Summer of Science

Quantum Computing & Information - P02

Shaik Awez Mehtab

Roll No: 23b1080

Mentor: Kanishk Modi

Midterm Report

IIT Bombay

July 16, 2024

Abstract

In this first half of the project, I learned the basics of quantum computing and information followed by the study of quantum noise and operations, the book [2] has helped me a lot in this process and I've done most of my study from this book. Starting with linear algebra, I grasped new "quantum mechanical" notation followed by several theorems and definitions relevant to quantum mechanics like commutators, anticommutators, tensor products and polar and singular value decompositions etc. I then built a basic foundation of quantum mechanics by learning its postulates along with density operators. Applications like superdense coding helped in gaining more insight. This was followed by an overview of quantum information and computing by studying about qubits, quantum gates, quantum circuits, the no-cloning theorem, quantum teleportation etc. I studied few quantum algorithms like Deutsch's algorithm, Deutsch-Jozsa algorithm followed by Stern-Gerlach experiment which provided an experimental overview of quantum information processing. I then studied about quantum noise and quantum operations where I first got to properly understand what "noise" means, atleast in classical sense.I then studied different approaches to understand quantum operations, most importantly operator-sum representation. This was followed by understanding few examples of quantum noise and operations using Bloch sphere, generated by bit-flip, phase-flip, bit-phase-flip, depolarizing channels. At the end, I've covered amplitude and phase damping. I wrote this report simultaneouly while I was studying, please forgive any typos.

Revised PoA

Timeline

- **Week 5:** Distance measures for quantum information, Entropy and Information.
- **Week 6:** Quantum Circuits.
- **Week 7:** The quantum Fourier transform and its applications.
- **Week 8:** Quantum Search Algorithms.

Resources

[1] MIT Open Learning Library. `https://openlearninglibrary.mit.edu/courses/course-v1:MITx+8.370.1x+1T2018/about` . Quantum Information Science I, Part 1.

[2] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2011.

Contents

Part I

Week 1

Chapter 1

Linear Algebra

Usual Linear algebra notation is changed to **Dirac Notation**.

We consider vector spaces in $\mathbb{C}^n$, with additive & multiplicative properties on $\mathbb{C}^n$ too, i.e $\mathbb{C}^n$ is *n* dimensional. Here, the scalars are **complex**.

1.1 Dirac Notation

A vector ψ is denoted by $|\psi\rangle$[1] .$|\psi\rangle$ is referred to as ket, it's vector dual $\langle\psi|$ is referred to as bra. $|\psi\rangle$ is nothing but an n-tuple of complex numbers, $(\psi_1, \cdots, \psi_n)$ or a column vector $\begin{bmatrix} \psi_1 \\ \vdots \\ \psi_n \end{bmatrix}$

few more dirac notations:

Notation	Description
z^*	Complex conjugate of z, $z \in \mathbb{C}$. eg.$(1+i)^* = 1-i$
$\langle\phi\|\psi\rangle$	Inner product (Dot product) of $\|\phi\rangle$ and $\|\psi\rangle$
$\|\psi\rangle \otimes \|\phi\rangle$ or $\|\psi\rangle\langle\phi\|$	Outer product of $\|\psi\rangle$ and $\|\phi\rangle$
$A^\dagger$	Hermitian conjugate or Adjoint of A, = $(A^T)^*$
$\langle\psi\|A\|\phi\rangle$	Inner product between ψ and $A\|\phi\rangle$ or, Inner product between $\|\phi\rangle$ and $A^\dagger\|\psi\rangle$

Table 1.1: Useful *Dirac* notations.

Revise spanning set, span, linear independence, basis, dimension. They're similar to the case of real scalars.

I'll write down few important definitions from linear algebra (and few facts too):

1.2 Linear Operators

A function A, from vector space V to W, which obey

$$A(\sum_i a_i |v_i\rangle) = \sum_i a_i A(|v_i\rangle)$$

[1] $|0\rangle$ isn't the zero vector, zero vector is denoted by just 0

in simpler words, operator on a linear combination of inputs is same linear combination of operator on each input

When based of inputs & output are specified, using coordinate vectors, A linear operator is equivalent to matrix multiplication.

1.2.1 Pauli Matrices

These are 4 fundamentally important matrices in quantum information & computation.

$$\sigma_0 = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \sigma_1 = \sigma_x = X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$
$$\sigma_2 = \sigma_y = Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \sigma_3 = \sigma_z = Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

1.3 Inner Products

This is a bit different from it's real counterpart, due to the reason below.

$$\langle v|w\rangle = (|v\rangle, |w\rangle) = \sum_i v_i^* w_i = \begin{bmatrix} v_1^* \cdots v_n^* \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

the conjugate comes from the fact that this should denoted length squared (for equal vectors). the formal reason is:

Inner product between two vectors $|v\rangle$ and $|w\rangle \in V$ is a function $(.,.) : VV \to \mathbb{C}$ which satisfies:

1. $(|v\rangle, \sum_i \lambda_i |w_i\rangle) = \sum_i \lambda_i (|v\rangle, |w_i\rangle)$
2. $(|v\rangle, |w\rangle) = (|w\rangle, |v\rangle)^*$
3. $(|v\rangle, |v\rangle) \geq 0$, 0 only when $|v\rangle = 0$

$$(\sum_i \lambda_i w_i, v) = \sum_i \lambda_i^* (w_i, v) \tag{1.1}$$

Any vector space that has a defined inner product is called an **inner product space** or **Hilbert space** (if finite dimension) in QIC.

Revise orthogonality, norm, orthonormal set, gram-schmidt process. they're same as in real case. We'll consider only orthonormal based for linear operators, and assume same spaces have same bases.

1.3.1 Dual of $|v\rangle$

it would've been evident from dot product that

$$\langle v| = (|v\rangle^T)^*$$

i.e if $|v\rangle = (v_1, \cdots, v_n)^T$, then

$$\langle v| = \begin{bmatrix} v_1^* \cdots v_n^* \end{bmatrix}$$

1.4 Outer Product

outer product of $|v\rangle \in V$ & $|w\rangle \in W$ is $|v\rangle \langle w|$

it can be seen that this is a matrix, hence it's a linear operator from W to V. let $|w'\rangle \in W$,

$$\underbrace{|v\rangle \langle w|}_{\text{operator}} |w'\rangle = |v\rangle \underbrace{\langle w|w'\rangle}_{\text{scalar}}$$

1.4.1 Completeness Relation

Continuing from yesterday, there's this not so trivial property using outer product, called **completeness relation**

$$\sum_i |i\rangle \langle i| = I$$

Here, $|i\rangle$'s form an orthonormal basis, they can be anything. one nice use of this is, if we have vector spaces V & W, which have bases $|v_1\rangle, \cdots, |v_m\rangle$ and $|w_1\rangle, \cdots, |w_n\rangle$ then any linear transform $A : V \to W$ which is of size nm can be written as a linear combination of

$$|w_1\rangle \langle v_1|, \cdots, |w_2\rangle \langle v_1|, \cdots, |w_n\rangle \langle v_m|$$

each of which are nm. The linear combination is given by

$$A = \sum_{ij} \underbrace{\langle w_j|A|v_i\rangle}_{\text{scalar}} \underbrace{|w_j\rangle \langle v_i|}_{\text{one of the basis matrix}}$$

eg. I can write the pauli matrix σ_0 as

$$1 * |0\rangle \langle 0| + 0 * |0\rangle \langle 1| + 0 * |1\rangle \langle 0| + 1 * |1\rangle \langle 1|$$

1.4.2 Cauchy Schwartz Inequality

For any two vectors $|v\rangle, |w\rangle$ of same dimension, the following inequality is always true

$$\langle v|v\rangle \langle w|w\rangle \geq \langle v|w\rangle \langle w|v\rangle = |\langle v|w\rangle|^2$$

here's a short proof I've written. $|0\rangle, \cdots, |n\rangle$ are orthonormal based I've considered/can be constructed by gram schmidt process.

1.5 Eigenvectors & Eigenvalues

it's the same as you know, just a slight change in notation. we represent both eigenvalue and eigenvector label as v.

$$\underbrace{A}_{\text{linear operator}} |v\rangle = \underbrace{v}_{\text{scalar}} |v\rangle$$

and the characteristic function is c$(\lambda) =$ det$(A - \lambda I)$. it depends only on the operator, not on how it's matrix is represented.

Eigenspace of an eigenvalue v is the set of vectors which have eigenvalue v. read diagonalisability, orthonormal decomposition, degenerate eigenvectors.

1.6 Special Operators

1.6.1 Adjoints & Hermitian operators

for any linear operator A on a vector space $V A^\dagger$, there exists a unique linear operator $A^\dagger$ such that, $\forall \left|v\right\rangle, \left|w\right\rangle \in V$

$$(\left|v\right\rangle, A\left|w\right\rangle) = (A^\dagger \left|v\right\rangle, \left|w\right\rangle)$$

I've deduced that $A^\dagger = (A^T)^*$ here, when A is represented as a matrix. This is actually same as dual of a vector. Few properties:

1. $(AB)^\dagger = B^\dagger A^\dagger$
2. $\left|v\right\rangle^\dagger = \left\langle v\right|$

Thus you can see $(A\left|v\right\rangle)^\dagger = \left\langle v\right| A^\dagger$, hence the above thing is evident.
It is **anti-linear.**

$$\left(\sum_i a_i A_i\right)^\dagger = \sum_i a_i^* A_i^\dagger$$

A ***Hermitian*** or a ***self-adjoint operator*** is a linear operator such that $A^\dagger = A$.

1.6.2 Projectors

these are nice. suppose W is a k-dimensional subspace of V, which is d-dimensional, we can always construct, using *gram-schmidt* process, an orthonormal basis $\left|1\right\rangle, \cdots, \left|d\right\rangle$ of V such that $\left|1\right\rangle, \cdots, \left|k\right\rangle$ is an orthonormal basis of W. Now we can project any vector in V onto W, using this operator:

$$P = \sum_{i=1}^{k} \left|i\right\rangle \left\langle i\right|$$

cross check this. This operator P is called the **projector**. It is also Hermitian. and we'll use P to refer to V. $Q = I - P$ is the **orthogonal complement** of P.

1.6.3 Normal Operators

An operator A is normal if $A^\dagger A = AA^\dagger$. all hermitian operators are normal.

Theorem 1 (Spectral Decomposition Theorem).

$$M \textit{ is a normal operator on } V \iff M \textit{ is diagonalisable}$$

diagonalisable w.r.t some basis in V. The proof is interesting to read, do check it out.

1.7 Tensor Products

Tensor product is a way of putting vector spaces together to form larger vector spaces. For example, if we have vector spaces V and W of dimension *m* & *n* respectively, the tensor product $V \otimes W$ represents an *mn* dimensional vector space.

Suppose $|v\rangle \in V$ & $|w\rangle \in W$ be two vectors, their tensor product $|v\rangle \otimes |w\rangle$ is also represented as $|v\rangle |w\rangle$, $|v, w\rangle$ or even $|vw\rangle$, and it lies in *mn* dimensional space $V \otimes W$

They have few nice properties:

1. For a scalar z,
$$z(|v\rangle \otimes |w\rangle) = (z\,|v\rangle) \otimes |w\rangle \tag{1.2}$$

2. For $|v_1\rangle , |v_2\rangle \in V$
$$(|v_1\rangle + |v_2\rangle) \otimes |w\rangle = |v_1\rangle \otimes |w\rangle + |v_2\rangle \otimes |w\rangle \tag{1.3}$$

3. For $|w_1\rangle , |w_2\rangle \in W$
$$|v\rangle \otimes (|w_1\rangle + |w_2\rangle) = |v\rangle \otimes |w_1\rangle + |v\rangle \otimes |w_2\rangle \tag{1.4}$$

4. If we have operators $A : V \to V$ & $B : W \to W$, we can define a new operator $A \otimes B : V \otimes W \to V \otimes W$, we can't write it in matrix form, but we know it's a linear operator and,
$$(A \otimes B)(|v\rangle \otimes |w\rangle) \equiv A\,|v\rangle \otimes B\,|w\rangle \tag{1.5}$$

5. We can define inner product in $V \otimes W$ as
$$(a_i\,|v_i\rangle \otimes |w_i\rangle , b_j\,|v_j\rangle \otimes |w_j\rangle) = a_i^* b_j \langle v_i|v_j\rangle \langle w_i|w_j\rangle \tag{1.6}$$
As we have a well defined inner product, this is a hilbert space too, we can use properties like adjoint, unitary, normal, hermitian matrices.

1.7.1 Kronecker Product

There is indeed a representation which would let us interpret tensor product in some manner, suppose we have two matrices $A_{m\times n}$, $B_{p\times q}$, their tensor product in this representation is:

$$A \otimes B = \begin{bmatrix} A_{11}B & A_{12}B & \cdots & A_{1n}B \\ A_{21}B & A_{22}B & \cdots & A_{2n}B \\ \vdots & \vdots & \vdots & \vdots \\ A_{m1}B & A_{m2}B & \cdots & A_{mn}B \end{bmatrix}_{mp\times nq} \tag{1.7}$$

for example, tensor product of Pauli matrices X & Y is represented as

$$X \otimes Y = \begin{bmatrix} 0\cdot Y & 1\cdot Y \\ 1\cdot Y & 0\cdot Y \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{bmatrix} \tag{1.8}$$

and another example,

$$\begin{bmatrix}1\\2\end{bmatrix}\otimes\begin{bmatrix}2\\3\end{bmatrix}=\begin{bmatrix}2\\3\\4\\6\end{bmatrix} \tag{1.9}$$

Using similar notations, $|\phi\rangle^{\otimes k}$ means tensor product of $|\phi\rangle$ with itself k times.

Just for practice, I'm writing down few tensor products, I, X, Y, Z are pauli matrices.

$$X\otimes Z=\begin{bmatrix}0&0&1&0\\0&0&0&-1\\1&0&0&0\\0&-1&0&0\end{bmatrix} \qquad I\otimes X=\begin{bmatrix}X&0\\0&X\end{bmatrix}=\begin{bmatrix}0&1&0&0\\1&0&0&0\\0&0&0&1\\0&0&1&0\end{bmatrix} \tag{1.10}$$

$$X\otimes I=\begin{bmatrix}0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\end{bmatrix} \tag{1.11}$$

Hence, tensor product **isn't commutative**. Few more facts,

1. $(A\otimes B)^* = A^*\otimes B^*$
2. $(A\otimes B)^T = A^T\otimes B^T$
3. $(A\otimes B)^\dagger = A^\dagger\otimes B^\dagger$
4. Tensor product preserves (a) Unitarity (b) Hermiticity (c) Positivity (d) Projection

1.8 Operator functions

If we have a function $f:\mathbb{C}^n\to\mathbb{C}^n$, we can define a corresponding matrix function, which acts on *normal matrices* as follows

$$A=\sum_a a\,|a\rangle\langle a| \implies f(A)\equiv\sum_a f(a)\,|a\rangle\langle a| \tag{1.12}$$

where $\sum_a a\,|a\rangle\langle a|$ is a spectral decomposition (exists since A is normal 1). Now we can calculate square root, logarithm, exponential of a positive, positive-definite, normal operator respectively. for eg.

$$\exp(\theta Z)=\begin{bmatrix}e^\theta&0\\0&e^{-\theta}\end{bmatrix} \tag{1.13}$$

I'll do another example, let $A=\begin{bmatrix}4&3\\3&4\end{bmatrix}$, let's find square root of it. Eigenvalues of A are 7, 1. Their corresponding eigenvectors which form an orthonormal basis of $\mathbb{R}^2$ are $(|0\rangle+|1\rangle)/\sqrt{2}=|a\rangle$ & $(|0\rangle-|1\rangle)/\sqrt{2}=|b\rangle$,

$$A=7\,|a\rangle\langle a|+|b\rangle\langle b| \tag{1.14}$$

$$\implies \sqrt{A}=\sqrt{7}\,|a\rangle\langle a|+|b\rangle\langle b| \tag{1.15}$$

$$=\begin{bmatrix}\frac{\sqrt{7}+1}{2}&\frac{\sqrt{7}-1}{2}\\\frac{\sqrt{7}-1}{2}&\frac{\sqrt{7}+1}{2}\end{bmatrix} \tag{1.16}$$

1.8.1 Trace

This is just the sum of diagonal elements of a matrix.

$$\text{tr}(A) = \sum_i A_{ii} \tag{1.17}$$

It's also defined as

$$\text{tr}(A) = \sum_i \langle i|A|i\rangle \tag{1.18}$$

i.e it doesn't depend on which coordinate I represent A on, traces have these nice properties:

1. It's *cyclic*, $\text{tr}(AB) = \text{tr}(BA)$
2. $\text{tr}(cA + dB) = c\text{tr}(A) + d\text{tr}(B), c, d \in \mathbb{C}$
3. if U is unitary, $\text{tr}(A) = \text{tr}(UAU^\dagger)$
4. $\text{tr}(A\,|\psi\rangle\,\langle\psi|) = \langle\psi|A|\psi\rangle$

1.8.2 Hilbert-Schmidt inner product

The vector space L_V of linear operators on a Hilbert space V can be converted into a Hilbert space by defining an inner product on $L_V \times L_V$ as

$$(A, B) \equiv \text{tr}(A^\dagger B) \tag{1.19}$$

this is known as the *Hilbert-Schmidt* or *Trace* inner product.

