Quantum Mechanics

Part II. Quantum Entanglement

Information Theory

■ Classical Information

■ Probability Theory

A **deterministic variable** x must represent the some *definite* value (even if the value is not necessarily known). Example: if $2\,x + 1 = 5$, then $x = 2$.

A **random variable** X defines a *support space* $\mathcal{X}$ in which each *possible values* $x \in \mathcal{X}$ is assigned a **probability** $p(x) \equiv \Pr(X = x)$.

- The **expectation value** (mean, average) of a random variable X is the *average* of all possible values *weighted* by their probabilities

$$\boxed{\langle x \rangle = \mathbb{E}[X] = \sum_{x \in \mathcal{X}} x\, p(x).} \tag{1}$$

- The **variance** of a random variable X is the expectation value of the *squared deviation* of X from its mean $\mathbb{E}[X]$.

$$\boxed{\mathrm{var}(X) = \mathbb{E}\Big[(X - \mathbb{E}[X])^2\Big] = \big\langle x^2 \big\rangle - \langle x \rangle^2.} \tag{2}$$

A *deterministic* variable x can be viewed as a special limit of a *random* variable X, whose probability is concentrated at one *fixed* value x_*.

$$p(x) = \Pr(X = x) = \begin{cases} 1 & x = x_* \\ 0 & x \neq x_* \end{cases}. \tag{3}$$

■ Observation Theory

A *Bayesian* (subjectivism) view of probability:

- Probability assignment is a means to describe our **state of knowledge** about the random variable X.
- The probability assignment can be *updated*, if our state of knowledge is *changed* by **observations** (given new evidence from observations).

A (full) **observation** of a random variable X *uniquely* determines a *outcome* value $x_* \in \mathcal{X}$, such outcome will be observed with the probability $p(x_*)$.

- Before observing $X = x_*$, the probability distribution $p(x)$ is called the **prior probability**.
- After observing $X = x_*$, the probability distribution will *collapse* to the **posterior probability** $p(x \mid X = x_*)$ (reads: the probability of x given the observation $X = x_*$)

$$\boxed{p(x \mid X = x_*) = \begin{cases} 1 & \text{if } x = x_*, \\ 0 & \text{else.} \end{cases}} \tag{4}$$

- If we continue to observe X again, we would obtain $X = x_*$ for sure, based on the updated probability $p(x \mid X = x_*)$.

Observation removes the *uncertainty* of a random variable, making it deterministic. *Repeated observations* will only *confirm* the previous observation outcome. The observation outcome is *verifiable* ⇒ it qualifies as a piece of *knowledge*.

Example:

Rolling a dice ($\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$) and observe $X = 2$.

- Prior probability:

$$\begin{array}{ccccccc} x & 1 & 2 & 3 & 4 & 5 & 6 \\ p(x) & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{array} \tag{5}$$

- Posterior probability:

$$\begin{array}{ccccccc} x & 1 & 2 & 3 & 4 & 5 & 6 \\ p(x \mid X = 2) & 0 & 1 & 0 & 0 & 0 & 0 \end{array} \tag{6}$$

Of course, the observation of $X = 2$ is only one possible outcome. If other possible outcome is observed instead, the posterior probability will be different.

▪ Information Theory

The amount of **information** that we gain from a specific observation (observing X and obtaining $X = x_*$) is the negative log-probability associated with the observation event

$$\boxed{I(X = x_*) = -\log \Pr(X = x_*) = -\log p(x_*).} \tag{7}$$

Intuitions:

- If $p(x_*) = 1$, we already know for certain that $X = x_*$, further observing $X = x_*$ will not bring us new knowledge, thus we are not gaining more information from the observation, and $I(X = x_*) = 0$ in this case.

- If $p(x_*) \to 0$, observing $X = x_*$ is a rare event. But if the rare event happens, it would bring us a big surprise (a lot of knowledge), because base on the observation, a large amount of other possibilities (of $X \neq x_*$) can be ruled out, which is very informative. So $I(X = x_*) \to +\infty$ in this case.

Examples:

- Tossing a coin ($\mathcal{X} = \{\text{head}, \text{tail}\}$) and observe $X = \text{head}$.

$$\begin{array}{ccc} x & \text{head} & \text{tail} \\ p(x) & \frac{1}{2} & \frac{1}{2} \end{array} \tag{8}$$

Information gained from observation

$$I(X = \text{head}) = -\log(1/2) = \log 2 = 1\,\text{bit}. \tag{9}$$

  - $\log 2$ information is also called one **bit** of information ($1\,\text{bit} = \log 2$ is widely used as the information unit in information science), it is the amount of information obtained from knowing the answer of a yes-or-no question.

- Observing an unknown bit string of length $n = 3$ and obtain (say) $X = 010$

$$\begin{array}{ccccccccc} x & 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ p(x) & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{array} \tag{10}$$

Information gained from observation

$$I(X = 010) = -\log(1/8) = 3\log 2 = 3\,\text{bit}. \tag{11}$$

  - The observation specifies one outcome from eight equally likely outcomes. To determine the bit string 010, equivalently, we need to ask *three* yes-or-no questions (is the 1st/2nd/3rd bit 0?). Answering each of the question will give us *one bit* of information about the unknown string. So the observation gives us totally *three bits* of information all at once.
  - In general, observing an unknown bit string of length n will provide n bits of information about the string, as it determines one outcome from 2^n equally likely outcomes. The probability associated with the particular observation is

$$p = \left(\frac{1}{2}\right)^n \Rightarrow -\log p = n\log 2 = n\,\text{bit} = I. \tag{12}$$

- So the information $I(X = x_*)$ that we gain from observing $X = x_*$ should be defined to be the negative logarithm of the probability $p(x_*)$ for such observation event to happen.

$$I(X = x_*) = -\log p(x_*) \tag{13}$$

Why logarithm? Probability is *multiplicative*, while information is *additive*. The **logarithm** is only correct function to convert multiplication to addition.

- Observing a random variable X ($\mathcal{X} = \{a, b, c, d\}$) associated with the following (prior) probability

$$\begin{array}{ccccc} x & a & b & c & d \\ p(x) & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \end{array}. \tag{14}$$

Information can be different for different observation outcomes

$$\begin{aligned} &I(X = a) = -\log(1/2) = \log 2 = 1\text{ bit}, \\ &I(X = b) = -\log(1/4) = 2\log 2 = 2\text{ bit}, \\ &I(X = c) = -\log(1/8) = 3\log 2 = 3\text{ bit}, \\ &I(X = d) = -\log(1/8) = 3\log 2 = 3\text{ bit}. \end{aligned} \tag{15}$$

However, different outcome happens with different probability. What is amount of information that we can obtained from observing X *on average*?

$$\begin{aligned} &I(X) = I(X = a)\, p(a) + I(X = b)\, p(b) + I(X = c)\, p(c) + I(X = d)\, p(d) \\ &= \left(1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 3 \times \frac{1}{8}\right)\text{bit} \\ &= 1.75\text{ bit}. \end{aligned} \tag{16}$$

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

$$\boxed{I(X) = -\langle \log p(x) \rangle = -\sum_{x \in \mathcal{X}} p(x) \log p(x).} \tag{17}$$

▪ Entropy

The **Shannon entropy** of a random variable X measures its uncertainty (randomness), which is defined based on its probability distribution

$$\boxed{S(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x).} \tag{18}$$

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

$$S(X) \geq 0. \tag{19}$$

  - $S(X) = 0$ means the value of X is known for certain (no randomness).
  - Large $S(X)$ indicates large uncertainty in X.
- Entropy can change under observation, as observation can remove (or reduce) the uncertainty of a random variable.

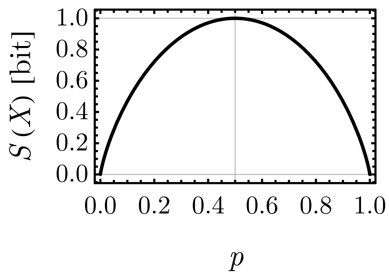
Example:

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

$$p(\text{false}) = 1 - p,$$
$$p(\text{true}) = p, \tag{20}$$

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

$$S(X) = -p \log p - (1-p)\log(1-p). \tag{21}$$



- $S(X) = 0$ 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.

