain

Main questions

- What are the ultimate limitations on communication imposed by physical laws?
- What are methods for achieving these limits?
- To address these questions, we need to consider quantum mechanics, and so we are naturally led to an intersection of information theory and quantum mechanics called *quantum information theory*
- What is different about quantum and "classical" information theory?
- What tasks can we achieve with quantum mechanics that we cannot without it? (long list: Bell inequalities, super-dense coding, teleportation, data locking, data hiding, quantum cryptography, etc.)

Prehistory of quantum information theory

- 1927 Heisenberg uncertainty principle
- 1935 Einstein–Podolsky–Rosen paper questioning compatibility of uncertainty principle and phenomenon of quantum entanglement / 1964 Bell's theorem as an answer / 2009 Berta *et al.* entropic uncertainty relation as another answer
- 1932 von Neumann quantum entropy / 1962 Umegaki quantum relative entropy / 1973 Lieb–Ruskai strong subadditivity of quantum entropy / 1975 Lindblad data-processing for quantum relative entropy
- 1970s theory of quantum measurements and similarity measures for quantum states — Helstrom, Holevo (Shannon Award 2016), Ozawa, Bures, Uhlmann, etc.

Information theory

- 1948 — Shannon set the foundations of information theory, defining notions like data compression and channel capacity and giving answers in terms of entropy and mutual information, resp.
- Shannon considered only classical physics (without quantum effects)
- His work (and that of others) ultimately led to questions like:
- "How do quantum effects enhance communication capacity?"
- "How do quantum effects enhance communication security?"
- "What are some quantum communication tasks that do not have a counterpart in the classical world?"

Tutorial overview

- Quantum states and channels
- Fundamental protocols: Bell / CHSH game, entanglement distribution, super-dense coding, quantum teleportation
- Distance measures for quantum states
- Information measures
- Quantum data compression
- Communication over quantum channels

Review of quantum formalism

- Let's begin by reviewing some basics of quantum information
- All we need to start understanding quantum information is how to represent states and evolutions of quantum systems.
- We do this by using *density matrices* and *quantum channels*.
- These ideas extend how we represent states of a classical system with probability distributions and evolutions of these classical systems with classical channels (conditional probability distributions).
- We'll find that the set of quantum states contains all classical states and is far richer, which is suggestive of why we can do things that are not possible in classical information theory.

Quantum states

Quantum states

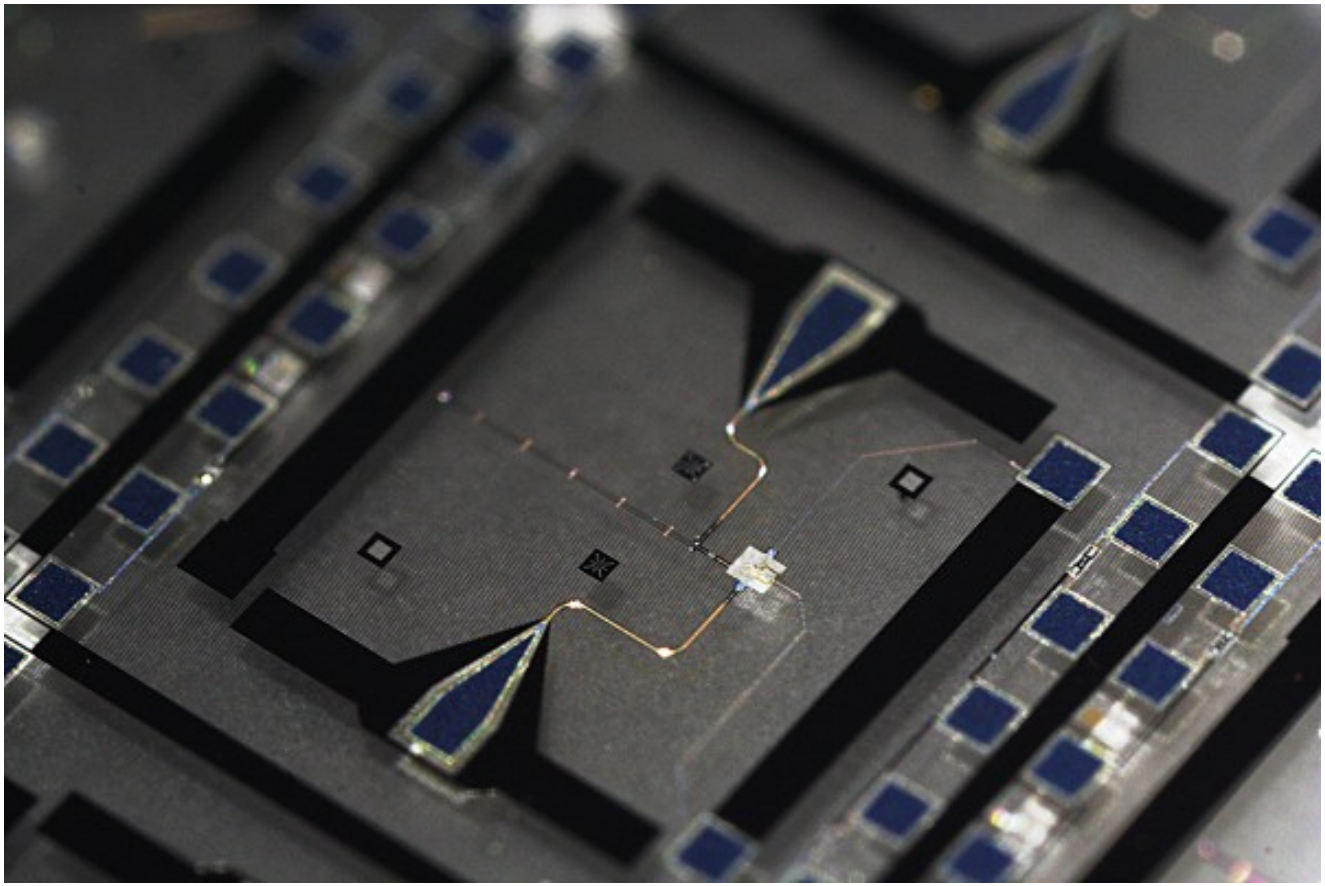
- The state of a quantum system is given by a square matrix called *the density matrix*, usually denoted by ρ, σ, τ, ω, etc.
- It should be positive semi-definite and have trace equal to one. That is, all of its eigenvalues should be non-negative and sum up to one. We write these conditions symbolically as $\rho \geq 0$ and $\text{Tr}\{\rho\} = 1$. Can abbreviate more simply as $\rho \in \mathcal{D}(\mathcal{H})$, to be read as "$\rho$ is in the set of density matrices."
- The dimension of the matrix indicates the number of distinguishable states of the quantum system.
- For example, a physical *qubit* is a quantum system with dimension two. A classical bit, which has two distinguishable states, can be embedded into a qubit.

Interpretation of density matrix

- The density matrix, in addition to a description of an experimental procedure, is all that one requires to predict the (probabilistic) outcomes of a given experiment performed on a quantum system.

- It is a generalization of (and subsumes) a probability distribution, which describes the state of a classical system. All probability distributions can be embedded into a quantum state by placing the entries along the diagonal of the density matrix.

Let's talk about qubits...



Superconducting phase qubit from
http://web.physics.ucsb.edu/~martinisgroup/photos.shtml,
taken by Erik Lucero

Examples of quantum states

- Let

$$|0\rangle \equiv \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \langle 0| \equiv \begin{bmatrix} 1 & 0 \end{bmatrix},$$

$$\text{so that density matrix } \rho_0 \equiv |0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

- Similarly, let

$$|1\rangle \equiv \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \langle 1| \equiv \begin{bmatrix} 0 & 1 \end{bmatrix},$$

$$\text{so that density matrix } \rho_1 \equiv |1\rangle\langle 1| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

- Then $\rho_0\rho_1 = 0$. The states ρ_0 and ρ_1 are orthogonal to each other, and, physically, this means that they are perfectly distinguishable.
- What we have done here is to embed classical bits into quantum bits. We can think of ρ_0 as '0' and ρ_1 as '1.'

Mixtures of quantum states

- Any probabilistic mixture of two quantum states is also a quantum state. That is, for $\sigma_0, \sigma_1 \in \mathcal{D}(\mathcal{H})$ and $p \in [0,1]$, we have

$$p\sigma_0 + (1-p)\sigma_1 \in \mathcal{D}(\mathcal{H}).$$

  The set of density matrices is thus convex.

- For our classical example, we find

$$\begin{aligned} p\rho_0 + (1-p)\rho_1 &= p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1| \\ &= \begin{bmatrix} p & 0 \\ 0 & 1-p \end{bmatrix}. \end{aligned}$$

- This is the statement that probabilistic classical bits can be embedded into quantum bits, and the probabilities appear along the diagonal of the matrix. Can we have other kinds of quantum states?

Superpositions of quantum states

- Construct the following unit vector as a *superposition* of $|0\rangle$ and $|1\rangle$:

$$|\psi\rangle \equiv \alpha|0\rangle + \beta|1\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix},$$

  where $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$. Note that $\langle\psi| = \begin{bmatrix} \alpha^* & \beta^* \end{bmatrix}$ and $\langle\varphi|\psi\rangle$ denotes the inner product of vectors $|\psi\rangle$ and $|\varphi\rangle$.
- The unit vector $|\psi\rangle$ leads to the following quantum state:

$$|\psi\rangle\langle\psi| = \begin{bmatrix} |\alpha|^2 & \alpha\beta^* \\ \beta\alpha^* & |\beta|^2 \end{bmatrix}.$$

- The difference between this quantum state and the others we've considered so far is the presence of off-diagonal elements in the density matrix (called *quantum coherences*).
- This state is physically distinct from $\begin{bmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{bmatrix}$.

Bloch sphere

- We can visualize the state of a qubit using the *Bloch sphere*. To see this, consider the Pauli matrices

$$I \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad X \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Y \equiv \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad Z \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

  The last three Pauli matrices have eigenvalues ± 1 and eigenvectors:

$$|\pm\rangle \equiv \frac{1}{\sqrt{2}}\left(|0\rangle \pm |1\rangle\right), \qquad |\pm_Y\rangle \equiv \frac{1}{\sqrt{2}}\left(|0\rangle \pm i|1\rangle\right), \qquad |0\rangle, |1\rangle.$$

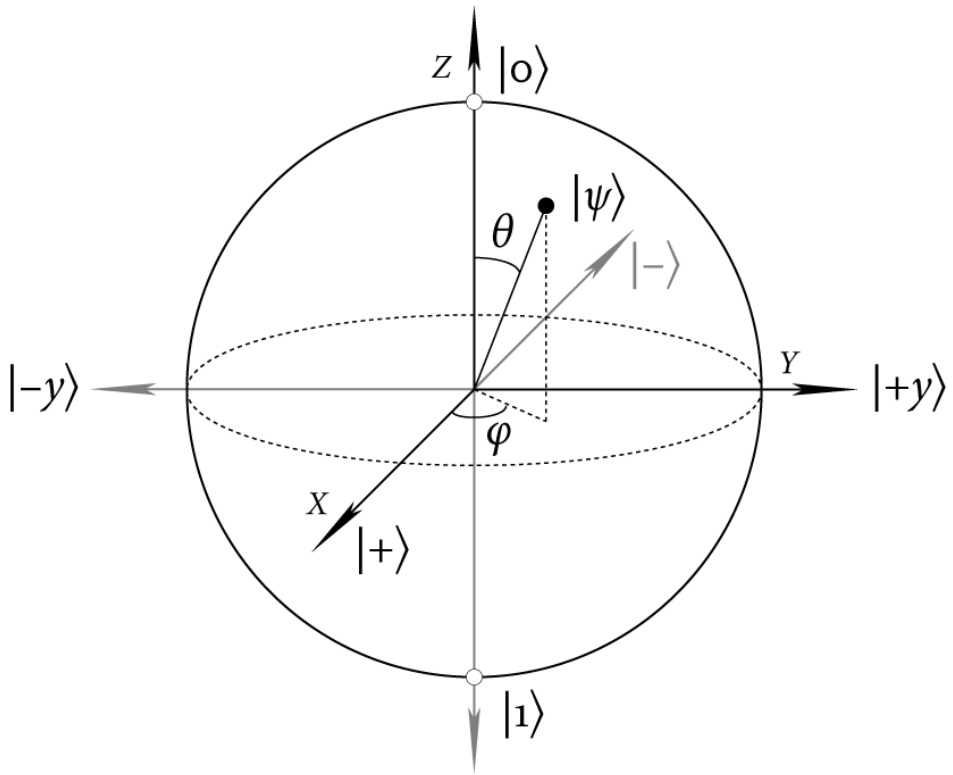
- We can write the density matrix ρ of a qubit in terms of three parameters r_x, r_y, and r_z:

$$\rho = \frac{1}{2}\left(I + r_x X + r_y Y + r_z Z\right),$$

  where $r_x^2 + r_y^2 + r_z^2 \leq 1$, which is the equation of a unit sphere in $\mathbb{R}^3$.

Bloch sphere

We can visualize the state of a qubit using the Bloch sphere:



- The *maximally mixed state* $I/2 = (|0\rangle\langle 0| + |1\rangle\langle 1|)/2$ is at the center.
- Classical states are on the line going from $|0\rangle$ to $|1\rangle$.
- A quantum state is *pure* if it is on the surface and otherwise *mixed*.