1.9 Commutator & Anti-commutator

The commutator between two operators A & B is defined as

$$[A, B] = AB - BA \tag{1.20}$$

if $[A, B] = 0 \implies AB = BA$, then we say A *commutes* with B. The *anti-commutator* between A & B is defined as

$$A, B = AB + BA \tag{1.21}$$

we say A *anti-commutes* with B if A, B = 0.

Theorem 2 (Simultaneous diagonalization theorem). *If A & B are Hermitian operators, then $[A, B] = 0 \implies \exists$ an orthonormal basis such that both A and B are simultaneously diagonalizable w.r.t same basis.*

Commutations & Anti-commutation for Pauli matrices

$$[X, Y] = 2iZ \qquad [Y, Z] = 2iX \qquad [Z, X] = 2iY \tag{1.22}$$

$$\tag{1.23}$$

we can write it using ϵ_{jkl}, the antisymmetric tensor, for which $\epsilon_{jkl} = 0$ except for $\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1$ and $\epsilon_{321} = \epsilon_{213} = \epsilon_{132} = -1$:

$$[\sigma_j, \sigma_k] = 2i\sum_{i=1}^{3} \epsilon_{jkl}\sigma_l \tag{1.24}$$

Also, $\{\sigma_i, \sigma_j\} = 0, \forall i \neq j, i.j \in 1, 2, 3$. whereas $\{\sigma_0, \sigma_j\} = 2\sigma_j$.

1.10 Polar and Singular value decompositions

Theorem 3 (**Polar decomposition**). *Let A be a linear operator on a vector space V. Then there exist unitary U and **positive operators** J and JK such that*

$$A = UJ = KU \tag{1.25}$$

where the unique positive operators J & K are $J \equiv \sqrt{A^\dagger A}$ and $K \equiv \sqrt{AA^\dagger}$, if A is invertible, U is unique.

$A = UJ$ is the left polar decomposition of A, and $A = KU$ is the right polar decomposition of A

Singular value decomposition is just polar decomposition & spectral decomposition.

Corollary 3.1 (**Singular value decomposition**). *Let A is a square matrix, Then $\exists$ unitary U & V, a diagonal matrix D with **non-negative entries** such that*

$$A = UDV \tag{1.26}$$

*diagonal elements of D are called **singular values** of A.*

Proof. By polar decomposition, A = SJ, where S is unitary and J is positive. By spectral decomposition, $J = TDT^\dagger$ where T is unitary and D has non negative entries. setting $U = ST$ and $V = T^\dagger$ we get the above thing. ☐

Chapter 2

Postulates of Quantum Mechanics

2.1 State Space

Postulate 2.1. *Any isolated physical system has a Hilbert Space associated with it known as the* ***state space****. The system is completely described by it's* ***state vector****, which is a unit vector in state space.*

The simplest quantum mechanical system, which we'll be concerned the most about, is a *qubit* which has a 2-dimensional state space. Hence an arbitrary state vector $|\psi\rangle$ can be represented by

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|\psi\rangle \tag{2.1}$$

where $\alpha, \beta \in \mathbb{C}$ and $|0\rangle\,, |1\rangle$ form an orthonormal basis of state space. Since state vector is a unit vector, $\langle\psi|\psi\rangle = 1 \implies |\alpha|^2 + |\beta|^2 = 1$, this is known as *normalization* of state vector.

For now, *qubit* is an abstract thing. We consider a fixed orthonormal basis $|0\rangle\,, |1\rangle$ apriori. These two states can be considered analogous to the bits 0 and 1, except that a state vector is a linear combination, or in other words, *superposition* of these bits. The linear combination $\sum_i \alpha_i\,|\psi_i\rangle$ is defined as the superposition of $|\psi_i\rangle$ with amplitude α_i for $|\psi_i\rangle$.

2.2 Evolution

Postulate 2.2. *The evolution of a* ***closed*** *system is described by a* ***unitary transformation****. The state of the system $|\psi\rangle$ at time t_1 is related to the state of the system $|\psi^{'}\rangle$ at time t_2 solely by the unitary transformation U which only depends on t_1, t_2*

$$|\psi^{'}\rangle = U\,|\psi\rangle \tag{2.2}$$

Quantum mechanics doesn't help us find the *state space* or *unitary operator*, it just assures us that any physical system would behave this way.

Few such operator on a qubit are pauli matrices. X is referred to as *NOT gate* or *bit-flip* since it turns $|0\rangle$ to $|1\rangle$ and vice versa. Z is known as *phase-flip* as it leaves $|0\rangle$ as it is, which inverting the phase (sign) of $|1\rangle$.

The second postulate just defines state of the system at discrete intervals t_1 and t_2, this can be generalised for *continuous* t.

Corollary 3.2. *Time evolution of a **closed** quantum system is described by **Schrödinger equation**,*

$$i\hbar\frac{d\,|\psi\rangle}{dt} = H\,|\psi\rangle \tag{2.3}$$

*where $\hbar$ is the reduced planck's constant, H is a fixed **Hermitian** operator known as the **Hamiltonian** of the closed system.*

Since Hamiltonian is a Hermitian operator, it has a spectral decomposition,

$$H = \sum_E E\,|E\rangle\,\langle E| \tag{2.4}$$

Here the eigenvectors $|E\rangle$ are conventionally called *energy eigenstates* or *stationary states* and E is called *energy* of the state $|E\rangle$. The lowest energy is known as *ground state energy* and the corresponding state is the *ground state.*

We can see a connection between corollary 2.2 and postulate 3.2, if we know the solution of Schrödinger's equation, which can be verified that it is

$$|\psi(t_2)\rangle = exp\left[\frac{-iH(t_2-t_1)}{\hbar}\right]|\psi(t_1)\rangle = U(t_1,t_2)\,|\psi(t_1)\rangle \tag{2.5}$$

if we set,

$$U(t_1,t_2) \equiv exp\left[\frac{-iH(t_2-t_1)}{\hbar}\right] \tag{2.6}$$

it can be shown that if K is a Hermitian operator, then $U = exp(iK)$ is a unitary operator. Thus, there's a one-to-one correspondence between the theorem and it's corollary.

If the system we're considering is not closed, we can approximate it's evolution using a time varying Hamiltonian. There's this example that when we consider a laser focused on an atom, the whole system is properly described by a Hamiltonian, but the behaviour of the atome alone, is seem to be approximately described by another Hamiltonian, this thing contains terms related to the laser intensity.

2.3 Quantum measurement

We've seen that in closed systems, the evolution of state vector is according to a unitary transform, but when we try to measure the state of the system, we need to interact with it, which leads to a change other than a unitary transform, which is described as

Postulate 2.3. *A **quantum measurement** is a collection $\{M_m\}$ of measurement operators. These operators act on the state space. The index m refers to the measurement outcome that we want using M_m. If the system is in state vector $|\psi\rangle$ before the measurement had been made, the probability that outcome is m is,*

$$p(m) = \langle\psi|\,M_m^\dagger M_m\,|\psi\rangle \tag{2.7}$$

After the measurement, the state changes to

$$\frac{M_m\,|\psi\rangle}{\sqrt{\langle\psi|\,M_m^\dagger M_m\,|\psi\rangle}} \tag{2.8}$$

All the M_i satisfy the ***completeness relation***

$$\sum_i M_i^\dagger M_i = I \tag{2.9}$$

This ensures the probability that some measurement occurs is 1,

$$\sum_i p(i) = \sum_i \langle\psi| M_i^\dagger M_i |\psi\rangle = 1 \tag{2.10}$$

As an example, suppose we have a closed system whose state vector $|\psi\rangle$ is currently $a|0\rangle + b|1\rangle$ and we want to make measurements using $M_0 = |0\rangle\langle 0|$ and $M_1 = |1\rangle\langle 1|$, we can calculate and see that probability that measurements 0 & 1 are made correct are $|a|$ and $|b|$ respectively, and the final states are $\frac{a}{|a|}|0\rangle$ if we make measurement 0 and $\frac{b}{|b|}|1\rangle$ if we make measurement 1.

So when there's an isolated physical system, it has a state vector, when you try to measure it, using the collection of measurement operators gives just a label. we can just get the probability that we get a specific probability, using equation 2.7

2.4 Distinguishing quantum states

When we have n systems, with n state vectors, we can only distinguish them by measuring them only if all the state vectors are orthonormal. This has a good proof.

Theorem 4. *No measurement distinguishing non-orthogonal states is possible*

Proof. We consider two states $|\psi_1\rangle, |\psi_2\rangle$ and assume that there exists a measurement. We distinguish these states by looking at the output label we get from the measurement. Taking the function $f(j)$ where j is the label. If the state is $|\psi_1\rangle$ then probability that our prediction is correct $\sum_{j;f(j)=1} \langle\psi_1| M_j^\dagger M_j |\psi_1\rangle = 1$ similarly for 2. To simlify this we define

$$E_i = \sum_{j;f(j)=i} M_j^\dagger M_j \tag{2.11}$$

so from what we've done above

$$\langle\psi_1| E_1 |\psi_1\rangle = 1 \, \langle\psi_2| E_2 |\psi_2\rangle = 1 \tag{2.12}$$

since we'll always measure $|\psi_1\rangle$'s label when the state is $|\psi_1\rangle$,

$$\langle\psi_1| E_2 |\psi_2\rangle = 0 \tag{2.13}$$

$$\implies \sqrt{E_2}\,|\psi_2\rangle = 0 \tag{2.14}$$

now let's focus on the orthogonal component of $|\psi_2\rangle$, by writing it as $|\psi_2\rangle = \alpha|\psi_1\rangle = \beta|\phi\rangle$, where $|\phi\rangle$ is orthonormal to $|\psi_1\rangle \implies |\beta| < 1$, since they're non-orthogonal.

But there's a contradiction,

$$\langle\psi_2| E_2 |\psi_2\rangle = \beta^2 \langle\phi| E_2 |\phi\rangle \tag{2.15}$$

$$\leq |\beta|^2 \tag{2.16}$$

$$< 1 \tag{2.17}$$

Thus by contradiction, our assumption is false $\implies$ we can't distinguish non-orthogonal states reliably. □

2.5 Projective measurements

This is the main thing we'll deal in quantum information and computation.

Postulate 2.4. *A projective measurement is an observable M, a Hermitian operator on the state space of system considered, having spectral decomposition*

$$M = \sum_m mP_m \tag{2.18}$$

where m is one of the possible outcomes. m is an eigenvalue of M, P_m is the projector onto the corresponding eigenspace. When we measure a state $|\psi\rangle$, the probability of getting the result m is

$$p(m) = \langle\psi| P_m |\psi\rangle \tag{2.19}$$

After measuring, the state of the system becomes

$$\frac{P_m |\psi\rangle}{\sqrt{p(m)}} \tag{2.20}$$

We can see projective measurement is a special kind of measurement, which do obey completeness relation, i.e $\sum_m M_m^\dagger M_m = I$. But M_m also are *Orthogonal projectors*, hence *Hermitian* and $M_m M_{m^‘} = \delta_{mm^‘} M_m$

These have few nice properties, we can find mean of measurements easily, which is

$$\begin{aligned} E(M) &= \sum_m mp(m) && (2.21)\\ &= \sum_m m \langle\psi| P_m |\psi\rangle && (2.22)\\ &= \langle\psi| \sum_m mP_m |\psi\rangle && (2.23)\\ &= \langle\psi| M |\psi\rangle && (2.24) \end{aligned}$$

this is a nice thing, we can also find the standard deviation of the outcomes we get.

$$\begin{aligned} \Delta(M) &= \sqrt{\langle(M - \langle M\rangle\rangle^2} && (2.25)\\ &= \sqrt{\langle M^2\rangle - \langle M\rangle^2} && (2.26) \end{aligned}$$

As you might have observed, we defined a new term *observable*, some still like to look at the projection measurement as a collection of operators, using the observable implicitly. Another phrase used is "measure in basis $|m\rangle$", which means the projectors $P_m = |m\rangle \langle m|$ are used for projective measurement.

If $\vec{v}$ be a three dimensional unit vector, then the measurement of the observable

$$\vec{v} \cdot \vec{\sigma} = v_1\sigma_1 + v_2\sigma_2 + v_3\sigma_3 \tag{2.27}$$

is historically known as the "*measurement of spin along $\vec{v}$ axis*"

Remark. Eigenvalues of $\vec{v} \cdot \vec{\sigma}$ are ± 1 and the corresponding projection operators are $P_\pm = \frac{I \pm \vec{v} \cdot \vec{\sigma}}{2}$

2.5.1 Heisenberg Uncertainty Principle

Suppose A, B be two Hermitian operators and $|\psi\rangle$ is a quantum state. Let $\langle\psi| AB |\psi\rangle = x + iy$ where $x, y \in \mathbb{R}$. Then $\langle\psi| [AB] |\psi\rangle = 2iy$ and $\langle\psi| \{AB\} |\psi\rangle = 2x$, therefore

$$(\langle\psi| [A, B] |\psi\rangle)^2 + (\langle\psi| \{A, B\} |\psi\rangle)^2 = 4 |\langle\psi| AB |\psi\rangle|^2 \tag{2.28}$$

Using Cauchy-Schwartz inequality, then substituting $A = C - \langle C\rangle$ and $B = D - \langle D\rangle$

$$\langle\psi| A^2 |\psi\rangle \langle\psi| B^2 |\psi\rangle \geq (\langle\psi| AB |\psi\rangle)^2 \tag{2.29}$$

$$\geq \frac{(\langle\psi| [A, B] |\psi\rangle)^2}{4} \tag{2.30}$$

$$\implies \Delta(C)\Delta(D) \geq \frac{\langle\psi| [C, D] |\psi\rangle}{2} \tag{2.31}$$

Hence Proved.

Note that the common misconception of Heisenberg uncertainity principle is that we "measure a state with some error" which is $\Delta(C), \Delta(D)$, but the actual thing is we have a large amount of $|\psi\rangle$'s and we measure each of them seperately to with observables C, D to get a set of values of C, D whose standard deviation is $\Delta(C), \Delta(D)$. This is because once we make a measurement, the quantum state changes because we've altered it.

2.6 POVM measurements

It stands for *Positive Operator-Valued Measure*. POVM is a formalism, and is used when we are concerned with the outcome more than the state of the output we're measuring. For eg. when we've done an experiment and just want to measure the final state.

If we're measuring a state $|\psi\rangle$ using measurement operators M_m, the probability that the output is m is $\langle\psi| M_m^\dagger M_m |\psi\rangle$, if we define

$$E_m = M_m^\dagger M_m \tag{2.32}$$

E_m satisfy $\sum_m E_m = I$, E_m is enough to find the probability of output m which is $\langle\psi| E_m |\psi\rangle$. The set $\{E_m\}$ is called the *POVM set*.

In projective measurements POVM set is same as the set of measurement operators, since $E_m = P_m^\dagger P_m = P_m$. In fact any measurement where the measurement operators and the POVM elements coincide is a projective measurement.

If we want the post-measurement state, we can conveniently *define* a POVM to be any set of operators $\{E_m\}$ such that (a) each E_m is positive (b) it satisfies *completeness relation*, i.e $\sum_m E_m = I$ In this way we can choose $M_m = U_m\sqrt{E_m}$ where U_m is a unitary operator.