- Rolling a dice ($\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$) and observe $X = 2$.

$$\begin{array}{cccccccl} x & 1 & 2 & 3 & 4 & 5 & 6 & \\ p(x) & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \text{(prior)} \\ p(x \mid X=2) & 0 & 1 & 0 & 0 & 0 & 0 & \text{(posterior)} \end{array} \tag{22}$$

  - Entropy of X *before* the observation (the *prior* entropy)

$$S(X_{\text{prior}}) = -\left(\frac{1}{6}\log\left(\frac{1}{6}\right)\right) \times 6 = \log 6. \tag{23}$$

  - Entropy of X *after* the observation (the *posterior* entropy)

$$S(X_{\text{post}}) = -((0\log 0) \times 5 + 1\log 1) = 0. \tag{24}$$

  Note: $0\log 0 = 0$ (as it should be understood as the limit $\lim_{x\to 0} x\log x = 0$), and $1\log 1 = 0$ (because $\log 1 = 0$).

  - Expected information gain from the observation

$$I(X) = \left(\frac{1}{6}\log\left(\frac{1}{6}\right)\right) \times 6 = \log 6. \tag{25}$$

Relation between **entropy** and **information**.

- Observation changes the probability distribution (i.e. the state of knowledge)

$$\begin{array}{cccc} & X_{\text{prior}} & \xrightarrow{\text{observe}} & X_{\text{post}} \\ \text{probability} & p(x) & & p(x \mid X) \\ \text{entropy} & S(X_{\text{prior}}) & & S(X_{\text{post}}) \\ & \text{more} & & \text{less} \end{array} \tag{26}$$

- The entropy change is the negative of the information gain

$$\Delta S = S(X_{\text{post}}) - S(X_{\text{prior}}) = -I(X). \tag{27}$$

Conclusion: entropy = negative information.

Quantum Information

Pure and Mixed States

A **pure state** describes a *deterministic* quantum system, modeled by

- a single **state vector** (ket state) $|\psi\rangle$,
- or the **state projector** (pure state **density matrix**)

$$\boxed{\hat{\rho} = \hat{\mathcal{P}}_\psi := |\psi\rangle\langle\psi|.} \tag{28}$$

A **mixed state** describes a *random* quantum system, consists of a random **ensemble** of *distinct* states, in which each state $|\phi\rangle$ appears with the **probability** $p(|\phi\rangle)$.

- The mixed state is modeled by a **density matrix** (density operator) ρ, defined as the ensemble *average* of the *state projectors*

$$\boxed{\hat{\rho} = \mathop{\mathbb{E}}_{\phi}[\hat{\mathcal{P}}_\phi] = \sum_\phi p(|\phi\rangle)\,|\phi\rangle\langle\phi|.} \tag{29}$$

- A density matrix should satisfy the following properties
  - **Hermitian**: $\hat{\rho}^\dagger = \hat{\rho}$.
  - **Normalization** (trace one): $\operatorname{Tr}\hat{\rho} = 1$.
  - **Positive (semi)definite**: $\forall\,|\psi\rangle : \langle\psi|\,\hat{\rho}\,|\psi\rangle \geq 0$.
- A *pure* state $|\psi\rangle$ can be viewed a special limit of a *mixed* state $\hat{\rho}$, when the density matrix **factorizes** as

$$\hat{\rho} = |\psi\rangle\langle\psi|. \tag{30}$$

  Meaning that the probability $p(|\phi\rangle)$ concentrated at on fixed state $|\phi\rangle = |\psi\rangle$. *Not* every density matrix can be expressed in the form of $|\psi\rangle\langle\psi| \Rightarrow$ A *density matrix* is more general than a *state vector*.

  - **Definition of mixed state**: a *mixed state* is a quantum state described by a density matrix $\hat{\rho}$ that can *not* be written as the *pure* state form $|\psi\rangle\langle\psi|$.

Example: *Polarization* of photon

- Examples of *pure* states:
  - σ^z basis states: *horizontal* and *vertical* polarizations

$$|\leftrightarrow\rangle \doteq \begin{pmatrix}1\\0\end{pmatrix},\ |\updownarrow\rangle \doteq \begin{pmatrix}0\\1\end{pmatrix}. \tag{31}$$

- σ^x basis states: 45° polarizations

$$|\nearrow\rangle \doteq \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix},\ |\searrow\rangle \doteq \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}. \tag{32}$$

- σ^y basis states: *circular* polarizations

$$|\circlearrowleft\rangle \doteq \frac{1}{\sqrt{2}}\begin{pmatrix}1\\i\end{pmatrix},\ |\circlearrowright\rangle \doteq \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\end{pmatrix}. \tag{33}$$

- Linear polarization along θ angle (with respect to x-axis)

$$|\theta\rangle \doteq \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix} \Rightarrow \hat{\rho}_\theta = |\theta\rangle\langle\theta| \doteq \begin{pmatrix}\cos^2\theta & \cos\theta\sin\theta\\ \cos\theta\sin\theta & \sin^2\theta\end{pmatrix}. \tag{34}$$

- *Natural light*: an ensemble of all possible polarizations with equal probability ⇒ *maximally mixed* state

$$\hat{\rho}_{\mathcal{N}} = \int \frac{d\theta}{2\pi}\hat{\rho}_\theta \doteq \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}. \tag{35}$$

This is an example of a *mixed* state.

Exc 1 Show that it is impossible to express $\hat{\rho}_{\mathcal{N}}$ as $|\psi\rangle\langle\psi|$, hence $\hat{\rho}_{\mathcal{N}}$ is indeed a mixed state.

The density matrix description of light polarization is also known as the **Jones matrix**.

Observable Expectation Value

Density matrix enables us to think about the expectation value of an observable O in a unified manner for both pure and mixed states.

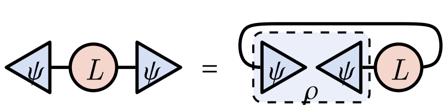
- Let $\hat{O}$ be a *Hermitian operator* describing the observable O,
- let $\hat{\rho}$ be the *density matrix* describing the state of a quantum system,
- observing O on the quantum system, the **expectation value** of O is given by

$$\boxed{\langle O\rangle = \operatorname{Tr}\hat{\rho}\,\hat{O}.} \tag{42}$$

Arguments:

- For pure state $\hat{\rho} = |\psi\rangle\langle\psi|$,

$$\langle O\rangle = \langle\psi|\,\hat{O}\,|\psi\rangle = \operatorname{Tr}|\psi\rangle\langle\psi|\,\hat{O} = \operatorname{Tr}\hat{\rho}\,\hat{O}. \tag{43}$$



- For mixed state $\hat{\rho} = \sum_\phi p(|\phi\rangle) |\phi\rangle \langle\phi|$,

$$\langle O\rangle = \sum_\phi p(|\phi\rangle) \langle\phi| \hat{O} |\phi\rangle = \sum_\phi p(|\phi\rangle) \operatorname{Tr} |\phi\rangle \langle\phi| \hat{O} = \operatorname{Tr} \hat{\rho} \hat{O}. \tag{44}$$

Quantum state tomography: reconstruction of the *density matrix* from (repeated) *measurements* on the systems taken from the *ensemble*. For a single qubit, by measuring $\langle\boldsymbol{\sigma}\rangle$, the density matrix can be reconstructed as

$$\boxed{\hat{\rho} = \frac{1}{2} (\mathbb{1} + \langle\boldsymbol{\sigma}\rangle \cdot \hat{\boldsymbol{\sigma}}).} \tag{45}$$

As $\hat{\rho}$ is the only solution of the density matrix that is *normalized* and *reproduces* the expectation values of all *measurements* on the qubit.

HW 1

> Check that the density matrix $\hat{\rho}$ in Eq. (45) is normalized $\operatorname{Tr} \hat{\rho} = 1$ and reproduces all measurement expectation values $\operatorname{Tr} \hat{\rho} \hat{\boldsymbol{\sigma}} = \langle\boldsymbol{\sigma}\rangle$.

Measurement

A *Bayesian* (subjectivism) view of quantum state.

- **Quantum state** (as generally modeled by the *density matrix* $\hat{\rho}$) describes our **state of knowledge** about the quantum system.
- The density matrix can be updated (the quantum state can collapse), if our state of knowledge is *changed* by **observations**.

