PHYS 130C

Part 2: Quantum Information Theory

From Classical to Quantum Information

■ Classical Probability

■ Random Variable and Probability

A **random variable** X is a variable that can take *random* values x in a set $\mathcal{X} = \{x_1, x_2, \ldots\}$ of all *possible values*, with an *assignment* of a **probability** $p(x)$ to each value $x \in \mathcal{X}$, quantifying the degree of *belief* or *certainty* to observe X taking on the value x.

- Only specifying the set $\mathcal{X}$ of possible values is not enough to define a random variable, assigning the *probability* $p(x)$ is essential.

  Example: X = a random image of muffin. The following images are all *possible*, but *not equally likely*.

- Probability assignment describes our current state of **knowledge** about the random variable. (It can be *subjective*.)
- The probability assignment should be *updated*, if our knowledge has been *changed* by **observations** (providing new evidence).

Example:

Tossing a coin. $x = \text{head}, \text{tail}$.

- **Prior probability** *before* observation

$$\begin{array}{ccc} x & \text{head} & \text{tail} \\ p(x) & \frac{1}{2} & \frac{1}{2} \end{array} \tag{1}$$

- **Posterior probability** *after* observing head up.

$$\begin{array}{ccc} x & \text{head} & \text{tail} \\ p(x) & 1 & 0 \end{array} \tag{2}$$

Observation removes uncertainty and provides **information**.

Properties:

- Positivity:

$$\boxed{p(x) \geq 0.} \tag{3}$$

- Normalization:

$$\boxed{\sum_{x \in \mathcal{X}} p(x) = 1.} \tag{4}$$

Expectation Value

The **expectation value** (mean, average, first-moment) of a random variable X:

$$\boxed{\langle x \rangle = \mathbb{E}[X] = \sum_{x \in \mathcal{X}} x\, p(x).} \tag{5}$$

Expectation value also can be defined for a *function* of the random variable $f(X)$,

$$\langle f(x) \rangle = \mathbb{E}[f(X)] = \sum_{x \in \mathcal{X}} f(x)\, p(x). \tag{6}$$

Properties:

- Double expectation

$$\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X]. \tag{7}$$

- Linearity (for two random variables X and Y and a constant α)

$$\begin{gathered} \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y], \\ \mathbb{E}[\alpha\, X] = \alpha\, \mathbb{E}[X]. \end{gathered} \tag{8}$$

More generally, for multiple random variables X_i ($i = 1, 2, \ldots$) linearly combined together,

$$\mathbb{E}\left[\sum_i \alpha_i\, X_i\right] = \sum_i \alpha_i\, \mathbb{E}[X_i]. \tag{9}$$

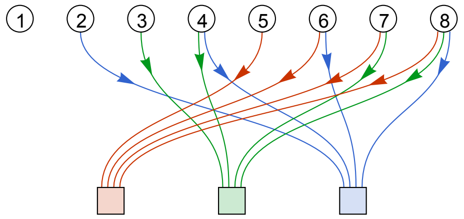
Information

Answering every independent **yes-or-no question** provides 1 **bit** of information.

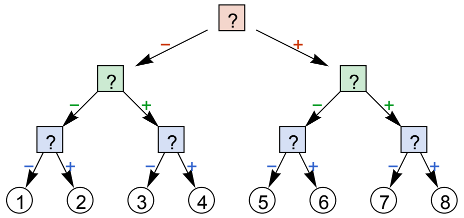
Example: Binary pooled testing

Among eight individuals, there is *one* (and *only one*) contracted COVID. How many tests are needed to identify the COVID positive individual?

- Sample collection (encoding scheme)



- Test result analysis (decoding scheme)



In a binary search, answering n independent yes-or-no questions will simultaneously do the following:

- Identify a unique outcome out of 2^n equally-likely possibilities,
- Collapse the probability from $p = \left(\frac{1}{2}\right)^n$ to $p = 1$ for the observed outcome,
- Provide n bits of information:

$$p = \left(\frac{1}{2}\right)^n \Rightarrow n = -\log_2 p = -\frac{\log p}{\log 2}. \tag{10}$$

  $\log 2 = 1$ bit is treated as an **information unit**.

The amount of **information** I gained from the *observation* of a **probability** p outcome is

$$\boxed{I = -\log p. \quad \text{(for a particular outcome)}} \tag{11}$$

Example: average information

Observing a random variable X with the following (prior) probability

$$\begin{array}{ccccc} x & a & b & c & d \\ p(x) & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \end{array}. \tag{12}$$

- Information gain can be different for different observation outcomes

$$\begin{aligned} I(a) &= -\log(1/2) = \log 2 = 1 \text{ bit}, \\ I(b) &= -\log(1/4) = 2\log 2 = 2 \text{ bit}, \\ I(c) &= -\log(1/8) = 3\log 2 = 3 \text{ bit}, \\ I(d) &= -\log(1/8) = 3\log 2 = 3 \text{ bit}. \end{aligned} \tag{13}$$

- However, different outcome happens with different probability. What is *average* amount of information that we can obtain from observing X (regardless of its outcome)?

$$\begin{aligned} I(X) &= I(a)\,p(a) + I(b)\,p(b) + I(c)\,p(c) + I(d)\,p(d) \\ &= \left(1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 3 \times \frac{1}{8}\right) \text{bit} \\ &= 1.75\ \text{bit}. \end{aligned} \tag{14}$$

In conclusion, given a random variable X, the *expected* **information** gained from a *full observation* of X is

$$\boxed{I(X) = -\langle \log p(x) \rangle = -\sum_{x \in \mathcal{X}} p(x) \log p(x).} \tag{15}$$

Entropy

The **Shannon entropy** measures the *uncertainty* (*lack* of information, ignorance) remained in a random variable X, determined by the *probability distribution* $p(x)$,

$$\boxed{S(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x).} \tag{16}$$

- Entropy is always non-negative (follows from $0 \leq p(x) \leq 1$)

$$\boxed{S(X) \geq 0.} \tag{17}$$

  - $S(X) = 0$ means the value of X is known for certain (no randomness).
  - Large $S(X)$ indicates large uncertainty in X.
- Entropy can be changed by *observation*, as observation can remove/reduce uncertainty from a random variable.

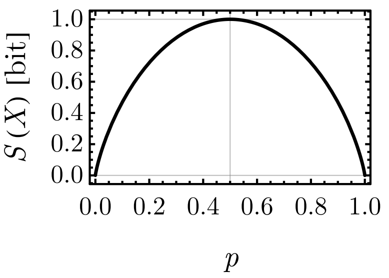
Example:

- A binary random variable X (with $\mathcal{X} = \{\text{false}, \text{true}\}$)

$$\begin{aligned} p(\text{false}) &= 1 - p, \\ p(\text{true}) &= p, \end{aligned} \tag{18}$$

  where $0 \leq p \leq 1$. Entropy of X

$$S(X) = -p \log p - (1 - p) \log (1 - p). \tag{19}$$



- $S(X) = 0$ when $p = 0$ ($X =$ false for sure) or $p = 1$ ($X =$ true for sure).
- $S(X)$ is maximized at $p = 1/2$, where X is most uncertain. The maximum entropy of a binary random variable is 1 bit.

▪ Mutual Information

Mutual information $I(X : Y)$ quantifies the amount of information shared between two random variables X and Y.

$$\boxed{I(X : Y) = S(X) + S(Y) - S(X, Y),} \tag{20}$$

where, given the *joint* distribution $p(x, y)$ for $(x, y) \in \mathcal{X} \times \mathcal{Y}$:

- $S(X) = -\sum_x p(x) \log p(x)$ is the entropy of X — the uncertainty in X regardless of Y, with $p(x) = \sum_y p(x, y)$ being the *marginal* distribution of X.
- $S(Y) = -\sum_y p(y) \log p(y)$ is the entropy of X — the uncertainty in Y regardless of X, with $p(y) = \sum_x p(x, y)$ being the *marginal* distribution of Y.
- $S(X, Y) = -\sum_{x,y} p(x, y) \log p(x, y)$ is the **joint entropy** of X and Y — the uncertainty of the joint distribution.

Alternatively,

$$I(X : Y) = S(X) - S(X \mid Y) = S(Y) - S(Y \mid X), \tag{21}$$

where

- $S(X \mid Y) = -\sum_{x,y} p(x, y) \log p(x \mid y)$ is the **conditional entropy** of X given Y — the remaining uncertainty in X after knowing Y.
- $S(Y \mid X) = -\sum_{x,y} p(x, y) \log p(y \mid x)$ is the **conditional entropy** of Y given X — the remaining uncertainty in Y after knowing X.

Exc 1

> Show that $S(X \mid Y) = S(X, Y) - S(Y)$ and $S(Y \mid X) = S(X, Y) - S(X)$, therefore Eq. (21) is consistent with the definition in Eq. (20).

Eq. (21) implies that $I(X : Y)$ measures the *reduction in uncertainty* of one variable due to *knowledge* of the other, hence the name “mutual information”.

- If X and Y are independent, $I(X : Y) = 0$.

- If X and Y are perfectly correlated, $I(X:Y)$ is maximized.

Key properties:

- **Non-negativity**: $I(X:Y) \geq 0$,
- **Symmetry**: $I(X:Y) = I(Y:X)$.

Example: X and Y are two binary random variables (classical bits), perfectly correlated with each other, described by the joint distribution $p(x, y)$:

$p(x, y)$		x		$p(y)$
		0	1	
y	0	$\frac{1}{2}$	0	$\frac{1}{2}$
	1	0	$\frac{1}{2}$	$\frac{1}{2}$
$p(x)$		$\frac{1}{2}$	$\frac{1}{2}$	

(22)

One can compute that

$$\begin{aligned} &S(X) = 1\,\text{bit},\ S(Y) = 1\,\text{bit},\ S(X,\ Y) = 1\,\text{bit}, \\ &I(X:Y) = (1 + 1 - 1)\,\text{bit} = 1\,\text{bit}. \end{aligned} \tag{23}$$

In fact, for two classical bits, the *maximal* mutual information between them is at most 1bit. However, as we will see later, the mutual information between two *quantum bits* can *exceed* this limit, due to *quantum entanglement*.

Quantum Density Matrix

Pure and Mixed States

In quantum mechanics, the state of a system is described by a **state vector** $|\psi\rangle$. However, in many realistic situations, due to our *ignorance*, we are not entirely sure about what the state really is, the state vector $|\psi\rangle$ itself can become a **random variable**.

- **Pure state**: A pure state is described by a *single* state vector $|\psi\rangle$, representing *maximum knowledge* about the system's quantum state.
- **Mixed state**: A mixed state describes a system in which we have only *partial information*, typically as a **statistical mixture of pure states**.

  — Meaning that the state of the system will be *randomly sampled* from an ensemble of states $\mathcal{E} = \{|\psi_1\rangle, |\psi_2\rangle, \ldots\}$, with an *assignment* of a **probability** p_i to each state $|\psi_i\rangle \in \mathcal{E}$, describing the likelihood of the system to be in that state.

Density Matrix

Density matrix (or **density operator**, or **state operator**) provides a unified mathematical description for both *pure* and *mixed* states.

- For a **pure state** $|\psi\rangle$, the corresponding density matrix is

$$\boxed{\rho = |\psi\rangle\langle\psi|,} \tag{24}$$

which is also the **projection operator** of the state.

- For a **mixed state** specified by a *statistical ensemble* of pure states $\{|\psi_i\rangle\}$ with p_i being the *probability* of the system being in the pure state $|\psi_i\rangle$, the corresponding density matrix is

$$\boxed{\rho = \sum_i p_i\, |\psi_i\rangle\langle\psi_i|.} \tag{25}$$

  - By properties of probability, $p_i \geq 0$ and $\sum_i p_i = 1$.
  - If p_i concentrate in one specific state, e.g. $p_1 = 1$ and $p_2 = p_3 = \ldots = 0$, then $\rho = |\psi_1\rangle\langle\psi_1|$ reduces to a pure state.
    ⇒ **Pure state** is a **limit** (ideal case) of **mixed states**.

In general, a density matrix should satisfy the following *defining properties*:

- **Hermitian**: $\rho^\dagger = \rho$.
- **Normalization** (trace identity): $\operatorname{Tr}\rho = 1$.
- **Positive (semi)definite**: $\forall\, |\psi\rangle : \langle\psi|\,\rho\,|\psi\rangle \geq 0$.

Exc 2

> Verify that the mixed state density matrix in Eq. (25) indeed satisfies all the properties above.

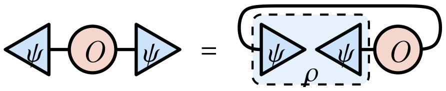
Observable Expectation Value

What is the point of the density matrix?

Motivation: an alternative way to think about the **expectation value** of an **observable** O:

- For pure state $|\psi\rangle$: there are two *equivalent* ways to express $\langle O\rangle$

$$\langle O\rangle = \langle\psi|\, O\, |\psi\rangle = \operatorname{Tr}|\psi\rangle\langle\psi|\, O = \operatorname{Tr}\rho\, O. \tag{26}$$



Comment: Introducing ρ does not seem to be necessary in this case, it is sufficient to work with $|\psi\rangle$.

- For mixed states, what should be the most reasonable approach to calculating $\langle O\rangle$?

with probability :	the system is in the state :	on which $\langle O\rangle$ should be :
p_1	$\lvert\psi_1\rangle$	$\langle\psi_1\rvert\, O\, \lvert\psi_1\rangle$
p_2	$\lvert\psi_2\rangle$	$\langle\psi_2\rvert\, O\, \lvert\psi_2\rangle$
p_3	$\lvert\psi_3\rangle$	$\langle\psi_3\rvert\, O\, \lvert\psi_3\rangle$
$\vdots$	$\vdots$	$\vdots$

therefore, the **ensemble average** must be given by

$$\begin{aligned}\langle O\rangle &= \sum_i p_i \langle\psi_i|\, O\, |\psi_i\rangle \\ &= \operatorname{Tr} \sum_i p_i |\psi_i\rangle\langle\psi_i|\, O.\end{aligned} \tag{27}$$

One can see, the density matrix $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ emerges in the expression exactly as Eq. (25), such that all the *state-dependent information* can be packed into ρ, and the **observable expectation value** can be universally evaluated by

$$\boxed{\langle O\rangle = \operatorname{Tr} \rho\, O,} \tag{28}$$

regardless whether the state ρ is pure or mixed.

