Quantum Cryptography

Anish Banerjee

Contents

§0. Notation

For any $n \in \mathbb{N} := \{1, 2, \ldots\}$, we define $[n]$ to be the set $\{1, 2, \ldots, n\}$.
For any distribution $\mathcal{D}$, $x \leftarrow \mathcal{D}$ denotes that x is sampled from the distribution $\mathcal{D}$.
Similarly for a set A, $x \xleftarrow{R} A$ denotes that x is a random, uniformly distributed element from A.
$\{A, B\} = AB + BA$ denotes the anti-commutator of operators A, B
$[A, B] = AB - BA$ denotes the commutator of operators A, B
The Pauli matrices are denoted using

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The Hadamard, Phase and T $(\pi/8)$ operators are denoted as

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \qquad T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$$

$\text{Bool}(n, m)$ denotes the set of functions from $\{0,1\}^n$ to $\{0,1\}^m$
The Bell basis is represented as

$$\left|\Psi_{xy}\right\rangle = \frac{|0x\rangle + (-1)^y |1\bar{x}\rangle}{\sqrt{2}}$$

and $\left|\phi^+\right\rangle = |\Psi_{00}\rangle$

Acknowledgements

These notes are primarily based on the edX course[VW16] on Quantum Cryptography. For some concepts, [NC10] has also been referred to. I would like to thank Arpon Basu for providing me with this LaTeX-template and teaching me how to make notes in it.

§1. Preliminaries

1.1 Some points about density matrices and partial trace

Definition 1.1. If a source prepares quantum states in a probabilistic manner, i.e. it prepares the quantum state ρ_X is p_x, then the net density matrix is written as $\rho = \sum_i p_i \rho_i$ and $\mathcal{E} = \{p_i, \rho_i\}$ is called an **ensemble** of states.

- Observe the difference between the mixed state $\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$ and the $|+\rangle$ state. This is the difference between a random toss of a coin and a superposition. We can figure out the difference by measuring them in the $\{|+\rangle, |-\rangle\}$ basis. In the first case, we have a probability $\langle +|\rho|+\rangle = 1/2$ of getting $|+\rangle$ while in the second case, the probability is 1.
- Time evolution of the density operator is given by $\rho_{t'} = U(t,t')\rho_t U(t,t')^\dagger$
- If a measurement is made on the state ρ using the measurement operators $\{M_m\}$ (equivalently using POVM $\{E_m\}$) then the probability of obtaining outcome m is given by

  $$p(m) = \mathrm{tr}\Big(M_m^\dagger M_m \rho\Big) = \mathrm{tr}\big(E_m \rho\big)$$

  and the post measurement state is

  $$\rho_m = \frac{M_m \rho M_m^\dagger}{p(m)}$$

- The density matrix of a composite of systems A and B can be written as

  $$\rho_{AB} = \sum_{i,j,k,l} c_{ij}^{kl} |i\rangle\langle j| \otimes |k\rangle\langle l|$$

  where i, j are basis elements of system A and k, l are basis elements of system B.
- In general, for a composite system, we cannot always decompose $\rho_{AB} = \rho_A \otimes \rho_B$ by using the partial trace.
- The partial trace can be computed as:

  $$\mathrm{tr}_B(\rho_{AB}) = \sum_{i,j,k,l} c_{ij}^{kl} |i\rangle\langle j| \otimes \mathrm{tr}(|k\rangle\langle l|) = \sum_{i,j,k,l} c_{ij}^{kl} |i\rangle\langle j| \delta_{kl} = \sum_{i,j,k} c_{ij}^{kk} |i\rangle\langle j|$$

- Partial trace wrt B can also be seen as performing the measurement on the subsystem B and forgetting the result. Consider a composite system ρ_{AB}. The probability that we obtain the outcome b when we measure system B in the basis $\{|b\rangle\}$ is:

  $$p_b = \mathrm{tr}\big((\mathbb{I} \otimes |b\rangle\langle b|)\rho_{AB}\big)$$

  and the post-measurement state is

  $$\rho_{AB}^b = \frac{(\mathbb{I} \otimes |b\rangle\langle b|)\rho_{AB}(\mathbb{I} \otimes |b\rangle\langle b|)}{p_b}$$

Taking the partial trace with respect to B:

$$\begin{aligned}\rho_A^b &= \mathrm{tr}_B(\rho_{AB}^b) \\ &= \frac{1}{p_b}\sum_{i,j,k,l} c_{ij}^{kl}\,|i\rangle\langle j|\,\mathrm{tr}(|b\rangle\langle b|k\rangle\langle l|b\rangle\langle b|) \\ &= \frac{1}{p_b}\sum_{i,j,k,l} c_{ij}^{kl}\,|i\rangle\langle j|\,\delta_{bk}\delta_{bl} \\ &= \frac{1}{p_b}\sum_{i,j} c_{ij}^{bb}\,|i\rangle\langle j|\end{aligned}$$

Now we can write the partial trace wrt B on system AB as

$$\mathrm{tr}_B(\rho_{AB}) = \sum_b p_b \rho_A^b$$

It can be easily expanded and seen to be same as the expression obtained above. Finally, the second method of computing partial trace is:

$$\rho_A = \sum_{i,j,b} c_{i,j}^{k,l}\,|i\rangle\langle j|\,\langle b|k\rangle\langle l|b\rangle$$

(since $\mathrm{tr}(|b\rangle\langle b|k\rangle\langle l|b\rangle\langle b|) = \langle b|k\rangle\langle l|b\rangle$)

$$= \sum_b (\mathbb{I}\otimes\langle b|)\rho_{AB}(\mathbb{I}\otimes|b\rangle)$$

- Partial trace is the unique operation which gives rise to the correct description of observable quantities for subsystems of a composite system.

  Let f be any mapping from density matrices for the composite system AB to the system A. We must have

  $$\mathrm{tr}\big(Mf(\rho_{AB})\big) = \mathrm{tr}\big((M\otimes I)\rho_{AB}\big)$$

  Let M_i be an orthonormal basis for the operators on the space of system A with respect to the Hilbert-Schmidt Inner product: $(X,Y) = \mathrm{tr}\left(X^\dagger Y\right)$. Then we can write:

  $$f(\rho_{AB}) = \sum_i M_i\,\mathrm{tr}\left(M_i^\dagger f(\rho_{AB})\right) = \sum_i M_i\,\mathrm{tr}\big((M_i\otimes I)\rho_{AB}\big)$$

  This is a unique operator, since the RHS is independent of f. Moreover, partial trace satisfies this property. Hence, partial trace is the required unique operator.

1.2 General Measurements

Definition 1.2 (Positive Operator Valued Measure)**.** Let $\{E_m\}$ be a set of positive semidefinite operators such that $\sum_m E_m = \mathbb{I}$. Then the set is called a POVM.

The probability of getting the outcome m is $p_m = \mathrm{tr}\big(E_m\rho\big)$
In the case of POVM's, it turns out that the information given by the operators $\{E_m\}$ is not sufficient fully to determine the post-measurement state. This is because the measurement may not fully collapse the state (the post measurement state may not be pure), **and as a consequence there remains the flexibility to apply an arbitary unitary in the post-measurement state, without affecting the outcome probabilities**.

Definition 1.3 (Kraus Operators)**.** Let $E = \{E_m\}$ be a given POVM on $\mathbb{C}^d$. A Kraus operator representation of E is a set of linear operators $M_m \in \mathcal{L}(\mathbb{C}^d, \mathbb{C}^d)$ such that $E_m = M_m^\dagger M_m$ for all m.

Remark. The Kraus decomposition *always exists* ($M_m = \sqrt{E_m}$) and is *not unique* (For any unitary U_m, $M'_m = U_m\sqrt{E_m}$)
We can write the post measurement state in terms of Kraus operators as:

$$\rho_m = \frac{M_m \rho M_m^\dagger}{\mathrm{tr}\left(M_m^\dagger M_m \rho\right)}$$

Remark. Consider the EPR state

$$|EPR\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

If we measure it using the projector $\Pi = |00\rangle\langle 00| + |11\rangle\langle 11|$ then the post measurement state is $\rho = |EPR\rangle\langle EPR|$. However, if we measure using the standard basis $\{|00\rangle, |11\rangle\}$ then the post-measurement state is

$$\rho' = \frac{|00\rangle\langle 00| + |11\rangle\langle 11|}{2}$$

This is one of the main advantages of using generalized measurements as opposed to basis measurements: they allow to compute certain simple quantities on multi-qubit states (such as the parity) without fully "destroying" the state, as happens when measuring in a basis.

§2. Quantum Tools and First Protocol

2.1 Classical One Time Pad

The encription and decription functions are:

$$Enc(k, m) = k \oplus m = c$$

$$Dec(k, c) = k \oplus c = k \oplus (k \oplus m) = m$$

Hence the one-time pad is correct.

Definition 2.1 (Shannon Secrecy)**.**

$$\forall m, e \; Pr[M = m | E = e] = Pr[M = m]$$

It can also be written as

$$Pr[E = e | M = m] = Pr[E = e]$$

because,

$$Pr[M = m, E = e] = \overbrace{Pr[M = m | E = e] \underbrace{Pr[E = e] = Pr[E = e | M = m]}_{equal} Pr[M = m]}^{equal}$$

The one-time pad is perfectly secure. For one bit, the probability that it gets encoded to 1 is $1/2$ and that it gets encoded to 0 is $1/2$ (since the key is randomly generated).
Disadvantages of One-Time Pad

- Can be used only once with perfect security
- Need a key as long as the message

2.2 Quantum One-Time Pad

Let $m \in \{0, 1\}$. The encryption and decryption functions are:

$$|e\rangle = X^k |m\rangle$$

$$|m\rangle = X^k |e\rangle$$

Writing it in the form of density matrices, we obtain:

$$\rho = \frac{1}{2} |m\rangle\langle m| + \frac{1}{2} X |m\rangle\langle m| X$$

which, irrespective of m, will always be the completely mixed state $\rho = \frac{\mathbb{I}_2}{2}$ So, any Eve, without the access to the keys, cannot decipher the state.
For doing the same in the $\{|+\rangle, |-\rangle\}$ basis, we can use the Z gate, since it acts as an X gate in the Hadamard basis. However Z cannot encrypt for the standard basis.
So, we must use two keys k_1, k_2 for encrypting our bits.

$$|e\rangle = X^{k_1} Z^{k_2} |m\rangle$$

Correctness: Bob can decript this using the inverse of $X^{k_1} Z^{k_2}$.
Perfect Security: We have the density matrix that Eve sees as

$$\rho = \frac{1}{4} [|m\rangle\langle m| + X |m\rangle\langle m| X + Z |m\rangle\langle m| Z + XZ |m\rangle\langle m| ZX]$$

Taking $|m\rangle$ as $\alpha |0\rangle + \beta |1\rangle$, we see that the expression reduces to the completely mixed state $\frac{\mathbb{I}_2}{2}$.

Remark. In this scheme, observe the difference between the density matrices seen by Eve and Bob:

$$\rho_{\text{EVE}} = \frac{\mathbb{I}}{2} \quad \text{(a fully mixed state)}$$

$$\rho_{\text{BOB}} = X^{k_1} Z^{k_2} \left|\psi\right\rangle \left\langle\psi\right| Z^{k_2} X^{k_1} \quad \text{(a pure state)}$$

Note that this is created due to the *knowledge difference* of Eve and Bob- Eve doesn't know the key bits whereas Bob does.

Remark. This can also be done by taking two qubits:

$$U_E \left|k\right\rangle \left|m\right\rangle = \left|k\right\rangle \left|m \oplus k\right\rangle$$

Send the second qubit to Bob. He can decript it using $U_E^\dagger$ and the key.

2.3 Classical-Quantum States

This will be extremely useful for cryptography where we have a classical system holding a key and a quantum attacker who has a qunatum system. Together a classical-quantum state.

Definition 2.2 (Classical State)**.** (An isolated) System X is in a **classical state** ρ_X when it is diagonal in the standard basis.

Note: Classical systems are generally labelled X,Y,Z whereas quantum systems are labelled A,B,C.

Definition 2.3 (Classical Quantum System)**.** A system of the form:

$$\rho_{XA} = \sum_x p_x \left|x\right\rangle \left\langle x\right|^X \otimes \rho_x^A$$

is said to be in a cq-state.

§3. Power of Entanglement

We say that a pure state is entangled if it cannot be written as the tensor product of two other states.
Similarly, for mixed states, ρ_{AB} is separable if $\exists\{p_i, \rho_i^A, \rho_i^B\}$ such that

$$\rho_{AB} = \sum_i p_i \rho_i^A \otimes \rho_i^B$$

3.1 Maximally Entangled States

$|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$

$$|\psi\rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle |i\rangle$$

is the maximally entangled state. It has some useful properties: (A^T denotes the transpose of A wrt the computational basis and $\bar{A}$ denotes the element-wise conjugate of A wrt the standard basis)

- $(A \otimes \mathbb{I}) |\psi\rangle = (\mathbb{I} \otimes B) |\psi\rangle \implies A = B^T$

  Proof.

$$\begin{aligned}
(A \otimes \mathbb{I}) |\psi\rangle &= (\mathbb{I} \otimes B) |\psi\rangle \\
\implies \sum_i (A |i\rangle) |i\rangle &= \sum_i |i\rangle B |i\rangle \\
\implies \langle k| \langle l| \sum_i (A |i\rangle) |i\rangle &= \langle k| \langle l| \sum_i |i\rangle B |i\rangle \\
\implies \langle k| A |l\rangle &= \langle l| B |k\rangle \\
\implies A_{kl} &= B_{lk}
\end{aligned}$$

■

- $\langle\psi| (A \otimes \mathbb{I}) |\psi\rangle = \frac{1}{d} \operatorname{Tr}(A)$
- $\langle\psi| (\bar{A} \otimes B) |\psi\rangle = \frac{1}{d} \operatorname{Tr}\left(A^\dagger B\right)$

The last two can be easily verified by expanding.

3.2 Purifications

Every mixed state arises as a reduced density matrix of a bigger pure state.
We can write any density matrix as, using the Spectral Decomposition:

$$\rho_A = \sum_{i=1}^{d_A} p_i |\psi_i\rangle \langle\psi_i|$$

Consider the pure state:

$$|\psi\rangle_{AB} = \sum_{i=1}^{d_A} \sqrt{p_i} |\psi_i\rangle_A \otimes |i\rangle_B$$

it has the density matrix:

$$\rho_{AB} = \sum_{i=1}^{d_A} p_i |\psi_i\rangle_A \langle\psi_i|_A \otimes |i\rangle_B \langle i|_B$$

The partial trace of this system wrt B is ρ_A. Here ρ_{AB} is a purification of ρ_A. Formally,

Definition 3.1 (Purifications)**.** A pure state $|\psi\rangle_{AB}$ is a **purification** of $\rho_A = \sum_i p_i |\psi_i\rangle \langle\psi_i|$ if for some system B with basis $\{|b_i\rangle\}$ and dimension $d_B = d_A$

$$|\psi\rangle_{AB} = \sum_{i=1}^{d_A} \sqrt{p_i} |\psi_i\rangle |b_i\rangle$$

From which it follows that

$$\mathrm{tr}_B(|\psi_{AB}\rangle \langle\psi_{AB}|) = \rho_A$$

Before moving further, recalling a few tools from linear algebra will be helpful. Some very basic but important points are listed below:

- A good way to understand the multiplication of matrices AX is as follows:

$$AX = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$$

  Now if take the output vectors for all the matrices X, then we obtain the column space of A, $C(A)$. In this case, it is a plane, but it can also be a line or $\mathbb{R}^3$.

- We can factorize $A = CR$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

  If it were just a rank 1 matrix, we could have obtained R, C as row and column vectors. Through this decomposition, we obtain the bases for the row and column spaces. Observe that R is also the RRE form of the matrix.

- Multiplication of two matrices A, B can also be seen as multiplication of columns of A with rows of B (Outer products)

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 5 & 1 & 3 \\ 4 & 1 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} \begin{bmatrix} 5 & 1 & 3 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} \begin{bmatrix} 4 & 1 & 6 \end{bmatrix} + \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} \begin{bmatrix} 7 & 8 & 9 \end{bmatrix}$$

- Rank of a matrix is the dimension of the column space, or the number of independent columns.
- The columns of an Unitary matrix form an orthonormal basis for the vector space

3.3 Polar and Singular Value Decompositions

Polar and singular value decompositions are useful ways of breaking linear operators into products of unitary and positive operators, which have more structure.

Theorem 3.1 (Polar Decomposition)**.** Let A be a linear operator on a vector space V. Then there exists unitary U and positive operators J and K such that

$$\begin{aligned} A &= UJ && \text{[Left Polar Decomposition]} \\ &= KU && \text{[Right Polar Decomposition]} \end{aligned}$$

where the unique positive operators J and K are defined by $J := \sqrt{A^\dagger A}, K := \sqrt{AA^\dagger}$. Moreover, if A is invertible then U is unique.

Proof. Since $J = \sqrt{A^\dagger A}$ is a positive operator, we can write its spectral decomposition $J = \sum_i \lambda_i |i\rangle\langle i|, \lambda_i \geq 0$. Define $|\psi_i\rangle := A|i\rangle$. Then $\langle\psi_i|\psi_i\rangle = \lambda_i^2$. For those i where $\lambda_i > 0$, define $|e_i\rangle = |\psi_i\rangle / \lambda_i$. Verify that $|e_i\rangle$ are orthonormal. Extend this orthonormal set to an orthonormal basis of V.
Define $U = \sum_i |e_i\rangle\langle i|$ (basis change). When $\lambda_i \neq 0$, we have $UJ|i\rangle = \lambda_i U|i\rangle = \lambda_i|e_i\rangle = A|i\rangle$ and when $\lambda_i = 0$ $UJ|i\rangle = 0 = A|i\rangle$. As the action of A and UJ on a basis agree, we have $A = UJ$.
Observing that $A = UJ = UJU^\dagger U = KJ$ where $K = UJU^\dagger$ gives the right polar decomposition. $AA^\dagger = K^2$ so that $K = \sqrt{AA^\dagger}$.
If A is invertible, then since $J = U^\dagger A$, $A^{-1}U$ is the inverse of J, so that $U = AJ^{-1}$ is unique.

■

The singular value decomposition (SVD) is a factorization of a real or complex matrix. It generalizes the eigendecomposition of a square normal matrix with an orthonormal eigenbasis to any $m \times n$ matrix.[Sin23].
It decomposes any matrix into three simple transformations:

1. a rotation $V^\dagger$
2. a scaling Σ along the rotated coordinate axes
3. a second rotation U

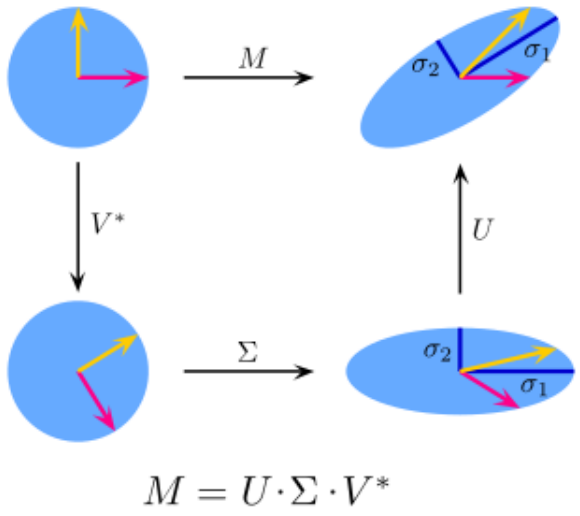


Figure 1: Singular Value Decomposition

Theorem 3.2 (Singular Value Decomposition)**.** Let M be a $m \times n$ matrix. Then it can be decomposed as:

$$M = U\Sigma V^\dagger$$

Where U and V are unitary matrices and Σ is a diagonal matrix with non-negative entries. In other words, any linear operator can be decomposed into

$$M = \sum_k |u_k\rangle\langle v_k|$$

Some points to note:

- The diangonal values of Σ are uniquely determined and are called the **singular values** of M.
- The number of non-zero singular values is the **rank** of the matrix.
- The SVD is not unique. It is always possible to choose the decomposition so that the singular values Σ_{ii} are in descending order. In this case, Σ (but not U and V) is uniquely determined by M.
- Since U and V are unitary, the columns of each of them form a set of orthonormal vectors, which can be regarded as basis vectors.

3.4 The Schmidt Decomposition

Schmidt Decomposition is a convenient 'normal form' for bipartite systems. It helps us to immediately determine if a bipartite system is entangled or not.

Theorem 3.3 (Schmidt Decomposition). Any pure bipartite state $|\psi\rangle_{AB}$ can be written in the form

$$|\psi\rangle_{AB} = \sum_{i=1}^{\min(d_A,d_B)} \lambda_i |i_A\rangle \otimes |i_B\rangle$$

for some bases $\{|i_A\rangle\}$ of system A and $\{|i_B\rangle\}$ of system B (Schmidt Vectors), and **non-negative** real numbers λ_i (Schmidt Coefficients) with $\sum_i \lambda_i^2 = 1$.

Proof. Any state $|\psi\rangle_{AB}$ can be written as

$$|\psi\rangle_{AB} = \sum_{i,j} c_{ij} |i\rangle_A \otimes |j\rangle_B$$

using the standard basis of systems A, B. The matrix $C = (c_{ij})$ can be decomposed using the SVD into

$$C = \sum_k \alpha_k |u_k\rangle \langle v_k|$$

$$c_{ij} = \langle i| C |j\rangle = \sum_k \alpha_k \langle i|u_k\rangle \langle v_k|j\rangle$$

Now,

$$\begin{aligned}
|\psi\rangle_{AB} &= \sum_{i,j} \langle i| C |j\rangle |i\rangle_A \otimes |j\rangle_B \\
&= \sum_{i,j,k} \alpha_k \langle i|u_k\rangle \langle v_k|j\rangle |i\rangle_A \otimes |j\rangle_B \\
&= \sum_k \alpha_k \left(\sum_i \langle i|u_k\rangle |i\rangle \right) \otimes \left(\sum_j \langle v_k|j\rangle |j\rangle \right) \\
&= \sum_k \alpha_k \left(\sum_i u_{ki} |i\rangle \right) \otimes \left(\sum_j v_{kj}^* |j\rangle \right) \\
&= \sum_k \alpha_k |u_k\rangle \otimes |v_k^*\rangle
\end{aligned}$$

Where $|v_k^*\rangle$ is the complex conjugate of $|v_k\rangle$, i.e., having the coefficients of the basis elements as the complex conjugate of the corresponding ones on $|v_k\rangle$

Currently, the α_k are complex numbers but we can conider them as

$$\alpha_k = |\alpha_k| e^{i\theta_k}$$

And take the phase factor inside the basis elements. If the $\{|u_k\rangle\}$ were a basis, then $\{e^{i\theta_k}|u_k\rangle\}$ are also a basis. The sum $\sum_i \lambda_i^2 = 1$ comes from the normalization condition. ■

Definition 3.2 (Schmidt rank). $\text{SR} = |\{i : \lambda_i \neq 0\}|$

Schmidt Coefficients characterize entanglement:

- $\text{SR} = 1 \Leftrightarrow$ Product State $\Leftrightarrow$ The subsystems ρ_A, ρ_B are pure
- $\text{SR} \neq 1 \implies$ Entangled state
- Higher the schmidt rank, more the entanglement

Local and Global data: The Schmidt decompostion divides the degrees of freedom into local and global categories:

- The sets $\{|u_i\rangle_A\}, \{|u_i\rangle_B\}$ are local data because they can be changed by applying unitaries locally on systems A and B respectively.
- The Schmidt Coefficients are global data because they cannot be changed locally.

Reduced Densities have same Eigenvalues
Computing the reduced density matrices:

$$\rho_A = \sum_k |\lambda_k|^2 |u_k\rangle \langle u_k|$$
$$\rho_B = \sum_k |\lambda_k|^2 |v_k\rangle \langle v_k|$$

We observe that both the density matrices have the same eigenvalues.
Uniqueness
The Schmit decompostion is not unique in general:

$$|\text{EPR}\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle = \frac{1}{\sqrt{2}}|{+}{+}\rangle + \frac{1}{\sqrt{2}}|{-}{-}\rangle$$

However, if the Schmidt Coefficients are unique, then the Schmidt Decomposition is also unique, upto a constant phase factor.
This provides us with an algorithm to calculate the Schmidt Decompostion of a state:

1. Calculate the reduced states
2. Perform spectral decomposition on the states. This gives us the schmidt coefficients and vectors.
3. Write the Schmidt Decomposition

Let $\text{Sch}(|\psi\rangle)$ denote the multiset of non-zero Schmidt coefficients of $|\psi\rangle$. Then we have the following lemma:

Lemma 3.4. Let $|\psi\rangle, |\phi\rangle, |\eta\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ with $|\psi\rangle = |\phi\rangle + |\eta\rangle$. Let

$$\rho_A = \text{Tr}_B\left[|\psi\rangle\langle\psi|\right] \quad \sigma_A = \text{Tr}_B\left[|\phi\rangle\langle\phi|\right] \quad \tau_A = \text{Tr}_B\left[|\eta\rangle\langle\eta|\right]$$
$$\rho_B = \text{Tr}_A\left[|\psi\rangle\langle\psi|\right] \quad \sigma_B = \text{Tr}_A\left[|\phi\rangle\langle\phi|\right] \quad \tau_B = \text{Tr}_A\left[|\eta\rangle\langle\eta|\right]$$

Suppose that $|\phi\rangle$ and $|\eta\rangle$ are "orthogonal on both subsystems" in the sense that $\sigma_A\tau_A = 0 = \sigma_B\tau_B$. Then

$$\text{Sch}(|\psi\rangle) = \text{Sch}(|\phi\rangle) \sqcup \text{Sch}(|\eta\rangle)$$

Proof. Observe that $\sigma_A\tau_A = 0 \implies \tau_A\sigma_A = 0$ by taking the complex conjugate. This implies $[\sigma_A, \tau_A] = 0$, so that they are diagonal in the same basis. Similarly $[\sigma_B, \tau_B] = 0$. So, we can write

$$|\phi\rangle = \sum_i \lambda_i |i\rangle_A |i\rangle_B \qquad |\eta\rangle = \sum_i \mu_i |i\rangle_A |i\rangle_B$$

Then $\sigma_A = \sum_i \lambda_i |i\rangle\langle i|_A$, $\tau_A = \sum_i \mu_i |i\rangle\langle i|_A$. Taking the product we observe that $\lambda_i\mu_i = 0$. Thus

$$|\psi\rangle = \sum_i (\lambda_i + \mu_i) |i\rangle_A |i\rangle_B$$

Where in each term at most one of λ_i or μ_i is non-zero. The result follows from this. ■

States with full Schmidt Rank

Let $|\psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$ be a state with Schmidt decomposition

$$|\psi\rangle = \sum_{i=1}^{d} \lambda_i |i\rangle_A |i\rangle_B$$

where $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_d$. $|\psi\rangle$ is said to have full Schmidt rank if $\lambda_d > 0$.

