Lecture Notes on Quantum Information Theory

Victor Kawasaki-Borruat

HS22 - ETHZ

Contents

1 Probability Theory Basics

1.1 Rényi Axioms

We will consider the probability of conditional events in a Boolean algebra of events, given Pr[C] ¿ 0. The **Rényi Axioms** are:

1. Positivity: $Pr[A|C] \geq 0$
2. Normalization: $Pr[A|C] = 1 \iff C \implies A$
3. Summation: $Pr[A \vee B|C] = Pr[A|C] + Pr[B|C]$ for $A \wedge B = 0$
4. Multiplication: $Pr[A \wedge B|C] = Pr[A|B \wedge C]Pr[B|C]$

We can derive a few more useful properties from these axioms, notably the **union bound**:

$$Pr[A \vee B] = Pr[A] + Pr[B] - Pr[A \wedge B] \tag{1}$$

which is useful as an inequality[1]:

$$Pr[\bigvee_{i=1}^{n} A_i] \leq \sum_{i=1}^{n} Pr[A_i] \tag{2}$$

Lastly, **Bayes' Rule** is a useful 'inverse probability' theorem

$$P[A|B \wedge C] = \frac{P[B|A \wedge C]}{P[B|C]} P[A|C] \tag{3}$$

1.1.1 Random Variables

Definition 1.1. *The probability mass function of a random variable Z is denoted:*

$$P_Z : \mathcal{Z} \to [0,1] : z \mapsto P_Z(z) = Pr[Z = z] \tag{4}$$

Definition 1.2. *The joint probability of two random variables is given by:*

$$P_{X,Y} : \mathcal{X} \times \mathcal{Y} \to [0,1] : (x,y) \mapsto P_{X,Y}(x,y)0Pr[X = x \wedge Y = y] \tag{5}$$

Definition 1.3. *The marginal probability mass function given a joint is given by:*

$$P_X(x) = \sum_{y \in \mathcal{Y}} P_{X,Y}(x,y) \tag{6}$$

Definition 1.4. *The conditional probability mass function for X given Y is given by:*

$$P_{X|Y=y}(x) = \frac{P_{X,Y}(x,y)}{P_Y(y)} \tag{7}$$

and forms itself a distribution of its own.

This definition leads to interpreting the marginal as the **average of the conditionals**, since

$$P_X(x) = \sum_{y \in \mathcal{Y}} P_{X|Y}(x|y) P_Y(y) \tag{8}$$

If we consider $Y = f(X)$, we can also define its probability distribution in accordance to P_X.

$$P_Y(y) = \sum_{x \in \mathcal{X}} I[f(x) = y] P_X(x) \tag{9}$$

[1] equality if the A_i are disjoint

1.2 Vector Representation

To help with our quantum theory formalisation, we will use the vector representation of Probability Theory. We will consider $P_Y(y)$ to be the **inner product of two vectors**.

- Let $P = (P_X(x))_{x \in \mathcal{X}}$ be a vectorization of the probabilities for a given ordering of $\mathcal{X}$
- Let $E(A) = (I[f(x) = y])_{x \in \mathcal{X}}$ be the event vector indicating the occurences of $Y = y$

Combining the two yields

$$Pr[A] = P \cdot E(A) \tag{10}$$

This implies that we only require the atoms of our Boolean algebra of events, as they forms a basis for our vector representation. Each event vector for atoms is a unit vector in $\{0,1\}^n \in \mathbb{R}^n$.

Example:

Let

$$E(A) = (1,0,0) \quad E(B) = (0,1,0) \quad E(C) = (0,0,1) \tag{11}$$

so by using logical OR, we can state that

$$E(A \vee B) = (1,1,0) \tag{12}$$

Definition 1.5. *The set of all probability distributions over $\mathbb{R}^n$ is denoted*

$$Prob(n) = \{(p_1, p_2, ..., p_n) \in \mathbb{R}^n : p_i \geq 0, \sum_{i=1}^{n} p_i = 1\} \tag{13}$$

This definition can be extended to certain spaces we are interested in, i.e. $Prob(X)$ denotes the set of all distributions over $\mathcal{X}$.

Definition 1.6. *The set of all events over $\mathbb{R}^n$ is denoted by*

$$Events(n) = \{(e_1, e_2, ..., e_n) \in \mathbb{R}^n : e_i \in \{0,1\} \quad \forall i\} \tag{14}$$

1.3 Convexity

Definition 1.7. *A **convex combination** is essentially an average of possible values.*

Example:

The expectation is a convex combination of all possible values of a random variable.

$$\langle Z \rangle = \sum_z z P_Z(z)$$

Definition 1.8. *A **convex set** is a set closed under convex combination of its elements, i.e.*

$$S = \{s_1, s_2 : \forall \lambda \in [0,1], \lambda s_1 + \bar{\lambda} s_2 \in S\} \tag{15}$$

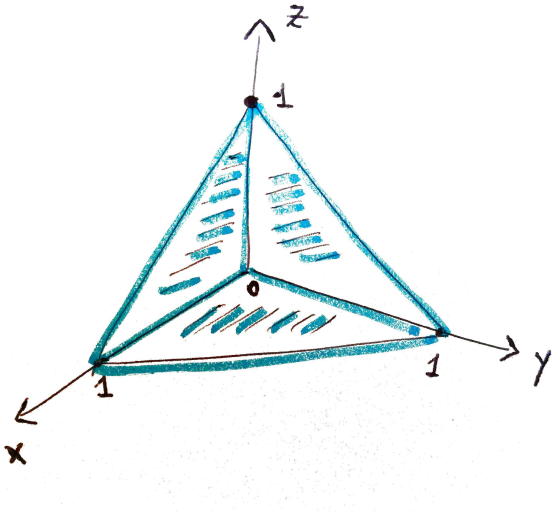
Remark 1. *$Prob(n)$ is a convex set in $\mathbb{R}^n$*

Definition 1.9. *Any arbitrary set can be extended to a convex set by taking all convex combinations of its elements. This is called its **convex hull**.*

Definition 1.10. *The **extreme points** of a convex set are the points that cannot be expressed as a nontrivial comvex combination of other elements.*

Example:
In $Prob(n)$, extreme points are deterministic probability distributions.

Definition 1.11. *A convex set in which every point has a unique convex decomposition is called a **simplex**. The set of probability distributions in $\mathbb{R}^n$ forms a simplex and can be embedded in $\mathbb{R}^{n-1}$ (see image below). In 3 dimensions, a tetrahedron would represent a 4D probability distribution.*



Definition 1.12. *A **convex function** $f : \mathcal{X} \to \mathbb{R}$ satisfies the following:*

$$f(\lambda x_1 + \bar{\lambda} x_2) \leq \lambda f(x_1) + \bar{\lambda} f(x_2) \quad \forall \lambda \in [0,1] \tag{16}$$

Proposition 1.13. *Convex functions satisfy **Jensen's inequality:***

$$f(\langle X \rangle) \leq \langle f(X) \rangle \tag{17}$$

Remark 2. *$Events(n)$ is not a convex set, but we can relax the conditions of the notion of events to that of stochastic events, which are events with values within $[0,1]$. This leads to the set of tests:*

Definition 1.14. *The set of tests is given by:*

$$Tests(n) = \{(t_1, t_2, ..., t_n) \in \mathbb{R}^n : 0 \leq t_i \leq 1\} \tag{18}$$

1.4 Independence

Definition 1.15. *Two random variables A and B are said **independent** if and only if*

$$Pr[A \wedge B] = Pr[A]Pr[B] \tag{19}$$

which is equivalent to

$$Pr[B|A] = P[B] \quad or \quad Pr[A|B] = Pr[A] \tag{20}$$

1.4.1 Practical Laws / Bounds

Definition 1.16. *The **Weak Law of Large Numbers** states that for $X_i \sim_{i.i.d.} P_X$ with $E[X] = \mu$, $Var(X) = \sigma^2$ the arithmetic mean tends in probability to the true mean as $n \to \infty$*

$$lim_{n\to\infty} Pr[\frac{1}{n}|\sum_{i=1}^{n} X_i - \mu| < \epsilon] = 1 \quad \forall \epsilon > 0 \tag{21}$$

Definition 1.17. ***Markov's inequality** states that*

$$Pr[X > \epsilon] \leq \frac{\langle X \rangle}{\epsilon} \tag{22}$$

Definition 1.18. ***Chebyshev's inequality** states that*

$$Pr[(Y - \mu)^2 \geq \epsilon] \leq \frac{\sigma^2}{\epsilon} \tag{23}$$

Definition 1.19. *The **Strong Law of Large Numbers** states that the arithemtic mean nears the true mean*

$$Pr[lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu] = 1 \tag{24}$$

2 Quantum Theory

Quantum experiments are typically carried out in 3 steps:

- Preparation / Setup
- Dynamics
- Measurement

2.1 Quantum Formalism

Recalling our probability theory vectorization, we had that

- $Pr_P[T] = T \cdot P$
- $P \in \mathbb{R}^n, \quad P \geq 0, \quad 1_n \cdot P \leq 1$
- $T \in \mathbb{R}^n, \quad T \geq 0, \quad T \leq 1_n$

such that our vector structure allows probability computation via the use of an inner product. Quantum probability will live within a **Hilbert space** $\mathbb{C}^d$, and probabilies are computed via the **Hilbert-Schmidt** inner product:

$$\langle S, T \rangle = Tr[S^* T] \tag{25}$$

This leads us to the following quantum formalism

- $Pr_\rho = Tr[\Lambda \rho]$
- $\rho \in Lin(\mathcal{H}), \quad \rho \geq 0, \quad Tr[\rho] = 1$
- $\Lambda \in Lin(\mathcal{H}), \quad \Lambda \geq 0, \quad \Lambda \leq \mathbb{I}$

This is very analogous to our classical version, in that ρ denotes a certain distribution over states, and Λ denotes events. Probabilities are also computed via an inner product.

2.1.1 Operators

Recalling the vector representation of probability theory, we will derive an analogous quantum probability formalism. Our space of interest is the **set of operators** on a Hilbert space $\mathcal{H} = \mathbb{C}^d$, denotes $\text{Lin}(\mathcal{H})$.

Definition 2.1 (Density Operator)**.** *An ensemble of quantum states $\{p_k, |\psi_k\rangle\}$ is described by a* ***density operator***

$$\rho = \sum_k p_k |\psi_k\rangle\langle\psi_k| \in Lin(\mathcal{H}) \tag{26}$$

A density operator can also be seen as the preparation of a quantum system, in multiple possible states, hence the analogy to a probability distribution.

Properties 2.2. *A density operator ρ satisfies the following:*

- *$Tr[\rho] = 1$*
- $\rho \geq 0$

Proposition 2.3 (Quantum Measurement / POVM)**.** *Quantum measurements (or effects) are described by a collection* $\{\Lambda(x)\}$ *of measurement operators* $\in Lin(\mathcal{H})$ *such that:*

- $\Lambda(x) \geq 0$
- $\Lambda(x) \leq \mathbb{I}$
- $\sum_x \Lambda(x) = \mathbb{I}$

The third condition yields a ***POVM*** *(positive operator-valued measure).*

Definition 2.4. *The* ***spectral decomposition*** *of any normal operator* A *(i.e.* $A^\dagger A = AA^\dagger$*) on* $\mathcal{H}$ *with respect to an orthonormal basis* $\{|b_k\rangle\}$ *can be done. Let* $\{\lambda_k, |\psi_k\rangle\}$ *be* A*'s eigenvalues resp. eigenvectors, then*

$$A = \sum_k \lambda_k |\psi_k\rangle\langle\psi_k| \tag{27}$$

Definition 2.5. *The set of states of a quantum system in* $\mathcal{H}$ *is denoted*

$$Stat(\mathcal{H}) = \{\sigma \in Lin(\mathcal{H}) : \sigma \geq 0, Tr[\sigma] = 1\} \tag{28}$$

Definition 2.6 (Pure States)**.** *A pure state in a quantum system is a state that cannot be written a a mixture of states. It is an extreme point of our density operator.*

Using the spectral decomposition, notice that any any positive operator M *can be decomposed as*

$$M = \sum_j \lambda_j |\psi_j\rangle\langle\psi_j| \tag{29}$$

If $Tr[M] = 1$*, then* M *becomes a density operator.* M***'s eigenvalues form a probability distribution.*** *Thus any projection operator associated* $|\psi\rangle\langle\psi|$ *associated to wavefunction* $|\psi\rangle$ *is a pure state.*

It follows that the set of states is the convex hull of the pure states.

Theorem 2.7. *A density operator* ρ *is a pure state if and only if*

$$Tr[\rho^2] = 1 \tag{30}$$

2.1.2 Trace of Operators

Proposition 2.8. *The trace of a linear operator* A *on* $\mathcal{H}$ *is defined as the trace on* ***any matrix representation*** *of* A*, i.e. for* $\{v_k\}$ *be an arbitrary ONB of* $\mathcal{H}$*,*

$$Tr[A] = \sum_k \langle v_k|A|v_k\rangle \tag{31}$$

Proof.

$$\begin{aligned} Tr[A] &= \sum_j \langle v_j|(\sum_{j',k} A_{k,j'}|w_k\rangle\langle w_{j'}|)|v_j\rangle = \sum_{j,j',k} A_{k,j'}\langle v_j|w_k\rangle\langle w_{j'}|v_j\rangle \\ &= \sum_{j,j',k} A_{k,j'}\langle w_{j'}|v_j\rangle\langle v_j|w_k\rangle = \sum_{k,j'} A_{k,j'}\langle w_{j'}|(\sum_j |v_j\rangle\langle v_j|)|w_k\rangle \\ &= \sum_{j',k} A_{k,j'}\langle w_{j'}|w_k\rangle = \sum_k A_{kk} \end{aligned}$$

□

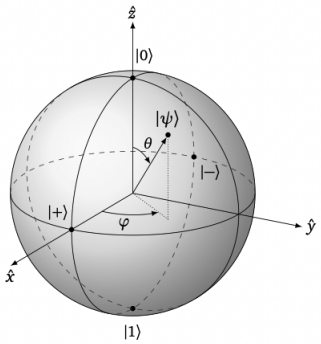


Figure 1: Bloch Sphere. Every pure qubit state lies on its *surface*.