A **projective measurement** of an observable O determines a *outcome* value $O_k \in$ eigenvalues of $\hat{O}$, with the probability $p(O = O_k \,|\, \rho)$. Recalled that every Hermitian operator $\hat{O}$ admits the following spectral decomposition

$$\hat{O} = \sum_k O_k \hat{\mathcal{P}}_{O=O_k}. \tag{46}$$

- Before observing $O = O_k$, the state of the quantum system is described by a **prior density matrix** $\hat{\rho}_{\text{prior}}$.
  - Based on this description, we can predict that the outcome O_k will appear with the probability given by

$$\boxed{p\big(O = O_k \,\big|\, \rho_{\text{prior}}\big) = \operatorname{Tr} \hat{\rho}_{\text{prior}} \hat{\mathcal{P}}_{O=O_k}.} \tag{47}$$

- After observing $O = O_k$, the quantum state will *collapse* to the **posterior density matrix** $\hat{\rho}_{\text{post}}$, which is given by

$$\boxed{\hat{\rho}_{\text{prior}} \to \hat{\rho}_{\text{post}} = \frac{\hat{\mathcal{P}}_{O=O_k}\,\hat{\rho}_{\text{prior}}\,\hat{\mathcal{P}}_{O=O_k}}{\operatorname{Tr}\hat{\rho}_{\text{prior}}\,\hat{\mathcal{P}}_{O=O_k}}.} \tag{48}$$

The denominator is a normalization factor to ensure $\operatorname{Tr}\hat{\rho}_{\text{post}} = 1$.

Dynamics

Time evolution (by time t) is implemented by a **unitary** operator $\hat{U}(t)$, under which the density matrix evolves as

$$\boxed{\hat{\rho}(t) = \hat{U}(t)\,\hat{\rho}(0)\,\hat{U}(t)^{\dagger}.} \tag{49}$$

For infinitesimal evolution

$$\hat{U}(dt) = e^{-i\,dt\,\hat{H}}, \tag{50}$$

Eq. (49) implies

$$\begin{aligned}\hat{\rho}(dt) &= e^{-i\,dt\,\hat{H}}\,\hat{\rho}(0)\,e^{i\,dt\,\hat{H}}\\ &= \hat{\rho}(0) - i\,dt\left[\hat{H}, \hat{\rho}(0)\right].\end{aligned} \tag{51}$$

Therefore the *time-evolution* of the *density matrix* follows the **von Neumann equation** (also known as the Liouville-von Neumann equation) ($\hbar$ is restored)

$$\boxed{i\,\hbar\,\partial_t\hat{\rho}(t) = \left[\hat{H}, \hat{\rho}(t)\right].} \tag{52}$$

- Here the density matrix is taken to be in the **Schrödinger picture**.
- Even though the *von Neumann equation* looks like the *Heisenberg equation* $i\,\hbar\,\partial_t\,\hat{O}(t) = \left[\hat{O}(t), \hat{H}\right]$ (which governs the operator evolution in the Heisenberg picture), but there is a crucial sign difference.
- However in the **Heisenberg picture**, the density matrix is *time-independent*, because the *state* does not evolve in the Heisenberg picture and the density matrix *follows the state*.

Example: Quantum decoherence.

Consider a single-qubit Hamiltonian $\hat{H} = \frac{\omega}{2}\,\hat{\sigma}^z$. Starting from the initial density matrix (in the diagonal basis of $\hat{H}$)

$$\hat{\rho}(0) \simeq \begin{pmatrix}\rho_{00} & \rho_{01}\\ \rho_{10} & \rho_{11}\end{pmatrix}. \tag{53}$$

Under time evolution (set $\hbar = 1$),

$$\hat{\rho}(t) \simeq \begin{pmatrix}\rho_{00} & \rho_{01}\,e^{-i\,\omega\,t}\\ \rho_{10}\,e^{i\,\omega\,t} & \rho_{11}\end{pmatrix}. \tag{54}$$

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

$$\begin{aligned}\overline{\rho} &= \frac{1}{T}\int_0^\infty \hat{\rho}(t)\,e^{-t/T}\,d t \\ &\doteq \begin{pmatrix} \rho_{00} & \frac{\rho_{01}}{1+i\,\omega\,T} \\ \frac{\rho_{10}}{1-i\,\omega\,T} & \rho_{11} \end{pmatrix} \xrightarrow{\omega\,T\gg 1} \begin{pmatrix} \rho_{00} & 0 \\ 0 & \rho_{11} \end{pmatrix}.\end{aligned} \tag{55}$$

The **off-diagonal** elements of the density matrix **decays** much more *quickly* than the **diagonal** elements, due to its fast oscillating phase (in this model).

- **Quantum Decoherence** (brief idea): the **loss** of *off-diagonal* density matrix elements (**quantum coherence**) over time in the *energy eigenbasis*.
- Decoherence generally takes a *pure* state to a *mixed* state (unless the pure state is an energy eigenstate) and is the fundamental source of **entropy production** (to be explained later).

■ Quantum Channel*

Transmitting a quantum state through a **quantum channel**, the density matrix $\hat{\rho}$ may undergo

- a **unitary evolution** (time evolution)

$$\hat{\rho} \to \hat{U}\,\hat{\rho}\,\hat{U}^\dagger, \tag{56}$$

- a **projective measurement** (measure and obtain a definite outcome)

$$\hat{\rho} \to \hat{\mathcal{P}}\,\hat{\rho}\,\hat{\mathcal{P}}, \quad \text{(without normalization)} \tag{57}$$

$$\hat{\rho} \to \frac{\hat{\mathcal{P}}\,\hat{\rho}\,\hat{\mathcal{P}}}{\mathrm{Tr}\left(\hat{\mathcal{P}}\,\hat{\rho}\,\hat{\mathcal{P}}\right)}, \quad \text{(with normalization)} \tag{58}$$

the normalization factor $\mathrm{Tr}\,\hat{\mathcal{P}}\,\hat{\rho}\,\hat{\mathcal{P}} = \mathrm{Tr}\,\hat{\rho}\,\hat{\mathcal{P}}$ is the probability to obtain the out come.

The *evolution* and *measurement* can be unified as **quantum operations**, described by a Kraus operator $\hat{K}$

$$\hat{\rho} \to \hat{K}\,\hat{\rho}\,\hat{K}^\dagger, \quad \text{(without normalization)} \tag{59}$$

$$\hat{\rho} \to \frac{\hat{K}\,\hat{\rho}\,\hat{K}^\dagger}{\mathrm{Tr}\left(\hat{K}\,\hat{\rho}\,\hat{K}^\dagger\right)}. \quad \text{(with normalization)} \tag{60}$$

A sequence of *quantum operations* put together forms a *quantum channel*.

Example: Quantum optics.

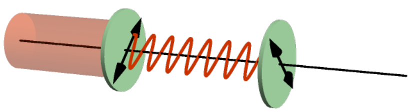
- Unitary evolution: *phase retarders* creates ϕ relative phase shift between horizontal and vertical polarizations

$$\hat{U}_\phi = e^{i \frac{\phi}{2} \hat{\sigma}^z} \doteq \begin{pmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{pmatrix}. \tag{61}$$

- Projective measurement: *polarizers* oriented along θ angle axis

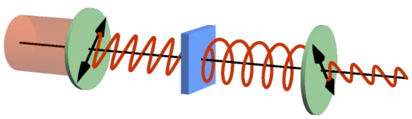
$$\hat{\mathcal{P}}_\theta = |\theta\rangle\langle\theta| \doteq \begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix}. \tag{62}$$

Natural light going through two perpendicular polarizers ⇒ no transmission.



$$\begin{aligned} \hat{\rho}' &= \hat{\mathcal{P}}_{-\pi/4}\hat{\mathcal{P}}_{\pi/4}\hat{\rho}\hat{\mathcal{P}}_{\pi/4}\hat{\mathcal{P}}_{-\pi/4} \\ &\doteq \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ &\Rightarrow \operatorname{Tr}\hat{\rho}' = 0. \end{aligned} \tag{63}$$

Insert a phase retarder between the polarizers ⇒ 1/4 transmission!



$$\begin{aligned} \hat{\rho}' &= \hat{\mathcal{P}}_{-\pi/4}\,\hat{U}_{\pi/2}\,\hat{\mathcal{P}}_{\pi/4}\hat{\rho}\hat{\mathcal{P}}_{\pi/4}\,\hat{U}^\dagger_{\pi/2}\,\hat{\mathcal{P}}_{-\pi/4} \\ &\doteq \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} e^{i\pi/4} & 0 \\ 0 & e^{-i\pi/4} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} e^{-i\pi/4} & 0 \\ 0 & e^{i\pi/4} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{8} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{1}{8} \end{pmatrix} \\ &\Rightarrow \operatorname{Tr}\hat{\rho}' = \frac{1}{4}. \end{aligned} \tag{64}$$

▪ von Neumann Entropy

As a *Hermitian* matrix, the **density matrix** also admits **spectral decomposition**

$$\boxed{\hat{\rho} = \sum_k p_k \hat{\mathcal{P}}_{\rho=p_k} \overset{*}{=} \sum_k p_k |\phi_k\rangle\langle\phi_k|.} \tag{65}$$

- When there is no degeneracy, eigen projector is simply $\hat{\mathcal{P}}_{\rho=p_k} = |\phi_k\rangle\langle\phi_k|$.
- The *eigenvectors* $|\phi_k\rangle$ form an *orthonormal basis*.
- The *eigenvalues* p_k has the physical meaning of *probability*, satisfying:
  - **Hermitian**: $\hat{\rho}^\dagger = \hat{\rho} \Leftrightarrow p_k \in \mathbb{R}$.
  - **Normalization** (trace one): $\mathrm{Tr}\,\hat{\rho} = 1 \Leftrightarrow \sum_k p_k = 1$.
  - **Positive (semi)definite**: $\forall\,|\psi\rangle : \langle\psi|\,\hat{\rho}\,|\psi\rangle \geq 0 \Leftrightarrow p_k \geq 0$.