2.7 Phase

This is kind of interesting and deja vu giving.

2.7.1 Global Phase

If $|\psi\rangle$ is a state, $\theta \in \mathbb{R}$, we say that $e^{i\theta} |\psi\rangle$ is equal to $|\psi\rangle$ upto the *global phase factor* $e^{i\theta}$. These two act quite similar when you try to measure them, because the output of m would occur with a probability of $\langle\psi| M_m^\dagger M_m |\psi\rangle$ and $\langle\psi| e^{-i\theta} M_m^\dagger M_m e^{i\theta} |\psi\rangle = \langle\psi| M_m^\dagger M_m |\psi\rangle$, which is the same. Thus, they're physically equivalent.

2.7.2 Relative Phase

If two amplitudes, a and b, *differ by a relative phase* if there is a real θ such that $a = be^{i\theta}$. Two states *differ by a relative phase* in *some basis* if each of the amplitudes in that basis is related by such a phase factor. Unlike global phase factors, relative phase factors differ from amplitude to amplitude, thus, there are physically observable differences in measurement statistics, and two states differing by a relative phase aren't physically equivalent.

2.8 Composite systems

It's useful in describing state of a composite system made up of multiple systems

Postulate 2.5. *The state space of a composite physical system is the tensor product of state spaces of component physical systems. If n physical systems are prepared in state* $|\psi_1\rangle, \cdots, |\psi_n\rangle$ *The composite system has the state* $|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$

Using this, we can prove a fact.

Remark. To implement a general measurement, it is sufficient to make projective measurements together with unitary dynamics

I'm not giving the entire proof. but assume that the state space we're considering is Q and we want to measure using operators M_m. We can consider a new *ancilla* system which has state space M. M has orthonormal basis $|m\rangle$ which are in one-to-one correspondence with the first output. Let's define an operator $U : QM \to QM$ as

$$U|\psi\rangle|0\rangle = \sum_m M_m|\psi\rangle|m\rangle \tag{2.33}$$

where $|0\rangle$ is some state in Q. It can be shown that U preserves dot product. i.e $\langle 0|\langle\phi|U^\dagger U|\psi\rangle|0\rangle = \langle\phi|\psi\rangle$ which implies that U extends over $Q \otimes M$ which can be proved too.

Now, if we consider two systems with projectors $P_m = I_Q \otimes |m\rangle\langle m|$ and measure using unitary dynamics and projective transformation, the probability of output m is

$$\begin{aligned} p(m) &= \langle\psi|\langle 0|U^\dagger P_m U|0\rangle|\psi\rangle & (2.34) \\ &= \sum_{m',m''} \langle\psi|\langle m'|M_{m'}^\dagger(I_Q \otimes |m\rangle\langle m|)M_{m''}|m''\rangle|\psi\rangle & (2.35) \\ &= \langle\psi|M_m^\dagger M_m|\psi\rangle & (2.36) \end{aligned}$$

which is just as measured directly from general measurement. also the state of the composite system after the measurement is

$$\frac{P_m U|\psi\rangle|0\rangle}{\sqrt{\langle\psi|U^\dagger P_m U|\psi\rangle}} = \frac{M_m|\psi\rangle|m\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}} \tag{2.37}$$

Since the state of the system M after measurement is $|m\rangle$, state of system Q is

$$\frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}} \tag{2.38}$$

also as measured directly from general measurement.

Quantum Entanglement

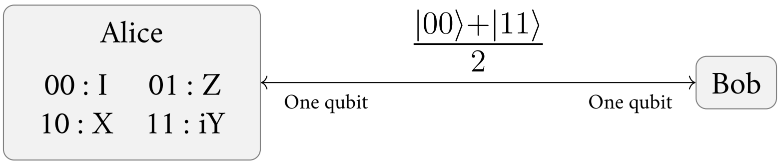
This is something you would've heard of before. Consider the two qubit state,

$$|\psi\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}} \tag{2.39}$$

we can never write it as a tensor product of two other qubit states, i.e $|\psi\rangle = |a\rangle |b\rangle$ can never be true. It's easily checkable. A state of a composite system like this, which can't be written as tensor product of it's component systems is an *entangled* state. Entangled states play a cruicial rules in quantum computation and information.

2.9 Application: Superdense Coding

This nicely uses the idea of entanglement and "applying" quantum gates. Suppose we have 'Alice' and 'Bob' again, who're away from each other. Alice wants two share two classical bits of information using one qubit to bob, can she do it?



Yes, using superdense coding. Here, Alice and Bob share a pair of externally setup qubits in entangled state

$$|\psi\rangle = \frac{|00\rangle + |11\rangle}{2} \tag{2.40}$$

Alice has one qubit, which she can alter and Bob has another. Any change Alice makes to her qubit changes the composite system, which can be measured by Bob. So she applies quantum gates to change the qubits, as given in the diagram, which result in the following changes.

$$00 : |\psi\rangle \xrightarrow{I} \frac{|00\rangle + |11\rangle}{2} \tag{2.41}$$

$$01 : |\psi\rangle \xrightarrow{Z} \frac{|00\rangle - |11\rangle}{2} \tag{2.42}$$

$$10 : |\psi\rangle \xrightarrow{X} \frac{|10\rangle + |01\rangle}{2} \tag{2.43}$$

$$11 : |\psi\rangle \xrightarrow{iY} \frac{-|10\rangle + |01\rangle}{2} \tag{2.44}$$

We can see that all the outputs are orthonormal, hence Bob can choose appropriate measurement operators and clearly distinguish each state, hence getting the information Alice had sent him. These are known as the *Bell states, Bell basis* or *EPR pairs*.

Thus, a single qubit can be used to transfer information which would've taken two classical bits.

2.10 The density operator

Till now we've been using the language of state vectors. The density operator is a different way of representing states, it's mathematically equivalent to state vectors. It's much more convenient to use in some scenarios. It's really useful while describing *individual subsystems* of a composite system.

2.10.1 Ensembles of quantum states

It's useful when we don't exactly know the state of a system. If the possible states are $|\psi_i\rangle$ with probabilities p_i, $\{p_i, |\psi_i\rangle\}$ is an *ensemble of pure states*. The density operator, also called as the *density matrix* is defined as

$$\rho \equiv \sum_i p_i |\psi_i\rangle \langle\psi_i| \tag{2.45}$$

We can now workout rewriting the postulates of quantum mechanics.

Suppose, the evolution of closed quantum system is described by a unitary operator U. The system is initially in the state $|\psi_i\rangle$ with probability p_i, after the evolution, system will be in the state $U|\psi_i\rangle$ with probability p_i, then the density operator evolves into

$$\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i| \xrightarrow{U} \sum_i p_i U |\psi_i\rangle \langle\psi_i| U^\dagger = U\rho U^\dagger. \tag{2.46}$$

We can also predict the probability of a measurement outcome m, when our system is defined by the density operator ρ, if the initial state was $|\psi_i\rangle$, then the probabililty of outcome m is

$$p(m|i) = \langle\psi_i|M_m^\dagger M_m|\psi_i\rangle = \text{tr}(M_m^\dagger M_m |\psi_i\rangle \langle\psi_i|) \tag{2.47}$$

This is the probability that the outcome would be m, if the state was $|\psi_i\rangle$ which has a probability of p_i, hence the overall probability of outcome being m is

$$p(m) = \sum_i p(m|i) p_i \tag{2.48}$$

$$= \sum_i p_i \text{tr}(M_m^\dagger M_m |\psi_i\rangle \langle\psi_i|) \tag{2.49}$$

$$= \text{tr}(M_m^\dagger M_m \rho) \tag{2.50}$$

Also we can find out the state of the system if the outcome was m, if the state was initially $|\psi_i\rangle$ which has probability p_i the final state would be

$$|\psi_i^m\rangle = \frac{M_m |\psi_i\rangle}{\sqrt{\langle\psi_i|M_m^\dagger M_m|\psi_i\rangle}} \tag{2.51}$$

Hence, the final density operator would be

$$\rho_m = \sum_i p(i|m) |\psi_i^m\rangle \langle\psi_i^m| = \sum_i p(i|m) \frac{M_m |\psi_i\rangle \langle\psi_i| M_m^\dagger}{\langle\psi_i|M_m^\dagger M_m|\psi_i\rangle} \tag{2.52}$$

By Bayes' theorem, $p(i|m) = p(m|i)p_i/p(m)$, subsituting this,

$$\rho_m = \sum_i p_i \frac{M_m |\psi_i\rangle \langle\psi_i| M_m^\dagger}{\text{tr}(M_m^\dagger M_m \rho)} \tag{2.53}$$

$$= \frac{M_m \rho M_m^\dagger}{\text{tr}(M_m^\dagger M_m \rho)} \tag{2.54}$$

A system whose state $|\psi\rangle$ is exactly known is said to be in *pure state*. The density operator is then simply $|\psi\rangle\langle\psi|$. Otherwise, the system is said to be in a *mixed state*. It's said to be in a *mixture* of different states in the ensemble of ρ. Suppose a quantum state is prepared from ρ_i with probability p_i, the system can then be described by the density operator $\sum_i p_i \rho_i$, which can be proved.

This is useful in mainly in the analysis of quantum noise, when we need to introduce ignorance into our knowledge of the quantum state. For example if we made a measurement and somehow lost/forgot our measurement, then the final state of the system would be *mixed*, given by

$$\rho = \sum_m p(m) \rho_m \tag{2.55}$$

$$= \sum_m \text{tr}(M_m^\dagger M_m \rho) \frac{M_m \rho M_m^\dagger}{\text{tr}(M_m^\dagger M_m \rho)} \tag{2.56}$$

$$= \sum_m M_m \rho M_m^\dagger \tag{2.57}$$

2.11 General properties of density operator

We can actually characterize what operators are density operators using

Theorem 5 (Characterization of density operators). *An operator ρ is a density operator associated with an ensemble $\{p_i, |\psi_i\rangle\}$ if and only if*

1. ***Trace condition*** *tr$(\rho) = 1$*
2. ***Positivity condition*** *ρ is a positive operator.*

Proof. Forward is easy and upto the reader. For the backward proof, a possible ensemble is $\{\lambda_i, |i\rangle\}$, where $\lambda_i, |i\rangle$ are eigenvalues are corresponding eigenvectors of ρ, it's upto the reader to check other conditions. □

With this in our hands, we can redefine our postulates with this density operator picture

Postulate 1. *Every physical system is associated with a Hilbert space known as it's state space. The system is completely described by it's density operator, ρ which is a positive operator with trace equal to one acting on the state space of the system. If the system is in state ρ_i with probability p_i, the density operator of the system is $\sum_i p_i \rho_i$.*

Postulate 2. *The evolution of a closed system is described by a unitary transformation, i.e state of the system ρ at time t_1 is related to the state ρ' at time t_2 by a unitary operator which only depends on t_1 and t_2.*

$$\rho' = U \rho U^\dagger \tag{2.58}$$

Postulate 3. *A quantum measurement is described by a set $\{M_m\}$ of measurement operators. These act on the state space of the system. The index m refers to the possible outcome of the measurement. If the state of the system is ρ just before the measurement, the probability of outcome m is*

$$p(m) = tr(M_m \rho M_m^\dagger) \tag{2.59}$$

The state of the system changes after the measurement is made, to ρ', given by

$$\rho' = \frac{M_m \rho M_m^\dagger}{tr(M_m \rho M_m^\dagger)} \tag{2.60}$$

The measurement operators satisfy the relation

$$\sum_m M_m M_m^\dagger = I. \tag{2.61}$$

Postulate 4. *State space of a composite system is the tensor product of it's component systems. If the components are prepared in state $\rho_1, \rho_2, \cdots, \rho_n$, state of the composite system is given by $\rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n$.*

Remark. If ρ is a density operator, then $\mathrm{tr}(\rho^2) \leq 1$. Equality holds when ρ is a pure state.

By now, you would've noticed that two *different* states can have the same density operator. To get to know what kind of connection there is between these states, we define $|\tilde{\psi_i}\rangle$ associated with the state $|\psi_i\rangle$, the latter is said to *generate* the density operator if $\rho = \sum_i |\tilde{\psi_i}\rangle \langle\tilde{\psi_i}|$. It's observable that $|\tilde{\psi_i}\rangle = \sqrt{p_i}\,|\psi_i\rangle$. Now we proceed for the theorem.

Theorem 6 (Unitary freedom in the ensemble of the density matrices.)**.** *Two sets $|\tilde{\psi_i}\rangle$ and $|\tilde{\phi_j}\rangle$ generate the same density operator if*

$$|\tilde{\psi_i}\rangle = \sum_j u_{ij} |\tilde{\phi_j}\rangle \tag{2.62}$$

where u_{ij} is a unitary matrix. The sets which is smaller is padded with 0.

In other words, if they two systems have possible states $|\psi_i\rangle$ and $|\phi_j\rangle$ with probabilities p_i and q_j respectively, they both have the same density operator if and only if

$$\sqrt{p_i}\,|\psi_i\rangle = \sum_j u_{ij} \sqrt{q_j}\,|\phi_j\rangle \tag{2.63}$$

where u_{ij} is a unitary matrix, and the smaller set is padded with 0.

2.12 Reduced Density Operator

This is most useful in describing *subsystems* of a composite quantum system. Suppose we have physical system A and B, whose state is described by the density operator ρ^{AB}. The reduced density operator for system A is defined by

$$\rho^A \equiv \mathrm{tr}_B(\rho^{AB}), \tag{2.64}$$

where tr_B is a map of operators known as the *partial trace* over system B. It's defined by

$$\mathrm{tr}_B(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|) \equiv |a_1\rangle\langle a_2|\,\mathrm{tr}(|b_1\rangle\langle b_2|), \tag{2.65}$$

This partial trace operation is defined only on a special subclass of operators on AB; the specification is completed by requiring in addition to Equation (2.65) that the partial trace be linear in its input.

Let's do an example to understand this better, suppose $\rho^{AB} = \rho \otimes \sigma$, where ρ is the density operator for system A, and σ for system B. Then

$$\rho^A = \mathrm{tr}_B(\rho \otimes \sigma) = \rho\mathrm{tr}(\sigma) = \rho \tag{2.66}$$

Which is intuitively correct. For an unintuitive example, let's take an entangled state which would further reveal many things. Let's take the bell state $AB\ \frac{|00\rangle+|11\rangle}{2}$, which is an entangled state and has the density operator

$$\rho^{AB} = \left(\frac{|00\rangle + |11\rangle}{2}\right)\left(\frac{\langle 00| + \langle 11|}{2}\right) \tag{2.67}$$

$$= \frac{|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|}{2} \tag{2.68}$$

Thus we can get ρ^B from this

$$\rho^B = \mathrm{tr}_A(\rho^{AB}) \tag{2.69}$$

$$= \frac{\mathrm{tr}(|0\rangle\langle 0|)\,|0\rangle\langle 0| + \mathrm{tr}(|1\rangle\langle 0|)\,|1\rangle\langle 0|}{2} \tag{2.70}$$

$$+ \frac{\mathrm{tr}(|0\rangle\langle 1|)\,|0\rangle\langle 1| + \mathrm{tr}(|1\rangle\langle 1|)\,|1\rangle\langle 1|}{2} \tag{2.71}$$

$$= \frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2} \tag{2.72}$$

$$= \frac{I}{2}. \tag{2.73}$$

This is actually a *mixed state*, since $\mathrm{tr}((I/2)^2) < 1$. It's wonderful how we can't find state vector of an entangled qubit, but we can find out the density operator of an entangled qubit. Also the composite system is in pure state, but the entangled qubit is in a mixed state, we don't know exactly what it is. This is a really strange property, we can know the exact state of a joint system, but still we can't figure out each subsystem's state exactly. A hallmark of quantum entanglement.

A physical justification for making such an identification is that the reduced density operator provides the correct measurement statistics for the measurements made on system A.

2.13 Schmidt decomposition and purifications

2.13.1 Schmidt decomposition

This is a tool like density operator which are useful for the study of composite quantum systems.