Lemma 3.5. Let $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ be a state with full Schmidt rank and for $A, B \in \mathbb{C}^d \otimes \mathbb{C}^d$

$$A \otimes \mathbb{I} |\psi\rangle = \mathbb{I} \otimes B |\psi\rangle$$

then $A = B^T$

Proof.

$$A \otimes \mathbb{I} |\psi\rangle = \mathbb{I} \otimes B |\psi\rangle \implies \sum_k \lambda_k (A |k\rangle) |k\rangle = \sum_k \lambda_k |k\rangle B |k\rangle$$

The Schmidt decompotition has a property that **the space associated with each Schmidt coefficient is unique**. Let $\lambda_1 = \lambda_2 = \cdots \lambda_r$. Then

$$\mathcal{S} = \text{Span}(|1\rangle \cdots |r\rangle) = \text{Span}(A |1\rangle \cdots A |r\rangle) = \text{Span}(B |1\rangle \cdots B |r\rangle)$$

So, we can block decompose A and B as

$$A = \begin{pmatrix} A^{(1)} & 0 \\ 0 & A' \end{pmatrix} \qquad B = \begin{pmatrix} B^{(1)} & 0 \\ 0 & B' \end{pmatrix}$$

Restricted to $\mathcal{S}$, we have $(A^{(1)} \otimes \mathbb{I}) |\psi\rangle = (\mathbb{I} \otimes B^{(1)}) |\psi\rangle$ where $|\psi\rangle$ is the maximally entangled state. Using property of completely entangled states Section 3.1, we have $A = B^T$. Proceeding in the same way, we have

$$A = \begin{pmatrix} A^{(1)} & 0 & 0 & \cdots \\ 0 & A^{(2)} & 0 & \cdots \\ 0 & 0 & A^{(3)} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \qquad B = \begin{pmatrix} B^{(1)} & 0 & 0 & \cdots \\ 0 & B^{(2)} & 0 & \cdots \\ 0 & 0 & B^{(3)} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

where each $A^{(i)} = \left(B^{(i)}\right)^T$ ■

Corollary 3.6. Let $|\psi\rangle$ be a state with full Schmidt rank. If $\langle\psi|(A\otimes B)|\psi\rangle = 1$ then $A = B^T$

Proof. $(A\otimes\mathbb{I})|\psi\rangle$ and $(\mathbb{I}\otimes B)|\psi\rangle$ have inner product 1, so $A = B^T$. ■

3.5 Uhlmann's Theorem

Uhlmann's theorem is a fundamental theorem in Quantum Information Theory. It answers the question - when is it the case that two pure states can be mapped from one to the other by acting only on parts of the state?
While choosing the purification of a state, we have a unitary freedom of choosing the basis of the system B. Consider the state

$$\rho = \frac{1}{3}|0\rangle\langle 0| + \frac{2}{3}|0\rangle\langle 1|$$

This has two purifications:

$$|\psi\rangle_{AB} = \sqrt{\frac{1}{3}}|0\rangle_A|0\rangle_B + \sqrt{\frac{2}{3}}|1\rangle_A|1\rangle_B$$

$$|\psi'\rangle_{AB} = \sqrt{\frac{1}{3}}|0\rangle_A|+\rangle_B + \sqrt{\frac{2}{3}}|1\rangle_A|-\rangle_B$$

There is an unitary operator which transforms from the $\{|0\rangle, |1\rangle\}$ basis to the $\{|+\rangle, |-\rangle\}$ (namely, the Hadamard Transform)

Theorem 3.7 (Uhlmann's Theorem)**.** Any two purifications $|\psi\rangle_{AB}, |\phi\rangle_{AB}$ of the same ρ_A are related by a unitary transform on B

Proof. $|\psi\rangle_{AB}, |\phi\rangle_{AB}$ can be written using the Schmidt Decomposition as:

$$|\psi\rangle_{AB} = \sum_k \lambda_k |u_k\rangle|v_k\rangle$$

$$|\psi\rangle_{AB} = \sum_k \mu_k |w_k\rangle|z_k\rangle$$

Observe that $\lambda_k = \mu_k$ because the reduced density matrices of system A will have the same eigenvalues. Moreover we can take $|u_k\rangle = |w_k\rangle$ as they differ only by a phase factor, which can be pushed to $|v_k\rangle$ and $|z_k\rangle$. Now since $|v_k\rangle$ and $|z_k\rangle$ form a basis of system B, so there must exist a unitary transform on B U_B which maps $|v_k\rangle \mapsto |z_k\rangle$ ■

Remark. Uhlmann's Theorem is not valid for *mixed* states. For example consider $\rho_{AB} = \frac{1}{2}(|0\rangle\langle 0|\otimes|0\rangle\langle 0| + |1\rangle\langle 1|\otimes|1\rangle\langle 1|)$ and the state $\sigma_{AB} = |EPR\rangle\langle EPR|$. Both have the same reduced state $\rho_A = \mathbb{I}_2/2$ but no unitary transform on system B can transform them into each other.

3.6 Using Entanglement to share a secret

3.6.1 Classical secret

Suppose Alice and Bob want to share a secret $s \in \{0,1\}$. Both Alice and Bob individually have some information a and b respectively about the secret, but they cannot discover anything independently about s using a, b. However, if they come together, then they can share their information to discover the secret. How can we design a protocol for this?
A classical solution is using the OTP. Take $a \xleftarrow{R} \{0,1\}$ be a random bit. Hence $b = a \oplus s$ will also be random. But they can discover $s = a \oplus b$.

Definition 3.3 (Bell Basis)**.** The basis defined by the orthonormal states:

$$\left\{ |\phi_{00}\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}, |\phi_{01}\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}, |\phi_{10}\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}, |\phi_{11}\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}, \right\}$$

or succintly for $a, b \in \{0, 1\}$

$$|\phi_{ab}\rangle = \frac{|ba\rangle + (-1)^a |ab\rangle}{\sqrt{2}},$$

1. The Bell basis is a basis for $\mathbb{C}^2$
2. The Bell basis is locally indistinguishable. All reduced states are the completely mized state $\frac{\mathbb{I}}{2}$
3. By Uhlmann's Theorem, the basis states are interconvertible using local unitaries

Now for sharing a 2 qubit secret, Alice and Bob can share a Bell pair. When they come together, they can measure in the Bell basis to obtain the secret ab.

3.6.2 Quantum Secret

Now suppose that the secret is a quantum state $|s\rangle \in \mathbb{C}^3$ (We need at least 3 parties for this scheme). Now define the states

$$|0\rangle_S \mapsto \frac{1}{\sqrt{3}}|000\rangle + |111\rangle + |222\rangle$$

$$|1\rangle_S \mapsto \frac{1}{\sqrt{3}}|012\rangle + |120\rangle + |201\rangle$$

$$|2\rangle_S \mapsto \frac{1}{\sqrt{3}}|021\rangle + |210\rangle + |102\rangle$$

It can be verified that $\rho_A = \rho_B = \rho_C = \mathbb{I}/3$ and that the states are orthonormal. But observe that if just Alice and Bob come together and measure their qubits obtaining outcomes a, b, then they can immediately disover the secret $s = (b - a)(mod 3)$, without any help from Charlie.
This is an example of a error correction code which protects against the loss of a qubit. We have the following theorem:

Theorem 3.8. For any n, k such that $n < 2k$, we can code k qubits into n with recovery possible from any k.

This also shows why we cannot code 1 qubit into 2 qubits, as then $n = 2k = 2$

3.7 Monogamy of Entanglement

Suppose Alice and Bob share an EPR state. Is it possible to create a joint state in which the strong correlations of the EPR pair are shared simultaneously between all three systems?
It turns out that the answer is no. Let $|\Psi\rangle_{ABC}$ be a state shared by A, B, C such that A and B share a maximally entangled pair. We will show that

$$|\Psi\rangle_{ABC} = \frac{|00\rangle + |11\rangle}{\sqrt{2}} \otimes |\phi\rangle$$

Sharing	Alice	Bob	Eve	Remark
Bit	0	0	0	Yes, if Alice and Bob share a bit then Eve can also share the same bit
Distribution	p=1/2 0 p=1/2 1	p=1/2 1 p=1/2 0	p=1/2 1 p=1/2 0	Yes, if Alice and Bob share a correlated distribution, then Eve can also share the same correlation as Bob. We can replace Bob and Eve and still have the same distribution
Qubit	$\|0\rangle$	$\|0\rangle$	$\|0\rangle$	Yes, Alice, Bob and Eve can share the same qubit
Entanglement	No! Entanglement is monogamous			

Table 1: What all can be shared?

On measuring in the standard basis, A and B measure $|00\rangle$ or $|11\rangle$ each with probability $\frac{1}{2}$. So,

$$|\Psi\rangle_{ABC} = \frac{1}{\sqrt{2}} |00\rangle (\alpha_0 |0\rangle + \alpha_1 |1\rangle) + \frac{1}{\sqrt{2}} |11\rangle (\beta_0 |0\rangle + \beta_1 |1\rangle)$$

In the Hadamard basis

$$|00\rangle = \frac{1}{2}(|+\rangle + |-\rangle)(|+\rangle + |-\rangle) = \frac{1}{2}(|++\rangle + |+-\rangle + |-+\rangle + |--\rangle)$$

$$|11\rangle = \frac{1}{2}(|+\rangle - |-\rangle)(|+\rangle - |-\rangle) = \frac{1}{2}(|++\rangle - |+-\rangle - |-+\rangle + |--\rangle)$$

so that

$$|\Psi\rangle_{ABC} = \frac{1}{2\sqrt{2}}(|++\rangle + |--\rangle)((\alpha_0 + \beta_0) |0\rangle + (\alpha_1 + \beta_1) |1\rangle) + \frac{1}{2\sqrt{2}}(|+-\rangle + |-+\rangle)((\alpha_0 - \beta_0) |0\rangle + (\alpha_1 - \beta_1) |1\rangle)$$

Since on measuring in the Hadamard basis too, A and B measure $|++\rangle$ or $|--\rangle$ each with probability $\frac{1}{2}$, the second term must be zero, so that $\alpha_0 = \beta_0$ and $\alpha_1 = \beta_1$. Thus

$$|\Psi\rangle_{ABC} = \frac{|++\rangle + |--\rangle}{\sqrt{2}} \otimes ((\alpha_0) |0\rangle + (\alpha_1) |1\rangle)$$

which is of the desired form.
More generally, let $\rho_{AB} = |\Psi\rangle \langle\Psi|$ be a pure state. Then any purification of ρ_{AB} will be $\rho_{ABC} = \rho_{AB} \otimes |\Phi\rangle \langle\Phi|$. This state $AB : C$ has SR $= 1$ and is not an entangled state. Thus we see that **any pure state cannot have an entangled purification**. In other words, **a pure entangled state cannot have pure reduced states**.

3.7.1 Quantifying Monogamy

There are several measures of entanglement. One of them is the Schmidt Rank. However, that is valid only for pure states.
We will come across several measures throughout the course. Here are the desired properties of any such measure $E(A : B)$ on the joint system AB:

1. **Vanishes for separable states:** $E(A : B) = 0$ if the AB is separable
2. **Non-increasing under Local Operations and Classical Communication (LOCC):** If any local operations are performed on an entangled state or classical information is shared between the component states, then the entanglement cannot increase
3. **Monogamy Inequality:** $E(A : B) + E(A : C) \leq E(A : BC)$

§4. Bell Inequalities and Non-Local Games

In a **non-local game**, two or more players cooperate with one another to maximize their chances of winning. A refree runs the game and all communication in the game is between the players and the referee - no direct communication between the players is permitted. The referee randomly selects a question (from a foretold qurey distribution $\mathcal{Q}$) for each player and sends it to him/her. Each player sends the answer back to the referee, and based on all the answers and questions, he determines if the players win or lose.

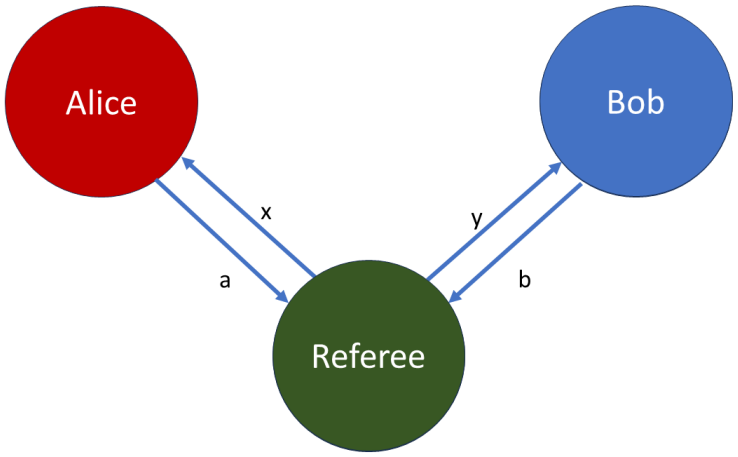


Figure 2: Non-Local Game

We can speak of classical strategies, where the players must behave classically, or quantum strategies which allow them to share an entangled state on which they can perform quantum measurements to determine their answers. The bounds on classical strategies (or local hidden variable theories) are typically known as **Bell Inequalities**, and quantum strategies (or experiments) that beat these bounds are said to *violate* a Bell inequality. We are particularly interested in such games where there exists a quantum strategy which performs better than any classical strategy. We want to maximize the probability of winning, i.e.:

$$p_{\text{win}} = \max_{\text{strategy}} \sum_{x,y} p(x,y) \sum_{a,b} \mathcal{V}(x,y,a,b) p(a,b|x,y)$$

Where $\mathcal{V}(x,y,a,b)$ is a winning indicator. Classically, Alice and Bob can use a deterministic strategy, viz functions $f_A(x) = a, f_B(x) = b$. Then $p(a,b|x,y) = 1$ iff $f_A(x) = a, f_B(x) = b$ and 0 otherwise. They can also use shared randomness (called hidden-variable in physics). This is modelled by giving them another string r that they share with probability $p(r)$. Then the functions depend on both r and the input provided: $p(a,b|x,y,r) = 1$ iff $f_A(x;r) = a, f_B(x;r) = b$. Note that once r is fixed, the strategy becomes deterministic.

$$p(a,b|x,y) = \sum_r p(r) p(a,b|x,y,r)$$

$$p_{\text{win}} = \max_{\text{strategy}} \sum_{x,y} p(x,y) \sum_{a,b} \mathcal{V}(x,y,a,b) \sum_r p(r) p(a,b|x,y,r)$$

$$= \max_{\text{strategy}} \sum_r p(r) \left(\sum_{x,y} p(x,y) \sum_{a,b} \mathcal{V}(x,y,a,b) p(a,b|x,y,r) \right)$$

The term in the brackets will be maximum for some specific value of r. So we can choose that value with probability 1 and this gives a deterministic strategy. This classical randomness doesn't give any advantage over the deterministic case.

There is a rich body of research on this area of non-local games, an excellent source for this is [Cle19]. In these notes, we will be looking at the CHSH game [CHSH69].

4.1 CHSH Game

The CHSH game is a non-local game used as a decisive test between quantum mechanics and hidden-local variable theories. In this game, the refree asks two questions $x, y \in \{0,1\}$ each to Alice and Bob, who in turn reply with $a, b \in \{0,1\}$. They are not allowed to communicate with each other during the game. They win if

$$a \oplus b = xy$$

.

x	y	Win	ab
0	0	0	00 or 11
0	1	0	00 or 11
1	0	0	00 or 11
1	1	1	01 or 10

Table 2: Winning the CHSH Game

Classical Strategy: The best deterministic classical strategy is to always output 00 for a and b, which wins with a probability 3/4. Further, we have seen before that a classical randomness provides no advantage. Thus the optimal classical strategy wins with probability 3/4
Quantum Strategy: In this case, Alice and Bob share an entangled state $|\phi_{00}\rangle$. Depending on the queries received, they measure their qubits in different bases according to Table 3:

	Input to Alice and Bob	**Basis to Measure**
x	0	$B_0 = \{|0\rangle, |1\rangle\}$
	1	$B_{\frac{\pi}{4}} = \{|+\rangle, |-\rangle\}$
y	0	$B_{\frac{\pi}{8}} = \{|\pi/8\rangle, |5\pi/8\rangle\}$
	1	$B_{\frac{-\pi}{8}} = \{|-\pi/8\rangle, |3\pi/8\rangle\}$

Table 3: Measurement Table

Suppose one of the qubits of the entangled pair is measured in the basis $B_\theta = \{|\theta\rangle, |\theta + \pi/2\rangle\}$ and the other is measured in B_ϕ. Then the probability of measuring $|\theta\rangle$ and $|\phi\rangle$ is

$$\left\| \langle \theta\phi | \phi_{00} \rangle \right\|^2 = \frac{\cos^2(\theta - \phi)}{2}$$

Consider the case when $x = y = 0$. Alice measures in basis B_0 and Bob measures in $B_{\frac{\pi}{8}}$. Then the probability of win is:

$$p_{\text{win}\,|\,00} = \left\| \left\langle 0\frac{\pi}{8} \middle| \phi_{00} \right\rangle \right\|^2 + \left\| \left\langle \frac{\pi}{2}\frac{5\pi}{8} \middle| \phi_{00} \right\rangle \right\|^2$$

$$= \cos^2\left(\frac{\pi}{8}\right) = \frac{1+\sqrt{2}}{2\sqrt{2}} \approx 0.85$$

Similarly, we can see for all other cases that the winning probability is the same. Hence

$$\Pr[\text{Win}_Q] \approx 0.85$$

4.1.1 What if the shared state is not fully entangled?

Consider the case when the shared state is the Werner state:

$$\rho_{AB} = p\,|\phi_{00}\rangle\langle\phi_{00}| + (1-p)\frac{\mathbb{I}}{2} \otimes \frac{\mathbb{I}}{2}$$

The probability of measuring $|\theta\phi\rangle$ on this state is:

$$\begin{aligned}\operatorname{tr}\Big(\rho_{AB}(|\theta\rangle\langle\theta|\otimes|\phi\rangle\langle\phi|)\Big) &= p\frac{\cos^2(\theta-\phi)}{2}+(1-p)\frac{1}{4}\operatorname{tr}\Big(|\theta\rangle\langle\theta|\otimes|\phi\rangle\langle\phi|\Big)\\ &= p\frac{\cos^2(\theta-\phi)}{2}+(1-p)\frac{1}{4}\operatorname{tr}(|\theta\rangle\langle\theta|)\operatorname{tr}\Big(|\phi\rangle\langle\phi|\Big)\\ &= p\frac{\cos^2(\theta-\phi)}{2}+(1-p)\frac{1}{4}\end{aligned}$$

Thus the new probability of success is now:

$$\Pr[\text{Win}_W] = p\cos^2\left(\frac{\pi}{8}\right)+(1-p)\frac{1}{2}\leq \Pr[\text{Win}_Q]$$

Example. Consider the case that Bob's bit got flipped due to some error. Now they share the state $|\phi_{01}\rangle$. Can they still get the same optimal probability with some classical processing of the input and outputs?
If they keep the protocol the same, then the probability of obtaining outcome $|\theta\phi\rangle$ is now

$$\frac{\sin^2(\theta+\phi)}{2}$$

So, if we follow the same procedure, then we won't get the optimal probability. Now instead, Bob flips the outcome he obtains. In this case, the difference between the measurement bases will be $3\pi/8$ and $\sin^2\left(\frac{3\pi}{8}\right)=\cos^2\left(\frac{\pi}{8}\right)$, resulting in similar probability.

4.1.2 Observable View

We can look at the CHSH game in another way using the notion of an *observable*:

Definition 4.1 (Observable)**.** Suppose we have a projective measurement $\{\Pi_a : a\in\Gamma\}$ for some finite set $\Gamma\in\mathbb{R}$ of measurement outcomes. Then the observable that corresponds to this measurement is a single Hermitian matrix:

$$A=\sum_{a\in\Gamma} a\Pi_a$$

The eigenspaces correspond to the projection operators and the eigenvalues correspond to the measurement outcomes. Expectation value of the observable on any state ρ is given by $\operatorname{tr}(A\rho)$

Now we make a small variation to the CHSH game: we identify the outputs given by Alice and Bob as $a\mapsto(-1)^a$ $b\mapsto(-1)^b$. Then the observables of Alice and Bob become:

$$A_0=|0\rangle\langle 0|-|1\rangle\langle 1|=\begin{pmatrix}1&0\\0&-1\end{pmatrix}=Z \qquad A_1=|+\rangle\langle +|-|-\rangle\langle -|=\begin{pmatrix}0&1\\1&0\end{pmatrix}=X$$

$$B_0=\left|\frac{\pi}{8}\right\rangle\left\langle\frac{\pi}{8}\right|-\left|\frac{5\pi}{8}\right\rangle\left\langle\frac{5\pi}{8}\right|=\frac{1}{\sqrt{2}}\begin{pmatrix}1&1\\1&-1\end{pmatrix}=H \qquad B_1=\left|-\frac{\pi}{8}\right\rangle\left\langle-\frac{\pi}{8}\right|-\left|\frac{3\pi}{8}\right\rangle\left\langle\frac{3\pi}{8}\right|=\frac{1}{\sqrt{2}}\begin{pmatrix}1&-1\\-1&-1\end{pmatrix}=\bar{H}$$

Observe that the projective measurements comprising the observables can be derived from them as

$$A_0^0=|0\rangle\langle 0|=\frac{1}{2}(\mathbb{I}+A_0) \qquad A_0^1=|1\rangle\langle 1|=\frac{1}{2}(\mathbb{I}-A_0)$$

Overall, we have to maximize the winning probability:

$$\Pr[\text{Win}_Q] = \frac{1}{4}\sum_{x,y}\sum_{a,b}\Pr[a \oplus b = x \cdot y]$$

Expanding using the expressions for the projective measurements, we obtain:

$$\Pr[\text{Win}_Q] = \frac{1}{2} + \frac{1}{8}\langle\phi_{00}| A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1 |\phi_{00}\rangle$$

Thus we have to maximize the expectation value of this observable:

$$S = A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1$$

For the classical case we have $S \leq 2$ and for the quantum case we have $S \leq 2\sqrt{2}$.

4.1.3 Tsirelson's Bound

Can we do better than 0.85 using any quantum strategy? Tsirelson's bound tells us that this is the maximum possible probability even for quantum computers.
Since all the observables defined above are measurements of the spin along some axis $\vec{a_i} \in \mathbb{R}^3$, $A_i = \vec{a_i} \cdot \vec{\sigma}$. We can show, by simple calculation and using $\{\sigma_i, \sigma_j\} = 0$ that

$$S^2 = (A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1)^2 = 4\mathbb{I} + [A_0, A_1] \otimes [B_1, B_0]$$

Moreover observe that

$$[A_0, A_1] = 2i(\vec{a_0} \times \vec{a_1}) \cdot \vec{\sigma}$$

Taking $\vec{a} = (\vec{a_0} \times \vec{a_1})$ and $\vec{b} = (\vec{b_0} \times \vec{b_1})$

$$\langle\psi| [A_0, A_1] \otimes [B_1, B_0] |\psi\rangle = \frac{-4}{2}\sum_{m,n\in\{0,1\}} \langle nn| \vec{a} \cdot \vec{\sigma} \otimes \vec{b} \cdot \vec{\sigma} |mm\rangle = -4(a_x b_x - a_y b_y + a_z b_z) \leq 4$$

Also, we know that $\langle A\rangle^2 \leq \langle A^2\rangle$ Hence

$$\langle S\rangle^2 \leq \langle S^2\rangle \leq 8$$

Thus we obtain the upper bound $S \leq 2\sqrt{2}$.
Though in the above proof, we assumed that the shared state was the EPR pair, we can do it more generally. Since the observables A_i, B_i have eigenvalues ± 1 the four terms of the commutator tensor product will have maximum value 1 each.

Another Proof

As above we have

$$\Pr[\text{Win}_Q] = \frac{1}{2} + \frac{1}{8}\langle\Psi| S |\Psi\rangle \leq \frac{1}{2} + \frac{1}{8}\|S\|$$

$$S^2 = 4\mathbb{I} + [A_0, A_1] \otimes [B_1, B_0]$$

$$\implies \left\|S^2\right\| \leq \|4\mathbb{I}\| + \left\|[A_0, A_1] \otimes [B_1, B_0]\right\|$$

Recall that $\|AB\| \leq \|A\|\|B\|$ and $\|A \otimes B\| = \|A\|\|B\|$. Since A_0, A_1, B_0, B_1 are binary observables squaring to identity, they have norm at most 1, so that

$$\left\|[A_0, A_1]\right\| = \|A_0A_1 - A_1A_0\| \leq \|A_0A_1\| + \|A_1A_0\| \leq 2$$

$$\left\|S^2\right\| \leq 4 + \left\|[A_0, A_1]\right\|\left\|[B_1, B_0]\right\| = 8$$

So that $\|S\| \leq 2\sqrt{2}$.

Yet another proof (using Sum-of-Squares)

Verify that

$$S = 2\sqrt{2}\mathbb{I} - \frac{q_0^2 + q_1^2}{\sqrt{2}}$$

where $q_0 = A_0 - \frac{B_0+B_1}{\sqrt{2}}$ and $q_1 = A_1 - \frac{B_0-B_1}{\sqrt{2}}$. The second term is positive, which gives us $S \leq 2\sqrt{2}\mathbb{I}$

4.1.4 Rigidity

The CHSH game has an interesting property that any strategy which obtains an optimal success probability (or close to it) must be "equivalent" to the strategy discussed above:

Theorem 4.1 (CHSH Rigidity)**.** Suppose we are given state $|\psi\rangle_{AB} \in \mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ and observables A_0, A_1 for Alice and B_0, B_1 for Bob such that the corresponding strategy has a success probability $p^*_{\text{CHSH}} = \cos^2(\pi/8)$ in the CHSH game. Then there exist local isometries $U_A : \mathbb{C}^{d_A} \to \mathbb{C}^2 \otimes \mathbb{C}^{d_{A'}}$ and $V_A : \mathbb{C}^{d_B} \to \mathbb{C}^2 \otimes \mathbb{C}^{d_{B'}}$ such that

- $(U_A \otimes V_B)|\psi\rangle_{AB} = |\phi_{00}\rangle \otimes |\text{junk}\rangle_{A'B'}$
- $(U_A \otimes V_B)(A_0 \otimes \mathbb{I}_B)|\psi\rangle_{AB} = ((Z \otimes \mathbb{I})|\phi_{00}\rangle) \otimes |\text{junk}\rangle_{A'B'}$
- $(U_A \otimes V_B)(A_1 \otimes \mathbb{I}_B)|\psi\rangle_{AB} = ((X \otimes \mathbb{I})|\phi_{00}\rangle) \otimes |\text{junk}\rangle_{A'B'}$
- $(U_A \otimes V_B)(\mathbb{I}_A \otimes B_0)|\psi\rangle_{AB} = ((\mathbb{I} \otimes H)|\phi_{00}\rangle) \otimes |\text{junk}\rangle_{A'B'}$
- $(U_A \otimes V_B)(\mathbb{I}_A \otimes B_1)|\psi\rangle_{AB} = ((\mathbb{I} \otimes \bar{H})|\phi_{00}\rangle) \otimes |\text{junk}\rangle_{A'B'}$

In this theorem, the notion of 'equivalence' is captured by the local isometries, because their range may not be the entire space. The dimensions d_A, d_B may not even be even! Also, the junk state is useless but unavoidable, as any strategy can be made to appear more complicated by extending the entangled state arbitarily, and make the players measurements act as identity on the extended space.