The density matrix $\hat{\rho}$ describes a *random* quantum system, in which each *pure state* $|\phi_k\rangle$ is *prepared* with *probability* p_k.

- $\hat{\rho}$ is a **pure state**, iff there is only a single k for which $p_k = 1$ (and all the other $p_j = 0$ for $j \neq k$). For example, if $p_1 = 1$ and $p_2 = p_3 = \ldots = 0$, then $\hat{\rho} = |\phi_1\rangle\langle\phi_1|$ is a pure state.
- Otherwise, for generic distribution p_k, $\hat{\rho}$ is a **mixed state**.

Given that the eigenvalues p_k of a density matrix $\hat{\rho}$ from a **probability distribution**, the **Shannon entropy** associated with this distribution is

$$S = -\sum_k p_k \log p_k. \tag{66}$$

This entropy is known as the **von Neumann entropy** of a density matrix $\hat{\rho}$

$$\boxed{S(\hat{\rho}) = -\mathrm{Tr}\,\hat{\rho}\log\hat{\rho}.} \tag{67}$$

Exc 2 Show that Eq. (67) is equivalent to Eq. (66) given Eq. (65).

Example: Computing von Neumann entropy. Given a single-qubit density matrix (for $-1 \leq m \leq 1$)

$$\hat{\rho} \simeq \frac{1}{2}\begin{pmatrix} 1 & m \\ m & 1 \end{pmatrix}, \tag{72}$$

calculate its von Neumann entropy $S(\hat{\rho})$.

- Step 1: diagonalize the density matrix (find the eigenvalues)

$$\hat{\rho} \simeq \frac{1}{2}\begin{pmatrix} 1 & m \\ m & 1 \end{pmatrix} \to \begin{pmatrix} \frac{1+m}{2} & 0 \\ 0 & \frac{1-m}{2} \end{pmatrix} \tag{73}$$

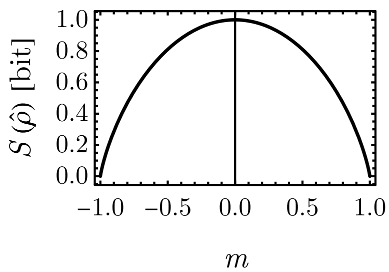
- Step 2: extract eigenvalues and arrange them into a probability distribution

$$p_1 = \frac{1+m}{2},\; p_2 = \frac{1-m}{2}. \tag{74}$$

  Make sure that $p_1 + p_2 = 1$.

- Step 3: evaluate the entropy using Eq. (66),

$$
\begin{aligned}
S(\hat{\rho}) &= -(p_1 \log p_1 + p_2 \log p_2)\\
&= -\left(\frac{1+m}{2} \log \frac{1+m}{2} + \frac{1-m}{2} \log \frac{1-m}{2}\right)\\
&= \log 2 - \frac{1}{2} \log\left(1 - m^2\right) - m \operatorname{arctanh} m.
\end{aligned}
$$



HW 2

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

Purity and Rényi Entropy

Purity of a density matrix $\hat{\rho}$ quantify to which degree the density matrix is pure/mixed,

$$\boxed{\operatorname{Tr} \hat{\rho}^2 = \sum_i p_i^2} \tag{76}$$

By construction, $\operatorname{Tr} \hat{\rho}^2 \in [0, 1]$. The criteria to determine if a density matrix $\hat{\rho}$ is pure or mixed is

$$\hat{\rho} \text{ is } \begin{cases} \text{pure} & \text{if } \operatorname{Tr} \hat{\rho}^2 = 1, \\ \text{mixed} & \text{if } \operatorname{Tr} \hat{\rho}^2 < 1. \end{cases} \tag{77}$$

Rényi entropy of a density matrix $\hat{\rho}$

$$\boxed{S^{(n)}(\hat{\rho}) = \frac{1}{1-n} \log \operatorname{Tr} \hat{\rho}^n.} \tag{78}$$

In terms of the *eigenvalues* p_i of the density matrix $\hat{\rho}$,

$$S^{(n)} = (1-n)^{-1} \log \sum_i p_i^n. \tag{79}$$

- 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. $S(\hat{\rho}) = \lim_{n\to 1} S^{(n)}(\hat{\rho})$.

Exc 3 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 $S^{(2)} = -\log \operatorname{Tr} \hat{\rho}^2$.
- $n = \infty$: **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).

$$\sum_i p_i^n = e^{(1-n)\, S^{(n)}} \quad (\text{for } n = 1, 2, \ldots, \dim \mathcal{H}). \tag{81}$$

▪ Entropy and Knowledge

Entropy *measures* our **ignorance** about the quantum system. If we describe a quantum system by:

- a **pure state**, we know that the system is in a *definite* quantum state ⇒ we have the **full knowledge** about the system ⇒ the system has no entropy, i.e. $S^{(n)}(\hat{\rho}) = 0$ (for all n).
- a **mixed state**, there are several possible states that the system can take (we are not sure) ⇒ we only have **partial knowledge** about the system ⇒ our *ignorance* gives rise to the *entropy* of the system.
- a **maximally mixed state** $\hat{\rho} = \mathbb{1} / \operatorname{Tr} \hat{\rho}$, we have **no knowledge** about the system, every state in the Hilbert space are *equally likely* ⇒ the entropy of the system is *maximized.*

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

$$\hat{\rho} \text{ is } \begin{cases} \text{pure} & \text{if } S^{(n)}(\hat{\rho}) = 0, \\ \text{mixed} & \text{if } S^{(n)}(\hat{\rho}) > 0, \end{cases} \text{ for } n = 1, 2, \ldots. \tag{82}$$

- Decoherence generally takes a *pure* state to a *mixed* state (unless the pure state is an energy eigenstate) and is the fundamental source of **entropy production** (to be explained later).

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

$$\ln \dim \mathcal{H} = S^{(0)} \geq S^{(1)} \geq S^{(2)} \geq \ldots \geq S^{(\infty)} \geq 0. \tag{83}$$

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

$$\forall\, i : p_i = \frac{1}{\dim \mathcal{H}} \Rightarrow \forall\, n \geq 0 : S^{(n)} = \log \dim \mathcal{H}. \tag{84}$$

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.

$$\hat{\rho} = \frac{1}{\dim \mathcal{H}}\, \mathbb{1}. \tag{85}$$

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: We have no knowledge about the qubit ⇒ no preferred spin direction, i.e. $\langle\boldsymbol{\sigma}\rangle = 0$. Then according to Eq. (45) (quantum state tomography),

$$\hat{\rho} = \frac{\mathbb{1}}{2} \doteq \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{86}$$

- Application: if the qubit basis corresponds to the **photon polarization**, then the density matrix in Eq. (86) describes the **natural light** ensemble of photons.
- All Rényi entropies are identically $\log 2$ for a maximally mixed qubit,

$$S^{(n)} = \frac{1}{1-n}\log\left(\frac{1}{2^n} + \frac{1}{2^n}\right) = \log 2 = 1\text{ bit}. \tag{87}$$

- 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:

state $\hat{\rho}$	pure ↔ maximally mixed		
purity $\mathrm{Tr}\,\hat{\rho}^2$	1	↔	$1/\dim\mathcal{H}$
entropy $S^{(n)}(\hat{\rho})$	0	↔	$\log\dim\mathcal{H}$
knowledge	max	↔	none

Quantum Entanglement

Two-Qubit Systems

Tensor Product Hilbert Space

Composition of Systems: The *Hilbert space* of a *combined* quantum system is the **direct product** of the *Hilbert space* of each *subsystem*.