- For mixed states, the **density matrix description** is indispensable, because the mixed state density matrix ρ cannot be decomposed into the form of $|\psi\rangle\langle\psi|$ (rank-1 matrix) in general, meaning that there is no state vector description for mixed states.

▪ Quantum State Tomography

How are density matrices determined?

Quantum state tomography refers to the reconstruction of the density matrix of an *unknown state* from *repeated* measurements on *identical* copies of the state.

- For a **single qubit**, by measuring $\langle X\rangle$, $\langle Y\rangle$, $\langle Z\rangle$, the density matrix can be reconstructed as

$$\rho = \frac{1}{2}(\mathbb{1} + \langle X\rangle X + \langle Y\rangle Y + \langle Z\rangle Z). \tag{29}$$

It is also convenient to combine Pauli operators into a vector of operators

$$\boldsymbol{\sigma} = (\sigma^x, \sigma^y, \sigma^z) := (X, Y, Z), \tag{30}$$

such that Eq. (29) can be written in a more compact form as

$$\boxed{\rho = \frac{1}{2}(\mathbb{1} + \langle\boldsymbol{\sigma}\rangle\cdot\boldsymbol{\sigma}).} \tag{31}$$

Exc 3

> Show that Eq. (31) is the density matrix that is consistent with the expectation values and properly normalized.

- For **multi-qubit system**, Eq. (29) can be generalized

$$\boxed{\rho = \frac{1}{2^N}\sum_P \langle P\rangle\, P.} \tag{32}$$

where

- P - sums over all Pauli operators in an N-qubit system (there are 4^N of them, including $\mathbb{1}$)
- $\langle P\rangle$ - the expectation value of the Pauli observable P, inferred from repeated measurement (however, measuring all 4^N observables requires exponentially amount of resources, hence a full quantum state tomography is not feasible for large systems).

Spectral Decomposition

Does density matrix correspond to a physical observable?

As a *Hermitian operator*, **density matrix** also correspond to a physical observable: the **probability** itself.

- Every Hermitian operator admits a **spectral decomposition**, so does the density matrix ρ

$$\boxed{\rho = \sum_i p_i \,|\phi_i\rangle \langle\phi_i|.} \tag{33}$$

- p_i are **eigenvalues** of ρ, with the physical meaning of *probability* for the system to take the $|\phi_i\rangle$ state, satisfying $p_i \geq 0$ and $\sum_i p_i = 1$,
- $|\phi_i\rangle$ are corresponding **eigenvectors** of ρ, forming a set of *orthonormal basis*, also known as the *natural orbitals*.
- Note that p_i and $|\phi_i\rangle$ may not necessary coincide with the mixed state ensemble in Eq. (25) that was originally used to construct the density matrix.

  Example: Consider a mixed state of a single qubit with $1/2$ probability to be $|0\rangle$, and $1/2$ probability to be $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Its density matrix is

$$\begin{aligned}\rho &= \frac{1}{2}\,|0\rangle\langle 0| + \frac{1}{2}\,|+\rangle\langle +| \\ &\doteq \frac{1}{4}\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}.\end{aligned} \tag{34}$$

  However, its eigen decomposition can look very different

$$\rho = p_1\,|\phi_1\rangle\langle\phi_1| + p_2\,|\phi_2\rangle\langle\phi_2|,$$

$$\begin{aligned} p_1 &= \tfrac{2+\sqrt{2}}{4} \approx 0.853553: \quad |\phi_1\rangle \doteq \begin{pmatrix} \frac{\sqrt{2+\sqrt{2}}}{2} \\ \frac{1}{\sqrt{2(2+\sqrt{2})}} \end{pmatrix} \approx \begin{pmatrix} 0.92388 \\ 0.38268 \end{pmatrix}, \\ p_2 &= \tfrac{2-\sqrt{2}}{4} \approx 0.146447: \quad |\phi_2\rangle \doteq \begin{pmatrix} -\frac{1}{\sqrt{2(2+\sqrt{2})}} \\ \frac{\sqrt{2+\sqrt{2}}}{2} \end{pmatrix} \approx \begin{pmatrix} -0.38268 \\ 0.92388 \end{pmatrix}. \end{aligned} \tag{35}$$

Lessons to learn:

- **Ambiguity in Mixed State Interpretation**: *Different* sets of pure states with *different* probabilities can lead to the *same* resulting density matrix, meaning there can be multiple possible *interpretations* of a mixed state as a combination of pure states.
- **Spectral decomposition** of density matrix provides a canonical interpretation of mixed state in terms of mixture of *distinct* (*orthogonal*) pure states.
- Distinction between
  - **Quantum superposition** (linear combination of **state vectors**): $|\psi\rangle = \alpha_1 |\psi_1\rangle + \alpha_2 |\psi_2\rangle + \ldots$ — the result is still a *pure* state.
  - **Statistical mixture** (convex sum of **density matrices**): $\rho = p_1 \rho_1 + p_2 \rho_2 + \ldots$ — the result is generally a *mixed* state.

  Example: An equal amplitude *quantum superposition* of $|0\rangle$ and $|1\rangle$ states is

  $$|+\rangle = \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle). \tag{36}$$

  Its corresponding density matrix is

  $$|+\rangle \langle+| \doteq \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{37}$$

  An equal probability *statistical mixture* of $|0\rangle$ and $|1\rangle$ would be

  $$\rho = \frac{1}{2} |0\rangle \langle 0| + \frac{1}{2} |1\rangle \langle 1| \doteq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{38}$$

  Eq. (37) and Eq. (38) differed by the off-diagonal matrix elements — the signature of **quantum coherence** in the density matrix.

Purity

Given a density matrix, is it pure or mixed?

The **eigenvalues** $\{p_i\}$ obtained from *spectral decomposition* of the **density matrix** ρ provide the key to determining whether the state is pure or mixed:

- **Pure state**: *only one* eigenvalue $p_1 = 1$, and all others are *zero*.

  In this case,

  $$\sum_i p_i^2 = 1 + 0 + 0 + \ldots = 1. \tag{39}$$

- **Mixed state**: *more than one* non-zero eigenvalues, i.e. $p_i > 0$ for multiple i.

  In this case, given that $\sum_i p_i = p_1 + p_2 + \ldots = 1$,

  $$\sum_i p_i^2 = p_1^2 + p_2^2 + \ldots < 1. \tag{40}$$

Purity: to quantify how close the state ρ is to being a pure state,

$$\boxed{\operatorname{Tr}\rho^2 = \sum_i p_i^2} \tag{41}$$

- By construction, $\operatorname{Tr}\rho^2 \in [0, 1]$.
- Purity $\operatorname{Tr}\rho^2$ is **invariant** under *basis transformation* $\rho \to U\,\rho\,U^\dagger$ (which is also seen from the fact that it only has to do with the eigenvalues $\{p_i\}$ not the eigenvectors of ρ)
- Purity provides a *basis-independent* approach to determine if a state ρ is pure or mixed:

$$\rho \text{ is } \begin{cases} \text{pure} & \text{if } \operatorname{Tr}\rho^2 = 1, \\ \text{mixed} & \text{if } \operatorname{Tr}\rho^2 < 1. \end{cases} \tag{42}$$

▪ von Neumann and Rényi Entropy

von Neumann entropy of a density matrix

$$\boxed{S^{(1)} = -\operatorname{Tr}\rho\log\rho.} \tag{43}$$

- Eq. (43) should be understood as

$$S^{(1)} = -\sum_i p_i \log p_i, \tag{44}$$

in terms of the *eigenvalues* p_i of the density matrix ρ. [Note: $0\ln 0$ should be treated as 0 in this calculation]

- This matches the **Shannon entropy** Eq. (16) of a *probability distribution* in the information theory.

HW 1

> Consider a generic single-qubit density matrix of the following form
> $\rho = \frac{1}{2}(\mathbb{1} + \boldsymbol{m}\cdot\boldsymbol{\sigma})$,
> where $\boldsymbol{m}$ is a three-component real vector. Calculate its von Neumann entropy $S^{(1)}$. Show that $S^{(1)} = 0$ when $|\boldsymbol{m}| = 1$, and $S^{(1)} = \log 2$ when $|\boldsymbol{m}| = 0$.

Rényi entropy of a density matrix

$$\boxed{S^{(n)} = \frac{1}{1-n}\log\operatorname{Tr}\rho^n.} \tag{45}$$

In terms of the *eigenvalues* p_i,

$$S^{(n)} = \frac{1}{1-n}\log\sum_i p_i^n. \tag{46}$$

- n is the **Rényi index**.
  - $n = 0$: **max-entropy**, simply counts the log of the Hilbert space dimension $S^{(0)} = \log\dim\mathcal{H}$.
  - $n \to 1$ limit: equivalent to the **von Neumann entropy**, i.e. $S^{(1)} = \lim_{n\to 1} S^{(n)}$.

Exc 4

> Show that in the $n \to 1$ limit, the Rényi entropy reduces to the von Neumann entropy.

- $n = 2$: the **2nd Rényi entropy** is directly related to **purity** by $S^{(2)} = -\log \operatorname{Tr} \rho^2$.
- $n = \infty$: **min-entropy**, lower bound of all Rényi entropies, $S^{(\infty)} = -\log \max_i p_i$.

- The **spectrum** of the **density matrix**, i.e. all *eigenvalues* p_i, can be *reconstructed* from the family of *Rényi entropies* (by solving the following equations, in principle).

$$\sum_i p_i^n = e^{(1-n)\, S^{(n)}} \quad (\text{for } n = 1, 2, \ldots, \dim \mathcal{H}). \tag{47}$$

Entropy and Knowledge

The *Rényi entropy* (including the *von Neumann entropy* as a special case) can characterize how much the *ensemble* is *mixed*.

$$\rho \text{ is } \begin{cases} \text{pure} & \text{if } S^{(n)} = 0, \\ \text{mixed} & \text{if } S^{(n)} > 0, \end{cases} \text{for } n = 1, 2, \ldots. \tag{48}$$

Pure state has **no entropy**. A *pure* state represents the *maximal knowledge* we can have of a system.

Entropy *measures* our **ignorance** about the quantum system. If the ensemble is *pure*, the system is in a *definite* quantum state, hence no entropy. If the ensemble is *mixed*, there are several possible states that the system can take, our *ignorance* is quantified by the *entropy*.

- **Jensen's inequality**: Rényi entropy is generally *decreasing* with the Rényi index,

$$\log \dim \mathcal{H} = S^{(0)} \geq S^{(1)} \geq S^{(2)} \geq \ldots \geq S^{(\infty)} \geq 0. \tag{49}$$

The *equality* is achieved (simultaneously) if all p_i are *equal*.

$$\forall\, i : p_i = \frac{1}{\dim \mathcal{H}} \Rightarrow \forall\, n \geq 0 : S^{(n)} = \log \dim \mathcal{H}. \tag{50}$$

In this case, all *Rényi entropies* reach the *maximum*, and the *ensemble* is **maximally mixed**. The *density matrix* is proportional to *identity matrix* for *maximally mixed* ensemble.

$$\rho = \frac{1}{\dim \mathcal{H}} \mathbb{1}. \tag{51}$$

Any quantum state can be realized with *equal possibility* in a *maximally mixed* ensemble ⇒ we are *completely ignorant* about the system ⇒ *entropy* is therefore *maximized*.

Maximally mixed qubit: SU(2) symmetric, no preferred spin direction, i.e. $\langle \boldsymbol{\sigma} \rangle = 0$. Then according to Eq. (31),

$$\rho = \mathbb{1}/2 \doteq \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{52}$$

- Application: if the qubit basis corresponds to the *left-circular* and *right-circular* **photon polarization**, then the density matrix in Eq. (52) describes the **natural light** ensemble of photons.
- All Rényi entropies are identically $\log 2$ for a maximally mixed qubit,

  $$S^{(n)} = \frac{1}{1-n} \log\left(\frac{1}{2^n} + \frac{1}{2^n}\right) = \log 2 = 1 \text{ bit}. \tag{53}$$

- This is the *maximal entropy* that a qubit could have: our ignorance about a qubit is at most 1 bit. This is why a *qubit* is called a **quantum bit**.

Let us conclude our discussion in the following table:

ensemble	pure	mixed	maximally mixed
entropy	0	⟷	$\log \dim \mathcal{H}$
knowledge	max	⟷	none

Quantum Entanglement

Product and Entangled States

Pure State Entanglement

Consider a *pure* quantum state of a composite system AB, described by the state vector $|\psi_{AB}\rangle$.

- $|\psi_{AB}\rangle$ is a **product state**, if it can be expressed as

  $$|\psi_{AB}\rangle = |\psi_A\rangle \otimes |\psi_B\rangle, \tag{54}$$

  i.e. tensor product of individual states of its subsystems A and B.
- $|\psi_{AB}\rangle$ is a **entangled state**, if it is *not* a product state, i.e.

  $$|\psi_{AB}\rangle \neq |\psi_A\rangle \otimes |\psi_B\rangle. \tag{55}$$

Key characters:

product state	entangled state
The state of one subsystem can be described *independently* of the other.	The subsystems are *correlated* in a way that cannot be described classically.
All information is contained *locally* in each separate subsystem.	Some information is *shared* between subsystems, and can not be accessed from either subsystem.
Measurement of one subsystem does not affect the state of the other subsystem.	Measurement on one subsystem can *instantaneously affect* the state of the other subsystem, a phenomenon known as **quantum nonlocality**.

Examples: For a two-qubit system,

- Suppose $|\psi_A\rangle = z_0\,|0\rangle + z_1\,|1\rangle$, $|\psi_B\rangle = w_0\,|0\rangle + w_1\,|1\rangle$, the most general *product state* must take the form of

$$\begin{aligned} &|\psi_A\rangle \otimes |\psi_B\rangle = (z_0\,|0\rangle + z_1\,|1\rangle) \otimes (w_0\,|0\rangle + w_1\,|1\rangle) \\ &= z_0\,w_0\,|00\rangle + z_0\,w_1\,|01\rangle + z_1\,w_0\,|10\rangle + z_1\,w_1\,|11\rangle. \end{aligned} \tag{56}$$

- The following states, known as **Bell states**, are *entangled states*:

$$\begin{aligned} &|\psi_{AB}^{++}\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}},\ |\psi_{AB}^{+-}\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}, \\ &|\psi_{AB}^{-+}\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}},\ |\psi_{AB}^{--}\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}. \end{aligned} \tag{57}$$

Exc 5

> Prove that states in Eq. (57) can not be written as product states like Eq. (56).