Higher dimensional quantum systems

- A density matrix can have dimension ≥ 2 and can be written as

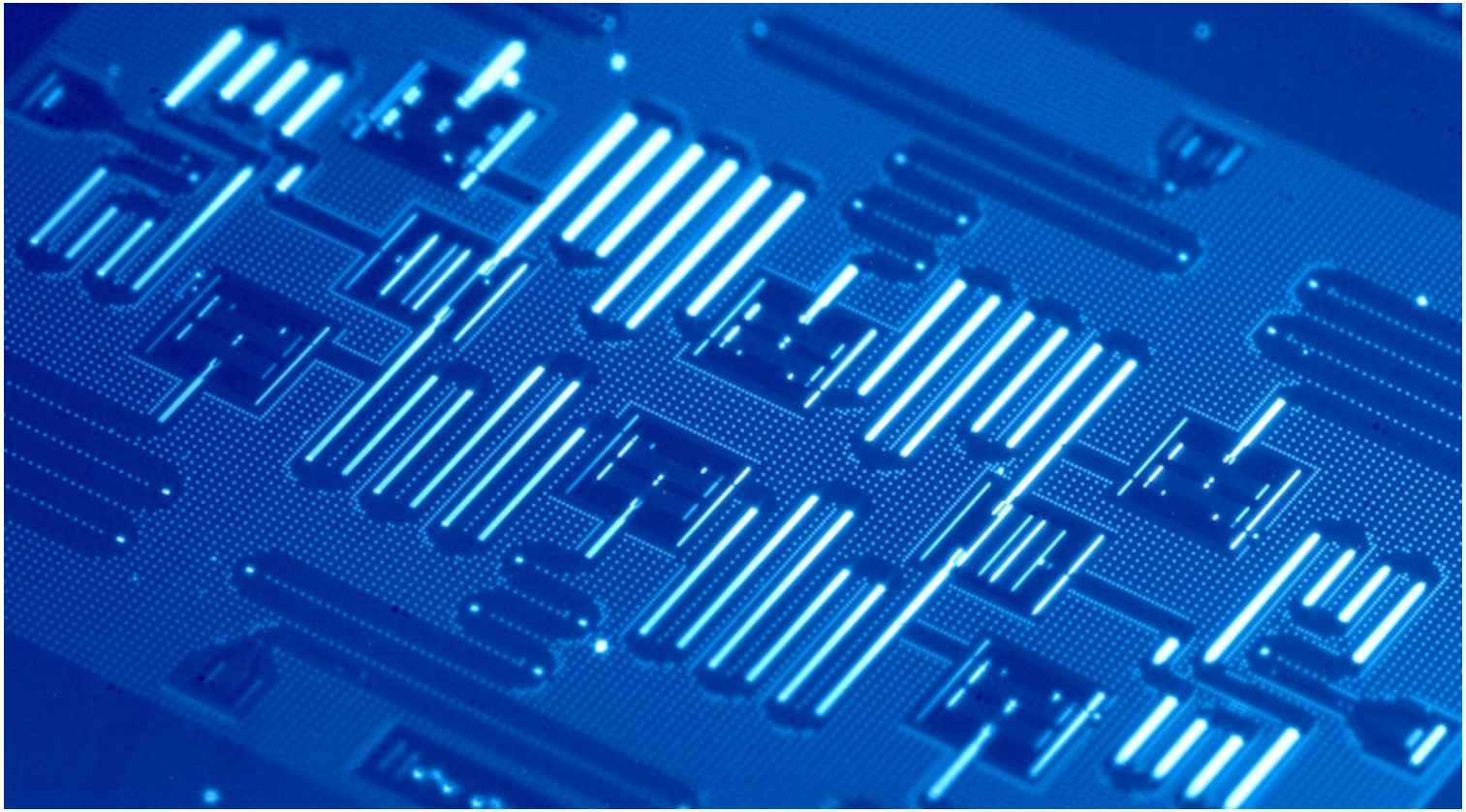
$$\rho = \sum_{i,j} \rho^{i,j} |i\rangle\langle j|,$$

  where $\{|i\rangle \equiv e_i\}$ is the standard basis and $\rho^{i,j}$ are the matrix elements.
- Since every density matrix is positive semi-definite and has trace equal to one, it has a spectral decomposition as

$$\rho = \sum_x p_X(x) |\phi_x\rangle\langle\phi_x|,$$

  where $\{p_X(x)\}$ are the non-negative eigenvalues, summing to one, and $\{|\phi_x\rangle\}$ is a set of orthonormal eigenvectors.
- A density matrix ρ is *pure* if there exists a unit vector $|\psi\rangle$ such that $\rho = |\psi\rangle\langle\psi|$ and otherwise it is *mixed*.

Multiple qubits...



IBM five-qubit universal quantum computer (released May 2016)

Composite quantum systems

- Just as we need more than one bit for information processing to become interesting, quantum information really only becomes interesting when multiple quantum systems can interact.
- We use Cartesian product to represent state of two or more bits:

$$(0,0),\ (0,1),\ (1,0),\ (1,1) \in \mathbb{Z}_2 \times \mathbb{Z}_2,$$

  but Cartesian product is not rich enough to capture quantum states.
- Consider that before we constructed a quantum state from a superposition of two unit vectors. So we could imagine constructing a quantum state from a superposition of vectors as

$$\alpha|0,0\rangle + \beta|0,1\rangle + \gamma|1,0\rangle + \delta|1,1\rangle,$$

  where $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$. But what are $|i,j\rangle$?

Tensor product

- We use the tensor product to represent multiple quantum systems.
- For vectors, it is defined as

$$\begin{bmatrix} a_1 \\ b_1 \end{bmatrix} \otimes \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \equiv \begin{bmatrix} a_1 \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \\ b_1 \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} a_1 a_2 \\ a_1 b_2 \\ b_1 a_2 \\ b_1 b_2 \end{bmatrix}.$$

- So, then with this definition, we have

$$|\varphi\rangle \equiv \alpha|0\rangle \otimes |0\rangle + \beta|0\rangle \otimes |1\rangle + \gamma|1\rangle \otimes |0\rangle + \delta|1\rangle \otimes |1\rangle = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix},$$

which leads to a two-qubit density operator $|\varphi\rangle\langle\varphi|$.

System labels

- Often it can be helpful to write system labels, which indicate which qubit Alice possesses and which Bob possesses:

$$|\varphi\rangle_{AB} \equiv \alpha|0\rangle_A \otimes |0\rangle_B + \beta|0\rangle_A \otimes |1\rangle_B + \gamma|1\rangle_A \otimes |0\rangle_B + \delta|1\rangle_A \otimes |1\rangle_B.$$

  We can also write the labels on the two-qubit density operator:

$$|\varphi\rangle\langle\varphi|_{AB}.$$

- Often we abbreviate the above more simply as

$$\alpha|00\rangle_{AB} + \beta|01\rangle_{AB} + \gamma|10\rangle_{AB} + \delta|11\rangle_{AB}.$$

Tensor product for matrices

- For matrices K and L, the tensor product is defined in a similar way:

$$
\begin{aligned}
K \otimes L &\equiv \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} \otimes \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} \\
&\equiv \begin{bmatrix} k_{11} \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} & k_{12} \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} \\ k_{21} \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} & k_{22} \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} \end{bmatrix} \\
&= \begin{bmatrix} k_{11}l_{11} & k_{11}l_{12} & k_{12}l_{11} & k_{12}l_{12} \\ k_{11}l_{21} & k_{11}l_{22} & k_{12}l_{21} & k_{12}l_{22} \\ k_{21}l_{11} & k_{21}l_{12} & k_{22}l_{11} & k_{22}l_{12} \\ k_{21}l_{21} & k_{21}l_{22} & k_{22}l_{21} & k_{22}l_{22} \end{bmatrix}.
\end{aligned}
$$

Properties of tensor product

- For vectors:

$$z(|\phi\rangle \otimes |\psi\rangle) = (z|\phi\rangle) \otimes |\psi\rangle = |\phi\rangle \otimes (z|\psi\rangle),$$
$$(|\phi_1\rangle + |\phi_2\rangle) \otimes |\psi\rangle = |\phi_1\rangle \otimes |\psi\rangle + |\phi_2\rangle \otimes |\psi\rangle,$$
$$|\phi\rangle \otimes (|\psi_1\rangle + |\psi_2\rangle) = |\phi\rangle \otimes |\psi_1\rangle + |\phi\rangle \otimes |\psi_2\rangle.$$

- Matrices acting on vectors:

$$(K \otimes L)(|\phi\rangle \otimes |\psi\rangle) = K|\phi\rangle \otimes L|\psi\rangle,$$
$$(K \otimes L)\left(\sum_x \lambda_x |\phi_x\rangle \otimes |\psi_x\rangle\right) = \sum_x \lambda_x K|\phi_x\rangle \otimes L|\psi_x\rangle,$$
$$\left(\sum_x \mu_x K_x \otimes L_x\right)(|\phi\rangle \otimes |\psi\rangle) = \sum_x \mu_x K_x|\phi\rangle \otimes L_x|\psi\rangle.$$

- Inner product: $(\langle\phi_1| \otimes \langle\psi_1|)(|\phi_2\rangle \otimes |\psi_2\rangle) = \langle\phi_1|\phi_2\rangle\langle\psi_1|\psi_2\rangle.$

Composite quantum systems

- If the state of Alice's system is ρ and the state of Bob's system is σ and they have never interacted in the past, then the state of the joint Alice-Bob system is

$$\rho_A \otimes \sigma_B.$$

- We use the system labels to say who has what.

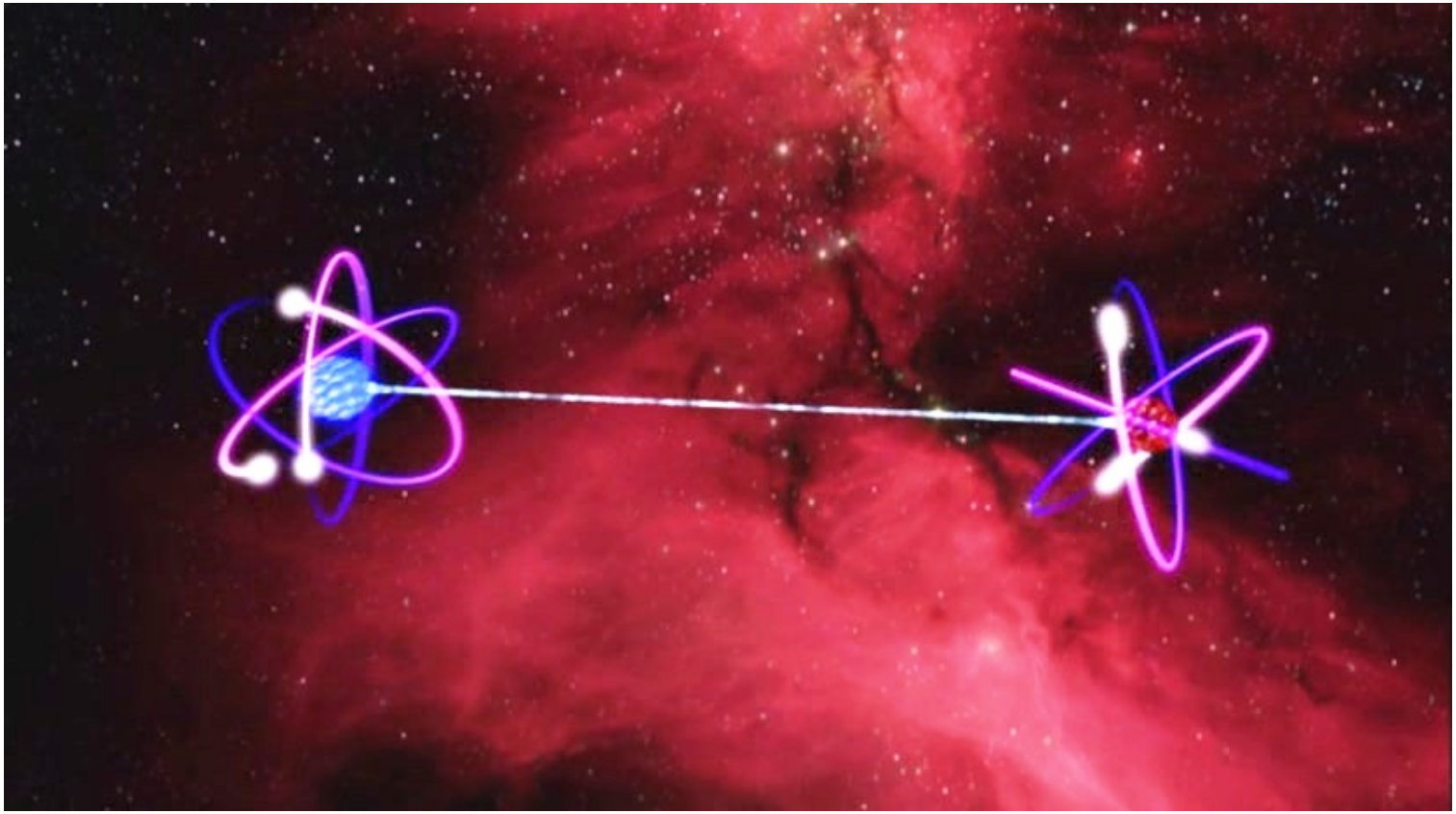
- For example, their state could be

$$|0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_B, \text{ or}$$
$$|1\rangle\langle 1|_A \otimes |1\rangle\langle 1|_B,$$

or a mixture of both, with $p \in [0, 1]$:

$$p|0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_B + (1-p)|1\rangle\langle 1|_A \otimes |1\rangle\langle 1|_B.$$

Quantum entanglement...



Depiction of quantum entanglement taken from
http://thelifeofpsi.com/2013/10/28/bertlmanns-socks/

Separable states and entangled states

- If Alice and Bob prepare states ρ_A^x and σ_B^x based on a random variable X with distribution p_X, then the state of their systems is

$$\sum_x p_X(x) \rho_A^x \otimes \sigma_B^x.$$

- Such states are called *separable states* and can be prepared using local operations and classical communication (no need for a quantum interaction between A and B to prepare these states).
- By spectral decomposition, every separable state can be written as

$$\sum_z p_Z(z) |\psi^z\rangle\langle\psi^z|_A \otimes |\phi^z\rangle\langle\phi^z|_B,$$

  where, for each z, $|\psi^z\rangle_A$ and $|\phi^z\rangle_B$ are unit vectors.
- *Entangled states* are states that cannot be written in the above form.

