Notes for Quantum Information
(TUM, WS 19/20)

Lecturer: Norbert Schuch
Typing: Yilun Yang

Contents

Chapter 1

Introduction

Quantum information is to study information processing with quantum mechanical systems. In the lecture, we will talk about

– What is quantum information / data?

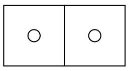
– How can we store / transmit / process it?

– What can we do with it?

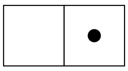
– How can we relate is physically?

Question Why do we study information processing with quantum mechanical systems? Isn't information independent of physical realization?

- *Landauer (1961): Erasing information releases heat.*



Particle in box at *unknown* position: 1 bit of information, entropy $S_0 = k \ln 2$.



Particle in box at *known* position: the bit erased, entropy $S_1 = 0$.

$$\Rightarrow \Delta S_{\text{sys}} = -k \ln 2 \Rightarrow \Delta Q_{\text{env}} = -T \Delta S_{\text{sys}} = kT \ln 2 \tag{1.1}$$

Erasing 1 bit releases $\Delta Q = kT \ln 2$ heat. Hence *information is physical.*

- Other motivation: *Moore's law.* The number of transistors / chips double every 18 months. This results in transistor size approaching atomic size, and we have to take into account quantum mechanics effects. It is better to use them!

1.1 Basic ideas of quantum information

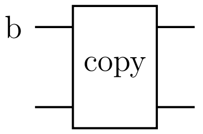
Quantum bits (qubits)

- Classical information
  basic unit: bit $b = 0, 1$ (2 possibilities).
  N bits: bits storing $b_1 \cdots b_N = 0 \cdots 0, 0 \cdots 01, \cdots$ (2^N possibilities!)

- Quantum information
  basic unit: quantum bit (qubit) $|b\rangle = \alpha |0\rangle + \beta |1\rangle$, with $\alpha, \beta \in \mathbb{C} \Rightarrow$ infinitely many possibilities!
  N qubits:

  $$|\boldsymbol{b}\rangle = \alpha_{0\cdots0} |0 \cdots 0\rangle + \alpha_{0\cdots01} |0 \cdots 01\rangle + \cdots \tag{1.2}$$

  2^N complex parameters! Can we store / extract ”infinitely much” information? How to quantify amount of information?

Cloning Can we copy information? Classically it works: 

,

but quantum mechanically NO! We expect

$$|0\rangle \xrightarrow{\text{copy}} |0\rangle \otimes |0\rangle \tag{1.3a}$$

$$|1\rangle \xrightarrow{\text{copy}} |1\rangle \otimes |1\rangle \tag{1.3b}$$

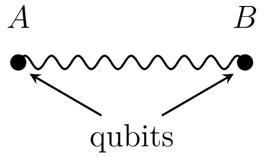
$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \xrightarrow{\text{copy}} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \tag{1.3c}$$

But from linearity, (1.3a)+(1.3b) $\Rightarrow$

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \quad \xrightarrow{\text{copy}} \tfrac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle) \tag{1.4}$$

Contradicting with (1.3c)! Quantum information cannot be cloned. This is called ”*No-cloning theorem*”. How can we then store / transmit quantum information? How can we deal with errors?

Entanglement, teleportation, Bell inequalities Consider two participants Alice (A) and Bob (B) sharing a quantum system with total state $|\psi\rangle_{AB} = \frac{1}{\sqrt{2}}(|0\rangle_A |1\rangle_B - |1\rangle_A |0\rangle_B)$.



1. Alice and Bob measure in $\{|0\rangle, |1\rangle\}$ basis: out comes are *perfectly anti-correlated*! Is this quantum feature? No, it happens also for classical variables.

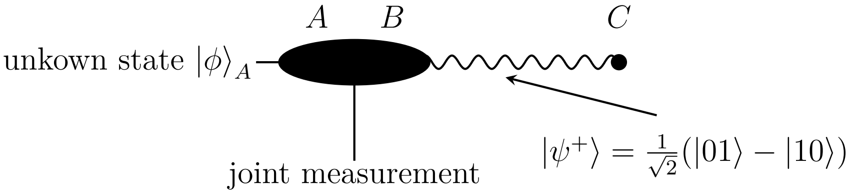
2. Out comes are also perfectly anti-correlated in *all* bases! Is this non classical?
   Still no: in local hidden variable model (LHV), each spin can also be described by a list of bits, one for each measurement direction (ϕ, θ) such that

   $$b_A(\phi,\theta) + b_B(\phi,\theta) = 1. \tag{1.5}$$

   $\Rightarrow$ perfect anti-correlated in *any basis* with ”classical model”.

3. But QM is actually incompatible with any LHV model (*Bell inequalities*)! QM cannot be described by a *local* and *realistic* model.

Teleportation Quantum information cannot be cloned — then how can one transport it over long distances?


A
B
C
unkown state $|\phi\rangle_A$
joint measurement
$|\psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$

Assume we have an unknown state $|\phi\rangle$ at A, and a correlated pair B and C separated by a long distance. Joint measurement of A and B will make $|\phi\rangle$ appear at C.

Notes

- The state of the system instead of the system itself is teleported.
- This does not allow faster-than-light communication. State in C is ”scrambled”; the measurement outcome is needed for unscrambling, which must be communicated classically.

Quantum computing By building computers acting on *qubits* rather than *classical bits* that can process *superposition* of exponentially many possibilities, can we obtain exponential speed-up?
It is tricky to *extract* information!
Shor (1994): quantum computer can factor numbers *exponentiall*y faster than any *known* classical method.

Quantum error correction Noise can destroy quantum information. How can we protect it?
Classically we can simply make copies or find some smarter methods. But quantum mechanically, copies can protect against only *bit flip*, but not *phase flip*:

$$\begin{aligned} |0\rangle &\to |000\rangle \\ |1\rangle &\to |111\rangle \\ \frac{|0\rangle + |1\rangle}{\sqrt{2}} &\to \frac{|000\rangle + |111\rangle}{\sqrt{2}} \end{aligned} \tag{1.6}$$

If there's phase flip $|0\rangle \to |0\rangle, |1\rangle \to -|1\rangle$ at one qubit, then

$$\frac{|000\rangle + |111\rangle}{\sqrt{2}} \to \frac{|000\rangle - |111\rangle}{\sqrt{2}} \approx \frac{|0\rangle - |1\rangle}{\sqrt{2}}! \tag{1.7}$$

Hence quantum error correction codes (QECC) is necessary.

Chapter 2

The Formalism, States, Measurements, Evolution

2.1 Pure states, unitary evolution, projective measurements

Quantum mechanical system A quantum mechanical system is the Hilbert space $\mathcal{H} \cong \mathbb{C}^d$. In quantum information, it is typically finite dimensional. We use the *ket-bra notation*:

$$\begin{aligned} |v\rangle \in \mathbb{C}^d &: \text{column vector} \\ \langle v| = (|v\rangle)^\dagger &: \text{row vector} \\ \langle w|v\rangle \in \mathbb{C} &: \text{scalar product} \end{aligned}$$

A state of the system is a vector $|\psi\rangle \in \mathcal{H}$ with $||\,|\psi\rangle\,||^2 = \langle\psi|\psi\rangle = 1$ and $|\psi\rangle \sim e^{i\phi}\,|\psi\rangle$, $\phi \in \mathbb{R}$.

Basis notation The *computational basis* of $\mathbb{C}^d$ is $\{|0\rangle\,, |1\rangle\,, \cdots |d-1\rangle\}$ with

$$|k\rangle \,\hat{=} e_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0, \end{pmatrix} \leftarrow k\text{'th position}, \quad |v\rangle = \sum_{k=0}^{d-1} v_k\,|k\rangle = \begin{pmatrix} v_0 \\ \vdots \\ v_{d-1} \end{pmatrix} \tag{2.1}$$

Linear operators $M : \mathbb{C}^d \to \mathbb{C}^d$ is *linear* if

$$M(\alpha |v\rangle + \beta |w\rangle) = \alpha M(|v\rangle) + \beta M(|w\rangle). \tag{2.2}$$

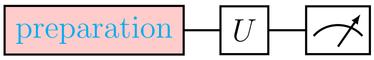
Write $M |v\rangle := M(|v\rangle)$.

Matrix notation / expansion

$$\begin{aligned} M =& (\sum_{i=0}^{d-1} |i\rangle \langle i|) M (\sum_{j=0}^{d-1} |j\rangle \langle i|) \\ =& \sum_{i,j=0}^{d-1} M_{ij} |i\rangle \langle j| = \begin{pmatrix} M_{00} & M_{01} & \cdots & M_{0,d-1} \\ M_{10} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ M_{d-1,0} & & & M_{d-1,d-1} \end{pmatrix} \end{aligned} \tag{2.3}$$

with $M_{ij} = \langle i|M|j\rangle$.

Quantum mechanics setup



1. Preparation: prepare a known initial state $|\phi\rangle \in \mathbb{C}^d$.

2. Evolution: act a unitary transform $U : \mathbb{C}^d \to \mathbb{C}^d$, $|\phi\rangle \mapsto U |\phi\rangle$, where $U^\dagger U = UU^\dagger = \mathbb{1}$.

Notes

(i) $\langle \phi|U^\dagger U|\phi\rangle = \langle \phi|\phi\rangle = 1 \Rightarrow$ the norm of the state is preserved if and only if U is unitary.

(ii) U can in principle be generated by time evolution with Hamiltonian.

3. Measurement: observable quantities $\equiv$ hermitian operator $A = A^\dagger$, with eigenvalue decomposition

$$A = \sum_n a_n E_n, \ E_n^2 = E_n = E_n^\dagger = |\psi_n\rangle \langle \psi_n| \,. \tag{2.4}$$

Measurement of A on state $|\phi\rangle$ leads to outcome a_n with probability $p_n = \langle \phi|E_n|\phi\rangle = |\langle \psi_n|\phi\rangle|^2$ and post-measurement state

$$|\phi_n\rangle = E_n |\phi_n\rangle /||E_n |\phi\rangle ||. \tag{2.5}$$

Note that $\sum_n p_n = \langle\phi|\sum_n E_n|\phi\rangle = \langle\phi|\phi\rangle = 1$.
The expectation value is

$$\langle\phi|A|\phi\rangle = \sum_n a_n \langle\phi|E_n|\phi\rangle \,. \tag{2.6}$$

Example

- The Hilbert space for a single *qubit* is $\mathcal{H} = \mathbb{C}^2$, with $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$, $|\alpha|^2 + |\beta|^2 = 1$.

- An *observable* is $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \underbrace{|0\rangle\langle 0|}_{E_0} - \underbrace{|1\rangle\langle 1|}_{E_1}$ with $a_0 = +1$ and $a_1 = -1$. The *measurement* will have outcome $a_0 = +1$ with probability $\langle\psi|E_0|\psi\rangle = |\alpha|^2$ and $a_1 = -1$ with probability $|\beta|^2$, respectively.
  Another observable is $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = |0\rangle\langle 1| + |1\rangle\langle 0| = \underbrace{|+\rangle\langle +|}_{E_+} - \underbrace{|-\rangle\langle -|}_{E_-}$ with $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$, the measurement will have outcome ± 1 with probability $|\langle\pm|(\alpha\,|0\rangle + \beta\,|1\rangle)\rangle|^2 = |\alpha \pm \beta|^2/2$.

- The *Hadamard gate* $H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is an *evolution* operator,

$$H\,|\psi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}(\alpha\,|0\rangle + \beta\,|1\rangle) = \frac{\alpha+\beta}{\sqrt{2}}\,|0\rangle + \frac{\alpha-\beta}{\sqrt{2}}\,|1\rangle \,. \tag{2.7}$$

  Measure this state in Z-basis $\{|0\rangle, |1\rangle\}$, and we will get outcome 0 with probability $|\alpha+\beta|^2/2$ and 1 with probability $|\alpha+\beta|^2/2$. H transforms between X and Z basis. In fact,

$$H = |+\rangle\langle 0| + |-\rangle\langle 1| = |0\rangle\langle +| + |1\rangle\langle -| = H^{\tau}. \tag{2.8}$$

2.2 Composite systems

Formalism Consider a system with two separate parts (”subsystems”) Alice (A) and Bob (B). The joint system has Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. Let $|i\rangle_A$ and $|j\rangle_B$ be basis of $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively, then $|i\rangle_A \otimes |j\rangle_B \equiv |i\rangle_A |j\rangle_B \equiv |i,j\rangle_{AB} \equiv |ij\rangle_{AB} \equiv |ij\rangle$ is a basis of $\mathcal{H}_{AB} = \mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B} \cong \mathbb{C}^{\mathbb{d}_\mathbb{A}\mathbb{d}_\mathbb{B}}$. A general state of the composite system would be

$$|\phi\rangle_{AB} = \sum_{i,j} c_{ij}\,|i\rangle_A\,|j\rangle_B \,. \tag{2.9}$$

What if Alice acts with M_A on her system and Bob with N_B in his? (M_A, N_B could be unitaries, measurements or doing nothing, i.e., $\mathbb{1}$.)
Consider first $|\phi\rangle_{AB} = |i\rangle_A \otimes |j\rangle_B$. Action of Alice should only change *her* system as if Bob wasn't there:

$$|i\rangle_A \mapsto M_A |i\rangle_A , \quad |i\rangle_A \otimes |j\rangle_B \mapsto (M_A |i\rangle_A) \otimes |j\rangle_B \tag{2.10}$$

and it is the same for Bob, jointly

$$|i\rangle_A \otimes |j\rangle_B \mapsto (M_A |i\rangle_A) \otimes (N_B |j\rangle_B). \tag{2.11}$$

By *linearity*,

$$|\phi\rangle_{AB} \mapsto M_A |i\rangle_A \otimes N_B |j\rangle_B \equiv (M_A \otimes N_B) |\phi\rangle_{AB} . \tag{2.12}$$

In matrix elements,

$$M_A \otimes N_B = \left(\begin{array}{c|c|c} M_{A00} \cdot N_B & \cdots & M_{A0,d-1} \cdot N_B \\ \hline \vdots & \ddots & \vdots \\ \hline M_{Ad-1,0} \cdot N_B & \cdots & M_{Ad-1,d-1} \cdot N_B \end{array}\right). \tag{2.13}$$

Example For a bipartite state

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) = \frac{1}{\sqrt{2}}(|+-\rangle - |-+\rangle), \tag{2.14}$$

if

1. both Alice and Bob measure Z:
$$\begin{aligned} |\langle 00|\psi\rangle|^2 = |\langle 11|\psi\rangle|^2 = 0, \\ |\langle 10|\psi\rangle|^2 = |\langle 01|\psi\rangle|^2 = 1/2. \end{aligned} \tag{2.15}$$

2. both Alice and Bob measure X:
$$\begin{aligned} |\langle + + |\psi\rangle|^2 = |\langle - - |\psi\rangle|^2 = 0, \\ |\langle + - |\psi\rangle|^2 = |\langle - + |\psi\rangle|^2 = 1/2. \end{aligned} \tag{2.16}$$

3. Alice measures X, Bob measures Z:
$$|\langle +0|\psi\rangle|^2 \quad = |\langle +1|\psi\rangle|^2 = |\langle -0|\psi\rangle|^2 = |\langle -1|\psi\rangle|^2 = 1/4. \tag{2.17}$$

The outcomes for Alice and Bob are separately completely random, but there is perfect anti-correlation when measured in *the same basis*.

2.3 Mixed States

Consider a bipartite state $|\psi\rangle_{AB} = \sum c_{ij} |i\rangle |j\rangle$. We have only access to A. How can we characterize the system by measurement on A?
The measurement M on A $\iff$ measurement $M_A \otimes \mathbb{1}_B$ on A+B.

$$\begin{aligned}\langle\psi|M_A \otimes \mathbb{1}_B|\psi\rangle &= \sum_{ij,i'j'} c^*_{i'j'} \langle i'| \langle j'|M_A \otimes \mathbb{1}_B|i\rangle |j\rangle c_{ij} \\ &= \sum_{ij,i'j'} c_{ij} c^*_{i'j'} \langle i'|M_A|i\rangle \langle j'|j\rangle \\ &= \sum_{ii'} (\sum_j c_{ij} c^*_{i'j}) \langle i'|M_A|i\rangle \\ &= \mathrm{tr}[\rho_A M]\end{aligned} \tag{2.18}$$

by defining ρ_A via

$$(\rho_A)_{ii'} = \sum_j c_{ij} c^*_{i'j} = CC^\dagger, \tag{2.19}$$

or equivalently

$$\rho_A = \sum_{ii'j} c_{ij} c^*_{i'j} |i\rangle \langle i'|, \tag{2.20}$$

with the trace: $\mathrm{tr}(X) = \sum_k |k|X|k\rangle$ for some orthonormal basis (ONB) $\{|k\rangle\}$. Note that trace is cyclic and thus basis-independent: $\mathrm{tr}(X) = \mathrm{tr}(U^\dagger U X) = \mathrm{tr}(UXU^\dagger)$.

Definition 1. *ρ_A is called the* ***density operator*** *or* ***density matrix*** *of a mixed state. It characterizes the systems that we only have* ***partial knowledge****.*

Properties of ρ_A

- $\rho_A = CC^\dagger \Rightarrow \rho_A^\dagger = (CC^\dagger)^\dagger = CC^\dagger = \rho_A$.
- ρ_A is positive semi-definite (all eigenvalues ≥ 0) or $\rho_A \geq 0$, since $\langle\phi|\rho_A|\phi\rangle = (C^\dagger |\phi\rangle^\dagger)(C^\dagger |\phi\rangle) \geq 0$, $\forall |\phi\rangle$.
- $\mathrm{tr}(\rho_A) = \sum_i (CC^\dagger)_{ii'} = \sum_{ij} C_{ij} C^*_{ij} = \langle\psi|\psi\rangle = 1$.

Note

As a consequence, for $0 < p < 1$, ρ and ρ' being density operators, $p\rho + (1-p)\rho'$ is also density operator $\Rightarrow$ density operators form a convex set!

Is ρ_A uniquely determined by

$$\mathrm{tr}(M\rho_A) = \langle\psi|_{AB} M \otimes \mathbb{1} |\psi\rangle_{AB}? \tag{2.21}$$

Yes. $\mathrm{tr}(X^\dagger Y)$ is a scalar product and the overlap of ρ_A with all hermitian M determines the hermitian part of ρ_A *entirely*! (As a consequence, all numbers in ρ_A are meaningful, and there is no phase ambiguity!)
For a *pure state* $|\phi\rangle_A$, $\rho_A = \langle\phi|M|\phi\rangle = \mathrm{tr}\,\langle\phi|M|\phi\rangle = \mathrm{tr}\left[M\,|\phi\rangle\,\langle\phi|\right] \Rightarrow \rho = |\phi\rangle\,\langle\phi|$, which is a projector onto $|\phi\rangle$.

Partial trace What about a *general* state ρ_{AB} in A + B (e.g., $\rho_{AB} = |\psi\rangle\,\langle\psi|$)?

$$\begin{aligned} \mathrm{tr}\left[(M\otimes\mathbb{1})\rho_{AB}\right] &= \sum_{ij,i'j'} \langle ij|M\otimes\mathbb{1}|i'j'\rangle\,\langle i'j'|\rho_{AB}|ij\rangle \\ &= \sum_{ii'j} \langle i|M|i'\rangle\,\langle i'j|\rho_{AB}|ij\rangle = \mathrm{tr}(M\rho_A), \end{aligned} \tag{2.22}$$

where we define the partial trace

$$\begin{aligned} \rho_A =& \sum_j |i'\rangle\,\langle i'j|\rho_{AB}|ij\rangle\,\langle i| \\ =& \sum_j (\mathbb{1}_A \otimes \langle j'|_B)\rho_{AB}(\mathbb{1}_A \otimes |j\rangle_B) \\ =& \sum_j \langle j|_B \rho_{AB} |j\rangle_B = \qquad\qquad : \mathrm{tr}_B\rho_{AB}. \end{aligned} \tag{2.23}$$

In components, $(\mathrm{tr}_B\rho_{AB})_{ii'} = \sum_j (\rho_{AB})_{(ij)(i'j)}$.
Is any density matrix ρ *physical*? Take $\rho = \sum_i \lambda_i\,|\phi_i\rangle\,\langle\phi_i|$ in eigenvalue decomposition, and let $|\psi\rangle_{AB} = \sum_i \sqrt{\lambda_i}\,|\phi_i\rangle_A\,|i\rangle_B$ (*purification* of ρ). Then

$$\begin{aligned} \mathrm{tr}_B\left[|\psi\rangle_{AB}\,\langle\psi|_{AB}\right] =& \mathrm{tr}_B\left[\sum_{ij} \sqrt{\lambda_i\lambda_j}\,|\phi_i\rangle\,\langle\phi_j| \otimes |i\rangle\,\langle j|\right] \\ =& \sum_i \lambda_i\,|\phi_i\rangle\,\langle\phi_i| = \rho \Rightarrow \text{yes!} \end{aligned} \tag{2.24}$$

The density matrix can serve as *alternative definition* of a state.