Question: Is the state $\frac{1}{2}\,(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$ entangled or not?

It is not obvious to see if a state is entangled or not ⇒ we need to develop *measures of entanglement*, such that by measuring these quantities, we can decide how much the state is entangled… (to be discussed later).

□ Mixed State Entanglement

For mixed states, the composite system AB is described by its density matrix ρ_{AB}

- ρ_{AB} is a **product state**, if the it can be written as tensor product of density matrices of its subsystems

$$\rho_{AB} = \rho_A \otimes \rho_B. \tag{58}$$

- ρ_{AB} is a **separable state**, if the it is a statistical mixture of product states

$$\rho_{AB} = \sum_i p_i\, \rho_A^{(i)} \otimes \rho_B^{(i)}. \tag{59}$$

- ρ_{AB} is an **entangled state**, if it is not separable, i.e.

$$\rho_{AB} \neq \sum_i p_i\, \rho_A^{(i)} \otimes \rho_B^{(i)}, \tag{60}$$

  - This is a subtle point: sometimes a density matrix may *appear entangled*, but there might exists a non-trivial *decomposition* that reveals the state is actually *separable*.
  - Determining whether a mixed state is separable or entangled is a computationally challenging task, and is known to be NP-hard!

Reduced Density Matrix

The **reduced density matrix** ρ_A provides an *effective description* of the state of the subsystem A, derived from the full density matrix ρ_{AB} of the composite system.

$$\boxed{\rho_A = \mathrm{Tr}_B\, \rho_{AB},} \tag{61}$$

where Tr_B denotes the **partial trace** over subsystem B.

- Mathematical definition of partial trace:

$$\mathrm{Tr}_B\, \rho_{AB} = \sum_i \langle i|_B\, \rho_{AB}\, |i\rangle_B, \tag{62}$$

  where $\{|i\rangle_B\}$ is any set of orthonormal basis for subsystem B.

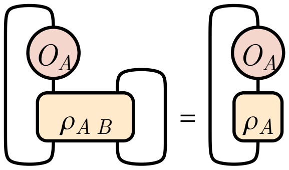
- Physical meaning: taking partial trace ignores the information related to subsystem B while retaining the information remaining in subsystem A. It corresponds to “tracing out” or averaging over the unobserved degrees of freedom of B.

Why is ρ_A an effective description? Why the partial trace?

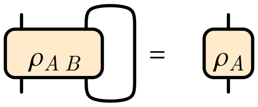
ρ_A captures the state of subsystem A in the sense that it encodes all the information needed to compute *expectation values* of observables acting *only* in A.

- *Existence* of ρ_A: For any observable O_A acting only in A, ρ_A must provide an equivalent way to evaluate $\langle O_A\rangle$ as using ρ_{AB}:

$$\langle O_A\rangle = \mathrm{Tr}(\rho_{AB}\, O_A \otimes I_B) = \mathrm{Tr}(\rho_A\, O_A). \tag{63}$$



For Eq. (63) to hold for any choice of O_A, we should require



which provides a construction: $\rho_A = \mathrm{Tr}_B\, \rho_{AB}$.

- *Uniqueness* of ρ_A: For any different state $\rho'_A \neq \rho_A$, there must exist at least one observable O_A, such that $\mathrm{Tr}\, \rho'_A\, O_A \neq \mathrm{Tr}\, \rho_A\, O_A$. By matching all observable expectation values $\langle O_A \rangle$ in Eq. (63), the state ρ_A is uniquely determined, meaning that $\rho_A = \mathrm{Tr}_B\, \rho_{AB}$ is the only valid construction.

▪ Entanglement Entropy

The **entanglement entropy** of the subsystem A in a joint state $|\psi_{AB}\rangle$ is given by

$$\boxed{S^{(1)}(A) = -\mathrm{Tr}\, \rho_A \log \rho_A.} \tag{64}$$

where ρ_A is the **reduced density matrix** of subsystem A obtained by *tracing out* subsystem B in the full **density matrix** $\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$, following the general approach in Eq. (61),

$$\boxed{\rho_A = \mathrm{Tr}_B\, |\psi_{AB}\rangle\langle\psi_{AB}|.} \tag{65}$$

One may also define a more general *Rényi version* as

$$\boxed{S^{(n)}(A) = \frac{1}{1-n} \log \mathrm{Tr}\, \rho_A^n,} \tag{66}$$

such that $S^{(1)}(A) = \lim_{n\to 1} S^{(n)}(A)$.

Question: why do we care about subsystem A only, what about B?

It turns out, as long as ρ_{AB} is *pure*,

$$S^{(n)}(A) = S^{(n)}(B), \tag{67}$$

meaning that it doesn't matter whether you quantify the *entanglement* between A and B by looking at the *entropy* in A or in B — they are the same.

Example I: take a **spin-singlet state** of two spin-1/2 particles,

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle). \tag{68}$$

- Full density matrix

$$|\psi\rangle\langle\psi| \doteq \frac{1}{2} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 & 0 \end{pmatrix} = \frac{1}{2} \left(\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ \hline 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right). \tag{69}$$

- *Partial trace* over qubit $B \Rightarrow$ *reduced density matrix* of qubit A

$$\begin{aligned}\rho_A &= \mathrm{Tr}_B\,|\psi\rangle\,\langle\psi| \\ &\doteq \frac{1}{2}\begin{pmatrix} \mathrm{tr}\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix} & \mathrm{tr}\begin{pmatrix}0 & 0\\ -1 & 0\end{pmatrix} \\ \mathrm{tr}\begin{pmatrix}0 & -1\\ 0 & 0\end{pmatrix} & \mathrm{tr}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.\end{aligned}$$

Note that ρ_A indeed describes a *maximally mixed* qubit.

- Compute the *entropy* of the *reduced density matrix*,

$$S(A) = -\mathrm{Tr}\,\rho_A \ln \rho_A = \ln 2 = 1\,\text{bit}. \tag{71}$$

Example II: take the **product state** of two spin-1/2 particles,

$$|\psi\rangle = \frac{1}{2}\,(|\uparrow\uparrow\rangle + |\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle + |\downarrow\downarrow\rangle). \tag{72}$$

- Full density matrix

$$\rho = |\psi\rangle\,\langle\psi| \doteq \frac{1}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix}\begin{pmatrix}1 & 1 & 1 & 1\end{pmatrix} = \frac{1}{4}\left(\begin{array}{cc|cc} 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1\\ \hline 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1\end{array}\right). \tag{73}$$

- *Partial trace* over qubit $B \Rightarrow$ *reduced density matrix* of qubit A

$$\begin{aligned}\rho_A &= \mathrm{Tr}_B\,\rho \\ &\doteq \frac{1}{4}\begin{pmatrix} \mathrm{tr}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} & \mathrm{tr}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} \\ \mathrm{tr}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} & \mathrm{tr}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}.\end{aligned} \tag{74}$$

- Compute the *entropy* of the *reduced density matrix*,

$$S(A) = -\mathrm{Tr}\,\rho_A \ln \rho_A = -(0 \ln 0 + 1 \ln 1) = 0\,\text{bit}. \tag{75}$$

Conclusion: The **entanglement entropy** characterizes the amount of **quantum entanglement** between subsystem A and its complement $\overline{A}$ (which is B here), given that the full system $A \cup \overline{A}$ is *pure*.

$\lvert\psi\rangle$ (pure)	product	entangled	maximally entangled
ρ_A	pure	mixed	maximally mixed
$S^{(n)}(A)$	0	$\longleftrightarrow$	$\ln \dim \mathcal{H}$
entanlement	none	$\longleftrightarrow$	max

For diagnostic purpose (to distinguish product state from entangled state), any *Rényi index* $n = 1, 2, \ldots$ will work.

Why entropy provides a measure of entanglement? Quantum entanglement: the *nonlocal* nature of *quantum information* in an *entangled* state (i.e. information shared jointly among subsystems) ⇒ separating out a subsystem would lead to *lost* of *information* ⇒ hence the *produc-*

tion of (entanglement) *entropy*.

Open questions: The *system* must be *pure*, otherwise there are other source of entropy productions. What about entanglement in a *mixed* state? Good to describe *bipartite* entanglement. What about *multipartite* entanglement?

Mutual Information

The **mutual information** between qubit A and qubit B is

$$\boxed{I(A:B) = S(A) + S(B) - S(A \cup B).} \tag{76}$$

Or more generally, one may define the *Rényi version*,

$$\boxed{I^{(n)}(A:B) = S^{(n)}(A) + S^{(n)}(B) - S^{(n)}(A \cup B).} \tag{77}$$

$I^{(n)}(A:B) =$ the amount of *information* shared by A and B.

Example: take the **spin-singlet state**, we have

$$\begin{aligned} &S^{(n)}(A) = S^{(n)}(B) = 1\,\text{bit}, \\ &S^{(n)}(A \cup B) = 0\,\text{bit}, \end{aligned} \tag{78}$$

hence 2 bit mutual information (regardless of the Rényi index n)

$$I^{(n)}(A:B) = S^{(n)}(A) + S^{(n)}(B) - S^{(n)}(A \cup B) = 2\,\text{bit}. \tag{79}$$

This is a surprising result!

- For classical systems, the *mutual information* between two *classical bits* will never exceed 1 bit. How can we tell more than 1 bit of information about B by measuring A?
- The maximal mutual information between two classical bits is achieved when they are perfectly correlated, e.g.

$$p(10) = p(01) = 1/2,\ \ p(11) = p(00) = 0. \tag{80}$$

- **Entanglement** is *more* than **correlation**: the *extra bit* of quantum information *shared* between qubits A and B is their *quantum entanglement*, that goes beyond the classical correlation.

For a two-qubit system, the *2nd Rényi* ($n = 2$) mutual information $I^{(2)}(A:B)$ between the two qubits is related to the *spin observables* in a relatively simple way

$$I^{(2)}(A:B) = \log\left(1 + \frac{\| \langle \boldsymbol{\sigma}_A \otimes \boldsymbol{\sigma}_B \rangle \|^2 - \| \langle \boldsymbol{\sigma}_A \rangle \|^2 \ \| \langle \boldsymbol{\sigma}_B \rangle \|^2}{\left(1 + \| \langle \boldsymbol{\sigma}_A \rangle \|^2\right)\left(1 + \| \langle \boldsymbol{\sigma}_B \rangle \|^2\right)}\right). \tag{81}$$

Note: $\| \langle \boldsymbol{\sigma}_A \otimes \boldsymbol{\sigma}_B \rangle \|^2 = \sum_{i,j=x,y,z} \left\langle \sigma_A^i \otimes \sigma_B^j \right\rangle^2$ and $\| \langle \boldsymbol{\sigma}_A \rangle \|^2 = \sum_{i=x,y,z} \left\langle \sigma_A^i \right\rangle^2$.

- **Classical state**: *statistical* superposition

$$\rho = \frac{1}{2} \left| \uparrow \downarrow \right\rangle \left\langle \uparrow \downarrow \right| + \frac{1}{2} \left| \downarrow \uparrow \right\rangle \left\langle \downarrow \uparrow \right|, \tag{82}$$

- Observables

$$\langle \boldsymbol{\sigma}_A \rangle = \langle \boldsymbol{\sigma}_B \rangle = (0, 0, 0), \\ \langle \boldsymbol{\sigma}_A \otimes \boldsymbol{\sigma}_B \rangle = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{83}$$

- Mutual information

$$I^{(2)}(A:B) = \log\left(1 + \| \langle \boldsymbol{\sigma}_A \otimes \boldsymbol{\sigma}_B \rangle \|^2\right) = \log(1+1) = \log 2 = 1 \text{ bit}. \tag{84}$$

- **Quantum state**: *quantum* superposition

$$\rho = |\psi\rangle \langle\psi|, \\ |\psi\rangle = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle). \tag{85}$$

- Observables

$$\langle \boldsymbol{\sigma}_A \rangle = \langle \boldsymbol{\sigma}_B \rangle = (0, 0, 0), \\ \langle \boldsymbol{\sigma}_A \otimes \boldsymbol{\sigma}_B \rangle = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{86}$$

- Mutual information

$$I^{(2)}(A:B) = \log\left(1 + \| \langle \boldsymbol{\sigma}_A \otimes \boldsymbol{\sigma}_B \rangle \|^2\right) = \log(1+3) = \log 4 = 2 \text{ bit}. \tag{87}$$

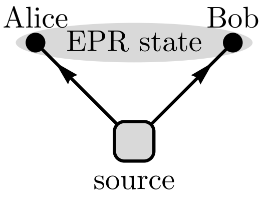
In a spin-singlet state, not only $\sigma_A^z \sigma_B^z$ is perfectly correlated, but $\sigma_A^x \sigma_B^x$ and $\sigma_A^y \sigma_B^y$ are *also* perfectly correlated. Such additional correlations (by changing *measurement basis*) can not be realized by classical bits. The additional information channel enables the *two-qubit* system to store all its *two bits* of *quantum information* purely as *shared information* between qubits, without using any “local access” of information.

EPR Pair and Bell Inequality

The **Bell states**, or the Einstein-Podolsky-Rosen **(EPR) pair states**, refers to two qubits (spins) in *maximally entangled pure* states. The **spin-singlet state** in Eq. (85) is one such example,

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle). \tag{88}$$

Suppose a machine can repeatedly *prepare* **spin-singlet EPR pairs** and *distribute* the spins separately to Alice and Bob,



Alice and Bob can measure their own spins and record the measurement outcome. After the measurement, the pair of spins are discarded. New EPR pairs will be acquired from the source.