Two quantum systems A and B are associated with the Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively,

$$\mathcal{H}_A = \mathrm{span}\,\{|i\rangle_A\},\ \mathcal{H}_B = \mathrm{span}\,\{|j\rangle_B\}, \tag{88}$$

the composite system $A \bigcup B$ will be associated with the Hilbert space

$$\mathcal{H}_{A\bigcup B} = \mathcal{H}_A \otimes \mathcal{H}_B = \mathrm{span}\,\{|i\rangle_A \otimes |j\rangle_B\} = \mathrm{span}\,\{|ij\rangle\}. \tag{89}$$

- Hilbert space **tensor product** ⇒ Hilbert space *dimension* **multiplies**

$$\dim\mathcal{H}_{A\bigcup B} = \dim\mathcal{H}_A \dim\mathcal{H}_B. \tag{90}$$

- Rule of **scalar product**: scalar product of two tensor product states equals the product of scalar products in each tensor Hilbert space

$$\langle ij \mid kl\rangle = \langle j|_B \otimes \langle i|_A\, |k\rangle_A \otimes |l\rangle_B = \langle i \mid k\rangle_A\, \langle j \mid l\rangle_B = \delta_{ik}\, \delta_{jl}. \tag{91}$$

The tensor product basis is still *orthonormal*.

- **Generic states** in $\mathcal{H}_{A\cup B}$

$$|v\rangle = \sum_{i,j} v_{ij}\, |ij\rangle. \tag{92}$$

where the vector element

$$v_{ij} = \langle ij|v\rangle. \tag{93}$$

- **Generic operators** in $\mathcal{H}_{A\cup B}$

$$\hat{O} = \sum_{i,j,k,l} |ij\rangle\, \hat{O}_{ij,kl}\, \langle kl|, \tag{94}$$

where the matrix (tensor) element

$$O_{ij,kl} = \langle ij|\, \hat{O}\, |kl\rangle. \tag{95}$$

- **Tensor product** of **states**. Suppose $|u\rangle = \sum_i u_i\, |i\rangle_A$, $|v\rangle = \sum_j v_j\, |j\rangle_B$

$$|u\rangle \otimes |v\rangle = \sum_{i,j} u_i\, v_j\, |i\rangle_A \otimes |j\rangle_B = \sum_{i,j} u_i\, v_j\, |ij\rangle. \tag{96}$$

  - Note: the **double index** ij labels a **single state** $|ij\rangle$.
  - Take two qubits for example

$$\begin{pmatrix} u_0 \\ u_1 \end{pmatrix} \otimes \begin{pmatrix} v_0 \\ v_1 \end{pmatrix} = \begin{pmatrix} u_0 \begin{pmatrix} v_0 \\ v_1 \end{pmatrix} \\ u_1 \begin{pmatrix} v_0 \\ v_1 \end{pmatrix} \end{pmatrix} = \left(\begin{array}{c} u_0\, v_0 \\ u_0\, v_1 \\ \hline u_1\, v_0 \\ u_1\, v_1 \end{array}\right). \tag{97}$$

- **Tensor product** of **operators**. Suppose $A = \sum_{i,j} |i\rangle_A\, A_{ij}\, \langle j|_A$, $B = \sum_{k,l} |k\rangle_B\, B_{kl}\, \langle l|_B$,

$$\begin{aligned} \hat{A} \otimes \hat{B} &= \sum_{i,j,k,l} A_{ij}\, B_{kl}\, |i\rangle_A\, |k\rangle_B\, \langle j|_A\, \langle l|_B \\ &= \sum_{i,j,k,l} A_{ij}\, B_{kl}\, |ik\rangle\, \langle jl|. \end{aligned} \tag{98}$$

  - Take two qubits for example

$$\begin{pmatrix} A_{00} & A_{01} \\ A_{10} & A_{11} \end{pmatrix} \otimes \begin{pmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{pmatrix}$$

$$= \begin{pmatrix} A_{00}\begin{pmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{pmatrix} & A_{01}\begin{pmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{pmatrix} \\ A_{10}\begin{pmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{pmatrix} & A_{11}\begin{pmatrix} B_{00} & B_{01} \\ B_{10} & B_{11} \end{pmatrix} \end{pmatrix}$$

$$= \left(\begin{array}{cc|cc} A_{00}\,B_{00} & A_{00}\,B_{01} & A_{01}\,B_{00} & A_{01}\,B_{01} \\ A_{00}\,B_{10} & A_{00}\,B_{11} & A_{01}\,B_{10} & A_{01}\,B_{11} \\ \hline A_{10}\,B_{00} & A_{10}\,B_{01} & A_{11}\,B_{00} & A_{11}\,B_{01} \\ A_{10}\,B_{10} & A_{10}\,B_{11} & A_{11}\,B_{10} & A_{11}\,B_{11} \end{array}\right).$$

▪ Two-Qubit States

Each qubit has *two* basis states $|0\rangle$ and $|1\rangle$ (forming a 2-dim Hilbert space) ⇒ two qubits together have *four* basis states

$$\begin{array}{cc|cc} & & \text{qubit } B & \\ & & |0\rangle & |1\rangle \\ \hline \text{qubit } A & |0\rangle & |00\rangle & |01\rangle \\ & |1\rangle & |10\rangle & |11\rangle \end{array} \tag{100}$$

The precise meaning of $|00\rangle$ is a **tensor product** of $|0\rangle_A$ and $|0\rangle_B$ states. In the *vector representation*,

$$|00\rangle = |0\rangle_A \otimes |0\rangle_B \doteq \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \left(\begin{array}{c} 1 \\ 0 \\ \hline 0 \\ 0 \end{array}\right). \tag{101}$$

Similarly,

$$\begin{aligned} |01\rangle &= |0\rangle_A \otimes |1\rangle_B \doteq \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \left(\begin{array}{c} 0 \\ 1 \\ \hline 0 \\ 0 \end{array}\right), \\ |10\rangle &= |1\rangle_A \otimes |0\rangle_B \doteq \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \left(\begin{array}{c} 0 \\ 0 \\ \hline 1 \\ 0 \end{array}\right), \\ |11\rangle &= |1\rangle_A \otimes |0\rangle_B \doteq \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \left(\begin{array}{c} 0 \\ 0 \\ \hline 0 \\ 1 \end{array}\right). \end{aligned} \tag{102}$$

These four *basis states* span the **two-qubit Hilbert space**.

A *generic state* in the *two-qubit Hilbert space* is a superposition of these four basis states,

$$|\psi\rangle = \psi_0\,|00\rangle + \psi_1\,|01\rangle + \psi_2\,|10\rangle + \psi_3\,|11\rangle \doteq \begin{pmatrix}\psi_0\\ \psi_1\\ \psi_2\\ \psi_3\end{pmatrix}. \tag{103}$$

Normalization is still expected: $\langle\psi\mid\psi\rangle = \sum_i |\psi_i|^2 = 1$.

- **Product state**: a state that can be *factorized* as a *tensor product* of *single-qubit states.*

Suppose $|u\rangle = u_0\,|0\rangle + u_1\,|1\rangle$ is a state of the *first* qubit and $|v\rangle = v_0\,|0\rangle + v_1\,|1\rangle$ is a state of the *second* qubit. A *two-qubit* **product state** takes the general form of

$$\begin{aligned}&|u\rangle \otimes |v\rangle = (u_0\,|0\rangle + u_1\,|1\rangle) \otimes (v_0\,|0\rangle + v_1\,|1\rangle)\\ &= u_0\,v_0\,|00\rangle + u_0\,v_1\,|01\rangle + u_1\,v_0\,|10\rangle + u_1\,v_1\,|11\rangle.\end{aligned} \tag{104}$$

The main feature of a *product state* is that each qubit behaves **independently** of the other: *measurement* or *unitary operation* of one qubit will *not affect* the other.

Not every state in the *two-qubit Hilbert space* can be written as *product state*. Why? Let us count the degrees of freedom:

- A generic state as $|\psi\rangle$ in Eq. (103) has *six* real parameters. $4\times 2-1-1=6$.
- A generic *product* state as $|u\rangle\otimes|v\rangle$ in Eq. (104) has only *four* real parameters. $(2\times 2-1-1)\times 2 = 4$.

A generic state has more freedom than a product state, the additional freedom has to do with **quantum entanglement**.

- **Entangled state**: any state that can *not* be factorized to *product states* are *entangled.*

Example: the state $\frac{1}{\sqrt{2}}\,(|01\rangle - |10\rangle)$ is entangled.

Exc 4

> Prove that $\frac{1}{\sqrt{2}}\,(|01\rangle - |10\rangle)$ can not be written as a product state.

Question: Is the state $\frac{1}{2}\,(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$ entangled?