Ensemble interpretation of density matrix Consider $|\psi\rangle_{AB} = \alpha\,|00\rangle + \beta\,|11\rangle \Rightarrow \rho_A = \begin{pmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{pmatrix} = |\alpha|^2\,|0\rangle\,\langle 0| + |\beta|^2\,|1\rangle\,\langle 1|$,

$$\mathrm{tr}\,[M\rho_A] = |\alpha|^2 \mathrm{tr}\,[M\,|0\rangle\,\langle 0|] + |\beta|^2 \mathrm{tr}\,[M\,|1\rangle\,\langle 1|]\,. \tag{2.25}$$

This can be interpreted as having $|0\rangle$ with probability $p_0 = |\alpha|^2$ and $|1\rangle$ with $p_1 = |\beta|^2$: *"ensemble interpretation"*. It this consistent with that $|\psi\rangle_{AB}$ is a *pure* state?
Let Bob do projective measurement in Z-basis:

$$|\psi\rangle = \alpha\,|00\rangle + \beta\,|11\rangle \xrightarrow{Z-\text{measurement on B}} \begin{cases} |\psi_0\rangle_A = |0\rangle_A\,,\ p_0 = |\alpha|^2; \\ |\psi_1\rangle_A = |1\rangle_A\,,\ p_1 = |\beta|^2. \end{cases} \tag{2.26}$$

Since Alice doesn't know the outcome, she will get an *ensemble*

$$\{(p_0; |0\rangle), (p_1; |1\rangle)\} = \begin{pmatrix} |\alpha|^2 & 0 \\ 0 & |\beta|^2 \end{pmatrix}. \tag{2.27}$$

Note

Bob's description is different. He knows the outcome and would describe the state as *either* $|0\rangle\,\langle 0|$ *or* $|1\rangle\,\langle 1|$.

But Bob could also measure in $|\pm\rangle$ basis!

$$|\psi\rangle = \alpha\,|00\rangle + \beta\,|11\rangle \xrightarrow{X-\text{meas.}} \begin{cases} |\psi_+\rangle_A = \alpha\,|0\rangle + \beta\,|1\rangle\,,\ p_+ = \dfrac{1}{2}; \\ |\psi_-\rangle_A = \alpha\,|0\rangle - \beta\,|1\rangle\,,\ p_- = \dfrac{1}{2}, \end{cases} \tag{2.28}$$

which gives a *different* ensemble

$$\rho_A = p_+\,|\psi_+\rangle\,\langle\psi_-| + p_-\,|\psi_-\rangle\,\langle\psi_-| \tag{2.29}$$

for the same state $\Rightarrow$ *ensemble interpretation is ambiguous*! The number of terms can vary, and the states can be non-orthogonal.

How are different ensembles related?

Theorem 1. *Let* $\rho = \sum_i p_i\,|\psi_i\rangle\,\langle\psi_i| = \sum_j q_j\,|\phi)j\rangle\,\langle\psi_j|$. *Then there exists a unitary* $U = (u_{ij})$ *such that*

$$\sqrt{p_i}\,|\psi_i\rangle = \sum_j u_{ij}\sqrt{q_j}\,|\phi_j\rangle \tag{2.30}$$

and vice versa. If there are less j's than i's, pad with zeros and vice versa.

Proof. $\Leftarrow$: Let $\sqrt{p_i}\,|\psi_i\rangle = \sum_j u_{ij}\sqrt{q_j}\,|\phi_j\rangle$. Then

$$\begin{aligned}\sum_i p_i\,|\psi_i\rangle\langle\psi_i| &= \sum_i(\sum_j u_{ij}\sqrt{q_j}\,|\phi_j\rangle)(\sum_{j'} u^*_{ij'}\sqrt{q_{j'}}\,\langle\phi_{i'}|)\\ &= \sum_{jj'}\sqrt{q_jq_{j'}}\,|\phi_j\rangle\langle\phi_{j'}|\,\underbrace{(\sum_i u^*_{ij'}u_{ij})}_{=\delta_{j,j'}}\\ &= \sum_j q_j\,|\phi_j\rangle\langle\phi_j|\,.\end{aligned} \tag{2.31}$$

$\Rightarrow$: see later (equivalence of purification).

□

2.4 Schmidt decomposition and purifications

2.4.1 Schmidt decomposition

Given $|\psi\rangle_{AB}$ a bipartite, let

$$\mathrm{tr}_B\,|\psi\rangle\langle\psi| = \rho_A = \sum p_i\,|i\rangle_A\langle i|_A \tag{2.32}$$

with $|i\rangle_A$ being ONB. Choose some ONB of B, expand

$$|\psi\rangle_{AB} = \sum_{i,j} c_{ij}\,|i\rangle_A\,|a_j\rangle_B = \sum_i |i\rangle_A\,\underbrace{(\sum_j c_{ij}\,|a_j\rangle_B)}_{=:|b_i\rangle_B} = \sum_i |i\rangle_A\,|b_i\rangle_B\,. \tag{2.33}$$

Insert this into Eq. (2.32), and we get

$$\langle b_{i'}|b_i\rangle = \delta_{i',i}p_i. \tag{2.34}$$

Hence $|i\rangle_B \equiv \frac{1}{\sqrt{p_i}}\,|b_i\rangle_B = \frac{1}{\sqrt{p_i}}\sum_j c_{ij}\,|a_j\rangle_B$ is an ONB for B.
We get the ***Schmidt decomposition***: any $|\psi\rangle_{AB}$ can be written as

$$|\psi\rangle_{AB} = \sum_i \lambda_i\,|i\rangle_a\,|i\rangle_B\,, \tag{2.35}$$

with ONBs $|i\rangle_A$ and $|i\rangle_B$. The $\lambda_i = \sqrt{p_i} \geq 0$ are called Schmidt coefficients.

Notes

$\rho_B = \mathrm{tr}_A \left|\psi\right\rangle \left\langle\psi\right| = \sum_i \lambda_i^2 \left|i\right\rangle_B \left\langle i\right|_B$, so $\left|i\right\rangle_B$ is an eigenbasis of ρ_B! If p_i is non-degenerate, Schmidt decomposition can be obtained by pairing eigenvectors of ρ_A and ρ_B. Another important consequence is that, for pure states $\left|\psi\right\rangle_{AB}$, ρ_A and ρ_B have the *same eigenvalues*!

Question How is the Schmidt decomposition related to other expansions $\left|\psi\right\rangle_{AB} = \sum_k \lambda_k \left|k\right\rangle_A \left|k\right\rangle_B = \sum_{ij} c_{ij} \left|x_i\right\rangle_A \left|y_j\right\rangle_B$ with ONBs $\left|x_i\right\rangle_A$ and $\left|y_j\right\rangle_B$? Assume that

$$\left|k\right\rangle_A = \sum_i u_{ik} \left|x_i\right\rangle_A\,,\ \left|k\right\rangle_B = \sum_j v_{jk}^* \left|y_j\right\rangle_B \tag{2.36}$$

(pad with zeroes if necessary), then we get

$$\begin{aligned}\sum_{ij} c_{ij} \left|x_i\right\rangle_A \left|y_j\right\rangle_B &= \sum_{ijk} \lambda_k u_{ik} v_{jk}^* \left|x_i\right\rangle_A \left|y_j\right\rangle_B \\ \xRightarrow{\text{lin. indep. of } \left|x_i\right\rangle \left|y_j\right\rangle} c_{ij} &= \sum_k u_{ik} \lambda_k v_{jk}^*, \\ \text{or } C &= UDV^\dagger,\end{aligned} \tag{2.37}$$

where U and V are unitaries, D only have λ_i on its diagonal. This is the ***singular value decomposition (SVD)*** of C.

Remark

Any two states $\left|\phi\right\rangle$ and $\left|\psi\right\rangle$ with identical Schmidt coefficients are related by local unitaries, i.e.,

$$\exists U, V : \left|\phi\right\rangle = U \otimes V \left|\psi\right\rangle . \tag{2.38}$$

(ONBs can be related with unitaries.) Hence the λ_i contain all non-local properties $\lambda_1 \geq \lambda_2 \geq \cdots$.

2.4.2 Purification

Definition 2. *Any states $\left|\psi\right\rangle_{AB}$ such that $\mathrm{tr}_B \left|\psi\right\rangle_{AB} \left|\psi\right\rangle_{AB} = \rho_A$ is called a* ***purification*** *of ρ_A.*

e.g., for $\rho_A = \sum_i p_i \left|\psi\right\rangle_i \left\langle\psi\right|_i$, $\sum_i \sqrt{p_i} \left|\psi\right\rangle_i \left|i\right\rangle$ is a purification.

Given two purifications $|\phi\rangle$ and $|\psi\rangle$ of ρ_A, what is their relation? Write them in *Schmidt form*:

$$\begin{aligned} |\phi\rangle &= \sum_i \lambda_i \left|\phi_i\right\rangle_A \left|\phi_i\right\rangle_B \\ |\psi\rangle &= \sum_i \mu_i \left|\psi_i\right\rangle_A \left|\psi_i\right\rangle_B \end{aligned} \tag{2.39}$$

with λ_i, μ_i in descending order. $\mathrm{tr}_B \, |\phi\rangle \, |\phi\rangle = \mathrm{tr}_B \, |\psi\rangle \, |\psi\rangle$ gives $\lambda_i = \mu_i$ and $\left|\phi_i\right\rangle_A = \left|\psi_i\right\rangle_A$ (up to a phase) if λ_i is non-degenerate. Now we can choose a local unitary U such that $U \left|\phi_i\right\rangle_B = \left|\psi_i\right\rangle_B$ for any i. Then

$$|\psi\rangle = (\mathbb{1} \otimes U) \, |\phi\rangle \, . \tag{2.40}$$

All purifications are related by a unitary on the purifying system.

Question How does a mixed state ρ_A evolve under a unitary U_A? Consider purification $|\psi\rangle_{AB}$, and $|\psi\rangle \mapsto (U_A \otimes \mathbb{1}) \, |\psi\rangle$. Then

$$\rho_A = \mathrm{tr}_B |\psi\rangle \, |\psi\rangle \mapsto \mathrm{tr}_B \left[(U_A \otimes \mathbb{1}) \, |\psi\rangle \, \langle\psi| \, (U_A^\dagger \otimes \mathbb{1}_B) \right] = U_A \rho_A U_A^\dagger . \tag{2.41}$$

Similarly, projective measurement E_n will give outcome a_n probability $p_n = \mathrm{tr}(E_n \rho_A)$, and the post measurement state is

$$\rho_{A,n} = \frac{1}{p_n} \mathrm{tr}_B \left[(E_n \otimes \mathbb{1}) \, |\psi\rangle \, \langle\psi| \, (E_n^\dagger \otimes \mathbb{1}) \right] = E_n \rho_A E_n^\dagger . \tag{2.42}$$

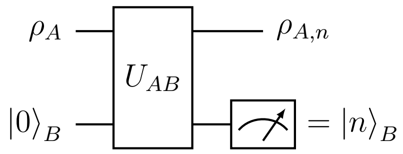
2.5 POVM measurements

Question We have seen that adding system B can lead to richer situations. What measurements can we realize by adding extra systems?
The idea is to

1. Add ”ancillary” B in state $|0\rangle$;

2. Act with unitary U_{AB};

3. Measure B in computational basis.



The un-normalized post-measurement states is

$$\begin{aligned} \rho_{A,n} &= \langle n|_B U \left(\rho_A \otimes |0\rangle \langle 0|_B\right) U^\dagger |n\rangle_B \\ &= M_n \rho_A M_n^\dagger \text{ with } M_n \equiv \langle n|_B U|0\rangle_B \,, \end{aligned} \tag{2.43}$$

and $p_n = \mathrm{tr}\rho_{A,n} = \mathrm{tr}(M_n^\dagger M_n \rho_A)$. Since $\sum_n M_n^\dagger M_n = \mathbb{1}$, this ensures $\sum_n p_n = \mathrm{tr}(\rho_A) = 1$.

Definition 3. *A set $\{F_n = M_n^\dagger M_n\}$ with $0 \le F_n \le 1$ and $\sum_u F_n = \mathbb{1}$ is called **positive operator-valued measure (POVM)**, and the corresponding measurement with outcome probabilities $p_n = \mathrm{tr}(M_n^\dagger M_n \rho) = \mathrm{tr}(F_n \rho)$ is a POVM measurement.*

Can any $\{M_n\}$ with $\sum_n M_n^\dagger M_n = \mathbb{1}$ be realized by extensions to unitaries and projective measurement? Yes! The extended unitary is

$$U = \begin{array}{ccccc} & |0\rangle_B & |1\rangle_B & \cdots & |d-1\rangle_B \\ \langle 0|_B & M_0 & M_1^\dagger & \cdots & M_{d-1}^\dagger \\ \langle 1|_B & M_1 & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \langle d-1|_B & M_{d-1} & \cdots & \cdots & \cdots \end{array} \tag{2.44}$$

Question Is POVM the most general measurement? Consider the most general linear model: set $\{F_n\}$ such that $p_n = \mathrm{tr}(F_n \rho)$. Assume $F_n = F_n^\dagger$. If not, write $F_n = \frac{1}{2}(F_n + F_n^\dagger) + \frac{1}{2}(F_n - F_n^\dagger)$ into the sum of hermitian and anti-hermitian part. Then $-i(F_n - F_n^\dagger)$ is hermitian, and hence

$$\mathrm{tr}\left[(F_n - F_n^\dagger)\rho\right] = i \cdot \mathrm{tr}\left[-i(F_n - F_n^\dagger)\rho\right] \in i\mathbb{R} \tag{2.45}$$

Since $p_n \ge 0$ and $\mathrm{tr}\left[(F_n + F_n^\dagger)\rho\right] \in \mathbb{R}$, $\mathrm{tr}\left[(F_n - F_n^\dagger)\rho\right] = 0$. Therefore we can choose F_n to be hermitian without loss of generality.
The conditions for $\{F_n\}$ are

(i) $1 = \sum_n p_n = \mathrm{tr}\left[(\sum_n F_n)\rho\right],\ \forall \rho \Rightarrow \sum_n F_n = \mathbb{1}$.

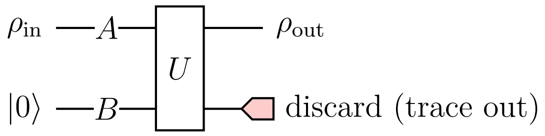
(ii) $0 \le p_n = \mathrm{tr}(F_n \rho) \Rightarrow F_n \ge 0$ (otherwise there exists $F_n |\phi\rangle = \lambda |\phi\rangle, \lambda < 0$ and we can choose $\rho = |\phi\rangle \langle \phi|$ so that $p_n < 0$)

$\Rightarrow F_n = M_n^\dagger M_n$ (e.g., write in eigenbasis $F_n = \sum_i \lambda |\phi_i\rangle \langle \phi_i|$, choose $M_n^\dagger = \left(\sqrt{\lambda_1}\,|\phi_1\rangle \quad \cdots \quad \sqrt{\lambda_{d-1}}\,|\phi_{d-1}\rangle\right)$).

Hence POVM is the most general measurement!

2.6 General evolution – superoperators

Question What is the *most general physical* map on density matrices (”superoperator”)?
Similarly, the idea is to add an ancillary



and the evolution is

$$\begin{aligned}\rho \mapsto \mathcal{E}(\rho) &= \operatorname{tr}_B \left[U(\rho \otimes |0\rangle \langle 0|)U^\dagger \right] \\ &= \sum_n \langle n|_B \, U \, |0\rangle_B \, \rho \, \langle 0|_B \, U^\dagger \, |n\rangle_B \\ &= \sum_n M_n \rho M_n^\dagger.\end{aligned} \tag{2.46}$$

Note that when we take the trace in different basis, choice of M_n is not unique. As before, $\sum_n M_n^\dagger M_n = \mathbb{1}_A$.

Definition 4. *We call*

$$\mathcal{E}(\rho) = \sum_n M_n \rho M_n^\dagger; \; \sum_n M_n^\dagger M_n = \mathbb{1} \tag{2.47}$$

the ***Kraus form*** *or* ***Kraus representation*** *of* $\mathcal{E}$*.*

Remark

Any such $\mathcal{E}$ can be realized by ancillaries and a unitary operator (or POVM). In fact, $\mathcal{E}$ can be seen as POVM where we ignore the result (measurement by environment).

Question Is this the most general physical evolution?
Conditions for a physical evolution $\mathcal{E}$ are

(i) Hermitian preserving: $\rho = \rho^\dagger \Rightarrow \mathcal{E}(\rho) = \mathcal{E}(\rho)^\dagger$.

(ii) Positivity preserving: $\rho \geq 0 \Rightarrow \mathcal{E}(\rho) \geq 0$.

(iii) Trace preserving: $\operatorname{tr}(\rho) = 1 \Rightarrow \operatorname{tr}(\mathcal{E}(\rho)) = 1$.

(iv) Linearity: $\mathcal{E}(\rho + \lambda\sigma) = \mathcal{E}(\rho) + \lambda\mathcal{E}(\sigma)$. Without linearity, ensemble interpretation breaks down.

(v) *Complete positivity*: $\rho_{AB} \geq 0 \Rightarrow (\mathcal{E}(A) \otimes \mathbb{1}_B)\rho_{AB} \geq 0$. $\mathcal{E}$ should still be a physical map when it acts on a part of a larger system.

Definition 5. *We call $\mathcal{E}$ satisfying (i)-(v) a **completely positive trace preserving (CPTP) map**, or a **quantum channel**.*

Are there maps which are *positive trace preserving* ((i)-(iv) but not CP? Yes. An example is the ”transpose channel”:

$$\begin{aligned} \mathcal{E}(\rho) &= \rho^{\tau} \\ (\mathcal{E} \otimes \mathbb{1})(\rho_{AB}) &= \rho_{AB}^{\tau_A} \text{ (”partial transpose”)} \end{aligned} \tag{2.48}$$

e.g., $|\Omega\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$,

$$\begin{aligned} (\mathcal{E} \otimes \mathbb{1})(|\Omega\rangle\langle\Omega|) &= (|\Omega\rangle\langle\Omega|)^{\tau_A} \\ &= \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|)^{\tau_A} \\ &= \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \ngeq 0 \end{aligned} \tag{2.49}$$

Remark

Positive but not CP maps can serve as *entanglement witnesses*: $(\mathcal{E} \otimes \mathbb{1})\rho \geq 0$ for all unentangled states, so $(\mathcal{E} \otimes \mathbb{1})\sigma \ngeq 0 \Rightarrow \sigma$ is entangled.

Clearly, $\mathcal{E}$ is in Kraus form $\Rightarrow$ $\mathcal{E}$ is a CPTP. Are also all CPTP maps of Kraus form?