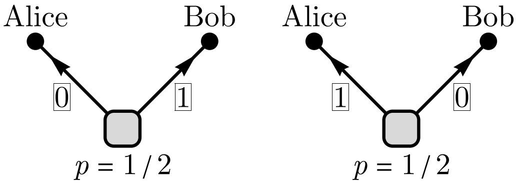
Let $\boldsymbol{\sigma}_A$ (or $\boldsymbol{\sigma}_B$) be the spin observable of Alice's (or Bob's) spin. On the spin-singlet state, their expectation values are

$$\begin{aligned}&\langle\boldsymbol{\sigma}_A\rangle = \langle\boldsymbol{\sigma}_B\rangle = (0, 0, 0),\\ &\langle\boldsymbol{\sigma}_A \otimes \boldsymbol{\sigma}_B\rangle = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.\end{aligned} \tag{89}$$

If Alice and Bob both measure σ^z, they will find

$$\sigma_A^z = -\sigma_B^z = \begin{cases} +1 & p = 1/2 \\ -1 & p = 1/2 \end{cases}. \tag{90}$$

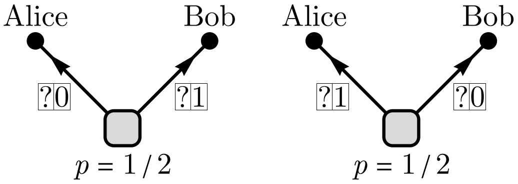
- *Quantum* explanation: can be inferred from $\langle\sigma_A^z\rangle = \langle\sigma_B^z\rangle = 0$ and $\langle\sigma_A^z\,\sigma_B^z\rangle = -1$.
- This is not too surprising: just a perfect correlation between two random variables. *Classically*, one may model the perfect correlation by a **hidden variable**:



If Alice and Both both measure σ^x, they will find

$$\sigma_A^x = -\sigma_B^x = \begin{cases} +1 & p = 1/2 \\ -1 & p = 1/2 \end{cases}. \tag{91}$$

- *Quantum* explanation: can be inferred from $\langle\sigma_A^x\rangle = \langle\sigma_B^x\rangle = 0$ and $\langle\sigma_A^x\,\sigma_B^x\rangle = -1$.
- To model this *classically*: we will need to introduce *another* hidden variable to encode the perfect correlation in σ^x channel.



As Alice and Bob can choose to measure either σ^z or σ^x at their *free will* ⇒ *Classically*, both hidden variables about σ^z and σ^x must be sent with the qubit. (Although a single state vector

$|\psi\rangle$ is sufficient to explain all situations in the quantum way).

If Alice measures σ_A^z and Bob measures σ_B^x, they will find independently that

$$\sigma_A^z = \begin{cases} +1 & p = 1/2 \\ -1 & p = 1/2 \end{cases}, \; \sigma_B^x = \begin{cases} +1 & p = 1/2 \\ -1 & p = 1/2 \end{cases}. \tag{92}$$

- *Quantum* explanation: can be inferred from $\langle \sigma_A^z \rangle = \langle \sigma_B^x \rangle = 0$ and $\langle \sigma_A^z \, \sigma_B^x \rangle = 0$.
- The *classical* hidden variables can reproduce this behavior only if they follow the joint distribution

Alice	Bob	p
00	11	$1/4$
01	10	$1/4$
10	01	$1/4$
11	00	$1/4$

(93)

So far so good. But Alice and Bob can also decide to measure σ^y, or more generally, *any* linear combination of their observables … What if Alice measures $\boldsymbol{n}_A \cdot \boldsymbol{\sigma}_A$ and Bob measures $\boldsymbol{n}_B \cdot \boldsymbol{\sigma}_B$? (where $\boldsymbol{n}_A$ and $\boldsymbol{n}_B$ are *unit* vectors) Their outcomes will follow the joint distribution

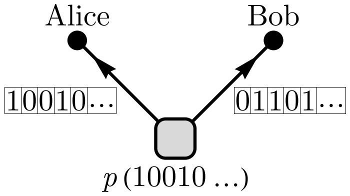
$\boldsymbol{n}_A \cdot \boldsymbol{\sigma}_A$	$\boldsymbol{n}_B \cdot \boldsymbol{\sigma}_B$	p
$+1$	$+1$	$(1 - \boldsymbol{n}_A \cdot \boldsymbol{n}_B)/4$
$+1$	-1	$(1 + \boldsymbol{n}_A \cdot \boldsymbol{n}_B)/4$
-1	$+1$	$(1 + \boldsymbol{n}_A \cdot \boldsymbol{n}_B)/4$
-1	-1	$(1 - \boldsymbol{n}_A \cdot \boldsymbol{n}_B)/4$

(94)

The probability that Alice and Bob obtain *opposite* outcomes is

$$p(\boldsymbol{n}_A \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_B \cdot \boldsymbol{\sigma}_B) = \frac{1 + \boldsymbol{n}_A \cdot \boldsymbol{n}_B}{2}. \tag{95}$$

- *Quantum* explanation: can be inferred from $\langle \boldsymbol{n}_A \cdot \boldsymbol{\sigma}_A \rangle = \langle \boldsymbol{n}_B \cdot \boldsymbol{\sigma}_B \rangle = 0$ and $\langle \boldsymbol{n}_A \cdot \boldsymbol{\sigma}_A \, \boldsymbol{n}_B \cdot \boldsymbol{\sigma}_B \rangle = -\boldsymbol{n}_A \cdot \boldsymbol{n}_B$.
- *Classically*, to reproduce all these, we will need *many* (could be infinitely many) hidden variables, one for each choice of the measurement axis $\boldsymbol{n}$. (This is ugly but not fatal yet.)



There should be complicated *correlation* among *hidden variables* in an *attempt* to match quantum predictions (but the attempt may fail). Suppose two of the hidden variables happen to determine the outcome of $\boldsymbol{n}_1 \cdot \boldsymbol{\sigma}$ and $\boldsymbol{n}_2 \cdot \boldsymbol{\sigma}$. After *marginalizing* (summing) over all the other hidden variables, the marginal distribution should be

Alice	Bob	p
...00...	...11...	$(1+\boldsymbol{n}_1\cdot\boldsymbol{n}_2)/4$
...01...	...10...	$(1-\boldsymbol{n}_1\cdot\boldsymbol{n}_2)/4$.
...10...	...01...	$(1-\boldsymbol{n}_1\cdot\boldsymbol{n}_2)/4$
...11...	...00...	$(1+\boldsymbol{n}_1\cdot\boldsymbol{n}_2)/4$

(96)

Now consider Alice and Bob can choose to measure any one of the *three* observables $\boldsymbol{n}_1\cdot\boldsymbol{\sigma}$, $\boldsymbol{n}_2\cdot\boldsymbol{\sigma}$ and $\boldsymbol{n}_3\cdot\boldsymbol{\sigma}$ (on their own qubits respectively, where $\boldsymbol{n}_{1,2,3}$ are *unit* vectors).

- *Classically*, there must be *three* hidden variables associated with the *three* observables, following some marginal distribution

Alice	Bob	p
...000...	...111...	p_1
...001...	...110...	p_2
...010...	...101...	p_3
...011...	...100...	p_4 .
...100...	...011...	p_5
...101...	...010...	p_6
...110...	...001...	p_7
...111...	...000...	p_8

(97)

The probability must sum up to 1, i.e.

$$p_1+p_2+\ldots+p_8=1. \tag{98}$$

- If Alice measures $\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_A$ and Bob measures $\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_B$, the probability that they obtain *opposite* outcomes is

$$p(\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_A=-\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_B)=p_1+p_2+p_7+p_8. \tag{99}$$

- If Alice measures $\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_A$ and Bob measures $\boldsymbol{n}_3\cdot\boldsymbol{\sigma}_B$, the probability that they obtain *opposite* outcomes is

$$p(\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_A=-\boldsymbol{n}_3\cdot\boldsymbol{\sigma}_B)=p_1+p_4+p_5+p_8. \tag{100}$$

- If Alice measures $\boldsymbol{n}_3\cdot\boldsymbol{\sigma}_A$ and Bob measures $\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_B$, the probability that they obtain *opposite* outcomes is

$$p(\boldsymbol{n}_3\cdot\boldsymbol{\sigma}_A=-\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_B)=p_1+p_3+p_6+p_8. \tag{101}$$

Put together,

$$\begin{aligned}&p(\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_A=-\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_B)+p(\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_A=-\boldsymbol{n}_3\cdot\boldsymbol{\sigma}_B)+p(\boldsymbol{n}_3\cdot\boldsymbol{\sigma}_A=-\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_B)\\&=3\,p_1+p_2+p_3+p_4+p_5+p_6+p_7+3\,p_8\\&=1+2\,p_1+2\,p_8\end{aligned} \tag{102}$$

This leads to a (version of) **Bell inequality**.

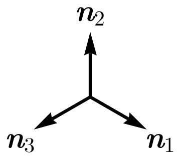
$$\boxed{p(\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_A=-\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_B)+p(\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_A=-\boldsymbol{n}_3\cdot\boldsymbol{\sigma}_B)+p(\boldsymbol{n}_3\cdot\boldsymbol{\sigma}_A=-\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_B)\geq 1.} \tag{103}$$

- Now what is the **quantum mechanical prediction**? Recall the *quantum* result in Eq. (95), the Bell inequality would require

$$\frac{1+\boldsymbol{n}_1\cdot\boldsymbol{n}_2}{2}+\frac{1+\boldsymbol{n}_2\cdot\boldsymbol{n}_3}{2}+\frac{1+\boldsymbol{n}_3\cdot\boldsymbol{n}_1}{2}\geq 1, \tag{104}$$

  for three unit vectors $\boldsymbol{n}_1$, $\boldsymbol{n}_2$ and $\boldsymbol{n}_3$.

Consider a special case, where the three vectors are 120° to each other in a plane.



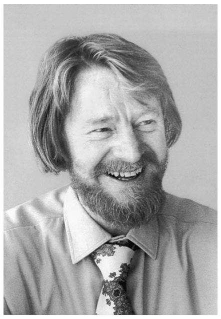
$$\boldsymbol{n}_1\cdot\boldsymbol{n}_2=\boldsymbol{n}_2\cdot\boldsymbol{n}_3=\boldsymbol{n}_3\cdot\boldsymbol{n}_1=-1/2. \tag{105}$$

Then Eq. (104) would require

$$\frac{1}{4}+\frac{1}{4}+\frac{1}{4}=\frac{3}{4}\not\geq 1, \tag{106}$$

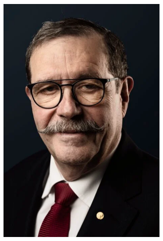
which is violates Bell inequality.

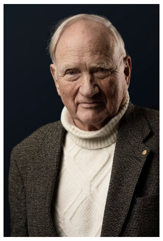
The *violation* of Bell inequality indicates that no classical model of *local hidden variables* can ever reproduce all the predictions of quantum mechanics. This is the **Bell's theorem**.



John S. Bell (1928–1990)

The Nobel (No-Bell) Prize in Physics 2022





Alain Aspect John F. Clauser Anton Zeilinger

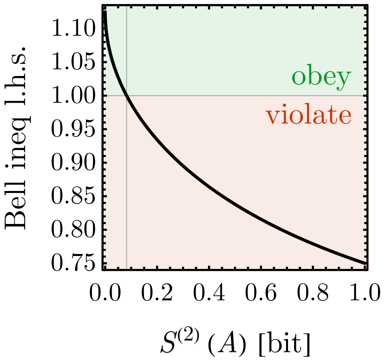
for experiments with entangled photons,
establishing the violation of Bell inequalities
and pioneering quantum information science

How strictly does Bell inequality tell us about entanglement?

HW 2

Consider a two-spin state $|\psi\rangle = \cos\,\alpha\,|\uparrow\downarrow\rangle - \sin\,\alpha\,|\downarrow\uparrow\rangle$, where $\alpha \in [0, \pi/4]$ is a phase angle to tune $|\psi\rangle$ from a product state to a maximally entangled state.
(i) Calculate the 2nd Rényi entanglement entropy $S^{(2)}(A)$ of qubit A (as a function of α).
(ii) Evaluate $\langle\psi|\,\boldsymbol{\sigma}_A\,|\psi\rangle$, $\langle\psi|\,\boldsymbol{\sigma}_B\,|\psi\rangle$ and $\langle\psi|\,\boldsymbol{\sigma}_A \otimes \boldsymbol{\sigma}_B\,|\psi\rangle$.
(iii) Let $\boldsymbol{n}_1$, $\boldsymbol{n}_2$, $\boldsymbol{n}_3$ be three unit vectors 120° to each other in the xz-plane, evaluate the left-hand-side of the Bell inequality
$p(\boldsymbol{n}_1 \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_2 \cdot \boldsymbol{\sigma}_B) + p(\boldsymbol{n}_2 \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_3 \cdot \boldsymbol{\sigma}_B) + p(\boldsymbol{n}_3 \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_1 \cdot \boldsymbol{\sigma}_B)$
as a function of α.

We can plot the l.h.s. of the Bell inequality v.s. the 2nd Rényi entanglement entropy for different α:



- For *pure* state, such as $|\psi\rangle$ in the above example, entanglement entropy $S^{(2)}(A) > 0 \Leftrightarrow$ the state is *entangled*. But the Bell inequality is *not always* violated. ⇒ It is an **entanglement witness**.
- For *mixed* state, entropy no longer provides a good measure of quantum entanglement. We had to rely on Bell inequalities and other entanglement witness.

Quantum Information Processing

Unitary Operations

Time Evolution

How does a quantum state evolve over time?

- **Pure state**: the *state vector* $|\psi(t)\rangle$ encodes complete information about the quantum system at time t. If $|\psi(0)\rangle$ is know at $t = 0$, there must be a *deterministic* rule to find $|\psi(t)\rangle$ at any time t:

$$\boxed{|\psi(t)\rangle = U(t)\,|\psi(0)\rangle,} \tag{107}$$

where $U(t)$ is the **time-evolution operato**r.

- **Mixed state**: the *density matrix* $\rho(t)$ describes a statistical ensemble of pure states $\{|\psi_i\rangle\}$ with probabilities $\{p_i\}$

$$\rho(t) = \sum_i p_i\,|\psi_i(t)\rangle\,\langle\psi_i(t)|. \tag{108}$$

  - Assumption: the probabilities $\{p_i\}$ remains unchanged under *deterministic* time-evolution. (The probability transfer process is stochastic.)
  - Since every pure state $|\psi_i(t)\rangle$ evolves independently under Eq. (107), the density matrix evolves as

$$\boxed{\rho(t) = U(t)\,\rho(0)\,U(t)^\dagger.} \tag{109}$$

Unitarity: the time-evolution operator $U(t)$ of an *isolated* quantum system must be *unitary*, to ensure **no information is lost** during the evolution (both forward and backward in time).