Example of entangled state

- A prominent example of an entangled state is the *ebit* (eee · bit):

$$|\Phi\rangle\langle\Phi|_{AB},$$

  where $|\Phi\rangle_{AB} \equiv \frac{1}{\sqrt{2}}(|00\rangle_{AB} + |11\rangle_{AB})$.
- In matrix form, this is

$$|\Phi\rangle\langle\Phi|_{AB} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

- To see that this is entangled, consider that for every $|\psi\rangle_A$ and $|\phi\rangle_B$

$$|\langle\Phi|_{AB}|\psi\rangle_A \otimes |\phi\rangle_B|^2 \leq \frac{1}{2}$$

- $\Rightarrow$ impossible to write $|\Phi\rangle\langle\Phi|_{AB}$ as a separable state.

Tool: Schmidt decomposition

Schmidt decomposition theorem

Given a two-party unit vector $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$, we can express it as

$$|\psi\rangle_{AB} \equiv \sum_{i=0}^{d-1} \sqrt{p_i}\, |i\rangle_A\, |i\rangle_B, \text{ where}$$

- probabilities p_i are real, strictly positive, and normalized $\sum_i p_i = 1$.
- $\{|i\rangle_A\}$ and $\{|i\rangle_B\}$ are orthonormal bases for systems A and B.
- $\left[\sqrt{p_i}\right]_{i\in\{0,\ldots,d-1\}}$ is the vector of Schmidt coefficients.
- Schmidt rank d of $|\psi\rangle_{AB}$ is equal to the number of Schmidt coefficients p_i in its Schmidt decomposition and satisfies

$$d \leq \min\left\{\dim(\mathcal{H}_A), \dim(\mathcal{H}_B)\right\}.$$

- State $|\psi\rangle\langle\psi|_{AB}$ is entangled iff $d \geq 2$.

Tool: Partial trace

- The trace of a matrix X can be realized as

$$\mathrm{Tr}\{X\} = \sum_i \langle i|X|i\rangle,$$

where $\{|i\rangle\}$ is an orthonormal basis.

- *Partial trace* of a matrix Y_{AB} acting on $\mathcal{H}_A \otimes \mathcal{H}_B$ can be realized as

$$\mathrm{Tr}_A\{Y_{AB}\} = \sum_i (\langle i|_A \otimes I_B) Y_{AB} (|i\rangle_A \otimes I_B),$$

where $\{|i\rangle_A\}$ is an orthonormal basis for $\mathcal{H}_A$ and I_B is the identity matrix acting on $\mathcal{H}_B$.

- Both trace and partial trace are linear operations.

Interpretation of partial trace

- Suppose Alice and Bob possess quantum systems in the state ρ_{AB}. We calculate the density matrix for Alice's system using partial trace:

$$\rho_A \equiv \text{Tr}_A\{\rho_{AB}\}.$$

- We can then use ρ_A to predict the outcome of any experiment performed on Alice's system alone.
- Partial trace generalizes marginalizing a probability distribution:

$$\begin{aligned}
&\text{Tr}_Y \left\{ \sum_{x,y} p_{X,Y}(x,y) |x\rangle\langle x|_X \otimes |y\rangle\langle y|_Y \right\} \\
&= \sum_{x,y} p_{X,Y}(x,y) |x\rangle\langle x|_X \, \text{Tr}\left\{ |y\rangle\langle y|_Y \right\} \\
&= \sum_x \left[\sum_y p_{X,Y}(x,y) \right] |x\rangle\langle x|_X = \sum_x p_X(x) |x\rangle\langle x|_X,
\end{aligned}$$

where $p_X(x) \equiv \sum_y p_{X,Y}(x,y)$.

Purification of quantum noise...

Artistic rendering of the notion of purification
(Image courtesy of seaskylab at FreeDigitalPhotos.net)

Tool: Purification of quantum states

- A purification of a state ρ_S on system S is a pure quantum state $|\psi\rangle\langle\psi|_{RS}$ on systems R and S, such that

$$\rho_S = \text{Tr}_R\{|\psi\rangle\langle\psi|_{RS}\}.$$

- Simple construction: take $|\psi\rangle_{RS} = \sum_x \sqrt{p(x)}|x\rangle_R \otimes |x\rangle_S$ if ρ_S has spectral decomposition $\sum_x p(x)|x\rangle\langle x|_S$.

- Two different states $|\psi\rangle\langle\psi|_{RS}$ and $|\phi\rangle\langle\phi|_{RS}$ purify ρ_S iff they are related by a unitary U_R acting on the reference system. Necessity:

$$\begin{aligned}\text{Tr}_R\{(U_R \otimes I_S)|\psi\rangle\langle\psi|_{RS}(U_R^\dagger \otimes I_S)\} &= \text{Tr}_R\{(U_R^\dagger U_R \otimes I_S)|\psi\rangle\langle\psi|_{RS}\} \\ &= \text{Tr}_R\{|\psi\rangle\langle\psi|_{RS}\} \\ &= \rho_S.\end{aligned}$$

To prove sufficiency, use Schmidt decomposition.

Uses and interpretations of purification

- The concept of purification is one of the most often used tools in quantum information theory.
- This concept does not exist in classical information theory and represents a radical departure (i.e., in classical information theory it is not possible to have a definite state of two systems such that the reduced systems are individually indefinite).
- Physical interpretation: Noise or mixedness in a quantum state is due to entanglement with an inaccessible reference / environment system.
- Cryptographic interpretation: In the setting of quantum cryptography, we assume that an eavesdropper Eve has access to the full purification of a state ρ_{AB} that Alice and Bob share. This means physically that Eve has access to every other system in the universe that Alice and Bob do not have access to!
- Advantage: only need to characterize Alice and Bob's state in order to understand what Eve has.

Quantum channels

Classical channels

- Classical channels model evolutions of classical systems.
- What are the requirements that we make for classical channels?
- 1) They should be linear maps, which means they respect convexity.
- 2) They should take probability distributions to probability distributions (i.e., they should output a legitimate state of a classical system when a classical state is input).
- These requirements imply that the evolution of a classical system is specified by a conditional probability matrix N with entries $p_{Y|X}(y|x)$, so that the input-output relationship of a classical channel is given by

$$p_Y = N\, p_X \qquad \Longleftrightarrow \qquad p_Y(y) = \sum_x p_{Y|X}(y|x) p_X(x).$$

Quantum channels

- Quantum channels model evolutions of quantum systems.
- We make similar requirements:
- A quantum channel $\mathcal{N}$ is a linear map acting on the space of (density) matrices:

$$\mathcal{N}(p\rho + (1-p)\sigma) = p\mathcal{N}(\rho) + (1-p)\mathcal{N}(\sigma),$$

  where $p \in [0,1]$ and $\rho, \sigma \in \mathcal{D}(\mathcal{H})$.
- We demand that a quantum channel should take quantum states to quantum states.
- This means that it should be trace (probability) preserving:

$$\text{Tr}\{\mathcal{N}(X)\} = \text{Tr}\{X\}$$

  for all $X \in \mathcal{L}(\mathcal{H})$ (linear operators, i.e., matrices).

Complete positivity

- Other requirement is complete positivity.
- We can always expand $X_{RS} \in \mathcal{L}(\mathcal{H}_{\mathcal{R}} \otimes \mathcal{H}_{\mathcal{S}})$ as

$$X_{RS} = \sum_{i,j} |i\rangle\langle j|_R \otimes X_S^{i,j},$$

and then define

$$(\mathrm{id}_R \otimes \mathcal{N}_S)(X_{RS}) = \sum_{i,j} |i\rangle\langle j|_R \otimes \mathcal{N}_S\left(X_S^{i,j}\right),$$

with the interpretation being that "nothing (identity channel) happens on system R while the channel $\mathcal{N}$ acts on system S."

- A quantum channel should also be completely positive:

$$(\mathrm{id}_R \otimes \mathcal{N}_S)(X_{RS}) \geq 0,$$

where id_R denotes the identity channel acting on system R of arbitrary size and $X_{RS} \in \mathcal{L}(\mathcal{H}_{\mathcal{R}} \otimes \mathcal{H}_{\mathcal{S}})$ is such that $X_{RS} \geq 0$.

Quantum channels: completely positive, trace-preserving

- A map $\mathcal{N}$ satisfying the requirements of linearity, trace preservation, and complete positivity takes all density matrices to density matrices and is called a *quantum channel*.
- To check whether a given map is completely positive, it suffices to check whether

$$(\mathrm{id}_R \otimes \mathcal{N}_S)(|\Phi\rangle\langle\Phi|_{RS}) \geq 0,$$

where

$$|\Phi\rangle_{RS} = \frac{1}{\sqrt{d}} \sum_i |i\rangle_R \otimes |i\rangle_S$$

and $d = \dim(\mathcal{H}_{\mathcal{R}}) = \dim(\mathcal{H}_{\mathcal{S}})$.

- Interpretation: the state resulting from a channel acting on one share of a maximally entangled state completely characterizes the channel.

Choi-Kraus representation theorem

Structure theorem for quantum channels

Every quantum channel $\mathcal{N}$ can be written in the following form:

$$\mathcal{N}(X) = \sum_i K_i X K_i^\dagger, \tag{1}$$

where $\{K_i\}$ is a set of Kraus operators, with the property that

$$\sum_i K_i^\dagger K_i = I. \tag{2}$$

The form given in (1) corresponds to complete positivity and the condition in (2) to trace (probability) preservation. This decomposition is not unique, but one can find a minimal decomposition by taking a spectral decomposition of $(\mathrm{id}_R \otimes \mathcal{N}_S)(|\Phi\rangle\langle\Phi|_{RS})$.

Examples of quantum channels

- Quantum bit-flip channel for $p \in [0, 1]$:

$$\rho \to (1 - p)\rho + pX\rho X.$$

- Quantum depolarizing channel for $p \in [0, 1]$:

$$\rho \to (1 - p)\rho + p\pi,$$

  where $\pi \equiv I/d$ (maximally mixed state).

- Quantum erasure channel for $p \in [0, 1]$:

$$\rho \to (1 - p)\rho + p|e\rangle\langle e|,$$

  where $\langle e|\rho|e\rangle = 0$ for all inputs ρ.

Unitary channels

- If a channel has one Kraus operator (call it U), then it satisfies $U^\dagger U = I$ and is thus a unitary matrix.[1]
- Unitary channels are ideal, reversible channels.
- Instruction sequences for quantum algorithms (to be run on quantum computers) are composed of ideal, unitary channels.
- So if a quantum channel has more than one Kraus operator (in a minimal decomposition), then it is non-unitary and irreversible.

[1] It could also be part of a unitary matrix, in which case it is called an "isometry."

Preparation channels

- Preparation channels take classical systems as input and produce quantum systems as output.
- A preparation channel $\mathcal{P}$ has the following form:

$$\mathcal{P}(\rho) = \sum_x \langle x|\rho|x\rangle \sigma^x,$$

  where $\{|x\rangle\}$ is an orthonormal basis and $\{\sigma^x\}$ is a set of states.
- Inputting the classical state $|x\rangle\langle x|$ leads to quantum output σ^x, i.e., it is just the map

$$x \to \sigma^x,$$

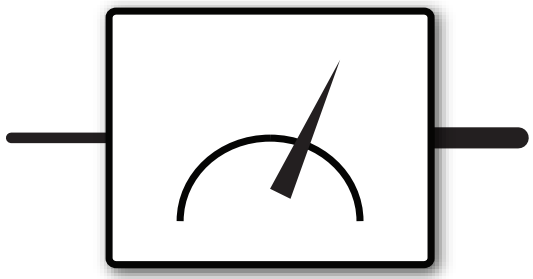
  where x is a classical letter. Sometimes called "cq" channel, short for "classical-to-quantum" channel.

Measurement channels

- Measurement channels take quantum systems as input and produce classical systems as output.
- A measurement channel $\mathcal{M}$ has the following form:

$$\mathcal{M}(\rho) = \sum_x \text{Tr}\{M^x \rho\} |x\rangle\langle x|,$$

  where $M_x \geq 0$ for all x and $\sum_x M^x = I$.
- Can also interpret a measurement channel as returning the classical value x with probability $\text{Tr}\{M^x \rho\}$.
- We depict them as



"Measuring an operator"

- Let G be a Hermitian operator with spectral decomposition

$$G = \sum_x \mu_x \Pi_x,$$

where μ_x are real eigenvalues and Π_x are projections onto corresponding eigensubspaces.

- We say that an experimenter "measures an operator G" by performing the following measurement channel:

$$\rho \to \sum_x \text{Tr}\{\Pi^x \rho\} |x\rangle\langle x|,$$

where $\{|x\rangle\}$ is an orthonormal basis.

Entanglement-breaking channels

- An entanglement-breaking channel $\mathcal{N}$ is defined such that for every input state ρ_{RS}, the output

$$(\mathrm{id}_R \otimes \mathcal{N}_S)(\rho_{RS})$$

  is a separable state.