Theorem 2. ***Choi–Jamiołkowski isomorphism**: Let $\mathcal{C} = \{\mathcal{E} | \mathcal{E}$ is CPTP$\}$ the space of CPTP maps on the density operators of $\mathbb{C}^d$, and $\Gamma = \{\sigma_{AB} \in \mathcal{B}(\mathbb{C}^d \otimes \mathbb{C}^d) | \sigma_{AB} \geq 0, \mathrm{tr}_A(\sigma_{AB}) = \frac{1}{d}\mathbb{1}\}$ ($\mathcal{B}$ means linear operators). Then, the map*

$$\hat{X} : \mathcal{E} \mapsto \sigma_{AB} = (\mathcal{E} \otimes \mathbb{1}_B)(|\Omega\rangle\langle\Omega|),\ |\Omega\rangle = \frac{1}{\sqrt{d}}\sum_{i=0}^{d-1} |i, i\rangle \tag{2.50}$$

defines the Choi–Jamiołkowski isomorphism between $\mathcal{C}$ and Γ, with σ_{AB} the Choi state of $\mathcal{E}$. The reverse map is given by

$$\hat{Y} : \sigma_{AB} \mapsto F,\ F(\rho) = d \cdot \mathrm{tr}_B\left[\sigma_{AB}(\mathbb{1} \otimes \rho^{\tau})\right] \tag{2.51}$$

Proof. we need to show

(i) $\hat{Y}\hat{X} = \mathbb{1}_{\mathcal{C}}$.

$$\begin{aligned}\hat{Y}(\hat{X}(\mathcal{E}))(\rho) &= d \cdot \mathrm{tr}_B \left[\hat{X}(\mathcal{E})(\mathbb{1} \otimes \rho^\tau)\right] \\ &= \sum_{ij} \mathrm{tr}_B \left[(\mathcal{E} \otimes \mathbb{1}_B)(|i\rangle \langle j| \otimes |i\rangle \langle j|)(\mathbb{1}_A \otimes \rho^\tau)\right] \\ &= \sum_{ij} \mathcal{E}(|i\rangle \langle j|)\mathrm{tr}_B(|i\rangle \langle j| \rho^\tau) \\ &= \mathcal{E}(\sum_{ij} \rho_{ij} |i\rangle \langle j|) = \mathcal{E}(\rho).\end{aligned} \tag{2.52}$$

(ii) $\hat{X}\hat{Y} = \mathbb{1}_\Gamma$.

$$\begin{aligned}\hat{X}(\hat{Y}(\sigma_{AB})) &= (\hat{Y}(\sigma_{AB}) \otimes \mathbb{1})(|\Omega\rangle \langle\Omega|) \\ &= \frac{1}{d}\sum_{ij}(\hat{Y}(\sigma_{AB}))(|i\rangle \langle j|) \otimes |i\rangle \langle j| \\ &= \sum_{ij} \mathrm{tr}_B \left[\sigma_{AB}(\mathbb{1} \otimes (|i\rangle \langle j|)^\tau)\right] \otimes |i\rangle \langle j| \\ &= \sum_{ij} \langle i|_B \sigma_{AB} |j\rangle_B \otimes |i\rangle_B \langle j|_B = \sigma_{AB}.\end{aligned} \tag{2.53}$$

(iii) $\mathrm{Im}\hat{X} = \{\hat{X}(\mathcal{E}) | \mathcal{E} \in \mathcal{C}\} \subset \Gamma$.

By construction, $\sigma_{AB} = \hat{X}(\mathcal{E}) \geq 0$. Since $\mathcal{E}$ preserves the trace,

$$\mathrm{tr}_A(\sigma_{AB}) = \frac{1}{d}\sum_{ij} \mathrm{tr}_A \left[\mathcal{E}(|i\rangle \langle j|) \otimes |i\rangle \langle j|\right] = \frac{1}{d}\mathbb{1}_B. \tag{2.54}$$

(iv) $\mathrm{Im}\hat{Y} \subset \mathcal{C}$.

Let $\sigma_{AB} \in \Gamma$. Write $\sigma_{AB} = \sum_k |\tilde{\psi_k}\rangle \langle \tilde{\psi_k}|$ ($|\tilde{\psi_k}\rangle$ are unnormalized). We have

$$|\tilde{\psi_k}\rangle = \sum_{ij} m_k^{ij} |j\rangle |i\rangle = \frac{1}{\sqrt{d}}\sum_i (M_k \otimes \mathbb{1}) |i\rangle |i\rangle = (M_k \otimes \mathbb{1}) |\Omega\rangle , \tag{2.55}$$

and hence

$$\begin{aligned}
\sigma_{AB} &= \sum_k (M_k \otimes \mathbb{1}) \left|\Omega\right\rangle \left\langle\Omega\right| (M_k \otimes \mathbb{1})^\dagger, \\
(\hat{Y}(\sigma_{AB}))(\rho) &= d \cdot \mathrm{tr}_B \left[\sigma_{AB}(\mathbb{1} \otimes \rho^\tau)\right] \\
&= d \cdot \mathrm{tr}_B \left[(\sum_k (M_k \otimes \mathbb{1}) \left|\Omega\right\rangle \left\langle\Omega\right| (M_k \otimes \mathbb{1})^\dagger)(\mathbb{1} \otimes \rho^\tau) \right] \\
&= d \sum_k M_k \mathrm{tr}_B \left[\left|\Omega\right\rangle \left\langle\Omega\right| (\mathbb{1} \otimes \rho^\tau)\right] M_k^\dagger \\
&= \sum_k M_k \rho M_k^\dagger.
\end{aligned} \tag{2.56}$$

Since $\mathrm{tr}_A \sigma_{AB} = \frac{1}{d}\mathbb{1}$,

$$\begin{aligned}
\mathbb{1} = d \cdot \mathrm{tr}_A \sigma_{AB} &= d \sum_k \mathrm{tr}_A \left[(M_k^\dagger M_k \otimes \mathbb{1}) \left|\Omega\right\rangle \left\langle\Omega\right| \right] \\
&= \sum_{ijk} \mathrm{tr}(M_k^\dagger M_k \left|i\right\rangle \left\langle j\right|) \otimes \left|i\right\rangle \left\langle j\right| \\
&\Rightarrow \left\langle j\right| M_k^\dagger M_k \left|i\right\rangle = \mathrm{tr}(M_k^\dagger M_k \left|i\right\rangle \left\langle j\right|) = \delta_{ij} \\
&\Rightarrow \sum_k M_k^\dagger M_k = \mathbb{1} \\
&\Rightarrow \hat{Y}(\sigma_{AB}) \in \mathcal{C}, \ \forall \sigma_{AB} \in \Gamma.
\end{aligned} \tag{2.57}$$

□

Note

the isomorphism still holds if we drop trace preserving and $\mathrm{tr}_A \sigma_{AB} = \frac{1}{d}\mathbb{1}$, respectively.

Corollary 1. *(from (iv)) All CPTP maps are of Kraus form (and can thus be realized with ancillary + unitary + tracing).*

2.7 Axioms ("mixed version")

1. States are linear operators ρ with

$$\rho \geq 0, \ \mathrm{tr}\rho = 1. \tag{2.58}$$

2. Evolutions are CPTP maps $\mathcal{E}(\rho) = \sum_n M_n \rho M_n^\dagger$ with

$$\sum_n M_n^\dagger M_n = \mathbb{1}. \tag{2.59}$$

3. Measurements act as

$$\rho \mapsto \rho_n = M_n \rho M_n^\dagger / \mathrm{tr}(M_n \rho M_n^\dagger), \tag{2.60}$$

   with probability $p_n = \mathrm{tr}(M_n \dagger M_n \rho)$ and $\sum_n M_n^\dagger M_n = \mathbb{1}$.

Chapter 3

Entanglement

3.1 Introduction

Consider a *bipartite pure state* $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. $|\psi\rangle$ is a ***product state*** if it can be written as

$$|\psi\rangle = |\phi\rangle_A \otimes |\phi\rangle_B \, ; \tag{3.1}$$

otherwise we say $|\psi\rangle$ is ***entangled***.

Characterization

- *Schmidt coefficients*: $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots)$.
  product states: $\boldsymbol{\lambda} = (1, 0, \cdots, 0)$;
  entangled states: $\boldsymbol{\lambda} = (\lambda_1, \lambda_2 \neq 0, \cdots)$.

- *Reduced density matrix* ρ_a, ρ_B.
  product states: $\rho_A = |\phi_A\rangle \langle\phi_A| \, , \rho_B = |\phi_B\rangle \langle\phi_B|$ are pure states, and conversely $|\psi\rangle$ can be determined by $|\psi\rangle = |\phi\rangle_A \otimes |\phi\rangle_B$;
  entangled states: ρ_A is mixed $\Rightarrow$ $\mathrm{tr}\rho_A^2 < 1$.

Entangled states are "different"

- One cannot describe two parts independently.

- Measurement outcomes are correlated.

- We will see that they are suitable for non-trivial tasks.

Aims of studying entanglement

– How *non-classical* are entangled states?

– What can we do with them? (”resource”)

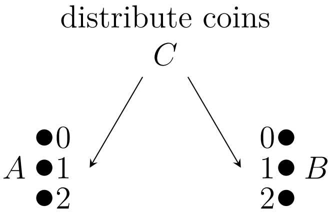
– How can we *quantify* amount of entanglement?

– How can we transform / manipulate entanglement?

– What about *entanglement of mixed states*?

3.2 Bell inequalities

How non-classical are entangled states? Consider the following game of A and B with coins:



1. A and B each get 3 coins in boxes (labeled by 0, 1, 2), prepared in the same way (deterministic or random) by C.

2. A and B can look at one coin each ($i, j = 0, 1, 2$). The result is +1 (head) or -1 (tail). We denote result by $a_i, b_j = \pm 1$.

3. A and B observe: if they look at the same coin, they always get the same outcome $a_i = b_i$.

4. Can A infers the value of 2 of her coins? The idea is that B looks at $j = i' \neq i$, and then they now a_i and $a_{i'}$. It works classically!

5. It implies that A and B can use this to estimate the probability $p(a_i = a_{i'})\ \forall i, i'$. Clearly, we must have

$$p(a_0 = a_1) + p(a_1 = a_2) + p(a_2 = a_0) \geq 1, \qquad (3.2)$$

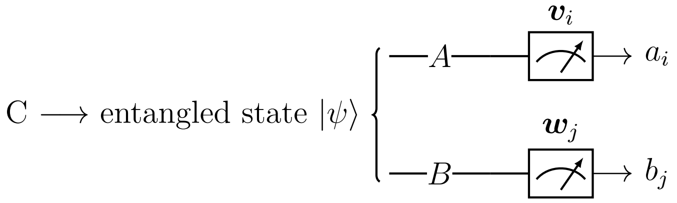
since in each instance of the game, at least 2 coins must be equal. Therefore

$$p(a_0 = b_1) + p(a_1 = b_2) + p(a_2 = b_0) \geq 1 \qquad (3.3)$$

is satisfied classically!

Eq. (3.3) is called a ***Bell inequality***. But in a quantum mechanical version of this game, the Bell inequality can be violated!

Quantum Mechanical version of the game



Choose $|\psi\rangle = |\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$. A and B will measure spin along some axes $\boldsymbol{v_i}$ and $\boldsymbol{w_j}$, i.e., the operators $\boldsymbol{v_i} \cdot \boldsymbol{\sigma}$ and $\boldsymbol{w_i} \cdot \boldsymbol{\sigma}$, with $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$. Since we have $(\boldsymbol{\sigma}^A + \boldsymbol{\sigma}^B)\,|\psi^-\rangle = 0$,

$$\begin{aligned}\langle\psi^-|(\boldsymbol{\sigma}^A \cdot \boldsymbol{v})(\boldsymbol{\sigma}^B \cdot \boldsymbol{w})|\psi^-\rangle &= -\,\langle\psi^-|(\boldsymbol{\sigma}^A \cdot \boldsymbol{v})(\boldsymbol{\sigma}^A \cdot \boldsymbol{w})|\psi^-\rangle \\ &= -\sum_{kl} v_k w_l \mathrm{tr}(\sigma_k^A \sigma_l^B \underbrace{\mathrm{tr}_B(|\psi^-\rangle\,\langle\psi^-|)}_{=\frac{1}{2}\mathbb{1}}) \\ &= -\sum_k v_k w_k = -\boldsymbol{v} \cdot \boldsymbol{w} = -\cos\theta\end{aligned} \tag{3.4}$$

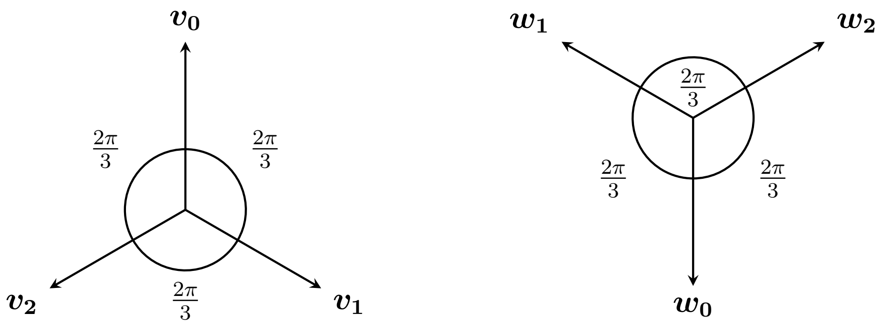
The measurement of A along $\boldsymbol{v}$ is the projections $E_{\pm 1}(\boldsymbol{v}) = \frac{1}{2}(\mathbb{1} \pm \boldsymbol{v} \cdot \boldsymbol{\sigma})$. Hence

$$\begin{aligned}p(\pm 1, \pm 1) &= \langle\psi^-|E_{\pm 1}(\boldsymbol{v})E_{\pm 1}(\boldsymbol{w})|\psi^-\rangle \\ &= \frac{1}{4}\,\langle\psi^-|\mathbb{1} \pm \boldsymbol{v} \cdot \boldsymbol{\sigma}^A \pm \boldsymbol{w} \cdot \boldsymbol{\sigma}^B + (\boldsymbol{\sigma}^A \cdot \boldsymbol{v})(\boldsymbol{\sigma}^B \cdot \boldsymbol{w})|\psi^-\rangle \\ &= \frac{1}{4}(1 - \cos\theta)\end{aligned} \tag{3.5}$$

and

$$p(\pm 1, \mp 1) = \frac{1}{4}(1 + \cos\theta). \tag{3.6}$$

Therefore $p_{\text{equal}} = \frac{1}{2}(1 - \cos\theta)$, $p_{\text{different}} = \frac{1}{2}(1 + \cos\theta)$.
Now let A measure along $\boldsymbol{v_0}$, $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ in the XZ-plane, separating $2\pi/3$ each, and B along $\boldsymbol{w_i} = -\boldsymbol{v_i}$.



- $i = j$: $p_{\text{equal}} = \frac{1}{2}(1 - \cos\pi) = 1$ $\surd$
- $i \neq j$: $p_{\text{equal}} = \frac{1}{2}(1 - \cos(\pm\frac{2}{3}\pi)) = \frac{1}{4}$.

$$\Rightarrow p(a_0 = b_1) + p(a_1 = b_2) + p(a_2 = b_0) = \frac{3}{4} < 1$$

$\Rightarrow$ Bell inequality violated!

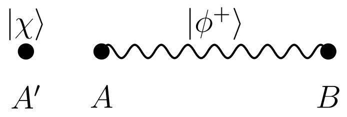
Assumptions

1. *Realism*: outcomes of measurements are "elements of reality" (i.e., have pre-determined values) even without measurement.
2. *Locality*: A and B's boxes cannot communicate once distributed.

$\Rightarrow$ QM predictions are *incompatible* with *local and realistic* description.
$\Rightarrow$ We need to give up either locality or realism (or both).

3.3 Applications of entanglement: teleportation, dense coding

3.3.1 Teleportation



Setup A and B share an entangled state $|\phi^+\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. A has unknown state $|\chi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$. A and B cannot send a quantum state, but can communicate classically "for free". Can A get $|\chi\rangle$ to B? Note that measurement of $|\chi\rangle$ would *break* the state and destroy the information $\Rightarrow$ we need teleportation!

Protocol

1. Combined measurement of A′ and A in *Bell basis*

$$\begin{aligned}
|\phi^+\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\
|\phi^-\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) = (Z \otimes \mathbb{1})\,|\phi^+\rangle = (\mathbb{1} \otimes Z)\,|\phi^+\rangle \\
|\psi^+\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) = (X \otimes \mathbb{1})\,|\phi^+\rangle = (\mathbb{1} \otimes X)\,|\phi^+\rangle \\
|\psi^-\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) = (ZX \otimes \mathbb{1})\,|\phi^+\rangle = (\mathbb{1} \otimes XZ)\,|\phi^+\rangle
\end{aligned} \tag{3.7}$$

Or $|\phi_{\alpha\beta}\rangle = (Z^\alpha X^\beta \otimes \mathbb{1})\,|\phi^+\rangle = (\mathbb{1} \otimes X^\beta Z^\alpha)\,|\phi^+\rangle$ $(\alpha, \beta = 0, 1)$.

The outcome probabilities for $|\phi_{\alpha\beta}\rangle$:

$$\begin{aligned}
&\rho_A = \mathrm{tr}_B(|\phi^+\rangle\langle\phi^+|_{AB}) = \frac{1}{2}\mathbb{1}, \\
&\langle\phi_{\alpha\beta}|\,|\chi\rangle\langle\chi|_{A'} \otimes \frac{1}{2}\mathbb{1}_A|\phi_{\alpha\beta}\rangle = \frac{1}{2}\mathrm{tr}\left[(|\chi\rangle\langle\chi|_{A'} \otimes \mathbb{1}_A)\,|\phi_{\alpha\beta}\rangle\langle\phi_{\alpha\beta}|\right] \\
=&\frac{1}{2}\mathrm{tr}_{A'}\left[(|\chi\rangle\langle\chi|_{A'} \otimes \mathbb{1}_A)\,\underbrace{\mathrm{tr}_A(|\phi_{\alpha\beta}\rangle\langle\phi_{\alpha\beta}|)}_{=\frac{1}{2}\mathbb{1}}\right] = \frac{1}{4}.
\end{aligned} \tag{3.8}$$

The probabilities for all 4 outcomes are equal, independent of $|\chi\rangle$. It is a good news since no information about $|\chi\rangle$ acquired means no disturbance. Then what is the state of B after measurement?

(i) Outcome for $|\phi^+\rangle = |\phi_{00}\rangle$ (unnormalized):

$$\begin{aligned}
&\langle\phi^+|_{AA'}\,(|\chi\rangle_{A'} \otimes |\phi^\dagger\rangle_{AB}) \\
=&\frac{1}{2}(\langle 00| + \langle 11|)_{AA'}(\alpha\,|0\rangle + \beta\,|1\rangle)_{A'}(|00\rangle + |11\rangle)_{AB} \\
=&\frac{1}{2}(\alpha\,|0\rangle_B + \beta\,|1\rangle_B).
\end{aligned} \tag{3.9}$$

State $|\chi\rangle$ appears in B!

(ii) General outcome:

$$\begin{aligned}
&\langle\phi_{\alpha\beta}|_{AA'}\,|\phi^\dagger\rangle_{AB} = \langle\phi^+|_{A'A}\,\mathbb{1}_{A'} \otimes Z_A^\alpha X_A^\beta\,|\phi^+\rangle_{AB} \\
=&\langle\phi^+|_{A'A}\,Z_A^\alpha X_A^\beta \otimes \mathbb{1}_B\,|\phi^+\rangle_{AB} \\
=&\langle\phi^+|_{A'A}\,\mathbb{1}_A \otimes X_B^\beta Z_B^\alpha\,|\phi^+\rangle_{AB}.
\end{aligned} \tag{3.10}$$

Hence

$$\begin{aligned}
&\langle\phi_{\alpha\beta}|_{AA'}\left(|\chi\rangle_{A'}\otimes|\phi^{\dagger}\rangle_{AB}\right)\\
=&X_B^{\beta}Z_B^{\alpha}\langle\phi^{+}|_{AA'}\left(|\chi\rangle_{A'}\otimes|\phi^{\dagger}\rangle_{AB}\right)\\
=&X_B^{\beta}Z_B^{\alpha}|\chi\rangle_{B}\,.
\end{aligned}\tag{3.11}$$

The average state of B is $\frac{1}{4}\sum_{\alpha\beta}X^{\beta}Z^{\alpha}|\chi\rangle\langle\chi|Z^{\alpha}X^{\beta}=\frac{1}{2}\mathbb{1}$. There is no information at Bob's side!