A useful interpretation of the trace of an effect acting on a density operator ρ is to consider ρ as a mixed state.

$$Tr[\Lambda\rho] = \sum_{j=1}^{n} \lambda_j Tr[\Lambda|\lambda_j\rangle\langle\lambda_j|] \tag{32}$$

represents the probability of ρ, but is also interpretable as the **avergage of conditional probabilities of Λ given various pure states $|\lambda_j\rangle\langle\lambda_j|$.**

2.1.3 (Non)-Convexity of QPT

Unlike classical probability distributions, the set of states $\text{Stat}(\mathcal{H})$ is not convex. An arbitrary mixed state does not correspond to a unique decomposition into pure states.

The set of effects $\{\Lambda(x)\}$ however, is convex, with extreme points being projection operators.

2.2 Qubits

A **qubit** is a two-dimensional quantum system. The most typical basis is the **computational basis**, denoted $\{|0\rangle, |1\rangle\}$. The pure states of a qubit a defined as:

$$|\psi\rangle = a|0\rangle + b|1\rangle, \quad |a|^2 + |b|^2 = 1 \tag{33}$$

2.2.1 Bloch Sphere and Vectors

Definition 2.9 (Bloch Sphere)**.** *A useful parametrization of a qubit $|\psi\rangle$ is via spherical coordinates:*

$$|\psi\rangle = \cos(\frac{\theta}{2})|0\rangle + \sin(\frac{\theta}{2})|1\rangle \tag{34}$$

*which yields a sphere a states called the **Bloch sphere**.*

Definition 2.10 (Bloch Vectors)**.** *Using our previous parametrization and the cardinal directions $\hat{x}, \hat{y}, \hat{z}$ of the Bloch sphere, we can describe any state with the **Bloch vector** $\hat{n}$:*

$$\hat{n} = \hat{x}\sin\theta\cos\varphi + \hat{y}\sin\theta\sin\varphi + \hat{z}\cos\theta \tag{35}$$

Remark 3. *A particularity of Bloch vectors is that $|\hat{n}\rangle$ and $|-\hat{n}\rangle$ are **orthogonal**. This is quite counter intuitive, but remember that $|0\rangle$ and $|1\rangle$ form an orthonormal basis of qubits.*

From this, we note that the 6 cardinal directions of the Bloch sphere $\{\pm\hat{x}, \pm\hat{y}, \pm\hat{z}\}$ form three orthogonal bases.

Definition 2.11 (Pauli Matrices)**.** [2] *The three bases directions $\{\pm\hat{x}, \pm\hat{y}, \pm\hat{z}\}$ are the eigenbases of the Pauli operators.*

$$\sigma_X = |0\rangle\langle 1| + |1\rangle\langle 0| \cong \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{36}$$

$$\sigma_Y = -i|0\rangle\langle 1| + i|1\rangle\langle 0| \cong \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \tag{37}$$

$$\sigma_Z = |0\rangle\langle 0| - |1\rangle\langle 1| \cong \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{38}$$

Definition 2.12 (Pauli Operators)**.** *Moreover, the three Pauli matrices, along with the identity matrix $\mathbb{I}_2$ form a basis for Hermitian operators in $\mathcal{H} = \mathbb{C}^2$.*

2.2.2 Pauli representation of density operators

Since the Pauli operators along with identity form a basis of $\mathrm{Lin}(\mathcal{H})$, any operator A can take the form:

$$A = a_0\mathbb{I} + \vec{a} \cdot \vec{\sigma} \tag{39}$$

where $\vec{a} = \hat{x}a_x + \hat{y}a_y + \hat{z}a_z$ and $\vec{\sigma} = \hat{x}\sigma_X + \hat{y}\sigma_Y + \hat{z}\sigma_Z$. This yields

$$A = a_0\mathbb{I} + a_x\sigma_X + a_y\sigma_Y + a_z\sigma_Z \tag{40}$$

The eigenvalues of A are $\lambda_\pm = a_0 + ||\vec{a}||$, with corresponding eigenstates $|\pm\hat{a}\rangle$, with $\hat{a} = \frac{\vec{a}}{||\vec{a}||}$. We may immediately conclude that projection operators ρ or the form $\rho = |\hat{n}\rangle\langle\hat{n}|$ can be written as:

$$\rho = |\hat{n}\rangle\langle\hat{n} = \frac{1}{2}(\mathbb{I} + \hat{n} \cdot \vec{\sigma}) \tag{41}$$

Any density operator on a qubit can be written as

$$\rho = \frac{1}{2}[\mathbb{I} + r \cdot \sigma] \tag{42}$$

where σ is the 3d vector $[\sigma_X, \sigma_Y, \sigma_Z]$.

Remark 4. *If ρ describes a pure state, then $|r| = 1$.*

2.3 Dirac Notation

2.3.1 Vectors

Definition 2.13 (Bra-Ket as a linear map)**.** *Consider ket as the mapping:*

$$|\psi\rangle : \mathbb{C} \to \mathcal{H} : z \mapsto z\psi \tag{43}$$

[2] basically spinors

Note that all expressions are compositions of linear maps. Thus, bra is the linear mapping:

$$\langle\psi| : \mathcal{H} \to \mathbb{C} : u \mapsto \langle\psi, u\rangle \tag{44}$$

The composition of the maps is:

$$\langle\phi| \cdot |\psi\rangle = \langle\phi|\psi\rangle : \mathbb{C} \to \mathbb{C} : z \mapsto z\langle\phi, \psi\rangle \tag{45}$$

The other composition is defined as

$$|\phi\rangle\langle\psi| : \mathcal{H} \to \mathcal{H} : \theta \mapsto |\phi\rangle\langle\psi|\theta\rangle \tag{46}$$

and if $\psi = \phi$, then this mapping is a projection of θ onto ψ.

Remark 5. *Notice that*

$$Tr[|\psi\rangle\langle\phi|] = \langle\phi|\psi\rangle \tag{47}$$

Also, any map $S \in Lin(\mathcal{H})$ can be written as a linear combination of outer products:

$$S = \sum_i |u_i\rangle\langle v_i| \tag{48}$$

Example:

$$\mathbb{I} = \sum_i |b_i\rangle\langle b_i| \tag{49}$$

for any choice of basis $\{|b_i\rangle\}_i$. Moreover, if $S = \sum_i |u_i\rangle\langle v_i|$, then

$$Tr[S] = \sum_i \langle v_i|u_i\rangle \tag{50}$$

Remark 6. *The trace operator Tr is cyclical, i.e.*

$$Tr[ST] = Tr[TS] \quad Tr[S^*] = Tr[S]^* \tag{51}$$

2.3.2 Matrices

Given an orthonormal basis $\{|b_k\rangle\}_{k=1}^d$, $S \in Lin(\mathcal{H})$ has components

$$S_{jk} = \langle b_j|S|b_k\rangle \tag{52}$$

Remark 7. *The outer product (matrix) $|b_j\rangle\langle b_k|$ is all-zero except for a 1 in the j-th row and the k-th column.*

2.3.3 Tensors

Tensors are essentially a product which is compatible with linearity of operators. In our case, we will define a particular product.

Definition 2.14. *In Dirac notation, for two bases $\{|b_j\rangle_A\}_{j=1}^{d_A}$ and $\{|b'_k\rangle_B\}_{k=1}^{d_B}$ of systems $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively, the basis of the **tensor product** of the two systems ($\mathcal{H}_A \otimes \mathcal{H}_B$) is denotes*

$$\{|b_j\rangle_A \otimes |b'_k\rangle_B\} \tag{53}$$

Properties 2.15. *For two vectors $|u\rangle \in \mathcal{H}_A, |v\rangle \in \mathcal{H}_B$ and two maps $S \in Lin(\mathcal{H}_A, \mathcal{H}'_A), T \in Lin(\mathcal{H}_B, \mathcal{H}'_B)$ we have that*

- $\langle u \otimes v, u' \otimes v' \rangle = \langle u|u'\rangle\langle v|v'\rangle$
- $S \otimes T \in Lin(\mathcal{H}_A \otimes \mathcal{H}'_A, \mathcal{H}_B \otimes \mathcal{H}'_B)$ *such that*

$$(S \otimes T) : (u \otimes v) \mapsto (Su) \otimes (Tv) \tag{54}$$

I.e. the maps act on their respective subsystems. From the above property it is also safe to say that

$$Lin(\mathcal{H}_\mathcal{A}, \mathcal{H}'_A) \otimes Lin(\mathcal{H}_B, \mathcal{H}'_\mathcal{B}) \cong Lin(\mathcal{H}_A \otimes \mathcal{H}'_A, \mathcal{H}_B \otimes \mathcal{H}'_B) \tag{55}$$

2.3.4 Partial Trace

Partial trace is very useful to 'trace out' subsystems of composite systems. We will focus on the **Dirac notation** to work out a good system and way of picturing it.

Definition 2.16 (Partial Trace)**.** *Let A, B be quantum systems $A \otimes B$ be their composite system. For states of the form $S_A \otimes T_B$, it is defined as*

$$Tr_B : S_A \otimes T_B \mapsto Tr[T_B]S_A \tag{56}$$

In particular, consider B has an orthonormal basis $\{j\}$. The partial trace over B of a composite system $A \otimes B$ is an operation

$$Tr_B : Lin(\mathcal{H}_A \otimes \mathcal{H}_B) \to Lin(\mathcal{H}_A)^3 : \rho_{AB} \mapsto \sum_j (\mathbb{I}_A \otimes \langle j|_B)\rho_{AB}(\mathbb{I}_A \otimes |j\rangle_B) \tag{57}$$

Moreover, the partial trace operator Tr_B commutes with left and right operations of an operator of the form $T_A \otimes \mathbb{I}_B$

$$\begin{aligned} Tr_B[S_{AB}(T_A \otimes \mathbb{I}_B)] &= Tr_B[S_{AB}]T_A \\ Tr_B[(T_A \otimes \mathbb{I}_B S_{AB}] &= T_A Tr_B[S_{AB}] \end{aligned} \tag{58}$$

Consider the case where we want to simply measure subsystem A with a set of POVMs $\{\Lambda_A(x)\}$. Then we can apply the partial trace on B as follows:

$$P(x) = Tr_{AB}[(\Lambda_A(x) \otimes \mathbb{I}_B)|\Phi\rangle\langle\Phi|_{AB}] = Tr_A[\Lambda_A(x) \sum_{j=1}^{d_B} \langle b_j|(|\Phi\rangle\langle\Phi|_{AB})|b_j\rangle] \tag{59}$$

Example:
For an idea of how to compute it, consider $\rho = |00\rangle\langle 00| = (|0\rangle \otimes |0\rangle)(\langle 0| \otimes \langle 0|) \in Lin(A \otimes B)$. Then

$$Tr_B(\rho) = \sum_{j \in 0,1} (\mathbb{I}_A \otimes \langle j|)\rho(\mathbb{I}_\mathbb{A} \otimes |j\rangle) = (\mathbb{I}_A \otimes \langle 0|)\rho(\mathbb{I}_\mathbb{A} \otimes |0\rangle) + (\mathbb{I}_A \otimes \langle 1|)\rho(\mathbb{I}_\mathbb{A} \otimes |1\rangle)$$

$$= (\mathbb{I}_A \otimes \langle 0|)(|0\rangle \otimes |0\rangle)(\langle 0| \otimes \langle 0|)(\mathbb{I}_\mathbb{A} \otimes |0\rangle) + (\mathbb{I}_A \otimes \langle 1|)(|0\rangle \otimes |0\rangle)(\langle 0| \otimes \langle 0|)(\mathbb{I}_\mathbb{A} \otimes |1\rangle)$$

$$= (|0\rangle \otimes \langle 0|0\rangle)(\langle 0| \otimes \langle 0|0\rangle) + (|0\rangle \otimes \langle 1|0\rangle)(\langle 0| \otimes \langle 0|1\rangle) = |0\rangle\langle 0| \in Lin(\mathcal{H}_A)$$

[3] is a morphism from the composite to the single system, as seen in ACT

Example:
Let $|\Psi\rangle_{AB} = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B)$ be the maximally entangled state. We can denote the bases $\{|0\rangle_A, |1\rangle_A\}, \{|0\rangle_B, |1\rangle_B\}$ in vector form:

$$|0\rangle_A = \begin{bmatrix}1\\0\end{bmatrix}_A, |1\rangle_A = \begin{bmatrix}0\\1\end{bmatrix}_A, |0\rangle_B = \begin{bmatrix}1\\0\end{bmatrix}_B, |1\rangle_B = \begin{bmatrix}0\\1\end{bmatrix}_B$$

The composite basis elements $\{|0\rangle_A \otimes |0\rangle_B, |1\rangle_A \otimes |1\rangle_B\}$ are thus

$$|0\rangle_A \otimes |0\rangle_B = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \quad |1\rangle_A \otimes |1\rangle_B = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$$

Thus,

$$|\Psi\rangle_{AB} = \begin{bmatrix}\frac{1}{\sqrt{2}}\\0\\0\\\frac{1}{\sqrt{2}}\end{bmatrix} \implies |\Psi\rangle\langle\Psi|_{AB} = \begin{bmatrix}\frac{1}{2} & 0 & 0 & \frac{1}{2}\\0 & 0 & 0 & 0\\0 & 0 & 0 & 0\\\frac{1}{2} & 0 & 0 & \frac{1}{2}\end{bmatrix}$$

Now, suppose we only wish to measure subsystem A for $|0\rangle_A$ or $|1\rangle_B$, so we will measure $|\psi\rangle_{AB}$ with $(\sigma_z)_A \otimes \mathbb{I}_B$. The marginal probabilities are as follows.

$$P(0) = Tr[(|0\rangle\langle 0| \otimes \mathbb{I}_B)|\Psi\rangle\langle\Psi|] = Tr_A[|0\rangle\langle 0| \cdot Tr_B[\mathbb{I}_B|\Psi\rangle\langle\Psi|]]$$

$$= Tr_A[|0\rangle\langle 0| \cdot Tr_B[|\Psi\rangle\langle\Psi|]] = Tr_A[|0\rangle\langle 0| \cdot \frac{1}{2}Tr_B[\begin{bmatrix}1 & 0 & 0 & 1\\0 & 0 & 0 & 0\\0 & 0 & 0 & 0\\1 & 0 & 0 & 1\end{bmatrix}]] = \frac{1}{2}$$

In Dirac notation, it becomes less cumbersome:

$$P(0) = Tr[(|0\rangle\langle 0| \otimes \mathbb{I}_B)|\Psi\rangle\langle\Psi|] = Tr_A[|0\rangle\langle 0| \cdot Tr_B[\mathbb{I}_B|\Psi\rangle\langle\Psi|]] = Tr_A[|0\rangle\langle 0| \cdot Tr_B[|\Psi\rangle\langle\Psi|]]$$

$$= Tr_A[|0\rangle\langle 0| \cdot Tr_B[\frac{1}{2}(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|)]] = Tr_A[|0\rangle\langle 0|]$$

2.4 Entanglement

Multiple quantum systems are described via the tensor product. For two systems A, B (or $\mathcal{H}_A, \mathcal{H}_B$), we can either denote the composite system as $AB, \mathcal{H}_{AB}$ or $\mathcal{H}_A \otimes \mathcal{H}_B$.