Remark. In practice, we would not be able to verify if the winning probability is $\cos^2 \pi/8$, but only upto some error $\delta > 0$. See [MYS12] for a robust analysis of the above theorem.

Proof. For proving this, we will require the notion of **principal angle** between two projections. In the two dimensional case, two unit vectors $|u\rangle, |v\rangle$ have an angle of $\theta \in [0, \pi/2]$ such that $\cos\theta = |\langle u|v\rangle|$. Upto a change of basis, we can always consider $u = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $v = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$. Defining $P = |u\rangle\langle u|$ and $Q = |v\rangle\langle v|$ we get

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad Q = \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix}$$

In higher dimensions, there may be several angles between two projectors. The principal angles are defined inductively:

$$\cos\theta_i = \sup_{\substack{|u_i\rangle \in P, |u_i\rangle \perp \text{Span}(|u_1\rangle, \ldots, |u_{i-1}\rangle) \\ |v_i\rangle \in P, |v_i\rangle \perp \text{Span}(|v_1\rangle, \ldots, |v_{i-1}\rangle)}} |\langle u_i|v_i\rangle|$$

where $|u_k\rangle, k \in [i-1]$ attains the supremum for $\cos\theta_k$. Jordan's lemma states that associated with the principal angles comes a very convenient simultaneous block decomposition of P and Q.

Lemma 4.2 (Jordan's Lemma)**.** Let P, Q be the projections on a separable Hilbert space $\mathcal{H}$. Then there exists an orthogonal decomposition

$$\mathcal{H} = \oplus_i \mathcal{S}_i$$

such that each S_i is a 1 or 2-dimensional subspace that is stable (invariant) by P and Q. Furthermore, whenever $\mathcal{S}_i$ is 2-dimensional, there is an orthonormal basis for it in which P and Q take the form

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad Q = \begin{pmatrix} c_i^2 & c_i s_i \\ c_i s_i & s_i^2 \end{pmatrix}$$

(restricted to the subspace $\mathcal{S}$) where $c_i = \cos\theta_i$ and $s_i = \sin\theta_i$, $\theta_i \in [0, \pi/2)$ are the principal angles between the projectors and may depend on S_i. In other words, there exists a basis of $\mathbb{C}^d$ in which P and Q are simultaneously block diagonal.

Informally, when only two projections are concerned, we can reduce the analysis to a 2-dimensional problem. Also, both P and Q are block diagonal matrices having the same block sizes with respect to a particular basis.

Proof. Consider $R = P + Q$. R is Hermitian and has an orthonormal set of eigenvectors, which forms a basis for $\mathcal{H}$. Now, let $|\phi\rangle$ be an eigenvector of R with eigenvalue λ.

$$Q|\phi\rangle = R|\phi\rangle - P|\phi\rangle = \lambda|\phi\rangle - P|\phi\rangle \tag{4.1}$$

Let $\mathcal{S} = \text{span}(|\phi\rangle, P|\phi\rangle)$. We take two cases:

1. $P|\phi\rangle = \mu|\phi\rangle$. Then $|\phi\rangle$ is a simultaneous eigenvector of P and Q due to Equation (4.1). Note that $\mu \in \{0, 1\}$ as P is a projector. So $\mathcal{S} = \text{span}(|\phi\rangle)$ is 1-dimensional and P, Q are either identity or 0 on $\mathcal{S}$.

2. $P|\phi\rangle$ is linearly independent of $|\phi\rangle$. This implies $Q|\phi\rangle$ is also linearly independent of $|\phi\rangle$ due to Equation (4.1). Then $\mathcal{S}$ is stable (invariant) under P, i.e., $P|\psi\rangle \in \mathcal{S}, \forall |\psi\rangle \in \mathcal{S}$. Moreover,

$$QP|\phi\rangle = Q(R - Q)|\phi\rangle = (\lambda - 1)Q|\phi\rangle$$

   so $\mathcal{S}$ is stable under Q too.

   Normalise the vectors $\{P|\phi\rangle, |\phi\rangle - P|\phi\rangle\}$ to obtain the orthonormal basis $\{|\psi_1\rangle, |\psi_2\rangle\}$ of $\mathcal{S}$. It is easy to check that $P|\psi_1\rangle = |\psi_1\rangle$ and $P|\psi_2\rangle = 0$, therefore, we can write $P = |\psi_1\rangle\langle\psi_1|$ restricted to $\mathcal{S}$.

   Let $|\Phi\rangle = Q|\phi\rangle$. Observe that $Q|\Phi\rangle = |\Phi\rangle$ and $Q|\Phi^\perp\rangle = 0$. In the basis of $\{|\psi_1\rangle, |\psi_2\rangle\}$ we can write $|\Phi\rangle = \cos\theta_i|\phi_1\rangle + e^{i\phi}\sin\theta_i|\phi_2\rangle = \cos\theta_i|\phi_1\rangle + \sin\theta_i|\phi_2\rangle$ where we overload $|\phi_2\rangle$ by absorbing the global phase $e^{i\phi}$ into it. Also, we can assume without loss of generality that $\theta \in (0, \pi/2]$. Take

$$\begin{aligned} Q &= |\Phi\rangle\langle\Phi| \\ &= (c_i|\psi_1\rangle + s_i|\psi_2\rangle)(c_i^*\langle\psi_1| + s_i^*\langle\psi_2|) \\ &= c_i^2|\psi_1\rangle\langle\psi_1| + c_i s_i(|\psi_1\rangle\langle\psi_2| + |\psi_2\rangle\langle\psi_1|) + s_i^2|\psi_2\rangle\langle\psi_2| \end{aligned}$$

   Thus we have, with respect to $\{|\psi_1\rangle, |\psi_2\rangle\}$

$$P_{|\mathcal{S}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad Q_{|\mathcal{S}} = \begin{pmatrix} c_i^2 & c_i s_i \\ c_i s_i & s_i^2 \end{pmatrix}$$

Finally, since $\mathcal{S}$ is stable by both P and Q, it is stable under $R = P + Q$, so it has a basis made of eigenvectors of R - the vector $|\phi\rangle$ we started from, and it's orthogonal in R. Proceeding in this way inductively lets us identify an eigenbasis of R such that its vectors are either isolated (stable by both P and Q) or in pairs (spanning a 2D subspace that is stable by both P and Q) ■

The proof of the rigidity theorem proceeds in the following steps:

1. **Reduction to qubit strategy** Consider an arbitrary strategy $\left(|\psi\rangle_{AB}, A_0, A_1, B_0, B_1\right)$. Apply Jordan's lemma to the projectors $P = \frac{1}{2}(\mathbb{I} + A_0)$ and $Q = \frac{1}{2}(\mathbb{I} + A_1)$. Then A_0 and A_1 are both block diagonal in some basis. This decomposition lets us reformulate Alice's strategy as follows: each of her two outcome projective measurement is equivalent to a measurement which:
   (a) Applies a *multiple outcome projective measurement* which projects on the individual blocks of the decomposition.
   (b) Depending on the block obtained as outcome, performs the basis measurement associated with the *restriction* of A_0 (or A_1) to that block.

   The same can be done with Bob's observables. This reformulation of an arbitrary strategy shows that it can always be reduced to a convex combination of qubit strategies, and it will be sufficient to analyze the latter.

2. **Optimal Strategies** We have shown above that $\|S\| \leq 2\sqrt{2}$. This claim is true for all strategies. But now using the above result, lets restrict ourselves to qubit strategies. The maximum is attained when

$$([A_0, A_1] \otimes [B_0, B_1])\,|\psi\rangle = -4\,|\psi\rangle$$

   Since A_0, A_1, B_0, B_1 are binary observables, $[A_0, A_1]^2 \leq 4\mathbb{I}$ and $[B_0, B_1]^2 \leq 4\mathbb{I}$

$$([A_0, A_1]^2 \otimes \mathbb{I})\,|\psi\rangle = (\mathbb{I} \otimes [B_0, B_1]^2)\,|\psi\rangle = 4\,|\psi\rangle$$

   This implies that both the observables of Alice and Bob anticommute:

$$A_0A_1 + A_1A_0 = 0 \quad B_0B_1 + B_1B_0 = 0$$

   This is strong constraint and implies the following:

Theorem 4.3. For any two Hermitian operators A, B acting on the finite-dimensional space $\mathbb{C}^2$ such that $A^2 = B^2 = \mathbb{I}$ and $AB = -BA$, there exists a unitary U acting on $\mathbb{C}^2$ such that

$$UAU^\dagger = X, \quad UBU^\dagger = Z$$

Proof. Since A, B satisfy $A^2 = B^2 = \mathbb{I}$, the only eigenvalues they can have are ± 1. Further, since $AB = -BA$, we cannot have either A or B as $\pm\mathbb{I}$. Thus, we can write their spectral decomposition as:

$$A = |a_+\rangle\langle a_+| - |a_-\rangle\langle a_-| \qquad B = |b_+\rangle\langle b_+| - |b_-\rangle\langle b_-|$$

Where $|a_\pm\rangle$ and $|b_\pm\rangle$ are their eigenvectors corresponding to the ± 1 eigenvalues respectively. Further, since A and B are Hermitian, their eigenvectors form an orthonormal basis of $\mathbb{C}^2$. Hence we have an unitary transformation that maps:

$$|b_+\rangle \mapsto |0\rangle \qquad |b_-\rangle \mapsto e^{i\theta_b}\,|1\rangle$$

Where θ_b is a phase to be determined later. Observe that

$$UBU^\dagger = U\,|b_+\rangle\langle b_+|\,U^\dagger - U\,|b_-\rangle\langle b_-|\,U^\dagger = |0\rangle\langle 0| - |1\rangle\langle 1| = Z$$

More explicitly, if we can think of $U = |0\rangle\,\langle b_+| + e^{i\theta_b}\,|1\rangle\,\langle b_-|$. Now, since $AB + BA = 0$, we have

$$\langle a_\pm|\,(AB + BA)\,|a_\pm\rangle = 0 \implies \langle a_\pm|\,B\,|a_\pm\rangle = 0 \implies |\,\langle b_+|a_\pm\rangle\,| = |\,\langle b_-|a_\pm\rangle\,|$$

Further, since $|\,\langle b_+|a_\pm\rangle\,|^2 + |\,\langle b_-|a_\pm\rangle\,|^2 = 1$,

$$|\,\langle b_+|a_\pm\rangle\,| = |\,\langle b_-|a_\pm\rangle\,| = \frac{1}{\sqrt{2}}$$

The action of U on $|a_\pm\rangle$ gives us the vector $\begin{pmatrix} \langle b_+|a_\pm\rangle \\ e^{i\theta_b}\langle b_-|a_\pm\rangle \end{pmatrix}$ in the standard basis. Without loss of generality, we can write this as

$$|a_+\rangle \mapsto \frac{e^{i\theta_{a+}}}{\sqrt{2}}\begin{pmatrix} 1 \\ e^{i\theta_b}e^{i\phi_{a+}} \end{pmatrix} \qquad |a_-\rangle \mapsto \frac{e^{i\theta_{a-}}}{\sqrt{2}}\begin{pmatrix} 1 \\ e^{i\theta_b}e^{i\phi_{a-}} \end{pmatrix}$$

For some phases $\theta_{a\pm}, \phi_{a\pm}$. These have a inner product 0, which gives

$$1 + e^{i\phi_{a+}-\phi_{a-}} = 0 \implies e^{i\phi_{a+}} = -e^{i\phi_{a-}}$$

Now, taking $\theta_b = -\phi_{a+}$ gives

$$|a_+\rangle \mapsto \frac{e^{i\theta_{a+}}}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = e^{i\theta_{a+}}|+\rangle \qquad |a_-\rangle \mapsto \frac{e^{i\theta_{a-}}}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = e^{i\theta_{a-}}|-\rangle$$

Finally, this implies

$$UAU^\dagger = |+\rangle\langle+| - |-\rangle\langle-| = X$$

Hence U is a unitary transformation that maps $UAU^\dagger = X$ and $UBU^\dagger = Z$. ■

From Theorem 4.3 we have unitaries U_A, U_B acting respectively on Alice and Bob's systems such that

$$U_A A_0 U_A^\dagger = Z U_A A_1 U_A^\dagger = X$$
$$U_B B_0 U_B^\dagger = H U_B B_1 U_B^\dagger = \bar{H}$$

Finally we can verify that for

$$S^* = Z\otimes H + Z\otimes \bar{H} + X\otimes H - X\otimes \bar{H}$$

has max eigenvalue $2\sqrt{2}$ with eigenvector $|\psi^+\rangle$.

3. **Putting Everything Together:** In summary, we start with an arbitrary strategy $(|\psi_{AB}\rangle, A_x, B_y)$. Let $\Pi^A = \{\Pi_0^A, \Pi_1^A \dots \Pi_{k_A}^A\}$ and $\Pi^B = \{\Pi_0^B, \Pi_1^B \dots \Pi_{k_B}^B\}$ be the projectors corresponding to the basis in which A_0, A_1 and B_0, B_1 are block diagonal. Then we can write $A_x = \sum_j \Pi_j^A A_x \Pi_j^A$ and $B_y = \sum_j \Pi_j^B B_y \Pi_j^B$ for $x, y \in \{0,1\}$. We know that any strategy can win with probability at most p^*_{CHSH}, hence all the qubit strategies $(\Pi_i^A \otimes \Pi_j^B |\psi_{AB}\rangle, \Pi_i^A A_x \Pi_i^A, \Pi_j^B B_y \Pi_j^B)$ must have success probability p^*_{CHSH}. Finally, we can use the unitary of Theorem 4.3 to write the operators in the desired form.

■

§5. Distance Measures for Qunatum Information

In qunatum information theory, there are two types of distance measures:

- **Static Measures:** Quantify how close two quantum states are
- **Dynamic Measures:** Quantify how well information has been preserved during a dynamic process.

Classically[1], two measures are popular:

1. **Trace Distance/Kolmogorov Distance/L_1 distance:**
$$D(p,q) := \frac{1}{2}\sum_x |p_x - q_x|$$
2. **Fidelity:**
$$F(p,q) := \sum_x \sqrt{p_x q_x}$$

We have similar measures for quantum information too!

5.1 Trace Distance

Suppose we have a device which is equally likely to produce the ideal (ρ_{KE}^{Ideal}) or the real state(ρ_{KE}^{Real}). How do we find out which state it has produced?
Define a POVM $\{M_{\text{real}}, M_{\text{ideal}}\}$ which is used to distinguish between the states $\rho_{KE}^{\text{ideal}}, \rho_{KE}^{\text{real}}$. Then the probability of distinguishing is:

$$\begin{aligned}
p_{\text{succ}} &= \Pr\left(\rho_{KE} = \rho_{KE}^{\text{real}}\right)\Pr\left(\text{output=real}|\rho_{KE} = \rho_{KE}^{\text{real}}\right) + \Pr\left(\rho_{KE} = \rho_{KE}^{\text{real}}\right)\Pr\left(\text{output=ideal}|\rho_{KE} = \rho_{KE}^{\text{ideal}}\right)\\
&= \frac{1}{2}\operatorname{tr}\left(M_{\text{real}}\rho_{KE}^{\text{real}}\right) + \frac{1}{2}\operatorname{tr}\left(M_{\text{ideal}}\rho_{KE}^{\text{ideal}}\right)\\
&= \frac{1}{2}\operatorname{tr}\left(M_{\text{real}}\rho_{KE}^{\text{real}}\right) + \frac{1}{2}\operatorname{tr}\left((\mathbb{I} - M_{\text{real}})\rho_{KE}^{\text{ideal}}\right)\\
&= \frac{1}{2} + \frac{1}{2}\operatorname{tr}\left(M_{\text{real}}(\rho_{KE}^{\text{real}} - \rho_{KE}^{\text{ideal}})\right)
\end{aligned}$$

Definition 5.1 (Trace Distance)**.** Given two states σ, ρ, the trace distance between them is defined as:

$$D(\rho,\sigma) := \max_{0\le M\le \mathbb{I}} \operatorname{tr}\big(M(\rho-\sigma)\big)$$

Where the maximum is taken over all POVM M. Note that $0 \le M \le \mathbb{I}$ by definition of POVM. It can also be written as:

$$D(\rho,\sigma) = \frac{1}{2}\operatorname{tr}|\rho-\sigma|$$

Where $|A| = \sqrt{A^\dagger A}$. $\operatorname{tr}|A|$ can be thought of as the sum of singular values of A, and is called the **trace norm** of A. Hence, **the trace distance between two states is half of the trace norm of their difference**.
We will show the equivalence of the above definitions in Theorem 5.1

Some points to note

[1] In classical information theory, an information source is modelled as a random variable P, a probability distribution over some source alphabet Σ. $p_x = \Pr(P = x)$

- In literature, it is also written as $\frac{1}{2}\|\rho-\sigma\|_{\text{tr}}$ or $\frac{1}{2}||\rho-\sigma||_1$
- Two quantum states ρ and σ are ϵ-close, if $D(\rho,\sigma)\leq\epsilon$. We also write this as $\rho\approx_\epsilon\sigma$
- If ρ and σ commute, then the quantum trace distance is equivalent to the classical one, i.e. it is the sum of the absolute difference of their corresponding eigenvalues.
- Since POVM elements are positve, the optimal M for trace distance will be **the projector on the positive eigenspace of** $\rho-\sigma$
- Using the above expression for the trace distance, we obtain that the maximum probability of distinguishing between two states ρ and σ is $\frac{1}{2}+\frac{1}{2}D(\rho,\sigma)=\frac{1}{2}+\frac{1}{4}\|\rho-\sigma\|_{\text{tr}}$. This is known as the **Holevo-Helstrom Theorem**.

Intuitive meaning:

Recall that the density matrix for a qubit ρ can be written as

$$\rho=\frac{I+\vec{r}\cdot\vec{\sigma}}{2}$$

Where $\vec{r}\in\mathbb{R}^3$ is the *Bloch Vector* of ρ; $\|\vec{r}\|\leq 1$ with equality only for pure states. Writing $\rho=\frac{I+\vec{r}\cdot\vec{\sigma}}{2}$ and $\sigma=\frac{I+\vec{s}\cdot\vec{\sigma}}{2}$, the trace distance becomes

$$\frac{1}{2}\operatorname{tr}|\rho-\sigma|=\frac{1}{4}\operatorname{tr}|(\vec{r}-\vec{s})\cdot\vec{\sigma}|$$

This has eigenvalues $\pm\|\vec{r}-\vec{s}\|$. We also know that if A is diagonalizable then $\operatorname{tr}|A|=\sum_i|\lambda_i|$, the sum of absolute values of eigenvalues of A. Hence,

$$D(\rho,\sigma)=\frac{\|\vec{r}-\vec{s}\|}{2}$$

i.e., **half of the distance between their Bloch vectors**. This property has several implications. For instance, since euclidean distance is invariant under rotation or reflection, we have the property (3) below.

Properties of Trace Distance

The trace distance is a metric over the space of density matrices.

1. **Non-Negativity:** $D(\rho,\sigma)\geq 0,\qquad D(\rho,\sigma)=0\iff\rho=\sigma$
2. **Symmetry:** $D(\rho,\sigma)=D(\sigma,\rho)$
3. **Triangle Inequality:** $D(\rho,\tau)\leq D(\rho,\sigma)+D(\sigma,\tau)$
4. **Invariance under Unitary Transforms:** $D(U\rho U^\dagger,U\sigma U^\dagger)=D(\rho,\sigma)$
5. **Convexity:** $D(\sum_i p_i\rho_i,\sigma)\leq\sum_i p_i D(\rho_i,\sigma)$

Theorem 5.1. Both the definitions in Definition 5.1 are equivalent. Namely

$$D(\rho,\sigma)=\frac{1}{2}\operatorname{tr}|\rho-\sigma|=\max_P\operatorname{tr}\big(P(\rho-\sigma)\big)$$

Where the maximization is taken alternately over all projectors P, or over all POVMs $P\leq\mathbb{I}$; the formula is valid in either case.

Proof. (Easier) Since $0 \leq P \leq \mathbb{I}$, we claim that the maximum will be attained when P is a projector on the positive eigenspace of $\rho - \sigma$. This is because for any general POVM (E_0, E_1),

$$\operatorname{tr}\big(E_0(\rho-\sigma)\big) = \operatorname{tr}\left(\left(\frac{E_0+E_1}{2} + \frac{E_0-E_1}{2}\right)(\rho-\sigma)\right) = \frac{1}{2}\operatorname{tr}(E_0-E_1)(\rho-\sigma)$$

Now, for any two normal matrices L, M we have $\operatorname{tr}(LM) \leq \|L\|_\infty \|M\|_1$. Using this in the above expression and the fact that $\|E_0 - E_1\|_\infty \leq 1$ because they are 2 element POVMs and are diagonal in the same basis, we obtain the upper bound. Taking E_0 to be the projector on the positive eigenspace, we see that the inequality will be attained. ■

Proof. [NC10] For the proof, we need this lemma:

Lemma 5.2. For any two states ρ, σ one may write $\rho - \sigma = Q - S$ where Q, S are positive operators with support on orthogonal vector spaces.

Proof. The spectral decomposition of $\rho - \sigma = UDU^\dagger$. Observe that since $\rho - \sigma$ is Hermitian, all the eigenvalues (diagonal elements of D) will be real. Decompose D into its positive and negative eigenvalues: $D = P - N$ where both P and N are diagonal matrices with non-negative entries. Moreover, $PN = 0$ is trivial. Takeing $Q = UPU^\dagger$ and $S = UNU^\dagger$ we are done. ■

Now coming back to the main result, we prove the equation for the maximization over all projectors. Using Theorem 5.2

$$|\rho - \sigma| = \sqrt{(Q-S)(Q-S)} = Q + S$$

since $\sqrt{A^\dagger A}$ is the sum of absolute values of the eigenvalues of A.

$$\frac{1}{2}\operatorname{tr}|\rho-\sigma| = \frac{1}{2}\operatorname{tr}(Q+S) = \operatorname{tr}(Q)$$

where the second equality follows from $\operatorname{tr}(Q-S) = \operatorname{tr}\big(\rho-\sigma\big) = 0 \implies \operatorname{tr}(Q) = \operatorname{tr}(S)$. Taking P to be the support on Q gives the result. Conversely let P be any projector. Then

$$\operatorname{tr}\big(P(Q-S)\big) \leq \operatorname{tr}(PQ) \leq \operatorname{tr}(Q)$$

Where the second inequality is because for $Q = \sum_i \lambda_i |i\rangle\langle i|, P = \sum_j |j\rangle\langle j|$,

$$\operatorname{tr}(PQ) = \sum_{ij} \lambda_i \left\|\langle i|j\rangle\right\|^2 \leq \sum_i \lambda_i = \operatorname{tr}(Q)$$

By definition of a projector, since j are elements of an orthonormal basis, $\sum_i \left\|\langle i|j\rangle\right\|^2 \leq 1$. ■

Definition 5.2 (ϵ-Ball of ρ)**.** Given any state ρ, its ϵ ball is the set of states within ϵ trace distance

$$\mathcal{B}^\epsilon(\rho) := \left\{\rho' | \rho' \geq 0, \operatorname{tr}\left(\rho'\right) = 1, D(\rho, \rho') \leq \epsilon\right\}$$

5.2 Fidelity

Definition 5.3 (Fidelity)**.** Given two states σ, ρ, the fidelity is defined as :

$$F(\rho, \sigma) := \text{tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]$$

Intuitive Meaning:

Consider the experiment in which we are trying to generate the state $|\psi\rangle$ but our machine produces the state ρ. To check the success probability, we define the set of POVM operators $\{M_{\text{succ}}, M_{\text{fail}}\}$ where $M_{\text{succ}} = |\psi\rangle\langle\psi|, M_{\text{fail}} = \mathbb{I} - M_{\text{succ}}$. Probability that we prepared the correct state is:

$$\Pr[\text{success}] = \text{tr}[M_{\text{succ}}\rho] = F(|\psi\rangle, \rho)^2$$

Properties of Fidelity:

1. For pure states $\rho = |\Psi\rangle\langle\Psi|, \sigma = |\Phi\rangle\langle\Phi|, \qquad F(\rho, \sigma) = \|\langle\Psi|\Phi\rangle\|$
2. If only one of the states (ρ) is pure then $F(\rho, \sigma) = \sqrt{\langle\Psi|\sigma|\Psi\rangle}$
3. F is not a metric since $F(\rho, \sigma) = 0$ doesn't imply $\rho = \sigma$
4. $F(\rho, \rho) = 1$. F is 1 for identical states and gets smaller and smaller as the states are different
5. **Invariant under Unitary Transforms:** $F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho, \sigma)$
6. **Bounded between 0 and 1:** $0 \leq F(\rho, \sigma) \leq 1$
7. **Symmetry:** $F(\rho, \sigma) = F(\sigma, \rho)$
8. **Multiplicative under Tensor Product:** $F(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) = F(\rho_1, \sigma_1)F(\rho_2, \sigma_2)$
9. **Relation to Trace Distance:** (Fuchs-van de Graaf Inequality)

$$1 - F(\rho, \sigma) \leq D(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)^2}$$

$$1 - D(\rho, \sigma) \leq F(\rho, \sigma) \leq \sqrt{1 - D(\rho, \sigma)^2}$$

Proof. 3. This follows from the fact that for any positive operator A with Spectral Decomposition $\sum_i \lambda_i |i\rangle\langle i|$ we have for unitary $U, UAU^\dagger = \sum_i \lambda_i U|i\rangle\langle i|U^\dagger = \sum_i \lambda_i |i'\rangle\langle i'|$ for another orthonormal basis $\{|i'\rangle\}$. So,

$$\sqrt{UAU^\dagger} = U\sqrt{A}U^\dagger$$

■

Another way to write the Fidelity is given by the Uhlmann's Theorem

Theorem 5.3 (Uhlmann's Theorem)**.** Suppose ρ, σ are two states of a quantum system Q. Introduce a second quantum system R which is a copy of Q. Then

$$F(\rho.\sigma) = \max_{|\psi\rangle|\phi\rangle} \left\|\langle\psi|\phi\rangle\right\|$$

Where the maximization is over all the purifications $|\psi\rangle$ of ρ and $|\phi\rangle$ of σ into RQ

5.3 Relationship between Trace Distance and Fidelity

For pure states, trace distance and fidelity are equivalent. Infact, we have

$$D(|\psi\rangle, |\phi\rangle) = \sqrt{1 - |\langle\phi|\psi\rangle|^2} = \sqrt{1 - F\left(|\psi\rangle, |\phi\rangle\right)}$$

Proof. Write $|\phi\rangle = \alpha |\psi\rangle + \beta \left|\psi^{\perp}\right\rangle$ and find the eignevalues of the 2×2 matrix in the basis of $|\phi\rangle$ and $\left|\psi^{\perp}\right\rangle$ ■

§6. Quantifying Information

What do we precisely mean when we say that Eve is ignorant about the key? We imply two properties:

1. All keys are equally likely: $\rho_K = \mathbb{I}/|K|$, where $|K|$ is the size of the key space
2. Eve doesn't have any correlation with the key: $\rho_{KE} = \rho_K \otimes \rho_E$

Ideally, we want that

$$\rho_{KE} = \frac{\mathbb{I}}{|K|} \otimes \rho_E$$

Consider the case that both the registers K, E are classical. Then we can write:

$$\begin{aligned}\rho_{KE} &= \sum_{k,e} \Pr[K = k, E = e] \, |k\rangle \langle k| \otimes |e\rangle \langle e| \\ &= \sum_k \Pr[K = k] \, |k\rangle \langle k| \otimes \sum_e \Pr\big[E = e | K = k\big] \, |e\rangle \langle e|\end{aligned}$$

As all keys are equally likely, $\Pr[K = k] = 1/|K|$ and as Eve should not know anything given the key $\Pr\big[E = e|K = k\big] = \Pr[E = e]$.

$$\rho_{KE} = \sum_k \frac{1}{|K|} \, |k\rangle \langle k| \otimes \sum_e \Pr[E = e] \, |e\rangle \langle e| = \frac{\mathbb{I}}{|K|} \otimes \rho_E$$

Same as the quantum case.