2. A communicates measurement outcome (α,β) to Bob.

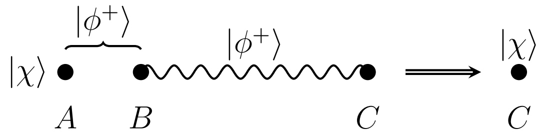
3. Bob applies $(Z^{\alpha}X^{\beta})$ on B to recover $|\chi\rangle$.

Notes

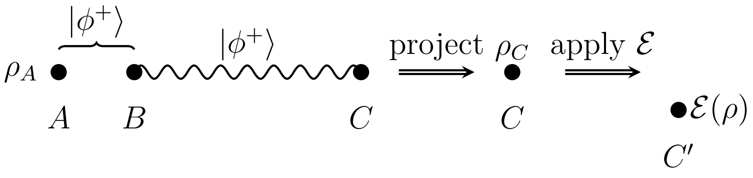
- No faster-than-light communication!
- Communicating 1 qubit requires 1 "e-bit" (maximal entangled state of 1+1 qubits) and 2 bits of classical communication.

Relations between teleportation and Choi–Jamiołkowski isomorphism

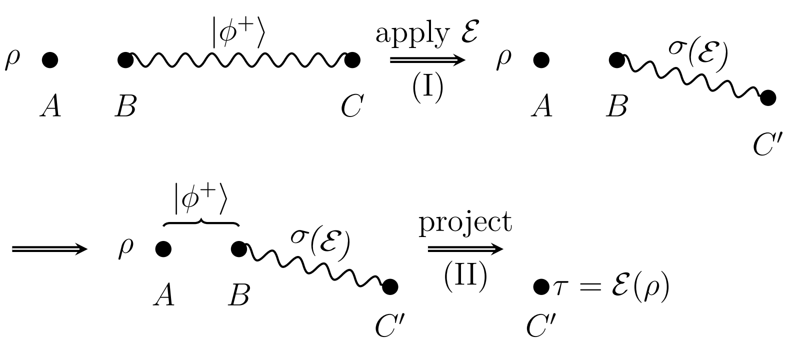
1. Consider "post selected teleportation" (projection onto $|\phi^{+}\rangle_{AB}$):



2. Protocol for applying $\rho\mapsto\mathcal{E}(\rho)$:



3. Exchange the order of projection and applying $\mathcal{E}$:



This is the Choi–Jamiołkowski isomorphism:

(I) is the $\mathcal{E} \mapsto \sigma$ map, and

(II) is the $\sigma \mapsto \mathcal{E}$ map.

3.3.2 Dense coding

See homework (A.3).

3.4 Entanglement conversion and quantification

3.4.1 Introduction and setup

Entanglement is what cannot be changed by ***local operations and classical communication (LOCC)***.

Question When can we convert entangled states into each other with LOCC? The problem is relevant about that

- different protocols might require different ("cheaper" / "more expansive") entangled states;

- it can be used to *quantify entanglement* in terms of some reference state: How many "e-bits" $|\phi^+\rangle = \frac{1}{\sqrt{}}(|00\rangle + |11\rangle)$ are contained in a state?

We've known that the same Schmidt coefficients $\iff$ related by local unitary $\iff$ same entanglement. What if Schmidt coefficients are different?

Example $|\chi\rangle = \sqrt{\frac{2}{3}}\,|00\rangle + \sqrt{\frac{1}{3}}\,|11\rangle$; $|\phi^+\rangle = \sqrt{\frac{1}{2}}\,|00\rangle + \sqrt{\frac{1}{2}}\,|11\rangle$.

1. Can we convert $|\phi^+\rangle \to |\chi\rangle$?
   A does POVM $\{M_0, M_1\}$. $M_0 = \begin{pmatrix} \sqrt{2/3} & 0 \\ 0 & \sqrt{1/3} \end{pmatrix}$, $M_1 = \begin{pmatrix} \sqrt{1/3} & 0 \\ 0 & \sqrt{2/3} \end{pmatrix}$.
   The post measurement states are $|\tilde{\psi}_k\rangle = M_k\,|\phi^+\rangle$:

   $$\begin{aligned} |\psi_0\rangle = |\chi\rangle\,, \quad |\psi_1\rangle = X \otimes X\,|\chi\rangle\,; \\ p_0 = p_1 = \frac{1}{2}. \end{aligned} \tag{3.12}$$

   Protocol: A does POVM and reads result to B. If it is 1, both apply X.
   Success probability: $p = p_0 + p_1 = 1$.
   Best possible: we cannot get > 1 copies, as POVM cannot increase Schmidt rank!

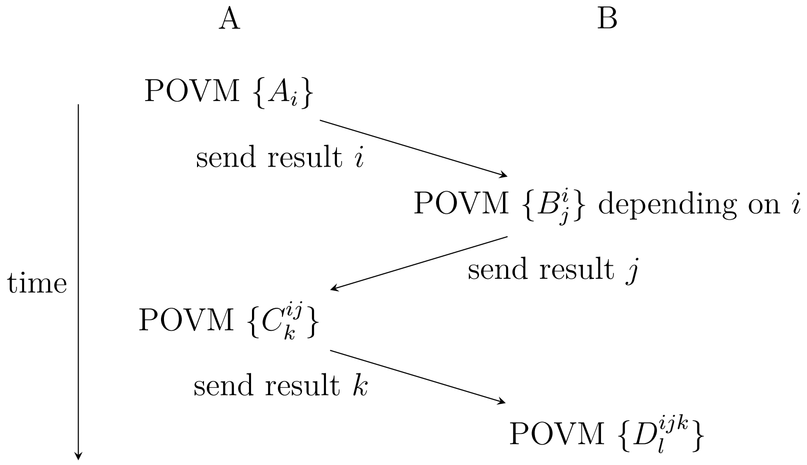
2. Can we do the converse: $|\chi\rangle \to |\phi^+\rangle$?
   (Will see it's the best possible): A does POVM $\{M_0, M_1\}$. $M_0 = \begin{pmatrix} \sqrt{1/2} & 0 \\ 0 & 1 \end{pmatrix}$, $M_1 = \begin{pmatrix} \sqrt{1/2} & 0 \\ 0 & 0 \end{pmatrix}$.

   $$\begin{aligned} |\psi_0\rangle = |\chi\rangle\,, \quad |\psi_1\rangle = |00\rangle \ \text{(no entanglement left!)}; \\ p_0 = \frac{2}{3}, \quad p_1 = \frac{1}{3}. \end{aligned} \tag{3.13}$$

   The success probability is 2/3. *Conversion is not reversible!*

General LOCC protocol

$$\rho \to \sum_{ijkl\cdots} (\cdots C_k^{ij} A_i) \otimes (\cdots D_l^{ijk} B_j^i)\rho \left[(\cdots C_k^{ij} A_i) \otimes (\cdots D_l^{ijk} B_j^i)\right]^\dagger \tag{3.14}$$



$\cdots$ (run for any number of rounds)

It has very complicated structure, but for *pure* stats, protocol can be replaced by *one-round* protocol with *one-way* communication: POVM $\{M_k\} \xrightarrow{k}$ unitary U_k , i.e., $|\psi\rangle \to |\tilde{\psi}_k\rangle = M_k \otimes U_k \, |\psi\rangle$.

General protocol for entanglement conversion

$$|\psi\rangle \to |\tilde{\psi}_k\rangle = M_k \otimes U_k \, |\psi\rangle \, ; \; p_k = | \, |\tilde{\psi}_k\rangle \, |^2 \tag{3.15}$$

For entanglement: given $|\psi\rangle$, $|\psi_k\rangle = |\tilde{\psi}_k\rangle \, / | \, |\tilde{\psi}_k\rangle \, |$ is fully characterized by *Schmidt coefficients*; U_k is irrelevant. Hence we can study instead the possible conversions

$$\rho_A \to \{p_k, \rho_{A,k}\}. \tag{3.16}$$

Under which conditions does there exist POVM $\{M_k\}$ such that $p_k \rho_{A,k} = M_k \rho_A M_k^\dagger$?

3.4.2 Single-copy protocols: majorization

Definition 6. *For $\lambda \in \mathbb{R}^d_{\geq 0}$, let $\lambda^\downarrow = (\lambda^\downarrow_1, \cdots, \lambda^\downarrow_d)$, $\lambda^\downarrow_1 \geq \lambda^\downarrow_2 \geq \cdots \geq 0$ denote the ordered version of λ. We say that λ **is majorized by** μ (or μ **majorizes** λ, denoted by $\lambda \prec \mu$) if and only if*

$$\sum_{i=1}^{k} \lambda^\downarrow_i \leq \sum_{i=1}^{k} \mu^\downarrow_i, \; \forall k = 1, \cdots, d, \; \textit{with equality for } k = d. \tag{3.17}$$

Theorem 3. *The following are equivalent:*

(i) $\lambda \prec \mu$*;*

(ii) there exist permutations P_i *and probabilities* q_i *such that* $\lambda = \sum_i q_i P_i \mu$*;*

(iii) there exists a doubly stochastic Q *(i.e.,* $Q_{ij} \geq 0$*,* $\sum_i Q_{ij} = \sum_j Q_{ij} = 1$*) such that* $\lambda = Q\mu$*.*

((ii) $\Longleftrightarrow$ (iii) follows from Birkhoff's theorem: every $Q \sum_i q_i P_i$.)
Intuition: $\lambda \prec \mu \iff \lambda$ can be obtained by random permutations of μ, which means it is "more random" (as a probability distribution). The largest (not random at all): $(1, 0, \cdots, 0)$; the smallest (most random): $(\frac{1}{d}, \cdots, \frac{1}{d})$.

Remarks

- Majorization defines a *partial order* on probability distributions.
- $\lambda \prec \mu$ means λ is more disordered than μ and in particular, has more entropy. It can be made rigorous by "Schur concavity / convexity": for a concave / convex $f(x)$, $F(\lambda) = \sum f(\lambda_i)$ fulfills $\lambda \prec \mu \iff F(\lambda) \gtreqless F(\mu)$.

Generalization to operators For a hermitian matrix A, $\lambda^{\downarrow}(A) =$ ordered eigenvalues of A.

Lemma 1. $\lambda^{\downarrow}(A+B) \prec \lambda^{\downarrow}(A) + \lambda^{\downarrow}(B)$.

Intuition: eigenvalues of $A+B$ are the most ordered if in the same basis.

Proof. Use Ky-Fan's maximum principle:

$$\sum_{j=1}^{k} \lambda_j^{\downarrow}(A) = \max_P \operatorname{tr}(AP), \ P \in \{\text{projections of rank } k\}. \tag{3.18}$$

Then

$$\begin{aligned} \sum_{j=1}^{k} \lambda_j^{\downarrow}(A+B) &= \max_P \operatorname{tr}((A+B)P) \leq \max_P \operatorname{tr}(AP) + \max_P \operatorname{tr}(BP) \\ &= \sum_{j=1}^{k} \lambda_j^{\downarrow}(A) + \lambda_j^{\downarrow}(B). \end{aligned} \tag{3.19}$$

□

Theorem 4. *(single-copy entanglement conversion) We can convert* $|\psi\rangle \to \{p_k, |\psi_k\rangle\}_{k=1}^{K}$ *by LOCC if and only if* $\lambda^{\downarrow}(\rho) \prec \sum_{k=1}^{K} p_k \lambda^{\downarrow}(\rho_k)$*, where* $\rho = \mathrm{tr}_A |\psi\rangle\langle\psi|$*,* $\rho_k = \mathrm{tr}_A |\psi_k\rangle\langle\psi_k|$*.*

Proof. "⇒": Protocol: A does POVM $\{M_k\}$. Without loss of generality, Bob's unitary $U_k = \mathbb{1}$ (only Schmidt coefficients matter!). Then

$$\begin{aligned}
&\sum_{k=1}^{K} p_k \lambda^{\downarrow}(\rho_k) = \sum_{k=1}^{K} \lambda^{\downarrow}(p_k\rho_k) \\
=&\sum_{k=1}^{K} \lambda^{\downarrow}\left(\mathrm{tr}_A\left[(M_k \otimes \mathbb{1})|\psi\rangle\langle\psi|(M_k^{\dagger} \otimes \mathbb{1})\right]\right) \\
\prec&\lambda^{\downarrow}\left(\mathrm{tr}_A\left[\sum_k M_k^{\dagger} M_k \otimes \mathbb{1}|\psi\rangle\langle\psi|\right]\right) = \lambda^{\downarrow}(\rho).
\end{aligned} \tag{3.20}$$

"⇐": $\lambda^{\downarrow}(\rho) \prec \sum_k p_k \lambda^{\downarrow}(\rho_k) \Rightarrow \exists P_j, q_j$ such that $\lambda^{\downarrow}(\rho) = \sum_{kj} p_k q_j P_j \lambda^{\downarrow}(\rho_k)$. Without loss of generality, assume that ρ and ρ_k are diagonal; otherwise we can add Bob's unitaries!
Define E_{kj} via $E_{kj}\sqrt{\rho} = \sqrt{p_k q_j}\sqrt{\rho_k}P_j^{\dagger}$. Then

$$\begin{aligned}
&\sqrt{\rho}(\sum_{kj} E_{kj}^{\dagger} E_{kj})\sqrt{\rho} = \sum_{kj} p_k q_j P_j \rho_k P_j^{\dagger} = \rho \\
&\Rightarrow \sum_{kj} E_{kj}^{\dagger} E_{kj} = \mathbb{1}.
\end{aligned} \tag{3.21}$$

(given that ρ is invertible; otherwise any E_{kj} on $\mathrm{Ker}(\rho)$will do.)

And $E_{kj}\rho E_{kj}^{\dagger} = p_k q_j \rho_k \Rightarrow \sum_j E_{kj}\rho E_{kj}^{\dagger} = p_k\rho_k \Rightarrow$ POVM for $\rho \mapsto \{p_k, \rho_k\}$! Note that we have several POVM operators labeled by j for the same outcome.

□

Review of the example We want to find the optimal rate for $(\frac{1}{2}, \frac{1}{2}) \leftrightarrow (\frac{2}{3}, \frac{1}{3})$:

$$\begin{aligned}
&(\frac{1}{2}, \frac{1}{2}) \prec (\frac{2}{3}, \frac{1}{3}) \qquad \checkmark \\
&(\frac{2}{3}, \frac{1}{3}) \prec \underbrace{\frac{2}{3}}_{\text{max. value!}} (\frac{1}{2}, \frac{1}{2}) + \frac{1}{3}(1,0) = (\frac{2}{3}, \frac{1}{3})
\end{aligned} \tag{3.22}$$

3.4.3 Asymptotic protocols

Single-copy protocol is not reversible, since entanglement will be lost. Can we do better with more copies?
For example,

$$\begin{aligned} |\chi\rangle^{\otimes 2} &\to p_2 |\phi^+\rangle^{\otimes 2} + p_1 |\phi^+\rangle \\ |\chi\rangle^{\otimes 3} &\to p_3 |\phi^+\rangle^{\otimes 3} + p_2 |\phi^+\rangle^{\otimes 2} + p_1 |\phi^+\rangle \\ &\cdots \end{aligned} \tag{3.23}$$

The average yielding of maximum entangled states is

$$\bar{p} = \frac{p_1 + 2p_2 + 3p_3 + \cdots}{k \text{ (number of copies)}} \tag{3.24}$$

Can we increase $\bar{p}$ by using more copies? Yes!

Requirements for asymptotic protocols

- Convert $|\phi^+\rangle^{\otimes M} \leftrightarrow |\chi\rangle^{\otimes N}$ with *rate* $N/M \to R > 0$ for $N, M \to \infty$.
- Success probability $p \to 1$ for $N \to \infty$.
- Conversion can be *imperfect*, as long as error $\to 0$ as $N \to \infty$.

Error measure $\delta = 1 - F$, $F = |\langle\psi|\phi\rangle|^2$: the fidelity.A good measure bounds the distance for any observable!

Form of $|\chi\rangle^{\otimes N}$ When $|\chi\rangle = \sum_x \sqrt{p(x)}\, |x\rangle_A |x\rangle_B$, $x = 1, \cdots, d$,

$$|\chi\rangle^{\otimes N} = \sum_{x_1, \cdots, x_N} \sqrt{p(x_1) \cdots p(x_N)}\, |x_1, \cdots x_N\rangle_A |x_1, \cdots x_N\rangle_B . \tag{3.25}$$

The probability of $x_1, \cdots x_N$ obeys independently identically distribution (i.i.d.), and the law of large numbers applies:

$$P(|\frac{1}{N} \sum_i x_i - E(x_i)| \geq \epsilon) \to 0, \forall \epsilon. \tag{3.26}$$

The most likely output is that x appears $Np(x)$ times, and hence

$$\begin{aligned} p(x_1, \cdots, x_N) = p(x_1) \cdots p(x_N) &\approx p(1)^{Np(1)} \cdots p(d)^{Np(d)} \\ \Rightarrow -\underbrace{\log}_{\text{base 2}} p(x_1, \cdots, x_N) &\approx N \underbrace{(-\sum_x p(x) \log p(x))}_{=H(p)\text{: Shannon entropy}} . \end{aligned} \tag{3.27}$$

Definition 7. *Asymptotically,* $P(|-\frac{1}{N}\log p(x_1,\cdots,x_N)-H(p)|\geq\epsilon)\to 0$. *We call all such* $(x_1,\cdots,x_N)$ ***ϵ-typical sequences***. *There are asymptotically* $\approx 2^{NH(p)}$ *typical sequences.*

Fix $\epsilon>0$. Define

$$|\theta_N\rangle := \sum_{x_1,\cdots,x_N\ \epsilon\text{-typical}} \sqrt{p(x_1)\cdots p(x_N)}\,|x_1,\cdots x_N\rangle_A\,|x_1,\cdots x_N\rangle_B \tag{3.28}$$

and $|\hat{\theta}_N\rangle := |\theta_N\rangle/|\,|\theta_N\rangle\,|$. We have

$$\langle\hat{\theta}_N|\chi^{\otimes N}\rangle = \frac{\sum_{\epsilon\text{-typical}} p(x_1,\cdots,x_N)}{\sqrt{\sum_{\epsilon\text{-typical}} p(x_1,\cdots,x_N)}} \xrightarrow{N\to\infty} 1 \tag{3.29}$$

and the number of terms in $|\hat{\theta}_N\rangle\approx 2^{NH(p)}$ (in fact $\leq 2^{N(H(p)+\epsilon)}$).

Protocol $|\phi^+\rangle^{\otimes M}\to|\chi\rangle^{\otimes N}$

- Use $M=N(H(p)+\epsilon)$ Bell pairs to prepare $|\hat{\theta}_N\rangle$. It is possible since $|\phi^+\rangle^{\otimes M}$ majorizes all distributions.
- $\frac{M}{N}\to H(p)+\epsilon\to H(p)$, and $|\hat{\theta}_N\rangle\to|\chi\rangle^{\otimes N}$ $\Rightarrow$ We can prepare $|\chi\rangle$ asymptotically at a cost $H(p)$ per copy!

Protocol $|\chi\rangle^{\otimes N}\to|\phi^+\rangle^{\otimes M}$

- Use $|\hat{\theta}_N\rangle$ instead of $|\chi\rangle^{\otimes N}$, since the fidelity $\to 1$.
- Schmidt coefficients approach flat distribution with $N(H(p)-\epsilon)$ terms asymptotically $\Rightarrow$ We can extract $M/N\to H(p)$ e-bits per copy of $|\chi\rangle$.