- Distinct states remain distinct:

$$\langle\phi(0)|\psi(0)\rangle = 0 \Rightarrow \langle\phi(t)|\psi(t)\rangle = \langle\phi(0)|\,U(t)^\dagger\,U(t)\,|\psi(0)\rangle = 0. \tag{110}$$

- Identical states remain identical:

$$\langle\psi(0)|\psi(0)\rangle = 1 \Rightarrow \langle\psi(t)|\psi(t)\rangle = \langle\psi(0)|\,U(t)^\dagger\,U(t)\,|\psi(0)\rangle = 1. \tag{111}$$

Let $\{|i\rangle\}$ be a set of *orthonormal basis*, i.e. $\langle i|j\rangle = \delta_{ij}$, Eq. (110) and Eq. (111) imply:

$$\langle i|\,U(t)^\dagger\,U(t)\,|j\rangle = \delta_{ij} \Leftrightarrow U\,(t)^\dagger\,U(t) = \mathbb{1}, \tag{112}$$

which, by definition, proves that $U(t)$ is unitary.

▪ Dynamic Equations

Every *unitary* operator is generated by some *Hermitian* operator. The Hermitian generator of the unitary time-evolution operator is called the **Hamiltonian** H, representing the energy observable of the quantum system

$$\boxed{U(t) = \exp\left(-\frac{i}{\hbar}\,H\,t\right).} \tag{113}$$

- **Schrödinger Equation** describes the time-evolution dynamics of a pure state

$$\boxed{i\,\hbar\,\partial_t|\psi(t)\rangle = H\,|\psi(t)\rangle.} \tag{114}$$

Exc 6

> Verify that Eq. (114) is consistent with Eq. (107) given Eq. (113).

- **von Neumann Equation** describes the time-evolution dynamics of a mixed state

$$\boxed{i\,\hbar\,\partial_t \rho(t) = [H, \rho(t)],} \tag{115}$$

where $[H, \rho] = H\,\rho - \rho\,H$ denotes the commutator.

Exc 7

> Verify that Eq. (115) is consistent with Eq. (109) given Eq. (113).

Example: single-qubit system governed by the Hamiltonian

$$H = \frac{\omega}{2} Z \doteq \frac{\omega}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{116}$$

- Initial density matrix (in the diagonal basis of H)

$$\rho(0) \doteq \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix}. \tag{117}$$

- Time-evolved density matrix (set $\hbar = 1$)

$$\rho(t) \doteq \begin{pmatrix} \rho_{00} & \rho_{01}\, e^{-i\,\omega\,t} \\ \rho_{10}\, e^{i\,\omega\,t} & \rho_{11} \end{pmatrix}. \tag{118}$$

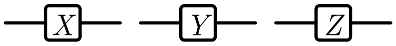
The **diagonal** elements are *invariant*, the **off-diagonal** elements *rotates* in time following $e^{\pm i\,\omega\,t}$ (with an angular frequency of ω).

▪ Quantum Circuits

Quantum computation leverages unitary evolution to manipulate quantum states and process quantum information. These unitary operations are realized through **quantum circuits**, which consist of **quantum gates** — the building block that apply unitary transformations to one or more qubits.

- **Single-Qubit Gates**: acts on a single qubit, represented as 2×2 unitary matrices.
  - **Pauli Gates**: implement bit flip or phase flip

$$X \doteq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\; Y \doteq \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\; Z \doteq \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{119}$$

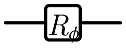


  - **Hadamard Gate**: creates superpositions, transforms between X and Z basis.

$$H \doteq \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{120}$$

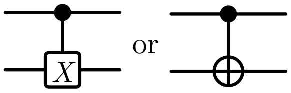
  - **Phase Gate**: introduces a relative phase rotation between $|0\rangle$ and $|1\rangle$.

$$R_\phi \doteq \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}. \tag{121}$$



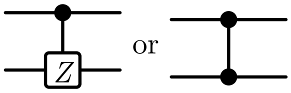
- **Two-Qubit Gates**: acts on two qubits, represented by 4 × 4 unitary matrices.
  - **CNOT Gate**: the first (control) qubit determines whether the second (target) qubit is bit flipped.

$$\text{CNOT} \doteq \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right). \tag{122}$$



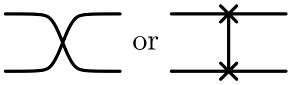
  - **CZ Gate**: the first (control) qubit determines whether the second (target) qubit is phase flipped.

$$\text{CZ} \doteq \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{array}\right). \tag{123}$$



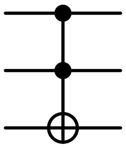
  - **SWAP Gate**: interchange two qubits.

$$\text{SWAP} \doteq \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right). \tag{124}$$



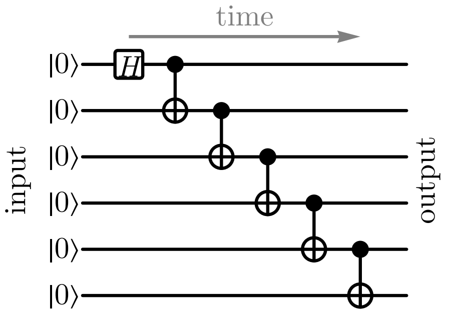
- **Three-Qubit Gates**: acts on three qubits, represented by 8 × 8 unitary matrices.
  - **Toffoli Gate** (Controlled-controlled-NOT): the first two (control) qubits jointly determines whether the last (target) qubit is bit flipped.

$$\text{CCNOT} \doteq \left(\begin{array}{cc:cc:cc:cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right).$$



This gate is often used in universal quantum computation.

Quantum gates can be combined to build **quantum circuits**. For example, the following quantum circuit



prepares a **Greenberger-Horne-Zeilinger (GHZ) state**, also known as the **Schrödinger cat state**:

$$|\text{GHZ}\rangle = \frac{|000\ldots0\rangle + |111\ldots1\rangle}{\sqrt{2}}. \tag{126}$$

Quantum circuits are graphically represented as **quantum circuit diagrams**, where

- **Wires**:
  - Horizontal lines represents qubits.
  - Time flows from left to right.
- **Gates**: Placed along the wires to represent unitary operations acting on the corresponding qubits.

Universal gate sets — a set of quantum gates is *universal* if any unitary operation can be *approximated* to *arbitrary accuracy* using only gates from the set. There are multiple choices:

- $\{H, R_\phi\} \cup \{\text{CNOT}\}$: Single-qubit rotation + entanglement
- $\{H, S = R_{\pi/2}, \text{CNOT}\} \cup \{T = R_{\pi/4}\}$: Clifford gates + magic

- $\{\text{CCNOT}\} \bigcup \{H\}$: Classical universality + quantum superposition

Quantum Measurements

Observables

In quantum mechanics, **observables** (physical quantities) are represented by *Hermitian operators* L, satisfying $L = L^\dagger$. Any Hermitian operator L admits a **spectral decomposition**:

$$\boxed{L = \sum_i |\lambda_i\rangle\, \lambda_i\, \langle\lambda_i|,} \tag{127}$$

where

- the *eigenvalues* $\lambda_i \in \mathbb{R}$ are real numbers, representing the possible **measurement outcomes**;
- the corresponding *eigen states* are $|\lambda_i\rangle$, forming an *orthonormal basis* $\langle\lambda_i|\lambda_j\rangle = \delta_{ij}$, also called the **measurement basis** of L.

Projective Measurement

Measurement of a Hermitian observable L is described by a set of *projection operators* P_λ, labeled by the observation value λ,

$$\boxed{P_\lambda = \sum_{\lambda_i=\lambda} |\lambda_i\rangle\, \langle\lambda_i|.} \tag{128}$$

- Hermiticity: $P_\lambda^\dagger = P_\lambda$,
- Orthogonality:

  $$P_\lambda\, P_{\lambda'} = \delta_{\lambda\lambda'}\, P_\lambda. \tag{129}$$

  In particular, $P_\lambda^2 = P_\lambda$ is the defining property of projection operators.
- Normalization condition:

  $$\sum_\lambda P_\lambda = \mathbb{1}. \tag{130}$$

Exc 8 Show that the projection operators in Eq. (128) satisfy Eq. (129) and Eq. (130).

- Projective measurement on **pure states**:
  - Pre-measurement state: $|\psi\rangle$.
  - Probability distribution: Measure L, the probability to observe $L = \lambda$ is

$$\boxed{p(\lambda) = \langle\psi|\, P_\lambda\, |\psi\rangle.} \tag{131}$$

  - Post-measurement state (with post-selection):

$$\boxed{|\psi\rangle \xrightarrow[\text{observe } \lambda]{\text{measure } L} |\psi'\rangle = \frac{P_\lambda |\psi\rangle}{\sqrt{\langle\psi| P_\lambda |\psi\rangle}}.} \tag{132}$$

- Projective measurement on **mixed states**:
  - Pre-measurement state: ρ.
  - Probability distribution: Measure L, the probability to observe $L = \lambda$ is

$$\boxed{p(\lambda) = \operatorname{Tr} \rho\, P_\lambda.} \tag{133}$$

  - Post-measurement state (with post-selection):

$$\boxed{\rho \xrightarrow[\text{observe } \lambda]{\text{measure } L} \rho' = \frac{P_\lambda\, \rho\, P_\lambda}{\operatorname{Tr} \rho\, P_\lambda}.} \tag{134}$$

Post-selection refers to selecting out the quantum system after a specific measurement outcome $L = \lambda$ is observed, effectively discarding cases of all other possible outcomes.

- With post-selection: the system will *collapse* to the specific post-measurement state

$$\rho \xrightarrow[\text{observe } \lambda]{\text{measure } L} \rho' = \frac{P_\lambda\, \rho\, P_\lambda}{\operatorname{Tr} \rho\, P_\lambda} \tag{135}$$

- Without post-selection: the system will be in a *statistical mixture* of all possible post-measurement states

$$\rho \xrightarrow[\text{ignore outcome}]{\text{measure } L} \rho' = \sum_\lambda p(\lambda) \frac{P_\lambda\, \rho\, P_\lambda}{\operatorname{Tr} \rho\, P_\lambda} = \sum_\lambda P_\lambda\, \rho\, P_\lambda. \tag{136}$$

Example: single-qubit system measured in Z basis. The Z observable

$$Z = |0\rangle\langle 0| - |1\rangle\langle 1| \tag{137}$$

has only two eigenvalues $+1$ and -1 (i.e. the possible measurement outcomes), corresponding to

$$\begin{aligned} P_{+1} &= |0\rangle\langle 0| \doteq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \\ P_{-1} &= |1\rangle\langle 1| \doteq \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned} \tag{138}$$

Given the prior density matrix

$$\rho \doteq \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix}, \tag{139}$$

measure the Z observable,

- with probability $p_{+1} = \operatorname{Tr} \rho\, P_{+1} = \rho_{00}$, the outcome $Z = +1$ will be observed, and the system collapses to

$$\rho' = \frac{P_{+1}\,\rho\, P_{+1}}{\operatorname{Tr} \rho\, P_{+1}} \doteq \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}; \tag{140}$$

- with probability $p_{-1} = \operatorname{Tr} \rho\, P_{-1} = \rho_{11}$, the outcome $Z = -1$ will be observed, and the system collapses to

$$\rho' = \frac{P_{-1}\,\rho\, P_{-1}}{\operatorname{Tr} \rho\, P_{-1}} \doteq \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}; \tag{141}$$

- without post selection, the post-measurement state will be

$$\rho' \doteq \rho_{00}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \rho_{11}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \rho_{00} & 0 \\ 0 & \rho_{11} \end{pmatrix}, \tag{142}$$

  where the off-diagonal matrix element disappeared in the density matrix ρ', a phenomenon known as **measurement-induced decoherence**.

HW 3

> A spin-1 particle has three S_z eigenstates, denoted as $|{+1}\rangle$, $|0\rangle$, and $|{-1}\rangle$. Suppose the particle was originally in the maximally mixed state, described by
> $\rho = \frac{1}{3}\,(|{+1}\rangle\langle{+1}| + |0\rangle\langle 0| + |{-1}\rangle\langle{-1}|)$.
> Consider measuring the S_x^2 observable,
> $S_x^2 = \frac{1}{2}\,(|{+1}\rangle + |{-1}\rangle)(\langle{+1}| + \langle{-1}|) + |0\rangle\langle 0|$.
> (i) Compute the eigenvalues of S_x^2, and construct the corresponding projection operators.
> (ii) What is the probability of each measurement outcome?
> (iii) Determine the post-measurement state ρ' fore each measurement outcome, and analyze whether the post-measurement state ρ' is pure or mixed in each cases.

▪ Indirect Measurement

Unitary evolution and **measurement** (state collapse) seem to be two irreconcilably different quantum dynamics. Are they related in any way?

Every measurement involves a *system* and an *apparatus*.

$$\begin{array}{cc|ccc} \multicolumn{2}{c|}{\otimes} & \multicolumn{3}{c}{\text{apparatus}} \\ & & |\varnothing\rangle & |+\rangle & |-\rangle \\ \hline \text{system} & |0\rangle & |0\,\varnothing\rangle & |0\,+\rangle & |0\,-\rangle \\ & |1\rangle & |1\,\varnothing\rangle & |1\,+\rangle & |1\,-\rangle \end{array} \tag{143}$$

- Consider the system is a qubit to be measured in Z basis (two basis states: $|0\rangle$ and $|1\rangle$).
- The apparatus has three states: $|\varnothing\rangle$ null (before measurement), $|+\rangle$ positive (indicating $Z = +1$ outcome), $|-\rangle$ negative (indicating $Z = -1$ outcome).
- The system and apparatus together is a *composite* quantum system of 6-dimensional Hilbert space, with the 6 basis states in Eq. (143) arranged in the following order

$$\{|0\,\emptyset\rangle, |0+\rangle, |0-\rangle, |1\,\emptyset\rangle, |1+\rangle, |1-\rangle\}. \tag{144}$$

The apparatus was designed such that when the measurement button is pressed, it will interact with the system by the *joint* **unitary evolution**, described by

$$\begin{aligned} U = {} & |0+\rangle\langle 0\,\emptyset| - |0\,\emptyset\rangle\langle 0+| + |0-\rangle\langle 0-| \\ & + |1-\rangle\langle 1\,\emptyset| - |1\,\emptyset\rangle\langle 1-| + |1+\rangle\langle 1+|. \end{aligned} \tag{145}$$

Exc 9

> Show that U in Eq. (145) is indeed unitary (and therefore can be realized by a time-evolution in principle).