- To determine whether a given channel is entanglement-breaking, it suffices to check whether the following state is separable:

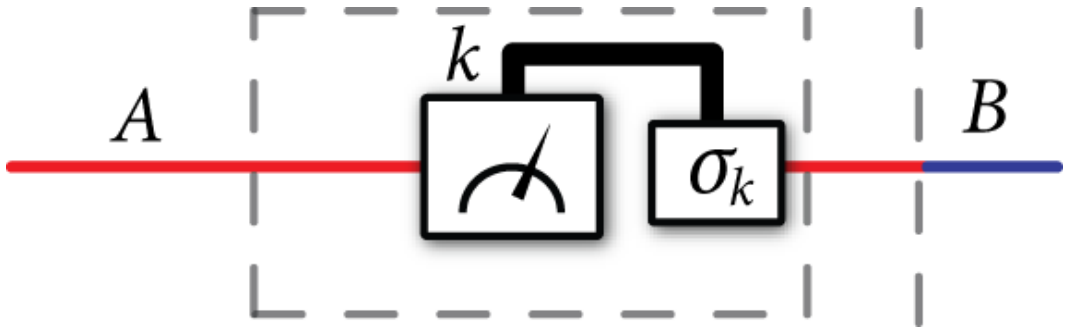
$$(\mathrm{id}_R \otimes \mathcal{N}_S)(|\Phi\rangle\langle\Phi|_{RS}).$$

Entanglement-breaking channels

- Every entanglement-breaking (EB) channel $\mathcal{N}$ can be written as a composition of a measurement $\mathcal{M}$ followed by a preparation $\mathcal{P}$:

$$\mathcal{N} = \mathcal{P} \circ \mathcal{M}.$$

- Thus, internally, every EB channel transforms a quantum system to a classical one and then back: $q \to c \to q$. In this sense, such channels are one step up from classical channels and inherit some properties of classical channels.



Purifications of quantum channels

- Recall that we can purify quantum states and understand noise as arising due to entanglement with an inaccessible reference system.
- We can also purify quantum channels and understand a noisy process as arising from a unitary interaction with an inaccessible environment.

Stinespring's theorem

For every quantum channel $\mathcal{N}_{A\to B}$, there exists a pure state $|0\rangle\langle 0|_E$ and a unitary matrix $U_{AE\to BE'}$, acting on input systems A and E and producing output systems B and E', such that

$$\mathcal{N}_{A\to B}(\rho_A) = \text{Tr}_{E'}\{U_{AE\to BE'}(\rho_A \otimes |0\rangle\langle 0|_E)(U_{AE\to BE'})^\dagger\}.$$

Construction of a unitary extension

- Standard construction of a unitary extension of a quantum channel: Given Kraus operators $\{K_i\}$ for $\mathcal{N}$ such that $\mathcal{N}(\rho) = \sum_i K_i \rho K_i^\dagger$, take

$$V = \sum_i K_i \otimes |i\rangle_{E'} \langle 0|_E .$$

- $V^\dagger V = I$, so we can fill in other columns such that matrix is unitary (call the result U).
- Then

$$U(\rho_A \otimes |0\rangle\langle 0|_E) U^\dagger = \sum_{i,j} K_i \rho K_j^\dagger \otimes |i\rangle\langle j|_{E'},$$

$$\text{and } \operatorname{Tr}_{E'}\{U(\rho_A \otimes |0\rangle\langle 0|_E) U^\dagger\} = \operatorname{Tr}_{E'} \left\{ \sum_{i,j} K_i \rho K_j^\dagger \otimes |i\rangle\langle j|_{E'} \right\}$$

$$= \sum_i K_i \rho K_i^\dagger = \mathcal{N}(\rho).$$

Summary of quantum states and channels

- Every quantum state is a positive, semi-definite matrix with trace equal to one.
- Quantum states of multiple systems can be separable or entangled.
- Quantum states can be purified (this notion does not exist in classical information theory).
- Quantum channels are completely positive, trace-preserving maps.
- Preparation channels take classical systems to quantum systems, and measurement channels take quantum systems to classical systems.
- Quantum channels can also be purified (i.e., every quantum channel can be realized by a unitary interaction with an environment, followed by partial trace). This notion also does not exist in classical information theory.

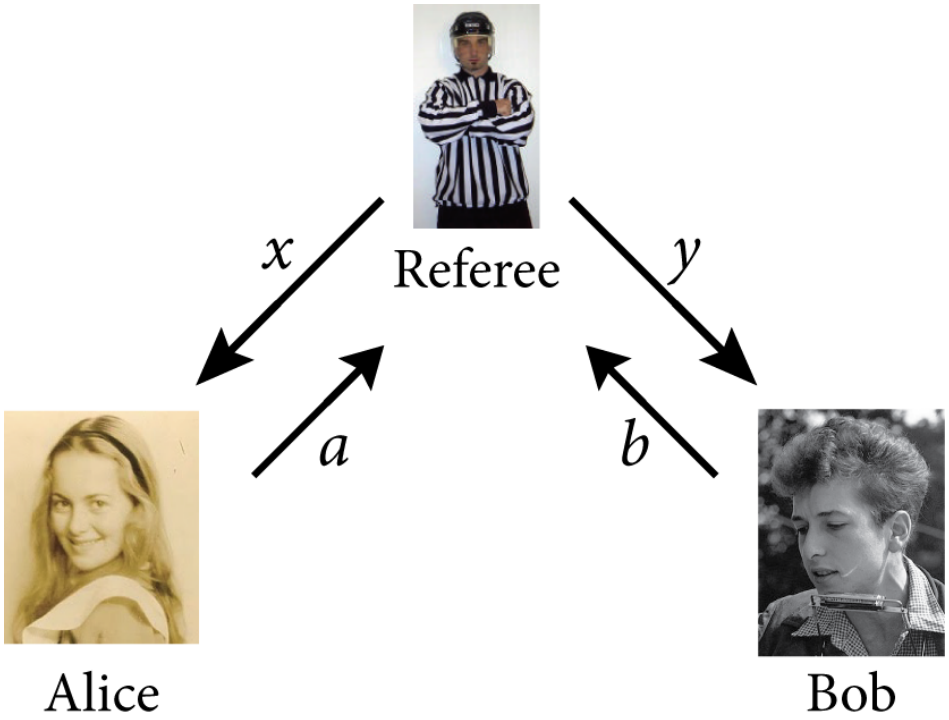
Fundamental protocols

Bell experiment / CHSH game

- How is quantum information different from classical information?
- One way to answer this question is to devise operational tasks for which a quantum strategy outperforms a classical one.
- The most famous is the Bell experiment / CHSH game.[2]
- The game involves two spatially separated parties (the players Alice and Bob) and a referee.

[2] A "loop-hole free" implementation of this experiment was conducted in 2015 (see arXiv:1508.05949).

Bell experiment / CHSH game



- Game begins with referee randomly picking bits x and y.
- Referee sends x and y to Alice and Bob, respectively.
- Alice replies with a bit a and Bob with a bit b.
- They win if and only if $a \oplus b = x \wedge y$.

Classical strategies

- The most general classical strategy allows for Alice and Bob to possess shared randomness before the game begins.
- However, can show that shared randomness does not help them win.
- Thus, to compute the winning probability with classical strategies, it suffices to consider deterministic classical strategies.

Deterministic classical strategies

- General deterministic strategy: $x \to a_x$ for Alice and $y \to b_y$ for Bob.
- The following table presents the winning conditions for the four different values of x and y using this deterministic strategy:

x	y	$x \wedge y$	$= a_x \oplus b_y$
0	0	0	$= a_0 \oplus b_0$
0	1	0	$= a_0 \oplus b_1$
1	0	0	$= a_1 \oplus b_0$
1	1	1	$= a_1 \oplus b_1$

- They cannot always win. (If they could, there would be a contradiction, because adding up 3rd column gives 1 while adding up 4th column gives 0.)
- The best they can do is to win only $3/4 = 0.75$ of the time!
- Strategy achieving this: Alice and Bob each always report back zero.

Quantum strategy

- Allow Alice and Bob to share two qubits in the state $|\Phi\rangle\langle\Phi|_{AB}$ before the game starts.
- If Alice receives $x = 0$, then she performs a measurement of Z. If she receives $x = 1$, then she performs a measurement of X. In each case, she reports the outcome as a.
- If Bob receives $y = 0$, then he performs a measurement of $(X + Z)/\sqrt{2}$. If he receives $y = 1$, then he performs a measurement of $(Z - X)/\sqrt{2}$. In each case, he reports the outcome as b.
- This quantum strategy has a winning probability of $\cos^2(\pi/8) \approx 0.85 > 0.75$ and thus represents a significant separation between classical and quantum information theory.

Loophole-free Bell test...

Picture of loophole-free Bell test at TU Delft
(Image taken from http://hansonlab.tudelft.nl/loophole-free-bell-test/)

Three fundamental protocols

- The three important noiseless protocols in quantum information theory are entanglement distribution, super-dense coding, and quantum teleportation.
- They are the building blocks for later core quantum communication protocols, in which we replace a noiseless resource with a noisy one.

Communication resources

Resources

- Let $[c \to c]$ denote a noiseless classical bit channel from Alice (sender) to Bob (receiver), which performs the following mapping on a qubit density matrix:

$$\rho = \begin{bmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{bmatrix} \to \frac{1}{2}\rho + \frac{1}{2}Z\rho Z = \begin{bmatrix} \rho_{00} & 0 \\ 0 & \rho_{11} \end{bmatrix}.$$

- Let $[q \to q]$ denote a noiseless quantum bit channel from Alice to Bob, which perfectly preserves a qubit density matrix.
- Let $[qq]$ denote a noiseless ebit shared between Alice and Bob, which is a maximally entangled state $|\Phi\rangle\langle\Phi|_{AB}$.
- Entanglement distribution, super-dense coding, and teleportation are non-trivial protocols for combining these resources.

Preparing a maximally entangled state of two qubits

- How to prepare a maximally entangled state?
- Alice begins by preparing two qubits in the tensor-product state:

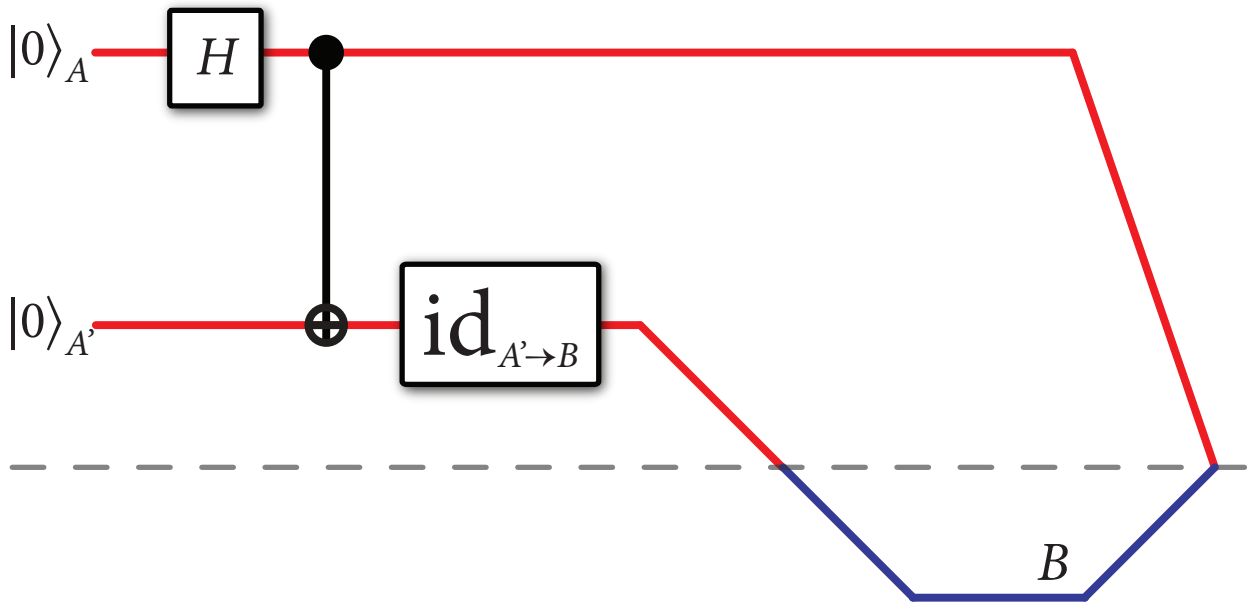
$$|0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_{A'}.$$

- Let $H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, which is a unitary matrix. Alice performs the unitary channel $H(\cdot)H^\dagger$ on her system A, leading to the global state

$$H_A|0\rangle\langle 0|_A H_A^\dagger \otimes |0\rangle\langle 0|_{A'}.$$

- Alice performs CNOT $= |0\rangle\langle 0|_A \otimes I_{A'} + |1\rangle\langle 1|_A \otimes X_{A'}$. This is a unitary called controlled-NOT, because it flips the second bit if and only if the first bit is one (these actions are done in superposition).
- After doing this, the state on AA' becomes $|\Phi\rangle\langle\Phi|_{AA'}$.

Entanglement distribution



- Alice performs local operations (the Hadamard and CNOT) and consumes one use of a noiseless qubit channel to generate one noiseless ebit $|\Phi\rangle\langle\Phi|_{AB}$ shared with Bob.
- Resource inequality: $[q \to q] \geq [qq]$.

Bell states

- Consider that, for a 2×2 matrix M_B,

$$\langle \Phi |_{AB} I_A \otimes M_B | \Phi \rangle_{AB} = \frac{1}{2} \operatorname{Tr}\{M_B\}.$$

- I has trace 2 and Pauli matrices X, Y, and Z are traceless. Multiplying any two of them of them gives another Pauli matrix.