Theorem 7 (**Schmidt decomposition**). *Suppose $|\psi\rangle$ is a pure state of a composite system, AB. Then there exist orthonormal state $|i_A\rangle$ for system A, and orthonormal states $|i_B\rangle$ of system B such that*

$$|\psi\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle\,, \tag{2.74}$$

*where λ_i are **non-negative real** numbers satisfying $\sum_i \lambda_i^2 = 1$ known as the **Schmidt coefficients**.*

Do note that it's not the linear combination of all possible tensor product of bases of A and B. It's just i, i.e linear. Only the tensor product of corresponding bases. This result is very useful. Let $|\psi\rangle$ be a pure state of a composite system, AB. Then by Schmidt decomposition, $\rho^A = \sum_i \lambda_i^2 |i_A\rangle \langle i_A|$ and $\rho^B = \sum_i \lambda_i^2 |i_B\rangle \langle i_B|$. Thus eigenvalues of ρ^A and ρ^B are identical, namely λ_i^2 for both density operators.

Proof. We'll consider when A and B have same dimension, it's also true when they're not of same dimension. Let $|j\rangle$ and $|k\rangle$ be fixed o.n.b for A and B respectively, then $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_{jk} a_{jk} |j\rangle |k\rangle , \tag{2.75}$$

where a_{jk} form a matrix a. By singular value decomposition, $a = udv$, where d is a diagonal matrix with non-negative elements, u and v are unitary matrices. Thus

$$|\psi\rangle = \sum_{ijk} u_{ji} d_{ii} v_{ik} |j\rangle |k\rangle , \tag{2.76}$$

If we define, $|i_A\rangle \equiv \sum_j u_{ji} |j\rangle$, $|i_B\rangle \equiv \sum_k v_i k |k\rangle$, and $\lambda_i \equiv d_{ii}$, it gives

$$|\psi\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle \tag{2.77}$$

$|i_A\rangle$ and $|i_B\rangle$ form an orthonormal set, from the unitarity of u and v and the orthonormality of $|j\rangle$ and $|k\rangle$. $\square$

Remark. If $|ABC\rangle$ is a three component quantum system. There are quantum states $|\psi\rangle$ of the system which can't be written as

$$|\psi\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle |i_C\rangle , \tag{2.78}$$

where $\lambda_i \in \mathbb{R}$ and $|i_A\rangle , |i_B\rangle , |i_C\rangle$ are the orthonormal bases of the respective systems.

The bases $|i_A\rangle$ and $|i_B\rangle$ are known as the *Schmidt bases* of systems A and B. The number of non-zero values λ_i is known as the *Schmidt number* of state $|\psi\rangle$. It in some sense relates to the "amount" of entanglement in $|\psi\rangle$. Schmidt number is preserved under unitary transformation, i.e Schmidt number of $U|\psi\rangle =$ Schmidt number of $|\psi\rangle$, since $U|\psi\rangle = \sum_i \lambda_i (U |i_A\rangle) |i_B\rangle$.

2.13.2 Purification

Suppose we're given a state ρ^A of a quantum system A, which might be *mixed* or *pure*. We can surely introduce a new system R and define a *pure state* $|AR\rangle$ for AR such that $A = \mathrm{tr}_R(|AR\rangle \langle AR|)$. Here R is known as the *reference system* has no physical significance, we just introduced it to *purify* our mixed state, i.e associate our mixed state with a pure state. This thing is known as *purification* of ρ^A.

To prove this, suppose ρ^A has the orthonormal decomposition $\sum_i p_i \left|i^A\right\rangle \left\langle i^A\right|$. We now introduce a system R which has the same state space as A, and has an orthonormal basis $\left|i^R\right\rangle$, to define a pure state of the combined system AR,

$$\left|AR\right\rangle \equiv \sum_i \sqrt{p_i} \left|i_A\right\rangle \left|i_R\right\rangle \tag{2.79}$$

Now the reduced density operator of system A, corresponding to the state $\left|AR\right\rangle$ is

$$\begin{aligned}
\mathrm{tr}_R(\left|AR\right\rangle \left\langle AR\right|) &= \sum_{ij} \sqrt{p_i p_j} \left|i^A\right\rangle \left\langle j_A\right| \mathrm{tr}(\left|i^R\right\rangle \left\langle j^R\right|) &(2.80)\\
&= \sum_{ij} \sqrt{p_i p_j} \left|i^A\right\rangle \left\langle j_A\right| \delta_{ij} &(2.81)\\
&= \sum_i p_i \left|i^A\right\rangle \left\langle i^A\right| &(2.82)\\
&= \rho^A &(2.83)
\end{aligned}$$

Thus, $\left|AR\right\rangle$ is the purification of ρ^A.

Part II

Week 2

Chapter 3

Overview of Quantum Computing

3.1 Global Perspectives

We'll first look into how QIC originated, how different fields like quantum physics, computer science, information theory, cryptography have contributed in it's development. No one would've thought we could process information and do computing using quantum mechanical systems few decades ago.

3.1.1 History of QIC

It all began in the turn of twentieth century, when the "then physics", now known as classical physics resulted in absurdities. Like ultraviolet catastrophe, involving infinite energies and the well known electron spiralling into nucleus in around 10^{-8} seconds. Small changes were made everytime, until it resulted in the modern theory of *quantum mechanics* after like a quarter of century. Now it's used in a wide range of fields from structure of DNA to semiconductors to study of nuclear fusion in stars.

Quantum Mechanics is a mathematical framework or a set of rules which base our physical theories. For example *quantum electrodynamics* is a fantastic theory built within the framework of quantum mechanics which describes the interaction between atoms and light. It contains it's own set of rules. The rules of quantum mechanics are simple, but even experts found them counter intuitive, one of the main aim of QIC is to build tools which sharpen our intuition of quantum mechanics.

For example, the problem of signalling faster than light, impossible according to Einstein. In quantum mechanics, it converted into the problem whether we *clone* (make a copy of) a quantum state. Even though it is possible in classical mechanics, the *no-cloning* theorem was proved in quantun mechanics. Few imperfect cloning devices were developed which helped us understand it more.

A historical thing which led to the development of QIC is problem of obtaining *complete control over single quantum systems*. The way we used quantum mechanics then had control over a large composite of quantum systems. For example, we could explain superconductivity incredibly well, but we had control over a huge bunch of atoms (quantum systems) to explain that. Slowly, few methods were developed, like isolating an atom by "trapping" it, electron tunnelling microscope which had control over single atoms etc. The reason we tried gaining control over a single quantum system was on a hunch. Most profound insights in science come when we try developing methods to explore new regimes of nature. Famous examples are radio astronomy, low temperature physics etc. In a similar way, there is a hope of discovering new phenomena if we have control over a single quantum system. QIC fits naturally in this problem. A deep understanding and the ability to harness the powers of QIC is essential to gain control over a single quantum system. Despite these efforts, quantum information processing systems have seen modest success, what we can do now is a dozen of operations on few qubits, few real world applications of *quantum cryptography*. It's a challenge now to develop computational methods to take it to a large scale.

Another big field, a great intellectual triumph, which contributed to QIC is *computer science* . It's history roots back to very old times. But the proper modern incarnation of computer science was announced by *Alan Turing*. *Turing* developed an abstract notion of what we now call as a programmable computer, the *Turing machine*. He also developed the notion of *Universal Turing machine* which could simulate any other turing machine. Another pioneer of computer science, *Alonzo Church*, developed *Church-Turing* thesis, which says if an algorithm can be done on any physical device, an equivalent algorithm is present for a Turing machine, which does the same task. He developed rigorous mathematicalental concepts which are equivalent to what class of algorithms that can be performed by a physical device.

Slowly electrical components like transistors started to develop and programmable computers were made. This growth was codified by *Moore's Law*, which states computer power would double for constant cost roughly every two years. But this law would eventually reach it's limits, as the components would shrink to maximum physical limit. Quantum mechanics would start interfering in the functioning of electrical devices. One possible solution to the failure of Moore's law is to move to a different computing paradigm, which is provided by QIC. A classical computer could simulate a quantum computer, but it's impossible to do it *efficiently*, since quantum computers provided way faster speed. Many reasearchers believe that no *conceivable* amount of development of classical computers would bridge the gap between the powers of a classical computer and powers of a quantum computer.

There are algorithms which are *efficient*, i.e in *computational complexity* terms, they can be solved in polynomial time, others are *inefficient algorithms*. An observation was made which was codified into *strengthened Church-Turing* thesis:-

Any algorithmic process can be simulated effectively using a turing machine.

But when *analog* computers were developed, theoretically they could solve few problems efficiently which couldn't be efficiently solved by a Turing machine, violating the aforementioned strengthened church-turing thesis. But realistically , considering the noises these couldn't solve the problems efficiently which turing machine couldn't solve efficiently. This consideration of noise was challenge in the development of QIC, which led to theories of *quantum error-correcting codes* and *fault-tolerant quantum computation*. Thus, quantum computers can tolerate a finite amount of noise while having the computational edge.

A major challenge faced by the Church-Turing thesis is when Robert Solovay and Volker Strassen showed that it is possible to find whether an integer is prime or not using a *randomized algorithm*. This could just determine whether a number is *probably* prime or composite with a *certainty*. At that time, no deterministic primality algorithm was known, neither is now. Hence this suggested that there are efficiently solvable problems which can't be solved efficiently using a deterministic Turing machine. Hence, the strengthened Church-Turing thesis was changed to

*Any algorithmic problem can be stimulated efficiently using a **probabilistic** turing machine.*

This kind of *ad hoc* modification meant that in future there could be some other computational model found which could solve problems efficiently which couldn't be solved efficiently by a Turing machine. Motivated to find a computational model which could simulate any other model of computation, In 1985, David Deutsch tried to look for physical theories which could be a foundation for Church-Turinh thesis which would be as secure as the physical theory, it should simulate any *arbitrary* physical system. Since all laws of physics are ultimately quantum mechanical laws, he considered devices based on principles of quantum mechanics, which were the quantum mechanical analogues of machines defined by Turing, forty-nine years earlier, which now we're considering as a quantum computer.

We still don't surely know whether Deutsch's notion of Universal quantum computer could solve any probleme efficiently. Maybe some mew strange effect from esoteric theories like quantum field theory, string theory, quantum gravity etc. might open a new dimension of computational models, which could solve problems not efficiently solvable by Deutsch's model. Still Deutsch gave a challenge to Church-Turing thesis. Slowly for few problems like finding prime factors of an integer and 'discrete logarithm' problem, it was shown that quantum computing is way superior to classical computing, as it provided an efficient solution while the latter couldn't. It's not possible to simulate quantum mechanical systems on a classical computer, while it is possible on a quantum computer, a problem with profound scientific and technological implications solved.

Coming up with good quantum algorithms which a quantum computer would solve more quickly than a classical computer is difficult. A pessimist might think it's because quantum computers can solve only the problems which have been discovered yet. But the actual reasons for why designing algorithms for quantum computers are (a) Human intuition is rooted in the classical world, when we use that intuition to construct algorithms, they'll be based on classical idea, whic wouldn't use true quantum effects to achieve a good quantum algorithm. Thus we have to turn off our classical intuition while designing quantum algorithms. (b) It's not enough to merely design a quantum mechanical algorithm, we have to make it *better* than any exisiting classical algorithm! for which we have to make use of true quantum aspects of quantum mechanics, which isn't a widespread interest because classical algorithms with comparable performace characteristics exist. This makes designing quantum algorithms a challenge!

Let's now change our focus to another big contributor to quantum computation and quantum information, information theory. In 1948, Claude Shannon laid the foundation of modern information and communication theory through his papers. The key step taken by Shannon was *to mathematically define the concept of information.* There are *different ways* to define fundamental things like this, which might have their widespread use. But the definition by Shannon is by far the most fruitful in terms of better understanding, a plethora of deep results and a theory with a rich structure which seems to reflect many (not all) real-world communication problems.

Shannon was interested in two key questions which he answered by proving the two fundamental theorems of information theory. Them being (a) what resources are required to send information reliably over a communication channel? Shannon's first theorem, *noiseless channel coding theorem*, quantifies the physical resources needed to store the output from an information source. (b) can information be transmitted while being protected against the noises in the communication channel? Shannon's second theorem, *noisy channel coding theorem*, quantifies how much information is possible to reliably sent over through a noisy channel. To achieve reliable transmission in presence of noise, Shannon showed that *error-correcting codes* could be used to protect the information. He gave the upper limit of protection it can give, but doesn't give the limiting codes. Many error-correcting codes are being researched and contructed these days, which are reliable enough to be used for computer modems, satellite systems etc. Do not that this is all classical physics.

Quantum information theory followed similar developmenet. First, Ben Schumacher provided an analogue to noiseless channel coding theorem, and defined a 'quantum bit' or 'qubit' as a tangible physical resource. However, no anallogue of noisy channel coding theorem is known till now. However their counterparts have been developed, as theory of quantum error correction, allowing quantum computers to effectively compute in presence of noise, also allowing transfer of data over *noisy quantum* channels effectively. Classical ideas of error-correction have been important in understanding quantum error-correction codes. Slowly, an important class of codes, CSS codes were developed. These were then subsumed *stabilizer codes*, these were felicitated by a deep understanding of classical linear coding theory.

Theory of quantum correcting codes was developed to protect quantum states, against noise, but if we tried to transmit *classical* information, few surprises were found. Like two classical bits of information being transferred by a single qubit, known as *superdense coding*. Another example, *distributed quantum computation*, where we consider two communicating quantum computers, in certain problems, these would take *exponentially lesser time* compared to classical computers connected. But these problems are of little practical application and have technical restrictions, which is a challenge in front of us.

There's also a *networked information theory* which deals with communications with network of many channels. There's not much known about the networked information theory, but the theory is very valuable and gaining insights into it would open up a lot of new things. For example, we have Alice and Bob, they want to send information through a noisy quantum channel which has zero capacity for quantum information. No information is possible to be communicated between them. But suppose we have two such channels operating synchronously, they still have zero capacity. But if we reverse one of the channels, then according to quantum mechanics, this network has non-zero capacity and we can transfer information between them.

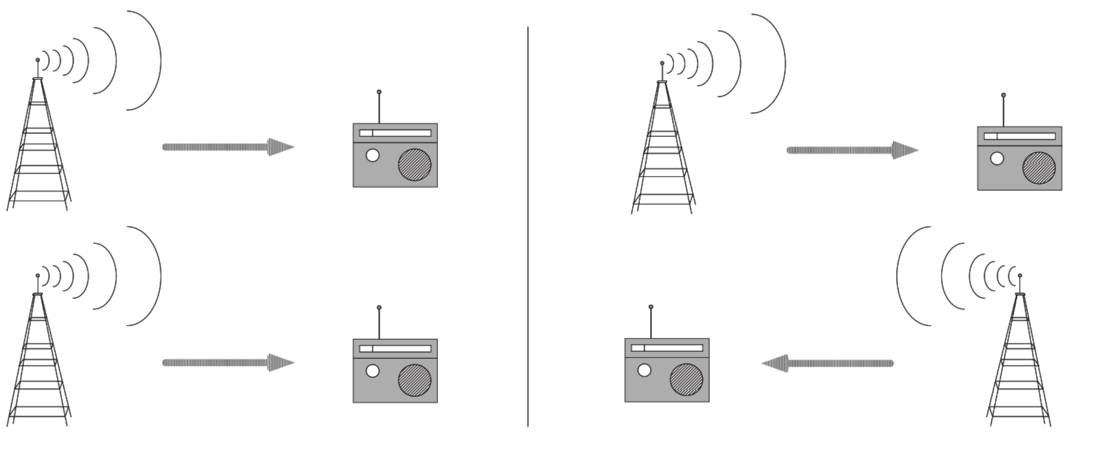


Figure 3.1: Reversing one of the channel makes the channel have a non-zero capacity

Let's finally move to the respected old art and science of *cryptography*. It's broadly described as the *communication* between two parties *who may not trust each other*. An important case for cryptography is when one party wants to send a secret message to another, without anyone eavesdropping. This is done through a *cryptographic protocol*. Two main ways this can be done is through (a) *public key cryptosystems* (b) *private key ecosystems*.

In *private key cryptosystems*, both party share a private key initially, using which they send *encrypted* messages to each other. Hence only these two can *decrypt* the message, no one else. But the main problem of this is, how are the private keys distributed? what if a third person eavesdrops while they're being distributed? He can get to know every secret message between them. An early discovery of QIC solves this, using *quantum cryptography* or *quantum key distribution*. The basic idea used from quantum mechanics is that if an eavesdropper tries to measure the private key when Alice and Bob are distributing it, they can see that the state of communication channel has changed. Hence they get to know someone's out there and restart it.