6.1 Min-Entropy

How do we quantify randomness? Suppose that a string $x_0 x_1 ... x_n \in \{0,1\}^n$ is choosen uniformly randomly, i.e.

$$\Pr[X = x] = \frac{1}{2^n}$$

$$\rho = \frac{\mathbb{I}}{2^n}$$

and a state

$$\rho' = \frac{\mathbb{I}}{2^{n-1}} \otimes |0\rangle \langle 0|$$

These two states only differ on the last qubit, hence it seems that the second state is random to a good approximation. But their trace distance is $1/2$, which is quite large (same as the distance between $|0\rangle \, and \, |+\rangle$!). So, trace distance isn't a very good measure to quantify approximate randomness.

Definition 6.1 (Information Content[Wil16])**.** The information content $i(x)$ of a particular realization x of the random variable X is a measure of the surprise that one has upon learning the outcome of the experiment.

$$i(x) = -\log\big(p_X(x)\big)$$

This definition captures our intuitive idea for surprise - it is higher for events with less probability

Definition 6.2 (Shannon Entropy)**.** The Shannon entropy of a discrete random variable X with probability distribution $p_X(x)$ is

$$H(X) := \mathop{\mathbb{E}}_x i(x) = -\sum_x p_x \log\big(p_x\big)$$

- When all strings are equally likely

$$H(X_1) = -\sum_x \frac{1}{2^n} \log\left(\frac{1}{2^n}\right) = n$$

- For $\rho' = \frac{\mathbb{I}}{2^{n-1}} \otimes |0\rangle\langle 0|$

$$H(X_2) = -\sum_{x\in\{0,1\}^n, x[-1]=0} \frac{1}{2^{n-1}} \log\left(\frac{1}{2^{n-1}}\right) = n-1$$

- Consider a probability distribution

$$Pr[X_3 = x] = \begin{cases} \frac{1}{2} & x = 111..111 \\ \frac{1}{2(2^n-1)} & Otherwise \end{cases}$$

  In this case, the entropy is $\approx n/2$. However, Eve has probability $1/2$ of guessing the correct key (by selecting 111...11). So Shannon Entropy too isn't a very good measure of approximate randomness.

Definition 6.3 (Min Entropy)**.**

$$H_{\min}(X) := -\log\left(\max_x p_x\right) = -\log\left(\Pr_{\text{guess}}(X)\right)$$

- The min entropy captures the worst case probability that the string is guessed correctly.
- It is exactly the quantity that tells us how many truly random bits we can, almost perfectly, produce from a randomness box with a certain amount of min entropy.
- In the example above for X_3 we have

$$H_{\min}(X_3) = 1$$

  which is independent of n. It means that as the size of our string increases, the min-entropy remains the same (though the Shannon entropy increases).

Lemma 6.1. The min-entropy satisfies the following bounds:

$$0 \le H_{\min}(X) \le H(X) \le \log|X|$$

Proof. The first inequality follows from $0 \le \Pr(x) \le 1$. The second inequality is because $\log\big(\Pr(X=x)\big) \le \log\big(\max \Pr(X=x)\big)$. The third inequality is because the entropy is maximum for a completely random string. Formally, it can be proved by the weighted AM GM inequality:

$$e^{H(X)} = \prod_x \left(\frac{1}{p_x}\right)^{p_x} \le |X|$$

■

Definition 6.4 (Conditional Min Entropy)**.**

$$H_{\min}(X|E) := -\log\left(\sum_e \Pr(E=e) \max_x \Pr\big(X=x|E=e\big)\right) = -\log\left(\Pr_{\text{guess}}(X|E)\right)$$

- The conditional min-entropy quantifies the maximum probability of guessing X given that Eve has access to a (quantum) register E.
- If Eve has quantum information about x then the joint state is the cq-state $\rho_{XE} = \sum_x p_x |x\rangle\langle x|^X \otimes \rho_x^E$. Then

$$\Pr_{\text{guess}}(X|E) = \max_{\{M_x\}} \sum_x \Pr(X = x) \operatorname{tr}\left(M_x^E \rho_x^E\right)$$

  Where the maximum is taken over all POVM $\{M_x\}$. E is also called the *side information* about x.
- **Largest and Smallest Values:** $0 \leq H_{\min}(X|E)_\rho \leq \log |X|$

  Proof. The first inequality follows from $0 \leq \Pr_{\text{guess}}(X|E) \leq 1$ and the second inequality is because the probability of guessing X with side information e will be at least the same as randomly guessing. ■
- **Conditioning Reduces Entropy:** $H_{\min}(X|EF)_\rho \leq H_{\min}(X|E)_\rho$

  This seems intuitive since if the amount of information with Eve increases, the probability of guessing the state correctly by her would increase. Hence the entropy decreases.
- **Reduction by one register:** $H_{\min}(X|E)_\rho \geq H_{\min}(X)_\rho - \log |E|$

Example. Let us calculate the conditional entropy of the state $\rho_{XE} = \frac{1}{2} |0\rangle \langle 0|^X \otimes |0\rangle \langle 0|^E + \frac{1}{2} |1\rangle \langle 1|^X \otimes |+\rangle \langle +|^E$.

$$\Pr_{\text{guess}}(X|E) = \max_{\{M_x\}} \sum_x \Pr(X = x) \operatorname{tr}\left(M_x^E \rho_x^E\right) = \max_{\{M_x\}} \left\{ \frac{1}{2} \operatorname{tr}(M_0 |0\rangle \langle 0|) + \frac{1}{2} \operatorname{tr}(M_+ |+\rangle \langle +|) \right\}$$

$$= \frac{1}{2}(1 + D(|0\rangle, |+\rangle)) = \frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\right)$$

In this case, since the dimension was 2, it was easy to calculate the optimal POVM operators, using the trace distance. However, it gets complex as the dimension increases. Nevertheless, finding the optimal success probability is a so-called semi-definite program (SDP) and can be evaluated efficiently (in the dimension of the states ρ_x^E) using for example Matlab or Julia.

A general quantum conditional min-entropy

Till now, we have considered X to be classical. However, in the fully general case, X can be quantum too.

Definition 6.5 (Quantum Conditional Min-Entropy[KRS09])**.** Given any bipartite density matrix ρ_{AX}, with A having dimension $|A|$, the conditional min entropy is

$$H_{\min}(A|E) := -\log[|A| \text{Dec}(A|E)]$$

$$\text{Dec}(A|E) := \max_{\Lambda_{E \to A'}} F((\mathbb{I}_A \otimes \Lambda_{E \to A'})\rho_{AE}, |\Phi\rangle \langle \Phi|_{AA'})^2$$

where $|\Phi_{AA'}\rangle := \frac{1}{|A|} \sum_{i=1}^{|A|} |a_i\rangle_A \otimes |a_i\rangle_{A'}$ is the maximally entangled state over AA' and the maximization is performed over all quantum channels Λ mapping $E \mapsto A'$

Smoothed Min-Entropy

Often, we don't know exactly what the state ρ'_{AE} is, the only information we may have is $\rho'_{AE} \approx_\epsilon \rho_{AE}$. In that case, Smoothed Min-Entropy is used:

Definition 6.6. Smoothed Conditional Min-Entropy

$$H^\epsilon_{\min}(X|E)_\rho := \max_{\rho' \in \mathcal{B}^\epsilon(\rho)} H_{\min}(X|E)_{\rho'}$$

6.2 The Uncertainity Game

The modern version of looking at the Uncertainity principle is in the form of games played between players.

6.2.1 Two player uncertainity game

Consider the following game between Alice and Eve:

1. Eve prepares a qubit in the state ρ_A and sends it to Alice
2. Alice samples $\Theta \leftarrow \{0,1\}$ and measures her qubit as:

$$\Theta = \begin{cases} 0 & \text{measure in the standard basis} \\ 1 & \text{measure in the hadamard basis} \end{cases}$$

   Let her measurement outcome be X
3. Alice sends Θ to Eve
4. Eve wins the game if she predicts X correctly

To see why this captures the essence of the uncertainty principle, note that if the measurements are incompatible, then there exists no state that Eve can prepare, that would allow her to guess the outcomes for both choices of measurements with certainty. Uncertainty can thus be quantified by a bound on the average probability that Eve correctly guesses X

$$\Pr_{\text{guess}}(X|\Theta) = \frac{1}{2}\left(\Pr_{\text{guess}}(X|\Theta=1) + \Pr_{\text{guess}}(X|\Theta=0)\right) \leq c$$

We will show that in the case where Eve knows only about the measurement basis, $c < 1$. Even if Eve has some extra classical register K, the inequality is unaffected. In terms of min entropy we have $H_{\min}(X|K\Theta) > 0$

Analysis

Let $\rho_A = \frac{\mathbb{I}+\vec{r}.\vec{\sigma}}{2}$. Then we have,

$$\text{tr}(\rho_A |0\rangle\langle 0|) = \frac{1+v_z}{2} \qquad \text{tr}(\rho_A |1\rangle\langle 1|) = \frac{1-v_z}{2}$$

$$\text{tr}(\rho_A |+\rangle\langle +|) = \frac{1+v_x}{2} \qquad \text{tr}(\rho_A |-\rangle\langle -|) = \frac{1-v_x}{2}$$

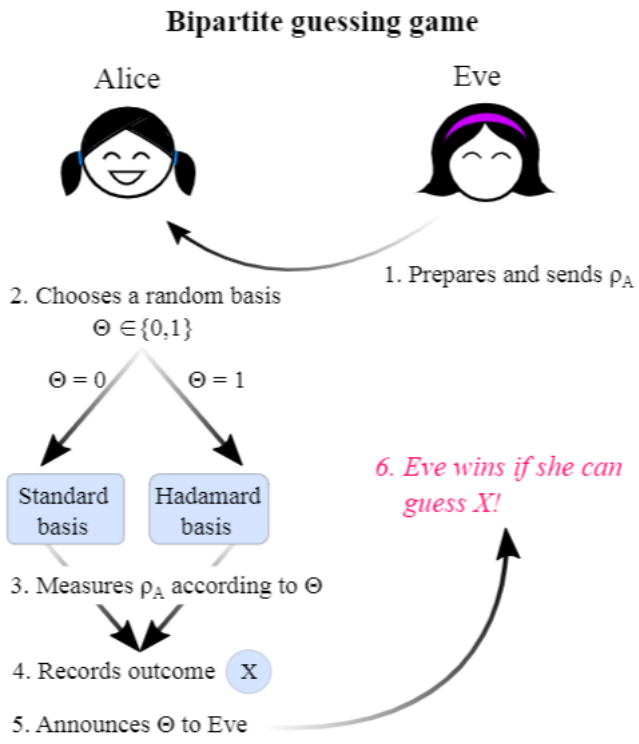


Figure 3: Bipartite Guessing Game (Taken from [VW16])

Where $v_x^2 + v_z^2 \leq 1$. Then we have,

$$\Pr_{\text{guess}}(X|\Theta) = \frac{1}{2}\left(\max\left\{\frac{1+v_z}{2}, \frac{1-v_z}{2}\right\} + \max\left\{\frac{1+v_x}{2}, \frac{1-v_x}{2}\right\}\right) = \frac{1}{4}(2 + |v_x| + |v_z|)$$

Using Cauchy-Schwarz inequality:

$$(|v_x| + |v_z|)^2 \leq (v_x^2 + v_z^2)(1+1) \leq 2$$

$$\Pr_{\text{guess}}(X|\Theta) = \frac{1}{4}(2+\sqrt{2}) \approx 0.85 < 1$$

Remark. If Eve can keep a qubit with herself, then she can win with probability 1! (EPR pair). In fact we have,

- **No entanglement** (Eve has just classical information): Large Uncertainity

$$H_{\min}(X|\Theta E) \approx 0.22$$

- **Some entanglement**: Some uncertainity

$$H^{\epsilon}_{\min}(X|\Theta E) \geq f_M(H_{\min}(A|E))$$

These are known as *entropic uncertainity relations with quantum side infromation*. This relation says that for any kind of measurements that Alice might perform there exists a bound by some function f_M of the amount of entanglement between A and E(M indicating that this function will depend on the measurements that Alice does).

- **Full entanglement**: No Uncertainity

$$H_{\min}(X|\Theta E) = 0$$

6.2.2 Three bases uncertainity game

This is similar to the above game, but here we can measure in the eigenbasis of Y as well. Thus, Alice measures in one of the three basis and returns $\Theta \in \{0,1,2\}$. Now Eve has to figure out the measurement outcome.

Analysis

Similar to above, we now see that the expression for maximum probability of success becomes:

$$\Pr_{\text{guess}}(X|\Theta) = \frac{1}{3}\left(\max\left\{\frac{1+v_z}{2}, \frac{1-v_z}{2}\right\} + \max\left\{\frac{1+v_y}{2}, \frac{1-v_y}{2}\right\} + \max\left\{\frac{1+v_x}{2}, \frac{1-v_x}{2}\right\}\right)$$

$$\Pr_{\text{guess}}(X|\Theta) = \frac{3+|v_x|+|v_y|+|v_z|}{6} \leq \frac{3+\sqrt{3}}{6} \approx 0.79$$

Which is less than the bipartite game. Intuitively this can be explained by us giving Alice more "incompatible" options. This makes it harder for Eve to find a state that has good overlap with all three vectors corresponding to the outcome "zero" in their respective bases.

6.2.3 Tripartite Uncertainity game

Consider the tripartite version of the game as shown in Fig. 4:

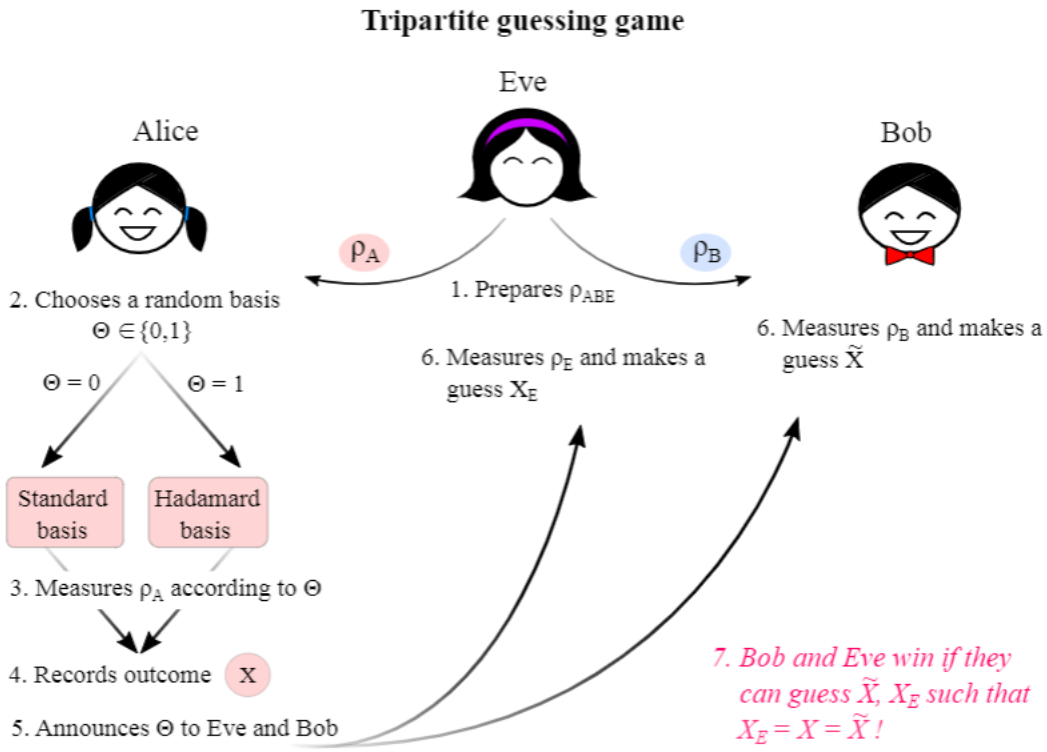


Figure 4: Tripartite Uncertainity Game (Taken from [VW16])

The main difference in this case is that Eve would prepare a three qubit state and give two qubits to Alice and Bob. We will see that if Alice and Bob have sufficient entanglement between them, then the maximum probability for Eve to guess Alice's measurement result is ≈ 0.85 - the same as when she lacked any quantum memory! This highlights an important property of entanglement - that it is monogamous, so these games are called **Monogamy of Entanglement games**[TFKW13]. Formally, we have the following definitions:

Definition 6.7 (Monogamy of Entanglement Game)**.** A monogamy of entanglemnt game $\mathcal{G}$ consists of a finite dimensional Hilbert space $\mathcal{H}_A$ and a list of measurements $\mathcal{M}^\theta = \{A_x^\theta\}_{x\in\mathcal{X}}$ on $\mathcal{H}_A$, indexed by $\theta \in \Theta$, where $\mathcal{X}$ and Θ are finite sets. Denote the n-fold repetition of game $\mathcal{G}$ as $\mathcal{G}^{\times n}$

In our case, we call the game $\mathcal{G}_{\text{BB84}}$ since Alice measures in the BB84 basis states. Observe that in this case Alice measures using a Wiesner state $A_x^\theta = \left|x^\theta\right\rangle\!\left\langle x^\theta\right| = H^\theta \left|x\right\rangle\!\left\langle x\right| H^\theta$.

Definition 6.8. A strategy $\mathcal{S}$ for a monogamy game $\mathcal{G}$ is a list

$$\mathcal{S} = \{\rho_{ABC}, B_x^\theta, C_x^\theta\}_{\theta\in\Theta, x\in\mathcal{X}}$$

Where the maximum is taken over all strategies $\mathcal{S}$. A strategy is called *pure* if ρ_{ABC} is pure and all the POVMs are projective.

Lemma 6.2. In the above maximization, it is sufficient to consider only pure strategies.

Proof. Purify ρ to $\left|\phi\right\rangle\!\left\langle\phi\right|$ by adding an auxilary register and use Naimark's Dialation Theorem to simulate the POVM by a projective measurement and unitary transform. ■

Definition 6.9. The winning probability for $\mathcal{G}$ with strategy $\mathcal{S}$ is denoted as

$$p_{\text{Win}}(\mathcal{G},\mathcal{S}) = \frac{1}{|\Theta|}\sum_\theta \text{Tr}\left[\Pi^\theta \rho_{ABC}\right] \qquad\qquad \Pi^\theta = \sum_{x\in\mathcal{X}} A_x^\theta \otimes B_x^\theta \otimes C_x^\theta$$

The optimal winning probability is

$$p_{\text{Win}}(\mathcal{G}) = \sup_{\mathcal{S}} p_{\text{Win}}(\mathcal{G},\mathcal{S})$$

Analysis

Before moving into the analysis, some more tools from linear algebra will be handy:

Definition 6.10 (Schatten ∞-Norm)**.** The Schatten ∞-norm of a operator A is defined as its largest singular value

$$\|A\| = \|A\|_\infty = s_1(A)$$

Properties:

1. $\|L\|^2 = \left\|L^\dagger L\right\| = \left\|LL^\dagger\right\|$

   Proof. Let the SVD of L be $L = U\Sigma V^\dagger$. Then $LL^\dagger = U\Sigma^2 U^\dagger$ ■

2. For PSD A, B, the singular values coincide with the eigenvalues and $A \leq B \implies \|A\| \leq \|B\|$
3. For PSD M and density matrix ρ, $\mathrm{Tr}(M\rho) \leq \|M\|$

   Proof. Let $M = \sum_i \lambda_i |i\rangle\langle i|, \rho = \sum_{ij} c_{ij} |i\rangle \langle j|$. Then

   $$\mathrm{Tr}(M\rho) = \mathrm{Tr}\left(\sum_{ij} \lambda_i c_{ij} |i\rangle \langle j|\right) = \sum_i \lambda_i c_{ii} \leq \lambda_{\max} \sum_i c_{ii} = \lambda_{\max}$$

   Infact, taking ρ to be the eigenvector corresponding to the maximum eigenvalue, we attain the bound. Thus we can write

   $$\|M\| = \max_\rho \mathrm{Tr}(M\rho)$$

   ■

4. If U is a unitary operator then $\|UA\| = \|AU\| = \|A\|$. In general, we never want any operator norm to be effected by multiplication by unitaries, since they only rotate or reflect the operator.
5. $\|A \oplus B\| = \max\{\|A\|, \|B\|\}$
6. $\|A \otimes B\| = \|A\| \|B\|$
7.

Lemma 6.3. Let A, B, L be linear operators in a Hilbert space $\mathcal{H}$ such that $A^\dagger A \geq B^\dagger B$. The $\|AL\| \geq \|BL\|$

Proof.

$$\begin{aligned} A^\dagger A \geq B^\dagger B &\implies L^\dagger A^\dagger A L \geq L^\dagger B^\dagger B L \\ &\implies \left\|L^\dagger A^\dagger A L\right\| \geq \left\|L^\dagger B^\dagger B L\right\| \\ &\implies \|AL\|^2 \geq \|BL\|^2 \\ &\implies \|AL\| \geq \|BL\| \end{aligned}$$

■

8. Using the above property if A, A', B, B' are PSD operators satisfying $A \geq A', B \geq B'$ then

   $$\left\|\sqrt{A}\sqrt{B}\right\| \geq \left\|\sqrt{A'}\sqrt{B}\right\| \geq \left\|\sqrt{A'}\sqrt{B'}\right\|$$

   In particular for projectors $\|AB\| \geq \left\|A'B'\right\|$.
9. We call two permutations $\pi_1, \pi_2 : [N] \to [N]$ to be orthogonal if $\forall i \in [N], \pi_1(i) \neq \pi_2(i)$ (Note that there must exist a set of N of them corresponding to the cyclic shifts, and this is maximal by PHP).