Asymptotically, we can dilute ($|\phi^+\rangle^{\otimes M}\to|\chi\rangle^{\otimes N}$) and distill ($|\chi\rangle^{\otimes N}\to|\phi^+\rangle^{\otimes M}$) at the same rate $H(p)$, with $p=(p_1,\cdots,p_d)$, $\sqrt{p_k}$ being Schmidt coefficients. It can be expressed in terms of the ***von Neumamn entropy***

$$S(\rho)=-\mathrm{tr}(\rho\log\rho),\quad H(p)=S(\mathrm{tr}_A(|\chi\rangle\langle\chi|))=S(\mathrm{tr}_B(|\chi\rangle\langle\chi|)). \tag{3.30}$$

The protocol allows to go reversibly between any to states $|\psi\rangle^{\otimes K}\leftrightarrow|\chi\rangle^{\otimes L}$ as long as $KS(\mathrm{tr}_B(|\psi\rangle\langle\psi|))=LS(\mathrm{tr}_B(|\chi\rangle\langle\chi|))$ by going via $|\phi^\dagger\rangle$.

Result The entropy of entanglement

$$E(|\chi\rangle)=S(\mathrm{tr}_A(|\chi\rangle\langle\chi|))=S(\mathrm{tr}_B(|\chi\rangle\langle\chi|)) \tag{3.31}$$

uniquely quantifies the amount of entanglement in a pure bipartite state.

3.5 Mixed state entanglement

3.5.1 Introduction

When is a mixed state ρ_{AB} entangled?

(i) If ρ_{AB} cannot be created by LOCC (or $\rho_{AB}^{\otimes N}$).

(ii) If we can extract e-bits from ρ_{AB} or $\rho_{AB}^{\otimes N}$.

(iii) If is helps us to perform some task better in an LOCC setting.

Clearly, there are more states with (i) than (iii) than (ii) ("$\geq$"). We use (i) as the definition.

Definition 8. *States which can be prepared by LOCC*

$$\rho = \sum_i p_i \rho_i^A \otimes \rho_i^B, \quad p_i \geq 0, \ \sum_i p_i = 1, \ \rho_i^A, \rho_i^B \textit{ are density matrices} \tag{3.32}$$

*are called **separable states**. ρ is **entangled** if it is **not separable**.*

Given ρ, how can we check if it is separable / entangled? We need to check over all decompositions

$$\rho \overset{?}{=} \sum_i p_i \rho_i^A \otimes \rho_i^B. \tag{3.33}$$

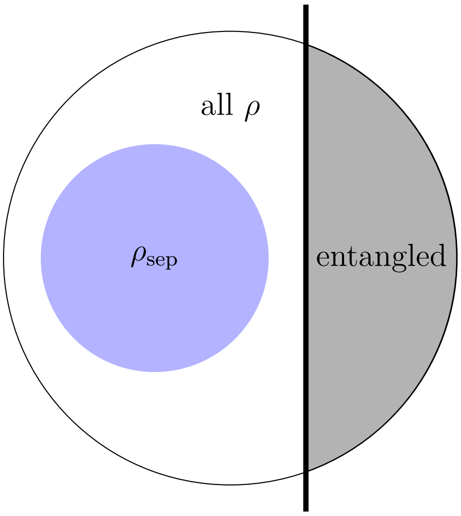
That is, we need to optimize over isometries! (Note that we can always choose ρ_i^* to be pure by further decomposite them.) In fact, the general problem is NP-hard ("exponentially" hard in dimension of space).

3.5.2 Entanglement witness

Let $\rho = \sum_i p_i \rho_i^A \otimes \rho_i^B$, $\sigma = \sum_j q_j \sigma_j^A \otimes \sigma_j^B$ be separable states. Then

$$\begin{aligned} \lambda\rho + (1-\lambda)\sigma &= \sum_k \chi_k^A \otimes \chi_k^B, \lambda \in [0,1] \\ r_k = (\lambda p_i, (1-\lambda) q_j), &\quad \chi_k^* = (\rho_i,^*, \sigma_j^*). \end{aligned} \tag{3.34}$$

$\Rightarrow \lambda\rho + (1-\lambda)\sigma$ is separable
$\Rightarrow$ separable states form a *convex set*!
Since the set of all density matrices is also convex, we can find *hyperplanes* such that all ρ on one side are entangled.



The general hyperplane is of the form

$$\operatorname{tr}(X\rho) + C = 0 \iff \operatorname{tr}\left[\underbrace{(X + C \cdot \mathbb{1})}_{:=W}\rho\right] = 0 \tag{3.35}$$

Definition 9. *An* ***entanglement witness*** *is a hermitian W such that*

$$\operatorname{tr}(W\rho) \geq 0 \;\forall \rho \textit{ separable, i.e., } \operatorname{tr}(W\rho) < 0 \Rightarrow \rho \textit{ entangled!} \tag{3.36}$$

Notes

- Need methods to show that ρ is separable $\Rightarrow \operatorname{tr}(W\rho) \geq 0$.
- A witness can only detect *certain entangled states.*
- A convex set is characterized by all tangent spaces $\Rightarrow$ there exists witness for any entangled state.
- Witnesses are *linear operators* $\Rightarrow$ they are experimentally measurable.

Example

$$W = \mathbb{F} := \sum_{i,j=1}^{d} |i,j\rangle\langle j,i| \;\text{("Flip").} \tag{3.37}$$

The "magic formula":

$$\begin{aligned}\mathrm{tr}\left[\mathbb{F}(A\otimes B)\right] &= \sum_{ij}\mathrm{tr}\left[|i,j\rangle\langle j,i|\,A\otimes B\right]\\ &= \sum_{ij}\langle j,i|A\otimes B|i,j\rangle = \sum_{ij}A_{ji}B_{ij}\\ &= \mathrm{tr}(AB)\end{aligned} \tag{3.38}$$

Let $\rho_{\text{sep}} = \sum_i p_i \rho_i^A \otimes \rho_i^B$:

$$\begin{aligned}\mathrm{tr}(W\rho_{\text{sep}}) &= \sum_i p_i \mathrm{tr}\left[\mathbb{F}(\rho_i^A\otimes\rho_i^B)\right]\\ &= \sum_i p_i \mathrm{tr}(\rho_i^A\rho_i^B) \geq 0\end{aligned} \tag{3.39}$$

Hence W is an entanglement witness. Which states does W detect?

- pure anti-symmetric state

$$|\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \Rightarrow \mathbb{F}\,|\psi^-\rangle = -\,|\psi^-\rangle \Rightarrow \langle\psi^-|\mathbb{F}|\psi^-\rangle = -1. \tag{3.40}$$

While for symmetric maximal entangled states ($|\chi\rangle = |\phi^+\rangle$, $|\phi^-\rangle$ and $|\psi^+\rangle$), $\mathbb{F}\,|\chi\rangle = |\chi\rangle \Rightarrow \langle\chi|\mathbb{F}|\chi\rangle = 1$.

- for mixed states, the general form is

$$\rho = \lambda\,|\psi^-\rangle\langle\psi^-| + (1-\lambda)\frac{\mathbb{1}}{4},\ \lambda\in[-\frac{1}{3},1], \tag{3.41}$$

and

$$\mathrm{tr}(\mathbb{F}\rho) = \lambda\underbrace{\langle\lambda^-|\mathbb{F}|\lambda^-\rangle}_{=-1} + (1-\lambda)\underbrace{\mathrm{tr}(\mathbb{F}\frac{\mathbb{1}}{4})}_{=\frac{1}{2}} = \frac{1}{2}(1-3\lambda). \tag{3.42}$$

Hence these states are entangled if $\lambda \geq 1/3$. The witness is already *optimal*: e.g., $\rho = |0\rangle\langle 0|\otimes|0\rangle\langle 0| \Rightarrow \mathrm{tr}(\mathbb{F}\rho) = 0$; the hyperplane touches the convex set!

There are other witnesses, for example, $W = \mathbb{1} - d\,|\Omega\rangle\langle\Omega|$, $|\Omega\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^d |i,i\rangle$.

3.5.3 Positive maps and the PPT criterion

Reminder A superoperator Λ is called positive if $\rho \geq 0 \Rightarrow \Lambda(\rho) \geq 0$. Usually we require Λ to be *completely positive*, i.e., $(\Lambda \otimes \mathbb{1})(\rho) \geq 0$ for $\rho \geq 0$. Here we will however be interested in the *positive but not complete positive* maps! Consider $\rho_{\text{sep}} = \sum_i p_i \rho_i^A \otimes \rho_i^B$, then

$$(\Lambda \otimes \mathbb{1})(\rho_{\text{sep}}) = \sum_i p_i \underbrace{\Lambda(\rho_i^A)}_{=\tilde{\rho}_i^A \geq 0} \otimes \rho_i^B = \sum_i p_i \tilde{\rho}_i^A \otimes \rho_i^B \geq 0. \tag{3.43}$$

i.e., $(\Lambda \otimes \mathbb{1})(\rho) \ngeq 0 \Rightarrow \rho$ is entangled.
The most important example is

$$\begin{aligned} \Lambda(\rho) &:= \rho^{\tau} \text{ (transpose)} \\ (\Lambda \otimes \mathbb{1})(\rho) &= \rho^{\tau_A} \text{ ("partial transpose")} \end{aligned} \tag{3.44}$$

i.e., $\rho = \sum_{iji'j'} \rho_{ij}^{i'j'} |i,j\rangle \langle i',j'| \to \rho^{\tau_A} = \sum_{iji'j'} \rho_{ij}^{i'j'} |i',j\rangle \langle i,j'|$. We have thus

Positive partial transpose (PPT) criterion

$$\rho^{\tau_A} \ngeq 0 \Rightarrow \rho \text{ is entangled.} \tag{3.45}$$

For example:

- $|\Omega\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^d |i,i\rangle$.

$$(|\Omega\rangle \langle\Omega|)^{\tau_A} = \frac{1}{d} \sum_{ij} (|i,i\rangle \langle j,j|)^{\tau_A} = \frac{1}{d} \sum_{ij} |j,i\rangle \langle i,j| (= \mathbb{F}), \tag{3.46}$$

  which is not positive.

- $\rho = \lambda |\Omega\rangle \langle\Omega| + (1-\lambda)\mathbb{1}/d^2$, $-\frac{1}{d^2-1} \leq \lambda \leq 1$ (isotropic state) for $d = 2$.

$$\begin{aligned} \rho &= \lambda \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix} + (1-\lambda) \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1+\lambda}{4} & 0 & 0 & \frac{\lambda}{2} \\ 0 & \frac{1-\lambda}{4} & 0 & 0 \\ 0 & 0 & \frac{1-\lambda}{4} & 0 \\ \frac{\lambda}{2} & 0 & 0 & \frac{1+\lambda}{4} \end{pmatrix} \\ \Rightarrow \rho^{\tau_A} &= \begin{pmatrix} \frac{1+\lambda}{4} & 0 & 0 & 0 \\ 0 & \frac{1-\lambda}{4} & \frac{\lambda}{2} & 0 \\ 0 & \frac{\lambda}{2} & \frac{1-\lambda}{4} & 0 \\ 0 & 0 & 0 & \frac{1+\lambda}{4} \end{pmatrix} \end{aligned} \tag{3.47}$$

It is positive if and only if $\frac{\lambda}{2} \leq \frac{1-\lambda}{4} \Rightarrow \lambda \leq \frac{1}{3}$.

Note

The criterion is independent of local unitaries on B and can therefore detect *all maximal entangled states*, which is stronger than entanglement witness! In fact, PPT criterion detects *all entangled stats* in dimension $d_A \times d_B = 2\times 2$ and 3×2. But there exist counterexamples in 3×3 and 4×2.

Another example: $\Lambda(\rho) = \mathrm{tr}(\rho)\mathbb{1} - \rho$.

$$(\Lambda \otimes \mathbb{1})(\rho) = \mathbb{1} \otimes \mathrm{tr}_A\rho - \rho = \mathbb{1} \otimes \rho_B - \rho \not\geq 0 \Rrightarrow \rho \text{ is entangled.} \tag{3.48}$$

This is called the reduction criterion:

$$\mathbb{1} \otimes \mathrm{tr}_A\rho \not\geq \rho \tag{3.49}$$

3.5.4 Relation between witness and positive maps

Theorem 5. *For each witness W, there is a positive map Λ which detects all states W detects (and in fact more).*

We can interpret a witness W as "bipartite state" (better as an operator) and map Λ is the Jamiołkowski map of "state" W^τ:

$$\Lambda(X) := \mathrm{tr}_A(W^\tau(X^\tau \otimes \mathbb{1})) = \mathrm{tr}_A(W(X \otimes \mathbb{1}))^\tau. \tag{3.50}$$

A and B are swapped here with regard to Chapter 2. Then,

$$\begin{aligned} \langle\phi|_B\Lambda(\rho)^\tau|\phi\rangle_B =& \langle\phi|\mathrm{tr}_A(W(\rho\otimes\mathbb{1}))|\phi\rangle \\ =&\mathrm{tr}(W(\underbrace{\rho\otimes|\phi\rangle\langle\phi|}_{\text{separable}})) \geq 0,\ \forall\rho\geq 0. \end{aligned} \tag{3.51}$$

i.e., Λ is a positive map and

$$\begin{aligned} \mathrm{tr}\left[W(A\otimes B)\right] &= \mathrm{tr}_B\left[\mathrm{tr}_A(W(A\otimes\mathbb{1}))B\right] = \mathrm{tr}\left[\Lambda(A)^\tau B\right] \\ &= \sum_{ij}\left[\Lambda(A)^\tau\right]_{ij} B_{ji} = d\left\langle\Omega|\Lambda(A)\otimes B|\Omega\right\rangle \\ &= d\left\langle\Omega|(\Lambda\otimes\mathbb{1})(A\otimes B)|\Omega\right\rangle \\ &\xRightarrow{\text{Linearity}} \mathrm{tr}(W\rho) = d\left\langle\Omega|(\Lambda\otimes 1)(\rho)|\Omega\right\rangle. \end{aligned} \tag{3.52}$$

i.e., $\mathrm{tr}(W\rho) < 0 \Rightarrow (\Lambda\otimes 1)(\rho) \not\geq 0$. Hence Λ detects all states which W detects.

Note

For example, when $W = \mathbb{F}$, $\Lambda(X) = \mathrm{tr}_A(\mathbb{F}(X^\tau \otimes \mathbb{1})) = X^\tau \Rightarrow$ PPT criterion! And PPT criterion is strictly stronger: $\mathbb{F}$ could not detect e.g. $|\Omega\rangle$!

Corollary 2. *A state is separable if and only if* $(\Lambda \otimes \mathbb{1})(\rho) \geq 0$ *for* ***all*** *positive* Λ.

3.5.5 Quantification of mixed state entanglement

How to *quantify* mixed state entanglement?

- Entanglement needed to create a state, e.g.,
  ”Entanglement of formation”:

$$E_F(\rho) = \min_{\{p_i, |\psi_i\rangle\}} \sum_i p_i E(|\psi_i\rangle) \quad \text{s.t.} \quad \sum_i p_i |\psi_i\rangle \langle\psi_i| = \rho. \tag{3.53}$$

  ”Entanglement of cost” (asymptotic cost per copy):

$$E_C(\rho) = \lim_{N\to\infty} \frac{1}{N} E_F(\rho^{\otimes N}) \tag{3.54}$$

- Extractable entanglement:
  ”Distillable entanglement”:

$$\begin{aligned} E_D(\rho) =& \text{max. rate } R = \frac{M}{N} \text{ achievable with LOCC protocol} \\ & \text{s.t. } \rho^{\otimes N} \to |\Omega\rangle\langle\Omega|^{\otimes M}. \end{aligned} \tag{3.55}$$

Note

$E_F \geq E_C \geq E_D$. Generally $E_C(\rho) \gneq E_D(\rho)$. For a process not reversible, there is no unique measure. For instance, ρ is called a PPT state if $\rho^{\tau_A} \geq 0$. LOCC map preserves PPT, and hence PPT states are undistillable, $E_D(\rho) = 0$. But for entangled PPT states $E_C(\rho) > 0$. The converse question: $\rho^{\tau_A} \ngeq 0 \overset{?}{\Rightarrow} E_D(\rho) > 0$ is a big open problem (existence of ”NPT bound entanglement”).

The problem is that E_F, E_C and E_D are very hard to compute. We want a *computable* entanglement measure E. The wishlist includes

(i) it is LOCC-monotonic: cannot be increased by LOCC;

(ii) has 0 value on separable states;

(iii) is additive: $E(\rho \otimes \sigma) = E(\rho) + E(\sigma)$;

(iv) is continuous: $\rho \approx \sigma \Rightarrow E(\rho) \approx E()\sigma$ (with some $\epsilon - \delta$ language);

(v) $E_D \leq E \leq E_C$;

(vi) coincides with $E(|\psi\rangle) = S(\mathrm{tr}_B |\psi\rangle \langle\psi|)$ on pure states.

It is (almost) impossible to get all; LOCC monotonicity is the most important.

Negativity - a computable entanglement measure

We have seen that ρ^{τ_A} has negative eigenvalues $\Rightarrow$ ρ is entangled. We can use this as the measure:

Definition 10. ***Negativity*** $\mathcal{N}$

$$\mathcal{N}(\rho) = \frac{1}{2}(\underbrace{\sum_i |\lambda_i(\rho^{\tau_A})|}_{:=|||\rho^{\tau_A}|_1} - 1) = -\sum_{\lambda_i<0} \lambda_i(\rho^{\tau_A}) \tag{3.56}$$

and ***log-negativity*** E_N

$$E_N(\rho) = \log ||\rho^{\tau_A}||_1. \tag{3.57}$$

Properties:

Negativity $\mathcal{N}$: (i)(ii)(iv) ((iii) and (vi) are violated).

Log-negativity E_N: (ii)(iii)(iv) ((i)! and (vi) are violated).

Chapter 4

Quantum Computation

4.1 The circuit model

4.1.1 Classical computation

Classical computers:

$$\begin{aligned}\text{Solve problems } &\equiv \text{ compute functions}\\ f : \{0,1\}^n &\to \{0,1\}^m \\ \underline{x} = (x_1, \cdots, x_n) &\mapsto f(\underline{x}) = \underline{y} = (y_1, \cdots, y_m).\end{aligned} \tag{4.1}$$

f depends on the *problem*, $\underline{x}$ encodes an *instance* of the problem.
E.g., *Multiplication* $(a, b) \mapsto a \cdot b$

$$\underline{x} = \underbrace{(\underline{x}_1, \underline{x}_2)}_{\text{encoded in binary}} \Rightarrow f(\underline{x}) = \underbrace{\underline{x_1 \cdot x_2}}_{\text{encoded in binary}} ; \tag{4.2}$$

Factorization $\underline{x}$: integers; $f(\underline{x})$: list of prime factors (with suitable encoding).

Each problem is encoded by a *family* of functions

$$f \equiv f^{(n)} = \{0,1\}^n \to \{0,1\}^m; \quad m = \text{poly}(n); \quad n \in \mathbb{N}. \tag{4.3}$$

What *ingredients* do we need to compute a general function f?

(i) $f : \{0,1\}^n \to \{0,1\}^m$, $f(\underline{x} = f_1(\underline{x}), \cdots, f_m(\underline{x})$, with $f_k : \{0,1\}^n \to \{0,1\}$. Hence we can focus on *Boolean* functions.

(ii) Let $L = \{\underline{y} | f(\underline{y}) = 1\} = \{\underline{y}^1, \cdots, \underline{y}^l\}$. Define

$$g_{\underline{y}}(\underline{x}) = \begin{cases} 0; \underline{x} \neq \underline{y} \\ 1; \underline{x} = \underline{y}. \end{cases} \tag{4.4}$$

Then $f(\underline{x}) = g_{y^1}(\underline{x}) \vee g_{y^2}(\underline{x}) \vee \cdots \vee g_{y^l}(\underline{x})$, where "$\vee$" is the *"logical OR"*: $0 \vee 0 = 0$, $0 \vee 1 = 1 \vee 0 = 1 \vee 1 = 1$. The logical or is associative.