In *public key cryptosystems*, each party makes it's own 'public key', which is made *available to the genreal public*. Using this key of Bob, Alice can encrypt the message. For a third person, it's *extremely difficult* to decrypt this. Not even Alice can decrypt this in real time. Only Bob, who has the keys to his public key personally, can decrypt it. This is most popular these days. The encryption is done through *RSA cryptosystem*. The key here is, it should be extremely decrypt the encryption only with the public key. For this, RSA uses something closely similar to factoring. But Shor's algorithm, run on a quantum computer can to this effectively. Few use discrete logarithm problem, which can also be done effectively by Shor's algorithm. This implies a need of quantum computing and quantum information.

The most striking of these in the study of *quantum entanglement*. It is a unique quantum mechanical *resource* which plays a key role in many applications of QIC. It's like iron to the classical world's bronze age. There's no proper theory of quantum entanglement as of now, a lot of effort is being made. Further study of quantum entanglement is going to lead to amazing new discoveries in quantum computation and quantum information.

3.2 Quantum Bits

A *bit* is a fundamental concept in classical computation and classical information. An analogous thing in QIC is a *quantum bit* or *qubit*. We'll treat it as a mathematical object but it also has correspondence with real world objects. Like a classical bit has 0, 1 as it's states, a qubit's analogous states are $|0\rangle$ and $|1\rangle$. In general a qubit can have the general state which is a *linear combination* or *superposition* of $|0\rangle$ and $|1\rangle$ as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle\,,\ \alpha, \beta \in \mathbb{C} \tag{3.1}$$

$|0\rangle\,, |1\rangle$ are known as *computational basis states*. Unlike a classical bit, we can't determine state of a qubit. When we try observing/measuring it, we get $|0\rangle$ with probability $|\alpha|^2$ and $|1\rangle$ with probability $|\beta|^2$. Thus $|\alpha|^2 + |\beta|^2 = 1$. Hence we can think of the state as unit vector in a complex vector space with the computational bases. This unobservablity of a qubit makes up the heart of QIC.

A state $|\psi\rangle$ of a qubit can be represented as,

$$|\psi\rangle = e^{i\gamma}\left(\cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle\right) \tag{3.2}$$

which equal to this by a global factor $e^{i\gamma}$,

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle \tag{3.3}$$

θ, φ define a point on a sphere known as the *Bloch sphere*; this thing describes many operations of qubits properly. But this doesn't generalize when it comes to multiple qubits.

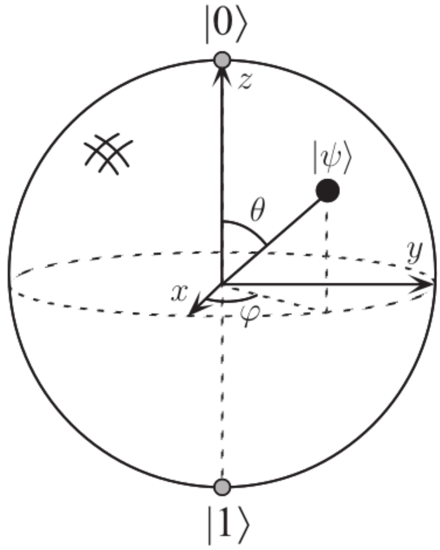


Figure 3.2: Bloch sphere representation of a qubit.

3.2.1 Multiple qubits

State vector $|\psi\rangle$ of a system of two qubits can be defined as,

$$|\psi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle \tag{3.4}$$

It can be seen that it has four *computational bases,* $|00\rangle, |01\rangle, |10\rangle, |11\rangle$. Even here when we measure the qubit, we get $x = (00, 01, 10, 11)$ with probability $|\alpha_x|^2$ and the qubit collapses to the state $|x\rangle$. By *normalization,* $\sum_{x\in\{0,1\}^2} |\alpha_x^2| = 1$, where $\{0, 1\}^2$ is set of all strings with each character 0 or 1. For this set of two qubits, if we try to measure a subset of this, i.e a qubit, say first qubit, the probability of getting the output 0 is $|\alpha_{00}|^2 + |\alpha_{01}|^2$ and after measuring, the system collapses to

$$|\psi'\rangle = \frac{\alpha_{00}|00\rangle + \alpha_{01}|01\rangle}{\sqrt{|\alpha_{00}|^2 + |\alpha_{01}|^2}} \tag{3.5}$$

This is a similar collapse *re-normalized* with a normalization constant.

Another interesting thing can be seen if we consider the Bell state,

$$|\psi\rangle = \frac{|00\rangle + |11\rangle}{2} \tag{3.6}$$

If we try to measure $|0\rangle$, we get the output 0 with probability $\frac{1}{2}$ and the bell state collapses to $|00\rangle$, and the output 1 with probability $\frac{1}{2}$ with the bell state collapsing to $|11\rangle$. We can see that if we get 0 we perfectly know about second qubit, i.e measurements are *correlated*. Also, a set of n systems can store 2^n amount of information, even though when we measure it, it collapses.

3.3 Quantum Computation

Similar to an electrical circuit, there's a *quantum circuit* which performs operations on qubits using *quantum gates*.

3.3.1 Single qubit gates

We can think of a NOT gate similar to it's classical analogue, which converts $|0\rangle$ to $|1\rangle$ and vice-versa. It's also linear, i.e NOT gate when acted on the state

$$\alpha|0\rangle + \beta|1\rangle \tag{3.7}$$

it interchanges $|0\rangle$ and $|1\rangle$ to give

$$\alpha|1\rangle + \beta|0\rangle \tag{3.8}$$

This fact is very important, which can be realised later. NOT gate can be represented by the matrix

$$X \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{3.9}$$

For a matrix U to be a *quantum gate*, it has to be unitary and this is the *only* constraint.

Few interesting gates are Z, which leaves $|0\rangle$ unchanged while flips the sign of $|1\rangle$,

$$Z \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{3.10}$$

The output is equal upto a relative phase factor, the other is the *Hadamard gate*, H given by

$$H \equiv \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{3.11}$$

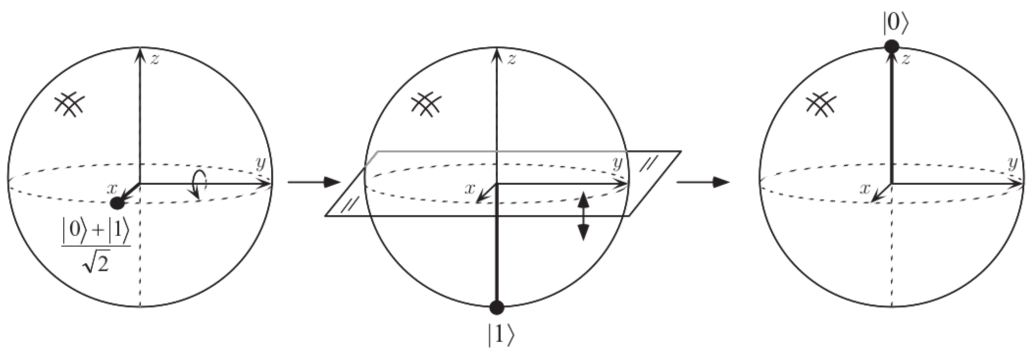


Figure 3.3: Hadmard gate acting on $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$ to give $|0\rangle$

This get takes $|0\rangle$ to $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$ and $|1\rangle$ to $\frac{|0\rangle-|1\rangle}{2}$. In the Bloch sphere point of view, this rotates the sphere first w.r.t $\hat{y}$ axis by $90°$ and then rotates w.r.t $\hat{x}$ axis by $180°$.

These qubit gates are summarised as

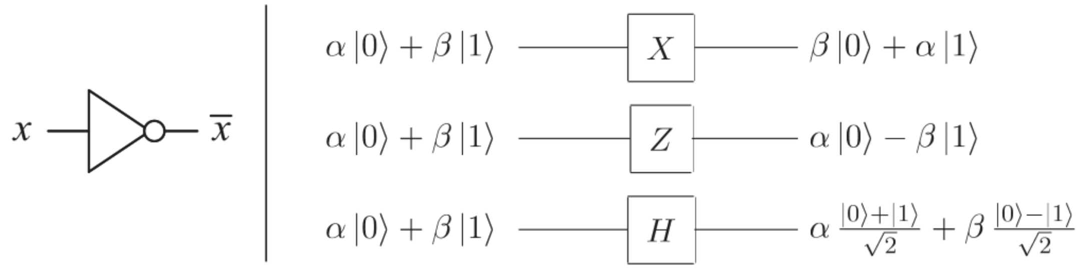


Figure 3.4: Single bit (left) and Qubit (right) logic gates.

We can see that any unitary matrix corresponds to a single qubit gate. Since, there are infinitely many unitary matrices, there are infinitely many single qubit gates. But we can represent this set only with a few single qubit gates. It can be shown that any unitary matrix U can be written as

$$U = e^{i\alpha}\begin{bmatrix} e^{-i\beta/2} & 0 \\ 0 & e^{i\beta/2} \end{bmatrix}\begin{bmatrix} \cos\frac{\gamma}{2} & \sin\frac{\gamma}{2} \\ \sin\frac{\gamma}{2} & -\cos\frac{\gamma}{2} \end{bmatrix}\begin{bmatrix} e^{-i\delta/2} & 0 \\ 0 & e^{i\delta/2} \end{bmatrix} \tag{3.12}$$

This can be seen as a rotation about $\hat{z}$ axis then a rotation about $\hat{y}$ axis followed by a rotation about $\hat{z}$ axis again and global phase shift. Any single qubit gate can be decomposed into a form like this. Also, any computation on a *finite* number of qubits can be generated by a *finite* set of gates known as the *universal* for quantum computation.

3.3.2 Multiple qubit gates

For classical bits, there are AND, OR, NAND, NOR, XOR gates. NAND gate is considered as the *universal gate* as any function of bits can be constructed using NAND gates. XOR gate can't do this. There's a special multi-qubit quantum gate, known as *controlled*-NOT gate or CNOT gate, which has a *control qubit* and a *target qubit*. If the control qubit is 0, nothing happens to target qubit, whereas if it's 1, the target qubit is flipped. i.e,

$$|00\rangle \to |00\rangle\,; |01\rangle \to |01\rangle\,; |10\rangle \to |11\rangle\,; |11\rangle \to |10\rangle \tag{3.13}$$

The CNOT gate is similar to the XOR gate, because it's just doing $|A, B\rangle \to |A, B \oplus A\rangle$, where $B \oplus A$ is just the XOR gate, is the sum modulo two. There are no similar generalisations for classical XOR, NAND because they're *irreversible* or *non-invertible*. Information is lost upon the action of these gates which is irretrievable. CNOT and single qubit gates are prototypes for all other gates because of the following *universality* result: *Any multiple qubit gate may be composed from* CNOT *and single qubit gates.*

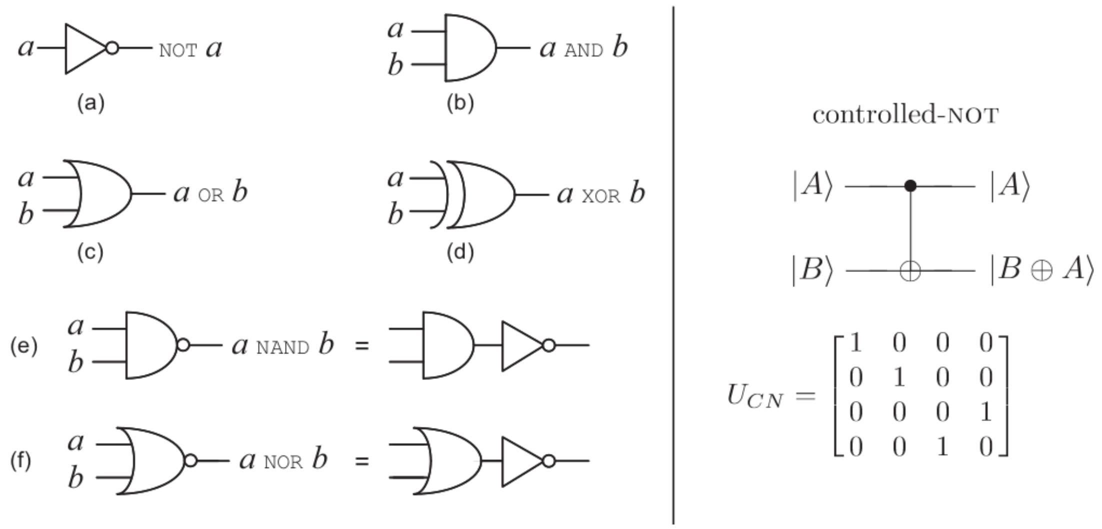


Figure 3.5: Figure on left shows common bit gates; The right figure shows the CNOT gate along with it's matrix representation, in the order of $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$.

3.3.3 Measurement in bases other than the computational basis

If we try to represent state of a system $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ in a different bases, $|a\rangle$ and $|b\rangle$ which are *orthonormal*as $\alpha'|a\rangle + \beta'|b\rangle$. By postulates of quantum mechanics, if we try to measure w.r.t basis $|a\rangle$, $|b\rangle$ we get *a* with probability $|\alpha'|^2$ and *b* with probability $|\beta'|^2$. We'll learn more about how efficient is it to measure w.r.t a basis in further study.

3.3.4 Quantum circuits

Similar to classical circuits, quantum circuits are made of wires, which don't necessarily represent physical wires, rather the passage of time or a physical photon moving in space. Let's first look at an example.

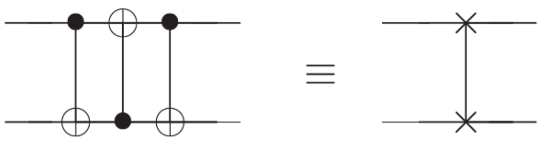


Figure 3.6: Circuit which swaps two qubits and an equivalent circuit since this is more useful and commonly used.

The above circuit swaps two qubits, which can be seen this way considering two qubits with states $|a\rangle$ and $|b\rangle$ as input,

$$
\begin{aligned}
|a, b\rangle &\to |a, b \oplus a\rangle && (3.14)\\
&\to |a \oplus (b \oplus a), b \oplus a\rangle = |b, b \oplus a\rangle && (3.15)\\
&\to |b, (b \oplus a) \oplus b\rangle = |b, a\rangle && (3.16)
\end{aligned}
$$

There are few features in classical circuits which aren't seen in quantum circuits. First one is *loops* are not allowed, i.e the circuit is *acyclic*. Second, classical circuit allows joining of wires, an operation known as FANIN. Here, the bitwise OR operator of inputs is taken. However, information is lost and thus this is irreversible, thus not being allowed in quantum circuits. Thirdly, the inverse of this, where multiple copies of a single bit is made in a classical circuits, known as FANOUT is not allowed in quantum circuits. We can show that it's impossible to copy a qubit.

There's another important gate, which is in some sense generalisation of CNOT gate, *controlled-U* gate. which is represented by this

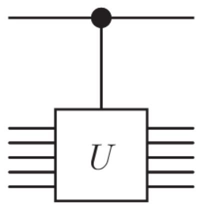


Figure 3.7: Controlled-U gate

This has similar logic, it's associated with a *unitary matrix* U and takes 1 qubit as the *control qubit*, and n qubits as *target qubits* as inputs. If the control qubit is set to 0, then nothing happens to target qubits. If control qubit is set to 1, then the gate U is applied to all the target qubits. In this sense, CNOT gate is the special case when $n = 1$ and $U = X$, as shown

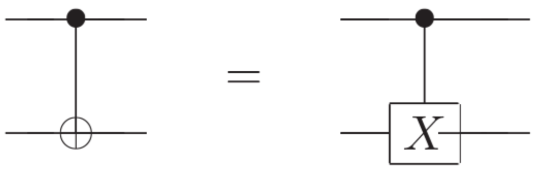


Figure 3.8: Two different representations of CNOT gate.

There's another important operation, known as measurement shown by the 'meter' symbol.

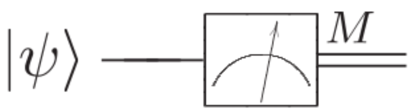


Figure 3.9: Quantum circuit symbol for measurement.