Operators are subscripted to denote which system they belong to. For example, $M_A \in Lin(\mathcal{H}_A)$ and $M_{AB} \in Lin(\mathcal{H}_{AB})$. We can also denote transformations of systems by writing $M_{AB|A} \in Lin(\mathcal{H}_A, \mathcal{H}_{AB})$.

Definition 2.17 (Product State)**.** *Let $|\psi\rangle \in \mathcal{H}_A, |\phi\rangle \in \mathcal{H}_B$ be pure states in their respective systems. We will first denote correspondence to a system by subscripting the kets, i.e. $|\psi\rangle_A, |\phi\rangle_B$. The* ***product state*** *of these two states in $\mathcal{H}_{\mathcal{AB}}$ is*

$$|\psi\rangle_A \otimes |\phi\rangle_B \tag{60}$$

All superpositions of product states are contained in $\mathcal{H}_{AB}$ since it is a vector space.

Definition 2.18 (Entangled State)**.** *Any composite state in $\mathcal{H}_{AB}$ that is* ***not a product state*** *is an entangled state.*

Example:
The state

$$|\Psi\rangle_{AB} = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B)$$

is an entangled state. It is actually referred to as the **maximally entangled state**.

The weirdness of entanglement comes from the fact that the state is composite. Consider $|\Psi\rangle_{AB}$ as above. Measuring is with the POVMs $\{\Pi_{AB}, \mathbb{I}_{AB} - \Pi_{AB}\}$ for $\Pi_{AB} = |\Psi\rangle\langle\Psi|$ is as good as deterministic, but measuring with $\sigma_Z \otimes \mathbb{I}_B$ is probabilistic. This is very strange, as the joint 'probability' is deterministic, but not the marginals...

2.4.1 Bell Bases and Weyl-Heisenberg Operators

For two spaces $\mathcal{H}_A, \mathcal{H}_B$ of orthonormal bases $\{|a_j\rangle\}_{j=1}^{d_A}, \{|b_j\rangle\}_{j=1}^{d_B}$, then $\{|a_j\rangle \otimes |b_k\rangle\}$ is an orthonormal basis of $\mathcal{H}_{AB}$.

Definition 2.19 (Bell Basis)**.** *The basis of entangled two qubit states can be described by*

$$\begin{aligned} |\Phi_{00}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \quad |\Phi_{01}\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \\ |\Phi_{10}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \quad |\Phi_{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \end{aligned} \tag{61}$$

Let $|\Phi\rangle$ denote $|\Phi_{00}\rangle$, then all 3 other states are describable as:

$$|\Phi_{jk}\rangle = \mathbb{I} \otimes (\sigma_X^j \sigma_Z^k)|\Phi\rangle \tag{62}$$

Definition 2.20 (Canonical Maximally Entangled State)**.** *For d-dimensional systems, we define this as:*

$$|\Phi\rangle_{AB} = \frac{1}{\sqrt{d}} \sum_{k \in \mathbb{Z}_d} |k\rangle_A \otimes |k\rangle_B \tag{63}$$

Definition 2.21 (Weyl-Heisenberg Operators)**.** *We define the shift and clock operators as:*

$$U = \sum |k+1\rangle\langle k|, \quad V = \sum \omega^k |k\rangle\langle k| \tag{64}$$

We notice that we can now define a basis for entangled states in d dimensions using:

$$|\Phi_{jk}\rangle_{AB} = \mathbb{I}_A \otimes U_B^j V_B^k |\Phi\rangle_{AB} \tag{65}$$

2.4.2 Classical-Quantum States

Density operators can also be used to represent classical information (diagonal operators). Composite states with a classical and quantum part can thus be represented with the tensor of each state.

$$\rho_X = \sum_{x \in \mathcal{X}} P_X(x) |b_x\rangle\langle b_x| \tag{66}$$

and suppose we have a quantum state ρ_A (a collection of $\varphi_A(x)$[4] states) that we can index with x, we thus get

$$\rho_{XA} = \sum_{x \in \mathcal{X}} P_X(x) |b_x\rangle\langle b_x| \otimes \varphi_A(x) \tag{67}$$

We can relate CQ states to **ensemble decompositions** of density operators. Consider

$$\rho_A = \sum_{x=1}^{n} \varphi_A(x) \tag{68}$$

with the φ_A subnormalized, so that $P_X(x) = Tr_A[\varphi_A(x)]$ forms a probability distribution. We can let $\phi_A(x) = \varphi_A(x)/P_X(x)$, and we thus have

$$\rho_A = \sum_{x \in \mathcal{X}} P_X(x) \phi_A(x) \tag{69}$$

2.4.3 Operator-Vector Isomorphism

This section will demonstrate that we can 'unravel' linear operators into bipartite vectors.

Definition 2.22 (Unnormalized Maximally Entangled State)**.** *Let $\{|a_j\rangle\}_{j=1}^{d_A}$ be an orthonormal basis of $\mathcal{H}_A$ and $\mathcal{H}_{A'} \cong \mathcal{H}_A$.*

$$|\Omega\rangle_{AA'} = \sum_{j=1}^{d_A} |a_j\rangle_A \otimes |a'_j\rangle_{A'} \tag{70}$$

Let $V : Lin(\mathcal{H}_A, \mathcal{H}_B) \to \mathcal{H}_A \otimes \mathcal{H}_B$ be a map from linear operators $M_{B|A}$ to bipartite vectors by acting as:

$$V : M_{B|A} \mapsto \mathbb{I}_A \otimes M_{B|A'} |\Omega\rangle_{AA'} \tag{71}$$

with inverse

$$V^{-1} : |\Psi\rangle_{AB} \mapsto_{AA'} \langle\Omega|\Psi\rangle_{A'B} \tag{72}$$

For more intuition, consider this:

$$\begin{aligned} M_{B|A} \in Lin(\mathcal{H}_{A'}, \mathcal{H}_B), \quad |\Omega\rangle_{AA'} \in Lin(\mathbb{C}, \mathcal{H}_A \otimes \mathcal{H}_{A'}) \\ \implies M_{B|A}|\Omega\rangle_{AA'} \in Lin(\mathbb{C}, \mathcal{H}_A \otimes \mathcal{H}_B) \cong \mathcal{H}_A \otimes \mathcal{H}_B \end{aligned} \tag{73}$$

For the inverse it is a bit trickier.

$$\begin{aligned} {}_{AA'}\langle\Omega|\Psi\rangle_{A'B} \in Lin(\mathcal{H}_A, \mathcal{H}_B) &= (\sum_j \langle b_j|_A \otimes \langle b_j|_{A'})(\sum_{kl} \Psi_{jk} |b_j\rangle_{A'} \otimes |b'_j\rangle_B) \\ &= \sum_{jk} \Psi_{jk} |b'_k\rangle_B \langle b_j|_A \in Lin(\mathcal{H}_A, \mathcal{H}_B) \end{aligned} \tag{74}$$

A good way to think about it is to consider the operator-vector isomorphism to be a 'bra-ket' inverter. Like it's going to transfer $|\cdot\rangle$ to $\langle\cdot|$ on subsystem A via application or inner product with $|\Omega\rangle_{AA'}$

pemnis

[4] we suppose that $\mathcal{H}_A \cong \mathcal{H}_X$

3 Quantum Channels

3.1 Classical and Quantum Channels

Just as with classical channels, quantum channels are meant to describe the change in an experimental setup due to time evolution, external interference, measurement, and so forth.

Quantum channels are supposed to relay quantum states to other quantum states, i.e. map density operators to density operators.

Definition 3.1 (Superoperator)**.** *A **superoperator** is a map $\mathcal{E}_{B|A}$ which acts on density operators in A and outputs density operators in B. Formally,*

$$\mathcal{E}_{B|A} : Lin(\mathcal{H}_A) \to Lin(\mathcal{H}_B) : \rho_A \mapsto \sigma_B \tag{75}$$

*The **set of superoperators** from $Lin(\mathcal{H}_A)$ to $Lin(\mathcal{H}_B)$ is denoted $Map(\mathcal{H}_A, \mathcal{H}_B)$.*

Definition 3.2 (Trace Preserving)**.** *A superoperator $\mathcal{E}$ is said **trace preserving** if*

$$Tr[\mathcal{E}[\rho_A]] = Tr[\rho_A] \quad \forall \rho_A \in Lin(\mathcal{H}_A) \tag{76}$$

Definition 3.3 (Complete Positivity)**.** *Unlike regular operators, a superoperator $\mathcal{E}_{B|A}$ can also be **completely positive**, if $\mathcal{E}_{B|A} \otimes \mathbb{I}_R$ is positive for all $\mathcal{H}_R$.*

Definition 3.4 (Quantum Channel)**.** *A **quantum channel** is a superoperator $\mathcal{E}_{B|A}$ that is*

1. *completely positive*
2. *trace-preserving*

*The **set of channels from** $Lin(\mathcal{H}_A)$ **to** $Lin(\mathcal{H}_B)$ is denoted $Chan(\mathcal{A}, \mathcal{H}_B)$ or $Map(\mathcal{A}, \mathcal{H}_B)$.*

3.2 Kraus Representation

The set of completely positive superoperators is closed under sum and addition. It is also reasonable to assume that general quantum operations $\mathcal{E} : Lin(\mathcal{H}) \to Lin(\mathcal{H})$ must satisfy:

1. $\mathcal{E}$ is completely positive
2. $0 \leq Tr[\mathcal{E}[\rho]] \leq 1$ since ρ is a density operator
3. $\mathcal{E}$ is linear, i.e. $\mathcal{E}[\sum_i p_i \rho_i] = \sum_i p_i \mathcal{E}[\rho_i], \quad \forall p_i \geq 0, \sum_i p_i = 1$

Theorem 3.5 (Kraus / Operator-Sum Representation)**.** *A superoperator $\mathcal{E}_{B|A} \in Map(\mathcal{H}_A, \mathcal{H}_B)$ is completely positive **if and only if** there exists a set of **Kraus operators** $\{K(j) \in Lin(\mathcal{H}_A, \mathcal{H}_B)\}_{j=1}^n$ such that*

$$\mathcal{E}_{B|A} : \rho_A \mapsto \sum_{j=1}^{n} K_{B|A}(j) \rho_A K_{B|A}^*(j) \tag{77}$$

Moreover, $\mathcal{E}_{B|A}$ is trace-preserving if

$$\sum_{j=1}^{n} K_{B|A}(j) K_{B|A}^*(j) = \mathbb{I}_A \tag{78}$$

This representation has very important implications, notably that **anything can be seen as a quantum channel**.

Example:
Any state $\rho \in Lin(\mathcal{H})$ is a channel $1 \mapsto \rho$. If $\rho = |\psi\rangle\langle\psi|$ is a pure state, then quite obviously, the Kraus operators are $\{|\psi\rangle\}$.

For an arbitrary ρ, any pure state ensemble decomposition $\{(P(x), |\psi_X\rangle)\}$ gives rise to Kraus operators

$$K(x) = \sqrt{P(x)}|\psi_x\rangle$$

Thus if a channels acts probabilistically on the input, we can easily represent its action via Kraus operators.

3.3 Single Qubit Quantum Noise Channels

Here we will explore a few examples of typical noise channels in operator-sum representation.

3.3.1 Bit-flip, Phase-flip and Bit-Phase-flip Channels

These channels are also called **Pauli channels**, as their Kraus operators are proportional to the Pauli matrices. Pauli channels thus always look like this:

$$\mathcal{E}_{Pauli}: \ \rho \mapsto \sum_{jk} P(j,k)\sigma_X^j \sigma_Z^k \rho \sigma_Z^k \sigma_X^j$$

for a distribution $P(j,k)$.

A quick reminder that the state of a single qubit can be vectorially described as

$$\rho = \frac{1}{2}(\mathbb{I} + \vec{r}\cdot\vec{\sigma}) = \frac{1}{2}\begin{bmatrix} 1+r_x & r_x - ir_y \\ r_x + ir_y & 1-r_z \end{bmatrix} \tag{79}$$

Definition 3.6 (Bit-flip Channel)**.** *The bit-flip channel describes a flip (via σ_X) of the qubit, which appears with probability p. I.e. with probability $1-p$, it is the identity channel, and with probability p it is σ_X. This gives rise to the Kraus operators:*

$$K(0) = \sqrt{1-p}\mathbb{I}, \quad K(1) = \sqrt{p}\sigma_X \tag{80}$$

Its action is thus

$$\begin{aligned} \mathcal{E}_{BF}[\rho] &= \sum_{j=0^1} K(j)\rho K^*(j) = K(0)\rho K^*(0) + K(1)\rho K^*(1) \\ = (1-p)\rho + p\sigma_X\rho\sigma_X &= \frac{(1-p)}{2}(\mathbb{I} + \vec{r}\cdot\vec{\sigma}) + \frac{p}{2}(\sigma_X^2 + r_x\sigma_X^3 + r_y\sigma_X\sigma_Y\sigma_X + r_z\sigma_X\sigma_Z\sigma_X) \\ &= \frac{(1-p)}{2}(\mathbb{I} + \vec{r}\cdot\vec{\sigma}) + \frac{p}{2}(\mathbb{I} + r_x\sigma_X - r_y\sigma_Y - r_z\sigma_Z) \\ &= \frac{1}{2}(\mathbb{I} + r_x\sigma_X + (1-2p)r_y\sigma_Y + (1-2p)r_z\sigma_Z) \end{aligned} \tag{81}$$

which is equivalent to a rescaling of the y and z components by $1-2p$.