(iii) $g_{\underline{y}}(\underline{x}) = (y_1 = x_1) \wedge (y_2 = x_2) \wedge \cdots$, where "$\wedge$" is the *"logical AND"*: $1 \wedge 1 = 1$, otherwise 0.

(iv)

$$(y_i = x_i) = \begin{cases} x_i, y_i = 1 \\ \neg x_i, y_i = 0. \end{cases} \tag{4.5}$$

"$\neg$" is the *"logical NOT"*, $\neg 1 = 0$, $\neg 0 = 1$.

As a result, $f(\underline{x})$ can be built from "AND", "OR", "NOT" gates and copy gate $x \mapsto (x, x)$. This is called a *universal gate set.*

Note

In fact ,"NAND" $\neg(x \wedge y)$ or "NOR" $\neg(x \vee y)$ are by themselves already universal together with the copy gate.

Circuit model of computation $f \equiv f^{(n)}$ can be built from a simple universal gate set without loops (i.e., gates are applied sequentially). (Technical point: the circuit family for $n \in \mathbb{N}$ must be generatable in an efficient way, e.g., a simple and fast program.)
The *hardness* of a problem is measured by the number $K(n)$ of gates needed to compute $f^{(n)}$ ($\hat{=}$ the number of time steps). It can be distinguished in different regimes:

- $K(n) \sim \text{ploy}(n)$: efficiently solvable (*class P*);
- $K(n) \gtrsim \text{ploy}(n)$, e.g. $K(n) \sim \exp(n)$: hard problem.

Questions

- Are there more hard or easy problems?
  Consider a random $f : \{0,1\}^n \to \{0,1\}$. The number of possible functions is $2^{(2^n)}$, but there are only $c^{\text{poly}(n)}$ circuits of length poly(n), where c is the number of universal gates! Therefore most f cannot be computed efficiently.
- Does computation power dependent on the gate set?
  No. By definition, any universal gate set can simulate any other gate set of few-qubit gates, with constant overhead!

- Are there other models of computation?

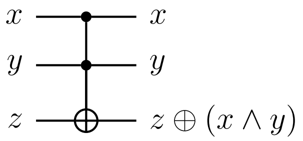
  - CPU+RAM
  - parallel computer
  - *Turing machine* (tape + head ≡ ”linear RAM”)
  - cellular automata

  ...

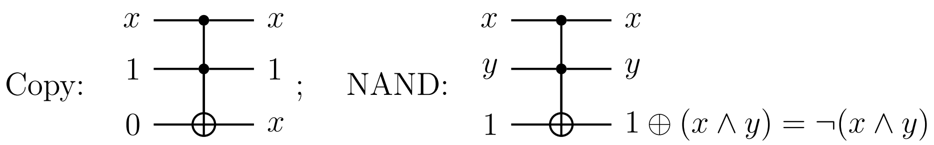
  But all known ”reasonable” models of computation can simulate each other with poly(n) overhead ⇒ (*Church–Turing thesis*) they have the same computation power! We will use the *circuit model* for quantum computer; the gates become *unitaries.*

- However, unitaries are reversible, while classical gates are not. So can we even fit classical computation into that?
  Yes! The classical computation can be turned reversible.

The *Toffoli gate* (⊕: ”XOR”, which gives the addition mod 2)



is reversible! Hence we can simulate AND, OR, NOT and copy using ancillaries, e.g.,



Any $f(\underline{x})$ can be computed reversibly with ancillaries and with essentially the same circuit:

$$f^R(\underline{x}, \underline{y}) = (\underline{x}, f(\underline{x}) \oplus \underline{y}) \tag{4.6}$$

The idea is to compute f reversibly with ancillaries, get XOR result with y and ”uncompute” everything, i.e., reset ancillaries. This can be optimized to use fewer ancillaries (See Preskill’s notes). Hence everything can be computed reversibly, but we still need the 3-qubit gate.

4.1.2 Quantum circuits

Model of quantum computation - circuit model

- System consists of qubits with tensor product structure.
- There exists universal gate set $S = \{U_,1 \cdots ,U_k\}$ of "small" (i.e., few-qubit) gates.
- Then build circuits:
- Input: *classical* input $|x_1\rangle \cdots |x_n\rangle = |x_1 \cdots x_n\rangle$ in computational basis $\{|0\rangle, |1\rangle\}$ and possibly *ancillaries* in $|0\rangle$ state.
- Output: Measure some (or all) qubits at the end in computational basis.

Notes

- Other measurements, e.g., POVM, can be simulated within the model.
- Measurements at earlier time can be always postponed.

Question Which gate set should we choose?

- Continuity of gates: the choice is much more rich!
- Different *notions of universality*:
  - *exact universality*: any $n-$ qubit gate U can be realized exactly.
  - *approximate universality*: any $n-$qubit gate U can be approximated by get set well.

    Theorem 6. *(Solovay-Kitaev) ϵ-approximation of n-qubit gate requires $O(poly(\log(1/\epsilon)))$ gates.*
- It turns out that 1 and 2-qubit gates alone are universal (in the classical case it requires 3 bits).
- Example of approximately universal gate sets: any random 2-qubit gate.
- Our *exact universal* gate sets:

(i) 1-qubit rotations about X and Z axes:

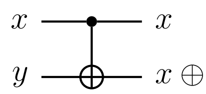
$$\begin{aligned} R_X(\phi) = e^{-iX\phi/2}; \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \cos\frac{\phi}{2} & -i\sin\frac{\phi}{2} \\ i\sin\frac{\phi}{2} & \cos\frac{\phi}{2} \end{pmatrix}, \; X^2 = \mathbb{1} \\ R_Z(\phi) = e^{-iZ\phi/2}; \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2} \end{pmatrix}, \; Z^2 = \mathbb{1} \end{aligned} \tag{4.7}$$

They can also be interpreted as rotations on *Bloch sphere* (see A.1), i.e., in $O(3) \cong SU(2)/\mathbb{Z}_2$, about X and Z direction by angle $\phi/2$. R_X and R_Z can generate any rotation in $O(3)$ with Euler angles:

$$U = R_x(\alpha)R_Z(\beta)R_X(\gamma), \; \forall U \in SU(2). \tag{4.8}$$

(ii) One 2-qubit gate (almost all would do!). Typically we choose the *controlled-NOT gate* (CNOT).

CNOT =



$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

CNOT yields y if and only if $x = 1$, which is the same as the classical XOR gate.

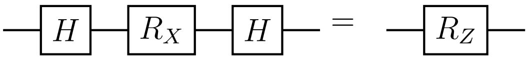
One can show that this gate set can create any n-qubit gate U exactly (of course not efficiently by counting parameters).

Some important gates and identities
(See this paper on arxiv: https://arxiv.org/abs/quant-ph/9503016 for more details on elementary gates.)

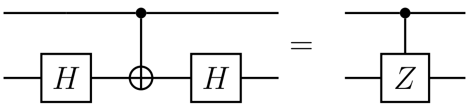
- Hadamard gate: $H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$; $H = H^\dagger$, $H^2 = \mathbb{1}$. We have

$$HR_X(\phi)H = R_Z(\phi), \quad HR_Z(\phi)H = R_X(\phi) \tag{4.9}$$

Computational notation:

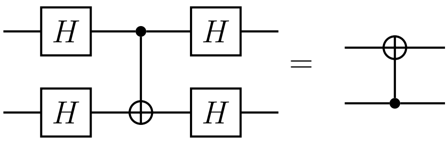


- The *"Controlled-Z"* (CZ) gate.

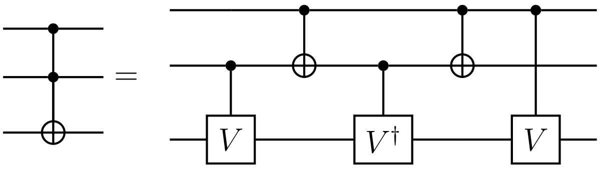


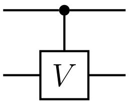
$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

-

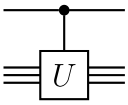


- *Toffoli* gate:

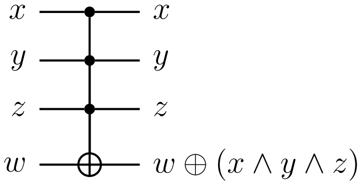


with $V = \frac{1-i}{2}(\mathbb{1} + iX)$ and  $= \begin{pmatrix} \mathbb{1} & 0 \\ 0 & V \end{pmatrix}$ being the *"controlled-V"* gate.

- If we can build a classical U, we can also build the *"controlled-U"* gate



just by replace every classical Toffoli gate (which is universal) by quantum Toffoli gate with three controls (we can actually build any n-control Toffoli gates):



Now we can name some more *approximately universal gate sets*:

- CNOT + 2 random 1-qubit gates;
- CNOT + H + $\underbrace{T}_{:=R_Z(\frac{\pi}{4}).(\text{"}\frac{\pi}{8}\text{ gate"})}$

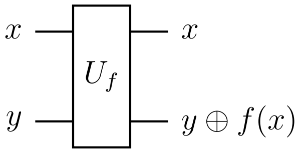
4.2 Oracle based algorithms

4.2.1 The Deutsch algorithm

Consider $f : \{0,1\} \to \{0,1\}$. Let f be ”very hard to compute” (e.g., a long circuit). We want to know: Is $f(0) \overset{?}{=} f(1)$?

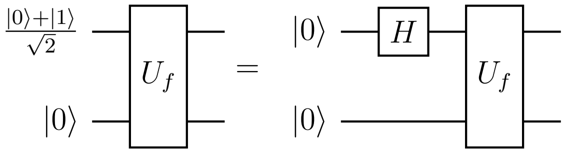
Question How often do we have to evaluate f (= run the circuit)? Classically we need 2 queries: $f(0)$ and $f(1)$. Can quantum mechanics do better? Consider a reversible implementation of f:

$$f^R(x,y) \mapsto (x, y \oplus f(x)) \tag{4.10}$$



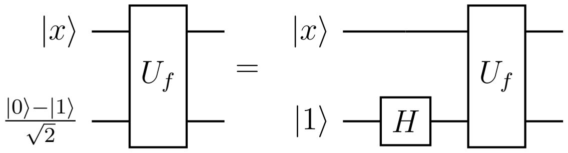
$, \; |x\rangle\, |y\rangle \mapsto |x\rangle\, |y \oplus f(x)\rangle.$

Can we try to input superposition states?



$$\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle\,|0\rangle + |1\rangle\,|0\rangle) \xmapsto{U_f} \frac{1}{\sqrt{2}}(|0\rangle\,|f(0)\rangle + |1\rangle\,|f(1)\rangle) \tag{4.11}$$

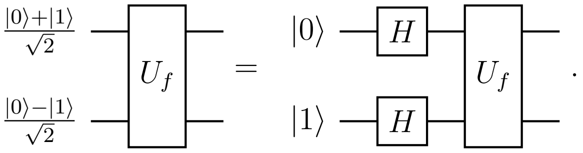
We have evaluated f on *both* inputs! But how can we *extract* the relevant information? If we measure one of the two qubits, the superposition will collapse!
Consider instead



.

$$\begin{aligned} |x\rangle\left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right) &\xmapsto{U_f} |x\rangle\left(\frac{|f(x)\rangle-|1\oplus f(x)\rangle}{\sqrt{2}}\right)\\ &= \begin{cases} |x\rangle\,\dfrac{|0\rangle-|1\rangle}{\sqrt{2}}, f(x)=0\\ |x\rangle\,\dfrac{|1\rangle-|0\rangle}{\sqrt{2}}, f(x)=1 \end{cases}\\ &= (-1)^{f(x)}\,|x\rangle\left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right) \end{aligned} \tag{4.12}$$

Hence the circuit


$\frac{|0\rangle+|1\rangle}{\sqrt{2}}$
$\frac{|0\rangle-|1\rangle}{\sqrt{2}}$
U_f
=
$|0\rangle$
$|1\rangle$
H
H
U_f
.

will give

$$\begin{aligned} |0\rangle\,|1\rangle &\xmapsto{H\otimes H} \left(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right)\\ &\xmapsto{U_f} \left(\frac{(-1)^{f(0)}\,|0\rangle+(-1)^{f(1)}\,|1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right) \end{aligned} \tag{4.13}$$

→ There is no entanglement created. (!)

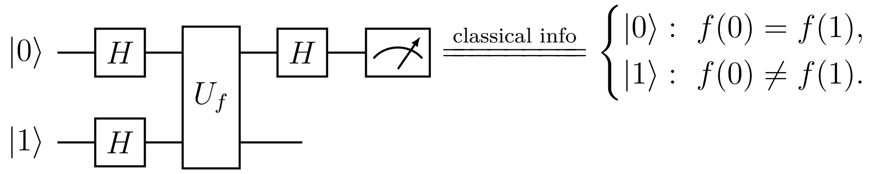
→ The second qubit leaves unchanged. (!!)

→ The first qubit gets a phase $(-1)^{f(x)}$.

⇒ this is called *"phase kickback"* technique.
Now we can measure the first qubit in $\{|+\rangle, |-\rangle\}$ basis, the result is

$$|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}} \Rightarrow f(0)=f(1);\ |-\rangle = \frac{|0\rangle-|1\rangle}{\sqrt{2}} \Rightarrow f(0)\neq f(1) \tag{4.14}$$

The complete circuit is



Only one application of U_f is sufficient ⇒ we have a speed up with regard to classical algorithm!

Note

The second qubit is never measured (and contains no information).

Main idea / points

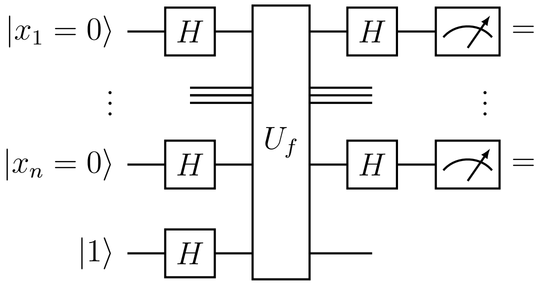
- use input $\sum |x\rangle$ to evaluate f on all inputs simultaneously.
- Need methods to read out relevant information!

4.2.2 The Deutsch-Jozsa algorithm

Consider $f : \{0,1\}^n \to \{0,1\}$ with the promise that

$$\begin{aligned} &\text{either } f(\underline{x}) = c \ \forall \underline{x} && (f \text{ is constant}), \\ &\text{or } |\{\underline{x}|f(\underline{x}) = 0\}| = |\{\underline{x}|f(\underline{x}) = 1\}| && (f \text{ is balanced}). \end{aligned}$$

Want to know: is f *constant or balanced*?
We can use the same idea: input $\sum |\underline{x}\rangle$ and $\frac{|0\rangle - |1\rangle}{\sqrt{2}}$ to $U_f : |\underline{x}\rangle\,|y\rangle \mapsto |\underline{x}\rangle\,|y \oplus f(x)\rangle$:



Before analyzing the circuit, we should first figure out: what is the action of $H^{\otimes n}$?

$$\begin{aligned} H : |x\rangle &\mapsto \frac{1}{\sqrt{2}} \sum_{y=0,1} (-1)^{xy} |y\rangle \\ H^{\otimes n} : |\underline{x}\rangle &\mapsto \frac{1}{\sqrt{2^n}} \sum_{y_1 \cdots y_n} (-1)^{x_1 y_1} \cdots (-1)^{x_n y_n} |y_1 \cdots y_n\rangle \\ &= \frac{1}{\sqrt{2^n}} \sum_{\underline{y}} (-1)^{\underline{x} \cdot \underline{y}} |\underline{y}\rangle\,, \end{aligned} \tag{4.15}$$

where $\underline{x}\cdot\underline{y} = (x_1y_1)\oplus(x_2y_2)\oplus\cdots\oplus(x_ny_n)$, which is the scalar product mod 2. The circuit now reads

$$\begin{aligned}
|\underline{0}\rangle\,|1\rangle &\xmapsto{H^{\otimes n}\otimes H} \sum_{\underline{x}} |\underline{x}\rangle\,(|0\rangle - |1\rangle) \text{ (we omit the normalization)}\\
&\xmapsto{U_f} \left(\sum_{\underline{x}} (-1)^{f(\underline{x})}\,|\underline{x}\rangle\right)(|0\rangle - |1\rangle)\\
&\xmapsto{H^{\otimes n}\otimes \mathbb{1}} \left(\sum_{\underline{y}} \underbrace{\sum_{\underline{x}} (-1)^{f(\underline{x})+\underline{x}\cdot\underline{y}}}_{=\circledast}\,|\underline{y}\rangle\right)(|0\rangle - |1\rangle).
\end{aligned} \tag{4.16}$$

$$\circledast = \begin{cases} (-1)^{\overbrace{f(\underline{x})}^{\text{constant!}}} \sum_{\underline{x}} (-1)^{\underline{x}\cdot\underline{y}} = (-1)^{f(\underline{x})}\delta_{\underline{y},\underline{0}}, & f \text{ constant;}\\ \sum_{\underline{x}} (-1)^{f(\underline{x})+\underline{x}\cdot\underline{0}} = \sum_{\underline{x}} (-1)^{f(\underline{x})} = \underline{0} \text{ when } \underline{y}=0, & f \text{ balanced.} \end{cases} \tag{4.17}$$

Thus when the output is $\underline{y} = 0$, f is constant; otherwise it is balanced. We achieve unambiguous discrimination without explicit evaluation of f!
What is the speed-up with regard to classical algorithm?

- *Quantum*: only 1 use of f.
- *Classical*: In the *worst* case, we need to test $2^n/2+1$ values of f. This is constant vs. exponential!
  But if we only want the right answer with high probability $p = 1-\epsilon$, then after k queries of f, the approximate probability of getting the same result for balanced f is (assume $k \ll 2^n$):
  $$p_{\text{error}} \approx 2(\frac{1}{2})^k \overset{!}{=} \epsilon \tag{4.18}$$
  and $k \sim \log(1/\epsilon)$. The speedup vs. *probabilistic* classical algorithm is much smaller; even for exponentially small error, $k \sim n$.

4.2.3 Simon's algorithm

$$f : \{0,1\}^n \to \{0,1\}^n \tag{4.19}$$

It is promised that $\exists\underline{a}$ such that $f(\underline{x}) = f(\underline{y})$ if and only if $\underline{x}\oplus\underline{a} = \underline{y}$ ("hidden periodicity). The problem is to find $\underline{a}$.

Classical We need to query $f(\underline{x_i})$ until $f(\underline{x_i}) = f(\underline{x_j})$ found. After k queries, we have $\sim k^2$ pairs and $P(f(\underline{x_i}) = f(\underline{x_j})) \sim 2^{-n}$. Hence $P_{\text{success}} \lesssim k^2 2^{-n}$ and we need $k \sim \exp(n)$ queries!