- These facts imply that the following set forms an orthonormal basis:

$$\{|\Phi\rangle_{AB},\ X_A|\Phi\rangle_{AB},\ Z_A|\Phi\rangle_{AB},\ Z_A X_A|\Phi\rangle_{AB}\}.$$

- So the following states are perfectly distinguishable:

$$\{|\Phi\rangle\langle\Phi|_{AB},\ X_A|\Phi\rangle\langle\Phi|_{AB}X_A,\ Z_A|\Phi\rangle\langle\Phi|_{AB}Z_A,\ Z_A X_A|\Phi\rangle\langle\Phi|_{AB}X_A Z_A\}.$$

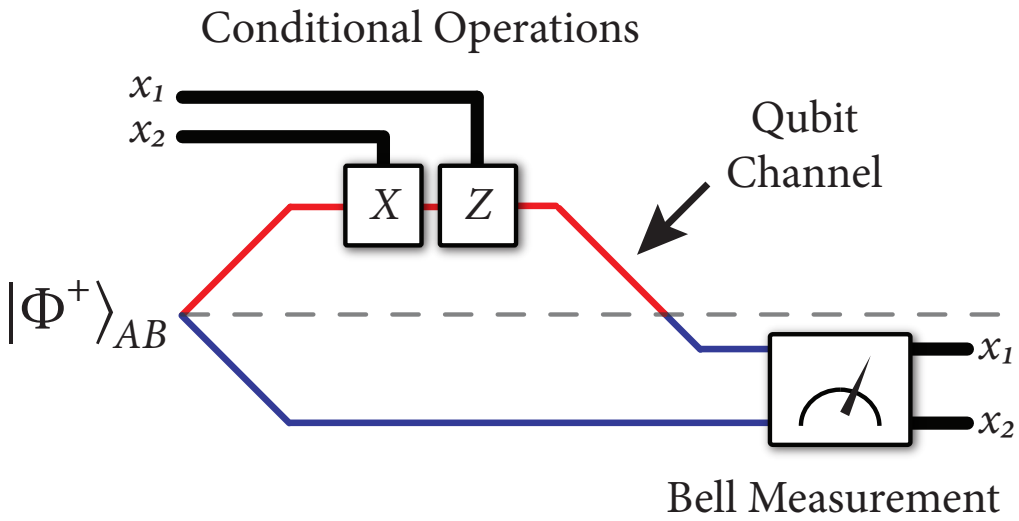
Bell measurement

- The measurement channel that distinguishes these states is called the *Bell measurement*:

$$\begin{aligned}\rho_{AB} \to & \operatorname{Tr}\{|\Phi\rangle\langle\Phi|_{AB}\rho_{AB}\}|00\rangle\langle 00| \\ & + \operatorname{Tr}\{X_A|\Phi\rangle\langle\Phi|_{AB}X_A\rho_{AB}\}|01\rangle\langle 01| \\ & + \operatorname{Tr}\{Z_A|\Phi\rangle\langle\Phi|_{AB}Z_A\rho_{AB}\}|10\rangle\langle 10| \\ & + \operatorname{Tr}\{Z_AX_A|\Phi\rangle\langle\Phi|_{AB}X_AZ_A\rho_{AB}\}|11\rangle\langle 11|.\end{aligned}$$

- This measurement can be implemented on a quantum computer by performing controlled-NOT from A to B, Hadamard on A, and then measuring A and B in the standard basis.

Super-dense coding



- Alice and Bob share an ebit. Alice would like to transmit two classical bits x_1x_2 to Bob. She performs a Pauli rotation conditioned on x_1x_2 and sends her share of the ebit over a noiseless qubit channel. Bob then performs a Bell measurement to get x_1x_2.
- Resource inequality: $[q \to q] + [qq] \geq 2[c \to c]$.

Algebraic trick for quantum teleportation

- Let $|\psi\rangle\langle\psi|$ be the state of a qubit where $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$.
- By using the algebra of the tensor product, can show that

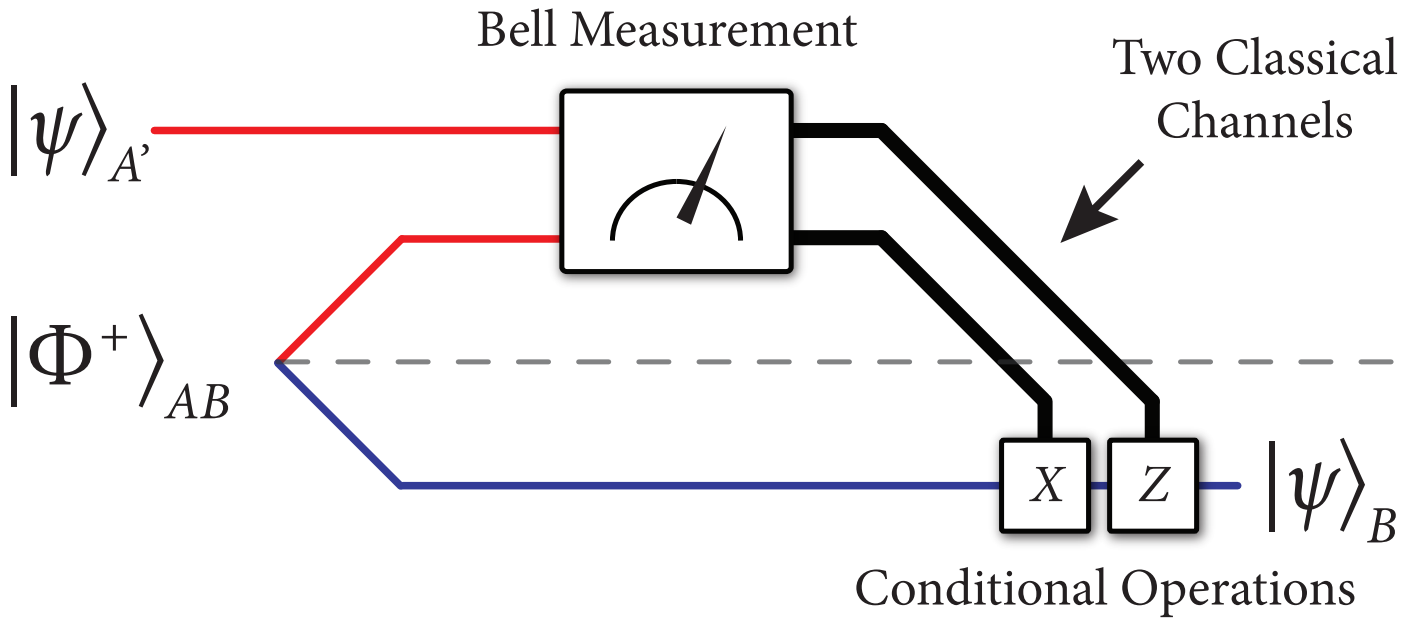
$$\begin{aligned} |\psi\rangle_{A'}|\Phi\rangle_{AB} \quad \propto \quad & |\Phi\rangle_{A'A}|\psi\rangle_B + X_A|\Phi\rangle_{A'A}X_B|\psi\rangle_B \\ & + Z_A|\Phi\rangle_{A'A}Z_B|\psi\rangle_B + Z_AX_A|\Phi\rangle_{A'A}X_BZ_B|\psi\rangle_B. \end{aligned}$$

- Performing the Bell measurement channel on systems AA' leads to the following state:

$$\begin{aligned} \frac{1}{4}\Big[& |00\rangle\langle 00|_{AA'} \otimes |\psi\rangle\langle\psi|_B + |01\rangle\langle 01|_{AA'} \otimes X_B|\psi\rangle\langle\psi|_B X_B \\ & + |10\rangle\langle 10|_{AA'} \otimes Z_B|\psi\rangle\langle\psi|_B Z_B \\ & + |11\rangle\langle 11|_{AA'} \otimes X_BZ_B|\psi\rangle\langle\psi|_B Z_BX_B\Big]. \end{aligned}$$

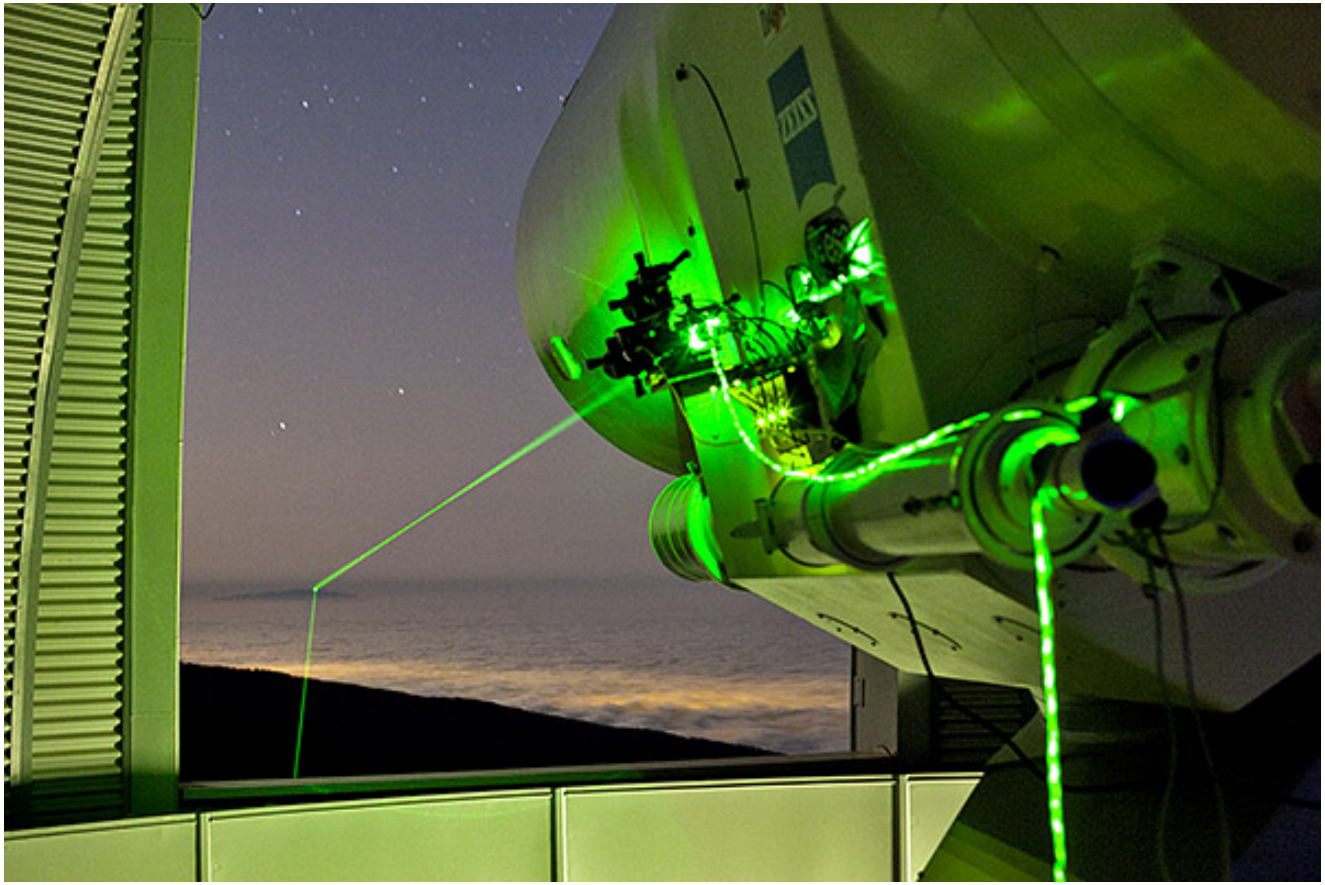
- Alice then sends the two classical bits in AA' to Bob. Bob can then undo the Pauli rotations and recover the state $|\psi\rangle\langle\psi|_B$.

Teleportation



- Alice would like to transmit an arbitrary quantum state $|\psi\rangle\langle\psi|_{A'}$ to Bob. Alice and Bob share an ebit before the protocol begins. Alice can "teleport" her quantum state to Bob by consuming the entanglement and two uses of a noiseless classical bit channel.
- Resource inequality: $2[c \to c] + [qq] \geq [q \to q]$.

Teleportation between Canary Islands...



Teleportation between two Canary Islands 143 km apart. Green lasers were used only for stabilization—invisible infrared photons were teleported (Image taken from http://www.ing.iac.es/PR/press/quantum.html)

Distance measures

Function of a diagonalizable matrix

- If an $n \times n$ matrix D is diagonal with entries $d_1, \ldots, d_n$, then for a function f, we define

$$f(D) = \begin{bmatrix} g(d_1) & 0 & \cdots & 0 \\ 0 & g(d_2) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & g(d_n) \end{bmatrix}$$

  where $g(x) = f(x)$ if $x \neq 0$ and $g(x) = 0$ otherwise.
- If a matrix A is diagonalizable as $A = KDK^{-1}$, then for a function f, we define

$$f(A) = Kf(D)K^{-1}.$$

- Evaluating the function only on the support of the matrix allows for functions such as $f(x) = x^{-1}$ and $f(x) = \log x$.

Trace distance

- Define the trace norm of a matrix X by $\|X\|_1 \equiv \mathrm{Tr}\{\sqrt{X^\dagger X}\}$.
- Trace norm induces *trace distance* between two matrices X and Y:

$$\|X - Y\|_1.$$

- For two density matrices ρ and σ, the following bounds hold

$$0 \leq \|\rho - \sigma\|_1 \leq 2.$$

  LHS saturated iff $\rho = \sigma$ and RHS iff ρ is orthogonal to σ.
- For commuting ρ and σ, trace distance reduces to variational distance between probability distributions along diagonals.
- Has an operational meaning as the bias of the optimal success probability in a hypothesis test to distinguish ρ from σ.
- Does not increase under the action of a quantum channel:

$$\|\rho - \sigma\|_1 \geq \|\mathcal{N}(\rho) - \mathcal{N}(\sigma)\|_1.$$

Fidelity

- Fidelity $F(\rho, \sigma)$ between density matrices ρ and σ is

$$F(\rho, \sigma) \equiv \|\sqrt{\rho}\sqrt{\sigma}\|_1^2.$$

- For pure states $|\psi\rangle\langle\psi|$ and $|\phi\rangle\langle\phi|$, reduces to squared overlap:

$$F(|\psi\rangle\langle\psi|, |\phi\rangle\langle\phi|) = |\langle\psi|\phi\rangle|^2.$$

- For commuting ρ and σ, reduces to Bhattacharyya coefficient of probability distributions along diagonals.
- For density matrices ρ and σ, the following bounds hold:

$$0 \leq F(\rho, \sigma) \leq 1.$$

  LHS saturated iff ρ and σ are orthogonal and RHS iff $\rho = \sigma$.
- Fidelity does not decrease under the action of a quantum channel $\mathcal{N}$:

$$F(\rho, \sigma) \leq F(\mathcal{N}(\rho), \mathcal{N}(\sigma)).$$

Uhlmann's theorem

- Uhlmann's theorem states that

$$F(\rho_S, \sigma_S) = \max_{U_R} |\langle \psi|_{RS} U_R \otimes I_S |\phi\rangle_{RS}|^2,$$

  where $|\psi\rangle_{RS}$ and $|\phi\rangle_{RS}$ purify ρ_S and σ_S, respectively.