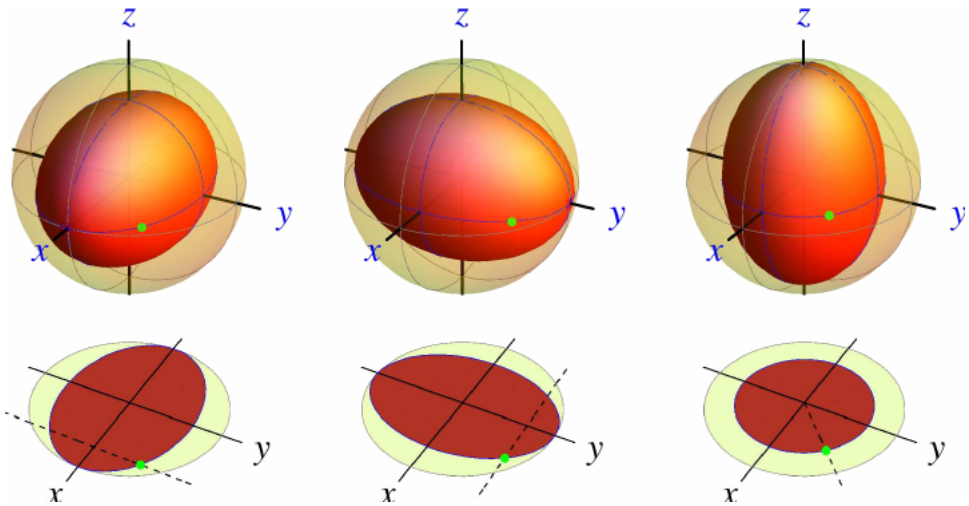


Figure 2: (Left) Bit-flip, (Middle) Phase-flip and (Right) Bit-Phase-flip

Definition 3.7 (Phase-flip Channel)**.** *Analogously to the bit-flip channel, it flips the phase with probability p using the Pauli Z σ_Z operator. Its Kraus operators are thus:*

$$K(0) = \sqrt{1-p}\mathbb{I}, \quad \sqrt{p}\sigma_Z \tag{82}$$

Its action is described as follows:

$$\mathcal{E}_{PF}[\rho] = \frac{1}{2}(\mathbb{I} + (1-2p)r_x\sigma_X + (1-2p)r_y\sigma_Z + r_z\sigma_Z) \tag{83}$$

which corresponds to a rescaling of the x and y components with factor $1-2p$.

Definition 3.8 (Bit-Phase-flip Channel)**.** *Finally, this channel is mediated via the Pauli Y σ_Y operator. Note that $\sigma_Y = i\sigma_Z\sigma_Z$. Thus its Kraus operators are*

$$K(0) = \sqrt{1-p}\mathbb{I}, \quad K(1) = \sqrt{p}\sigma_Y \tag{84}$$

and its action is

$$\mathcal{E}_{BPF}[\rho] = \frac{1}{2}(\mathbb{I} + (1-2p)r_x\sigma_X + r_y\sigma_Y + (1-2p)r_z\sigma_Z) \tag{85}$$

Visually, all these channels describe a 'compression' along their respective Pauli axis, see their action on the Bloch sphere in Fig 2.

3.3.2 Depolarizing Channel

This channels maps any qubit to the maximally entangled qubit ($\frac{1}{2}\mathbb{I}$) with probability p, i.e. the action of the channel is

$$\mathcal{E}_D[\rho] = (1-p)\rho + \frac{p}{2}\mathbb{I} \tag{86}$$

We can convert this into operator-sum representation using the relation

$$\mathbb{I} = \frac{1}{2}(\rho + \sigma_X\rho\sigma_X + \sigma_Y\rho\sigma_Y + \sigma_Z\rho\sigma_Z) \tag{87}$$

thus the action of the channel becomes

$$\begin{aligned}\mathcal{E}_D[\rho] &= (1-p)\rho + \frac{p}{4}(\rho + \sigma_X\rho\sigma_X + \sigma_Y\rho\sigma_Y + \sigma_Z\rho\sigma_Z) \\ &= (1-\frac{3p}{4})\rho + \frac{p}{4}(\sigma_X\rho\sigma_X + \sigma_Y\rho\sigma_Y + \sigma_Z\rho\sigma_Z)\end{aligned} \tag{88}$$

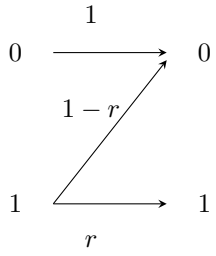
and since all Pauli operators are **self-adjoint**, then the Kraus operators of the depolarizing channel are:

$$K(0) = \sqrt{1-\frac{3p}{4}}\mathbb{I}, \quad K(1) = \sqrt{\frac{p}{4}}\sigma_X, \quad K(2) = \sqrt{\frac{p}{4}}\sigma_Y, \quad K(3) = \sqrt{\frac{p}{4}}\sigma_Z \tag{89}$$

Moreover, the depolarizing channel is a combination of the bit-flip, phase-flip and bit.phase-flip channels. It rescales the entire Bloch sphere with a factor of $(1-2p)^2$.

3.3.3 Amplitude Dampening Channel

This channel is somehow analogous to the classical Z-channel:



The amplitude dampening channel describes **spontaneous emission**, in which an excited state $|1\rangle$ jumps to the ground state $|0\rangle$ with probability $p \in [0,1]$. The Kraus operators of this channel are

$$K(0) = \begin{bmatrix}1 & 0\\ 0 & \sqrt{1-p}\end{bmatrix} \quad K(1) = \begin{bmatrix}0 & \sqrt{p}\\ 0 & 0\end{bmatrix} \tag{90}$$

3.4 Choi Isomorphism

In classical information theory, we can define a joint distribution between the input and output space. This is also doable in the case of quantum channels, and on top of that via the **Choi isomorphism**, the action of a quantum channel is made describable as a marginalization over the input.

Definition 3.9 (Choi Map)**.** *For $\mathcal{H}_A \cong \mathcal{H}_{A'}$, the Choi map turns quantum operation into bipartite states:*

$$C : Map(\mathcal{H}_A, \mathcal{H}_B) \to Lin(\mathcal{H}_A \otimes \mathcal{H}_B) : \mathcal{E}_{B|A} \mapsto \mathcal{E}_{B|A'}[\Omega_{AA'}] \tag{91}$$

Definition 3.10 (Choi Operator)**.** *The Choi operator of a quantum operation $\mathcal{E}_{B|A}$ is the resulting of applying the choi map to it:*

$$C(\mathcal{E}_{B|A}) = E_{AB} \in Lin(\mathcal{H}_A \otimes \mathcal{H}_B) \tag{92}$$

Remark 8. *Like in the operator-vector isomorphism, it is more intuitive to think of it as 'swapping' bra's and ket's in the Kraus operators of the channel. Applying then channel to $\Omega_{AA'}$ will 'collapse' the opposing bra-kets into 'aligned' kets.*

Theorem 3.11 (Choi Isomorphism of Superoperators and Bipartite Operators)**.** *The Choi Map C is an isomorphism between $Map(\mathcal{H}_A, \mathcal{H}_B)$ and $Lin(\mathcal{H}_A \otimes \mathcal{H}_B)$: The inverse is C^{-1} acts as follows:*

$$C^{-1}(M_{AB}) : \rho_A \mapsto Tr_A[\mathcal{T}_A[\rho_A]M_{AB}] \tag{93}$$

where $\mathcal{T}_A$ denotes the transpose over system A.

Proof. TODO (bruh) □

Remark 9. *It turns out that Choi operators that correspond to completely positive superoperators and positive semidefinite bipartite operators. This implies that the study of channels is reduced to the study of positive semidefinite operators.*

Theorem 3.12 (Choi Representation)**.** *A superoperator $\mathcal{E}_{B|A}$ is completely positive **if and only if** it's corresponding Choi operator $E_{AB} = C(\mathcal{E}_{B|A}) \geq 0$.*

*Moreover, it is trace-preserving **if and only if** $Tr_B[E_{AB}] = \mathbb{I}_A$*

Proof. TODO (bruh^2) □

3.5 CQ and QC Channels

As we have seen in Quantum Probability Theory, classical probability distributions are encapsulated in diagonal quantum density operators. Let $Lin(\mathcal{H}_X)$ and $Lin(\mathcal{H}_A)$ denote a classical system and a quantum system respectively.

Definition 3.13 (Classical-Quantum Channel)**.** *A device that can **prepare one of many quantum state** turns classical information into quantum information, where the input $|x\rangle\langle x|_X$ is mapped to $\rho_A(x)$, for an orthonormal basis $\{|x\rangle_X\}$.*

Considerations respective to off-diagonal inputs $|x\rangle\langle y|$ are simply discarded by the channel.

*The **Kraus operators of such a CQ channel are**:*

$$\{K_{A|X}(x) = |\psi_x\rangle_A\langle x|_X\} \tag{94}$$

Definition 3.14 (Quantum-Classical Channels)**.** *On the other hand, measurements are a good example of QC channels, as they produce classical information out of quantum information. The mneasurement can be seen as a channel taking a density operator ρ_A as an input, and maps it to an elements of a probability distribution.*

$$\mathcal{M}_{X|A} : \rho_A \mapsto \sum_{x \in \mathcal{X}} Tr_A[\Lambda_A(x)\rho_A]|x\rangle\langle x|_X \tag{95}$$

for a POVM $\{\Lambda_A(x) : x \in \mathcal{X}\}$.

Remark 10. *By the Kraus representation, it is actually true that **all CQ channels are state preparations and all QC channels are measurements.***

Proposition 3.15. *The Kraus representation Theorem also implies that every quantun channel can be regarded as a quantum measurement, followed by forgetting the measurement result, i.e.*

$$\mathcal{E}_{B|A} = Tr_X \circ Q_{XB|A} \tag{96}$$

Proof. Let $\{K(x)\}_{x=1}^{n}$ be the Kraus operators of $\mathcal{E}_{B|A}$, then $\{\hat{K}(X) = K(x) \otimes |x\rangle_X\}$ are also a valid set of Kraus operators for an orthonormal basis $\{|x\rangle_X\}$.

The quantum instrument is supposed to output

$$Q_{XB|A}[\rho_A] = \sum_{x=1}^{n} |x\rangle\langle x|_X \otimes K_{B|A}(x)\rho_A K_{B|A}^*(x) \tag{97}$$

Then clearly $\mathcal{E}_{B|A} = Tr_X \circ Q_{XB|A}$ □

3.6 Stinespring

This section will explain how to interpret measurements **and** quantum channels as unitary dynamics. Simply by 'expanding' their degrees of freedom.

3.6.1 Purification

Definition 3.16. *A **purification** of ρ_A is a normalized composite state $|\Psi_{AB}\rangle$ such that*

$$Tr[|\Psi\rangle\langle\Psi|_{AB}] = \rho_A \tag{98}$$

System B can be referred to as the purifying system. Notice that the above formula 'forces' Ψ_{AB} to be a product state (hence pure state), since tracing out B yields ρ_A.

We can now consider mixed states as pure states on a higher-dimensional system. By eigendecomposition, we can break down any state into its eigenvectors and 'append' a basis element of B to each eigenvectors.

$$\rho_A = \sum_{k=1}^{r} \lambda_k |\xi_k\rangle\langle\xi_k| \implies \Psi_{AB} = \sum_{k=1}^{r} \lambda_k |\xi_k\rangle\langle\xi_k| \otimes |b_k\rangle\langle b_k| \tag{99}$$

for any orthonormal basis of B $\{|b_k\rangle\}_{k=1}^{r}$, where r is the number of non-zero eigenvalues of ρ_A. This implies that the system B must have dimension at least as large as the rank of ρ_A.

Remark 11. *If ρ_A is represented in an ensemble decomposition as*

$$\rho_A = \sum_{z=1}^{r} \sqrt{P_Z(z)} |\phi_z\rangle_A \tag{100}$$

then it can be purified as follows

$$|\Psi\rangle_{AB} = \sum_{z=1}^{r} \sqrt{P_Z(z)} |\phi_z\rangle_A \otimes |b_z\rangle_B \tag{101}$$

*This above remark points to the fact that **pure state ensembles and purifications are as good as the same concept.***

Definition 3.17. *The **canonical purification** is very useful, and is as follows:*

$$\rho_A \mapsto |\Psi\rangle_{AA'} = (\sqrt{\rho_A} \otimes \mathbb{I}_{A'})|\Omega_{AA'}\rangle \tag{102}$$

Proposition 3.18 (Schmidt Decomposition)**.** *For any $|\Psi\rangle_{AB}$ there exist orthonormal bases $\{|\xi_j\rangle_A\}_{j=1}^{d_A}$ and $\{|\eta_j\rangle_B\}_{j=1}^{d_B}$, $n \leq \min\{d_A, d_B\}$ and **Schmidt coefficients** $\{s_k\}_{k=1}^{n}$ such that*

$$|\Psi\rangle_{AB} = \sum_{j=1}^{n} s_k |\xi_j\rangle_A \otimes |\eta_j\rangle_B \tag{103}$$

This implies that any composite state in $\mathcal{H}_{AB}$ can be considered as a purification of some state in $\mathcal{H}_A$ or $\mathcal{H}_B$.

Notice that if we consider $\psi\rangle_{AR}$ and $|\varphi\rangle_{AB}$ and their Schmidt decompositions, we get that

$$Tr_R[\psi_{AR}] = Tr_B[\varphi_{AB}] = \sum_{k=1} s_k^2 |\xi_k\rangle\langle\xi_k|_A \tag{104}$$

Now suppose that $\rho_A = |\phi\rangle\langle\phi|_A$ can be purified into [5]

$$|\psi\rangle_{AB} = \sum_{k=1}^{n} s_k |\xi_k\rangle_A \otimes |\eta_k\rangle_B \quad \& \quad |\psi'\rangle_{AC} = \sum_{k=1}^{n} s_k |\xi_k\rangle_A \otimes |\eta'_k\rangle_C \tag{105}$$

then notice that we can related both purifications by using a transformation

$$V_{C|B} : \mathcal{H}_B \to \mathcal{H}_C : |\eta_k\rangle_B \mapsto |\eta'_k\rangle_C$$

Since both bases $\{|\eta_k\rangle_B\}$ and $\{|\eta'_k\rangle_C\}$ are orthonormal, then it is safe to conclude that $V_{C|B}$ must **be an isometry**.