Most importantly, under the unitary evolution

$$\boxed{\begin{aligned} |0\,\emptyset\rangle &\to U\,|0\,\emptyset\rangle = |0+\rangle, \\ |1\,\emptyset\rangle &\to U\,|1\,\emptyset\rangle = |1-\rangle, \end{aligned}} \tag{147}$$

which

- leaves the qubit state *unchanged*,
- and flips the apparatus to $|+\rangle$ (or $|-\rangle$) state *controlled by* the state of the qubit being $|0\rangle$ (or $|1\rangle$),

thereby realizing the Z observable measurement.

More generally, assuming the initial state of the qubit is a superposition state

$$|\psi\rangle = \psi_0\,|0\rangle + \psi_1\,|1\rangle, \tag{148}$$

the composite system starts from a *product* state

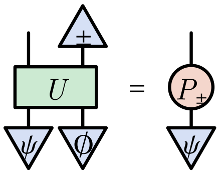
$$|\psi\rangle \otimes |\emptyset\rangle = \psi_0\,|0\,\emptyset\rangle + \psi_1\,|1\,\emptyset\rangle, \tag{149}$$

evolves into an *entangled* state

$$U(|\psi\rangle \otimes |\emptyset\rangle) = \psi_0\,|0+\rangle + \psi_1\,|1-\rangle. \tag{150}$$

Interpretations:

- If the apparatus reads + (or −), the qubit is in the $|0\rangle$ (or $|1\rangle$) state. Moreover, the probability for the measurement outcome $Z = +1$ to appear is $|\psi_0|^2$, exactly the same as the probability of the qubit being observed in the $|0\rangle$ state. — No explicit collapse is needed. The qubit and the apparatus evolves into an **entangled** reality under *unitary* time evolution.
  - When an external observer *reads* the apparatus, they will be involved into the entangled state. When the observer publishes their observation to the world, the entire world will be entangled … — This is the **multi-worlds interpretation** of quantum mechanics.
- Observing the apparatus still **collapses** the joint quantum state: the measurement of the *qubit* can be equivalently implemented by measuring the *apparatus* that has *interacted* with it, i.e. an **indirect measurement**. Collapsing the apparatus (by projecting the apparatus to $|\pm\rangle$ states) *induces* a corresponding **projection operator** $P_\pm$ acting on the qubit:



$$\boxed{P_{\pm} = \langle \pm|\, U\, |\emptyset\rangle \Rightarrow \begin{cases} P_{+} = |0\rangle\langle 0| \\ P_{-} = |1\rangle\langle 1| \end{cases}.} \tag{151}$$

Exc 10 Given U in Eq. (145), calculate $P_{\pm}$ in Eq. (151).

which are exactly the projection operators describing the **projective measurement** of the qubit in the Z basis. — State collapse can back propagate from apparatus to system.

- The collapse of the system is *caused* by the collapse of the apparatus, that the **causality** appears to flow *backward* in time in quantum measurement.
- If the apparatus is subsequently observed by an external observer, the underlying cause of the collapse is further traced back to the observer. This implies a **recursive observer hierarchy**, such that the origin of collapse can be recursively traced back forever.

Does the last entity to look at the entire system **collapse** the quantum state, or does it just get **entangled**? Or is there even a last observer? — These are still open (philosophical) questions about quantum measurements.

Generalized Measurement

- **Projective measurement** is disruptive: providing maximal information about the measurement outcome at the cost of fully collapse the state.
- **Generalized measurement** is a less disruptive form of quantum measurement that only *partially* disturb the state, but providing only *limited* information about the observable.

Consider *indirect measurement*, suppose the system-apparatus interaction is described by the following Hamiltonian

$$H = i\ g\ (|0+\rangle\langle 0\,\emptyset| - |0\,\emptyset\rangle\langle 0+| + |1-\rangle\langle 1\,\emptyset| - |1\,\emptyset\rangle\langle 1-|), \tag{152}$$

which is such designed that it will generate the desired unitary operation U in Eq. (145), via the following time evolution

$$U = e^{-i\,H} \tag{153}$$

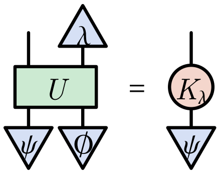
with the coupling $g = \pi/2$.

Exc 11 Show that H in Eq. (152) generates U in Eq. (145) with $g = \pi/2$.

- With generic coupling strength g, the initial states will be transformed by U as

$$\begin{aligned}|0\,\emptyset\rangle &\to U\,|0\,\emptyset\rangle = \cos\,g\,\,|0\,\emptyset\rangle + \sin\,g\,\,|0+\rangle,\\ |1\,\emptyset\rangle &\to U\,|1\,\emptyset\rangle = \cos\,g\,\,|1\,\emptyset\rangle + \sin g\,|1-\rangle.\end{aligned} \tag{154}$$

- Now if we observe the apparatus, there are three possible outcomes: $|\emptyset\rangle$ null, $|+\rangle$ positive, $|-\rangle$ negative.
  - Each outcome on the *apparatus* is associated with an effective **Kraus operator** acting on the *system*:



$$\begin{aligned}K_\emptyset &= \langle\emptyset|\;U\;|\emptyset\rangle = \cos g\begin{pmatrix}1&0\\0&1\end{pmatrix},\\ K_+ &= \langle +|\;U\;|\emptyset\rangle = \sin g\begin{pmatrix}1&0\\0&0\end{pmatrix},\\ K_- &= \langle -|\;U\;|\emptyset\rangle = \sin g\begin{pmatrix}0&0\\0&1\end{pmatrix}.\end{aligned} \tag{155}$$

Exc 12

> Calculate the operator K_λ.

  - The Kraus operator specifies how the system should response to the observation outcome of the apparatus.

$$\boxed{|\psi\rangle \xrightarrow[\text{in state }|\lambda\rangle]{\text{observe apparatus}} \frac{K_\lambda\,|\psi\rangle}{\sqrt{\langle\psi|\,K_\lambda^\dagger\,K_\lambda\,|\psi\rangle}}.} \tag{156}$$

  - The probability to observe apparatus in the state $|\mu\rangle$ should be given by

$$\boxed{p(\mu) = \langle\psi|\,K_\mu^\dagger\,K_\mu\,|\psi\rangle.} \tag{157}$$

    The normalization condition $\sum_\lambda p(\lambda) = 1$ requires that

$$\sum_\lambda K_\lambda^\dagger\,K_\lambda = \mathbb{1}, \tag{158}$$

    which is indeed satisfied by the Kraus operators K_λ in Eq. (155).

- **Weak measurement** corresponds to the limit when the coupling strength $g \to 0$. In this limit, the probability to read out ± from the apparatus is *small*:

$$p(\pm) \sim g^2, \tag{159}$$

meaning that

  - the measurement provides little information about the system,

- but the system is also rarely collapsed by measurement.

Quantum Channels

Kraus Theorem

What is the most general way a quantum state can evolve?

Quantum channel describes the most general way a quantum state can evolve, incorporating both unitary dynamics and measurement (and more).

Mathematically, a quantum channel $\mathcal{E}$ is a **completely positive trace-preserving (CPTP) map** that acts on a density matrix ρ:

$$\rho \to \rho' = \mathcal{E}(\rho). \tag{160}$$

- **Completely Positive**: the map should preserve the positivity of ρ, even when extended to a larger system.

$$(\mathcal{E} \otimes \mathrm{id})\,(\rho) \geq 0. \tag{161}$$

- **Trace-Preserving**: the total probability should remains 1.

$$\mathrm{Tr}\,\mathcal{E}(\rho) = 1. \tag{162}$$

Kraus' theorem: every CPTP map can be written as

$$\boxed{\mathcal{E}(\rho) = \sum_\lambda K_\lambda\,\rho\,K_\lambda^\dagger,} \tag{163}$$

with a set of **Kraus operators** K_λ satisfying

$$\sum_\lambda K_\lambda^\dagger\,K_\lambda = \mathbb{1}. \tag{164}$$

Examples:

- **Unitary evolution**: only one Kraus operator, and it is unitary $K = U$,

$$\mathcal{E}(\rho) = U\,\rho\,U^\dagger. \tag{165}$$

- **Projective measurement** (without post-selection): Kraus operators are projection operators $K_\lambda = P_\lambda$, see Eq. (136), as

$$\mathcal{E}(\rho) = \sum_\lambda P_\lambda\,\rho\,P_\lambda. \tag{166}$$

- **Generalized measurement** (without post-selection): generic Kraus operator, falls back to Eq. (163).

Quantum Decoherence

Quantum coherence is a property that allows *superposition* of states, enabling *interference* and *entanglement*. However, when a quantum system interact with environment, coherence can degrade — a process called **decoherence**.

- Decoherence explains why quantum system often *appear classical.*
- Decoherence models real-world *noise* in quantum computation and communication.

Consider a single qubit in a coherent superposition state

$$\begin{aligned} &|\psi\rangle = \psi_0\,|0\rangle + \psi_1\,|1\rangle, \\ &\rho = |\psi\rangle\,\langle\psi| \doteq \begin{pmatrix} \psi_0^*\,\psi_0 & \psi_1^*\,\psi_0 \\ \psi_0^*\,\psi_1 & \psi_1^*\,\psi_1 \end{pmatrix} = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix}. \end{aligned} \tag{167}$$

The quantum coherence refers to the non-vanishing *off-diagonal* matrix element $\rho_{01} = \psi_1^*\,\psi_0$ and $\rho_{10} = \psi_0^*\,\psi_1$ in ρ.

Mechanisms of Decoherence:

- **Time-Averaging**: Under unitary time evolution driven by the Hamiltonian

$$H = Z \doteq \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{168}$$

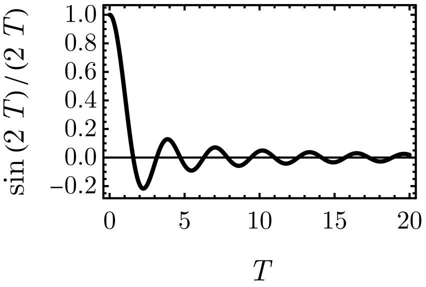
the density matrix evolves as

$$\rho(t) = e^{-i\,H\,t}\,\rho\,e^{i\,H\,t} \doteq \begin{pmatrix} \rho_{00} & \rho_{01}\,e^{-2\,i\,t} \\ \rho_{10}\,e^{2\,i\,t} & \rho_{11} \end{pmatrix}. \tag{169}$$

Suppose we do not have a good resolution of time, averaging the density matrix over a long enough time window, the effective state is

$$\overline{\rho} = \frac{1}{2\,T}\int_{-T}^{T}\rho(t) = \begin{pmatrix} \rho_{00} & \rho_{01}\,\frac{\sin(2\,T)}{2\,T} \\ \rho_{10}\,\frac{\sin(2\,T)}{2\,T} & \rho_{11} \end{pmatrix}. \tag{170}$$

The off-diagonal matrix element will decay toward 0 as $T \to \infty$:



Key points:

- Loss of coherence due to the inability to resolve fast oscillations over time.
- A classical averaging effect and does not involve an external interaction with the system.

- **Measurement-Induced**: Measure the qubit in Z basis and forget about the measurement outcome (measurement without post-selection). The state will undergo the following channel

$$\rho \to \mathcal{E}(\rho) = \sum_{\lambda=\pm} P_\lambda\,\rho\,P_\lambda, \tag{171}$$

where $P_\pm = (\mathbb{1} \pm Z)/2$. As a result,

$$\mathcal{E}(\rho) \doteq \begin{pmatrix} \rho_{00} & 0 \\ 0 & \rho_{11} \end{pmatrix}, \tag{172}$$

the off-diagonal matrix elements are removed.

Key points:

- Loss of coherence due to interaction with environment, effectively "measuring" the system and collapsing it into the Z-basis.
- Explicitly involves interaction with an external system.

In general, regardless of mechanism, **quantum decoherence** can be effectively described using **quantum channels**.

- **Depolarizing Channel**

For a single qubit:

$$\boxed{\mathcal{E}(\rho) = (1-p)\,\rho + \frac{p}{2}\,\mathbb{1}.} \tag{173}$$

Krause operators:

$$K_0 = \sqrt{1-p}\; I,\; K_1 = \sqrt{\frac{p}{2}}\; X,\; K_2 = \sqrt{\frac{p}{2}}\; Y,\; K_3 = \sqrt{\frac{p}{2}}\; Z. \tag{174}$$

- **Phase Damping (Dephasing) Channel**

For a single qubit:

$$\boxed{\mathcal{E}(\rho) = (1-p)\,\rho + p\,Z\,\rho\,Z.} \tag{175}$$

Krause operators:

$$K_\pm = \sqrt{\frac{1-p}{2}}\; I \pm \sqrt{\frac{p}{2}}\; Z. \tag{176}$$

Lindbladian Dynamics

While *quantum channels* provide a most general *discrete* description of **quantum dynamics** in *open* quantum systems, we seek a *continuous-time* formulation of how the density matrix ρ evolves.

Consider a small time step $d t$, under which

$$\rho(t + d t) = \mathcal{E}(\rho(t)) = \sum_{\lambda=0}^{n} K_\lambda\,\rho(t)\,K_\lambda^\dagger, \tag{177}$$

where $\mathcal{E}$ is a close-to-identity channel described by the following Kraus operators

$$\begin{aligned} K_0 &= \mathbb{1} - \mathbb{i}\, H\, \mathbb{d}t - \frac{1}{2} \sum_{i=1}^{n} L_i^\dagger\, L_i\, \mathbb{d}t, \\ K_i &= \sqrt{\mathbb{d}t}\; L_i \;\; (\text{for } i = 1, \ldots, n), \end{aligned} \tag{178}$$

where

- H is the **Hamiltonian**, driving the unitary time evolution;
- L_i are **Lindblad (jump) operators**, describing dissipation effects in open quantum systems.

Exc 13

> Show that the set of Kraus operators in Eq. (178) satisfies the normalization condition $\sum_{\lambda=0}^{n} K_\lambda^\dagger\, K_\lambda = \mathbb{1}$ to the leading order in $\mathbb{d}t$.