- A core theorem used in quantum Shannon theory, and in other areas such as quantum complexity theory and quantum error correction.

- Since it involves purifications, this theorem has no analog in classical information theory.

Relations between fidelity and trace distance

- Trace distance is useful because it obeys the triangle inequality, and fidelity is useful because we have Uhlmann's theorem.
- The following inequalities relate the two measures, which allows for going back and forth between them:

$$1 - \sqrt{F(\rho, \sigma)} \leq \frac{1}{2}\|\rho - \sigma\|_1 \leq \sqrt{1 - F(\rho, \sigma)}.$$

- A distance measure which has both properties (triangle inequality and Uhlmann's theorem) is $\sqrt{1 - F(\rho, \sigma)}$.

Information measures

Entropy and information...

Entropy and information can be discomforting...

Quantum relative entropy

- One of the most fundamental information measures is the quantum relative entropy, defined for a state ρ and a positive semi-definite matrix σ as
$$D(\rho\|\sigma) \equiv \text{Tr}\{\rho[\log_2 \rho - \log_2 \sigma]\},$$
when $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ and as $+\infty$ otherwise.
- It does not increase under the action of a quantum channel $\mathcal{N}$:
$$D(\rho\|\sigma) \geq D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)).$$
- If $\text{Tr}\{\rho\} \geq \text{Tr}\{\sigma\}$, then
$$D(\rho\|\sigma) \geq 0,$$
with equality holding iff $\rho = \sigma$.
- Quantum Pinsker inequality: $D(\rho\|\sigma) \geq \frac{1}{2 \ln 2}\|\rho - \sigma\|_1^2$.

Children of quantum relative entropy

Relative entropy as "parent" entropy

Many entropies can be written in terms of relative entropy:

- $H(A)_\rho \equiv -D(\rho_A \| I_A) = -\operatorname{Tr}\{\rho_A \log_2 \rho_A\}$ (entropy)
- $H(A|B)_\rho \equiv -D(\rho_{AB} \| I_A \otimes \rho_B)$ (conditional entropy)
- $I(A;B)_\rho \equiv D(\rho_{AB} \| \rho_A \otimes \rho_B)$ (mutual information)
- $I(A\rangle B)_\rho \equiv D(\rho_{AB} \| I_A \otimes \rho_B)$ (coherent information)

Equalities

- $H(A|B)_\rho = H(AB)_\rho - H(B)_\rho$
- $I(A\rangle B)_\rho = -H(A|B)_\rho$
- $I(A;B)_\rho = H(A)_\rho + H(B)_\rho - H(AB)_\rho$
- $I(A;B|C)_\rho \equiv H(AC)_\rho + H(BC)_\rho - H(ABC)_\rho - H(C)_\rho$
- $I(A;B|C)_\rho = H(B|C)_\rho - H(B|AC)_\rho$

Evaluating quantum entropy

- How do we evaluate the formula for quantum entropy of a state ρ_A?
- Consider spectral decomposition:

$$\rho_A = \sum_x p_X(x)|x\rangle\langle x|_A.$$

- Then, with $\eta(x) = -x \log_2(x)$,

$$\begin{aligned} H(A)_\rho &= \text{Tr}\{\eta(\rho_A)\} = \text{Tr}\left\{\sum_x \eta(p_X(x))|x\rangle\langle x|_A\right\} \\ &= \sum_x \eta(p_X(x)) \,\text{Tr}\{|x\rangle\langle x|_A\} = \sum_x \eta(p_X(x)) = H(p_X). \end{aligned}$$

- Quantum entropy of ρ_A is equal to Shannon entropy of eigenvalues.
- $\Rightarrow$ Entropy of a pure state is equal to zero.

Bipartite pure-state entanglement

- Let $|\psi\rangle\langle\psi|_{AB}$ be a pure state.
- By Schmidt decomposition theorem, we know that

$$|\psi\rangle_{AB} = \sum_x \sqrt{p_X(x)}|x\rangle_A \otimes |x\rangle_B,$$

  for prob. distribution p_X and orthonormal bases $\{|x\rangle_A\}$ and $\{|x\rangle_B\}$.
- $\Rightarrow$ Eigenvalues of marginal states $\text{Tr}_B\{|\psi\rangle\langle\psi|_{AB}\}$ and $\text{Tr}_A\{|\psi\rangle\langle\psi|_{AB}\}$ are equal.
- Thus, $H(A)_\rho = H(B)_\rho$ if ρ_{AB} is a pure state.
- Exercise: For a tripartite pure state $|\phi\rangle\langle\phi|_{ABC}$,

$$H(A|B)_\phi + H(A|C)_\phi = 0.$$

Conditional quantum entropy can be negative

- One of the most striking differences between classical and quantum information theory: conditional quantum entropy can be negative.
- Consider the conditional quantum entropy of the ebit $|\Phi\rangle\langle\Phi|_{AB}$.
- The global state is pure, while the marginal $\text{Tr}_A\{|\Phi\rangle\langle\Phi|_{AB}\}$ is maximally mixed.
- This implies that $H(AB)_\Phi = 0$ and $H(B)_\Phi = 1$, and thus

$$H(A|B)_\Phi = -1.$$

- If a state σ_{AB} is separable, then one can show that $H(A|B)_\sigma \geq 0$. So a negative conditional entropy implies that a state is entangled (signature of entanglement).

Strong subadditivity

Strong subadditivity

Let ρ_{ABC} be a tripartite quantum state. Then

$$I(A;B|C)_\rho \geq 0.$$

Equivalent statements (by definition)

- Entropy sum of two individual systems is larger than entropy sum of their union and intersection:

$$H(AC)_\rho + H(BC)_\rho \geq H(ABC)_\rho + H(C)_\rho.$$

- Conditional entropy does not decrease under the loss of system A:

$$H(B|C)_\rho \geq H(B|AC)_\rho.$$

Monogamy of entanglement

- By employing strong subadditivity and the Schmidt decomposition, we see that
$$H(A|B)_\rho + H(A|C)_\rho \geq 0.$$
- This is a nontrivial statement for quantum states, given that $H(A|B)_\rho$ can be negative.
- Thus, if $H(A|B)_\rho < 0$, implying that Alice is entangled with Bob, then it must be the case that $H(A|C)_\rho$ is large enough such that the sum is non-negative.
- Often called "monogamy of entanglement," because it says that Alice cannot be strongly entangled with both Bob and Charlie.

Quantum data compression

Quantum information source

- We model a quantum information source as an ensemble of pure states: $\{p_X(x), |\phi_x\rangle\langle\phi_x|\}$.
- The source has expected density matrix

$$\rho = \sum_x p_X(x)|\phi_x\rangle\langle\phi_x|. \tag{3}$$

- Every density matrix has a spectral decomposition:

$$\rho = \sum_z p_Z(z)|z\rangle\langle z|,$$

where p_Z is a probability distribution and $\{|z\rangle\}$ is an O.N. basis. This decomposition in general is different from the one in (3).

Quantum data compression protocols

- Inspired by Shannon, we consider independent calls of the quantum information source and allow for compression schemes that have slight error which vanishes in the limit of many calls of the source.
- An (n, R, ε) quantum data compression scheme consists of an encoding channel $\mathcal{E}^n$, with output system W, and a decoding channel $\mathcal{D}^n$ such that

$$\frac{1}{n} \log_2 \dim(\mathcal{H}_W) \leq R,$$

- and

$$\sum_{x^n} p_{X^n}(x^n) F(|\phi_{x^n}\rangle\langle\phi_{x^n}|, (\mathcal{D}^n \circ \mathcal{E}^n)[|\phi_{x^n}\rangle\langle\phi_{x^n}|]) \geq 1 - \varepsilon.$$

- A rate R is achievable if for all $\varepsilon \in (0, 1)$ and sufficiently large n, there exists an (n, R, ε) quantum compression scheme.
- Quantum data compression limit $=$ infimum of achievable rates.

Quantum data compression theorem

- The quantum data compression limit of a source $\{p_X(x), |\phi_x\rangle\langle\phi_x|\}$ is equal to the quantum entropy of $\rho = \sum_x p_X(x)|\phi_x\rangle\langle\phi_x|$.

- Focus on achievability part. To prove it, we use the notion of quantum typicality.

Quantum typicality

- Given a density matrix ρ with spectral decomposition $\sum_z p_Z(z)|z\rangle\langle z|$, define its (n,δ)-typical subspace by

$$T^{\rho}_{n,\delta} \equiv \text{span}\left\{|z^n\rangle : \left|-\frac{1}{n}\log_2 p_{Z^n}(z^n) - H(\rho)\right| \leq \delta\right\}, \text{ where}$$

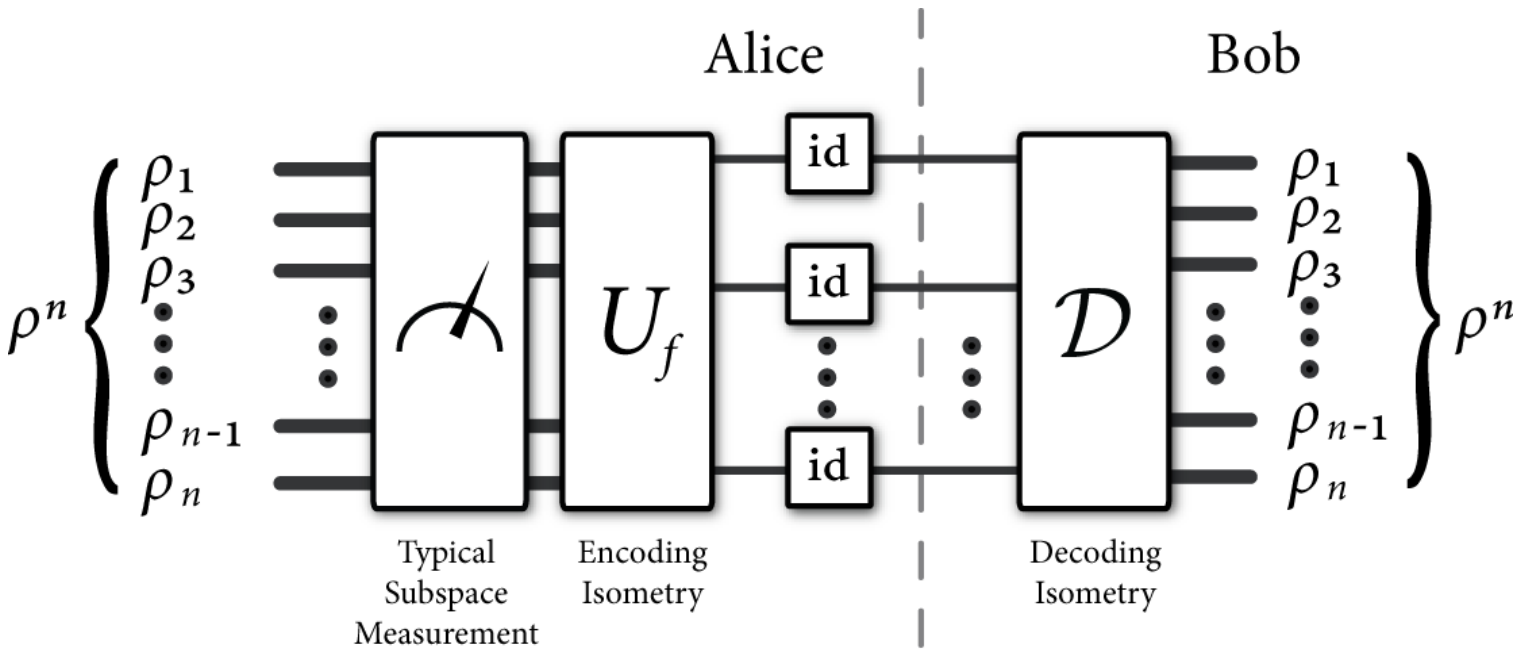
$$p_{Z^n}(z^n) \equiv p_Z(z_1)\cdots p_Z(z_n), \qquad |z^n\rangle \equiv |z_1\rangle \otimes \cdots \otimes |z_n\rangle.$$

- Let $\Pi^{\rho}_{n,\delta}$ denote the projection onto $T^{\rho}_{n,\delta}$.
- Then,

$$\begin{aligned}
\text{Tr}\{\Pi^{\rho}_{n,\delta}\rho^{\otimes n}\} &\geq 1-\varepsilon,\\
(1-\varepsilon)2^{n[H(\rho)-\delta]} \leq \text{Tr}\{\Pi^{\rho}_{n,\delta}\} &\leq 2^{n[H(\rho)+\delta]},\\
2^{-n[H(\rho)+\delta]}\Pi^{\rho}_{n,\delta} \leq \Pi^{\rho}_{n,\delta}\rho^{\otimes n}\Pi^{\rho}_{n,\delta} &\leq 2^{-n[H(\rho)-\delta]}\Pi^{\rho}_{n,\delta}.
\end{aligned}$$

- Inequalities with ε are true for all $\varepsilon \in (0,1)$ and sufficiently large n.