Proposition 3.19. *All purifications of an arbitrary ρ_A are related by isometry or unitary.*

Remark 12. *Using the canonical purification, we can now write **any** purification as follows:*

$$|\Psi'\rangle_{AR} = (\sqrt{\rho_A} \otimes V_{R|A'})|\Omega_{AA'}\rangle \tag{106}$$

3.6.2 Steering

We have just seen that pure state ensembles can be recovered from purifications with the correct measurements. We can extend this concept to **handpicking our ensemble via measurements.** We will see how to pick those measurements,.

Proposition 3.20. *(Unitary relation of ensemble decompositions) For any density operators ρ, let $\{(p_k, |\varphi_k\rangle)\}$ and $\{(q_k, |\phi_k\rangle)\}$ be two different ensemble decompositions of ρ. Then there exists a $n \times n$ unitary matrix U such that*

$$\sqrt{q_j}|\phi\rangle = \sum_{k=1}^{n} U_{jk} \sqrt{p_k}|\varphi_k\rangle \tag{107}$$

*This implies that ensemble decompostions of ρ are **all** related by a unitary operator.*

By this unitary relation, every purification can generate every possible ensemble decomposition by suitable measurement of the purifying system.

[5] via Schmidt

If we have the purifications

$$\Psi_{AB} = \sum_{k=1}^{n} p(x)|\phi_A(x)\rangle\langle\phi_A(x)| \otimes |b_k\rangle\langle b_k| \quad \& \quad \Psi'_{AB} = \sum_{k=1}^{n} q(x)|\varphi_A(x)\rangle\langle\varphi_A(x)| \otimes |b'_k\rangle\langle b'_k| \tag{108}$$

then measuring B with $\Pi_k = |b_k\rangle\langle b_k|$ yields $\{(p_k|\phi_k)\}$ and measuring B with $\Pi'_k = U^*|b_k\rangle\langle b_k|U$ will yield $\{(q_k|\varphi_k)\}$.

Proposition 3.21. *To steer the canonical purification $|\psi\rangle_{AA'}$ to a CQ state by measurement on A', we measure using $\Lambda_{A'}(x)^T$*

Proposition 3.22. *If $|\Psi\rangle_{AB}$ is a purification and $\rho_{XA} = \sum_{k=1}^{n} P_X(x)|x\rangle\langle x| \otimes \rho_A(x)$ is a CQ extension of ρ_A, then there exists a measurement $\mathcal{M}_{X|B}$ on Ψ_{AB} such that*

$$\mathcal{M}_{X|B}[\Psi_{AB}] = \rho_{XA}$$

Moreover, we know that

$$|\Psi\rangle_{AB} = (\sqrt{\rho_A} \otimes V_{B|A'})|\Omega\rangle_{AA'}$$

which relates the purification to the canonical via $V_{B|A}$, so we can relate the measurements via $V_{B|A}$ as follows:

$$\Gamma_B(x) = V_{B|A}\Lambda_A(x)^T V^*_{B|A}$$

such that measuring $\Psi_{AB} with \Gamma_B(x)$ allows us to steer.

3.6.3 Dilation

The dilation of a quantum channel leads to the **Stinespring representation**. It is very analogous to the Choi isomorphism, but for channels (i.e. superoperators). Using the Choi isomorphism, we get:

$$\mathcal{E}_{B|A}[\rho_A] = Tr_A[E_{BA}\rho_A^T] = Tr_A R[V_{BR|A}\rho_A(V^*_{BR_A})] \tag{109}$$

This implies that the action of a quantum channel is nothing more than purifying our density operator to the *environment* R, applying uniform dynamics (via isometries) and tracing out the environment.

Theorem 3.23. *(Strinespring) A map $\mathcal{E}_{B|A}$ is completely positive if and only if there exists $\mathcal{H}_R$ and $V_{BR|A}$ with*

$$\mathcal{E}_{B|A}[\rho_A] = Tr_R[V_{BR|A}\rho_A V^*_{BR|A}] \tag{110}$$

The smallest d_r is no larger than $d_A d_B$, and the map is trace preserving if and only if $V_{BR|A}$ is an isometry.

Proposition 3.24. *(Stinespring Isometries from Kraus Operators) We can directly contruct the Stinespring representation out of a set of Kraus operators:*

$$V_{BR|A} = \sum_{x=1}^{n} K_{B|A}(x) \otimes |b_x\rangle_R \tag{111}$$

Proposition 3.25. *(Unitary Relation of Kraus representations) Let $\{K(i)\}_{i=1}^n$ and $\{K'(j)\}_{j=1}^m$ be two Kraus representations of the the same superoperator $\mathcal{E}$ and set $l = \max\{m, n\}$. There exists a unitary $l \times l$ matrix U such that*

$$K'(j) = \sum_i U_{ji}K(i) \tag{112}$$

This is like a vector - matrix-vector equation.

Definition 3.26. *(Channel Complement) Let $\mathcal{E}_{B|A}$ be a quantum channel. By Stinespring, we can define the* ***channel complement*** *$\mathcal{E}_{|A}[\rho_A] = Tr_B[V_{BR|A}\rho_A V^*_{BR|A}]$. I.e. we are observing the information that is 'left' in the environment upon action of the channel.*

3.7 Relating Kraus, Choi and Stinespring

We have now all the tools to compare the three representations we've covered.

Proposition 3.27. *For any two dilations $V_{BR|A}$ and $V_{BR'|A}$ of $\mathcal{E}_{B|A}$, they are related by a partial isometry $W_{R'|R}$ such that $V_{BR'|A} = W - R'|RV_{BR|A}$. If $dim(\mathcal{H}_{R'}) > dim(\mathcal{H}_R)$, then it is an isometry. If it is an equality, they are related by unitary transformation.*

4 Quantum Communication

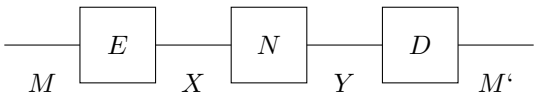
4.1 Basic Resources

The idea in resource simulation is to see how closely we can approximate the identity channel, or more interestingly, an information processing channel.

We will observe how much **classical information** can be transmitted over both **classical** and **quantum channels** & vice-versa.

4.1.1 Classical Information & Channels

We consider a basic chain of encoder-noise-decoder channel, and wish to see how well we can approximate identity: a typical Information Theory problem.



The probability of getting a corresponding input-output pair (M, M') is given by:

$$P(M = M') = \frac{1}{|M|} \sum_{m \in \mathcal{M}} W_{M'|M}(m, m) \tag{113}$$

i.e. the entire channel does not impede on the information transmission.

We can decompose $W_{M'|M}$ into $D \circ N \circ E$, yielding

$$P(M = M') = \frac{1}{|M|} \sum_{m \in \mathcal{M}} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} D_{M'|Y}(m, y) N_{Y|X}(y, x) E_{X|M}(x, m) \tag{114}$$

Since $E_{X|M}(x, m) \leq 1$ for all x and m, we get

$$P(M = M') \leq \frac{1}{|M|} \sum_{m \in \mathcal{M}} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} D_{M'|Y}(m, y) N_{Y|X}(y, x) = \frac{|X|}{|M|} \tag{115}$$

which is easily explained by the fact that we can consider E to be the communication bottleneck, thus we could not transmit more than what E can produce; namely $|X|$ symbols.

On the other hand, we can consider that $N_{Y|X}$ could produce errors, and thus

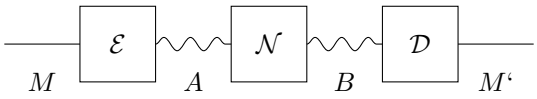
$$P(M = M') \leq \frac{1}{|M|} \sum_{m \in \mathcal{M}} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} D_{M'|Y}(m, y) E_{X|M}(x, m) = \frac{|Y|}{|M|} \tag{116}$$

Proposition 4.1. *The **upper bound on classical information over classical channels** is*

$$P(M = M')_{classical} \leq \frac{\min(|X|, |Y|)}{|M|} \tag{117}$$

4.1.2 Classical Information & Quantum Channels

Our communication system now looks like



- The encoder channel $\mathcal{E}_{A|M}$ creates a quantum density operator $\rho_A(m)$ from classical message $m \in \mathcal{M}$.
- The quantum noise channel $\mathcal{N}_{B|A}$ is an arbitrary quantum channel (trace-preserving & CP): $\rho_A(m) \mapsto \theta_B(m)$
- The decoder channel $\mathcal{D}_{M'|B}$ applies a measurement from POVM element $\Lambda_B(m)$ and outputs $Tr[\Lambda_B(m)\theta_B(m)]$ i.e. it measures by 'looking for' m

The probability of successful transmission is this

$$P(M = M') = \frac{1}{|M|} \sum_m Tr[\Lambda_B(m)\mathcal{N}_{B|A}[\rho_A(m)]] \tag{118}$$

Proposition 4.2. *Again, considering the bottlenecks in $\mathcal{E}_{A|M}$ or in $\mathcal{E}_{B|A}$, we get the* ***bound of information transmission over quantum channels***

$$P(M = M')_{quantum} \leq \frac{min(|A|, |B|)}{|M|} \tag{119}$$

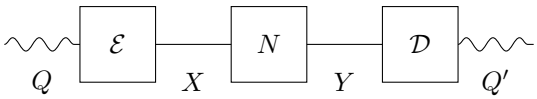
Thus, we have shown that if $|A| = |X|$, we get the same bounds, implying that a single qubit channel cannot transport more than one bit of information. In this regard, quantum channels and classical channels are seemingly equivalent.

4.1.3 Quantum Information & Classical Channels

When quantum information is being sent, we will have to adapt our notion of 'correct transmission'. A good 'metric' of a correct transmission of information would be that **the channel has successfully sent an entangled qubit**. This yields the following probability of success:

$$P_{agree}(\mathcal{N}_Q) = Tr[\Phi_{QQ'}\mathcal{N}_{\mathcal{Q}}[\Phi_{QQ'}]] \tag{120}$$

i.e. this quantity represents the probability of an entangled qubit to be transmitted through the channel. In the following case,



suppose that

- $\mathcal{E}_{X|Q}[\rho_Q] = Tr[\Lambda_Q(x)\rho_Q(x)]$ for a given $x \in \mathcal{X}$
- $N_{Y|X}$ is a regular noise channel
- $\mathcal{D}_{Q'|Y}[y] = \rho_Q(y)$ for the received classical $y \in \mathcal{Y}$

Definition 4.3 (Separable State)**.** *A quantum state ψ in a composite quantum system is said* ***separable*** *if it can be factored into individual states of subsystems.*

An entangled state is by definition not separable

Definition 4.4 (PPT States)**.** *A composite state is said* ***PPT*** *if it has positive partial transpose, i.e. state ρ_{AB} such that $\mathcal{T}_A[\rho_{AB}] \geq 0$. Separable states are necessarily PPT.*

Proposition 4.5 (PPT Bound)**.** *For any PPT state σ_{AB} and the maximally entangled state Φ_{AB}, it follows that*

$$Tr[\Phi_{AB}\sigma_{AB}] \leq \frac{1}{|A|} \tag{121}$$

We will consider the action of our channel on the maximally entangled state $\Phi_{QQ'}$.

$$\mathcal{W}_{Q'|Q}[\Phi_{QQ'}] = \sum_{xy} N_{Y|X}(y,x)\rho_Q(y) \otimes Tr_Q[\Lambda_Q(x)\Phi_{QQ'}] \tag{122}$$

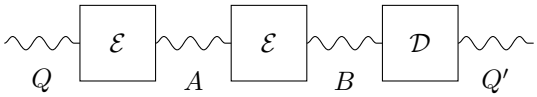
The output is a separable state, since both quantum-classical channels actions are positive, and so is N. Thus, by the PPT bound, we get

$$Tr[\Phi_{QQ'}\mathcal{N}_Q[\Phi_{QQ'}]] \leq \frac{1}{|Q|} \tag{123}$$

This implies that quantum information cannot be efficiently transmitted over classical channels at all!

4.1.4 Quantum Information & Channels

Finally, our communication system is



This case is a bit different. We consider the input to be a composite state $|\Phi_{QQ'}\rangle$, and the output to be a slightly altered state $|\Phi_{QQ'}\rangle$. Thus, we want to measure the 'distance' between $|\Phi_{QQ'}\rangle$ and $|\Phi_{QQ'}\rangle$. This can be done by projecting the initial state onto the output, yielding

$$P(\rho_{QQ'} = \Phi_{QQ'}) = \langle\Phi_{QQ'}|\rho_{QQ'}|\Phi_{QQ'}\rangle = Tr_{QQ'}[\Phi_{QQ'}\rho_{QQ'}] \tag{124}$$

Since $\Phi_{QQ'}$ is a rank-one density operator, then it can count as a POVM to test whether $\rho_{QQ'}$ is a match or not.

To analyse the system, we will need two important inequalities.