Quantum Start with $\frac{1}{\sqrt{2}} \sum_{\underline{x}} |\underline{x}\rangle = H^{\otimes n} |\underline{0}\rangle$:

$$\left(\frac{1}{\sqrt{2}} \sum_{\underline{x}} |\underline{x}\rangle_A \right) |\underline{0}\rangle_B \overset{U_f}{\longmapsto} \frac{1}{\sqrt{2}} \sum_{\underline{x}} |\underline{x}\rangle_A |f(\underline{x})\rangle_B . \tag{4.20}$$

Now measure $B \Rightarrow$ it collapses onto *random* $|f(\underline{x_0})\rangle$. A is collapsed to

$$c \sum_{\underline{x}: f(\underline{x}) = f(\underline{x_0})} |\underline{x}\rangle = \frac{1}{\sqrt{2}} (|\underline{x_0}\rangle + |\underline{x_0} \oplus \underline{a}\rangle). \tag{4.21}$$

How can we extract $\underline{a}$? Direct measurement in computational basis collapses to $|\underline{x_0}\rangle$ or $|\underline{x_0} \oplus \underline{a}\rangle$.
Let's apply $H^{\otimes n}$ again:

$$\begin{aligned} \frac{1}{\sqrt{2}} (|\underline{x_0}\rangle + |\underline{x_0} \oplus \underline{a}\rangle) \mapsto & \frac{1}{\sqrt{2^{n+1}}} \sum_{\underline{y}} \underbrace{\left[(-1)^{\underline{x_0} \cdot \underline{y}} + (-1)^{(\underline{x_0} \oplus \underline{a}) \cdot \underline{y}} \right]}_{= \begin{cases} 2 \cdot (-1)^{\underline{x_0} \cdot \underline{y}}, & \underline{a} \cdot \underline{y} = 0; \\ 0, & \underline{a} \cdot \underline{y} = 1; \end{cases}} |\underline{y}\rangle \\ = & \frac{1}{\sqrt{2^{n-1}}} \sum_{\underline{y}: \underline{a} \cdot \underline{y} = 0} (-1)^{\underline{x_0} \cdot \underline{y}} |\underline{y}\rangle . \end{aligned} \tag{4.22}$$

Measuring $|\underline{y}\rangle$ gives a random $\underline{y}$ such that $\underline{a} \cdot \underline{y} = 0$. $n-1$ linearly independent such $\underline{y}$ allow to determine $\underline{a}$, which requires $O(n)$ random $\underline{y}$. Hence $\underline{a}$ is found in $O(n)$ steps, which has exponential speedup with regard to classical algorithm!

Notes

- We don't even need to measure B (outcome is never used again!)
- $H^{\otimes n}$ can be understood as Fourier transformation over $\mathbb{Z}_2^{\otimes n}$. Will use quantum Fourier transformation to find periodicity later!

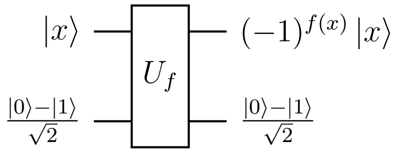
4.3 Grover's algorithm

For many hard computational problems, it is possible to *check solution efficiently*, but we don't know how to *find* it. They are so called *"NP problems"*. Examples include graph coloring, factoring, 3-SAT, Hamiltonian path, tiling problems and so on.

Reformulation We can compute $f(x) \in \{0,1\}$, $x \in \{0, 1, \cdots, N-1\}$; $f(x)$ is the *verifier* for a solution x, where $f(x) = 1$ means that the solution is correct. We want to find some x_0 such that $f(x_0) = 1$. It can be interpreted as "database search": we want to find "marked element" x_0 in an unstructured database.
Classically, $O(N)$ queries to f is needed for an *unstructured search*, (i.e., without using properties of f). While for a *quantum computer*, we will show that $O(\sqrt{N})$ queries is enough. It is only a quadratic speedup, but for a very large class of relevant problems.

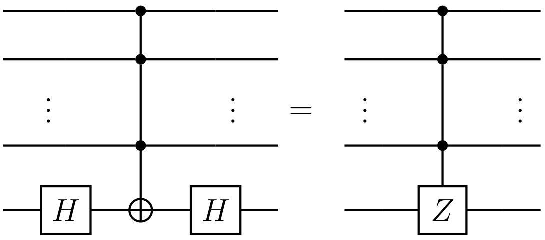
Ingredients

1. Oracle $O_f : |x\rangle \mapsto (-1)^{f(x)} |x\rangle = (-1)^{\delta_{x,x_0}} |x\rangle$. It can be built via


$|x\rangle$
$\frac{|0\rangle-|1\rangle}{\sqrt{2}}$
U_f
$(-1)^{f(x)} |x\rangle$
$\frac{|0\rangle-|1\rangle}{\sqrt{2}}$

   i.e., $O_f = \mathbb{1} - 2 |x_0\rangle \langle x_0|$ flips the amplitude of the *wanted* element.

2. Unitary $O_0 : |x\rangle \mapsto (-1)^{\delta_{x,0}} |x\rangle$. This corresponds to $(n-1)$-CZ gate:



   Again, $O_0 = \mathbb{1} - 2 |0\rangle \langle 0|$. Also define $O_\omega := H^{\otimes n} O_0 H^{\otimes n} = \mathbb{1} - 2 |\omega\rangle \langle \omega|$, where $|\omega\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle$. (We assumed here $N = 2^n$, but it is not necessary; every search problem can be trivially extended to $N = 2^n$.)

Algorithm Start from $|\psi_0\rangle = |\omega\rangle = H^{\otimes n} |0\rangle$. Apply the *Grove's iteration*

$$\begin{aligned} G &= -H^{\otimes n} O_0 H^{\otimes n} O_f = -O_\omega O_f, \\ |\psi_k\rangle \mapsto |\psi_{k+1}\rangle &= G |\psi_k\rangle = -O_\omega O_f |\psi_k\rangle . \end{aligned} \tag{4.23}$$

Observation There are only 2 special vectors in O_f and O_ω: $|x_0\rangle$ and $|\omega\rangle$. Hence we can analyze everything in two dimensional space spanned by $|x_0\rangle$ and $|\omega\rangle$! Define

$$\begin{aligned} |\alpha\rangle &:= |x_0\rangle \\ |\beta\rangle &:= \frac{1}{\sqrt{N-1}} \sum_{x \neq x_0} |x\rangle \propto |\omega\rangle - \frac{1}{\sqrt{N}} |x_0\rangle \end{aligned} \tag{4.24}$$

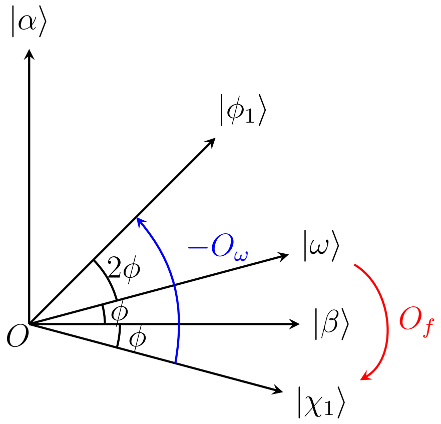
such that $\langle \alpha, \beta \rangle = 0$. Any state can be written as

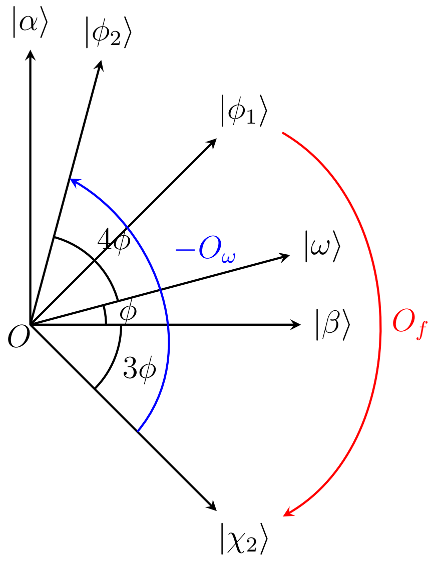
$$a\,|\alpha\rangle + b\,|\beta\rangle = x\,|\omega\rangle + y\,|\omega^\perp\rangle\,, \tag{4.25}$$

where $\langle \omega | \omega^\perp \rangle = 0$. The effects of O_f and $-O_\omega$ read

$$\begin{aligned} O_f(a\,|\alpha\rangle + b\,|\beta\rangle) &= -a\,|\alpha\rangle + b\,|\beta\rangle\,, \\ (-O_\omega)(x\,|\omega\rangle + y\,|\omega^\perp\rangle) &= x\,|\omega\rangle - y\,|\omega^\perp\rangle\,. \end{aligned} \tag{4.26}$$

They are reflections about $|\beta\rangle$ and $|\omega\rangle$ respectively. What happens if we start with $|\psi_0\rangle = |\omega\rangle$?





Let $|\chi_k\rangle = O_f(-O_\omega O_f)^{k-1}\,|\omega\rangle$ and $|\phi_k\rangle = (-O_\omega O_f)^k\,|\omega\rangle$. The first round is

$$\begin{aligned} |\omega\rangle &= \sin\phi\,|\alpha\rangle + \cos\phi\,|\beta\rangle \\ |\chi_1\rangle &:= O_f\,|\omega\rangle \\ |\phi_1\rangle &:= -O_\omega\,|\chi_1\rangle = \sin(3\phi)\,|\alpha\rangle + \cos(3\phi)\,|\beta\rangle \end{aligned} \tag{4.27}$$

and after the second iteration,

$$|\phi_2\rangle = \sin(5\phi)\,|\alpha\rangle + \cos(5\phi)\,|\beta\rangle\,. \tag{4.28}$$

Finally

$$|\phi_k\rangle = \sin((2k+1)\phi)\,|\alpha\rangle + \cos((2k+1)\phi)\,|\beta\rangle\,. \tag{4.29}$$

We want $\phi_k = (2k+1)\phi \approx \frac{\pi}{2}$, which yields $|\alpha\rangle = |x_0\rangle$! The initial state

$$\begin{aligned}
|\omega\rangle &= \frac{1}{\sqrt{N}}\,|\alpha\rangle + \sqrt{\frac{N-1}{N}}\,|\beta\rangle = \sin\phi\,|\alpha\rangle + \cos\phi\,|\beta\rangle \\
\Rightarrow \tan\phi &= \frac{\sin\phi}{\cos\phi} = \frac{1}{\sqrt{N-1}} \\
\Rightarrow \phi &\approx \frac{1}{\sqrt{N}} \text{ for large } N \\
\Rightarrow \text{Need } k &\approx \frac{\pi}{4}\sqrt{N} \Rightarrow O(\sqrt{N}) \text{ calls of } f \text{ is sufficient.}
\end{aligned} \tag{4.30}$$

There's *quadratic speedup* with regard to classical algorithms for *general search problems*!

Notes

- When there are K solutions, the same method works with $O(\sqrt{N/K})$ steps.
- This can be adopted to the case where K is unknown.

4.4 The quantum Fourier transformation, period finding and Shor's factoring algorithm

Recall In Simon's algorithm, we use $H^{\otimes n}$ as Fourier transformation over $\mathbb{Z}_2^{\otimes n}$ to find period in $\mathbb{Z}_2^{\otimes n}$.

- Can we construct a general quantum Fourier transformation?
- Can it be implemented efficiently?
- Are there any applications?

4.4.1 The quantum Fourier transformation

The Fourier transformation (FT) on $\mathbb{C}^N$ is

$$\begin{aligned} & x = (x_0, \cdots, x_{N-1}) \in \mathbb{C}^N \\ & x \mapsto y \in \mathbb{C}^N, \ y_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_j e^{2\pi i jk/N}. \end{aligned} \tag{4.31}$$

Similarly, we can define

Definition 11. *Quantum Fourier transformation (QFT):*

$$|j\rangle \mapsto \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{2\pi i jk/N} |k\rangle \tag{4.32}$$

The direct computational cost of FT is $O(N^2)$ which is exponential in the number of bits n $(N = 2^n)$. The *fast Fourier transformation (FFT)* algorithm has time cost $O(N \log N)$, but it is still $O(\exp(n))$. We will show that QFT can be implemented in $O(n^2)$ steps. There's *exponential speedup*!
First rewrite QFT in binary form:

1. $N = 2^n$;

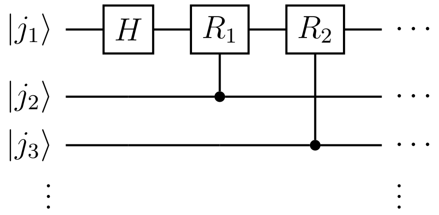
2. Write j in binary: $j = j_1 \cdots j_N = j_1 2^{n-1} + \cdots + j_n 2^0$;

3. Decimal point: $0.j_l \cdots j_m = \frac{j_1}{2^1} + \cdots \frac{j_m}{2^{m-l+1}}$.

Then

$$\begin{aligned} |j\rangle \mapsto & \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} e^{2\pi i jk/2^n} |k\rangle \\ = & \frac{1}{2^{n/2}} \sum_{k_1=0,1} \cdots \sum_{k_n=0,1} e^{2\pi i j(\sum_{l=1}^n k_l 2^{-l})} |k_1, \cdots, k_n\rangle \\ = & \bigotimes_{l=1}^n \left(\frac{1}{\sqrt{2}} \sum_{k_l=0,1} e^{2\pi i j k_l 2^{-l}} |k_l\rangle \right) \\ = & \bigotimes_{l=1}^n \frac{|0\rangle + e^{2\pi \overbrace{ij2^{-l}}^{=j_1 \cdots j_{n-l} . j_{n-l+1} \cdots j_n}} |1\rangle}{\sqrt{2}} = \bigotimes_{l=1}^n \frac{|0\rangle + e^{2\pi i 0.j_{n-l+1} \cdots j_n} |1\rangle}{\sqrt{2}} = \bigotimes_{l=1}^n |\tilde{j_l}\rangle . \end{aligned} \tag{4.33}$$

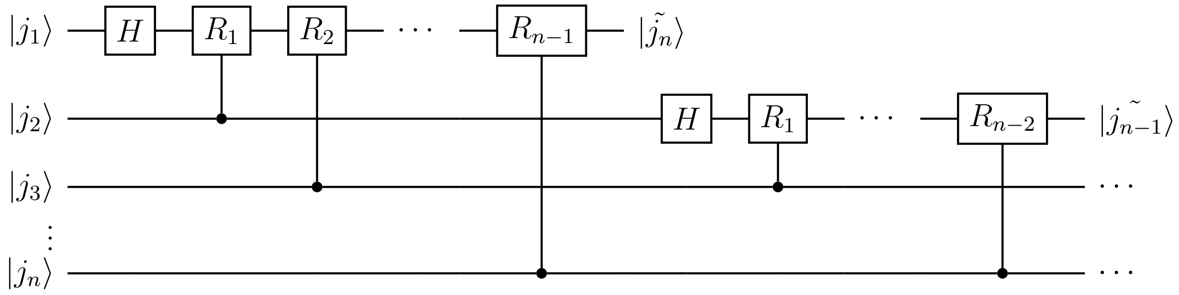
How to implement this map? We start with the right most term $|\tilde{j_n}\rangle = \frac{|0\rangle + e^{2\pi i 0.j_1 \cdots j_n} |1\rangle}{\sqrt{2}}$. The circuit acts as



with $R_d = \begin{pmatrix} 1 & 0 \\ 0 & e^{2\pi i 2^{-(d+1)}} \end{pmatrix}$.

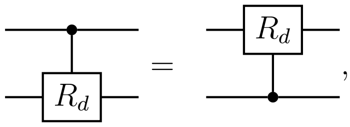
$$\begin{aligned} H :& |j_1\rangle \mapsto |0\rangle + e^{2\pi i 0.j_1} |1\rangle \\ C - R_1 :& (|0\rangle + e^{2\pi i 0.j_1} |1\rangle) |j_2\rangle \mapsto (|0\rangle + e^{2\pi i 0.j_1 j_2} |1\rangle) |j_2\rangle \\ C - R_2 :& (|0\rangle + e^{2\pi i 0.j_1 j_2} |1\rangle) |j_3\rangle \mapsto (|0\rangle + e^{2\pi i 0.j_1 j_2 j_3} |1\rangle) |j_3\rangle \\ & \vdots \text{ etc.} \end{aligned} \tag{4.34}$$

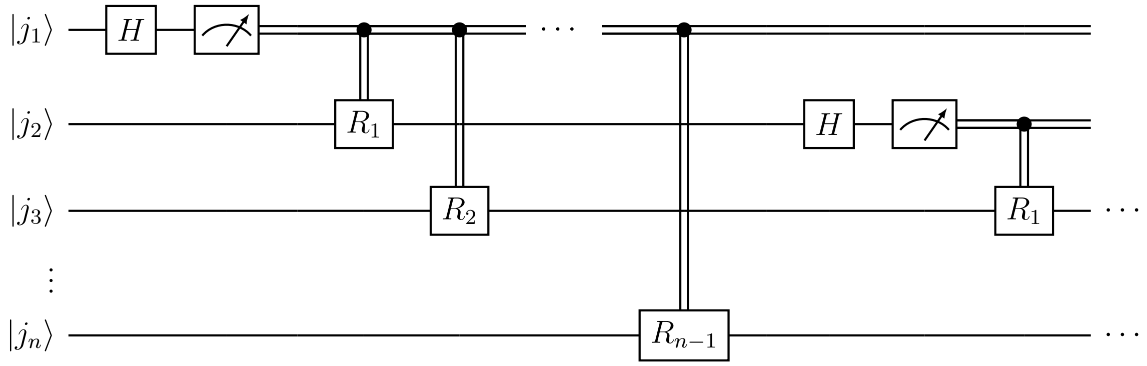
Hence we obtain the n-th qubit of QFT on the first qubit. Continue for $(j_2, \cdots, j_n)$, $(j_3, \cdots, j_n)$, $\cdots$:



Totally we need $n(n+1)/2 = O(n^2)$ gates!

Notes

- Output is in *reverse order* (but reordering needs only $\sim O(n)$ operations).
- Since , we can reverse the C-R_d gates. The upper controls act in computational basis: if we measure after QFT, we can measure after H and control R_d-gates *classically*! Then only 1-qubit gates are needed!



(Double wire represents classical control)

4.4.2 Period finding

Consider $f : \{0,1\}^n \to \{0,1\}^n$ such that $\exists r > 0$, $f(x) = f(x+r)$ (and otherwise $f(x) \neq f(y)$). Can we find r?
Use $U_f : |x\rangle |y\rangle \mapsto |x\rangle |y \oplus f(x)\rangle$:

1. $\frac{1}{2^{n/2}} \sum_x |x\rangle_A |0\rangle_B \mapsto \frac{1}{2^{n/2}} \sum_x |x\rangle_A |f(x)\rangle_B$;

2. Measure B $\to A$ collapses to $|\psi\rangle = \frac{1}{\sqrt{k_0}} \sum_{k=0}^{k_0-1} |x_0 + kr\rangle$. (As for Simon, we could omit this step!)

3. Apply QFT and measure in computational basis:
$$\begin{aligned} |\psi\rangle &\mapsto \frac{1}{2^{n/2}\sqrt{k_0}} \sum_{k=0}^{k_0-1} \sum_{l=0}^{2^n-1} e^{2\pi i (x_0+kr)l/2^n} |l\rangle \\ &= \sum_{l=0}^{2^n-1} \underbrace{\left(\sum_{k=0}^{k_0-1} \frac{1}{\sqrt{2^n k_0}} e^{2\pi i k r l/2^n} \right)}_{=:a_l} |l\rangle \end{aligned} \tag{4.35}$$
and $|a_l|^2$ is the probability of outcome l.

4. If $r \ll 2^n$, many values of k can be taken and it is almost periodic. Hence $|a_l|^2$ is peaked around l such that $rl/2^n \approx$ integers.

5. Explicit analysis of a_l shows: with high probability, we obtain l such that $l/2^n \approx s/r$. If $r \ll 2^n$, this can be used to determine s/r with high probability. If s and r are coprime – this happens with large enough probability, related to the density of primes – we can infer r!

This is the *quantum algorithm for period finding*!

4.4.3 Application: factoring

One use of the period finding is *factoring*. Given N (not prime), we want to find non-trivial r such that $r|N$.

Algorithm

1. Select random $2 \leq a < N$. If $\gcd(a, N) > 1$, it's already done! (gcd can be efficiently computed by Euclid's algorithm.) So we can assume $\gcd(a, N) = 1$.

2. Let r be the smallest r such that $a^r \equiv 1 \mod N$.
   (Existence: $\exists x, y : a^x \equiv a^y \mod N \Rightarrow a^x(1 - a^{y-x}) \equiv 0 \mod N \Rightarrow N|a^x(1 - a^y) \Rightarrow N|1 - a^{y-x} \Rightarrow a^{y-x} \equiv 1 \mod N$.)
   r is the period of $f_{N,a}(x) = a^x \mod N$. With $x = x_{m-1}2^{m-1} + \cdots + x_0 2^0$, $f_{N,a}(x)$ can be computed efficiently:

   $$a^x \equiv (a^{(2^{m-1})})^{x_{m-1}} \cdots (a^{(2^0)})^{x_0} \mod N, \tag{4.36}$$

   Hence r can be *found efficiently* with a *quantum computer*!