Lemma 6.4. Let $A_1 \dots A_N$ be PSD operators and let $\{\pi^k\}_{k\in[N]}$ be a set of N mutually orthogonal permutations of $[N]$. Then

$$\left\|\sum_{i=1}^N A_i\right\| \leq \sum_{k=1}^N \max_{i\in[N]} \left\|\sqrt{A_i}\sqrt{A_{\pi^k(i)}}\right\|$$

Proof. Define X as $(X)_{ij} = \delta_{j1}\sqrt{A_i}$

$$X = \begin{pmatrix} \sqrt{A_1} & 0 & \dots & 0 \\ \sqrt{A_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{A_N} & 0 & \dots & 0 \end{pmatrix}$$

$$X^\dagger X = \begin{pmatrix} \sum_i A_i & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \qquad XX^\dagger = \begin{pmatrix} A_1 & \sqrt{A_1}\sqrt{A_2} & \dots & \sqrt{A_1}\sqrt{A_N} \\ \sqrt{A_2}\sqrt{A_1} & A_2 & \dots & \sqrt{A_2}\sqrt{A_N} \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{A_N}\sqrt{A_1} & \sqrt{A_N}\sqrt{A_2} & \dots & A_N \end{pmatrix}$$

Using property 1

$$\left\|X^\dagger X\right\| = \left\|\sum_i A_i\right\| = \left\|XX^\dagger\right\|$$

Define matrices

$$(D_k)_{ij} = \delta_{j\pi^k(i)}\sqrt{A_i}\sqrt{A_j}$$

Then we can decompose $XX^\dagger = \sum_k D_k$, and using the triangle inequality $\left\|XX^\dagger\right\| \leq \sum_k \|D_k\|$. Now each D_k has only one non-zero block in each row and column. Hence using permutation matrices, we can convert them into block diagonal form:

$$D_k' = \bigoplus_{i\in[N]} \sqrt{A_i}\sqrt{A_{\pi^k(i)}}$$

Using property of direct sum, we get the required result. ■

Theorem 6.5. For any $n \in \mathbb{N}$

$$p_{\text{Win}}(\mathcal{G}_{\text{BB84}}^{\times n}) = \left(\frac{1}{2} + \frac{1}{2\sqrt{2}}\right)^n$$

Proof. Consider the state $|\phi\rangle := \cos\frac{\pi}{8}|0\rangle + \sin\frac{\pi}{8}|1\rangle$ ■

From Fig. 5, observe that $\langle +|\phi\rangle = \langle 0|\phi\rangle = \cos\frac{\pi}{8}$, which gives a winning probability of $\cos^2\frac{\pi}{8}$. Using this state for all the repetitions, Bob and Charlie can win with the above probability. Thus,

$$p_{\text{Win}}(\mathcal{G}_{\text{BB84}}^{\times n}) \geq \left(\frac{1}{2} + \frac{1}{2\sqrt{2}}\right)^n$$

Now we show that this is optimal. For any strategy $\mathcal{S}$

$$p_{\text{Win}}(\mathcal{G}, \mathcal{S}) = \frac{1}{2^n}\operatorname{Tr}\left[\sum_\theta \Pi^\theta \rho_{ABC}\right] \leq \frac{1}{2^n}\left\|\sum_\theta \Pi^\theta\right\| \leq \frac{1}{2^n}\sum_k \max_\theta \left\|\Pi^\theta \Pi^{\pi^k(\theta)}\right\|$$

Let $\mathcal{T}$ be the set of indices where θ and θ' differ, $\bar{\mathcal{T}}$ be its complement. Let $t = h(\theta, \theta')$ be the hamming distance between θ and θ' Define

$$P = \sum_x \left|x_{\mathcal{T}}^\theta\right\rangle\!\left\langle x_{\mathcal{T}}^\theta\right| \otimes \mathbb{I}_{\bar{\mathcal{T}}} \otimes B_x^\theta \otimes \mathbb{I}_C$$

$$Q = \sum_x \left|x_{\mathcal{T}}^{\theta'}\right\rangle\!\left\langle x_{\mathcal{T}}^{\theta'}\right| \otimes \mathbb{I}_{\bar{\mathcal{T}}} \otimes \mathbb{I}_B \otimes C_x^{\theta'}$$

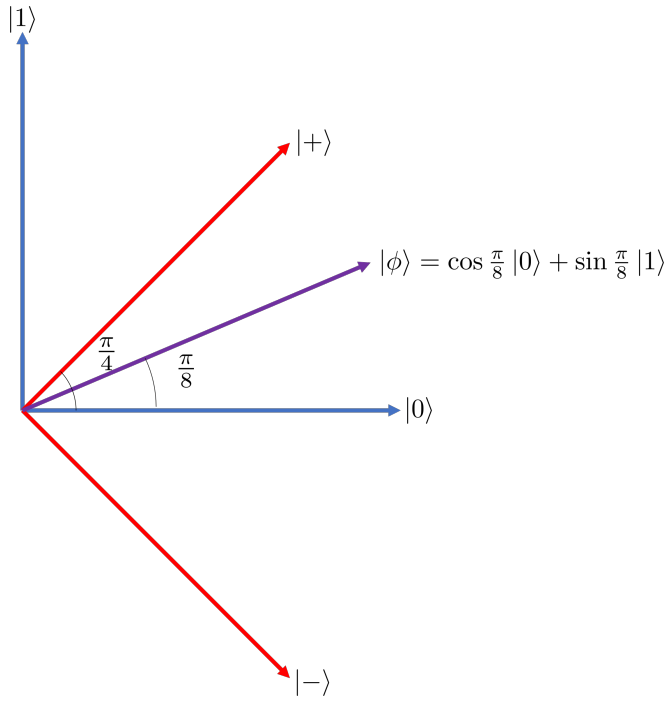


Figure 5: The states

Observe that $\Pi^\theta \leq P, \Pi^{\theta'} \leq Q$ so that $\left\|\Pi^\theta \Pi^{\theta'}\right\|^2 \leq \|PQ\|^2 \leq \|PQP\|$

$$\begin{aligned}
PQP &= \sum_{xyz} \left|x_{\mathcal{T}}^\theta\right\rangle\!\left\langle x_{\mathcal{T}}^\theta\right| \left|y_{\mathcal{T}}^{\theta'}\right\rangle\!\left\langle y_{\mathcal{T}}^{\theta'}\right| \left|z_{\mathcal{T}}^\theta\right\rangle\!\left\langle z_{\mathcal{T}}^\theta\right| \otimes \mathbb{I}_{\bar{\mathcal{T}}} \otimes P_x^\theta P_z^\theta \otimes Q_y^{\theta'} \\
&= \sum_{xyz} \left|x_{\mathcal{T}}^\theta\right\rangle\!\left\langle x_{\mathcal{T}}^\theta\right| \left|y_{\mathcal{T}}^{\theta'}\right\rangle\!\left\langle y_{\mathcal{T}}^{\theta'}\right| \left|z_{\mathcal{T}}^\theta\right\rangle\!\left\langle z_{\mathcal{T}}^\theta\right| \otimes \mathbb{I}_{\bar{\mathcal{T}}} \otimes \delta_{xz} P_x^\theta \otimes Q_y^{\theta'} \\
&= \sum_{xy} \left\|\left\langle x_{\mathcal{T}}^\theta \middle| y_{\mathcal{T}}^{\theta'}\right\rangle\right\|^2 \left|x_{\mathcal{T}}^\theta\right\rangle\!\left\langle x_{\mathcal{T}}^\theta\right| \otimes \mathbb{I}_{\bar{\mathcal{T}}} \otimes P_x^\theta \otimes Q_y^{\theta'} \\
&= 2^{-t} \sum_{xy} \left|x_{\mathcal{T}}^\theta\right\rangle\!\left\langle x_{\mathcal{T}}^\theta\right| \otimes \mathbb{I}_{\bar{\mathcal{T}}} \otimes P_x^\theta \otimes Q_y^{\theta'} \\
&= 2^{-t} \sum_{x} \left|x_{\mathcal{T}}^\theta\right\rangle\!\left\langle x_{\mathcal{T}}^\theta\right| \otimes \mathbb{I}_{\bar{\mathcal{T}}} \otimes P_x^\theta \otimes \mathbb{I}_C
\end{aligned}$$

Where the third step is because $x_{\mathcal{T}}^\theta$ and $y_{\mathcal{T}}^{\theta'}$ will be diagonal. Since the sum itself is a projector, $\|PQP\| \leq 2^{-t}$. Thus $\left\|\Pi^\theta \Pi^{\theta'}\right\| \leq 2^{-t/2}$. Now

$$\frac{1}{2^n} \sum_k \max_\theta \left\|\Pi^\theta \Pi^{\pi^k(\theta)}\right\| \leq \frac{1}{2^n} \sum_k \max_\theta \frac{1}{2^{h(\theta, \pi^k(\theta))/2}}$$

We need to find the permutations which minimize this hamming distance. Since the permutations must be orthogonal, the optimal permutations are given when $h(\theta, \pi^k(\theta))$ is the same for all θ, taking $\pi^k(\theta) = \theta \oplus k$. There are $\binom{n}{t}$ permutations with hamming weight t. So,

$$\frac{1}{2^n} \sum_{t=0}^{n} \binom{n}{t} \frac{1}{2^{t/2}} = \left(\frac{1}{2} + \frac{1}{2\sqrt{2}}\right)^n$$

Which gives the desired result.

§7. Privacy Amplification

Privacy amplification is generally the last task in any cryptographic protocol. It is useful because it helps us reduce the task of generating a key which is uniformly random and uncorrelated with the evaesdropper to the weaker task of generating a key which has some amount of uncertainity. Assume that Alice and Bob wish to agree on a secret random bitstring X (their secret key) and have at their disposal a perfect (and authenticated [2]) public channel and an imperfect private channel.
As Eve can listen to the communication between Alice and Bob, she can keep some side information E which can be a quantum state in general.
Input: State ρ_{XE} with $H_{\min}(X|E) \geq k$. If Eve has a direct copy of X then nothing can be done. So, there must be some amount of uncertainity in the initial state characterized by the min-entropy.
Goal: Generate random bitstrings R_A and R_B (with Alice and Bob respectively) such that the following are satisfied:

1. ***Correctness:***
$$\Pr[R_A \neq R_B] \leq \epsilon_c$$
2. ***Security:***
$$\left\| \rho_{R_A KE} - \frac{\mathbb{I}_{R_A}}{2^{-|R_A|}} \otimes \rho_{KE} \right\|_{\text{tr}} \leq \epsilon_s$$
where K is the side information which Eve may gain using the public communication between Alice and Bob

Note that the public communication between Alice and Bob is making our lives a bit harder. Is it really necessary? Indeed it is! Otherwise, if Alice and Bob create a random key out of X then the function f that they use to generate $R_A = f(X)$ must be deterministic. We assume that all deterministic functions are publicly known, so Eve too knows it. Now Eve can just keep the first bit of $f(X)$ and this renders the secret keys no longer random!

Example. *(a) Suppose $X \in \{0,1\}^3$ is uniformly distributed and $E = X_1 \oplus X_2$. Give a protocol for privacy amplification that outputs two secure bits (without any communication).*

Output $X_1 \oplus X_3$ and $X_2 \oplus X_3$. This is secure because Eve has no information about X_3 and hence no information about the output. Since X_3 is uniformly random, both the output bits will be uniformly random. Note that here if Eve even had X_1, X_2 still she won't be able to figure out the output bits.

(b) What if $E = (X_1 \oplus X_2; X_2 \oplus X_3) \in \{0,1\}^2$, can you still do it? If not, give a protocol extracting just one bit.

(c) Suppose the eavesdropper is allowed to keep any two of the bits of X as side information. Give a protocol for Alice and Bob to produce a Z which contains a single bit that is always uniformly random, irrespective of which two bits of X are stored by the eavesdropper.

Output $b = X_1 \oplus X_2 \oplus X_3$. WLOG suppose Eve knows X_1 and X_2. Then $\Pr[b = 0|X_1, X_2] = 1/2$

(d) How about an R that contains two bits — can they do it?

7.1 Randomness Sources

7.1.1 IID Sources

A classical iid source $X \in \{0,1\}^n$ has a distribution $\{p_x\}$ which has a product form: there is a distribution $\{p_0, p_1\}$ on a single bit such that for all $(x_1, \ldots, x_n) \in \{0,1\}^n$,

$$\Pr[X = (x_1, x_2 \ldots, x_n)] = \Pr[X_1 = x_1] \Pr[X_2 = x_2] \ldots \Pr[X_n = x_n] = p_{x_1} p_{x_2} \ldots p_{x_n}$$

[2] meaning that when Alice(or Bob) receives a message she has the guarantee that it came directly from Bob (or Alice). We will also assume the channel is noiseless, which in practice is ensured by a proper use of error-correcting codes.

In other words, all the constituent X_i's are Bernoulli random variables. Such sources are also called **Von Neumann** sources. Let us consider an example of a Von Newmann source with $p_0 = 1/4$ and let $Z = X_1 \oplus X_2 \cdots \oplus X_n \in \{0,1\}^n$. For n=2,

$$\Pr[Z = 0] = p_0^2 + p_1^2 = 0.625$$
$$\Pr[Z = 1] = 0.375$$

This is not as good as uniform but still better! For general n:

$$\Pr[Z = 0] = \sum_{r \text{ is even}} \binom{n}{r} p_0^r p_1^{n-r} = \frac{(p_0+p_1)^n + (p_0-p_1)^n}{2} = \frac{1}{2} + (-1)^n \frac{1}{2^{n+1}}$$

$$\Pr[Z = 1] = \frac{1}{2} - (-1)^n \frac{1}{2^{n+1}}$$

Which tends to the uniform distribution as n becomes very large. The trace distance with the uniform distribution is $(0.5)^{n+1} \leq \epsilon$ for $n = O(\log(1/\epsilon))$

7.1.2 Independent Bit Sources

Here, all bits are chosen independently but different bits may be chosen from different distributions. It turns out that taking the pairity of bits results in a bit that is close to unifrom as $n \to \infty$.
Let X be an independent n-bit source such that $\delta < \Pr\left[X_j = 0\right] < 1 - \delta$ for some $1/2 > \delta > 0$ and all $j \in \{1, \ldots, n\}$. Then for $Z = X_1 \oplus X_2 \cdots \oplus X_n \in \{0,1\}^n$

$$2^{n-1}\delta^n < \Pr[Z = 0] < 2^{n-1}(1-\delta)^n$$
$$2^{n-1}\delta^n < \Pr[Z = 1] < 2^{n-1}(1-\delta)^n$$

Then the trace distance of Z from the uniform distribution is:

$$\frac{1}{2}\left[|\Pr[Z=0] - \frac{1}{2}| + |\Pr[Z=1] - \frac{1}{2}|\right] = \frac{1}{2}|\Pr[Z=0] - \Pr[Z=1]| < 2^{n-2}|(1-\delta)^n - \delta^n|$$

7.1.3 Bit Fixing Sources

Bit-fixing sources are special cases of independent sources where some of the bits are fixed and the remaining are uniformly random. For example $\Pr\left[X = (1,0,1)\right] = \Pr\left[X = (1,1,1)\right] = \frac{1}{2}$ is a bit-fixing source where first and third bits are fixed to 1 and the second bit is uniformly random.
Just like the above two cases, here too the pairity of all the bits is a uniformly random bit. In fact in this case, as long as one of the bits is not fixed, we will get a uniformly random pairity bit. (This can be easily proved by induction)

7.1.4 General Sources

In general, the sources may produce strings in which all bits are independent of the other. For example, consider an *adverserial bit fixing source*, where the fixed bits can depend on the bits before them, $\Pr\left[X = (1,0,0)\right] = \Pr\left[X = (1,1,1)\right] = \frac{1}{2}$ is a bit-fixing source where first bit is fixed, the second bit is uniformly random and the third bit is equal to the second bit. Now, taking Z as the pairity of all the bits doesn't help. For any fixed choice of a subset of bits, there exists an adversarial bit-fixing source such that only one bit is fixed, but nevertheless the parity of the bits in the chosen subset is a constant (fix the last bit to be the pairity of the random bits before it).

Remark. In all the examples we have seen above, the function Ext is deterministic, i.e, it doesn't introduce any additional randomness over that already present in X. These extractors are called *deterministic extractors*.

Definition 7.1 (K-source)**.** A source X is called a K source if $H_{\min}(X) \geq K$. A cq-state ρ_{XE} is called a K source if $H_{\min}(X|E)_\rho \geq K$

Amplifying a weak secret

Let's say that Alice and Bob have a shared a uniformly random key X of n bits amoung themselves. However, Eve has a copy of pn bits of X for $0 < p < 1$. For instance,

$$X = X_1 X_2 \dots X_n$$

$$E = \perp X_2 \perp \dots X_n$$

In other words, we can say that Eve has access to a set of np equations $\{X \cdot e_{i_k}\}_{k \in [np]}$
Alice and Bob can generate a random key using

$$R = X_1 \oplus X_2 \cdots \oplus X_n$$

This will be perfectly random and uncorrelated with Eve as long as $p < 1$. However, this contains only one bit. How can we amplify this to get more bits?
As we have seen before, we need to incorporate randomness in order to do better. Alice generates $Y_1, Y_2, \dots, Y_m \in \{0,1\}^n$ at random passes them to Bob via the public channel. Now both of them generate $R_i = Y_i \cdot X$ and output $R = R_1 R_2 \dots R_m \in \{0,1\}^m$.
Even if we assume that Eve is aware of $Y_i \cdot X$ for $i \in [m-1]$, X can be any of the $2^{n-np-(m-1)}$ vectors, as Eve has a total of $m - 1 + np$ equations to determine X. Thus, as long as $m < n(1-p)$, we can extract m bits of randomness from X.

7.2 Randomness Extractors

The topic of randomness extractors is crucial in several domains of theoretical computer science. In particular, they are used in the design of pseudorandom generators, randomness-efficient algorithms, and in the study of randomness extractors themselves. Apart from cryptography, they are also used in graph theory, combinatorics, etc.

A randomness extractor is a function Ext which takes a random variable $X \in \{0,1\}^n$ such that $H_{\min}(X|E) \geq K$ and a uniformly random seed $Y \in \{0,1\}^d$ as input and outputs a random variable $\text{Ext}(X,Y) = Z \in \{0,1\}^m$ which is close to uniform: $\rho_{ZE} \approx 2^{-m} \mathbb{I} \otimes \rho_E$. In other words, it is a function which takes a *weak source* of randomness and outputs a *strong source* of randomness. We will be looking at extractors in which Eve also has access to the random seed Y. This is called a *strong extractor*:

$$\rho_{ZYE} \approx 2^{-m} \mathbb{I} \otimes \rho_{YE}$$

We also have *weak extractors* where Eve doesn't know about Y.

Observe that the task of randomness extraction is similar to that of privacy amplification. Alice generates a uniformly random bitstring Y (this is uniform and independent from Eve) in her lab and passes it to Bob via the public channel. Then they use $\text{Ext}(X,Y)$ to generate their secret key Z. This protocol will be correct ($\epsilon_c = 0$) and its security requirement

$$\left\| \rho_{R_A KE} - 2^{-|R_A|} \mathbb{I}_{R_A} \otimes \rho_{KE} \right\|_{\text{tr}} \leq \epsilon_s$$

will be the correctness requirement for the randomness extractor.

Definition 7.2 (Strong Seeded Extractors)**.** $\text{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (K, ϵ) strong seeded extractor if

$$H_{\min}(X|E) \geq K \qquad \left\| \rho_{\text{Ext}(X,Y)YE} - 2^{-m} \mathbb{I} \otimes \rho_{YE} \right\|_{\text{tr}} \leq \epsilon$$

Thw goals of the extractor are:

1. **Maximize m**: The more randomness we can extract, the better. Eve has at least K bits of uncertainity about X.

$$m \approx K$$

2. **Minimize d**: The seed should be as short as possible, so that it can be efficiently transmitted in the protocol.

$$d \approx \log(n/\epsilon)$$

Remark. **Min-entropy is the upper bound on the extractable randomness** No strong extractor can have $m > H_{\min}(X|E)$. Suppose $Ext(X|y) = f(X)$. Then the probability that we can guess $f(X)$ will be upper bounded by the probability that we can guess X. Thus

$$H_{\min}(f(X)|E) \leq H_{\min}(X|E)$$

This implies that output of $\text{Ext}(X, Y)$ conditioned on y can be uniformly random on at most $H_{\min}(X|E)$ bits.
Let us look at the example we discussed above.

$$X : n \text{ bits}$$
$$E : pn \text{ bits}$$
$$Y = (Y_1, Y_2 \ldots, Y_m) : nm \text{ uniformly random bits}$$
$$\text{Ext}(X, Y) = (X \cdot Y_1, X \cdot Y_2, \ldots, X \cdot Y_m) \in \{0,1\}^m$$

Observe that $H_{\min}(X|E) = -\log\left(1/2^{n-np}\right) = n(1-p)$, i.e, the number of bits of n which Eve doesn't know about. We will call these the *free bits*. We want to extract $m = n(1-p)$ bits of randomness.
The function Ext can be considered as a matrix multiplication over the binary field in which Y_i's form the rows of the matrix A and X is the column vector. Hence, we can write

$$\text{Ext}(X, Y) = \begin{bmatrix} & & \\ & Y_1 & \\ & Y_2 & \\ & \vdots & \\ & Y_m & \\ & & \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ X_n \end{bmatrix}$$

For Eve, np bits of X are known, so we will call the corresponding np columns of the Y *fixed*. If

$$\dim(\text{span}(\text{Non-fixed colums of } X)) \geq m$$

then Eve will see a uniformly random distribution.
The above construction is essentially taking the parity of random subsets of bits from the source, where the subset is determined by the seed. The only problem with this extractor is that it takes quite a long seed.

A Combinatorial View of Extractors

Let us consider a combinatorial view of randomness extraction as shown in Fig. 6. We will be ignoring the side information E here. Since $H_{\min}(X)_\rho \geq K$, we have $\Pr_{\text{guess}}[X = x] \leq 2^{-K}$. As a first step, this can be seen as any subset of the domain of size 2^K in which X is distributed uniformly randomly.
Since the size of the domain is larger than that of the codomain, Ext must be a many one function. This gives another reason for Ext to use a random seed. If Ext were deterministic then we can choose a subset of all the domain points which map to the same point of the codomain as the subset - in that case Ext's range will be far from random. The random seed Y gives a random map (depicted by the different coloured lines)[3].

7.3 Universal Hash Functions

[3] we could even have taken Y to be the subset which we want to choose from $\{0,1\}^n$. But this will be too costly since we will need n bits of seed!

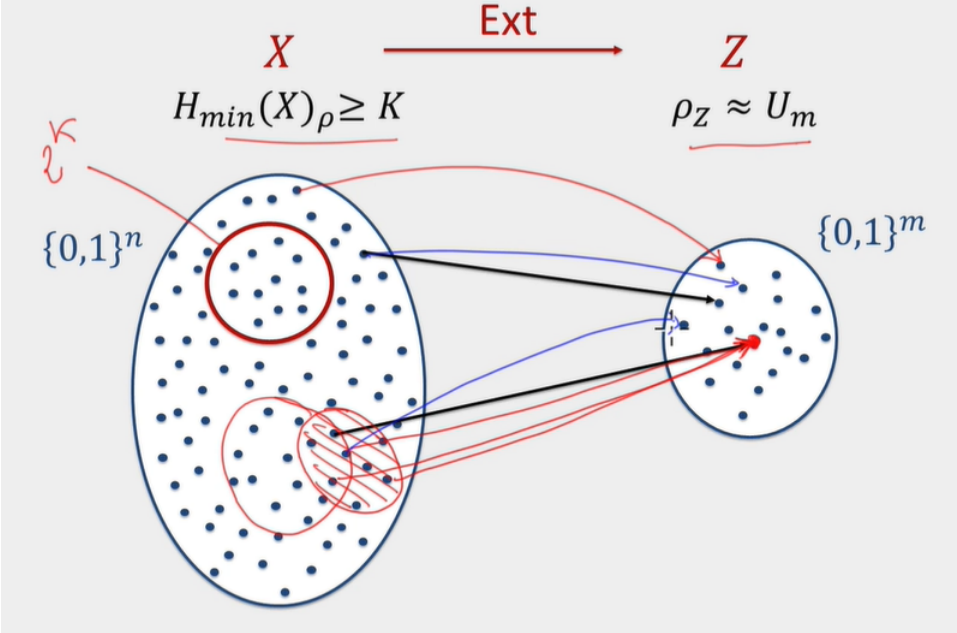


Figure 6: Combinatorial View of Extractors (Taken from the lecture video from [VW16])

Definition 7.3 (1-Universal Family)**.** A family of hash functions $\mathcal{H} = \{f_y : \{0,1\}^n \to \{0,1\}^m\}, m \leq n$ is called a 1-Universal family if for every x, z

$$\Pr_y[f_y(x) = z] = 2^{-m}$$

In other words, for fixed x and uniformly distributed Y $f_Y(x)$ has a uniform distribution.

For example, consider the family $\mathcal{H} = \{f_y : \{0,1\}^n \to \{0,1\}^n\}$ $f_y(x) = x \oplus y$ (bitwise XOR). Then since y is a uniformly random bit, $f_Y(x)$ will also be random. However, note that here we are utilizing the randomness of the seed Y instead of extracting it from X. Moreover, it is a weak seeded extractor, since with the information of y, the adversary can recover $x = y \oplus z$.
To obtain a strong seeded extractor, we need the following:

Definition 7.4 (2-Universal Family)**.** A family of hash functions $\mathcal{H} = \{f_y : \{0,1\}^n \to \{0,1\}^m\}, m \leq n$ is called a 2-Universal family if for every $x, x'(x \neq x'), z, z'$

$$\Pr_y[f_y(x) = z \wedge f_y(x') = z'] = 2^{-2m}$$

This condition would be satisfied if $f_y(x)$ and $f_y(x')$ were jointly chosen uniformly and independently at random in $\{0,1\}^m$. This is a stronger condition than that for 1-Universal family: we now require that the pair of random variables $(f_Y(x); f_Y(x'))$, for Y uniformly distributed, are jointly uniform.

The above family $f_y(x) = x \oplus y$ is not a 2-universal family:

$$\begin{aligned}\Pr_y[f_y(x) = z \wedge f_y(x') = z'] &= \Pr_y[x \oplus y = z \wedge x' \oplus y = z'] \\ &= \Pr_y[x \oplus y = z \wedge x' \oplus x = z' \oplus z'] \\ &= 0 \qquad\qquad \text{if } x' \oplus x \neq z' \oplus z'\end{aligned}$$

Consider $f_y(x) = x \cdot y \in \{0,1\}$

$$\Pr_y[f_y(x) = z \wedge f_y(x') = z'] = \Pr_y[x \cdot y = z \wedge x' \cdot y = z']$$

These can be thought as a system of linear equations

$$\begin{bmatrix} x \\ x' \end{bmatrix} \begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} z \\ z' \end{bmatrix}$$

And has 2^{n-2} solutions. So,

$$\Pr_y[x \cdot y = z \wedge x' \cdot y = z'] = \frac{2^{n-2}}{2^n} = \frac{1}{4}$$

Definition 7.5 (2-Universal Extractor)**.** Let $\mathcal{F} = \{f_y : \{0,1\}^n \to \{0,1\}^m, y \in \{0,1\}^d\}$ be a family of 2 uniform hash functions. Then the 2-universal extractor formed using these functions is $\text{Ext}_{\mathcal{F}} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$

$$\text{Ext}_{\mathcal{F}}(x, y) = f_y(x)$$

Definition 7.6 (Collision Probability)**.** The collision probability of a random variable X is the probability that two different samples from it are equal, defined as

$$\text{CP}(X) = \Pr_{x \leftarrow X, y \leftarrow X}[x = y] = \sum_x (\Pr[X = x])^2$$

7.4 Pretty Good Measurement

The topic of pretty good measurement is used in several areas of quantum computing. We will be using it for the proof of the leftover hash lemma. Consider the state $\sum_x |x\rangle\langle x| \otimes \rho_x^E$, where ρ_x^E are not normalized (their trace is p_x). What is the optimal probability that Eve succeeds in finding x?

$$\Pr[\text{Eve wins}] = \max_{M_x} \sum_x \text{tr}\left(M_x \rho_x^E\right)$$

For the case $x \in \{0,1\}$

$$\begin{aligned}
\Pr[\text{Eve wins}] &= \operatorname{tr}\left(M_0\rho_0^E\right) + \operatorname{tr}\left(M_1\rho_1^E\right) \\
&= \operatorname{tr}\left(\frac{\overbrace{(M_0+M_1)}^{\mathbb{I}}}{2}(\rho_0^E+\rho_1^E)\right) + \operatorname{tr}\left(\frac{(M_0-M_1)}{2}(\rho_0^E-\rho_1^E)\right) \\
&= \frac{1}{2} + \frac{1}{2}\operatorname{tr}\Big((M_0-M_1)(\rho_0^E-\rho_1^E)\Big) \\
&= \frac{1}{2} + \frac{1}{2}D(\rho_0^E,\rho_1^E)
\end{aligned}$$

However, for larger $|X|$, no closed form for the above expression is known. The pretty good measurement helps us to approximate the above.