Substitute Eq. (178) into Eq. (177) and expand to the leading order of $\mathbb{d}t$, we arrive at the **Lindblad master equation**:

$$\boxed{\partial_t \rho = -\mathbb{i}\,[H, \rho] + \sum_{i=1}^{n} \left(L_i\, \rho\, L_i^\dagger - \frac{1}{2} \left\{ L_i^\dagger\, L_i, \rho \right\} \right).} \tag{179}$$

Exc 14

> Derive Eq. (179).

- The first term $-\mathbb{i}\,[H, \rho]$ is inherited from the **von Neumann equation** of mixed state unitary dynamics.
- The second term describes the **dissipation effect**: $L_i\, \rho\, L_i^\dagger$ captures the change of density matrix under dissipation and $-\frac{1}{2}\left\{L_i^\dagger\, L_i, \rho\right\}$ takes care of the trace preservation.

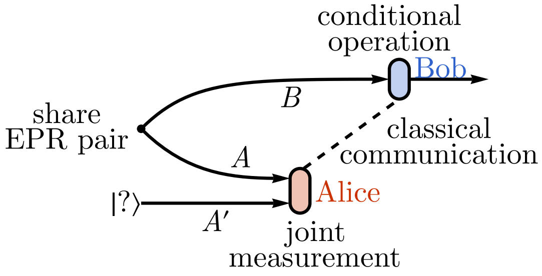
HW 4

> The density matrix of a single qubit takes the following form
> $\rho \doteq \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix}$.
> Consider the qubit evolves under Lindblad dynamics with $H = 0$, and a single Lindblad operator
> $L = \sqrt{\lambda} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$,
> where $\lambda > 0$ is the dissipation strength.
> (i) Derive the explicit form of the differential equations for the density matrix elements ρ_{00}, ρ_{11}, ρ_{01}, ρ_{10}.
> (ii) Show that the diagonal matrix elements ρ_{00}, ρ_{11} do not evolve in time, while the off-diagonal matrix elements ρ_{01}, ρ_{10} decays exponentially in time, realizing quantum decoherence.
> (iii) Analyze how the decoherence time is affected by the dissipation strength.

Quantum Protocols

Quantum Teleportation

Quantum teleportation transfers (unknown) quantum states from one location to another using *quantum entanglement* resource and *classical communication.*



- The model involves three qubits: A', A and B.
  - Establish **quantum entanglement**: A and B qubits are prepared in an EPR state, and shared between Alice and Bob respectively

$$|\text{EPR}\rangle_{AB} = \frac{1}{\sqrt{2}}(|\uparrow_A \uparrow_B\rangle + |\downarrow_A \downarrow_B\rangle) \doteq \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\0\\1\end{pmatrix}. \tag{183}$$

  - A' qubit is in an *unknown* quantum state in Alice's possession

$$|\psi\rangle_{A'} = \alpha\,|\uparrow_{A'}\rangle + \beta\,|\downarrow_{A'}\rangle \doteq \begin{pmatrix}\alpha\\\beta\end{pmatrix}, \tag{184}$$

  (Alice knowns that the single-qubit pure state must take this form, but does not need to know what α, β are.)

  - The three-qubit system is in a joint state

$$|\Psi\rangle = |\psi\rangle_{A'} \otimes |\text{EPR}\rangle_{AB} \doteq \begin{pmatrix}\alpha\\\beta\end{pmatrix} \otimes \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\0\\1\end{pmatrix} \tag{185}$$

- Goal: to teleport the quantum state $|\psi\rangle$ from A' to B without handing the qubit A' to Bob.

Protocol:

- Alice makes a **joint measurement** of A' and A, by observing $\sigma^x_{A'}\,\sigma^x_A$ and $\sigma^z_{A'}\,\sigma^z_A$ (note that they are commuting observables that can be measured simultaneously). There are four possible outcomes

$$\begin{array}{r|llll} \sigma^x_{A'}\,\sigma^x_A & +1 & +1 & -1 & -1 \\ \sigma^z_{A'}\,\sigma^z_A & +1 & -1 & +1 & -1 \\ P_{ab} & P_{++} & P_{+-} & P_{-+} & P_{--} \end{array}, \tag{186}$$

each corresponds to a projection operator (labeled by the measurement outcomes $a := \sigma^x_{A'}\,\sigma^x_A = \pm 1,\; b := \sigma^z_{A'}\,\sigma^z_A = \pm 1$)

$$\begin{aligned} P_{ab} &= \frac{\mathbb{1} + a\,\sigma^x_{A'}\,\sigma^x_A}{2}\;\frac{\mathbb{1} + b\,\sigma^z_{A'}\,\sigma^z_A}{2} \\ &\doteq \frac{1}{4}\begin{pmatrix} 1+b & 0 & 0 & a\,(1+b) \\ 0 & 1-b & a(1-b) & 0 \\ 0 & a(1-b) & 1-b & 0 \\ a\,(1+b) & 0 & 0 & 1+b \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned} \tag{187}$$

- After the measurement, the three-qubit state will collapse to

$$|\Psi\rangle \xrightarrow{\sigma^x_{A'}\,\sigma^x_A = a,\, \sigma^z_{A'}\,\sigma^z_A = b} \frac{P_{ab}\,|\Psi\rangle}{\text{normalization}\ldots}, \tag{188}$$

more explicitly as

$$P_{ab}\,|\Psi\rangle \doteq \frac{1}{4\sqrt{2}}\begin{pmatrix} \alpha\,(1+b) \\ \beta\,a\,(1+b) \\ \beta\,a\,(1-b) \\ \alpha\,(1-b) \\ \beta\,(1-b) \\ \alpha\,a\,(1-b) \\ \alpha\,a\,(1+b) \\ \beta\,(1+b) \end{pmatrix}. \tag{189}$$

Let us enumerate all four cases

$$\begin{array}{c|cccc} \sigma^x_{A'}\,\sigma^x_A & +1 & +1 & -1 & -1 \\ \sigma^z_{A'}\,\sigma^z_A & +1 & -1 & +1 & -1 \\ P_{ab}\,|\Psi\rangle & \begin{pmatrix} \alpha \\ \beta \\ 0 \\ 0 \\ 0 \\ 0 \\ \alpha \\ \beta \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ \beta \\ \alpha \\ \beta \\ \alpha \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} \alpha \\ -\beta \\ 0 \\ 0 \\ 0 \\ 0 \\ -\alpha \\ \beta \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ -\beta \\ \alpha \\ \beta \\ -\alpha \\ 0 \\ 0 \end{pmatrix} \end{array}, \tag{190}$$

It turns out that they can all be written as the tensor product state between $A'\,A$ and B as

$\sigma^x_{A'}\,\sigma^x_A$	$\sigma^z_{A'}\,\sigma^z_A$	$P_{ab}\,\lvert\Psi\rangle$
$+1$	$+1$	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$

$+1$	-1	$\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} \beta \\ \alpha \end{pmatrix}$
-1	$+1$	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix} \otimes \begin{pmatrix} \alpha \\ -\beta \end{pmatrix}$
-1	-1	$\begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} -\beta \\ \alpha \end{pmatrix}$

or in terms of ket state notation as

$$\begin{aligned} P_{++} \left|\Psi\right\rangle &= \frac{1}{\sqrt{2}} \left(\left|\uparrow_{A'} \uparrow_A\right\rangle + \left|\downarrow_{A'} \downarrow_A\right\rangle\right) \otimes \left(\alpha \left|\uparrow_B\right\rangle + \beta \left|\downarrow_B\right\rangle\right), \\ P_{+-} \left|\Psi\right\rangle &= \frac{1}{\sqrt{2}} \left(\left|\uparrow_{A'} \downarrow_A\right\rangle + \left|\downarrow_{A'} \uparrow_A\right\rangle\right) \otimes \left(\beta \left|\uparrow_B\right\rangle + \alpha \left|\downarrow_B\right\rangle\right), \\ P_{-+} \left|\Psi\right\rangle &= \frac{1}{\sqrt{2}} \left(\left|\uparrow_{A'} \uparrow_A\right\rangle - \left|\downarrow_{A'} \downarrow_A\right\rangle\right) \otimes \left(\alpha \left|\uparrow_B\right\rangle - \beta \left|\downarrow_B\right\rangle\right), \\ P_{--} \left|\Psi\right\rangle &= \frac{1}{\sqrt{2}} \left(\left|\uparrow_{A'} \downarrow_A\right\rangle - \left|\downarrow_{A'} \uparrow_A\right\rangle\right) \otimes \left(-\beta \left|\uparrow_B\right\rangle + \alpha \left|\downarrow_B\right\rangle\right). \end{aligned} \tag{192}$$

- Alice will tell Bob her measurement outcome

  $$(\sigma_{A'}^x \sigma_A^x = a, \sigma_{A'}^z \sigma_A^z = b), \tag{193}$$

  via a **classical communication channel** (e.g. by making a phone call).

  - If $(a, b) = ++$, the qubit B is in the state

    $$\alpha \left|\uparrow_B\right\rangle + \beta \left|\downarrow_B\right\rangle = \left|\psi\right\rangle_B. \tag{194}$$

    There is nothing more Bob needs to do. The state $\left|\psi\right\rangle$ has been teleported to B successfully.

  - If $(a, b) = +-$, the qubit B is in the state

    $$\beta \left|\uparrow_B\right\rangle + \alpha \left|\downarrow_B\right\rangle, \tag{195}$$

    Bob will apply a σ_B^x operator to the qubit B

    $$\sigma_B^x(\beta \left|\uparrow_B\right\rangle + \alpha \left|\downarrow_B\right\rangle) = \alpha \left|\uparrow_B\right\rangle + \beta \left|\downarrow_B\right\rangle = \left|\psi\right\rangle_B, \tag{196}$$

    then the qubit B is converted to the state $\left|\psi\right\rangle$.

  - If $(a, b) = -+$, the qubit B is in the state

    $$\alpha \left|\uparrow_B\right\rangle - \beta \left|\downarrow_B\right\rangle, \tag{197}$$

    Bob will apply a σ_B^z operator to the qubit B

$$\sigma_B^z(\alpha\,|\uparrow_B\rangle - \beta\,|\downarrow_B\rangle) = \alpha\,|\uparrow_B\rangle + \beta\,|\downarrow_B\rangle = |\psi\rangle_B, \tag{198}$$

then the qubit B is converted to the state $|\psi\rangle$.

- If $(a, b) = +-$, the qubit B is in the state

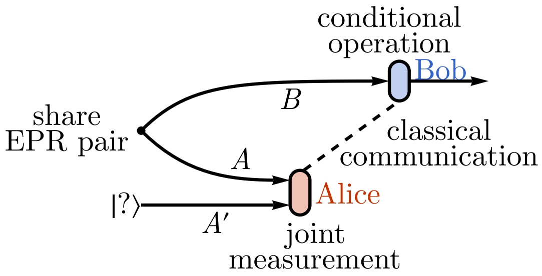
  $$-\beta\,|\uparrow_B\rangle + \alpha\,|\downarrow_B\rangle, \tag{199}$$

  Bob will apply the composite operator $\sigma_B^z\,\sigma_B^x$ to the qubit B

  $$\sigma_B^z\,\sigma_B^x(-\beta\,|\uparrow_B\rangle + \alpha\,|\downarrow_B\rangle) = \alpha\,|\uparrow_B\rangle + \beta\,|\downarrow_B\rangle = |\psi\rangle_B, \tag{200}$$

  then the qubit B is converted to the state $|\psi\rangle$.

Summary:



- Alice and Bob establish shared a *entanglement resource* (e.g. a EPR pair).
- For any unknown state $|\psi\rangle_{A'}$ handed to Alice, she measures $\sigma_{A'}^x\,\sigma_A^x$ and $\sigma_{A'}^z\,\sigma_A^z$.
- Alice tells Bob her the measurement outcomes by *classical communication*.
- Depending on the information, Bob performs conditional operation on his qubit

$\sigma_{A'}^x\,\sigma_A^x$	$\sigma_{A'}^z\,\sigma_A^z$	Bob’ s operation
$+1$	$+1$	$\mathbb{1}_B$
$+1$	-1	σ_B^x
-1	$+1$	σ_B^z
-1	-1	$\sigma_B^z\,\sigma_B^x$

. (201)

- After the operation, the state $|\psi\rangle$ will appear on Bob’s qubit.

Comments:

- The original state $|\psi\rangle$ on qubit A' is destroyed in the process → **No-Cloning Theorem**: it is impossible to create an identical copy of an arbitrary unknown state. You can only teleport an unknown state but not duplicate it.
- **Traversable wormhole** ⇔ interstellar **quantum teleportation**.
  - Entanglement resource: wormhole ⇔ entangled pairs of black holes
  - Classical communication: classical interaction between two black holes

Quantum Search

What is a search problem?

- Given a **bit string** x and a **query function** $Q(x)$ that tells if x is the target.

$$\begin{array}{ccc} x & Q(x) & \\ 0000 \to & 0 & \text{no} \\ 0001 \to & 0 & \text{no} \\ \vdots & \vdots & \\ 0110 \to & 1 & \text{yes!} \\ \vdots & \vdots & \end{array} \tag{202}$$

- Assuming there is only one target bit string to look for, the goal is to find it!

Complexity of unstructured search (over N entities)

- Classical computer ~ $O(N)$,
- Quantum computer (Grover's algorithm) ~ $O(N^{1/2})$. The quantum speedup can be significant when N is large.

Key idea: *quantum superposition* allows search to be carried out in *parallel.*

$$\begin{aligned} |s\rangle &= \frac{1}{\sqrt{N}} (|0000\rangle + |0001\rangle + \ldots + |0110\rangle + \ldots) \\ &= \frac{1}{\sqrt{N}} \sum_x |x\rangle. \end{aligned} \tag{203}$$

- Initially, all bit string states $|x\rangle$ have *equal* probability $1/N$ in a uniform superposition state $|s\rangle$.
- Grover's algorithm iteratively *enhance* the probability of the *target* state by *quantum interference.*
- After about $N^{1/2}$ steps, the target state emerges as the probability ~1 state.

The **Grover's algorithm** involves two steps in each iteration. Both steps are implemented as *unitary* operations.