As expected, this outputs a probabillistic classical bit M (which is 0 with probability $|\alpha|^2$ or 1 with probability $|\beta|^2$) if a qubit with state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is given as input. Thus, the output is distinguished from a qubit by drawing a double-line wire.

3.3.5 Qubit copying circuit?

First we'll look at a classical example, suppose we have a bit x which we want to copy, we'll use a CNOT gate with x and a 'scratchpad bit' 0 as first and second inputs respectively, it'll copy x succesfully as shown

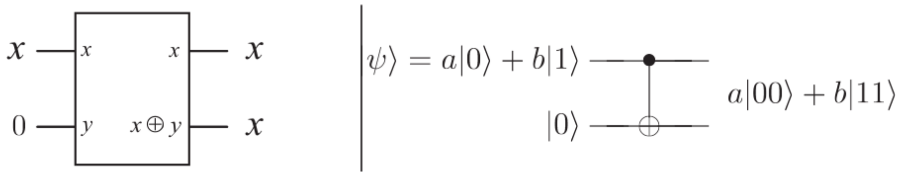


Figure 3.10: Trying to copy a classical bit (left) and a qubit (right)

But if we try copying a qubit with state $|\psi\rangle = a|0\rangle + b|1\rangle$ in a similar manner using a CNOT gate, we can give $|0\rangle$ as an input along with it as

$$[a|0\rangle + b|1\rangle]\,|0\rangle = a|00\rangle + b|10\rangle \tag{3.17}$$

applying the CNOT gate gives $a|00\rangle + b|11\rangle$ as the output. We'd expect the output $|\psi\rangle\,|\psi\rangle$ if we copied $|\psi\rangle$ succesfully. But that would mean the output should be

$$|\psi\rangle\,|\psi\rangle = a^2|00\rangle + ab|01\rangle + ab|10\rangle + b^2|11\rangle \tag{3.18}$$

But as we've seen above, it's only possible when $ab = 0$, i.e $a = 0$ or $b = 0$ i.e we want to copy either $|0\rangle$ or $|1\rangle$. Hence we can't copy $|\psi\rangle$. It can be shown in general that it is impossible to copy a general qubit with unkown quantum state, if we construct a cloning gate, it can copy different state qubits $|a\rangle$ and $|b\rangle$ only if they are orthogonal. This is known as the *no-cloning* theorem.

There's another way to look at it. An unkown qubit contains some 'hidden' information that isn't directly accessible with a measurement. What's happening above is we get $a|00\rangle + b|11\rangle$, which if we measure we obtain either 0 or 1 using which the other qubit is completely determined. Thus, no extra information is gained. But if we were able to copy $|\psi\rangle$, then the other qubit contains some more information, contradicting our knowledge, thus a copy can't be created.

3.3.6 Example: Bell states

Let's take a look at a circuit which has a Hadamard gate on first input followed by a CNOT gate, as shown

In	Out
$\|00\rangle$	$(\|00\rangle + \|11\rangle)/\sqrt{2} \equiv \|\beta_{00}\rangle$
$\|01\rangle$	$(\|01\rangle + \|10\rangle)/\sqrt{2} \equiv \|\beta_{01}\rangle$
$\|10\rangle$	$(\|00\rangle - \|11\rangle)/\sqrt{2} \equiv \|\beta_{10}\rangle$
$\|11\rangle$	$(\|01\rangle - \|10\rangle)/\sqrt{2} \equiv \|\beta_{11}\rangle$

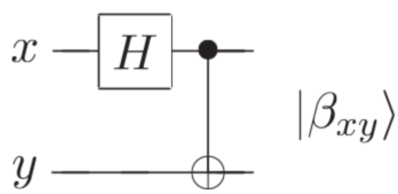


Figure 3.11: Quantum circuit which creates Bell states, along with it's quantum 'truth table'

As shown in the truth table, the Hadmard gate first transforms the first qubit into a superposition after which the CNOT gate inverts the target only when control is 1. The output states are

$$|\beta_{00}\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}} \tag{3.19}$$

$$|\beta_{01}\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}} \tag{3.20}$$

$$|\beta_{10}\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}} \tag{3.21}$$

$$|\beta_{11}\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}} \tag{3.22}$$

These are known as the *Bell states* or *EPR states* or *EPR pairs* as their strange properties were found by Einstein, Podolsky and Rosen. Generalizing it,

$$|\beta_{xy}\rangle = \frac{|0y\rangle + (-1)^x |1\bar{y}\rangle}{\sqrt{2}} \tag{3.23}$$

where $\bar{y}$ is the negation of y, $|\beta_{xy}\rangle$ denotes the output when the input given is $|xy\rangle$.

3.3.7 Example: quantum teleportation

It's a non-trivial, surprising and a fun thing! Quantum teleportation is a technique using which we can move quantum states around even in the absence of a quantum communication channel. Consider Alice and Bob, who've met long ago and have shared an EPR pair. Now Alice has a new qubit in state $|\psi\rangle$ which she doesn't know and has to deliver to Bob. She can only send *classical* information.

Intuitively, it's pretty bad situation. It's not possible for her to get to know the state $|\psi\rangle$ completely as she only has one copy of the qubit. Also if she knew the state, it's impossible to transfer this information to Bob, as $|\psi\rangle$ takes values in *continuous* space, hence requiring infinite amount of time. Quantum teleportation provides a way for it.

In short, what happens is she measures this state $|\psi\rangle$ along with the half of her EPR pair. She gets a result in 00, 01, 10 or 11 which she then sends to Bob. Using this information Bob can perform some operations and completely retrieve the state $|\psi\rangle$ without even getting to know it. Shown using a quantum circuit

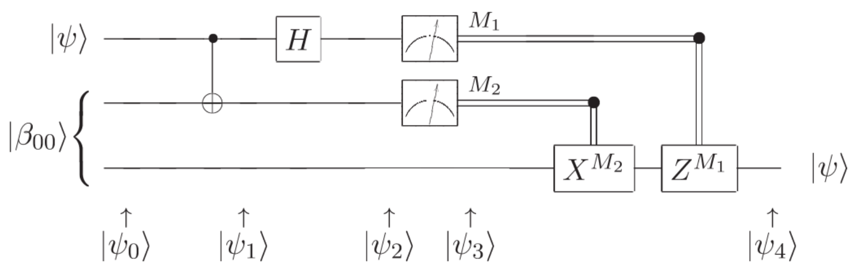


Figure 3.12: Quantum circuit showing the teleportation process

Suppose Alice and Bob are sharing $|\beta_{00}\rangle$ and the $|\psi\rangle$, assuming second input is from alice, overall input to the circuit is

$$
\begin{aligned}
|\psi_0\rangle &= |\psi\rangle\,|\beta_{00}\rangle && (3.24)\\
&= (\alpha\,|0\rangle + \beta\,|1\rangle)\left(\frac{|00\rangle + |11\rangle}{\sqrt{2}}\right) && (3.25)\\
&= \frac{1}{\sqrt{2}}\left[\alpha\,|0\rangle\,|00\rangle + \alpha\,|0\rangle\,|11\rangle + \beta\,|1\rangle\,|00\rangle + \beta\,|1\rangle\,|11\rangle\right] && (3.26)
\end{aligned}
$$

which when sent through CNOT gate on first two qubits gives $|\psi_1\rangle$

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}\left[\alpha\,|0\rangle\,(|00\rangle + |11\rangle) + \beta\,|1\rangle\,(|10\rangle + |01\rangle)\right] \tag{3.27}$$

The first qubit is then sent through Hadamard gate, which upon simplifying gives

$$|\psi_2\rangle = \frac{1}{2}\left[\alpha(|0\rangle + |1\rangle)(|00\rangle + |11\rangle) + \beta(|0\rangle - |1\rangle)(|10\rangle + |01\rangle)\right] \tag{3.28}$$

Now Alice measures the first two qubits of this whole system, giving outputs as M_1 and M_2 respectively. Based on what is measured as M_1M_2 we can know what is the final state of the qubit Bob has, which is

$$
\begin{aligned}
00 &\longmapsto |\psi_3(00)\rangle \equiv [\alpha\,|0\rangle + \beta\,|1\rangle] && (3.29)\\
01 &\longmapsto |\psi_3(01)\rangle \equiv [\alpha\,|1\rangle + \beta\,|0\rangle] && (3.30)\\
10 &\longmapsto |\psi_3(10)\rangle \equiv [\alpha\,|0\rangle - \beta\,|1\rangle] && (3.31)\\
11 &\longmapsto |\psi_3(11)\rangle \equiv [\alpha\,|1\rangle - \beta\,|0\rangle] && (3.32)
\end{aligned}
$$

Now Alice sends these two classical bits to Bob, according to which he knows what kind of transformation to apply on his qubit to transform it to the qubit Alice had. We can verify that he has to apply the matrix $Z^{M_1}X^{M_2}$ to get $|\psi_4\rangle$ i.e $|\psi\rangle$. Note that the order of matrix multiplication is opposite to the time flow as shown in the circuit diagram.

As we had to transfer classical bits, which can't be done faster than the speed of light, quantum teleportation also can't be done faster than light and is limited by the speed of transfer of the classical bits as faster than light travel could possibly result in sending information backwards in time.

Also note that we haven't created a copy of $|\psi\rangle$ which would contradict our *no-cloning theorem*. It's because the origin qubit we had in state $|\psi\rangle$ changes into the state $|0\rangle$ or $|1\rangle$ depending on the value of M_1 at the end of the process. Quantum teleportation shows that two classical bits can become a resource at least equal to one qubit information.

3.4 Quantum Algorithms

We'll explore what kind of problems a quantum computer can solve better than a classical computer. How a quantum computer can solve classical problems, we'll also look through well known quantum algorithms.

3.4.1 Classical algorithms on a quantum computer

It turns out that we can simulate any classical circuit using a quantum circuit. It should be possible as physicists believe that everything is explainable using quantum mechanics. We'll kind of prove it in the further discussion. Also, the reason we don't directly use quantum gates to simulate classical gates is that quantum gates are inherently *reversible*, whereas many classical gates such as NAND gates are inherently irreversible.

Any classical circuit can be replaced by a reversible elements known as the *Toffoli gate*. Toffoli gate is a reversible gate which takes in three input bits a, b, c. If we both the first two input bits (*control* bits) a, b are 1, it flips the third bit (*target* bit) else it doesn't affect it. It's equivalent to $(a, b, c) \longrightarrow (a, b, c \oplus ab)$ and is represented by this

Inputs			Outputs		
a	b	c	a'	b'	c'
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	1	1	0	1
1	1	0	1	1	1
1	1	1	1	1	0

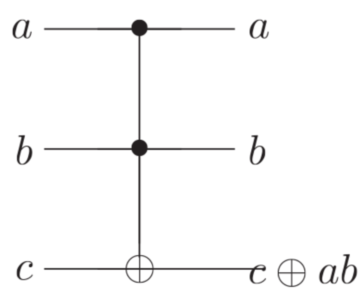


Figure 3.13: Left: Truth table for the Toffoli gate, Right: circuit representation of the Toffoli gate.

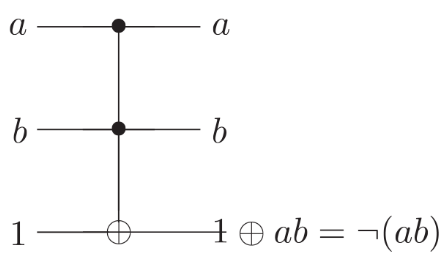


Figure 3.14: NAND gate simulated using toffoli gate.

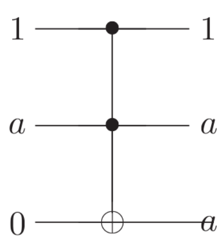


Figure 3.15: FANOUT gate simulated using Toffoli gate

Toffoli gate can also simulate NAND and FANOUT gates as shown above. Even though we've shown the classical implementation of toffoli gate above, it can be implemented as a quantum logic gate. It works in a similar way as the classical toffoli gate. For example flipping third qubit of $|110\rangle$ to give $|111\rangle$. It can also be proved that this is a valid unitary quantum gate by constructing the corresponding 8 by 8 matrix, U. Thus, the quantum toffoli gate can simulate irreversible classical gates, ensuring that quantum computers are capable of what classical computers can do. Not only a classical computer, a quantum computer can simulate a non-deterministic computer, i.e which can generate random bits. A simple way it can do it is preparing a state $|0\rangle$ then applying the Hadamard gate to it, giving $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$ which upon measuring gives $|0\rangle$ or $|1\rangle$ with 50% probability, thus generating a random number and efficiently simulating a non-deterministic classical computer. A quantum computer is way more advantageous than this and can perform much more powerful functions as we'll discuss further.

3.4.2 Quantum parallelism

Quantum parallelism is a fundamental feature of quantum computers. They can evaluate a function $f(x)$ *simultaneously* at many different values.

Suppose $f : \{0,1\} \rightarrow \{0,1\}$ be a function with a one-bit domain and range. A way of computing this function using a quantum computer is to generate a transform using quantum logic gates which transforms a given state $|x, y\rangle \longrightarrow |x, y \oplus f(x)\rangle$. When $y = 0$ the state of the second qubit is just $f(x)$. We name this map as U_f as shown

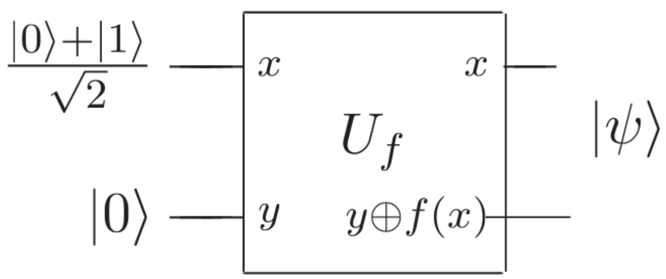


Figure 3.16: Quantum circuit for calculating $f(0)$ and $f(1)$ simultaneously. U_f is a quantum circuit which takes $|x, y\rangle$ to $|x, y \oplus f(x)\rangle$

Now if we want to compute both $f(0)$ and $f(1)$ simultaneously, we'll use the superposition state $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$, which can be generated from $|0\rangle$ using a Hadmard gate, and apply U_f on it. We then get

$$\frac{|0, f(0)\rangle + |1, f(1)\rangle}{\sqrt{2}} \tag{3.33}$$

which contains information about both $f(1)$ and $f(2)$. Thus we've in some sense evaluated both $f(0)$ and $f(1)$ parallely using a single circuit. In classical parallelism, we would have had to construct multiple circuits to evaluate the function at multiple points parallely. We can generalize it to a function which take *multiple* bits as inputs. First we'll apply *Hardmard transform* sometimes known as *Welsh-Hadmard transform* parallely on n qubits. Circuit for $n = 2$ is as shown

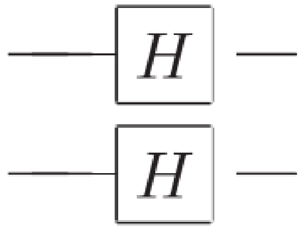


Figure 3.17: Two Hadmard gates ($H^{\otimes 2}$) acting on two qubits parallely.

For $n = 2$ we get

$$\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right) = \frac{|00\rangle + |01\rangle + |10\rangle + |11\rangle}{2} \tag{3.34}$$

This operation can be represented by $H^{\otimes 2}$. For n qubits, as $H^{\otimes n}$. Performing Hadmard transform on n qubits all initialised to $|0\rangle$ gives

$$\frac{1}{\sqrt{2^n}} \sum_x |x\rangle \tag{3.35}$$

Where x is all the possible combinations of n 0's and 1's.

Now if we want to use quantum parallelism to compute $f(x)$ which takes multi-bit values at multiple values, we first prepare a quantum system of $n+1$ qubits with state $|0\rangle^{\otimes n} |0\rangle$ which upon applying Hadmard transform on first n qubits gives 3.35. This followed by U_f produces

$$\frac{1}{\sqrt{2^n}} \sum_x |x\rangle |f(x)\rangle \tag{3.36}$$

Do note that we can still *measure* the value of f(x) only at one input. For quantum parallelism to be useful, we need some method to *extract* more informatioin from superposition states like $\sum_x |x\rangle |f(x)\rangle$.