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)

Two-Qubit Operators

Any *physical observable* of a two-qubit system is represented as a *Hermitian operator* acting on the two-qubit Hilbert space.

- **Single-qubit observables** (each as a $3\times 4\times 4$ tensor, extended from the $3\times 2\times 2$ Pauli tensor):

$$\begin{aligned}\hat{\boldsymbol{\sigma}}_A &= \left(\hat{\sigma}_A^x, \hat{\sigma}_A^y, \hat{\sigma}_A^z\right),\\ \hat{\boldsymbol{\sigma}}_B &= \left(\hat{\sigma}_B^x, \hat{\sigma}_B^y, \hat{\sigma}_B^z\right).\end{aligned} \tag{106}$$

- **Two-qubit observables** (joint measurements, $3\times3\times4\times4$ tensor):

$$\hat{\boldsymbol{\sigma}}_A \otimes \hat{\boldsymbol{\sigma}}_B = \begin{pmatrix} \hat{\sigma}_A^x\,\hat{\sigma}_B^x & \hat{\sigma}_A^y\,\hat{\sigma}_B^x & \hat{\sigma}_A^z\,\hat{\sigma}_B^x \\ \hat{\sigma}_A^x\,\hat{\sigma}_B^y & \hat{\sigma}_A^y\,\hat{\sigma}_B^y & \hat{\sigma}_A^z\,\hat{\sigma}_B^y \\ \hat{\sigma}_A^x\,\hat{\sigma}_B^z & \hat{\sigma}_A^y\,\hat{\sigma}_B^z & \hat{\sigma}_A^z\,\hat{\sigma}_B^z \end{pmatrix}. \tag{107}$$

The precise meaning of $\hat{\sigma}_A^x$ (4×4 matrix):

$$\hat{\sigma}_A^x \otimes \mathbb{1}_B \doteq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \left(\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right). \tag{108}$$

The precise meaning of $\sigma_A^z\,\sigma_B^y$ (4×4 matrix):

$$\hat{\sigma}_A^z \otimes \hat{\sigma}_B^y \doteq \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \left(\begin{array}{cc|cc} 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & i \\ 0 & 0 & -i & 0 \end{array}\right). \tag{109}$$

The *single-qubit observables* $\hat{\boldsymbol{\sigma}}_A$, $\hat{\boldsymbol{\sigma}}_B$, *two-qubit observables* $\hat{\boldsymbol{\sigma}}_A \otimes \hat{\boldsymbol{\sigma}}_B$ together with the *identity observable* $\mathbb{1}$ (altogether $3+3+3\times3+1 = 16$ observables) form the **complete set of observables** for a two-qubit system, i.e. any physical observables of a two-qubit system must be a linear superposition of these 16 *basis observables*.

A Two-Qubit Model

Two-qubit Heisenberg model. Consider *two qubits* governed by the *Hamiltonian*

$$H = \frac{J}{4}\,\hat{\boldsymbol{\sigma}}_A \cdot \hat{\boldsymbol{\sigma}}_B = \frac{J}{4}\left(\sigma_A^x\,\sigma_B^x + \sigma_A^y\,\sigma_B^y + \sigma_A^z\,\sigma_B^z\right). \tag{110}$$

First write down the matrix representation,

$$H \doteq \frac{J}{4}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{111}$$

Then diagonalize the Hamiltonian.

- Eigenvalue $E_s = -3\,J/4$: a unique eigenstate ⇒ **spin-singlet** state

$$|s\rangle = \frac{1}{\sqrt{2}}\left(|01\rangle - |10\rangle\right). \tag{112}$$

- Eigenvalue $E_t = J/4$: three degenerated eigenstates ⇒ **spin-triplet** states (there is a basis freedom here, we make the following choice)

$$|t_1\rangle = \frac{1}{\sqrt{2}} (|01\rangle + |10\rangle),$$
$$|t_2\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle),$$
$$|t_3\rangle = \frac{1}{\sqrt{2}} (|00\rangle - |11\rangle).$$

The *lowest energy eigenstate* is called the **ground state**, the rest of the eigenstates are **excited states**. In this model, assuming $J > 0$, the *ground state* is the *spin-singlet* state.

▪ The Spin-Singlet State

Use the *vector representation* of the *spin-single state*

$$|s\rangle = \frac{1}{\sqrt{2}} (|01\rangle - |10\rangle) \simeq \frac{1}{\sqrt{2}} (\,0\ 1\ -1\ 0\,)^{\intercal}. \tag{114}$$

- Expectation value of **single-qubit** observables

$$\begin{aligned} \langle s|\, \hat{\boldsymbol{\sigma}}_A \,|s\rangle &= (0, 0, 0), \\ \langle s|\, \hat{\boldsymbol{\sigma}}_B \,|s\rangle &= (0, 0, 0). \end{aligned} \tag{115}$$

- Expectation value of **two-qubit** observables

$$\langle s|\, \hat{\boldsymbol{\sigma}}_A \otimes \hat{\boldsymbol{\sigma}}_B \,|s\rangle = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{116}$$

There is something unusual!

- $|s\rangle$ is a **pure state** of the *two-qubit* system ⇒ the system is in a *definite* quantum state, *entropy* of the *entire system* = 0 ⇒ we have the *full knowledge* about the system.
- However $\langle s|\, \hat{\boldsymbol{\sigma}}_A \,|s\rangle = 0$ implies nothing is know about qubit A, because qubit A is in a **maximally mixed state** with maximal *entropy* of the *subsystem* (1bit) ⇒ we are *completely ignorant* about the subsystems. (Same argument applies for qubit B)

The phenomenon that we may know *everything* about a *quantum system* yet *nothing* about its *subsystems* is a demonstration of **quantum entanglement**.

- **Classical information** is stored *locally* (bit-by-bit) in every single classical bit. Knowing the entire system = knowing the state of every classical bit.
- **Quantum information** can be stored *jointly* in the *interrelations* among qubits, but *not locally* in single qubits. Knowing the entire system does not imply the knowledge of its subsystem.

■ Entanglement Entropy

The **entanglement entropy** of the qubit A in a two-qubit state $|\psi\rangle$ is given by

$$\boxed{S(\hat{\rho}_A) = -\operatorname{Tr} \hat{\rho}_A \log \hat{\rho}_A.} \tag{117}$$

where $\hat{\rho}_A$ is the **reduced density matrix** of qubit A obtained by *tracing out* qubit B in the full **density matrix** $|\psi\rangle\langle\psi|$

$$\boxed{\hat{\rho}_A = \operatorname{Tr}_B |\psi\rangle\langle\psi|.} \tag{118}$$

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

$$\boxed{S^{(n)}(\hat{\rho}_A) = \frac{1}{1-n} \log \operatorname{Tr} \hat{\rho}_A^n.} \tag{119}$$

Example I: take the **spin-singlet state**

$$|s\rangle = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle). \tag{120}$$

- Full density matrix

$$\hat{\rho} = |s\rangle\langle s| \doteq \frac{1}{2} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 & 0 \end{pmatrix} = \frac{1}{2} \left(\begin{array}{cc|cc} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ \hline 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right). \tag{121}$$

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

$$\hat{\rho}_A = \operatorname{Tr}_B |s\rangle\langle s|$$

$$\doteq \frac{1}{2} \begin{pmatrix} \operatorname{tr}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} & \operatorname{tr}\begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} \\ \operatorname{tr}\begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} & \operatorname{tr}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{122}$$

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

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

$$S(\hat{\rho}_A) = -\operatorname{Tr} \hat{\rho}_A \log \hat{\rho}_A = \log 2 = 1 \text{ bit}. \tag{123}$$

Example II: take the **product state**

$$|\psi\rangle = \frac{1}{2} (|00\rangle + |01\rangle + |10\rangle + |11\rangle). \tag{124}$$

- Full density matrix

$$\hat{\rho} = |\psi\rangle\langle\psi| \doteq \frac{1}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix}(1\ 1\ 1\ 1) = \frac{1}{4}\left(\begin{array}{cc|cc}1&1&1&1\\1&1&1&1\\\hline 1&1&1&1\\1&1&1&1\end{array}\right). \tag{125}$$

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

$$\begin{aligned}\hat{\rho}_A &= \mathrm{Tr}_B\,\hat{\rho}\\ &\doteq \frac{1}{4}\begin{pmatrix}\mathrm{tr}\begin{pmatrix}1&1\\1&1\end{pmatrix} & \mathrm{tr}\begin{pmatrix}1&1\\1&1\end{pmatrix}\\ \mathrm{tr}\begin{pmatrix}1&1\\1&1\end{pmatrix} & \mathrm{tr}\begin{pmatrix}1&1\\1&1\end{pmatrix}\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1&1\\1&1\end{pmatrix}.\end{aligned} \tag{126}$$

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

$$S(\hat{\rho}_A) = -\mathrm{Tr}\ \hat{\rho}_A \log \hat{\rho}_A = -(0\log 0 + 1\log 1) = 0\ \text{bit}. \tag{127}$$

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*.

pure state $|\psi\rangle$	product	$\leftrightarrow$	maximally entangled
$\hat{\rho}_A$	pure	$\leftrightarrow$	maximally mixed
EE $S^{(n)}(\hat{\rho}_A)$	0	$\leftrightarrow$	$\log\dim\mathcal{H}_A$
entanglement	none	$\leftrightarrow$	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) $\Rightarrow$ separating out a subsystem would lead to *lost* of *information* $\Rightarrow$ hence the *production* of (entanglement) *entropy*.