Quantum data compression



- Main idea for quantum data compression: measure typical subspace. Successful with probability $1 - \varepsilon$.
- If successful, perform a unitary that rotates typical subspace to space of dimension $\leq 2^{n[H(\rho)+\delta]}$ (represented with $n[H(\rho) + \delta]$ qubits).
- Send qubits to Bob, who then undoes the compression unitary.
- Scheme is guaranteed to meet the fidelity criterion.

Classical communication

Classical communication code

- Suppose that Alice and Bob are connected by a quantum channel $\mathcal{N}_{A\to B}$ and that they are allowed to use it n times. The resulting channel is $\mathcal{N}_{A\to B}^{\otimes n}$, with Kraus operators that are tensor products of the individual Kraus operators.
- An (n, R, ε) classical comm. code consists of an encoding channel $\mathcal{E}_{M'\to A^n}$ and a decoding measurement channel $\mathcal{D}_{B^n\to\hat{M}}$ such that:

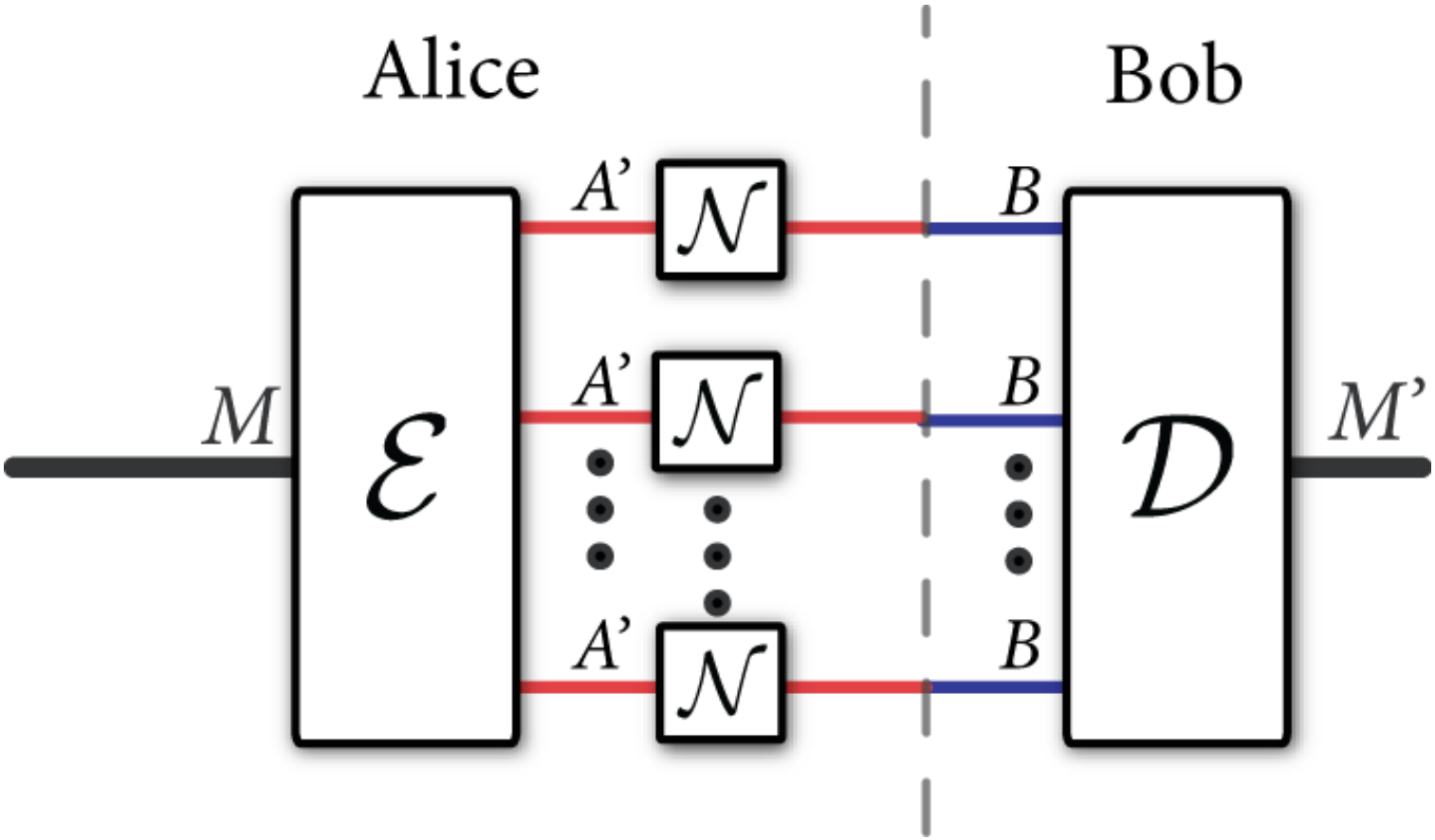
$$F(\overline{\Phi}_{M\hat{M}}, (\mathcal{D}_{B^n\to\hat{M}} \circ \mathcal{N}_{A\to B}^{\otimes n} \circ \mathcal{E}_{M'\to A^n})(\overline{\Phi}_{MM'})) \geq 1 - \varepsilon,$$

where

$$\overline{\Phi}_{M\hat{M}} \equiv \frac{1}{\dim(\mathcal{H}_M)} \sum_m |m\rangle\langle m|_M \otimes |m\rangle\langle m|_{\hat{M}},$$

- and $\frac{1}{n}\log_2(\dim(\mathcal{H}_M)) \geq R$.
- Note that $\overline{\Phi}_{M\hat{M}}$ represents a classical state, and the goal is for the coding scheme to preserve the classical correlations in this state.

Schematic of a classical communication code



Classical capacity

- A rate R for classical communication is *achievable* if for all $\varepsilon \in (0, 1)$ and sufficiently large n, there exists an (n, R, ε) classical communication code.

- The classical capacity $C(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is equal to the supremum of all achievable rates.

What is known about classical capacity

- Lower bound on classical capacity:

$$\chi(\mathcal{N}) \leq C(\mathcal{N})$$

$$\text{where } \chi(\mathcal{N}) = \max_{p_X(x),\rho_A^x} I(X;B)_\omega,$$

$$\omega_{XB} \equiv \sum_x p_X(x)|x\rangle\langle x|_X \otimes \mathcal{N}(\rho_A^x).$$

- For some special channels, we know that $\chi(\mathcal{N}) = C(\mathcal{N})$.
- But it is also known that there exists a channel for which

$$\chi(\mathcal{N}) < C(\mathcal{N}).$$

This superadditivity phenomenon is due to quantum entanglement.

Achievability part: Random coding

- Borrow the idea of random coding from Shannon, but then we need to figure out a decoding channel.
- Consider an ensemble $\{p_X(x), \rho_A^x\}$ that Alice can pick at the channel input. This leads to the output ensemble

$$\{p_X(x), \sigma_A^x \equiv \mathcal{N}_{A\to B}(\rho_A^x)\}.$$

- So pick classical codewords randomly according to $p_X(x)$. This leads to a codebook $\{x^n(m) \equiv x_1(m)\cdots x_n(m)\}_{m\in[\dim(\mathcal{H}_M)]}$.
- The channel output after sending the mth message is

$$\sigma_{B^n}^{x^n(m)} \equiv \sigma_{B_1}^{x_1(m)} \otimes \cdots \otimes \sigma_{B_n}^{x_n(m)}.$$

Achievability part: Sequential decoding

- To every channel output $\sigma_{B^n}^{x^n(m)}$, there exists a conditionally typical projector Π_m, with properties similar to those of the typical projector.
- A sequential decoding strategy consists of performing a sequence of binary tests using conditionally typical projectors, asking "Is it the first message? Is it the second message? etc." until there is a "hit."
- When sending the mth message, the success probability in decoding it using this strategy is

$$\text{Tr}\{\Pi_m \hat{\Pi}_{m-1} \cdots \hat{\Pi}_1 \sigma_{B^n}^{x^n(m)} \hat{\Pi}_1 \cdots \hat{\Pi}_{m-1} \Pi_m\},$$

  where $\hat{\Pi}_i \equiv I - \Pi_i$.
- This implies that the error probability is

$$1 - \text{Tr}\{\Pi_m \hat{\Pi}_{m-1} \cdots \hat{\Pi}_1 \sigma_{B^n}^{x^n(m)} \hat{\Pi}_1 \cdots \hat{\Pi}_{m-1} \Pi_m\}.$$

Error Analysis

- The expected channel output with respect to the code distribution is $\sigma_B = \sum_x p_X(x)\sigma_B^x$, which has a typical projection Π_σ.
- The error probability will ultimately change just slightly by incorporating this projection into the analysis:

$$\text{Tr}\{\Pi_\sigma \sigma_{B^n}^{x^n(m)} \Pi_\sigma\} - \text{Tr}\{\Pi_m \hat{\Pi}_{m-1} \cdots \hat{\Pi}_1 \Pi_\sigma \sigma_{B^n}^{x^n(m)} \Pi_\sigma \hat{\Pi}_1 \cdots \hat{\Pi}_{m-1} \Pi_m\}.$$

- Using a quantum version of the union bound, this can be bounded from above by

$$2\sqrt{\text{Tr}\{(I - \Pi_m)\Pi_\sigma \sigma_{B^n}^{x^n(m)} \Pi_\sigma\} + \sum_{i=1}^{m-1} \text{Tr}\{\Pi_i \Pi_\sigma \sigma_{B^n}^{x^n(m)} \Pi_\sigma\}}$$

  The two terms above are exactly analogous to similar error terms that arise in the analysis of Shannon's channel coding theorem.
- By taking an expecation with respect to the code distribution, we can then analyze this error.

- Error to bound:

$$2\sqrt{\mathbb{E}_{\mathcal{C}}\{\mathrm{Tr}\{(I-\Pi_m)\Pi_\sigma \sigma_{B^n}^{X^n(m)}\Pi_\sigma\}\} + \sum_{i=1}^{m-1}\mathbb{E}_{\mathcal{C}}\{\mathrm{Tr}\{\Pi_i\Pi_\sigma \sigma_{B^n}^{X^n(m)}\Pi_\sigma\}\}}$$

- The first term can be made small using properties of typicality.
- The second term can be made small by choosing the code rate to be smaller than the mutual information $I(X;B) = H(B) - H(B|X)$. Consider that

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}}\{\mathrm{Tr}\{\Pi_i\Pi_\sigma \sigma_{B^n}^{x^n(m)}\Pi_\sigma\}\} &= \mathrm{Tr}\{\mathbb{E}_{X^n(i)}\{\Pi_i\}\Pi_\sigma \mathbb{E}_{X^n(m)}\{\sigma_{B^n}^{X^n(m)}\}\Pi_\sigma\} \\
&= \mathrm{Tr}\{\mathbb{E}_{X^n(i)}\{\Pi_i\}\Pi_\sigma \sigma^{\otimes n}\Pi_\sigma\} \\
&\leq 2^{-n[H(B)-\delta]}\,\mathrm{Tr}\{\mathbb{E}_{X^n(i)}\{\Pi_i\}\Pi_\sigma\} \\
&\leq 2^{-n[H(B)-\delta]}\mathbb{E}_{X^n(i)}\{\mathrm{Tr}\{\Pi_i\}\} \\
&\leq 2^{-n[H(B)-\delta]}2^{n[H(B|X)+\delta]} \\
&= 2^{-n[I(X;B)-2\delta]}.
\end{aligned}$$

Conclusion of achievability part

- As long as we pick $\dim(\mathcal{H}_M) = 2^{n[I(X;B)-3\delta]}$, then there exists a code with small error probability, which we can make approach zero by picking n larger and larger.
- We can then expurgate the code if we wish to go from average to maximal error probability (throw away the worse half of the codewords, as in the classical case).
- So the Holevo information $I(X;B)$ is an achievable rate.

Converse theorem

- The converse part of the theorem establishes the regularized Holevo information as an upper bound on classical capacity:

$$C(\mathcal{N}) \leq \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n}).$$

- For some channels, such as entanglement-breaking channels, the following collapse happens for all n:

$$\frac{1}{n}\chi(\mathcal{N}^{\otimes n}) = \chi(\mathcal{N}).$$

- But we know it does not happen in general. That is, it is known that there exists a channel for which

$$\chi(\mathcal{N}) < \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n}).$$

- So there still remains quite a bit to understand about classical capacity.

Entanglement-assisted comm.