Definition 7.7 (Pretty Good Measurement (PGM))**.** Given a collection of PSD matrices $\{\rho_x\}$, the PGM associated with the collection is

$$M_x = \rho^{-1/2}\rho_x\rho^{-1/2}$$

where $\rho = \sum_x \rho_x$ and the inverse is the Moore-Penrose pseudo-inverse, i.e. $0^{-1} = 0$

Relation between PGM and optimal guessing probability: Let $\{N_x\}$ be an optimal measurement for Eve. Then

$$\begin{aligned}
\Pr_{\text{Guess}}[X|E] &= \sum_x \operatorname{Tr}\left[N_x\rho_x^E\right] \\
&= \sum_x \operatorname{Tr}\left[(\rho^{1/4}N_x\rho^{1/4})(\rho^{-1/4}\rho_x^E\rho^{-1/4})\right] \\
&\le \sum_x \sqrt{\operatorname{Tr}\left[\rho^{1/2}N_x\rho^{1/2}N_x\right]}\sqrt{\operatorname{Tr}\left[\rho^{-1/2}\rho_x^E\rho^{-1/2}\rho_x^E\right]} \\
&\le \sqrt{\underbrace{\sum_x \operatorname{Tr}\left[\rho^{1/2}N_x\rho^{1/2}N_x\right]}_{\le 1}}\sqrt{\underbrace{\sum_x \operatorname{Tr}\left[\rho^{-1/2}\rho_x^E\rho^{-1/2}\rho_x^E\right]}_{\text{PGM}(X|E)}} \\
&\le \sqrt{\text{PGM}(X|E)}
\end{aligned}$$

Where step 3 follows from the Cauchy-Schwarz inequality for matrices:

$$\operatorname{Tr}[AB] \le \sqrt{\operatorname{Tr}\left[AA^\dagger\right]}\sqrt{\operatorname{Tr}\left[BB^\dagger\right]}$$

and step 4 is the normal Cauchy-Schwarz inequality. In the last step, to bound the first term, we use submultiplicativity of trace for PSD operators - $\operatorname{Tr}[AB] \le \operatorname{Tr}[A]\operatorname{Tr}[B]$ (to prove this, write their spectral decomposition):

$$\sum_x \operatorname{Tr}\left[\rho^{1/2}N_x\rho^{1/2}N_x\right] \le \sum_x \underbrace{\operatorname{Tr}\left[N_x\rho\right]}_{\le 1}\operatorname{Tr}[N_x] \le \operatorname{Tr}\left[\sum_x N_x\right] = 1$$

Thus we obtain that $\text{PGM}(X|E)$ is **at least the square** of the optimal guessing probability.

7.5 Leftover Hash Lemma

Lemma 7.1 (Leftover Hash Lemma)**.** Let n and $k \leq n$ be arbitrary integers, $\epsilon \geq 0$ and $m \leq K - 2\log(1/\epsilon)$. Let $\mathcal{F}$ be a family of 2-universal hash functions. Then $\text{Ext}_{\mathcal{F}}$ is a (K, ϵ)-strong seeded randomness extractor.

Proof. We analyze it in two cases: with and without side information:

Without Side Information

1. **From trace distance to collision probability :** We bound the trace distance for the extractor $Z = \text{Ext}(X, Y)$ utilizing a source with min-entropy K and seed $Y \approx U_d$.

$$\begin{aligned}
D(\rho_{ZY}, 2^{-m}\mathbb{I} \otimes 2^{-d}\mathbb{I}) &= \frac{1}{2}\sum_{y,z}\left|\underbrace{\Pr[Y=y, Z=z]}_{p_{yz}} - \frac{1}{2^{m+d}}\right| \\
&\leq \frac{1}{2}\sqrt{2^{m+d}\sum_{y,z}\left|p_{yz} - \frac{1}{2^{m+d}}\right|^2} \\
&= \frac{1}{2}\sqrt{2^{m+d}\sum_{y,z}\left|p_{yz} - \frac{1}{2^{m+d}}\right|^2} \\
&= \frac{1}{2}\sqrt{2^{m+d}\sum_{y,z}\left(p_{yz}^2 - \frac{2}{2^{m+d}}p_{yz} + \frac{1}{2^{2(m+d)}}\right)} \\
&= \frac{1}{2}\sqrt{2^{m+d}\left(\sum_{y,z}p_{yz}^2 - \frac{2}{2^{m+d}} + \frac{1}{2^{m+d}}\right)} \\
&= \frac{1}{2}\sqrt{2^{m+d}\left(\sum_{y,z}p_{yz}^2 - \frac{1}{2^{m+d}}\right)} \\
&= \frac{1}{2}\sqrt{2^{m+d}\left(\text{CP}(YZ) - \frac{1}{2^{m+d}}\right)}
\end{aligned}$$

Where step 2 follows from the Cauchy-Schwarz (or AM-RMS) inequality,

$$\boxed{\sum_{i=1}^{N} a_i \leq \sqrt{N\sum_{i=1}^{N} a_i^2}}$$

Below, we have derived the bound

$$\text{CP}(YZ) \leq \frac{1}{2^{d+m}} + \frac{1}{2^{d+K}}$$

Putting it in the equation, we obtain

$$D(\rho_{ZY}, 2^{-m}\mathbb{I} \otimes 2^{-d}\mathbb{I}) \leq \frac{1}{2^{1+(K-m)/2}} \leq \epsilon$$

Which gives $m \leq K - 2\log(1/\epsilon)$

2. **Bounding the collision probability** The collision probability for our extractor based on 2-universal hash function $X \to Z = f_Y(X)$, where $Y \approx U_d$ will be bounded as

$$\begin{aligned}
\mathrm{CP}(YZ) &= \sum_{y,z} \Pr\big[Y = y, Z = z\big]^2 \\
&= \sum_{y,z} \Pr\big[Y = y\big]^2 \Pr\big[Z = z | Y = y\big]^2 \\
&= \frac{1}{2^{2d}} \sum_{y,z} \Pr\big[Z = z | Y = y\big]^2 \\
&= \frac{1}{2^{2d}} \sum_{y,z} \left(\sum_{x : f_y(x) = z} p_x \right)^2 \\
&= \frac{1}{2^{2d}} \left(\underbrace{\sum_{y,z} \sum_{x \neq x' : f_y(x) = f_y(x') = z} p_x p_{x'}}_{A} + \underbrace{\sum_{y,z} \sum_{x : f_y(x) = z} p_x^2}_{B} \right)
\end{aligned}$$

To bound A, we use the property of 2-universal family that 2^{-m} fraction of the total functions map a given input to the same output.

$$\begin{aligned}
A &= \sum_{y,z} \sum_{x \neq x' : f_y(x) = f_y(x') = z} p_x p_{x'} \\
&= \sum_{y} \sum_{x \neq x' : f_y(x) = f_y(x')} p_x p_{x'} \\
&= \frac{2^d}{2^m} \sum_{x \neq x'} p_x p_{x'} \\
&\leq \frac{2^d}{2^m}
\end{aligned}$$

To bound B, we use $\sum_x p_x^2 \leq (\max_x p_x) \sum_x p_x = \max_x p_x = 2^{-K}$

$$\begin{aligned}
B &= \sum_{y,z} \sum_{x : f_y(x) = z} p_x^2 \\
&= \sum_{y} \sum_{x} p_x^2 \\
&= 2^d \sum_{x} p_x^2 \\
&\leq \frac{2^d}{2^K}
\end{aligned}$$

Putting it all together

$$\mathrm{CP}(YZ) \leq \frac{1}{2^{2d}} \left(\frac{2^d}{2^m} + \frac{2^d}{2^K} \right) = \frac{1}{2^{d+m}} + \frac{1}{2^{d+K}}$$

■

7.6 Privacy Amplification using Extractors

Let Ext be a (K, ϵ) strong seeded extractor. Then the protocol for privacy amplification is

1. Alice and Bob share a weak secret X which may be correlated with Eve.
2. Alice chooses a random string Y and computes $R_A = \text{Ext}(X, Y)$. She send Y to Bob.
3. Bob computes $R_B = \text{Ext}(X, Y)$.

Correctness: $R_A = R_B$ by construction
Security: $D(\rho_{R_A YE}, 2^{-m}\mathbb{I} \otimes \rho_{YE}) = D(\rho_{\text{Ext}(X,Y)YE}, 2^{-m}\mathbb{I} \otimes \rho_{YE}) \leq \epsilon$

§8. Quantum Key Distribution : The BB84 Protocol

8.1 Assumptions

We assume that the labs of Alice and Bob are the only parts of the world where Eve has no access, otherwise she could just copy the key generated at the end of the protocol. As usual, we desire two properties from the protocol:

- **Correctness:** $\Pr[K_A \neq K_B] \leq \epsilon_c$
- **Security:** $\rho_{K_A E} \approx_{\epsilon_s} U_{K_A} \otimes \rho_E$

To achieve such a key distribution protocol, we will consider the following communicationchannels which Alice and Bob may have access to:

1. **A classical channel:** Alice and Bob can send classical bits in either direction over this channel. Eve has complete access to this channel. In particular, she can read all messages, modify them, and even impersonate Alice (or Bob).
2. **A classical authenticated channel (CAC):** A classical communication channel with one extra guarantee: Alice and Bob know that the message originated unaltered from Alice or Bob respectively. This means that while the channel is not secret because Eve can still read all the messages that travel across, she cannot impersonate Alice or Bob or alter messages traveling over the channel.
3. **A classical secret channel:** A classical communication channel in which Eve cannot learn any information about the messages traveling across. Yet, while she cannot hope to gain any information about the message, Eve could impersonate Alice or Bob.
4. **A classical secret and authenticated channel:** A classical communication channel combining both guarantees above.
5. **A quantum communication channel:** A channel where Alice may send quantum information (in particular, in the form of qubits) to Bob, where Eve has full access to all the quantum communication

Remark.

- We can model the special classical channel to be a binary symmetric channel between Alice and Eve, where with probability p Eve gets the same bit which Alice sent. So, the min-entropy for one bit is $H_{\min}(X|E) = \log(1/p)$. For n bits, $H_{\min}(X|E) = n\log(1/p)$. From Theorem 7.1, we can extract $m = n\log(1/p) - 2\log(1/\epsilon)$ bits of the key.
- We may also model Eve with limited memory. Suppose she can store only m bits. Then by the chaining property of min-entropy, we have $H_{\min}(X|E) \geq H_{\min}(X) - \log(|E|) = n - m$.
- Till now, we assumed that the channel between Alice and Bob was perfect, meaning no errors could occur on this channel, but this assumption is too strong. In reality, Alice may need to share some error correcting information at the end of the protocol. As long as that information is not too large we will still have a large min-entropy at the source $H_{\min}(X|EC) \geq H_{\min}(X|E) - \log(|C|)$. This step is called **information reconcillation**, and must be performed before privacy amplification.

8.2 Noiseless BB84

First let us the protocol without noise, so that there are no errors induced in the messages shared between Alice and Bob. For $x, \theta \in \{0,1\}^n$ define the Wiesner state

$$\left|x^\theta\right\rangle = H^{\theta_1}|x_1\rangle \otimes H^{\theta_2}|x_2\rangle \cdots \otimes H^{\theta_n}|x_n\rangle$$

Set $N = 4n + \eta$ where $\eta << 1$ is a small positive real number.

1. Alice samples $\widetilde{x_A} \leftarrow \{0,1\}^N$ and $\theta_A \leftarrow \{0,1\}^N$. She sends $\left|\widetilde{x_A}^{\theta_A}\right\rangle$ to Bob over the quantum channel.

2. Bob samples $\theta_B \leftarrow \{0,1\}^N$ and measures $\left|\widetilde{x_A}^{\theta_A}\right\rangle$ in the H^{θ_B} basis to obtain $\widetilde{x_B}$.
3. Bob sends a recipt that he has received all the qubits over CAC.
4. **Synchronization:** Alice and Bob share θ_A, θ_B with each other over CAC. They retain only those indices of $\widetilde{x_A}, \widetilde{x_B}$ where $\theta_A = \theta_B$. This happens with probability $1/2$ and they thus have x_A and x_B of length $\approx 2n$ at the end of this step.
5. **Test:** Alice samples a random set [4] $T \subseteq [N]$ of indices and sends it to Bob over CAC. Then they share $x_{A,T}$ and $x_{B,T}$ with each other, where $x_{A,T}$ is x_A restricted to the index set T. If $x_{A,T} \neq x_{B,T}$ they abort the protocol.
6. **Privacy Amplification:** Alice sends a random bitstring $Y \leftarrow \{0,1\}^d$ to Bob and they perform Privacy Amplification to obtain $R_A = \text{Ext}(x_{A,S}, Y)$, $R_B = \text{Ext}(x_{B,S}, Y)$ where $S = [N]/T$ is the complement of T.

Why is step 3 necessary? If suppose, we remove the it, then Eve with a sufficiently large quantum memory can store the qubits sent in the first step. On receiving θ_A, she can measure the qubits to obtain the exact same x_A. Then she can re-create the state $\left|x_A^{\theta_A}\right\rangle$ and send it to Bob with θ. As Eve has an exact copy of x_A, the protocol cannot succeed. But due to the no-cloning theorem, she cannot keep a copy of the state and send a copy to Bob.

8.3 Noisy BB84

Our assumption of the absence of noise is extremely unrealistic. Moreover, even if $x_A \neq x_B$ at even one position, then with a very high probability the keys generated after privacy amplilfication will be unequal. Practically, Alice and Bob will need to perform information reconcillation before the privacy amplification step to ensure the strings $x_{A,S}$ and $x_{B,S}$ are equal with a high probability. So, we add that after step 5 of the protocol:

1. Alice samples $\widetilde{x_A} \leftarrow \{0,1\}^N$ and $\theta_A \leftarrow \{0,1\}^N$. She sends $\left|\widetilde{x_A}^{\theta_A}\right\rangle$ to Bob over the quantum channel.
2. Bob samples $\theta_B \leftarrow \{0,1\}^N$ and measures $\left|\widetilde{x_A}^{\theta_A}\right\rangle$ in the H^{θ_B} basis to obtain $\widetilde{x_B}$.
3. Bob sends a recipt that he has received all the qubits over CAC.
4. **Synchronization:** Alice and Bob share θ_A, θ_B with each other over CAC. They retain only those indices of $\widetilde{x_A}, \widetilde{x_B}$ where $\theta_A = \theta_B$. This happens with probability $1/2$ and they thus have x_A and x_B of length $\approx 2n$ at the end of this step.
5. **Test:** Alice samples a random set $T \subseteq [N]$ of indices and sends it to Bob over CAC. Then they share $x_{A,T}$ and $x_{B,T}$ with each other, where $x_{A,T}$ is x_A restricted to the index set T.
6. **Information Reconcillation:** Alice and Bob calculate the error rate δ

$$\delta = \frac{|\{i : x_A[i] = x_B[i], i \in [T]\}|}{|T|}$$

   If δ is above a certain threshold, they abort the protocol. Otherwise, they exchange some error correcting information C accross CAC to obtain $\widehat{x_{A,S}}, \widehat{x_{B,S}}$ from $x_{A,S}$ and $x_{B,S}$ where $S = [N]/T$ is the complement of T.
7. **Privacy Amplification:** Alice sends a random bitstring $Y \leftarrow \{0,1\}^d$ to Bob and they perform Privacy Amplification to obtain $R_A = \text{Ext}(\widehat{x_{A,S}}, Y)$, $R_B = \text{Ext}(\widehat{x_{B,S}}, Y)$.

[4] where each element is chosen with probability $1/2$

8.4 Security Analysis

8.4.1 Purified BB84

We consider a purified version of the protocol for the security analysis. Instead of sending the states $|x_A\rangle^{\theta_A}$, Alice prepares the EPR pairs $|\phi^+\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$ and sends one qubit of the pair to Bob. Then she measures each qubit of her part of the pair in the standard or the hadamard basis. The corresponding part with Bob collapses to the same state that is measured by Alice. Thus this is equivalent to the original protocol. This helps us in the security analysis because if Alice and Bob were able to test for the presence of entanglement between their qubits, then (intuitively) by the monogamy of entanglement they would be able to certify that their systems are uncorrelated with Eve's. Here is the Purified BB84 protocol:

1. Alice prepares the states $|\phi^+\rangle^{\otimes n}$ and sends one qubit of each bell pair to Bob. Then she samples $\theta_A \leftarrow \{0,1\}^N$ and measures her qubits in the basis determined by θ_A to obtain x_A.
2. Bob samples $\theta_B \leftarrow \{0,1\}^N$ and measures $\left|\widetilde{x_A}^{\theta_A}\right\rangle$ in the H^{θ_B} basis to obtain $\widetilde{x_B}$.
3. Bob sends a recipt that he has received all the qubits over CAC.
4. **Synchronization:** Alice and Bob share θ_A, θ_B with each other over CAC. They retain only those indices of $\widetilde{x_A}, \widetilde{x_B}$ where $\theta_A = \theta_B$. This happens with probability $1/2$ and they thus have x_A and x_B of length $\approx 2n$ at the end of this step.
5. **Test:** Alice samples a random set $T \subseteq [N]$ of indices and sends it to Bob over CAC. Then they share $x_{A,T}$ and $x_{B,T}$ with each other, where $x_{A,T}$ is x_A restricted to the index set T.
6. **Information Reconcillation:** Alice and Bob calculate the error rate δ
$$\delta = \frac{|\{i : x_A[i] = x_B[i], i \in [T]\}|}{|T|}$$
If δ is above a certain threshold, they abort the protocol. Otherwise, they exchange some error correcting information C accross CAC to obtain $\widehat{x_{A,S}}, \widehat{x_{B,S}}$ from $x_{A,S}$ and $x_{B,S}$ where $S = [N]/T$ is the complement of T.
7. **Privacy Amplification:** Alice sends a random bitstring $Y \leftarrow \{0,1\}^d$ to Bob and they perform Privacy Amplification to obtain $R_A = \text{Ext}(\widehat{x_{A,S}}, Y)$, $R_B = \text{Ext}(\widehat{x_{B,S}}, Y)$.

8.4.2 More power to Eve

Modeling the power of Eve is generally challenging because she can manipulate the qubits in the quantum channel in any way she chooses. To address this, we modify the protocol to assume that Eve prepares a pure state ρ_{ABE} and shares it with Alice and Bob. Alice and Bob then perform a 'test' to verify that the state they received from Eve is indeed an EPR pair. This verification involves performing the projective measurement $|\phi^+\rangle\langle\phi^+|$.

0. Upon receiving their N respective qubits from Eve, Alice and Bob jointly measure each pair of qubits using the two-outcome PVM $\{|\phi^+\rangle\langle\phi^+|_{AB}, \mathbb{I}_{AB} - |\phi^+\rangle\langle\phi^+|_{AB}\}$ where $|\phi^+\rangle_{AB}$ denotes the EPR pair on Alice and Bob's joint system. If the number of pairs of qubits that were not found to equal $|\phi^+\rangle_{AB}$is larger than δn they abort.
1. Alice prepares the states $|\phi^+\rangle^{\otimes n}$ and sends one qubit to Bob. Then she samples $\theta_A \leftarrow \{0,1\}^N$ and measures her qubits in the basis determined by θ_A to obtain x_A.
2. Bob samples $\theta_B \leftarrow \{0,1\}^N$ and measures $\left|\widetilde{x_A}^{\theta_A}\right\rangle$ in the H^{θ_B} basis to obtain $\widetilde{x_B}$.
3. Bob sends a recipt that he has received all the qubits over CAC.
4. **Synchronization:** Alice and Bob share θ_A, θ_B with each other over CAC. They retain only those indices of $\widetilde{x_A}, \widetilde{x_B}$ where $\theta_A = \theta_B$. This happens with probability $1/2$ and they thus have x_A and x_B of length $\approx 2n$ at the end of this step.

5. **Test:** Alice samples a random set $T \subseteq [N]$ of indices and sends it to Bob over CAC. Then they share $x_{A,T}$ and $x_{B,T}$ with each other, where $x_{A,T}$ is x_A restricted to the index set T.
6. **Information Reconcillation:** Alice and Bob calculate the error rate δ

$$\delta = \frac{|\{i : x_A[i] = x_B[i], i \in [T]\}|}{|T|}$$

   If δ is above a certain threshold, they abort the protocol. Otherwise, they exchange some error correcting information C accross CAC to obtain $\widehat{x_{A,S}}, \widehat{x_{B,S}}$ from $x_{A,S}$ and $x_{B,S}$ where $S = [N]/T$ is the complement of T.
7. **Privacy Amplification:** Alice sends a random bitstring $Y \leftarrow \{0,1\}^d$ to Bob and they perform Privacy Amplification to obtain $R_A = \text{Ext}(\widehat{x_{A,S}}, Y)$, $R_B = \text{Ext}(\widehat{x_{B,S}}, Y)$.

The projective measurement appears to be an entangling measurement, raising the question of how Alice and Bob can implement it locally. We will now explore how they can achieve this using local measurements and classical communication.
Recall the 'Test' step in which they match some random indices of the two states. The probability that they measure the same outcome in the standard basis is $\text{Tr}\left[(|00\rangle\langle 00| + |11\rangle\langle 11|)\rho_{AB}\right]$ and that they measure the same outcome in hadamard basis is $\text{Tr}\left[(|{+}{+}\rangle\langle{+}{+}| + |{-}{-}\rangle\langle{-}{-}|)\rho_{AB}\right]$. Simple calculation shows that they are equivalent to measuring with the operators

$$\Pi_0 = \left|\phi^+\right\rangle\left\langle\phi^+\right| + |\Psi_{01}\rangle\langle\Psi_{01}|$$

$$\Pi_1 = \left|\phi^+\right\rangle\left\langle\phi^+\right| + |\Psi_{10}\rangle\langle\Psi_{10}|$$

So, this 'Test' phase is effectively equivalent to measuring with these 'entangled' projectors! To formally express the relation between the two steps, let $\rho_{A_jB_j}$ be a two qubit state, the j^{th} pair of qubits sent by Eve to Alice and Bob. Let p_j be the probability of passing the 'Test' step, averaged over the choice of a uniformly random (but identical for both A_j and B_j) basis in which to perform the test. Writing $\rho_{A_jB_j}$ in the Bell basis:

$$\rho_{A_jB_j} = \sum_{x,y\in\{0,1\}} q_{xy} \left|\Psi_{xy}\right\rangle\left\langle\Psi_{xy}\right|$$

Probability of winning in the test phase will be

$$p_j = \frac{1}{2}(\text{Tr}\left[\Pi_0\rho_{A_jB_j}\right] + \text{Tr}\left[\Pi_1\rho_{A_jB_j}\right])$$

so that $q_{00} \geq 2p_j - 1$

A concentration inequality

We just showed that if a state $\rho_{A_jB_j}$ passes the test phase, then with a high probability it was the EPR pair. But for generating the raw key, we use the bits other that those used in the test step - how do we make any guarantees for them? The key idea was that we chose the set T at random. For that case, we have the following inequality:

Theorem 8.1 ([TL17]). Let $m = n + k$ and consider m binary random variables $X_1, X_2 \ldots X_m$ where X_i may be arbitrarily correlated. Let T be a uniformly random subset of size k of $[m]$. Then for any $\delta > 0$

$$\Pr\left[\sum_{j\in T} X_j \leq \delta k \wedge \sum_{j\in[m]\setminus T} X_j \geq (\delta+\nu)n\right] \leq e^{-2\nu^2 \frac{nk^2}{(n+k)(k+1)}}$$

Assume for simplicity that the number of rounds in which Alice and Bob make the same basis choice is $R = 2n$ and $|T| = n$. For each $j \in R$, introduce the indicator random variable Z_j which is 1 if the measurement results are unequal and 0 otherwise. Using the above inequality for $m = 2n, k = n, \nu = \delta$ we have

$$\Pr\left[\sum_{j\in T} Z_j \leq \delta n \wedge \sum_{j\in[m]\setminus T} Z_j \geq 2\delta n\right] \leq e^{-\delta^2 n}$$

This bound implies that the probability that the test performed passes, but the outcomes obtained in the non-tested rounds do not match in a fraction larger than 2δ of these rounds, exponentially small in n. Denoting the event $\sum_{j\in T} Z_j > \delta n$ as ABORT we obtain

$$\Pr\left[\sum_{j\in[m]\setminus T} Z_j \geq 2\delta n \middle| \overline{\text{ABORT}}\right] \leq \frac{e^{-\delta^2 n}}{\Pr\left[\overline{\text{ABORT}}\right]}$$

Writing the bound in this way points to an important subtlety in how the security of QKD is defined - the probability of not aborting should not be too small. Otherwise the right hand side gets a blowup.

8.5 Authentication

In our protocol, it is always assumed that Alice and Bob have access to Classical Authenticated Channels (CACs). Practically, CACs are implemented using message authentication codes (MACs) or digital signatures, which are essentially unforgeable signatures attached to messages. However, for MACs to work, Alice and Bob must have a secret key established beforehand. This is a common requirement for all key-generation protocols: a (hopefully smaller) key must be available beforehand. QKD protocols offer a feature called **Everlasting Security**. This means that even if the authenticity of the channel is compromised after the key has been generated, the adversary gains no advantage. Consequently, authentication schemes based on computational assumptions can also be used; they only need to be secure for the duration of the protocol.

§9. Device Independent Quantum Key Distribution

DIQKD guarantees a stronger notion of security, where the state as well as the measurement device used by Alice and Bob are prepared by the adversary. Particularly, we make only the following assumptions:

1. Alice and Bob's labs are perfectly isolated: once the protocol begins, no information enters or exits the labs other than those pertaining to the protocol.
2. The random number generators used by Alice and Bob are perfect.
3. The devices used by Alice and Bob to perform the measurements are arbitrary.
4. At the end of the protocol, the devices used by Alice and Bob are discarded, so that Eve cannot obtain them.

9.1 The protocol

The protocol is based on the rigidity property of CHSH game discussed earlier Section 4.1.4. Note however that in the honest optimal strategy for the CHSH game, Alice and Bob never measure in the same basis. But for generating the key, we will prefer to measure in the same basis so that the outcomes are perfectly correlated. So, Alice's device gets 2 inputs (for measuring in Z or X basis) and Bob's device gets 3 inputs (for measuring in $H, \bar{H}$ or Z basis). When $(\theta_A, \theta_B) = (0, 2)$, both of them are expected to measure in the same basis and obtain the same outcome.

1. Alice chooses a uniformly random string $\theta_A \leftarrow \{0,1\}^n$. Bob chooses a uniformly random string $\theta_B \leftarrow \{0,1,2\}^n$. They measure their qubits in the corresponding basis, to obtain $x_A, x_B \in \{0,1\}^n$
2. Alice and Bob share their choice of the basis over the CAC.
3. Alice selects a random subset $T \in [n]$ of size $n/2$ and announces T to Bob. They set
$$T' = \{j \in T : \theta_B \in \{0,1\}\} \qquad T'' = \{j \in T : \theta_B = 2\} \qquad R = \{j \notin T : \theta_A = 0 \wedge \theta_B = 2\}$$
4. They share $x_{A,T}, x_{B,T}$ over CAC and calculate the following:
$$p_{\text{win}} = \frac{\{j \in T' : x_{A,T}[j] \oplus x_{B,T}[j] = \theta_A[j] \cdot \theta_B[j]\}}{|T'|}$$
$$p_{\text{match}} = \frac{\{j \in T'' : x_{A,T}[j] = x_{B,T}[j]\}}{|T''|}$$
If $p_{\text{win}} < \cos^2 \pi/8 - \delta$ or $p_{\text{match}} < 1 - \delta$, they abort.
5. Finally they perform Information Reconcillation and Privacy Amplification taking $x_{A,R}, x_{B,R}$ as the raw keys.