- **Quantum query** (in parallel)

  $$U_Q |x\rangle = (-1)^{Q(x)} |x\rangle. \tag{204}$$

  It simply marks the target state with a minus sign. Let $|t\rangle$ be the *target* state, Eq. (204) can also be written as

  $$U_Q |x\rangle = |x\rangle - 2 |t\rangle \langle t|x\rangle. \tag{205}$$

- **Grover diffusion**

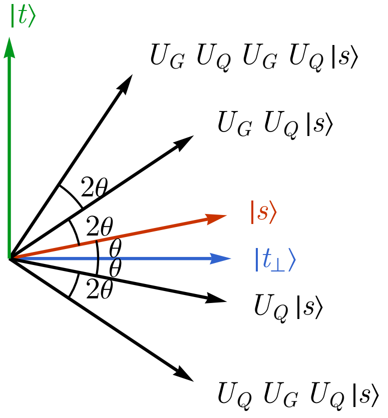
  $$U_G |x\rangle = 2 |s\rangle \langle s|x\rangle - |x\rangle, \tag{206}$$

  which reflect any state about the *source* state $|s\rangle$.

In the two-dimensional Hilbert space spanned by $|t\rangle$, $|s\rangle$, we can define the state $|t_\perp\rangle$ orthogonal to $|t\rangle$

$$|t_\perp\rangle = \frac{1}{\sqrt{N-1}} \sum_{x \neq t} |x\rangle. \tag{207}$$

- In each iteration, the quantum query $U_Q = \mathbb{1} - 2\,|t\rangle\,\langle t|$ reflects any state vector about $|t_\perp\rangle$, and the Grover diffusion $U_G = 2\,|s\rangle\,\langle s| - \mathbb{1}$ reflects any state vector about $|s\rangle$.



- Two reflections make a rotation of $2\,\theta$ angle, where the θ denotes the angle between the two reflection axes (i.e. the angle between $|s\rangle$ and $|t_\perp\rangle$) and is given by

$$\sin\theta = \langle s|t\rangle = \frac{1}{\sqrt{N}}, \tag{208}$$

- After k steps of Grover iterations, the source state $|s\rangle$ is rotated away from $|t_\perp\rangle$ by $(2\,k+1)\,\theta$.

- To rotate the state from $|s\rangle$ to $|t\rangle$ (which is $\pi/2$ from $|t_\perp\rangle$), the total rotation angle must accumulates to $\pi/2$

$$\frac{\pi}{2} = (2\,k+1)\,\theta, \tag{209}$$

such that the solution for k is given by

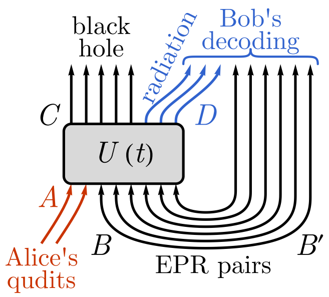
$$k = \frac{1}{2}\left(\frac{\pi}{2\,\theta} - 1\right) = \frac{1}{2}\left(\frac{\pi}{2\,\arcsin N^{-1/2}} - 1\right). \tag{210}$$

- In the large N limit (when there are many bit strings to search),

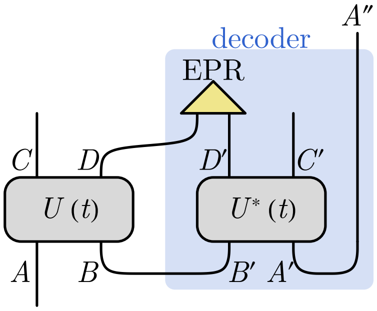
$$k \approx \frac{\pi}{4}\sqrt{N}, \tag{211}$$

which is indeed of order $N^{1/2}$ as mentioned.

Hayden-Preskill problem: can we recover the object that has fallen into a black hole? Yes. Combining quantum search and quantum teleportation (Yoshida, Kitaev, arXiv:1710.03363).



- Resources needed:
  - **Entanglement** resource: the black hole (on Alice's side) must be *entangled* with another black hole (on Bob's side) in the lab (effectively forming a wormhole).
  - **Quantum computation** resource: a strong enough *quantum computer* to simulate the quantum dynamics of black hole and to perform quantum search.
- Basic idea: *Hawking radiation* is an efficient *encoding* of the object that falls into the black hole (black hole is very much like a **quantum hash function**).
  - Collect the Hawking radiation a few moments after the object has fallen into the black hole on Alice's side.
  - Try to decode the Hawking radiation to recover the quantum information of the object ⇔ quantum teleportation of the object through the wormhole.



  - The key idea to search for a *collision* of Hawking radiations (hash keys) between the two entangled black holes. This relies on the quantum search algorithm.
  - Once the Hawking radiations from both black holes are made the same (same = perfectly entangled), the object will reemerge from Bob's black hole, as a result of quantum teleportation.

■ Quantum Error Correction

■ Five-Qubit Code

Quantum decoherence posts a serious threat to *quantum information processing*.

- Qubits *couple* to the environment and *decohere* inevitably.

- In the extreme case, a qubit can become **maximally mixed** ⇒ An **erasure error**: the quantum information of the qubit is fully *scrambled* with the *environment*, *as if* the information has been *erased*.

Quantum error correction: protecting the quantum information from errors by *spreading* the information into a *highly entangled* quantum many-body state (which we have access to).

$$\textit{one}\text{ logical qubit }\underset{\text{decoded to}}{\overset{\text{encoded in}}{\rightleftharpoons}}\ \textit{many}\text{ physical qubits.} \tag{212}$$

- **Logical qubit**: the information theoretic qubit (software level), whose basis states are denoted as $|\mathbf{0}\rangle$, $|\mathbf{1}\rangle$ (using *boldface*).
- **Physical qubit**: the actual qubit implemented on quantum devices (hardware level).

Even if some physical qubits are corrupted or erased, one can still retrieve the logical qubit from the rest of the physical qubits.

Five-qubit code: a quantum error correction code that encodes **one** *logical qubit* into **five** *physical qubits*, where the logical qubit is protected against the *erasure* of *any* **two** physical qubits.

- The *logical qubit* states $|\mathbf{0}\rangle$, $|\mathbf{1}\rangle$ span a 2-dimensional **code subspace** in the 2^5-dimensional physical qubit Hilbert space.
- The *code subspace* is specified by four **commuting Pauli operators** on the physical qubits:

$$\begin{aligned} M_1 &= X_1\, Z_2\, Z_3\, X_4,\\ M_2 &= X_2\, Z_3\, Z_4\, X_5,\\ M_3 &= X_3\, Z_4\, Z_5\, X_1,\\ M_4 &= X_4\, Z_5\, Z_1\, X_2. \end{aligned} \tag{213}$$

What about the last $X_5\, Z_1\, Z_2\, X_3$ in the cyclic arrangement? It turns out to be $\prod_{i=1}^4 M_i$ and is therefore not independent.

- These operators M_i are called **stabilizers**, as they stabilize the *logical qubit* as their *common eigenstates* of eigenvalue +1, i.e.

$$\forall\ i : M_i\, |\mathbf{0}\rangle = |\mathbf{0}\rangle,\ \ M_i\, |\mathbf{1}\rangle = |\mathbf{1}\rangle. \tag{214}$$

Recall that we can simultaneously diagonalize *commuting* operators by constructing a **many-body Hamiltonian**, e.g.

$$\begin{aligned} H &= -M_1 - M_2 - M_3 - M_4\\ &= -X_1\, Z_2\, Z_3\, X_4 - X_2\, Z_3\, Z_4\, X_5 - X_3\, Z_4\, Z_5\, X_1 - X_4\, Z_5\, Z_1\, X_2. \end{aligned} \tag{215}$$

  - The **code subspace** = the **common eigenspace** of *stabilizers* that $\forall\ i : M_i = +1$ = the **ground state subspace** of the *Hamiltonian H*.
  - The code subspace is *two-dimensional* ⇒ can encode a **logical qubit**. How do we know? 5 qubits, 4 stabilizers: each stabilizer **halves** the Hilbert space ⇒ the remaining space dimension:

$$\frac{2^5}{2^4} = 2. \tag{216}$$

- Within the code subspace, a choice of the basis is (can be obtained by diagonalize H)

$$\begin{aligned} |\mathbf{0}\rangle &= \frac{1}{4}\,(|00000\rangle - |00011\rangle + |00101\rangle - |00110\rangle + |01001\rangle + |01010\rangle - |01100\rangle - |01111\rangle - \\ &\quad |10001\rangle + |10010\rangle + |10100\rangle - |10111\rangle - |11000\rangle - |11011\rangle - |11101\rangle - |11110\rangle), \\ |\mathbf{1}\rangle &= \frac{1}{4}\,(-|00001\rangle - |00010\rangle - |00100\rangle - |00111\rangle - |01000\rangle + |01011\rangle + |01101\rangle - |01110\rangle - \\ &\quad |10000\rangle - |10011\rangle + |10101\rangle + |10110\rangle - |11001\rangle + |11010\rangle - |11100\rangle + |11111\rangle). \end{aligned} \tag{217}$$

- **Logical gates**: quantum gates that effectively operate on the logical qubit

$$\begin{aligned} &\boldsymbol{Z}\,|\mathbf{0}\rangle = |\mathbf{0}\rangle,\ \boldsymbol{Z}\,|\mathbf{1}\rangle = -|\mathbf{1}\rangle, \\ &\boldsymbol{X}\,|\mathbf{0}\rangle = |\mathbf{1}\rangle,\ \boldsymbol{X}\,|\mathbf{1}\rangle = |\mathbf{0}\rangle. \end{aligned} \tag{218}$$

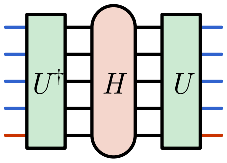
- $\boldsymbol{Z}$ and $\boldsymbol{X}$ must *commute* with all stabilizers (to remain in the code subspace), yet not any product of stabilizers (to be nontrivial). One canonical choice is

$$\begin{aligned} &\boldsymbol{Z} = Z_1\, Z_2\, Z_3\, Z_4\, Z_5, \\ &\boldsymbol{X} = X_1\, X_2\, X_3\, X_4\, X_5, \\ &\boldsymbol{Y} = i\,\boldsymbol{X}\,\boldsymbol{Z} = Y_1\, Y_2\, Y_3\, Y_4\, Y_5 \end{aligned} \tag{219}$$

- It is hard to decohere the logical qubit, because $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$ are non-local. ⇒ Their couplings to the environment are typically weak.

A **diagrammatic** understanding: the *unitary matrix* U that diagonalize the Hamiltonian H can be viewed as a *quantum circuit*,

$$U^\dagger\, H\, U = E. \tag{220}$$



The *quantum circuit* U should also simultaneously diagonalize all the stabilizers. With a proper basis choice, one can find U, such that

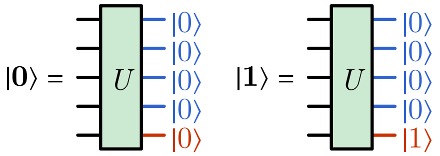
$$\begin{aligned} &U^\dagger\, M_1\, U = \tilde{Z}_1, \\ &U^\dagger\, M_2\, U = \tilde{Z}_2, \\ &U^\dagger\, M_3\, U = \tilde{Z}_3, \\ &U^\dagger\, M_4\, U = \tilde{Z}_4. \end{aligned} \tag{221}$$

As a result, the Hamiltonian transforms to

$$U^\dagger\, H\, U = -\tilde{Z}_1 - \tilde{Z}_2 - \tilde{Z}_3 - \tilde{Z}_4. \tag{222}$$

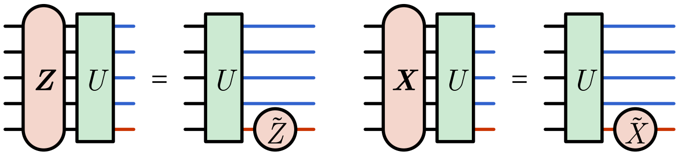
- The first four qubits are *pinned* by the Hamiltonian to $\left|\tilde{0}\,\tilde{0}\,\tilde{0}\,\tilde{0}\right\rangle$ to lower the energy ⇒ **syndrome qubits**.
- The last qubit is free ⇒ **logical qubit**.

The quantum circuit *encodes* the *logical* qubit into five *physical* qubits, given the *syndrome* qubits pinned to $\left|\tilde{0}\,\tilde{0}\,\tilde{0}\,\tilde{0}\right\rangle$. This is how Eq. (217) was obtained.



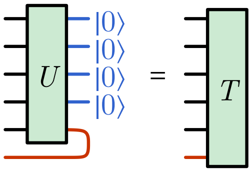
In addition, the *logical gates* do acts on the *logical qubit* as expected,

$$\begin{aligned} U^\dagger\, \boldsymbol{Z}\, U &= \tilde{Z}_5, \\ U^\dagger\, \boldsymbol{X}\, U &= \tilde{X}_5. \end{aligned} \tag{223}$$

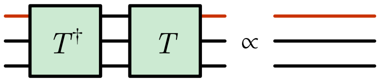


- Logical gates will not take the system out of the *code subspace*, as they will not touch the syndrome qubit.
- If any of the syndrome qubit is *flipped*. ⇒ The system is carried out of the *code subspace* (**excitation** created). ⇒ Signals an error. ⇒ Correct the error by applying appropriate unitary operations based on the syndrome.

How well is the logical qubit protected? Take the *unitary circuit*, pin the *syndrome qubits* and bend around the *logical qubit* → a six-leg tensor T describing how the *logical* and the *physical* qubits are related



It is a **perfect tensor**, because of an amazing property: T is proportional to a **unitary** matrix from *any half* of legs to the rest *half* of legs.



Treat T as a many-body state (after normalization) ⇒ it describes a *pure state* of *six* qubits, where *any* set of three qubits is *maximally entangled* with the complementary set of three qubits. Such states have been called **absolutely maximally entangled** states.

The the nth Rényi entanglement entropy of any m qubits in the six-qubit state $|T\rangle$ is

$$S^{(n)}(m) = \min(m, 6 - m) \log 2. \tag{224}$$

Exc 15 Use the perfect tensor property to show Eq. (224).

The **mutual information** between the *logical qubit* and any m *physical qubits* is given by

$$I^{(n)}(1 : m) = \begin{cases} 0 & m \leq 2, \\ 2 \log 2 & m \geq 3. \end{cases} \tag{225}$$

Exc 16 Verify Eq. (225) given Eq. (224).

The *five-qubit code* has the property that

- any *two* (or less) qubits have *no* information about the logical qubit.
- any *three* (or more) qubits have *complete* information about the logical qubit.

Therefore the logical qubit is protected against *erasure error* up to *two* qubits.