3.4.3 Deutsch's algorithm

Using this, we can get 'global information' about $f(x)$. It combines quantum parallelism with another quantum property known as *interference*. To see how this works, let's create our first qubit as superposition $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$ using Hadmard gate on $|0\rangle$ and our second qubit as superposition $\frac{|0\rangle-|1\rangle}{\sqrt{2}}$ using Hadmard gate on $|1\rangle$. Now let's use this circuit

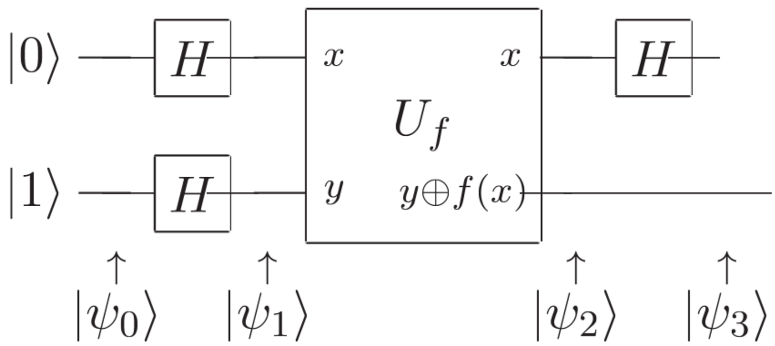


Figure 3.18: Quantum circuit implementing Deutsch's algorithm

Our input to this will be

$$|\psi_0\rangle = |0\rangle\,|1\rangle \tag{3.37}$$

When Hadmard gates are acted upon them, they become

$$|\psi_1\rangle = \left[\tfrac{|0\rangle+|1\rangle}{\sqrt{2}}\right]\left[\tfrac{|0\rangle-|1\rangle}{\sqrt{2}}\right] \tag{3.38}$$

With a little thought, it can be seen that applying U_f on $|x\rangle\,(|0\rangle-|1\rangle)/\sqrt{2}$ gives $(-1)^{f(x)}\,|x\rangle\,(|0\rangle-|1\rangle)/\sqrt{2}$, Hence applying U_f to $|\psi_1\rangle$ gives

$$|\psi_2\rangle = \begin{cases} \pm\left[\tfrac{|0\rangle+|1\rangle}{\sqrt{2}}\right]\left[\tfrac{|0\rangle-|1\rangle}{\sqrt{2}}\right] & \text{if } f(0)=f(1) \\ \\ \pm\left[\tfrac{|0\rangle-|1\rangle}{\sqrt{2}}\right]\left[\tfrac{|0\rangle-|1\rangle}{\sqrt{2}}\right] & \text{if } f(0)\neq f(1) \end{cases} \tag{3.39}$$

Applying the final Hadamard gate on this give $|\psi_3\rangle$

$$|\psi_3\rangle = \begin{cases} \pm\,|0\rangle\left[\tfrac{|0\rangle-|1\rangle}{\sqrt{2}}\right] & \text{if } f(0)=f(1) \\ \pm\,|1\rangle\left[\tfrac{|0\rangle-|1\rangle}{\sqrt{2}}\right] & \text{if } f(0)\neq f(1) \end{cases} \tag{3.40}$$

This can be condensed more by observing that $f(0) = f(1) \implies 0 = f(0)\oplus f(1)$ and $f(0)\neq f(1) \implies 1 = f(0)\oplus f(1)$, Thus

$$|\psi_3\rangle = \pm\,|f(0)\oplus f(1)\rangle\left[\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right] \tag{3.41}$$

Thus, we can measure $f(0)\oplus f(1)$ by measuring the first qubit. This is really interesting since we found out $f(0)\oplus f(1)$ which is a *global* property of $f(x)$ by just evaluating $f(x)$ *once*! Thus distinguishes a quantum parallelism from even a classical computer. Even a probabilistic classical computer could determine $f(0)$ or $f(1)$ with probability $\frac{1}{2}$ each, a quantum computer evaluates both at once. We can get global information about function by cleverly chosing operation and gates so that the two evaluations *interfere* with each other to give us information like this, which is not possible in classical computers.

3.4.4 The Deutsch-Jozsa algorithm

This is a generalisation to the Deutsch's algorithm. It's application, known as *Deutsch's problem* is described as following. Suppose Alice and Bob are far apart, Alice is in Amsterdam and Bob is in Boston. Alice sends a number x out of $\{0,\cdots,2^n-1\}$ through a letter to Bob. Bob computes a function $f(x)$ which yields the result of either 0 or 1 and mails it back to Alice. Alice needs to find out whether the function $f(x)$ is *constant* or *balanced* i.e 0 for half the values and 1 for the other half by sending as many mails as she can. Using classical means we can figure out that she would need $2^{n-1}+1$ mails to be sent in the worst case.

We can offer a much better solution using qubits if we're sure that Bob is willing to calculate $f(x)$ using a unitary transform U_f in a single interaction. To see how this works, Alice has a set of n qubits each prepared as $|0\rangle$ in which she stores the query $(\sim x)$ and another qubit prepered as $|1\rangle$ where she stores the answer $(\sim f(x))$. She then applies Hadmard transform on all $n+1$ qubits after which she sends these to Bob. Bobs first applies U_f on them and then Alice applies Hadmard transform on the first n qubits such that she gets information about $f(x)$ and thus the solution. We'll look through how this happens. Take a look at the circuit

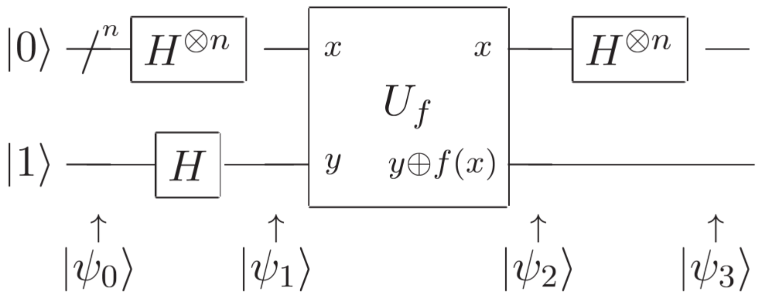


Figure 3.19: Implementation of Deutsch-Jozsa algorithm, the cross represents n qubits

So the input state is

$$|\psi_0\rangle = |0\rangle^{\otimes n} |1\rangle \tag{3.42}$$

When the Hadmard transform is applied, the query qubits turn into a superposition of $|x\rangle$ where x is binary representation of all numbers in $\{0, \cdots, 2^n - 1\}$

$$|\psi_1\rangle = \sum_x \frac{|x\rangle}{\sqrt{2^n}} \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \tag{3.43}$$

This is sent to Bob, he applies U_f to $|\psi_1\rangle$ which results in $|\psi_2\rangle$ using similar result as in 3.38

$$|\psi_2\rangle = \sum_x \frac{|x\rangle}{\sqrt{2^n}} (-1)^{f(x)} \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \tag{3.44}$$

Now comes the tricky part when Alice applies Hadmard gate on first n qubits. Let's just see what happens when Hadamard transform is applied to $|x\rangle$ which is just a binary string (representation of some number $\in \{0, \cdots, 2^n - 1\}$). We get another superposition of all possible basis states $|z\rangle$ whose quotients are $(-1)^{xz}$ which is bitwise multiplication/AND of x and z along with the normalization constant.

$$H|x\rangle = \sum_z \frac{(-1)^{xz} |z\rangle}{\sqrt{2^n}} \tag{3.45}$$

Thus the final output Alice gets is

$$|\psi_3\rangle = \sum_x \sum_z \frac{(-1)^{xz+f(x)} |z\rangle}{2^n} \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \tag{3.46}$$

Looking at first n qubits, coefficient of $|0\rangle^{\otimes n}$ turnes out to be just $\sum_x \frac{(-1)^{f(x)}}{2^n}$. Now note that if $f(x)$ was constant $\sum_x (-1)^{f(x)}$ would be $\pm 2^n$, which would mean coefficient of $|0\rangle^{\otimes n}$ is ± 1. Since norm of state of a system should be 1 coefficients of all other $|z\rangle$ would be 0. Thus, Alice would only measure $|0\rangle^{\otimes n}$. Now in the case when $f(x)$ is balanced, coefficient of $|0\rangle^{\otimes n}$ would be 0. Hence, Alice would never get the measurement result as 0.

Alice gets $|0\rangle^{\otimes n}$ upon measuring $|\psi_3\rangle \iff f(x)$ is constant

Thus by measuring the query register Alice can get to know the answer in just one correspondence. Let's summarise this algorithm

Algorithm 1: Deutsch-Jozsa

Inputs : A black box U_f which performs the transformation $|x\rangle |y\rangle \longrightarrow |x\rangle |y \oplus f(x)\rangle$, $x \in \{0, \cdots, 2^n - 1\}$ and $f(x) \in \{0, 1\}$. It's guaranteed that $f(X)$ is either *balanced* or *constant*.

Outputs : 0 if and only if f is constant.

Runtime: One evaluation of U_f, always succeeds.

Procedure

$|0\rangle^{\otimes n} |1\rangle$ `// initialize state`

$\longrightarrow \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle \left[\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right]$ `// create superposition using Hadmard gates`

$\longrightarrow \sum_x (-1)^{f(x)} |x\rangle \left[\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right]$ `// calculate` f `using` U_f

$\longrightarrow \sum_z \sum_x \frac{(-1)^{xz+f(x)}|z\rangle}{2^n} \left[\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right]$ `// perform Hadmard transform`

$\longrightarrow z$ `// measure to obtain output`

This is a seed for quantum algorithms. But it has it's own problems, such it has no real world application. It can be solved in a more realistic manner using a probabilistic computer.

3.4.5 Quantum algorithms summarized

Broadly speaking, quantum computers are good at solving three classes of problems. First one based upon quantum versions of Fourier transform, Deutsch-Jozsa algorithm, Shor's algorithm and discrete logarithm come under this class. Second class is quantum search algorithms. Third class of algorithms is quantum simulation, using which a quantum computer can simulate a quantum system. Let's look at each of these

Quantum algorithms based on Fourier transform

The discrete Fourier transform is described as transforming the set $x_0, \cdots, x_{N-1}$ of N complex numbers into the set of complex numbers $y_1, \cdots, y_{N-1}$ according to this

$$y_k \equiv \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{\frac{2\pi ijk}{N}} x_j \tag{3.47}$$

It's widely known in science because turns a much complex problem into an easier one. There's a more generalized theory of fourier transforms which we won't be discussing. The Hadamard gates we've used in Deutsch-Jozsa algorithm and also the famous Shor's algorithm and discrete logarithm involve some kind of fourier transform. To go into the quantum realm, we can define a linear transform U on n qubits whose action on the basis states $|k\rangle$, $0 \leq k \leq 2^n - 1$ is defined as

$$|j\rangle \longrightarrow \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} e^{\frac{2\pi ijk}{2^n}} |k\rangle \tag{3.48}$$

It can be checked that this transform is unitary, and can be realised as a quantum circuit. On a general qubit, this is

$$|\psi\rangle = \sum_{j=0}^{2^n-1} x_j |j\rangle \longrightarrow \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} \left[\sum_{j=0}^{2^n-1} e^{\frac{2\pi ijk}{2^n}} x_j \right] |k\rangle = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} y_k |k\rangle \tag{3.49}$$

Which can be seen as the vector form of fourier transform on the set x_j and $N = 2^n$.

A classical computer would perform the fourier transform in $N \log N = n2^n$ steps, whereas a quantum computer would just take $(\log N)^2 = n^2$ steps. Again the same problem arises that a lot of information is "hidden" and not accessible. When we try to measure it the state just collapses to either $|0\rangle$ or $|1\rangle$ making us loose our information. Hence, more cleverness is needed to harness the proper power of quantum computation. We'll come back to more applications of Fourier Transform.

Quantum search algorithms

A quantum computer can perform search algorithms, where we have to find an element from a collection of n random elements with a specific property in $\sqrt{n}$ steps. Whereas a classical computer takes n steps. Even though this advantage is less than what we had in fourier transform's case, this field has a wide variety of applications compared to the latter one.

Quantum simulation

Simulating quantum system, as expected, is better done by quantum computers than classical ones. For a classical computers the number of complex numbers needed to describe a system grows *exponentially* whereas it grows *linearly* for a quantum computer i.e a quantum system having n components would require c^n bits of memory for a classical computer where c depends on the system and accuracy we need. Whereas a quantum computer would require only kn bits of memory. Again, the c^n bits of information isn't accessible using the kn bit quantum comupter. We'd still just be able to access kn bits of memory. How to extract more information is still partially understood. Still quantum simulation would be an important applications of QIC as we coould understand about complex chemical molecules which generally occur in biology.

The power of quantum computation

We'll try to get some idea of how powerful quantum computers can be compared to classical computers, yet there is still a possibility they are not much powerful than a classical computer. *Computational complexity theory* deals with how difficult can it be to solve a problem. Most basic idea of it is of *complexity class* which is a set of problems all having some kind of similarity in how resource intensive they are.

Two most important complexity classes are **P** and **NP**. **P** is the set of problems which can be quickly solved using a classical computer. **NP** is the set of problems whose solutions can be quickly checked. For example, there's no algorithm for factoring two integers that's in **P** on a classical computer. But we can check if p is a factor of n, hence it's an **NP** problem.

It's clear that **P** is a subset of **NP** since the ability to solve a problem implies the ability to check potential solutions. But there are no known problems that are "surely" not in **P** but are in **NP**. This is a famous open problem in computer science

$$\mathbf{P} \overset{?}{\neq} \mathbf{NP} \tag{3.50}$$

There's an important subclass of **NP** problems, known as **NP**-complete problems. Many hard and highly important problems are known to be **NP**-complete. In some sense, **NP**-complete problems are "atleast" as hard as **NP** problems. Any algorithm solving **NP**-complete problems can be adapted to solve **NP** hard problems, with a little overhead. If $\mathbf{P} \neq \mathbf{NP}$, then there would be no **NP**-complete problem efficiently solvable by a classical computer.

It is not known whether quantum computers can solve all **NP** problems. They can solve few problems like factoring which aren't surely known to be in **NP** but are believed to be in. There's an interesting result that a simple variant of quantum parallelism can't solve all **NP** problems. But still, this does not rule out that there might be some deeper structure which would let us solve all **NP** problems on a quantum computer efficiently.

Another important complexity class **PSPACE** is the set of problems which are efficient on space but not necessarily on time. Even though *not proved*, **PSPACE** is believed to be bigger than both **P** and **NP**. Finally, **BPP** is the set of problems that can be solved using randomized algorithms with some bound probability error (say 1/4). They're more believed to be the class that can be solved by a classical computer. Even though **P** is more studied.

BQP is the class of all computational problems that can be solved efficiently on a quantum computer with a bound probability error. **BQP** is knownn surely to be bigger than **P** but not bigger than **PSPACE**. It's not exactly known where it fits between **P**, **NP**, **PSPACE**. This gives an important fact that if *quantum computers are strictly powerful than classical computers* then **PSPACE** is strictly bigger than **P**. The latter part has been really hard to prove by computer scientists, lowering the hope that quantum computers are more powerful than classical computers.

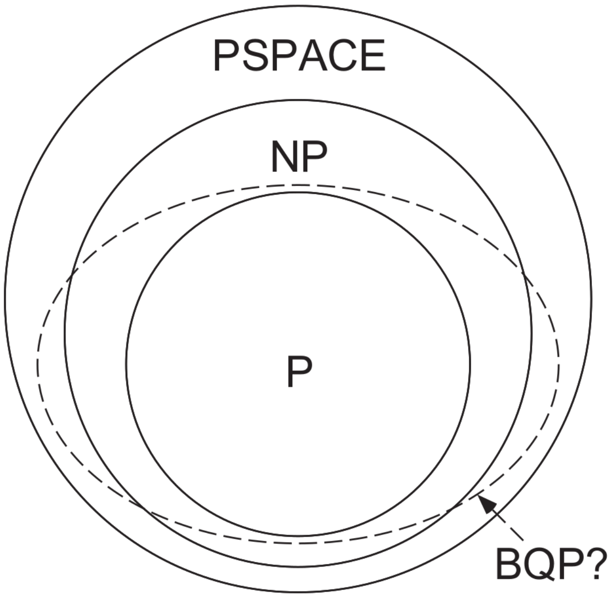


Figure 3.20: We surely know that quantum computers can solve all **P** problems quickly and also that they cannot solve any problem outside **PSPACE** but we don't know where quantum computers fit between **P** and **PSPACE**. We don't even know whether **P** is equal to **PSPACE**.