9.2 Security

Our goal is to show that there is an $\epsilon > 0$ (error) and $\kappa > 0$ (key-rate) such that for any strategy of Eve, specified by an initial state ρ_{ABE} of the devices and a choice of measurements to be made at every step in the protocol, either Alice and Bob abort with probability larger than ϵ or Alice's outcomes $x_{A,R}$ satisfy
$$H^{\epsilon}_{\min}(X_{A,R}|EK) \geq \kappa n$$
where K denotes all the communication exchanged on the CAC during the protocol. For analyzing, the proof, we study another guessing game based on CHSH.

9.2.1 CHSH-based Guessing Game

Consider a guessing game between Alice, Bob and Eve where Alice gets and input $\theta_A \in \{0,1\}$ and Bob gets an input $\theta_B \in \{0,1,2\}$. The players produce the outcome $x_A, x_B, z \in \{0,1\}$ respectively. They win iff the following hold:

1. $\theta_B \in \{0,1\}$ then $x_A \oplus x_B = \theta_A \cdot \theta_B$
2. $\theta_A = 0 \wedge \theta_B = 2$ then $z = x_A$

Lemma 9.1. Consider an arbitrary strategy for the players in CHSH guessing game. Let p_{win} be the probability that the first test passes and p_{id} be the probability that the second test passes. Suppose that $p_{\text{win}} \geq \cos^2 \pi/8 - \delta$. Then $p_{\text{id}} \leq \frac{1}{2} + 2\sqrt{\delta}$.

Proof. See [PAB+09] and [VV14] ■

9.2.2 Collective Attacks

In collective attacks, we assume that the initial state prepared by Eve is of the form $\rho_{ABE}^{\otimes n}$ and the devices are memoryless, that is they perform the same measurement every time. So, the behavior in each round is independent of that in other rounds.
Let $Z_1, Z_2 \ldots Z_k$ where $k = |T'|$ be the binary random variables such that $Z_i = 1$ if the CHSH condition is satisfied in that round. Then the empirical winning probability $p_{\text{win}} = \frac{1}{k}\sum_i Z_i$. Let the true quantity be $\hat{p}_{\text{win}}$. Recall the Chernoff bound:

Lemma 9.2 (Chernoff Bound)**.** Let $X_1 \ldots X_n$ be **iid Bernoulli random variables** with $\mathbb{E}[X_i] = \mu$. Then

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \alpha\mu\right] \leq 2e^{-\frac{\alpha^2\mu n}{3}}$$

First, we show that the size of $|T'|$ will be close to $n/6$ with high probability. Let X_i denote the binary random variable which is 1 if the i^{th} round is chosen for T'. Then $X \sim B(1, 1/6)$ and $\mu = 1/6$. Then $|T'| = \sum_{i=1}^{n} X_i$. Using the Chernoff bound with

$$\Pr\left[\left|\frac{|T'|}{n} - \mu\right| > \alpha\mu\right] \leq 2e^{-\frac{\alpha^2\mu n}{3}}$$

Setting $\alpha = 1/4$ we see that $|T'| < n/8$ with probability $\leq 2e^{-n/288}$. Assume that this is not the case. Then using the bound again in Z_j,

$$\Pr\left[\left|\frac{1}{|T'|}\sum_{j\in T'}^{n} Z_j - \hat{p}_{\text{win}}\right| > \alpha\hat{p}_{\text{win}}\right] \leq 2e^{-\frac{\alpha^2\hat{p}_{\text{win}}|T'|}{3}}$$

$$\implies \Pr\left[p_{\text{win}} > (1+\alpha)\hat{p}_{\text{win}}\right] \leq 2e^{-\frac{\alpha^2\hat{p}_{\text{win}}|T'|}{3}}$$

Using our lower bound on $|T'|$ and $\hat{p}_{\text{win}} \geq \cos^2 \pi/8$

$$\Pr\left[p_{\text{win}} > (1+\alpha)\hat{p}_{\text{win}}\right] \leq 2e^{-\frac{\alpha^2 n}{C}}$$

for some constant C.

Thus, except for probability exponentially small in n and given that the protocol doesn't abort, it must be that $\hat{p}_{\text{win}} \geq \frac{p_{\text{win}}}{1+\alpha} \geq \cos^2 \pi/8 - 2\delta$ setting $\alpha = \delta$. Using Theorem 9.1, $p_{\text{id}} \leq \frac{1}{2} + 2\sqrt{2\delta}$ which gives a direct bound on the guessing probability of the device

$$H_{\min}(X_j|E) \geq -\log\left(\frac{1}{2} + 2\sqrt{2\delta}\right) \geq 1 - C\sqrt{\delta}$$

for some small constant C. Finally, with the assumption that the device behaves independently and identically in all the rounds, we can extend the argument to the rounds which were not tested to get $H_{\min}(X_R|E) \geq |R|(1 - C\sqrt{\delta})$.

9.2.3 Coherent Attacks

Now, we move to the more general case where devices can have memory and Eve can prepare an arbitrary state. These result in the following limitations respectively:

1. We cannot directly infer properties of the devices in the rounds used for the raw key from its behavior in the testing rounds.
2. We can no longer claim that the min-entropy adds up across rounds, as there is no additivity or chain-rule for min-entropy.

The first difficulty can be handled by using a variant of the concentration bound (called **martingale inequalities**) that applies to processes which may have memory, but still a sequential nature and satisfy certain regularity properties.
There are two ways in which the second difficulty can be tackled:

1. **Quantum de-Finiti Theorem:** This theorem roughly states that if we have a channel $\Phi_{AB} : \rho_{A^nB^nE} \to (k_A, k_B)$ where we can randomly permote the rounds at the start of the protocol and randomly unpermute them at the end to obtain the same result (i.e. it is invariant under permutations), then $\epsilon-$security of Φ_{AB} on $\rho_{ABE}^{\otimes n}$ implies $\epsilon'-$security of Φ_{AB} on $\rho_{A^nB^nE}$. However, ϵ' depends on local dimension of A, B systems, so that dimension must be bounded. Thus, this theorem cannot be used to conclude security in the most general setting.
2. **Entropy Accumulation Theorem:** The EAT gives conditions under which min-entropy "accumulates", and these conditions are satisfied by our setup. The main conditions are that the outputs are generated sequentially in each round and are only a function of the state of the devices in that round. Moreover the test, when it is performed should be a deterministic function of the inputs and outputs in the round.

Finally, we conclude by stating the final result:

Theorem 9.3. For the DIQKD protocol, there is a $0 < \kappa < 1$ and $C \geq 1$(depending on δ) such that the following hold for $l = \kappa n$ and $\epsilon \leq 2^{-Cn}$:

1. There is an implementation of the devices such that the protocol does not abort with probability at least $1 - \epsilon$.
2. For any implementation of the devices, either the protocol aborts with probability larger than $1 - \epsilon$, or conditioned on not aborting Alice and Bob each produces a key of length l such that $\Pr[K_A \neq K_B] \leq \epsilon$ and $(1 - \Pr(\text{ABORT}))D(\rho_{K,E}, U_l \otimes \rho_E) \leq \epsilon$ where E is the side information of Eve at the end of the protocol.

§10. Multi-party Cryptography

Quantum Cryptographic protocols can be roughly divided into two categories:

1. The use of quantum information to implement classical tasks. Eg. QKD is used for distributing keys
2. The use of quantum information to implement genuinely quantum tasks, i.e. where atleast one of the inputs or outputs is quantum. Eg. Quantum secret sharing, where the goal is to distribute a qubit a qubit among several participants in such a way that at least a certain number of them need to come together to reconstruct the secret.

In this section, we will discuss some of these protocols.

10.1 Secure Function Evaluation

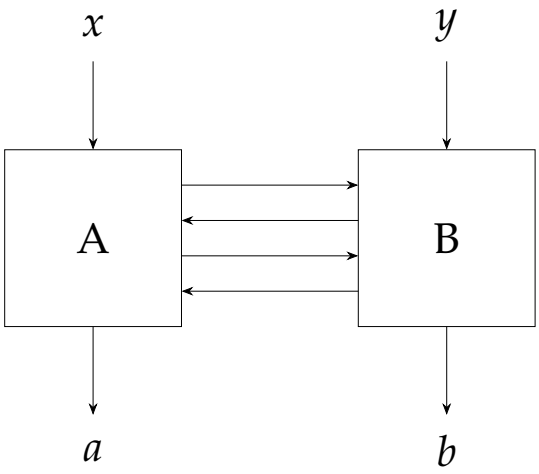


Figure 7: Secure Function Evaluation

Definition 10.1 (Secure Function Evaluation)**.** SFE is a task between two parties Alice and Bob who respectively own the inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. They interact over a communication channel and output $a \in \mathcal{A}$ and $b \in \mathcal{B}$. The given protocol is said to be a secure protocol calculating the functions $f_A : \mathcal{X} \times \mathcal{Y} \to \mathcal{A}$ for Alice and $f_B : \mathcal{X} \times \mathcal{Y} \to \mathcal{B}$ for Bob if it satisfies the following properties:

- **Correctness:** For honest Alice and Bob $f_A(x, y) = a, f_B(x, y) = b$
- **Security against cheating Bob:** If Alice is honest, then Bob cannot learn about x any more than from $f_B(x, y)$
- **Security against cheating Alice:** If Bob is honest, then Alice cannot learn about y any more than from $f_A(x, y)$

Making the notion of security precise is quite involved, so we will survey the main ideas at a high level.

The simulation paradigm

Definition 10.2 (Ideal Functionality)**.** Let f_a, f_B be a pair of functions pertaining to a SFE task. Then the ideal functionality is a box which takes as input x and y and outputs $f_A(x, y)$ and $f_B(x, y)$.

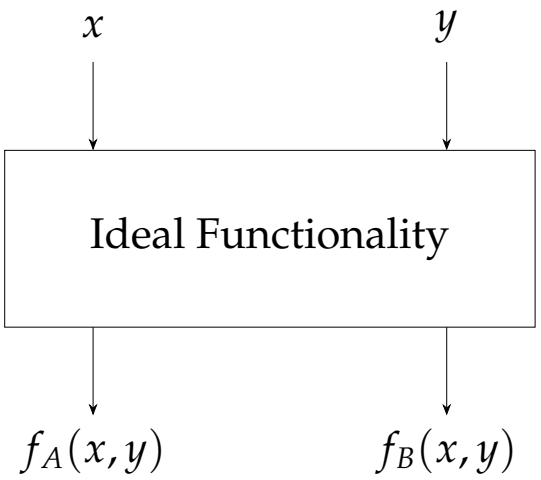


Figure 8: Ideal Functionality

Definition 10.3 (Security against cheating Bob and honest Alice)**.** A SFE protocol is said to be secure against cheating Bob if there exists a simulator which, by controlling Bob in an interaction with the ideal functionality, is able to generate a distribution on outputs that is indistinguishable from the outputs generated by Bob in the protocol.

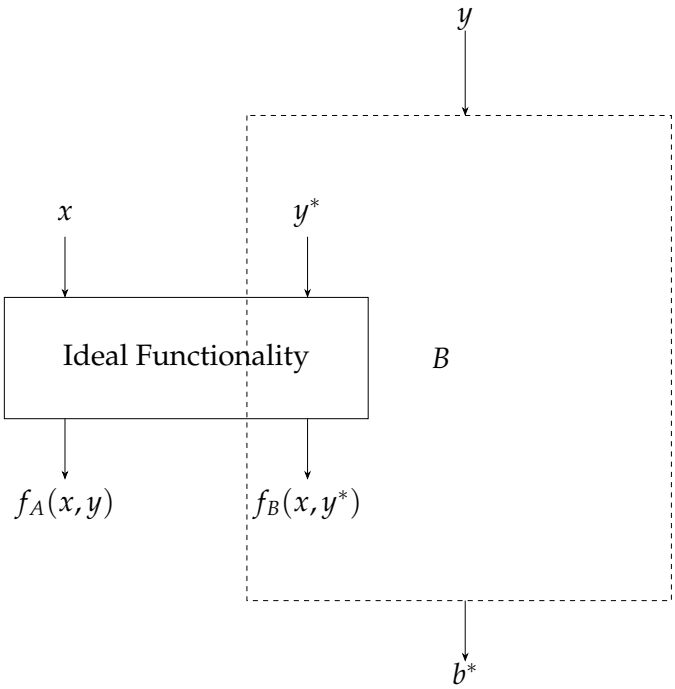


Figure 9: Simulator Paradigm

10.2 Oblivious Transfer

Oblivious Transfer is a SFE protocol in which:

$$\mathcal{X} = \{0,1\}^l \times \{0,1\}^l, \mathcal{Y} = \{0,1\}, f_A((x_0,x_1),y) = \perp, f_B((x_0,x_1),y) = x_y$$

Here Alice has two bitstrings x_0, x_1. Bob wants to retrieve one of those inputs but doesn't want Alice to know which one. Similarly, Alice wants to ensure that Bob doesn't get to know anything more than one of her bitstrings.

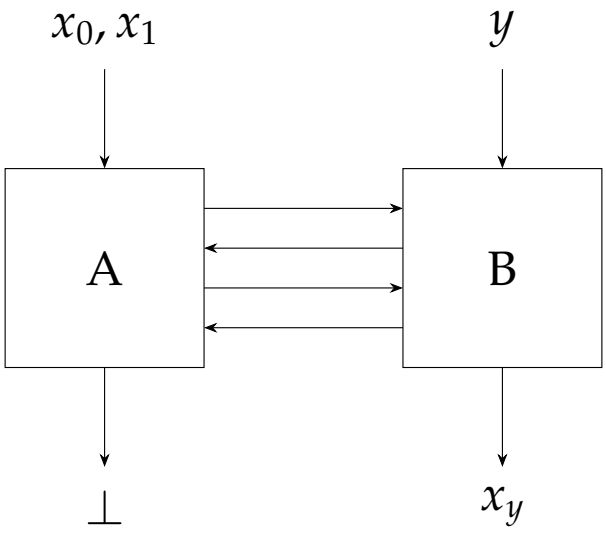


Figure 10: Oblivious Transfer

The notion of security demands that Alice gets to know nothing about Bob's input, as the Ideal Functionality gives no output to her. Similarly, Bob learns only one of the bitstrings, as he can use the Ideal Functionality exactly once. It turns out that 1-2 OT is universal for SFE, meaning that we can achieve any SFE task using it multiple times. For example, consider the following protocol for computing the distributed AND function. Here, we want Alice and Bob to output σ and s respectively so that $\sigma \oplus s = x \wedge y$.

Distributed AND function

1. Bob samples $a \leftarrow \{0,1\}$ and computes $b = a \oplus y$. He sends a to Alice.
2. Alice samples $\sigma \leftarrow \{0,1\}$ and $s_{b'} = \sigma \oplus (x \wedge (a \oplus b'))$ for $b' \in \{0,1\}$.
3. Alice and Bob use 1-2 OT with inputs $(s_0, s_1), b$ so that Bob receives $s = s_b$.
4. Alice outputs σ and Bob outputs s.

Figure 11: Distributed AND using 1-2OT

- **Correctness:** If Alice and Bob play honestly, then $s = s_b = \sigma \oplus (x \wedge (a \oplus (a \oplus y))) = \sigma \oplus x \wedge y$ so that $\sigma \oplus s = x \wedge y$.
- **Security against cheating Bob:** Bob receives one of the bitstrings $s_{b'} = \sigma \oplus (x \wedge (a \oplus b'))$. Since a uniformly random bit σ is xored, it appears uniformly random.
- **Security against cheating Alice:** Alice just learns a random bit a from Bob, which gives her no information about his input y.

Unfortunately, perfect 1-2 OT cannot be achieved without making computational assumptions in the classical or quantum world. Still, consider the following protocol:

Quantum Oblivious Transfer

1. Alice samples $x_A \leftarrow \{0,1\}^{2n}, \theta_A \leftarrow \{0,1\}^{2n}$ and sends $\left|x_A^{\theta_A}\right\rangle$ to Bob.
2. Bob samples $\theta_B \leftarrow \{0,1\}^{2n}$ and measures the states in θ_B basis to obtain x_B. He informs Alice that he has completed the measurements.
3. Alice sends θ_A to Bob.
4. Bob prepares the sets of indices
$$I = \{i : \theta_A[i] = \theta_B[i]\} \qquad I_y = I \qquad I_{1-y} = [2n]\backslash I$$
and sends (I_0, I_1) to Alice.
5. Alice sends $t_0 = s_0 \oplus x_{A,I_0}$ and $t_1 = s_1 \oplus x_{A,I_1}$ to Bob.
6. Alice outputs $\perp$ and Bob outputs $t_y \oplus x_{B,I_y}$.

Figure 12: Quantum Oblivious Transfer

- **Correctness:** evident since $x_{A,I_y} = x_{B,I_y}$ as they measured in the same basis.
- **Security against cheating Alice:** since θ_B is chosen at random, Alice only recieves the pair (I_0, I_1) which is a uniformly random partition of $[2n]$ containing no information about y. *So anything a dishonest Alice could do in this protocol can be simulated by an interaction with the ideal functionality, where the simulator would replace Bob's message (I_0, I_1)(which is not provided by the ideal functionality) with a uniformly random choice.*
- **Security against cheating Bob:** A malicious Bob can store all the received qubits and lie about performing the measurements. After receiving θ_A from Alice, he can measure in the same basis, obtaining $x_A = x_B$. So, he can recover both s_0 and s_1!

  There are two ways to solve this problem: by making some physical assumptions about the quantum memory of Bob or by making him commit his basis θ_B and outcome x_B before Alice reveals θ_A (using bit commitment).

Rabin's OT

Rabin's OT is closely related to the 1-2OT we discussed above. Here Alice wants to send a message m to Bob but Bob receives it with probability $\frac{1}{2}$. Alice remains unaware if the message has been delivered or not.
This can be made using 1-2OT: set $x_0 = m$, $x_1 \leftarrow \{0,1\}^{|m|}$ and $y \leftarrow \{0,1\}$.
Further, Rabin's OT can be used to construct 1-2 OT. For simplicity assume the messages m_0, m_1 are bits.

Rabin's OT to 1-2OT

1. Alice performs Rabin's OT n times with Bob for random bits.
2. Bob prepares index sets I_b and $I_{\bar{b}}$ of size $n/3$ where I_b has indices where he successfully received the bit while $I_{\bar{b}}$ has indices where he didn't receive the bit. He sends (I_0, I_1) to Alice.
3. Alice sends $\left(m_0 \oplus \left(\oplus_{i\in I_0} r_i\right), m_1 \oplus \left(\oplus_{i\in I_1} r_i\right)\right)$ to Bob.
4. Using the recovered bits corresponding to indices I_b, Bob recovers m_b.

Figure 13: Rabin's OT to 1-2OT

10.3 Bit Commitment

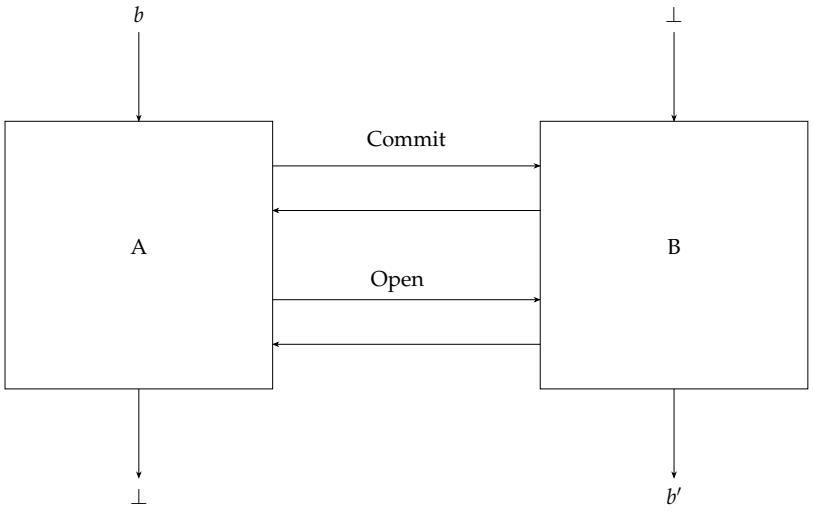


Figure 14: Bit Commitment

Definition 10.4 (Bit Commitment)**.** Bit Commitment is a task between two parties Alice (the committer) and Bob (the receiver). Alice gets an input b and outputs nothing. Bob gets no input but outputs a bit b'. A bit commitment has two phases - commit and open, satisfying the following properties:

1. **Correctness:** If both Alice and Bob are honest then $b = b'$.
2. **Hididng:** For any malicious Bob, the state of Bob at the end of the commit phase is independent of b.
3. **Binding:** For any three possible malicious behaviour A, A_0, A_1, the probabilities p_b of Bob outputting $b' = b$ after interacting with A in the commit phase and A_b in the open phase satisfies $p_0 + p_1 \leq 1$.

The bindidng property intuitively tries to capture the idea that once Alice has comitted to a bit running A, she should not be able to come up with two different behaviour A_0 and A_1 such that probability of Bob getting the bit in both cases is more than $1/2$.

Example. A bit commitment scheme for Yao's Millionare Problem

Remark. Suppose we have a bit commitment scheme in which is not binding, meaning in the opening phase, Alice can convince Bob of b being 0 or 1 as she pleases. In the classical case, it means that she possesses the certificates for both cases beforehand. If she wanted, she could have sent both to Bob (a rewinding type argument). It is possible to have two certificates simulataneously by creating a copy of the certificate in the classical case, but not in the quantum scenario.

10.3.1 Unviersality of Bit Commitment

We have already seen that quantum OT can be made secure using bit commitment. In this section, we will show that bit commitment is equivalent to oblivous transfer **in the quantum case**. Bit commitment is not universal for classical multi-party computation. The OT protocol must be quantum, even if the bit commitment is classical.

- **Correctness:** Evident from the construction.
- **Hididng:** For any malicious Bob, the state of Bob at the end of the commit phase is independent of b, as he receives no output.

Bit Commitment from 1-2OT

1. **Commit Phase:** Bob chooses $s_0, s_1 \leftarrow \{0,1\}^l$. Alice uses the bit b to perform 1-2 Oblivious Transfer with Bob.
2. **Open Phase:** Alice reveals $\hat{b}, \hat{s}$ to Bob. He accepts if $\hat{s} = s_{\hat{b}}$.

Figure 15: Bit Commitment from 1-2OT

- **Binding:** For any three possible malicious behaviour A, A_0, A_1, the probabilities p_b of Bob outputting $b' = b$ after interacting with A in the commit phase and A_b in the open phase satisfies $p_0 + p_1 \leq 1 + 2^{-l}$, because to change her commitment, Alice has to guess the other random bitstring of Bob. So this scheme is $\epsilon-$ binding for $\epsilon = 2^{-l}$

10.3.2 Impossiblity of Bit Commitment

Unfortunately, perfectly secure bit commitment is not possible even in the quantum world. Suppose the state shared by Alice and Bob after the commit phase is $\left|\psi^b_{AB}\right\rangle$ where b is the bit which Alice commits to. We can, without loss of generality, consider the state to be pure, using a purification and giving the auxiliary system to Alice. Due to the hiding property, we require that $\rho^0_B = \rho^1_B$. So, by Uhlmann's theorem Section 3.5, there exists a local unitary U_A on Alice's system which maps $\left|\psi^0_{AB}\right\rangle$ to $\left|\psi^1_{AB}\right\rangle$. Thus, she can cheat arbitrarily after the commit phase!

10.3.3 Computationally Secure Commitments

Here, we show a computationally secure bit commitment scheme based on PRGs. Let $G : \{0,1\}^n \to \{0,1\}^{3n}$ be a PRG. Consider the following protocol:

Computationally Secure Bit Commitment

Alice gets input b and outputs $\perp$
Bob gets input $\perp$ and outputs b'

1. **Commit Phase**
   (a) Bob selects $r \leftarrow \{0,1\}^{3n}$ and sends it to Alice.
   (b) Alice selects $s \leftarrow \{0,1\}^n$ and sends $\sigma = br \oplus G(s)$ to Bob.
2. **Open Phase**
   (a) Alice sends s to Bob.
   (b) Bob computes $\sigma \oplus G(s)$. If it is equal to 0^{3n}, he outputs 0 otherwise 1.

Figure 16: Computationally Secure Bit Commitment

Correctness is clear. The hiding property depends on Bob's ability to distinguish between $G(s)$ and $G(s) \oplus r$, which we can show to be negligible by the security of the PRG. Formally, argue that if Bob can distinguish $G(s)$ and $G(s) \oplus r$ for random r with probability $\frac{1}{2} + \epsilon$ then there exists a reduction breaking the PRG game with probability at least $\frac{1}{2} + \frac{\epsilon}{2}$.
If Alice wants to break the binding property, she must come up with a s, s' such that $G(s) \oplus G(s') = r$ for a random $3n$ bitstring r. This can happen with a probability at most $\frac{2^{2n}}{2^{3n}} = 2^{-n}$ which is negligible. This is where we require the PRG to be length tripling.
Observe that the binding property holds statistically while the hiding is computational.