3. Assume r is even: $a^r \equiv 1 \mod N \iff N|(a^r - 1) \iff N|(a^{r/2} - 1)(a^{r/2} + 1)$ and $N \nmid (a^{r/2} - 1)$ (otherwise $r/2$ is the smallest possible instead).
   $\Rightarrow$ Either $N|(a^{r/2} + 1)$ or N has non-trivial common factors with both $a^{r/2} \pm 1$.
   The latter $\Rightarrow 1 \neq \gcd\ (N, a^{r/2} + 1)|N \Rightarrow$ found a non trivial factor of N!
   The algorithm is successful as long as (i) r is even and (ii) $N \nmid (a^{r/2}+1)$. It can be shown to happen with probability $p \geq \frac{1}{2}$ for a random choice of a (unless $N = p^k$, p being a prime; but this can be checked by taking logarithms).

This efficient quantum algorithm for factoring is called "Shor's algorithm".

Chapter 5

Quantum Error Correction

5.1 Introduction

- The coupling of quantum systems to *environment* induces *errors*.
- Classical computers are *"macroscopic"* so that the errors are unlikely to occur.
- For quantum computers, we need qubits = "single" quantum systems, which are fragile; the system also have to be coupled to the environment to realize gates!

So, can we protect quantum information from noise?

Classical error correction We *copy* information, e.g., encode a bit in 3 copies:

$$\begin{aligned} 0 &\mapsto \hat{0} = 000 \\ 1 &\mapsto \hat{1} = 111 \end{aligned} \tag{5.1}$$

A bit may flip with some (small) probability p, so the typical 0 or 1 bit flipped correction is the majority vote, i.e.

$$\begin{aligned} 000, 001, 010, 100 &\mapsto 000 \\ 111, 110, 101, 011 &\mapsto 111 \end{aligned} \tag{5.2}$$

$p_{\text{error}} = P(\geq 2 \text{ flips}) = p^3 + 3p^2(1-p) \leq 3p^2 < p$ for $p < 1/3$! Hence the effective error probability is decreased.
The method can be improved by

- using more bits;
- using smarter codes (encoding k bits);
- nesting (”concatenating”) codes.

Quantum error correction The general potential problems are

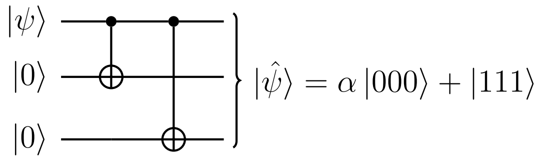
- not being able to copy qubits (and even if it is possible, how do we compare them?);
- different *types of errors*, e.g., X (bit flip) or Z (phase flip);
- errors can be *continuous*;
- *measuring* qubits *destroys quantum information*!

5.1.1 The 3-qubit bit flip code

We can copy qubits in fixed basis:

$$\begin{aligned} |0\rangle &\mapsto |\hat{0}\rangle = |000\rangle \\ |1\rangle &\mapsto |\hat{1}\rangle = |111\rangle \end{aligned} \tag{5.3}$$

i.e., $|\phi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \xrightarrow{\text{encoding}} |\hat{\psi}\rangle = \alpha\,|000\rangle + \beta\,|111\rangle$. The *encoding circuit* is



Bit flip error Consider a *bit flip error* $|\hat{\psi}\rangle \mapsto X_i\,|\hat{\psi}\rangle$ on qubit i. Can we correct for *one bit flip error* on unknown qubit? The problem is that measuring all qubits destroys quantum information! So we need measurements which only returns information about location of the error (independent of encoded state $|\psi\rangle$).

Definition 12. ***Syndrome measurement*** *with projectors:*

”no flip”: $P_0 = |000\rangle\,\langle 000| + |111\rangle\,\langle 111|$

"first qubit flipped": $P_1 = |100\rangle\langle 100| + |011\rangle\langle 011|$

"second qubit flipped": $P_2 = |010\rangle\langle 010| + |101\rangle\langle 101|$

"third qubit flipped": $P_3 = |001\rangle\langle 001| + |110\rangle\langle 110|$

Measuring $\{P_\alpha\}$ reveals only 2 bits of information and the rest one qubit is untouched. By inspection, the obtained information is the location of the flip. For instance, after X_1 flip, the measurement will return P_1, and the post measurement state $\alpha|100\rangle + \beta|011\rangle$ state can be recovered by flipping the first qubit. (By linearity, this also works for parts of large entangled states)

Continuous error What about *continuous errors*, such as

$$|\hat{\psi}\rangle \mapsto e^{i\theta X_i}|\hat{\psi}\rangle = (\cos\theta\mathbb{1} + i\sin\theta X_i)|\hat{\psi}\rangle \text{ ?} \tag{5.4}$$

For $|\hat{\psi}\rangle = \alpha|000\rangle + \beta|111\rangle$,

$$|\hat{\psi}\rangle \xmapsto[\text{qubit 1}]{\text{error, e.g.}} \underbrace{\cos\theta(\alpha|000\rangle + \beta|111\rangle)}_{\text{syndrome}P_0} + \underbrace{i\sin\theta(\alpha|100\rangle + \beta|011\rangle)}_{\text{syndrome}P_1} \tag{5.5}$$

The syndrome measurement collapses the state to

$$\begin{aligned} p = \cos^2\theta : P_0 \Rightarrow \alpha|000\rangle + |111\rangle\,, &\text{ there is no correction } \checkmark \\ p = \sin^2\theta : P_1 \Rightarrow \alpha|100\rangle + |011\rangle\,, &\text{ then flip bit 1 } \checkmark \end{aligned} \tag{5.6}$$

The measurement of error syndrome P_i collapses error onto "digital" error – no error or bit flip – and it is hence sufficient to study discrete errors!

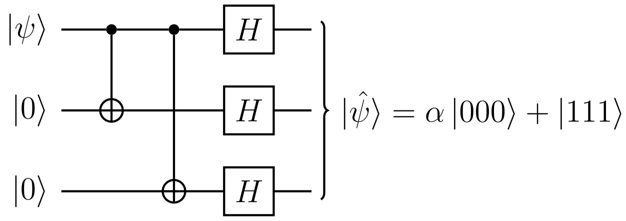
Phase flip error What about Z error?

$$\alpha|000\rangle + \beta|111\rangle \xrightarrow[\text{on some qubit}]{Z\text{ error}} \alpha|000\rangle - \beta|111\rangle \tag{5.7}$$

This is still in *code space* (i.e., space of valid $|\hat{\psi}\rangle$) so the error is *not detectable*, but it has changed $|\hat{\psi}\rangle$! The 3-qubit bit flip code cannot protect against "phase flip" Z. In fact, the phase flip Z_i acts as logical Z on the encoded qubit.

5.1.2 3-qubit phase flip code

We have $Z\left|+\right\rangle = \left|-\right\rangle$, $Z\left|-\right\rangle = \left|+\right\rangle$. The Z error $\hat{=}$ bit flip error in $\left|\pm\right\rangle$ basis. Hence encoding $|\hat{0}\rangle = |+++\rangle$ and $|\hat{1}\rangle = |---\rangle$ will protect against Z errors!

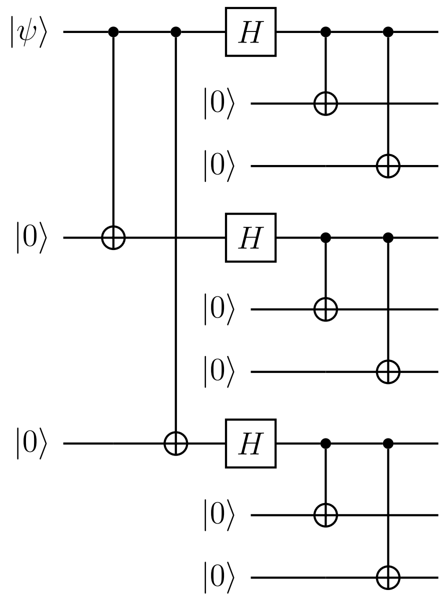


The syndrome measurement $\tilde{P_\alpha} := H^{\otimes 3} P_\alpha H^{\otimes 3}$ recovering $HX_\alpha H = Z_\alpha$. The problem is that there is no protection against bit-flip errors and X_i acts as X on the logical qubit.

5.2 The 9-qubit Shor code

Solution We can concatenate (nest) 3-qubit bit flip with 3-qubit phase flip code! It is called the *9-qubit Shor code.*

$$\begin{aligned} |0\rangle &\mapsto |+\rangle\,|+\rangle\,|+\rangle \mapsto \frac{((|000\rangle + |111\rangle)(|000\rangle + |111\rangle)\,|000\rangle + |111\rangle)}{2\sqrt{2}} \\ |1\rangle &\mapsto |-\rangle\,|-\rangle\,|-\rangle \mapsto \frac{((|000\rangle - |111\rangle)(|000\rangle - |111\rangle)\,|000\rangle - |111\rangle)}{2\sqrt{2}} \end{aligned} \tag{5.8}$$



It can correct *any* single qubit Pauli errors:

(i) X_i error is corrected on the "inner" layer;

(ii) Z_i error is corrected on the "outer" layer;

(iii) $Y_i \propto X_i Z_i$: X_i and Z_i errors are corrected independently.

5.3 Quantum error correcting code

Definition 13. *Quantum error correcting code (QECC) is defined by* ***a code space*** *$C \subset \mathcal{H}$ (elements of which are "code words") and choose a basis $|\hat{i}\rangle$.*

noise model a CPTP map

$$\mathcal{E}(\rho) = \sum_\alpha E_\alpha \rho E_\alpha^\dagger; \ \sum_\alpha E_\alpha^\dagger E_\alpha = \mathbb{1}. \tag{5.9}$$

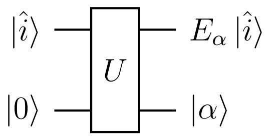
Recovery procedure Measurement + recovery $\leftrightarrow$ another CP map $\mathcal{R}$ requiring that $\mathcal{R}(\mathcal{E}(\rho)) = \rho \ \forall \rho = |\psi\rangle\langle\psi|$, $|\psi\rangle \in C$.

Necessary conditions Under which conditions on C and $\mathcal{E}$ does such an $\mathcal{R}$ exist?

(i) Environment carries no information about ρ for any $\rho = \sum_{ij} \rho^{ij} |\hat{i}\rangle |\hat{j}\rangle$ in C. By linearity, $\underbrace{\langle\hat{i}|E_\alpha^\dagger E_\alpha|\hat{i}\rangle}_{=P(\alpha)} = c_\alpha$, which is independent of $\hat{i}$.

Recall:

$$|\hat{i}\rangle\langle\hat{i}| \mapsto \mathcal{E}(|\hat{i}\rangle\langle\hat{i}|) \iff$$



(ii) orthogonal states must remains orthogonal ($\mathcal{R}$ cannot make states more orthogonal!)

$$\mathcal{E}(|\hat{i}\rangle\langle\hat{i}|) \perp \mathcal{E}(|\hat{j}\rangle\langle\hat{j}|) \text{ for } \langle\hat{i}|\hat{j}\rangle = 0. \tag{5.10}$$

i.e.,

$$\begin{aligned} \delta_{ij} &\propto \operatorname{tr}\left[\mathcal{E}(|\hat{i}\rangle\langle\hat{i}|)(\mathcal{E}(|\hat{j}\rangle\langle\hat{j}|))\right] \\ &= \sum_{\alpha\beta} \operatorname{tr}\left(E_\alpha |\hat{i}\rangle\langle\hat{i}| E_\alpha^\dagger E_\beta |\hat{j}\rangle\langle\hat{j}| E_\beta^\dagger\right) \\ &= \sum_{\alpha\beta} |\langle\hat{i}|E_\alpha^\dagger E_\beta\rangle|^2 \end{aligned} \tag{5.11}$$

(iii) (i) + (ii) ⇒

$$\langle\hat{i}|E_\alpha^\dagger E_\beta|\hat{j}\rangle = c_{\alpha\beta}\delta_{ij}, \; c_{\alpha\beta} = c_{\beta\alpha}^* \tag{5.12}$$

This is the *quantum error correction condition.* The above is also *sufficient*: Use the gauge degree of freedom

$$\sum_\alpha E_\alpha \rho E_\alpha^\dagger = \sum_\beta F_\beta \rho F_\beta^\dagger \text{ with } F_\beta = \sum_\alpha \underbrace{V_{\beta\alpha}}_{\text{isometry}} E_\alpha \tag{5.13}$$

to choose F_β such that $c_{\alpha\beta}$ becomes diagonal,

$$\langle\hat{i}|F_\alpha^\dagger F_\beta|\hat{j}\rangle = \lambda_\alpha \delta_{\alpha\beta}\delta_{ij}. \tag{5.14}$$

i.e., different F_α can be distinguished by measurement and undone:

$$\underbrace{\left(\frac{1}{\lambda_\beta}\sum_{\hat{i}} |\hat{i}\rangle\langle\hat{i}| F_\beta^\dagger\right)}_{\text{Kraus operator of recovery map}} F_\alpha |\hat{j}\rangle = \delta_{\alpha\beta} |\hat{j}\rangle\,. \tag{5.15}$$

Note

For a single-qubit error,

$$E_\alpha = \sum_{ks} \omega_{\alpha ks} \underbrace{\sigma_s^k}_{\sigma^k \text{ on qubit } s}, \tag{5.16}$$

i.e., $\langle\hat{i}|(\sigma_s^k)^\dagger \sigma_r^l|\hat{j}\rangle \propto \delta_{ij} \Rightarrow \langle\hat{i}|E_\alpha^\dagger E_\beta|\hat{j}\rangle \propto \delta_{ij}$.
i.e., error correction condition for Pauli's ⇒ for *any* single qubit error!
In particular, robust to depolarizing channel

$$\mathcal{E}(\rho) = p\rho + \frac{1-p}{3}(X\rho X + Y\rho Y + Z\rho Z) \tag{5.17}$$

on every qubit ⇒ robust to *any* 1-qubit noise! And it's similar for k-qubits errors and k-fold Pauli products!

Basic properties of QECC We focus on "binary codes": encode k qubits in $n > k$ qubits.

Definition 14. ***Distance*** *d: the smallest number of Pauli's ($\neq \mathbb{1}$) in E_α such that*

$$\langle\hat{i}|E_\alpha|\hat{j}\rangle \neq \lambda_\alpha \delta_{ij}. \tag{5.18}$$

Notation:$[\![\underbrace{n}_{\text{phys. qubits}}, \underbrace{k}_{\text{encoded qubits}}, \underbrace{d}_{\text{distance}}]\!]$ - code.
How many 1-qubit errors t can a distance d code correct? For E_α, E_β with $\leq t$ Pauli's:

$$\langle\hat{i}| \underbrace{E_\alpha^\dagger E_\beta}_{\leq 2t \text{ Pauli's}} |\hat{j}\rangle \overset{?}{=} c_{\alpha\beta}\delta_{ij} \iff 2t+1 \leq d \tag{5.19}$$

i.e., for $d = 3$, we can correct all 1-qubit codes.

Note

If location of errors is *known*, $E_\alpha^\dagger E_\beta$ has $\leq t$ Pauli's ⇒ $t + 1 \leq d$.
Or: code can correct t errors in *unknown* locations ⟺ t can correct $2t$ errors in *known* locations.

Are there constraints on $[\![n, k, d]\!]$? The simplest case is $k = 1$, $d = 3$. What is minimal n? We claim that $n \geq 5$.

Proof. Use the no-cloning theorem. Assume $n = 4$ is possible.

$$\begin{aligned} &|\psi\rangle \xrightarrow{\text{encode}} |\hat{\psi}\rangle = |\underbrace{\circ\circ}_{\text{trance out}} \quad \circ\circ\rangle \to | \quad \underbrace{\star\star}_{\text{put random states}} \quad \circ\circ\rangle \\ &\xrightarrow[\Rightarrow\text{recoverable!}]{\text{2 errors in known location}} |\psi\rangle \end{aligned} \tag{5.20}$$

So we can recover $|\psi\rangle$ from any two known positions:

$$|\psi\rangle \xrightarrow{\text{encode}} \begin{array}{l} \left.\begin{array}{c}\circ\\ \circ\end{array}\right\} \to |\circ\circ\star\star\rangle \to |\psi\rangle \\ \left.\begin{array}{c}\circ\\ \circ\end{array}\right\} \to |\star\star\circ\circ\rangle \to |\psi\rangle \end{array} \tag{5.21}$$

This means we make two copies of $|\psi\rangle$, which violates the no-cloning theorem! □

$[\![5,1,3]\!]$ code indeed exists (Raymond Laflamme et. al., 1996) and it is optimal!

Appendices

Appendix A

Homework Materials

A.1 Bloch sphere

For pure states Any pure state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ can be written as

$$|\psi\rangle = e^{i\delta}\left(\cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle\right), \tag{A.1}$$

where $0 \leq \theta \leq \pi$, $0 \leq \phi < 2\pi$ and $e^{i\delta}$ is an irrelevant global phase. The angles θ and ϕ can be interpreted as spherical coordinates describing a point on a sphere, which is the so-called *Bloch sphere*.
Eq. (A.1) implies that

$$|\psi\rangle\langle\psi| = \frac{1}{2}(\mathbb{1} + \boldsymbol{v}\cdot\boldsymbol{\sigma}), \text{ with } \boldsymbol{v} \in \mathbb{R}^3,\ |\boldsymbol{v}| = 1. \tag{A.2}$$

The vector $\boldsymbol{v} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ is exactly the point on the Bloch sphere, which is called the *Bloch sphere representation* of the state $|\psi\rangle$.
For any two states $|\psi\rangle$ and $|\phi\rangle$ that are orthogonal to each other, they are described by opposite Bloch vectors.

For mixed states Any hermitian 2 matrix ρ with $\mathrm{tr}\rho = 1$ can be written as

$$\rho = \frac{1}{2}(\mathbb{1} + \boldsymbol{r}\cdot\boldsymbol{\sigma}), \text{ with } \boldsymbol{r} \in \mathbb{R}^3. \tag{A.3}$$

ρ is a density operator if and only if $|\boldsymbol{r}| \leq 1$, and is pure if and only if $|\boldsymbol{r}| = 1$.

A.2 Examples of quantum channels

1. *Trace.*

$$\mathcal{E}(\rho) = \mathrm{tr}\rho. \tag{A.4}$$

2. *Dephasing channel.*

$$\mathcal{E}(\rho) = (1-p)\rho + pZ\rho Z. \tag{A.5}$$

   The action on the Bloch vector is

$$(r_x, r_y, r_z) \mapsto ((1-2p)r_x, (1-2p)r_y, r_z). \tag{A.6}$$

3. *Amplitude damping channel.* The Kraus operators are

$$\Pi_0 = \sqrt{\gamma}\,|0\rangle\langle 1|\,,\ \Pi_1 = |0\rangle\langle 0| + \sqrt{1-\gamma}\,|1\rangle\langle 1|\,, \tag{A.7}$$

   which describes a decay from $|1\rangle$ to $|0\rangle$ with decay rate γ.

4. *Twirling operation.*

$$\mathcal{E}(\rho) = \frac{1}{4}(\rho + X\rho X + Y\rho Y + Z\rho Z) = \frac{1}{2}\mathbb{1} \tag{A.8}$$

   for any input.

A.3 Dense coding

Assume Alice and Bob share free entangled Bell states. Using the fact that Alice can transform the shared Bell state into any other Bell state by acting on her own part, we can set up a protocol that transmit two classical bits α and β by sending only one quantum bit.
α and β can be encoded by

$$|\phi_{\alpha\beta}\rangle = (Z^\alpha X^\beta \otimes \mathbb{1})\,|\phi^+\rangle\,. \tag{A.9}$$

If Alice sends her own qubit to Bob, Bob can measure in the Bell state basis to get α and β.
The protocol is called *dense coding*, or sometimes *super dense coding*. Together with teleportation, we can show that both protocols are optimal given that shared entanglement is free.