Proposition 4.6 (Pinching Inequalities)**.** *For an arbitrary system A and the pinching map $\mathcal{P}_A$ in an arbitrary orthonormal basis, we have for any $S_{AB} \geq 0$*

$$S_{AB} \leq |A|\mathcal{P}_A[S_{AB}] \tag{125}$$

and for any classical-quantum operator S_{XB}, we have

$$S_{XB} \leq \mathbb{I}_X \otimes S_B \tag{126}$$

Combining the two properties yields a very useful inequality for all positive S_{AB}:

$$S_{AB} \leq |A|\mathbb{I} \otimes S_B \tag{127}$$

We know that $\rho_A \leq \mathbb{I}_A$, so the probability of successful transmission of the maximally entangled state can be written as

$$\begin{aligned} P(\rho_{QQ'} = \Phi_{QQ'}) &= Tr_{QQ'}[\Phi_{QQ'}\mathcal{D}[\mathcal{N}[\mathcal{E}[\Phi_{QQ'}]]]] \\ &\leq Tr_{QQ'}[\Phi_{QQ'}\mathcal{D}[\mathcal{N}[|A|\mathbb{I}_A \otimes \Phi_{Q'}]]] = Tr_{QQ'}[\Phi_{QQ'}\mathcal{D}[\mathcal{N}[\frac{|A|}{|Q|}\mathbb{I}_{AQ'}]]] \\ &= \frac{|A|}{|Q|}Tr_Q[\mathbb{I}_Q\mathcal{D}[\mathcal{N}[\mathbb{I}_A]Tr_Q[\frac{1}{|Q'|}\mathbb{I}_Q\Phi_Q]] = \frac{|A|}{|Q|^2}Tr_Q[\mathcal{D}\circ\mathcal{N}[\mathbb{I}_A]] = \frac{|A|^2}{|Q|^2} \end{aligned} \tag{128}$$

Again, by considering the bottleneck on $\mathcal{N}_{B|A}$, we can upper bound it by $\mathcal{N}_{B|A}[\rho_A] \leq \mathcal{N}_{B|A}[\mathbb{I}_A]$ thus by trace-preserving property,

$$\begin{aligned} Tr_{QQ'}[\Phi_{QQ'}\mathcal{D}_{Q|B} \circ \mathcal{N}_{B|A}[\rho_{AQ'}]] &= Tr_{QQ'}[\Phi_{QQ'}\mathcal{D}_{Q|B}[\rho_{BQ'}]] \\ &\leq Tr_{QQ'}[\Phi_{QQ'}\mathcal{D}_{Q|B}[|B|\mathbb{I}_B \otimes \Phi_{Q'}]] = ... = \frac{|B|^2}{|Q|^2} \end{aligned} \tag{129}$$

Proposition 4.7. *Thus, the probability of successful transmission of quantum information over quantum channels is upper bounded by*

$$P(agree)_{quantum-quantum} \leq \frac{\min(|A|^2, |B|^2|)}{|Q|^2} \tag{130}$$

4.2 Superdense Coding and Teleportation

Assisted communication is a technique that requires an extra resource, in this case a quantum state $\rho_{TT'}$, in which the system T is available to the sender and T' is available to the receiver.

Definition 4.8 (Superdense Coding)**.** *This protocol aims to transmit two classical bits over a noiseless one-qubit quantum channel, while exploiting the shared entanglement of the state $\Phi_{TT'}$.*

1. *Alice wants to send two bits (j, k) to Bob*
2. *Alice applies $(\sigma_X^j\sigma_Z^k)_T$ to the T subsystem of $|\Phi\rangle_{TT'}$*
3. *Alice transmits the system T over the noiseless quantum channel*
4. *Bob jointly measures TT' in the Bell basis, and Bob can thus reconstruct the two bits by inferring from the output Bell state*

Proposition 4.9. *Superdense coding is a form of assisted classical communication over a quantum channel, and thus satisfies the bound*

$$P_{agree} \leq \frac{\min\{|A|^2, |B|^2, |A||T|, |B||T|, |A||T'|, |B||T'|\}}{|M|} \tag{131}$$

Definition 4.10 (Teleportation)**.** *This protocol puts the Bell measurement at the sender and the Pauli operators at the receiver, while also exploiting the shared entangled system $|\Phi\rangle_{QQ'}$. The aim is to transmit one qubit of quantum information over a two bit classical channel.* ***The main goal is to transmit a full qubit state using 2 classical bits. The entanglement in necessary to the successful transmission.***

1. *Alice measures the system AT in the Bell basis (A is her quantum system she wishes to transmit)*
2. *Alice gets the measurement output (j,k) an transmits them classically*
3. *Bob then computes $(\sigma_X^j \sigma_Z^k)_{T'}$ from the received bits*
4. *Bob applies the measurement on T' of $|\Phi\rangle_{TT'}$, and can thus infer A*

Proposition 4.11. *Quantum teleportation is a form of assisted quantum communication over classical channels, and thus satisfies;:*

$$P_{agree} \leq \min\{\frac{|X|}{|Q|^2}, \frac{|Y|}{|Q|^2}, \frac{|T|}{|Q|}, \frac{|T'|}{|Q|}\} \tag{132}$$

4.3 Information Disturbance

This quick section will simply show why information cannot simply be 'created'.

Suppose a quantum instrument $\mathcal{Q}_{XA|A}$. If by tracing out the classical X output, we obtain the identity, then necessarily,

$$\mathcal{Q}_{XA|A} = P_X \otimes \mathbb{I}_A \tag{133}$$

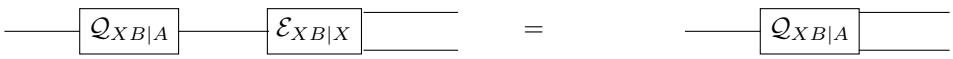
Proposition 4.12. *Let $\mathcal{Q}_{XB|A}$ be a quantum instrument such that*

$$Tr_B \circ \mathcal{Q}_{XB|A} = \mathcal{P}_{X|A}$$

where $\mathcal{P}$ is a pinch map for an arbitrary basis of A. Then there exists a set of density operators $\varphi_B(x)$ such that

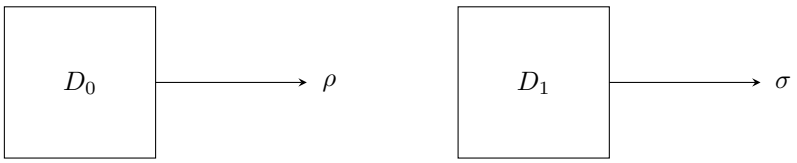
$$\mathcal{Q}_{XB|A} = \mathcal{E}_{XB|X} \circ \mathcal{P}_{X|A}$$

for $\mathcal{E}_{XB|X} : |x\rangle\langle x|_X \mapsto |x\rangle\langle x| \otimes \varphi_B(x)$



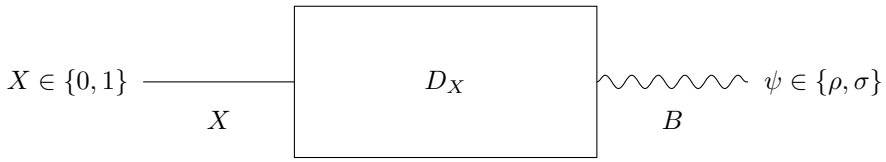
4.4 Discriminating States & Channels

The idea of this section is to distinguish two quantum states ρ and σ. Imagine we have two devices D_0 and D_1, which respectively produce ρ and σ.



Now imagine that **we do not know** which of the two systems is chosen, and a quantum state is output, how do we figure out which one it is?

Formally, we will let $X \sim P_X$ be a classical random variable taking value in $\{0,1\}$, and ρ, σ are states in a quantum system $\mathcal{H}_B$. We can thus reformulate our diagram as



To approximate the resource (D_0 or D_1), we can use a QC channel, which will hopefully correctly 'reconstruct' X. This is a simple measurement channel, whose POVM elements are $\{\Lambda, \mathbb{I} - \Lambda\}$ corresponding to ρ resp. σ.

Definition 4.13 (Conditional Probability of error)**.** *The probability of an incorrect guess (or reconstruction) of X is denoted by*

$$P_{error} = P[\hat{X} \neq X] = \begin{cases} Tr[(\mathbb{I} - \Lambda)\rho] & \text{if } X\text{=}0 \\ Tr[\Lambda\sigma] & \text{if } X\text{=}1 \end{cases} \tag{134}$$

Suppose $P[X = 0] = p$, then the probability of guessing correctly becomes

$$P_{guess} = pTr[\Lambda\rho] + (1-p)Tr[(\mathbb{I} - \Lambda)\sigma] = (1-p) + Tr[\Lambda(p\rho - (1-p)\sigma)] \tag{135}$$

We however lack a way to find this optimal measurement Λ^*. We will present two approaches to find it.

4.4.1 Bayesian Hypothesis Testing

Classical Bayesian hypothesis testing is done by obeserviing the **likelihood ratio**.

$$\frac{P(X_1|B)}{P(X_2|B)} = \frac{P(B|X_1)P(X_1)}{P(B|X_2)P(X_2)} \begin{cases} > 1 \implies X_1 \\ < 1 \implies X_2 \end{cases} \tag{136}$$

Returning to our case, we have

$$\frac{p\rho}{(1-p)\sigma} \begin{cases} > 1 \implies \rho \\ < 1 \implies \sigma \end{cases} \tag{137}$$

Since by assumption Λ must correspond to ρ and $\mathbb{I} - \Lambda$ must correspond to σ, then it makes sense to have Λ 'project' onto the space where $\frac{p\rho}{(1-p)\sigma} > 1$ holds.

$$\begin{aligned} \{\frac{p\rho}{(1-p)\sigma}\} > 1 &\equiv \{p\rho - (1-p)\sigma > 0\} \\ &\equiv \{\rho - \frac{1-p}{p}\sigma > 0\} \end{aligned} \tag{138}$$

We thus define

$$\Lambda(\gamma) = \{\rho - \gamma\sigma > 0\} \tag{139}$$

the projection onto the positive space of $\{\rho - \gamma\sigma\}$, which is maximal in $\frac{1-p}{p}$. Formulated as **variational problems**, we wish to maximize

$$f(M) = \max_{\Lambda} \{Tr[\Lambda M] : 0 \leq \Lambda \leq \mathbb{I}, \Lambda \in Lin(\mathcal{H})\} \tag{140}$$

Since $\Lambda \leq \mathbb{I} \implies \Lambda M \leq M$, we can instead look for the minimal $\Theta \geq M$.

$$f^{\dagger}(M) = \min_{\Theta} \{Tr[\Theta] : \Theta \geq M, \Theta \in Lin(\mathcal{H})\} \tag{141}$$

By considering the eigenbasis of M, we can write $M = \{M > 0\}M + \{M \leq 0\}M ? \{M\}_+ + \{M\}_-$, that M is the sum of it's projective spaces. Since the contribution of $\{M\}_-$ will only decrease the value of $Tr[M]$, it is now clear that $\Lambda(\frac{1-p}{p})$ is the optimal solution.

Since $Tr[M] = Tr[\{M\}_+] + Tr[\{M\}_-]$ and $||M||_1 = Tr[\{M\}_1] - Tr[\{M\}_-]$, our objective function is

$$f(M) = \frac{1}{2}(||M||_1 + Tr[M])$$

and we find that the optimal probability of guessing is

$$P(X|B)_{guess} = \frac{1}{2}(1 + ||p\rho - (1-p)\sigma||_1)$$

4.4.2 Neyman-Pearson Hypothesis Testing

This method does not use priors. Rather first observe the **testing region**

$$\mathcal{R}(\rho, \sigma) = \{(\alpha, \beta) : Tr[\Lambda\rho] = \alpha, Tr[\Lambda\sigma] = \beta, 0 \leq \Lambda \leq \mathbb{I}\} \tag{142}$$

and we optimize (by minimizing) one error probability for a **fixed success probability**. In our case, $Tr[\Lambda\rho]$ is fixed, so we will minimize $Tr[\Lambda\sigma]$. This yields the **boundary of our testing region**, over which the minimum is the minimum error $Tr[\Lambda\sigma]$.

$$\beta_{\alpha}(\rho, \sigma) = \min_{\Lambda}\{Tr[\Lambda\sigma] : Tr[\Lambda\rho] = \alpha, 0 \leq \Lambda \leq \mathbb{I}\} \tag{143}$$

However, the optimal Λ for our case is not very clear, so we will try to reformulate by maximizing over other parameters. Consider a positive density operator Θ. It then holds that

$$Tr[\Lambda\Theta] \leq Tr[\Theta] \implies mTr[\Lambda\Theta] \geq m\alpha - Tr[\Theta]$$

Furthermore, we will then require that

$$Tr[\Lambda\sigma] \geq mTr[\Lambda\rho] - Tr[\Theta] \implies \sigma \geq m\rho - \Theta$$

Since we wish to minimize σ, we can make ends meet by maximizing $m\rho - \Theta$ since $\Lambda \geq 0$ will never incur any sign changes.

Our problem is now

$$\beta^{\dagger}(\rho, \sigma) = \max_{m,\Theta}\{m\alpha - Tr[\Theta] : m\rho - \Theta \leq \sigma : \Theta \geq 0, \Theta \in Lin(\mathcal{H}), m \in \mathbb{R}\} \tag{144}$$

4.4.3 Semidefinite Programming

Both problems can be set up as semidefinite progams. We will first need a quick description. A semidefinite program is an optimization problem that aims to find the infimum (resp. supremum) of the trace od a matrix product, subject to inequality constraints. Formally,

$$\inf_X Tr[AX] \quad s.t. \quad \mathcal{L}[X] \leq B \tag{145}$$

is the **primal** of a problem, and

$$\sup_X Tr[BY] \quad s.t. \quad \mathcal{L}^*[Y] \geq A \tag{146}$$

is the **dual** statement.

In our case of distinguishability, we use a lot of **block matrices** to define the multiple conditions.

4.4.4 Distinguishability Formula

After our derivation of how to discriminate states, we can now write it in closed form for $p = \frac{1}{2}$.

Definition 4.14 (Distinguishability of States)**.** *For two quantum states ρ, σ we have*

$$\begin{aligned} \delta(\rho, \sigma) &= \max_{\Lambda} \{Tr[\Lambda(\rho - \sigma)] : 0 \leq \Lambda \leq 1\} \\ &= \min_{\theta} \{Tr[\theta] : \theta \geq \rho - \sigma, \theta \geq 0\} \\ &= \frac{1}{2} ||\rho - \sigma||_1 \end{aligned} \tag{147}$$

i.e. it is the maximal difference in trace that ρ and σ have under measurement Λ (interpretable as probability of not mixing them up).

Definition 4.15 (Distinguishability of Channels)**.** *Two channels $\mathcal{E}, \mathcal{F}$ are simply maps from operators to operators. We will thus compare their output states given the same input*

$$\delta(\mathcal{E}, \mathcal{F}) = \max_{\rho_{AR}} \max_{\Lambda_{BR}} \{Tr[\Lambda_{BR}(\mathcal{E}[\rho_{AR}] - \mathcal{F}[\rho_{AR}])]\} \tag{148}$$

which is analogous to state distinguishability.