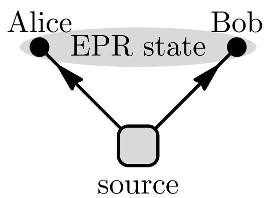
Open questions: The *system* must be *pure*, otherwise there are other source of entropy productions. What about entanglement in a *mixed* state? Good to describe *bipartite* entanglement. What about *multipartite* entanglement?

▪ EPR Pair and Bell Inequality

Bell states: *maximally entangled pure* states of two qubits. Also known as Einstein-Podolsky-Rosen **(EPR) pair** states. The *spin-singlet* state in Eq. (112) is one example:

$$|\mathrm{EPR}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \tag{128}$$

Suppose a machine can repeatedly *prepare* such EPR pairs and *distribute* the qubits separately to Alice and Bob,



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

- Alice defines her set of observables:

$$\hat{\boldsymbol{\sigma}}_A = \left(\hat{\sigma}_A^x, \hat{\sigma}_A^y, \hat{\sigma}_A^z\right). \tag{129}$$

- Bob defines his set of observables:

$$\hat{\boldsymbol{\sigma}}_B = \left(\hat{\sigma}_B^x, \hat{\sigma}_B^y, \hat{\sigma}_B^z\right). \tag{130}$$

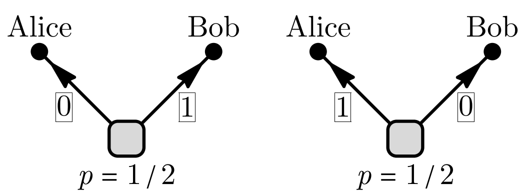
- The observables are *perfectly anti-correlated* between Alice and Bob

$$\begin{aligned}&\langle \text{EPR}|\, \hat{\boldsymbol{\sigma}}_A \,|\text{EPR}\rangle = \langle \text{EPR}|\, \hat{\boldsymbol{\sigma}}_B \,|\text{EPR}\rangle = (0, 0, 0),\\ &\langle \text{EPR}|\, \hat{\boldsymbol{\sigma}}_A \otimes \hat{\boldsymbol{\sigma}}_B \,|\text{EPR}\rangle = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.\end{aligned} \tag{131}$$

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

$$\sigma_A^z = -\sigma_B^z = \begin{cases} +1 & p = 1/2 \\ -1 & p = 1/2 \end{cases}. \tag{132}$$

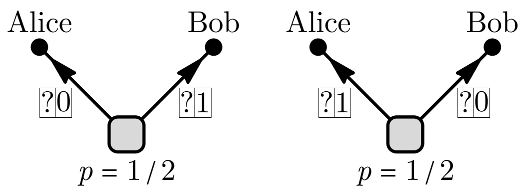
- *Quantum* explanation: can be inferred from $\langle \sigma_A^z \rangle = \langle \sigma_B^z \rangle = 0$ and $\langle \sigma_A^z \, \sigma_B^z \rangle = -1$.
- This is not too surprising: just a perfect anti-correlation between two random variables. *Classically*, one may model the perfect anti-correlation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



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

$$\sigma_A^x = \sigma_B^x = \begin{cases} +1 & p = 1/2 \\ -1 & p = 1/2 \end{cases}. \tag{133}$$

- *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 $|\text{EPR}\rangle$ state 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

$$\sigma_A^z = \begin{cases} +1 & p = 1/2 \\ -1 & p = 1/2 \end{cases}, \; \sigma_B^x = \begin{cases} +1 & p = 1/2 \\ -1 & p = 1/2 \end{cases}. \tag{134}$$

- *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

(135)

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

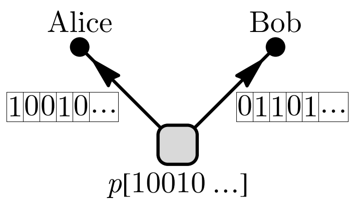
$\boldsymbol{n}_A \cdot \boldsymbol{\sigma}_A$	$\boldsymbol{n}_B \cdot \boldsymbol{\sigma}_B$	p
+1	−1	$(1 + \boldsymbol{n}_A \cdot \boldsymbol{n}_B)/4$
+1	+1	$(1 - \boldsymbol{n}_A \cdot \boldsymbol{n}_B)/4$
−1	−1	$(1 - \boldsymbol{n}_A \cdot \boldsymbol{n}_B)/4$
−1	+1	$(1 + \boldsymbol{n}_A \cdot \boldsymbol{n}_B)/4$

(136)

The probability that Alice and Bob obtain the same outcome is

$$p(\boldsymbol{n}_A \cdot \boldsymbol{\sigma}_A = \boldsymbol{n}_B \cdot \boldsymbol{\sigma}_B) = \frac{1 + \boldsymbol{n}_A \cdot \boldsymbol{n}_B}{2}. \tag{137}$$

- *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. (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

$$\begin{array}{lll} \text{Alice} & \text{Bob} & p \\ \hline \ldots 00 \ldots & \ldots 11 \ldots & (1 + \boldsymbol{n}_1 \cdot \boldsymbol{n}_2)/4 \\ \ldots 01 \ldots & \ldots 10 \ldots & (1 - \boldsymbol{n}_1 \cdot \boldsymbol{n}_2)/4 \\ \ldots 10 \ldots & \ldots 01 \ldots & (1 - \boldsymbol{n}_1 \cdot \boldsymbol{n}_2)/4 \\ \ldots 11 \ldots & \ldots 00 \ldots & (1 + \boldsymbol{n}_1 \cdot \boldsymbol{n}_2)/4 \end{array}. \tag{138}$$

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

$$\begin{array}{lll} \text{Alice} & \text{Bob} & p \\ \hline \ldots 000 \ldots & \ldots 111 \ldots & p_1 \\ \ldots 001 \ldots & \ldots 110 \ldots & p_2 \\ \ldots 010 \ldots & \ldots 101 \ldots & p_3 \\ \ldots 011 \ldots & \ldots 100 \ldots & p_4 \\ \ldots 100 \ldots & \ldots 011 \ldots & p_5 \\ \ldots 101 \ldots & \ldots 010 \ldots & p_6 \\ \ldots 110 \ldots & \ldots 001 \ldots & p_7 \\ \ldots 111 \ldots & \ldots 000 \ldots & p_8 \end{array}. \tag{139}$$

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

$$p_1 + p_2 + \ldots + p_8 = 1. \tag{140}$$

- 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 the opposite outcome is

$$p(\boldsymbol{n}_1 \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_2 \cdot \boldsymbol{\sigma}_B) = p_1 + p_2 + p_7 + p_8. \tag{141}$$

- 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 the opposite outcome is

$$p(\boldsymbol{n}_2 \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_3 \cdot \boldsymbol{\sigma}_B) = p_1 + p_4 + p_5 + p_8. \tag{142}$$

- 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 the opposite outcome is

$$p(\boldsymbol{n}_3 \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_1 \cdot \boldsymbol{\sigma}_B) = p_1 + p_3 + p_6 + p_8. \tag{143}$$

Put together,

$$\begin{aligned} &p(\boldsymbol{n}_1 \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_2 \cdot \boldsymbol{\sigma}_B) + p(\boldsymbol{n}_2 \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_3 \cdot \boldsymbol{\sigma}_B) + p(\boldsymbol{n}_3 \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_1 \cdot \boldsymbol{\sigma}_B) \\ &= 3\,p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + 3\,p_8 \\ &= 1 + 2\,p_1 + 2\,p_8 \end{aligned} \tag{144}$$