§11. Evading Impossiblity by Physical Assumptions

Since perfect bit commitment is impossible even in the quantum world, we need to make some assumption on the power of Alice or Bob. For example, we have the following options:

1. The states ρ_B^0 and ρ_B^1 are distinguishable but Bob doesn't have the power to distinguish them.

2. The unitary U_A which Alice uses to change her commitment $\left(\left|\psi_{AB}^0\right\rangle \mapsto \left|\psi_{AB}^1\right\rangle\right)$ cannot be implemented efficiently. This problem relates to the complexity of implementing Uhlmann's transformations and is a key component in several quantum information tasks (see Fig. 17). Since we know that bit commitments can be constructed from one-way functions and can be broken using Uhlmann's transformation, we have the following in increasing order of hardness:

$$\begin{array}{c}\text{Inverting}\\ \text{One-Way Functions}\end{array} \leq \begin{array}{c}\text{Breaking a}\\ \text{Commitment Scheme}\end{array} \leq \begin{array}{c}\text{Implementing}\\ \text{Uhlmann Transformations}\end{array}$$

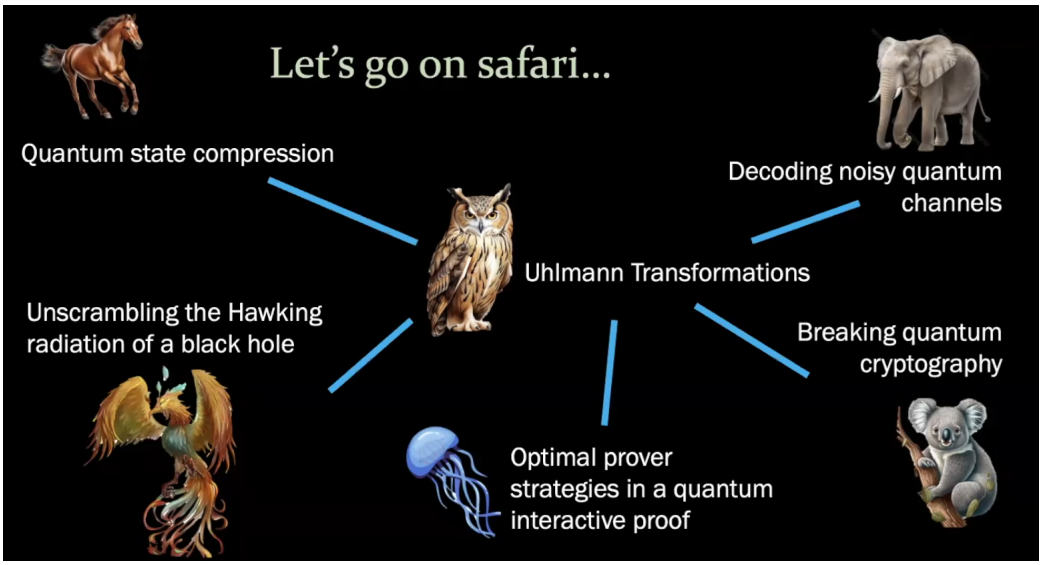


Figure 17: Implications of Uhlmann's Theorem (Slide by Henry Yuen)

3. Alice doesn't hold a purification of the state - that is with some third party.

   For example, let us study the following geometric protocol for bit commitment, based on the third option. We will assume that a third party called the Wizard prepares the states for Alice and Bob. Alice commits to a real number r in this protocol.

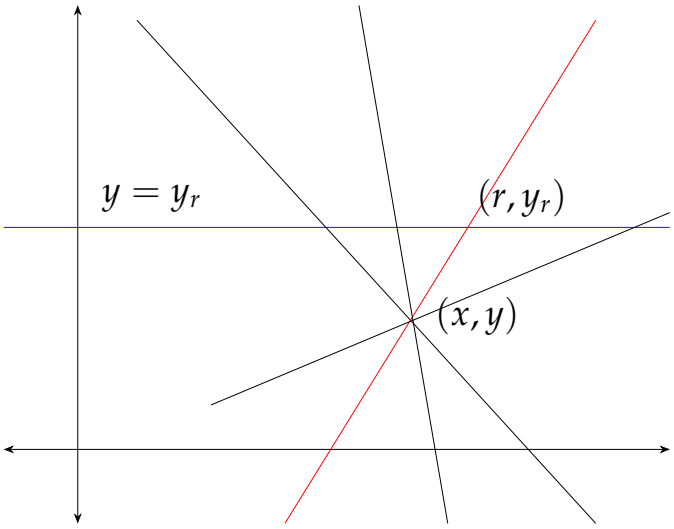
Bit Commitment using Lines and Points

(a) The Wizard samples three real numbers a, b, x from some distribution. He sends (a, b) to Alice and $(x, y = ax + b)$ to Bob.

(b) **Commit phase:** Alice computes $y_r = ar + b$ and sends it to Bob.

(c) **Open phase:** Alice sends r, a, b to Bob. Bob checks if (x, y) and (r, y_r) lie on the line. If not, he reports that Alice was cheating. Otherwise he outputs r.

Figure 18: Bit Commitment protocol where the purification is with a third party



- **Correctness:** If both Alice and Bob are honest then Bob outputs r which Alice had commited to.
- **Hididng:** For any malicious Bob, the state of Bob at the end of the commit phase is independent of r, because there can be infinitely many lines which pass through the point (x, y) as well as intersect the line $y = y_r$. This may however fail with a small probability, if $y_r = y$.
- **Binding:** Alice has no idea about the point which Bob has received from the Wizard. So, if she wants to change her mind to $\hat{r}$, then she has to come up with a line which contains $(\hat{r}, y_r)$ as well as Bob's point (x, y) and is on the line $y = ax + b$. This is not possible.

11.1 Criteria for the Assumptions

We must remember the following criterion while making the assumptions:

1. **What resources do we need to execute the protocol?**

   It should be easy to impelement by the honest parties.

2. **What resources do we need to break the protocol?**

   Any adversary must require significantly high computational power to break the protocol.

3. **How long do the security guarantees remain valid after the assumptions are lifted?**

   We want everlasting security, so that even if the adversary later gets access to more resources, still he is unable to retroactively break the protocol.

11.2 The Noisy Storage Model

In our analyses, we will be using the noisy storage model. In principle, this model allows us to achieve security for any cryptographic model where Alice and Bob don't trust each other. We assume that the adversary can only store q (noisy) qubits at one particular point during the protocol. Otherwise, the adversary remains all powerful: he/she may perform arbitrary quantum operations/computations, arbitrary encoding and decoding procedures, and store an infinite amount of classical information. Before, or after the execution of the protocol, however, the attacker is allowed to have an arbitrary quantum memory, and even a quantum computer. In particular, this means that if tomorrow we can build better quantum memories, security can nevertheless never be broken retroactively. But during the *waiting times*, the adversary can only keep the qubits in a noisy quantum storage.
Fact: Using q qubits, we can store no more than q qubits or classical bits reliably.
Thus if the number of protocols N used in our protocol are larger than q, then the adversary cannot use any encoding procedure to store all of them.

11.3 A protocol for 1-2 Oblivious Transfer

Now, let us discuss a protocol for 1-2 OT secure in the noisy storage model. This is very similar to the QKD protocols we discussed before and the protocol discussed in Section 10.2.

1-2 OT in the noisy storage model

Alice Input: (x_0, x_1) each N bits long
Bob Input: y Output: x_y

1. Alice samples $x_A \leftarrow \{0,1\}^{4N}$ and $\theta_A \leftarrow \{0,1\}^{4N}$. She sends $\left|x_A^{\theta_A}\right\rangle$ to Bob over the quantum channel.
2. Depending on y, Bob measures all the received qubits in the standard basis $(y = 0)$ or in the hadamard basis $(y = 1)$
3. Alice and Bob wait for some time Δt
4. Alice sends θ_A to Bob.
5. Bob computes the index set $I_y = \{i|\theta_A[i] = y\}$.
6. Alice too computes $I_0 = \{i|\theta_A[i] = 0\}$ and $I_1 = \{i|\theta_A[i] = 1\}$ sends $s_0 \oplus x_{A,I_0}$ and $s_1 \oplus x_{A,I_1}$ to Bob.
7. Bob computes x_{A,I_y} and xors it with $s_y \oplus x_{A,I_y}$ to obtain s_y.

Figure 19: 1-2 OT in the noisy storage model

We may have to pad s_0, s_1 in the above with zeros. Using a chernoff bound, it can be shown that the sets I_0 and I_1 will be of size at least N with a high probability.

- **Correctness:** evident since for the subset I_y , they measured in the same basis.
- **Security against cheating Alice:** since θ_B is chosen at random, Alice only recieves the pair (I_0, I_1) which is a uniformly random partition of $[2n]$ containing no information about y. So anything a dishonest Alice could do in this protocol can be simulated by an interaction with the ideal functionality, where the simulator would replace Bob's message (I_0, I_1)(which is not provided by the ideal functionality) with a uniformly random choice.
- **Security against cheating Bob:** Assume that Bob can store at most q qubits. Intuitively, due to the waiting time introduced in the protocol, Bob can no longer reliably store all the received qubits! Formally, if we can bound $H_{\min}(X_{A,I_0}X_{A,I_1}|K, Q, \theta_A)$ (where K is the classical register of Bob and Q is his quantum register), then we can use Privacy Amplification to ensure that Bob learns more than one of the two inputs to Alice.

$$H_{\min}(X_{A,I_0}X_{A,I_1}|K, Q, \theta_A) \geq H_{\min}(X_{A,I_0}X_{A,I_1}|K, \theta_A) - q$$

Now the analysis reduces to what we did for QKD using the guessing game.

$$H_{\min}(X_{A,I_0}X_{A,I_1}|K, \theta_A) = n\left(\log\left(\frac{1}{2} + \frac{1}{2\sqrt{2}}\right)\right) \approx 0.22n$$

To complete the arguemnt by bounding the individual min-entropies of X_{A,I_0} and X_{A,I_1}, we use the **min-entropy splitting lemma**, which states that there exists a subset of indices for which the individual min-entropies is at least approximately half of the joint entropy. This bound on the min-entropy ensures that Bob can learn only one of the inputs.

§12. Delegation of Quantum Computation

Definition 12.1 (Delegated Computation)**.** In the task of delegated computation, a verifier $\mathcal{V}$ has an input $(x, \mathcal{C})$, where x is a classical input and $\mathcal{C}$ is a classical description of a quantum circuit. $\mathcal{V}$ has a multiple round interaction with a quantum prover $\mathcal{P}$. At the end, $\mathcal{V}$ gives a classical output y or aborts. The protocol is called:

- **Correct:** If $y = \mathcal{C}(x)$ with high probability.
- **Blind:** If $\mathcal{P}$ has no information about $(x, \mathcal{C})$ after the protocol.
- **Verifiable:** If for any deviating prover, $\mathcal{V}$ aborts or outputs $y = \mathcal{C}(x)$

Though neither of the above properties implies the other, in practice often verifiability follows from blindness by arguing that $\mathcal{V}$ can run some dummy tests for which they already know the answer, but $\mathcal{P}$ has no means of distinguishing between a dummy and a real round.
There have been many approaches to achieve efficient delegation of quantum computation, for instance, using verifiable delegation of quantum circuits, delegation using measurement based quantum computation(MBQC), delegation using two entangled quantum provers [RUV12]. We will now look at one of the many approaches to delegating quantum computation.

12.1 Verifiable Delegation of Quantum Circuits

First, let us recall a few points:

- The complexity class BQP captures the notion of efficient quantum computation.

Definition 12.2 (Bounded Error Quantum Polynomial Time)**.** A promise problem $\mathcal{A} = (\mathcal{A}_{\text{Yes}}, \mathcal{A}_{\text{No}}, \mathcal{A}_{\text{Invalid}})$ is said to be in BQP if there is a P-uniform quantum circuit family $\{Q_n\}$ and a polynomial q such that for all $x \in \{0,1\}^n$, Q_n runs on registers A and B of n and $q(n)$ qubits respectively. The first qubit B_1 of B is conventionally taken to be the output qubit. After running Q_n this qubit is measured to obtain outcome y which satisfies the following:

  - **Completeness:** $x \in A_{\text{Yes}} \implies y = 1$ with probability at least $2/3$.
  - **Soundness:** $x \in A_{\text{No}} \implies y = 1$ with probability at least $1/3$.
  - **Invalid:** $x \in A_{\text{Invalid}}$ then $y = 0$ or 1 arbitrarily.

- The Pauli Group $\mathscr{P}$ is the group generated by the Pauli matrices $\mathscr{P} = \langle \{I, X, Y, Z\} \rangle$.
- The Clifford Group $\mathscr{C}$ is the normalizer of the Pauli group, i.e. $\mathscr{C} = \mathcal{N}(\mathscr{P}) = \{U | P \in \mathscr{P} \implies UPU^\dagger \in \mathscr{P}\}$. We also think of it as the group generated by $\{H, S, \text{CNOT}\}$. For multiple qubits, we take the tensor product of these groups on single qubits.
- By the Solovay Kitaev theorem, we can restrict the set of gates to some simple set of gates called "universal gate sets." Two examples are

$$\mathcal{G}_1 = \left\{ G = \begin{pmatrix} \cos \pi/8 & -\sin \pi/8 \\ \sin \pi/8 & \cos \pi/8 \end{pmatrix}, \text{CNOT} \right\} \qquad \mathcal{G}_2 = \{H, T, \text{CNOT}\}$$

- There doesn't exist a universal gate set consisting of only Clifford gates.

- Analogous to the universal turing machine, we have a universal circuit theorem:

Theorem 12.1 (Universal Circuit)**.** For any integer n and size parameter s there exists a fixed (universal) circuit $\mathcal{C}_U$ acting on $n+m$ qubits with $m = \text{poly}(n,s)$ such that for any circuit $\mathcal{C}$ and any input $x \in \{0,1\}^n$, there exists $z \in \{0,1\}^m$ such that $\mathcal{C}(x)$ and $\mathcal{C}_U(x,z)$.

To delegate quantum computation, $\mathcal{V}$ encrypts the input state ρ to obtain $\tilde{\rho}$ and sends it to $\mathcal{P}$, who performs some circuit $\tilde{\mathcal{C}}(\tilde{\rho}) = \widetilde{\mathcal{C}(\rho)}$ and returns it to $\mathcal{V}$. Now $\mathcal{V}$ can decrypt this to obtain the result. For example, let us consider the quantum one-time pad and see the effect of different operations on it:

$$\text{Enc}((x,z),\rho) = X^x Z^z \rho (X^x Z^z)^\dagger = X^x Z^z \rho Z^z X^x$$

Pauli Operations

In general, we can write the operation of any element of the Pauli group as the operation of $X^{x_1}Z^{z_1}$ on ρ. We want $\mathcal{P}$ to perform the following transformation:

$$X^x Z^z \rho Z^z X^x \mapsto X^x Z^z X^{x_1} Z^{z_1} \rho Z^{z_1} X^{x_1} Z^z X^x = X^{x \oplus x_1} Z^{z \oplus z_1} \rho Z^{z \oplus z_1} X^{x \oplus x_1}$$

Using $X^a Z^b = (-1)^{a \cdot b} Z^b X^a$. We can safely ignore the phase factors here as they get cancelled. It turns out that for Pauli operations, $\mathcal{V}$ doesn't need $\mathcal{P}$: it can just adjust the key as:

$$x \mapsto x \oplus x_1 \qquad\qquad z \mapsto z \oplus z_1$$

and decrypt to obtain the result.

Clifford Operations

Let C be a Clifford operator. Then as $C^\dagger$ is also Clifford, $C^\dagger X C = X^{a_1} Z^{b_1}$ and $C^\dagger Z C = X^{a_2} Z^{b_2}$ where a_1, a_2, b_1, b_2 are bits. Then ignoring global phases,

$$X^x Z^z C^c = C^c C^{c\dagger} X^x C^c C^{c\dagger} Z^z C^c = C^c X^{x - x \wedge c} (X^{a_1} Z^{b_1})^{x \wedge c} (X^{a_2} Z^{b_2})^{c \wedge z} Z^{z - z \wedge c} = C^c X^{x'} Z^{z'}$$

Thus, applying a Clifford on an unencrypted state is the same as applying it on encrypted state and updating the key accordingly.

The T Gate and Magic states

The T gate is a non-clifford gate, and it cannot be handled in the simple way above by applying a gate on the encrypted state and updating the key. Fortunately, magic states come to our rescue! Informally, these are certain pure ancilla states (independent of the input) that aid us in our computation. One example is

$$|\pi/4\rangle = T|+\rangle = \frac{|0\rangle + e^{i\pi/4}|1\rangle}{\sqrt{2}}$$

Preparing this state itself requires the T gate but note that we only need to apply the gate to a fixed known input state, which is independent of the actual state ρ on which we want to apply T. The preparation of single qubit states is a relatively easy task which $\mathcal{V}$ can perform using a single qubit quantum computer.
Consider the following circuit (for simplicity, consider the input state to be a pure state $|\psi\rangle$):

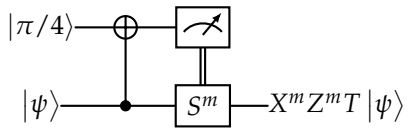


Figure 20: Teleporting a T gate

This performs the transformation

$$\begin{aligned}
|\pi/4\rangle |\psi\rangle &= \frac{|0\rangle + e^{i\pi/4}|1\rangle}{\sqrt{2}}(\alpha|0\rangle + \beta|1\rangle) \\
&\to \alpha\frac{|0\rangle + e^{i\pi/4}|1\rangle}{\sqrt{2}}|0\rangle + \beta\frac{|1\rangle + e^{i\pi/4}|0\rangle}{\sqrt{2}}|1\rangle \\
&= |0\rangle\frac{\alpha|0\rangle + e^{i\pi/4}\beta|1\rangle}{\sqrt{2}} + |1\rangle\frac{e^{i\pi/4}\alpha|1\rangle + \beta|0\rangle}{\sqrt{2}} \\
&= \frac{|0\rangle T|\psi\rangle + |1\rangle TX|\psi\rangle}{\sqrt{2}} \\
&\to X^m Z^m T|\psi\rangle
\end{aligned}$$

Where the last step follows because $STX = XZT$ upto a global phase[5]. If instead, we had performed the same operations on an encrypted $X^x Z^z |\psi\rangle$, we would obtain

$$X^m Z^m T X^x Z^z |\psi\rangle = X^m Z^m (ZSXZ)^x Z^z T|\psi\rangle = X^m S^x Z^{m+x} X^x Z^{z+x} T|\psi\rangle = S^x X^{m+x} Z^{m+z} T|\psi\rangle$$

which is $S^x X^{x'} Z^{z'}$. Observe that this gives us an extra phase factor dependent on the key. This can be removed by introducing randomness and applying a one-time pad on the magic state as well (see [Bro18] for details).

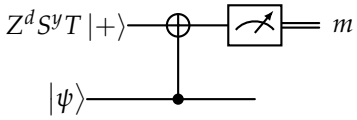
12.1.1 Blindness

Blindness of the input state is ensured by the encryption, but for the circuit, it seems that we must disclose it to $\mathcal{P}$. This can be resolved by using the universal circuit, and encoding the required circuit as a part of the input, which is then encrypted. Also, the interaction involving the T gates needs to be done with some randomness, refer [Bro18].

12.1.2 Verifiability

The intuition behind verifiability is to introduce some test rounds for which $\mathcal{V}$ already knows the answer. First, consider the following technique of delegating a T gate using something like a one-time pad

1. $\mathcal{V}$ samples $d, y \leftarrow \{0,1\}$ and sends $Z^d S^y T|+\rangle$ to $\mathcal{P}$.
2. $\mathcal{P}$ applies the following circuit:



   And sends the measurement outcome m to $\mathcal{V}$
3. $\mathcal{V}$ sends back $x = y \oplus m$ and $\mathcal{P}$ applies S^x to the resultant state in the second register.

[5] In this section, all the equalities among operators are true upto a global phase

It is easy to verify that the state with the prover after step 2 is $Z^d S^y T X^m |\psi\rangle$. Now on applying $S^{y\oplus m}$ observe:

$$S^{y\oplus m} Z^d S^y T X^m = S^{y\oplus m+y} Z^d T X^m = Z^{y(m\oplus 1)} S^m Z^d T X^m = Z^{y(m\oplus 1)\oplus d\oplus m} X^m T$$

Where we used $STX = ZXT$ and $S^{y\oplus m+y} = Z^{y(m\oplus 1)} S^m$. Now, $\mathcal{V}$ runs (X or Z) 'test' or 'computation' rounds according to their choice. From the blindness property, $\mathcal{P}$ cannot distinguish between these rounds. The idea is to convert the circuit into identity in a test round, without the prover noticing. For the Clifford gates, this can be easily done by adjusting the updates made to the key by $\mathcal{V}$. Let us look at how to handle the T gate in a X-test run: Let $|\psi\rangle = X^a |0\rangle$ and replace the magic state $Z^d S^y T |+\rangle$ by $X^d |0\rangle$. Then the transformations are:

$$|d\rangle |a\rangle \to |d \oplus a\rangle |a\rangle \implies m = d \oplus a$$

Thus, the measurement outcome is deterministically related to d, a which $\mathcal{V}$ can compute and match. The state $|\psi\rangle$ is left unchanged. A similar trick works for the Z-test.
Now we need to argue that any corruption done by the server will be discovered in the X or Z test rounds. For this we require the **Pauli Twirl** technique.

Lemma 12.2 (Pauli Twirl)**.** Let ρ be a single qubit density matrix and $P, P' \in \mathscr{P}$. Then

$$\frac{1}{4} \sum_{Q\in\mathscr{P}} (Q^\dagger P Q)\rho(Q^\dagger (P')^\dagger Q) = P\rho P^\dagger \delta_{PP'}$$

The same result holds for $n-$qubit Pauli operators.

Proof. Take $P = X^x Z^z$ and $P' = X^{x'} Z^{z'}$. Also, since Pauli's are Hermitian, we can ignore the †.

$$\begin{aligned}
\sum_{Q\in\mathscr{P}} (Q^\dagger P Q)\rho(Q^\dagger (P')^\dagger Q) &= \sum_{Q\in\mathscr{P}} (QPQ)\rho(QP'Q) \\
&= P\rho P' + XPX\rho XP'X + YPY\rho YP'Y + ZPZ\rho ZP'Z \\
&= P\rho P' + (-1)^{x\oplus x'} P\rho P' + (-1)^{x\oplus x'\oplus z\oplus z'} P\rho P' + (-1)^{z\oplus z'} P\rho P' \\
&= 4P\rho P\delta_{PP'}
\end{aligned}$$

Where the last step can be verified by taking 4 cases for $x \oplus x'$ and $z \oplus z'$. The multi-qubit case is similar. ■

Let the desired circuit be $\tilde{C} = \tilde{C}_m \dots \tilde{C}_1$ but $\mathcal{P}$ deviates from the desired implementation and ends up implementing $U_m \tilde{C}_m \dots U_1 \tilde{C}_1$. This can be also written as $U\tilde{C}$ where U encapsulates the entire deviation. The overall process can be written as $c(Q)U\tilde{C}Q\rho Q^\dagger \tilde{C}^\dagger U^\dagger c(Q)^\dagger$ where Q is the one-time pad and $c(Q)$ is the desired correction. We also know that

$$c(Q)\tilde{C}Q\rho Q^\dagger \tilde{C}^\dagger c(Q)^\dagger = c(Q)QC\rho C^\dagger Q^\dagger c(Q)^\dagger = C\rho C^\dagger$$

where C is the real circuit we want to implement. Further, let $U \sum_{P\in\mathscr{P}} \alpha_P P$ be the decomposition of U in terms of Paulis.

$$\begin{aligned}
\sum_{Q\in\mathscr{P}} c(Q)U\tilde{C}Q\rho Q^\dagger \tilde{C}^\dagger U^\dagger c(Q)^\dagger &= \sum_{Q\in\mathscr{P}} c(Q)Uc(Q)^\dagger c(Q)\tilde{C}Q\rho Q^\dagger \tilde{C}^\dagger c(Q)^\dagger c(Q)U^\dagger c(Q)^\dagger \\
&= \sum_{Q\in\mathscr{P}} c(Q)Uc(Q)^\dagger C\rho C^\dagger c(Q)U^\dagger c(Q)^\dagger \\
&= \sum_{P\in\mathscr{P}} |\alpha_P|^2 PC\rho C^\dagger P^\dagger
\end{aligned}$$

The last step follows form the Pauli twirl. Thus the deviating unitary reduces to Paulis, and these will be detected in the X or Z test rounds.

References

[Bro18] Anne Broadbent. How to verify a quantum computation. *Theory of Computing*, 14(1):1–37, 2018. URL: http://dx.doi.org/10.4086/toc.2018.v014a011, doi:10.4086/toc.2018.v014a011.

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969. URL: https://link.aps.org/doi/10.1103/PhysRevLett.23.880, doi:10.1103/PhysRevLett.23.880.

[Cle19] Richard Cleve. Qic 890 entanglement and nonlocal effects lecture notes. https://cleve.iqc.uwaterloo.ca/resources/Qic890LectureNotes2019Apr22(V22).pdf, 2019. Version 22, April 22, 2019.

[KRS09] Robert Konig, Renato Renner, and Christian Schaffner. The operational meaning of min- and max-entropy. *IEEE Transactions on Information Theory*, 55(9):4337–4347, sep 2009. URL: https://doi.org/10.1109%2Ftit.2009.2025545, doi:10.1109/tit.2009.2025545.

[MYS12] M McKague, T H Yang, and V Scarani. Robust self-testing of the singlet. *Journal of Physics A: Mathematical and Theoretical*, 45(45):455304, October 2012. URL: http://dx.doi.org/10.1088/1751-8113/45/45/455304, doi:10.1088/1751-8113/45/45/455304.

[NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge, 2010. URL: http://mmrc.amss.cas.cn/tlb/201702/W020170224608149940643.pdf.

[PAB+09] Stefano Pironio, Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, and Valerio Scarani. Device-independent quantum key distribution secure against collective attacks. *New Journal of Physics*, 11(4):045021, April 2009. URL: http://dx.doi.org/10.1088/1367-2630/11/4/045021, doi:10.1088/1367-2630/11/4/045021.

[RUV12] Ben W. Reichardt, Falk Unger, and Umesh Vazirani. A classical leash for a quantum system: Command of quantum systems via rigidity of chsh games, 2012. URL: https://arxiv.org/abs/1209.0448, arXiv:1209.0448.

[Sin23] Singular Value Decomposition. Singular value decomposition, 2023. URL: https://en.wikipedia.org/wiki/Singular_value_decomposition.

[TFKW13] Marco Tomamichel, Serge Fehr, Jędrzej Kaniewski, and Stephanie Wehner. A monogamy-of-entanglement game with applications to device-independent quantum cryptography. *New Journal of Physics*, 15(10):103002, oct 2013. URL: https://doi.org/10.1088%2F1367-2630%2F15%2F10%2F103002, doi:10.1088/1367-2630/15/10/103002.

[TL17] Marco Tomamichel and Anthony Leverrier. A largely self-contained and complete security proof for quantum key distribution. *Quantum*, 1:14, July 2017. URL: http://dx.doi.org/10.22331/q-2017-07-14-14, doi:10.22331/q-2017-07-14-14.

[VV14] Umesh Vazirani and Thomas Vidick. Fully device-independent quantum key distribution. *Physical Review Letters*, 113(14), September 2014. URL: http://dx.doi.org/10.1103/PhysRevLett.113.140501, doi:10.1103/physrevlett.113.140501.

[VW16] Thomas Vidick and Stephaine Wehner. edx quantum cryptography, 2016. URL: https://www.edx.org/course/quantum-cryptography.

[Wil16] Mark Wilde. Quantum information theory. In *Quantum Information Theory*, pages xi–xii. Cambridge University Press, nov 2016. URL: https://doi.org/10.1017%2F9781316809976.001, doi:10.1017/9781316809976.001.