Proposition 4.16 (Properties of distinguishability)**.** *For all states ρ, σ, θ and channels $\mathcal{E}, \mathcal{E}', \mathcal{F}, \mathcal{F}', \mathcal{G}$*

1. *$\delta(\rho, \theta) \leq \delta(\rho, \sigma) + \delta(\sigma, \theta)$ (triangle inequality)*
2. *$\delta(\mathcal{E}, \mathcal{F}) \leq \delta(\mathcal{G} \circ \mathcal{E}, \mathcal{G} \circ \mathcal{F})$ (DPI)*
3. *$\delta(\mathcal{E} \circ \mathcal{F}, \mathcal{E}' \circ \mathcal{F}') \leq \delta(\mathcal{E} \circ \mathcal{E}', \mathcal{F} \circ \mathcal{F}')$ (monotonicity)*

Another way to distinguish channels can be done **via Choi** and in SDP form:

Proposition 4.17. *(SDP Form of Channel Distinguishability) Let E_{BA} and E'_{BA} be the Choi operators of the channels $\mathcal{E}$ and $\mathcal{E}'$ we are trying to distinguish. The the distinguishability is*

$$\delta(\mathcal{E}, \mathcal{E}') = \max_{\rho_A, \Gamma_{BA}} \left\{ Tr[\Gamma_{BA}(E_{BA} - E'_{BA})] : Tr[\rho_A] = 1, \Gamma_{BA} \leq \mathbb{I}_B \otimes \rho_A^T, \rho_A \geq 0 \right\} \tag{149}$$

i.e. we are comparing the Choi operators, and ensuring that they are positive. What's nice about this is the independence to the environment R.

4.5 Optimal Receivers or Classical Information

Consider a CQ Channel $\mathcal{N}_{B|X}$, and the associated 'joint' state

$$\rho_{XB} = \sum_{x \in \mathcal{X}} P_X(x) |x\rangle\langle x|_X \otimes \varphi_B(x)$$

We will be interested here in ways of determining x by measuring system B with $\Lambda_B(x)$. Associated to $\Lambda_B(x)$, we can consider a measurement $\mathcal{M}_{X'|B}$, which would allow the formulation

$$P_{guess}(X|B) = \sum_x P_X(x) Tr[\Lambda_B(x) \varphi_B(x)] = Tr[\Pi_{XX'} \mathcal{M}_{X'|B}[\rho_{XB}]] \tag{150}$$

where $\Pi_{XX'} = \sum_x |xx'\rangle\langle xx'|_{XX'}$ is a measurement of the event $x = x'$. We are essentially hoping to find a measurement which will yield the joint of $XX' =$ the marginal of X.

A bit of rewriting gives us:

$$P_{guess}(X|B) = 1 - \delta(\mathcal{M}_{X'|B}[\rho_{XB}], P_{XX'}) = \max_{\Lambda}\{\sum_x P_X(x) Tr[\Lambda_B(x)\varphi_B(x)]\}$$

Proposition 4.18. *For any CQ state ρ_{XB}, the optimal guessing probability is given by*

$$P_{guess}(X|B) = \max_{\Lambda}\{Tr[\lambda_{XB}\rho_{XB}]\}, \quad Tr_X[\Lambda_{XB}\rho_{XB}] = \mathbb{I}_B, \quad \Lambda_{XB} \geq 0 \tag{151}$$

the dual form is

$$P_{guess}(X|B) = \min_{\sigma_B}\{Tr[\sigma_B]\}, \quad \mathbb{I}_X \otimes \sigma_B \geq \rho_{XB} \tag{152}$$

4.5.1 Pretty Good Measurement

$$\begin{aligned} P^{PGM}_{guess}(X|B)_\tau &= \sum_x P_X(x) Tr[\Lambda_B(x)\varphi_B(x)] \\ &= Tr[(\mathbb{I} \otimes \tau_B^{-1/2})\tau_{XB}(\mathbb{I} \otimes \tau_B^{1/2})\tau_{XB}] \end{aligned} \tag{153}$$

Proposition 4.19. *Moreover, the PGM satisfies*

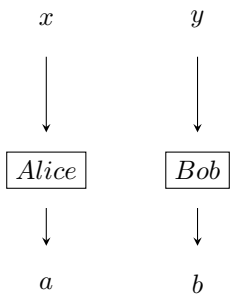
$$P^{PGM}_{guess}(X|B)_\rho \geq P_{guess}(X|B)^2_\rho \tag{154}$$

4.6 CHSH Game & Bell's Theorem

This section will act as a constructive proof that quantum mechanics can offer statistical predictions that are incompatible with classical probabilistic models.[6]

4.6.1 CHSH Game Setup & Classical Strategy

Alice and Bob each receive one bit x and y respectively, and will each output one bit a and b respectively. Alice and Bob are **not allowed to communicate**, and win if $a \oplus b = xy$.



Clearly, there is a higher chance that $xy = 0$, so the best strategy is for Alice and bob to both pick $a = b = 0$. This strategy will allow them to have

$$Pr[A + B = XY] \leq \frac{3}{4} \tag{155}$$

[6] basically Bell's theorem statement

4.6.2 Using Quantum Mechanics to Increase Win Probability

We will start by defining the following variables

$$a'_x = (-1)^{a_x} \quad b'_y = (-1)^{b_y}$$

We can now let these variables be 'assigned' to particular Bloch vectors, i.e. we are now working with qubits rather than bits.

$$a'_x \to \hat{a}_x \cdot \vec{\sigma} \quad b'_y \to \hat{b}_y \cdot \vec{\sigma}$$

Observing the quantities $\langle a'_x b'_y \rangle$, we can define winning conditions: i.e. we win if $\langle a'_x b'_y \rangle = 1$ for all cases except for $x = 1, y = 1$, where $\langle a'_x b'_y \rangle = -1$. Thus,

$$Pr[A + B = XY]_{quantum} = \frac{1}{8}[4 + \langle a'_0 b'_0 \rangle + \langle a'_0 b'_1 \rangle + \langle a'_1 b'_0 \rangle - \langle a'_1 b'_1 \rangle] \tag{156}$$

Exploiting the entangled state $|\Phi_{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$, we know that the average is

$$\langle \Phi_{11}|(\hat{a} \cdot \vec{\sigma})(\hat{b} \cdot \vec{\sigma})|\Phi_{11}\rangle = -\hat{a} \cdot \hat{b} \tag{157}$$

thus yielding

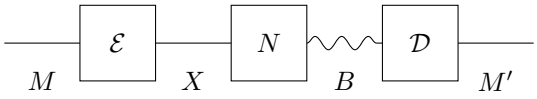
$$Pr[A + B = XY]_{quantum} = \frac{1}{8}[4 - \hat{a}_0 \cdot \hat{b}_0 - \hat{a}_0 \cdot \hat{b}_1 - \hat{a}_1 \cdot \hat{b}_0 + \hat{a}_1 \cdot \hat{b}_1] \tag{158}$$

For a correct choice of those values, we have that

$$Pr[A + B = XY]_{quantum} \approx 85\% \geq \frac{3}{4}$$

4.7 Channel Coding

Definition 4.20. *A (k, ϵ) **protocol** for (classical) communication over noisy channel $\mathcal{N}_{X|B}$ is a pair $(\mathcal{E}, \mathcal{D})$ such that $|M| = k$ and $\delta(\mathcal{D} \circ \mathcal{N} \circ \mathcal{E}, \mathbb{I}) \leq \epsilon$ i.e. it is rather easily distinguishable.*



Here we will derive an upper bound on k for (k, ϵ) coding schemes. The objective is to show that this bound **only depends on the channel** i.e. whatever the encoder and the decoder,

$$k \leq f(\mathcal{N}, \epsilon) \tag{159}$$

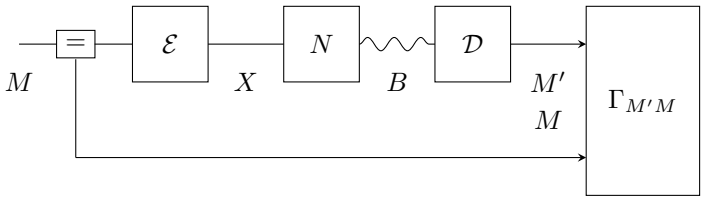
Remark 13. *Notice that if we derive an upper bound for the average case, i.e. not worst error, then this average error $\leq$ worst-case error. Thus, from now on (k, ϵ) will designate the average error protocol.*

Remark 14. *Also note that every (classical) stochastic map is a convex combination of deterministic maps.*

Considering the code (k, ϵ) is rather good (small ϵ), then the input M should resemble M' and it becomes thus 'easy' to distinguish the input-output distribution $P_{M'M}$ from any product distribution $Q_{M'} \otimes P_M$

4.7.1 Converse Bound

Our communication system now looks as follows:



with M uniformly distributed.

Let $\Gamma_{M'M} = \sum |m,m\rangle\langle m,m|_{M'M}$ be a measurement at the end of the system which checks for $P_{M'M}$. **By assumption of a (k,ϵ) code, we have**

$$Tr[\Gamma_{M'M} P_{M'M}] \geq 1 - \epsilon \tag{160}$$

Step 1: Defining the measurement Λ_{BM}

Now let ω_{BM} be the product state of the message and the output of $\mathcal{N}_{B|X}$, we thus have

$$Tr[\Gamma_{M'M} P_{M'M}] = Tr[\Gamma_{M'M} \mathcal{D}_{M'|B}[\omega_{BM}]] = Tr[\Lambda_{BM}\omega_{BM}] \geq 1 - \epsilon \tag{161}$$

for $\Lambda_{BM} = \mathcal{D}^*_{M'|B}[\Gamma_{M'M}]$

Step 2: Bounding the trace

Since Λ_{BM} is feasible for **any** product state τ_{BM}, we can write via Neyman-Pearson[7] that

$$\beta_{1-\epsilon}(\omega_{BM}, \tau_{BM}) \leq Tr[\Lambda_{BM}\tau_{BM}] \tag{162}$$

Step 3: Using uniform prior

Let σ_B be an arbitrary density operator on $\mathcal{H}_B$, then by setting

$$\tau_{BM} = \sigma_B \otimes \pi_M \tag{163}$$

we can now split the trace into

$$\begin{aligned} Tr[\Lambda_{BM}(\sigma_B \otimes \pi_M)] &= Tr[\Gamma_{M'M}(\mathcal{D}^*_{M'|B}[\sigma_B] \otimes \pi_M)] \\ = \frac{1}{|M|} Tr[\sum_m |m,m\rangle\langle m,m|_{M'M}(Q_{M'} \otimes \mathbb{I}_M)] &= \frac{1}{k}\sum_m \langle m|Q_{M'}|m\rangle_{M'} = \frac{1}{k} \end{aligned} \tag{164}$$

[7] $\beta_{1-\epsilon}(\omega_{BM}, \tau_{BM}) = \min_{\Lambda}\{Tr[\Lambda\tau_{BM}] : Tr[\Lambda\omega_{BM}] = 1-\epsilon\}$ is minimal thus less than or equal

Thus, we have that

$$\beta_{1-\epsilon}(\omega_{BM}, \sigma_b \otimes \pi_M) \leq \frac{1}{k} \tag{165}$$

Step 4: Restricting from protocol to code

Since a code is simply 'smaller' than a protocol, we can define

$$\omega_{BX} = \mathcal{N}_{B|X=x} \otimes \frac{1}{k} \sum_m |x_m\rangle\langle x_m| \tag{166}$$

Thus the mapping via the inverse of $\mathcal{E}$ is possible (since codes are injective). Using the DPI, we get

$$\beta_{1-\epsilon}(\omega_{BX}, \omega_X \otimes \sigma_B) \leq \beta_{1-\epsilon}(\omega_{BM}, \pi_M \otimes \sigma_B) \tag{167}$$

Thus

$$\max_{\sigma_B} \{\beta_{1-\epsilon}(\omega_{BX}, \omega_X \otimes \sigma_B)\} \leq \frac{1}{k} \tag{168}$$

However this is still encoder-dependent... By minimizing over P_X we remove then encoder dependence, and thus get our desired bound. □

Remark 15. *For symmetric channels, a uniform P_X achieves these optimal bounds.*

4.7.2 Achievability

This bound is the exact opposite of before. We want to show that for a fixed channel $\mathcal{N}$, there exists a (k, ϵ) code such that

$$k \geq g(\mathcal{N}, \epsilon) \tag{169}$$

By **Shannon's random coding argument**, we can show that there exists a (k, ϵ) code by considering the **average error**. If the average error probability is smaller than ϵ, then there must exist at least one ϵ-good code.

Remark 16. *This is **not encoder design!** This random coding argument only means that the choice of encoder is random, but the encoders are still deterministic.*

Definition 4.21. *A **pretty good measurement (PGM)** is a measurement $\Lambda_B(x)$ defined by an ensemble*

$$\tau_{XB} = \sum_x P_X(x)|x\rangle\langle x|_X \otimes \varphi_B(x) \tag{170}$$

The pretty good measurement is thus

$$\Lambda_B(x) = \tau_B^{-1/2} P_X(x) \varphi_B(x) \tau_B^{1/2}, \quad \tau_B = Tr_X[\tau_{XB}] = \sum_x P_X(x)\varphi_B(x) \tag{171}$$

i.e. the PGM is essentially a measurement onto the support of $\{\varphi_B(x)\}$, which is the output of the noisy channel.[8]

The previous statement is backed by the fact that $\sum_x \Lambda_B(x) = \tau_B^{-1/2} \tau_B \tau_B^{1/2}$

[8] Is very similar to a nearest-neighbour decoding scheme

From this definition, we can now write out the probability of guessing correctly using the PGM:

$$\begin{aligned} P_{guess}^{PGM}(X|B)_\tau &= \sum_x P_X(x) Tr[\Lambda_B(x)\varphi_B(x)] \\ &= Tr[(\mathbb{I}\otimes\tau_B^{-1/2})\tau_{XB}(\mathbb{I}\otimes\tau_B^{1/2})\tau_{XB}] \end{aligned} \tag{172}$$

Proposition 4.22. *For any CQ channel $\mathcal{N}$ and error ϵ, there exists a (k,ϵ) code with*

$$\frac{1}{k} \leq \min_{\eta\in[0,1]} \max_{P_X} \{\frac{1}{\eta}\beta_{1-\epsilon}(\omega_{BX}, \omega_B\otimes\omega_X)\} \tag{173}$$

for $\omega_{BX} = \sum_x \varphi_B(x)\otimes|x\rangle\langle x|_X$

4.8 Coding for i.i.d. Channels

We've already covered the rate of a channel, but now we will cover the capacity.