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

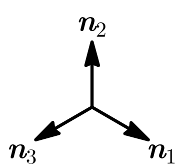
$$\boxed{p(\boldsymbol{n}_1 \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_2 \cdot \boldsymbol{\sigma}_B) + p(\boldsymbol{n}_2 \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_3 \cdot \boldsymbol{\sigma}_B) + p(\boldsymbol{n}_3 \cdot \boldsymbol{\sigma}_A = -\boldsymbol{n}_1 \cdot \boldsymbol{\sigma}_B) \geq 1.} \tag{145}$$

- Now what is the **quantum mechanical prediction**? Recall the *quantum* result in Eq. (137), the Bell inequality would require

$$\frac{1 + \boldsymbol{n}_1 \cdot \boldsymbol{n}_2}{2} + \frac{1 + \boldsymbol{n}_2 \cdot \boldsymbol{n}_3}{2} + \frac{1 + \boldsymbol{n}_3 \cdot \boldsymbol{n}_1}{2} \geq 1, \tag{146}$$

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.



$$\boldsymbol{n}_1 \cdot \boldsymbol{n}_2 = \boldsymbol{n}_2 \cdot \boldsymbol{n}_3 = \boldsymbol{n}_3 \cdot \boldsymbol{n}_1 = -1/2. \tag{147}$$

Then Eq. (146) would require

$$\frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} \geq 1, \tag{148}$$

which is not true.

The *violation* of the 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

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

Alain Aspect   John F. Clauser   Anton Zeilinger

How does Bell inequality tell us about entanglement?

Consider a two qubit state, parametrized by a phase angle α,

$$|\psi\rangle = \cos\alpha\,|01\rangle - \sin\alpha\,|10\rangle \doteq \begin{pmatrix} 0 \\ \cos\alpha \\ -\sin\alpha \\ 0 \end{pmatrix}. \tag{149}$$

```
ψ = {0, Cos[α], -Sin[α], 0};
```

- There are two limits:
  - $\alpha = \pi/4 \Rightarrow \cos\alpha = \sin\alpha = \frac{1}{\sqrt{2}} \Rightarrow |\psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ is the EPR state (maximally entangled).
  - $\alpha = 0 \Rightarrow \cos\alpha = 1$ and $\sin\alpha = 0 \Rightarrow |\psi\rangle = |01\rangle$ reduces to the product state (no entanglement).

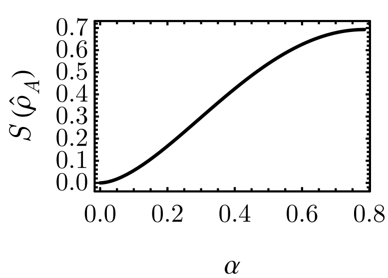
As α is tuned between these two limit, the amount of **quantum entanglement** can be adjusted. There must be a point where the amount of entanglement ("quantumness") is no longer sufficient to support the violation of the Bell inequality.

- **Entanglement entropy**

$$S(\hat{\rho}_A) = -\cos^2\alpha\log\cos^2\alpha - \sin^2\alpha\log\sin^2\alpha \tag{150}$$

```
ρ = ψ⊗ψ;
ρA = TensorContract[ArrayReshape[ρ, {2, 2, 2, 2}], {{2, 4}}];
SA = Total[-# Log[#] & /@ Eigenvalues@ρA]
```

```
-Cos[α]^2 Log[Cos[α]^2] - Log[Sin[α]^2] Sin[α]^2
```



- **Observable** expectation values

$\langle\psi|\,\hat{\boldsymbol{\sigma}}_A\,|\psi\rangle = (0, 0, \cos 2\alpha),$

$$\langle\psi|\,\hat{\boldsymbol{\sigma}}_B\,|\psi\rangle = (0, 0, -\cos 2\,\alpha),$$

$$\langle\psi|\,\hat{\boldsymbol{\sigma}}_A\otimes\hat{\boldsymbol{\sigma}}_B\,|\psi\rangle = \begin{pmatrix} -\sin 2\,\alpha & 0 & 0 \\ 0 & -\sin 2\,\alpha & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

```
Table[ψ.KroneckerProduct[PauliMatrix[a], PauliMatrix[b]].ψ, {a, 0, 3}, {b, 0, 3}] //
  FullSimplify // TableForm
```

1	0	0	-Cos[2 α]
0	-Sin[2 α]	0	0
0	0	-Sin[2 α]	0
Cos[2 α]	0	0	-1

- If Alice measures $\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_A$ and Bob measures $\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_B$, the probability for their measurement outcomes to be opposite is given by

$$p(\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_A = -\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_B) = \langle\psi|\,\hat{\mathcal{P}}_{\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_A=-\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_B}\,|\psi\rangle, \tag{152}$$

where the projection operator $\hat{\mathcal{P}}_{\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_A=-\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_B}$ is given by

$$\begin{aligned}\hat{\mathcal{P}}_{\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_A=-\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_B} &= \sum_{s=\pm1}\hat{\mathcal{P}}_{\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_A=s}\,\hat{\mathcal{P}}_{\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_B=-s}\\ &= \sum_{s=\pm1}\frac{\mathbb{1}+s\,\boldsymbol{n}_1\cdot\hat{\boldsymbol{\sigma}}_A}{2}\,\frac{\mathbb{1}-s\,\boldsymbol{n}_2\cdot\hat{\boldsymbol{\sigma}}_B}{2}\\ &= \frac{\mathbb{1}-\,\boldsymbol{n}_1\cdot(\hat{\boldsymbol{\sigma}}_A\otimes\hat{\boldsymbol{\sigma}}_B)\cdot\boldsymbol{n}_2}{2}.\end{aligned} \tag{153}$$

Therefore

$$p(\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_A = -\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_B) = \frac{1-\,\boldsymbol{n}_1\cdot\langle\psi|\,\hat{\boldsymbol{\sigma}}_A\otimes\hat{\boldsymbol{\sigma}}_B\,|\psi\rangle\cdot\boldsymbol{n}_2}{2}. \tag{154}$$

Suppose the unit vectors $\boldsymbol{n}_1$, $\boldsymbol{n}_2$, $\boldsymbol{n}_3$ are still placed in the xz-plane with 120° to each other

$$\begin{aligned}\boldsymbol{n}_1 &= (\cos\theta, 0, \sin\theta),\\ \boldsymbol{n}_2 &= (\cos(\theta+2\,\pi/3), 0, \sin(\theta+2\,\pi/3)),\\ \boldsymbol{n}_2 &= (\cos(\theta-2\,\pi/3), 0, \sin(\theta-2\,\pi/3)).\end{aligned} \tag{155}$$

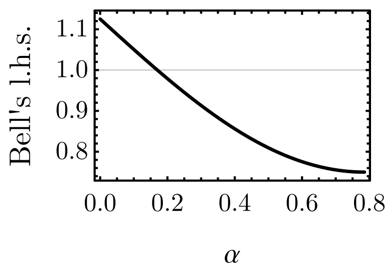
Then the left-hand-side (l.h.s.) of the Bell inequality reads

$$\begin{aligned}&p(\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_A = -\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_B) + p(\boldsymbol{n}_2\cdot\boldsymbol{\sigma}_A = -\boldsymbol{n}_3\cdot\boldsymbol{\sigma}_B) + p(\boldsymbol{n}_3\cdot\boldsymbol{\sigma}_A = -\boldsymbol{n}_1\cdot\boldsymbol{\sigma}_B)\\ &= \frac{3}{8}\,(3-\sin 2\,\alpha).\end{aligned} \tag{156}$$

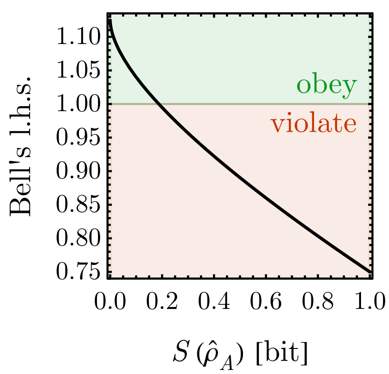
```
Simplify@Total[
  (1 - #1.DiagonalMatrix[{-Sin[2 α], -1}].#2) / 2 &@@@ Thread@{#, RotateLeft[#]} &@
   CirclePoints[{1, θ}, 3]]
```

$-\frac{3}{8}\left(-3+\text{Sin}[2\,\alpha]\right)$

It turns out that this result is independent of the choice of θ.



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 $S(\hat{\rho}_A) > 0$ as long as $\alpha > 0 \Leftrightarrow$ the state is *entangled*. But the Bell inequality is *not always* violated. $\Rightarrow$ 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.