Entanglement-assisted classical communication code

- Now allow for Alice and Bob to share entanglement before communication begins. From super-dense coding, we know that entanglement can double the classical capacity of a noiseless qubit channel. What about in general?
- An (n, R, ε) entanglement-assisted classical comm. code consists of an encoding channel $\mathcal{E}_{M' T_A \to A^n}$, a decoding measurement channel $\mathcal{D}_{B^n T_B \to \hat{M}}$, and an entangled state $\Psi_{T_A T_B}$ such that:

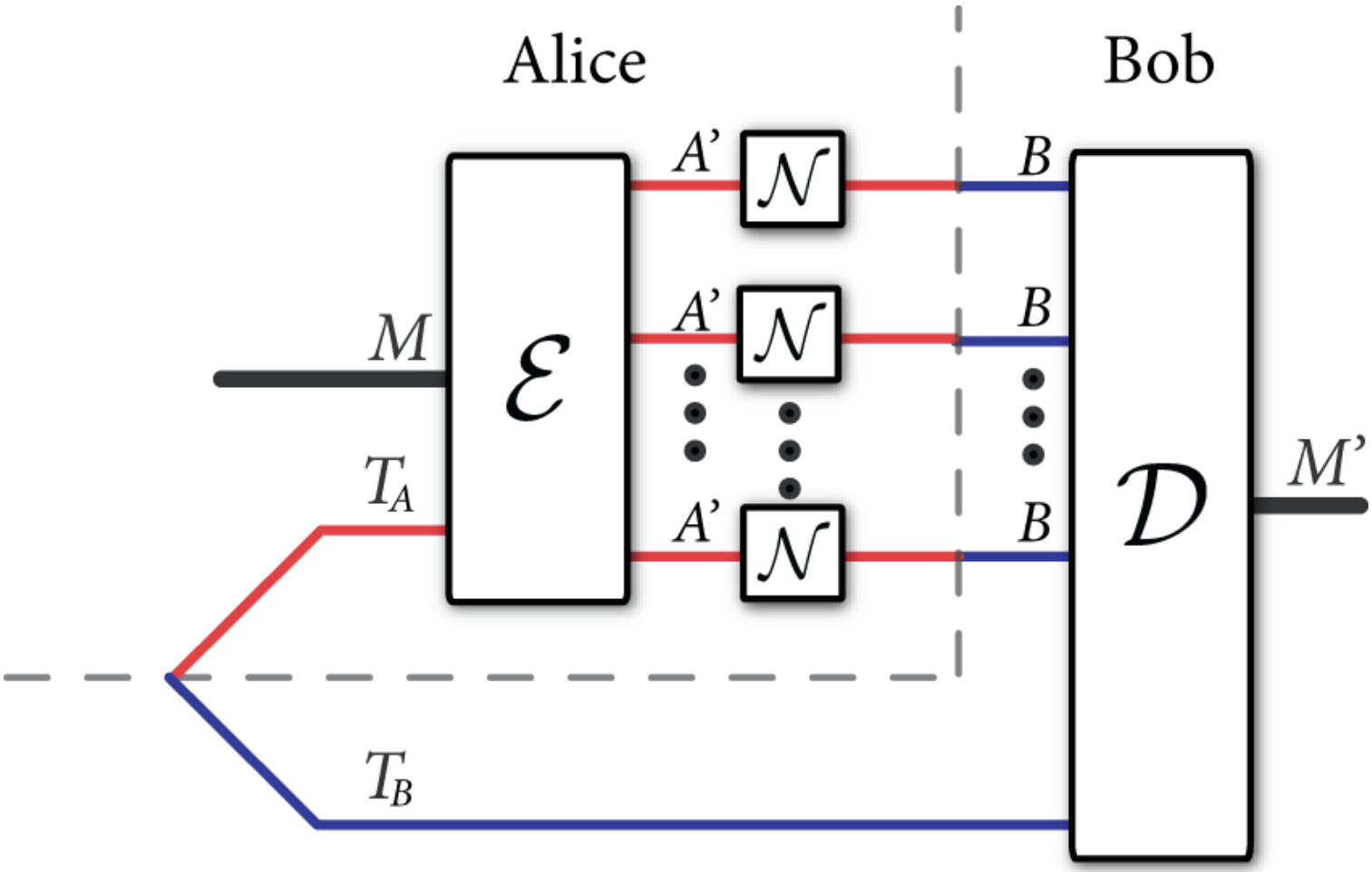
$$F(\overline{\Phi}_{M\hat{M}}, (\mathcal{D}_{B^n T_B \to \hat{M}} \circ \mathcal{N}_{A \to B}^{\otimes n} \circ \mathcal{E}_{M' T_A \to A^n})(\overline{\Phi}_{MM'} \otimes \Psi_{T_A T_B})) \geq 1 - \varepsilon,$$

where

$$\overline{\Phi}_{M\hat{M}} \equiv \frac{1}{\dim(\mathcal{H}_M)} \sum_m |m\rangle\langle m|_M \otimes |m\rangle\langle m|_{\hat{M}},$$

- and $\frac{1}{n} \log_2(\dim(\mathcal{H}_M)) \geq R$.
- The goal again is for the coding scheme to preserve the classical correlations in the state $\overline{\Phi}_{M\hat{M}}$.

Schematic of an EA classical communication code



Entanglement-assisted classical capacity

- A rate R for entanglement-assisted (EA) classical communication is *achievable* if for all $\varepsilon \in (0,1)$ and sufficiently large n, there exists an (n, R, ε) EA classical communication code.

- The EA classical capacity $C_{\text{EA}}(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is equal to the supremum of all achievable rates.

What is known about entanglement-assisted capacity

- Entanglement-assisted capacity theorem:

$$C_{\mathrm{EA}}(\mathcal{N}) = I(\mathcal{N})$$

$$\text{where } I(\mathcal{N}) = \max_{\phi_{RA}} I(R;B)_{\omega},$$
$$\omega_{RB} \equiv \mathcal{N}_{A\to B}(\phi_{RA}).$$

- Thus, this problem is completely solved!
- $C_{\mathrm{EA}}(\mathcal{N})$ does not change if there is a quantum feedback channel from Bob to Alice. We even know strong converse theorems for this setting as well. In these senses, the entanglement-assisted capacity represents the fully quantum analog of Shannon's channel capacity theorem.

Entanglement-assisted coding (simple version)

- Allow Alice and Bob to share a maximally entangled state $|\Phi\rangle\langle\Phi|_{AB}$.
- They then induce the following ensemble by Alice applying a randomly selected, generalized Pauli operator to her input:

$$\left\{ d^{-2}, \left(\mathcal{N}_{A\to B'} \otimes \mathrm{id}_B\right)\left(|\Phi^{x,z}_{AB}\rangle\langle\Phi^{x,z}_{AB}|\right)\right\}.$$

  where $|\Phi^{x,z}\rangle_{AB} = X(x)_A \, Z(z)_A \, |\Phi\rangle_{AB}$. (This is the same ensemble from super-dense coding if $\mathcal{N}$ is the identity channel.)

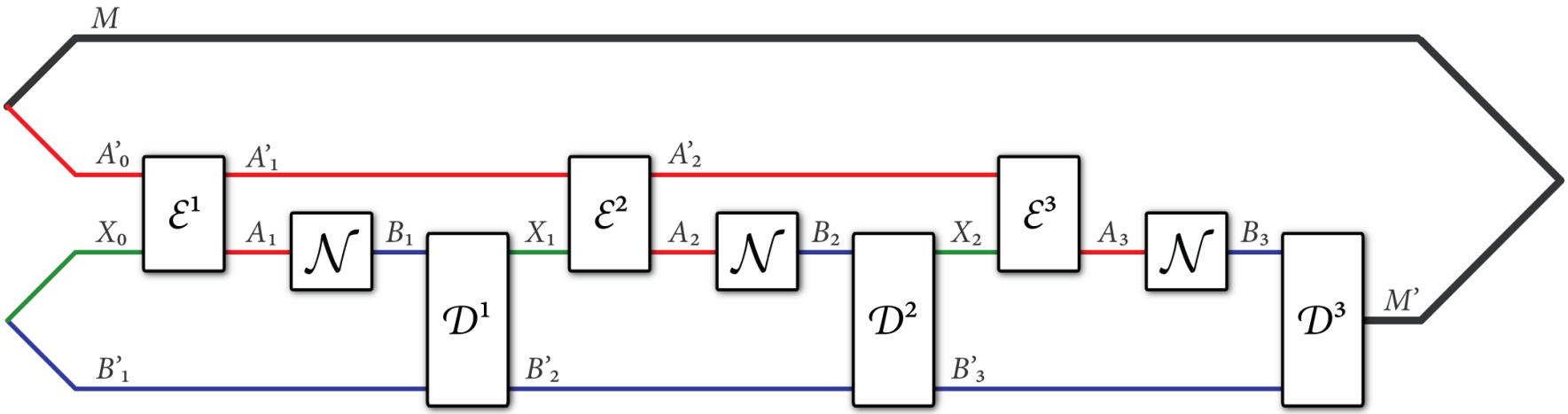
- By previous achievability result and some entropy manipulations, we can conclude that the mutual information $I(B';B)_{\mathcal{N}(\Phi)}$ is achievable.
- More general argument establishes that $I(B';B)_{\mathcal{N}(\phi)}$ is achievable, where ϕ_{AB} is a pure bipartite state. So then $C_{\mathrm{EA}}(\mathcal{N}) \geq I(\mathcal{N})$.

Entanglement-assisted converse theorem

- Employ data processing and the chain rule for conditional mutual information to conclude that

$$C_{\mathrm{EA}}(\mathcal{N}) \leq I(\mathcal{N}).$$

- Can even establish this bound when there is a quantum feedback channel of unlimited dimension connecting Bob to Alice, a setup like



Quantum communication

Quantum communication code

- Now Alice would like to transmit quantum information intact to or generate entanglement with Bob, perhaps for some distributed quantum computation.
- An (n, R, ε) quantum communication code consists of an encoding channel $\mathcal{E}_{M' \to A^n}$ and a decoding channel $\mathcal{D}_{B^n \to \hat{M}}$ such that:

$$F(\Phi_{M\hat{M}}, (\mathcal{D}_{B^n \to \hat{M}} \circ \mathcal{N}_{A \to B}^{\otimes n} \circ \mathcal{E}_{M' \to A^n})(\Phi_{MM'})) \geq 1 - \varepsilon,$$

  where $\Phi_{M\hat{M}}$ is the maximally entangled state:

$$\Phi_{M\hat{M}} \equiv \frac{1}{\dim(\mathcal{H}_M)} \sum_{m,m'} |m\rangle\langle m'|_M \otimes |m\rangle\langle m'|_{\hat{M}},$$

- and $\frac{1}{n} \log_2(\dim(\mathcal{H}_M)) \geq R$.
- The goal now is for the coding scheme to preserve the quantum correlations in the state $\Phi_{M\hat{M}}$.

Quantum capacity

- A rate R for quantum communication is *achievable* if for all $\varepsilon \in (0,1)$ and sufficiently large n, there exists an (n, R, ε) quantum communication code.

- The quantum capacity $Q(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is equal to the supremum of all achievable rates.

What is known about quantum capacity

- Coherent information lower bound on quantum capacity:

$$I_c(\mathcal{N}) \leq Q(\mathcal{N})$$

$$\text{where } I_c(\mathcal{N}) = \max_{\phi_{RA}} I(R\rangle B)_\omega,$$

$$\omega_{RB} \equiv \mathcal{N}_{A\to B}(\phi_{RA}).$$

- If a quantum channel is degradable (meaning that the receiver can simulate the channel from the input to the environment), then

$$I_c(\mathcal{N}) = Q(\mathcal{N}).$$

  A number of interesting quantum channels have this property.
- Quantum capacity is not known for most non-degradable channels. It also exhibits a striking effect called *superactivation*: there exist zero-quantum capacity channels such that they can combine to have a non-zero quantum capacity. (This does not occur for the basic setups in classical information theory.)

Achieving the coherent information

- There are now many coding methods known for achieving the coherent information rate.
- Perhaps the most prominent is known as the *decoupling method*.
- Suppose that Alice, Bob, and Eve share a tripartite pure entangled state $|\psi\rangle\langle\psi|_{RBE}$ after Alice transmits her share of the entanglement with the reference through a noisy channel.
- Then if the reduced state ψ_{RE} on the reference system and Eve's system is approximately decoupled, meaning that

$$\|\psi_{RE} - \psi_R \otimes \sigma_E\|_1 \leq \varepsilon,$$

where σ_E is arbitrary state, this implies that Bob can decode quantum information that Alice intended to send to him. Can show that decoupling is possible as long as qubit rate $\approx$ coherent information.

Decoupling method

- Why does this work? Suppose the state is exactly decoupled. Then one purification of the state ψ_{RE} is the state $|\psi\rangle\langle\psi|_{RBE}$ that they share after the channel acts.
- Another purification of $\psi_{RE} = \psi_R \otimes \sigma_E$ is $|\psi\rangle\langle\psi|_{RB_1} \otimes |\sigma\rangle\langle\sigma|_{B_2E}$, where $|\psi\rangle\langle\psi|_{RB_1}$ is the original state that Alice sent through the channel and $|\sigma\rangle\langle\sigma|_{B_2E}$ is some other state that purifies the state σ_E of the environment.
- All purifications are related by isometries and Bob possesses the purification of R and E,
- $\Rightarrow$ There exists some unitary $U_{B\to B_1B_2}$ such that

$$U_{B\to B_1B_2}|\psi\rangle_{RBE} = |\psi\rangle_{RB_1} \otimes |\sigma\rangle_{B_2E}\,.$$

  This unitary is then Bob's decoder!
- Thus, the decoupling condition implies the existence of a decoder for Bob, so that it is only necessary to show the existence of an encoder that decouples the reference from the environment.

Future directions

Open questions

- It might be difficult to find a general formula for quantum capacity. Some suspect that the quantity is uncomputable.
- Other capacities: private capacity, locking capacity, data hiding capacity (some results known but many questions remain).
- Constructing codes for quantum channels. Major open question for quantum polar codes is to find an efficiently implementable decoder.
- Network quantum information theory: Some results known for multiple access, broadcast, interference, relay channels. Major open question is to prove the existence of a quantum simultaneous decoder (special cases known, but general case is open).
- Strong converses and 2nd-order asymptotics. Some results known. Major open question to establish strong converse property for quantum capacity of degradable channels. Open: 2nd-order asymptotics for entanglement-assisted capacity of all channels.

Other topics

- Capacities of Gaussian quantum channels. These model practical communication channels. A number of open questions remain here. (see Shannon lecture of Holevo).
- Covert communication over quantum channels. (Informal workshop on Wednesday afternoon).
- Quantum channels with memory.
- Security of quantum cryptography (bringing theoretical security proofs closer to experimental implementations).
- Reformulating thermodynamics in the quantum regime using some tools of quantum information theory.
- Quantifying entanglement (resource theory of entanglement).
- Strengthenings of fundamental quantum entropy inequalities.