Definition 4.23. *The **capcaity** of a CQ channel $\mathcal{N}_{B|X}$ is the limit as $\epsilon\to 0$ of the ϵ-capacity.*

Proposition 4.24. *For any CQ channel $\mathcal{N}_{B|X}$ and associated 'joint' state $\omega_{XB} = \sum_x P_X(x)|x\rangle\langle x|_X\otimes\varphi_B(x)$, we have*

$$C(\mathcal{N}_{B|X}) = \max_{P_X}\{I(X;B)_\omega\} \tag{174}$$

The optimization over P_X is convex, so mixtures of optimal distribtions are also optimal

4.9 Entropy & Capacity

We are focusing on classical communication over quantum channels, and have already seen that the converse combined with the PGM yield a tight bound for error probability, given the size of a message space.

Now, we will consider *iid* channels, and more specifically n-folds of them. I.e. if $\mathcal{E}$ is a channel, then $\mathcal{E}^{\otimes n}$ is just like sending n bits over the same channel simultaneously.

Definition 4.25. *The **coding rate** R of a channel is defined as*

$$R = \frac{\log(k)}{n}\left[\frac{bits}{use}\right] \tag{175}$$

This begs the question: **what is the optimal rate?**

Definition 4.26. *For a fixed ϵ, the optimal rate is denoted $C(\mathcal{N},\epsilon)$. Asymptotically it is called the **capacity**, and is defined as*

$$C(\mathcal{N}) = \lim_{\epsilon\to 0} C(\mathcal{N},\epsilon) \tag{176}$$

Using the fact that our channels are now considered iid, and assuming that $/eta = \epsilon/2$, then we have

$$C(\mathcal{N},\epsilon) \geq \lim_{n\to\infty} -\frac{1}{n}\log(\beta_{1-\epsilon/2}(\omega_{BX}^{\otimes n}, (\omega_B\otimes\omega_X)^{\otimes n})) = {}^9 D(\omega_{BX}, \omega_B\otimes\omega_X) \tag{177}$$

where $D(\omega_{BX},\omega_B\otimes\omega_X)$ denotes the **relative entropy**.

$$D(\rho,\sigma) = Tr[\rho(\log\rho - \log\sigma)] \tag{178}$$

We thus have

$$C(\mathcal{N},\epsilon) \geq \max_{P_X} D(\omega_{BX},\omega_B\otimes\omega_X) = \max_{P_X}\mathcal{I}(X;B)_\omega \tag{179}$$

[9] via AEP

Definition 4.27. *The **entropy** in a classical random variable X is defined as*

$$H(X)_P = \sum_X P_X(x) \log\left(\frac{1}{P_X(x)}\right) \tag{180}$$

Definition 4.28. ***Quantum entropy*** *is the entropy in a quantum state ρ over a system A, defined as*

$$H(A)_\rho = -Tr\left[\rho \log(\rho)\right] \tag{181}$$

Notice that

$$H(A)_\rho = -D(\rho_A, \mathbb{I}_A) = \log(|A|) - D(\rho_A, \pi_A) \tag{182}$$

Proposition 4.29 (Klein & Stein's Lemmas)**.**

$$D(\rho, \sigma) \geq 0, \quad = \textit{iff} \quad \rho = \sigma \tag{183}$$

$$\forall \alpha \in [0,1], \lim_{n\to\infty} -\frac{1}{n} \log(\beta_\alpha(\rho^{\otimes n}, \sigma^{\otimes n}) = {}^{10} D(\rho^{\otimes n}, \sigma^{\otimes n}) \tag{184}$$

Proposition 4.30. *The properties of entropy:*

1. $H(A)_\rho \geq 0$
2. $H(A)_{U\rho U`*} = H(A)_\rho$ *for unitary* U
3. $H(A)_\rho \leq \log(|A|)$
4. $H(A)_\rho \geq \sum_x P_X(x) H(A)_{\rho(x)}$ *for* $\rho = \sum_x P_X(x)\rho(x) \implies H(A)_\rho$ *is concave!*
5. $H(A)_\sigma \geq H(A)_\rho$ *for* $\sigma = \sum_x \prod(x)\rho\prod(x)$ *a projective measurement*

The last property implies that the entropy increases if we measure ρ with σ, and then forget the measurement result.

Definition 4.31. *The **joint entropy** of two systems is*

$$H(AB)_\rho = -D(\rho_{AB}, \mathbb{I}_{AB}) \tag{185}$$

This is implies that for ρ_{AB} such that $\rho_A = Tr_B[\rho_{AB}]$ and $\rho_B = Tr_A[\rho_{AB}]$, the following holds:

1. $H(A)_\rho = H(B)_\rho$ for ρ a pure state
2. $H(AB)_\rho \leq H(A)_\rho + H(B)_\rho$, with equality if $\rho_{AB} = \rho_A \otimes \rho_B$
3. $H(AB)_\rho \geq |H(A)_\rho - H(B)_\rho|$

By extending the purification argument, if we have a tripartite state ρ_{ABC}, then by purification it holds that

$$H(B)_\rho = H(AC)_\rho \tag{186}$$

[10] basically the AEP. The avg log error probability tends to relative entropy

4.9.1 Log and Tensor Product

A useful property is

$$\log(\rho \otimes \sigma) = (\log(\rho) \otimes \mathbb{I}) + (\mathbb{I} \otimes \log(\sigma)) \tag{187}$$

thus, the relative entropy of ρ_{AB} and $\rho_A \otimes \rho_B$ is

$$\begin{gathered} D(\rho_{AB}, \rho_A \otimes \rho_B) = Tr[\rho_{AB}(\log(\rho_{AB} - \log(\rho_A \otimes \rho : B))] \\ = -H(AB)_\rho - Tr[\rho_{AB}(\log(\rho_A \otimes \mathbb{I}_B)] - Tr[\rho_{AB}(\mathbb{I}_A \otimes \log(\rho_B))] \\ = -H(A)_\rho - Tr_A[\rho_A \log(\rho_A)] - Tr_B[\rho_B \log(\rho_B)] \end{gathered} \tag{188}$$

4.9.2 Conditional Entropy & Mutual Information

Definition 4.32. *The **conditional entropy** of A given B is defined as*

$$H(A|B)_\rho = -D(\rho_{AB}, \mathbb{I}_A \otimes \rho_B) = \log(|A|) - D(\rho_{AB}, \pi_A \otimes \rho_B) \tag{189}$$

Definition 4.33. *The **mutual information** between systems A and B is defined as*

$$\mathcal{I}(A;B)_\rho = D(\rho_{AB}, \rho_A \otimes \rho_B) \tag{190}$$

Proposition 4.34. *The **chain rules of entropy** are:*

1. $H(A|B)_\rho = H(AB)_\rho - H(B)_\rho$
2. $\mathcal{I}(A;B) = H(A) + H(B) - H(AB) = H(A) - H(A|B)$
3. $H(A|B) = -H(A|C)$ *for* ρ_{ABC} *a pure state*
4. $|H(A|B)| \leq \log(|A|)$
5. $H(X|B) \geq 0$ *for* ρ_{XB} *a CQ state*

4.10 Entropic Uncertainty Relations

We will construct here two uncertainty relations, which will be used to argue that a particular game cannot be won. Here is a model of two versions of the same game

1. Bob prepares a qubit A as he wishes and delivers it to Alice
2. Alice randomly chooses $M = X$ or Z and announces what she got
3. Bob must emit a guess for the result of A when measuring it in M (depending on Alice's result)
4. Alice performs the measurement M
5. Bob wins if his guess was correct

1. Bob prepares a qubit A as he wishes and delivers it to Alice
2. Bob must emit a guess for the result of A when measuring it in M for both cases of $M = X$ or $M = Z$
3. Alice randomly chooses $M = X$ or Z and announces what she got
4. Alice performs the measurement M
5. Bob wins if his guess was correct

Clearly, Bob can always win version one, by preparing a **bipartite state** $|\Phi_{00}\rangle_{AB}$. This is obvious since Bob can just 'replicate' Alice's announcement on B and know what A will produce by entanglement.

In version 2 however, the fact that X and Z are orthogonal principles makes it more difficult. He could prepare a **tripartite state** ρ_{ABC} and store his guesses in B resp. C, but Bob's measurements do not commute. I.e. measuring X and then Z will not yield enough information to make an adequate guess. This leads us to the following entropy uncertainty relations, explaining each version of the game respectively:

Proposition 4.35. *For any tripartite state ρ_{ABC},*

$$\begin{aligned} H(X_A|B)_\rho + H(Z_A|B)_\rho &\geq \log|A| + H(A|B)_\rho \\ H(X_A|B)_\rho + H(Z_A|C)_\rho &\geq \log|A| \end{aligned} \tag{191}$$

where

$$H(Z_A|B)\rho = H(A|B)_{\mathcal{P}_A[\rho_{AB}]}{}^{11} \quad H(X_A|B)\rho = H(A|B)_{\tilde{\mathcal{P}}_A[\rho_{AB}]}{}^{12}$$

*This implies that storing **one** guess in B allows complete certainty if ρ_{AB} is entangled (entropy of $-\log|A|$), but two orthogonal guesses cannot be stored in BC.*

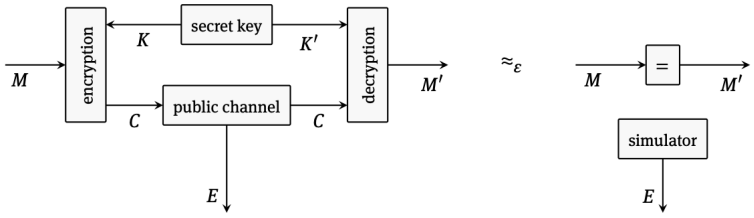
These uncertainty relations allow us to place constraints on incompatible observables. Lastly,

Proposition 4.36. *For a pure ρ_{ABC}, if either ρ_{AB} or ρ_{AC} is CQ with classical part A in the X or Z basis, then the second uncertainty relation is satisfied with equality.*

4.11 Quantum Key Distribution

The idea of quantum key distribution is to privately communicate a key to then perform classical encryption.

To ensure information-theoretic secure communication, we want the secret key channel via a public channel to simulate a identity channel along with a simulator:



mathematically,

$$\delta(\mathcal{N}_{M'E|M}, P_E \otimes I_{M'M}) \leq \epsilon \tag{192}$$

[11] pinches in the $|z\rangle$ basis
[12] pinches in the $|x\rangle$ basis

4.11.1 Problem Setup

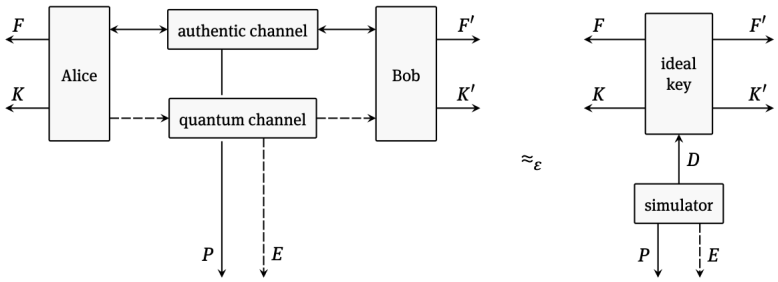
In this section, we will consider the eavesdropper Eve to have access to 2 different attacks:

- impersonation attacks: Eve acts as Alice or Bob
- jamming attacks: Eve floods the quantum channel

The protocol of QKD considers that Alice & Bob have an output K and K', indicating their produced keys, as well as a flag bit each: F and F', which indicate whether their output (K or K') are good or not.

Definition 4.37. *A **QKD protocol** has as output a length-l key such that for any channel $\mathcal{N}_{BE|A}$, there exists a simulator leading to output $\theta_{KK'FPE}$ such that*

$$\delta(\omega_{KK'FPE},]\theta_{KK'FPE}) \leq \epsilon$$



4.11.2 BB84 Protocol

The main idea of this protocol is that Alice and Bob exploit the fact that measuring a qubit will collapse it to the measuring basis to check whether someone has eavesdropped on their communication on an assumed to be **authenticated quantum channel**. The steps can be boiled down to

1. Generate 'raw' keys with qubits
2. Reconcile information with the observed rate of error
3. Privacy amplification to make the output keys

Definition 4.38. *The BB84 protocol is defined by*

- *6 integers (n, n_x, n_z, s, t, l)*
- *2 real numbers: $q \in [0, 1]$ and $\delta \in (0, \frac{1-q}{2})$*

and follows the steps:

1. *Alice randomly generates 2 n-bit strings Y and W, and generates an n-qubit string where W selects the basis and Y selects the state:*

$$Y = 0|W = 0 \sim |0\rangle, \quad Y = 1|W = 0 \sim |1\rangle, \quad Y = 0|W = 1 \sim |+\rangle, \quad Y = 1|W = 1 \sim |-\rangle \tag{193}$$

2. *Bob generates the random bit string W' and measures Alice's qubit string with W' as the same basis selector. He stores his measurement results in Y'*

3. *Alice announces W and Bob compares it to W'. He checks which indices of his matches and if there are less than n_x and n_z for their respective bases the protocol is aborted. If there are enough, he saves the indices and sends them to Alice. They now have a shared sequence, called the **raw keys**.*

4. *They then perform information reconciliation and privacy amplification (skipped here) to construct the **real keys***

The general idea is that Alice send n qubits to Bob over an authenticated channel, which are randomly in the X, Z basis, and random over $(|0\rangle, |1\rangle)$ or $(|+\rangle, |-\rangle)$. Bob will also measure these incoming bits randomly. After that, Alice only reveals the *sequence of bases* she used, and Bob discards those in which he measured in the wrong basis. They use this reduced bitstring as a key for something like OTP.