We can already see that the *theory* and notions of quantum computing pose a challenge to traditional computational methods. Another important challenge is that quantum computing is believed to be *experimentally realizable*, primarily because nature works according to quantum mechanics.

3.5 Experimental quantum information processing

We'll look into experimentally proving that quantum computing can be done. Let's start with the "Stern-Gerlach" experiment which provides evidence to the existence of qubits in nature.

3.5.1 The Stern-Gerlach experiment

In the original experiment performed in 1921-22, hot 'silver atoms' were beamed from an over through a magnetic field which deflected it. We'll consider the hydrogen atom version performed in 1927 as it's much simpler. A hydrogen atom consists of a proton and an electron spinning around which makes up the atomic dipole moment. By constructing the apparatus properly, it's possible to make the atom deflect by an amount dependent on the $\hat{z}$ component of atomic dipole moment. It's naturally expected that the atomic dipole moments would be randomly oriented and we'd get a continuous distribution. But the atoms were deflecting at specific angles. This was then explained by the *quantization* principle, which was trending at that time. But quantum mechanics had predicted that the net atomic dipole of an atom must be zero, which is classically strange though, but there were two beams seen, one deflected up and the other deflected down.

This puzzle was explained by a new quantity called *spin* which every hydrogen atom is supposed to be associated with. This spin is posited to make *extra contribution* to the magnetic dipole moment.

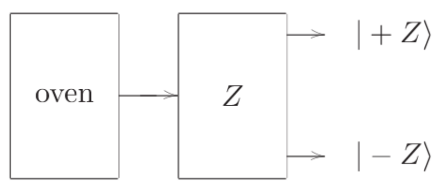


Figure 3.21: Original Stern-Gerlach setup.

We can assume that the spins $|{+Z}\rangle$, $|{-Z}\rangle$ leads to upper and lower deflection respectively. Now if the apparatus modiied that the upper beam is then passed through another apparatus *tipped sideways* so that it deflects according to $\hat{x}$ axis.

Part III

Week 3

Chapter 4

Quantum noise and quantum operations

We'll be discussing about *open systems*. These are unlike *closed* systems, which don't interact with outside environment. These can suffer with unwanted interactions with the outside environment. These unwanted interations show up as *noise* in quantum information systems. We'll describe *quantum operations formalism*, a powerful set of tools enabling us to describe *open* quantum systems and their behaviour. These can describe not only systems weakly coupled with the environment, but also the system which are strongly coupled with the environment. They can also describe closed systems which are suddenly opened and are subject to measurement. They are good at describing *discrete* state changes, i.e changes of state from ρ to ρ' without explicit mention of passage of time.

4.1 Classical noise and Markov processes

To understand about noise, let's consider a simple situation where a hard drive of a computer stores a single bit, which can be 0 or 1. Due to external stray magnetic fields, this bit is subject to change after a long time. To quantify it, let p be the probability that the bit will flip, then $1 - p$ will be the probability that the bit remains the same.

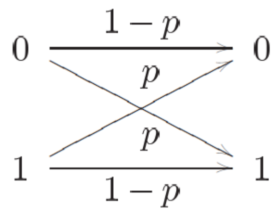


Figure 4.1: A bit flip might occur with probability p

To figure out this probability p, we need to understand two things (a) How the magnetic fields are distributed in the enviroment. Assuming the owner isn't crazy and placing a magnet near the disk, we can sample the magnetic field near the disk similar to the one the disk is in. (b) How this environment's magnetic field interacts with the disk. This can be done by the well established *Maxwell's equations*.

Suppose p_0, p_1 be the probabilities that the bit is in state 0 and 1 initially. Let q_0, q_1 be the corresponding probabilities after the noise has occured and X, Y be the initial and final states respectively then

$$P(Y = y) = \sum_x P(Y = y|X = x)P(X = x) \tag{4.1}$$

where the conditional probability $P(Y = y|X = x)$ is known as the *transition probability*. Thus

$$\begin{bmatrix} q_0 \\ q_1 \end{bmatrix} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \end{bmatrix} \tag{4.2}$$

Suppose we make a quantum circuit made of two faulty NOT gates. It takes an input state X converts it into an intermediate bit Y which is finally converted to Z. If we make a physically reasonable assumption that the noise produced at first NOT gate independent of the noise produced at the second NOT gate, it results in a stochastic process $X \longrightarrow Y \longrightarrow Z$ of a special type known as *Markov process*. Often in multi-stage processes, it's a good assumption to use Markov processes.

For a single stage process, output probabilities $\overrightarrow{q}$ are related to the input probabilities $\overrightarrow{p}$ according to

$$\overrightarrow{q} = E\overrightarrow{p} \tag{4.3}$$

E is a matrix of transition probabilities known as the *Evolution matrix*. Thus the final state of system is linearly related to it's initial state. This linearity is echoed in quantum noise's description, where probability distributions are replaced by density operators. In addition to this, E follows two more rules (a) All of it's entries must be positive, so that $E\overrightarrow{p}$ is also positive and a valid probability distribution. This is known as *positivity*. (b) Sum of the entries in each column should be 1. This is known as *completeness*.

4.2 Quantum operations

4.2.1 Overview

Quantum operations formalism is a general tool for describing evolution of quantum systems in a wide variety of circumstances, include stochastic changes, similar to how Markov processes describing classical stochastic changes. Similar to how classical bits are represented using probability distributions, here we'll use *density operators*. The transformation looks like

$$\rho' = \mathcal{E}(\rho) \tag{4.4}$$

Here, ρ is the initial density operator describing the state of the system. Similarly, ρ' is the final state upto a normalization factor. A few known transforms $\mathcal{E}$ are unitary transformations and measurements $\mathcal{E}(\rho) = U\rho U^\dagger$ and $\mathcal{E}_m(\rho) = M_m\rho M_m^\dagger$.

In our further study, we're going to look at quantum operations *three* seperate ways. The first way is based on the idea of studying dynamics as the result of interaction between a system and environment. This method is *concrete* and easy to relate to real world. But it is mathematically inconvenient. Second way is euivalent to the first one but is mathematically convenient using a powerful mathematical representation for quantum operations known as the *operator-sum representation*. Third way is equivalent to the first two, it's via a set of *physically motivated axioms* that we'd expect these maps to satisfy. It's advanage is that it shows that quantum dynamics will be described by quantum operations in a wide range of circumstances. But this is neither concrete nor mathematically convenient.

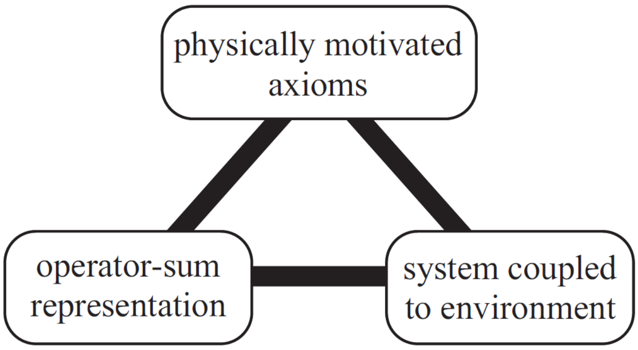


Figure 4.2: Three ways we're going to look at quantum operations, each provide their own advantages

4.2.2 Environment and quantum operations

By the postulate of quantum mechanics, dynamics of a *closed* system is described by just a unitary transform which can be thought of as a box. For dynamics of an *open* system, the system can be thought of as a composite closed system of the system we're considering, *principal system* and the *environment*.

Suppose the system is in state ρ, then the final state would be $\mathcal{E}(\rho)$ which need not be a unitary transform. We *assume* that system and environtment are in a product state $\rho \otimes \rho_{env}$. Evolution of this closed system is described by a unitary transform U and we use partial trace to find final state of the system $\mathcal{E}(\rho)$

$$\mathcal{E}(\rho) = \text{tr}_{\text{env}} \left[U\rho \otimes \rho_{\text{env}} U^\dagger \right] \tag{4.5}$$

This is the first *definition* of quantum operations. If U doesn't depend on the environment, then final state is simply $\tilde{U}\rho\tilde{U}^\dagger$ where $\tilde{U}$ is just part of transform on our principal state.

We assumed that environment and principal system are in a product state, but this might not be the case always. In cases of practical interest, this assumption is reasonable. The experimentalist while preparing an open state clears all the *correlations* between the state and environment.

You might thin the envionment would have near-infinite dimensional Hilbert space, but it turns out that for principal system with d-dimensional Hilbert space, it's enough to model the environment with no more that d^2 dimensions.

4.2.3 Operator sum representation

How an open system evolves can be understood with an operator acting on the Hilbert space of principal system. This representation of quantum operations is known as *Operator sum representation*.

Let this environment be in a pure state $\rho_{\text{env}} = |e_0\rangle \langle e_0|$, which even if it's not, can be made pure. Let $|e_k\rangle$ be an orthonormal basis for environment. The operation can be represented as

$$\mathcal{E}(\rho) = \sum_k \langle e_k| U \left[\rho \otimes |e_0\rangle \langle e_0|\right)] U^\dagger |e_k\rangle \tag{4.6}$$

$$= \sum_k E_k \rho E_k^\dagger \tag{4.7}$$

where $E_k = \langle e_k| U |e_0\rangle$ is an operator acting on principal state known as an *operation element.* These have a nice property known as *completeness relation* coming from the fact that tr$(\mathcal{E}(\phi)) = 1$ since that's just another valid state

$$\begin{aligned} 1 &= \text{tr}(\mathcal{E}(\rho)) && (4.8)\\ &= \text{tr}\left(\sum_k E_k \rho E_k^\dagger\right) && (4.9)\\ &= \text{tr}\left(\sum_k E_k E_k^\dagger \rho\right) && (4.10) \end{aligned}$$

As it's true for all ρ,

$$\sum_k E_k E_k^\dagger = I \tag{4.11}$$

This is satisfied by operations which are *trace-preserving*. We'll see later non trace preserving operations are also equivalent to this.

This representation gives us *intrinsic* means to characterize dynamics of the principal system. We don't need to know how the environment works, just E_k are sufficient. Next up we'll see it's physical interpretation, making a physical model from this representation and also how to define this representation when our assumptions faii.

Physical inerpretation of operator-sum representation

Suppose a unitary transform U is applied followed by a measurement of the environment, in basis $|e_k\rangle$. The measurement only affects the environment, thus state of the system if the output is k is

$$\begin{aligned} \rho_k &\propto \text{tr}_E\left(|e_k\rangle \langle e_k| U(\rho \otimes |e_0\rangle \langle e_0|)U^\dagger |e_k\rangle \langle e_k|\right) && (4.12)\\ &= \langle e_k| U\left(\rho \otimes |e_0\rangle \langle e_0|\right) U^\dagger |e_k\rangle && (4.13)\\ &= E_k \rho E_k^\dagger && (4.14) \end{aligned}$$

which upon normalizing gives,

$$\rho_k = \frac{E_k \rho E_k^\dagger}{\text{tr}(E_k \rho E_k^\dagger)} \tag{4.15}$$

with probability

$$\text{tr}(E_k \rho E_k^\dagger) \tag{4.16}$$

thus the final output is

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger \tag{4.17}$$

Thus we can see that a quantum operation changes the input randomly into a state $\frac{E_k \rho E_k^\dagger}{\text{tr}(E_k \rho E_k^\dagger)}$ with probability tr$(E_k \rho E_k^\dagger)$ which is very similar to a noisy channel.

Measurement and operator-sum representation

We can have non trace preserving operations if we allow *measurement* of the prinicpal state-environment system after the operator U is applied. Suppose the principal system is represented by Q and the environment by E, both are initially independent, environment is in the state σ. Then

$$\rho^{QE} = \rho \otimes \sigma \tag{4.18}$$

Now a unitary transform U is applied, then a projective measurement is made using P_m. We assume no measurement is made is same as the outcome is $m = 0$ and $P_0 \equiv I$. The final state of QE is

$$\frac{P_m U(\rho \otimes \sigma)U^\dagger P_m}{\text{tr}(P_m U(\rho \otimes \sigma)U^\dagger P_m)} \tag{4.19}$$

State of only the principal system Q is

$$\frac{\text{tr}_E(P_m U(\rho \otimes \sigma)U^\dagger P_m)}{\text{tr}(P_m U(\rho \otimes \sigma)U^\dagger P_m)} \tag{4.20}$$

We can define a map $\mathcal{E}_m(\rho)$ as

$$\mathcal{E}_m(\rho) = \text{tr}_E(P_m U(\rho \otimes \sigma)U^\dagger P_m) \tag{4.21}$$

therefore the final state of Q is $\frac{\mathcal{E}_m(\rho)}{\text{tr}(\mathcal{E}_m(\rho))}$ and $\text{tr}(\mathcal{E}_m(\rho))$ is the probability of measurement outcome m. Suppose $|e_k\rangle$ is an orthonormal basis of E and $\sigma = \sum_j q_j |j\rangle$ is an ensemble of initial state of environment then we observe that

$$\mathcal{E}_m(\rho) = \sum_{jk} q_j \text{tr}_E(|e_k\rangle \langle e_k| P_m U(\rho \otimes \sigma)U^\dagger P_m |e_k\rangle \langle e_k|) \tag{4.22}$$

$$= \sum_{jk} E_{jk} \rho E_{jk}^\dagger \tag{4.23}$$

where

$$E_j k \equiv \sqrt{q_j} \langle e_j| U |j\rangle \tag{4.24}$$

using which we can calculate the operators for the operator sum representation if σ of E is known. Even though we don't make projective measurements, respective measurement probabilities are $\text{tr}(\mathcal{E}_m(\rho))$

System-environment models for any operator-sum representation

Now a natural question is for a given set of operators $\{E_k\}$ is it possible to find a reasonable *model environment system and dynamics* which give rise to the quantum operation described by these operators. It turns out that for a set of trace preserving or non trace preserving quantum operation $\mathcal{E}$, there exists a model environment E staring with a pure state $|e_0\rangle$, and model dynamics specified by unitary operator U and projector P onto E such that

$$\mathcal{E}(\rho) = \text{tr}_E(PU(\rho \otimes |e_0\rangle \langle e_0|)U^\dagger P) \tag{4.25}$$

4.2.4 Axiomatic approach to quantum operations

We'll try to write down physically motivated axioms which quantum operations obey. We'll start over from scratch using these axioms. We'll then prove that a map $\mathcal{E}$ satisfies these axioms *if and only if* it has an operator sum representation.

We *define* a quantum operation $\mathcal{E}$ as a map from the set of density operators of the input space Q_1 to the set of density operators for the output space Q_2, which satisfy the following (for convenience we'd take $Q_1 = Q_2 = Q$):

1. $\text{tr}(\mathcal{E}(\rho))$ denotes the probability that the map described by $\mathcal{E}$ occurs. Thus $0 \leq \text{tr}(\mathcal{E}(\rho)) \leq 1$.

2. $\mathcal{E}$ is a *convex-linear map* on the set of density matrices i.e

$$\mathcal{E}\left(\sum_i p_i \rho_i\right) = \sum_i p_i \mathcal{E}(\rho_i). \tag{4.26}$$

3. $\mathcal{E}$ is a *completely-positive map* i.e $\mathcal{E}(A)$ is positive for any positive A. Also, if we introduce an extra system $\mathcal{R}$, then $\mathcal{I} \otimes \mathcal{E}(A)$ is positive for any positive A in $\mathcal{R}Q_1$, where $\mathcal{I}$ is the identity map on $\mathcal{R}$.

The following theorem binds the axiomatic approach and the operator-sum representation

Theorem 8. *The map $\mathcal{E}$ satisfies the above axioms if and only if*

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger \tag{4.27}$$

for some set of operators $\{E_k\}$ mapping the input Hilbert space to the output Hilbert space and satisfying $\sum_k E_k E_k^\dagger \leq I$

Freedom in the operator-sum representation

A quanum operation can be described by two different set of operators. For example take $\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger$ and $\mathcal{F}(\rho) = \sum_k F_k \rho F_k^\dagger$ and

$$E_1 = \frac{I}{\sqrt{2}} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad E_2 = \frac{Z}{\sqrt{2}} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{4.28}$$

$$F_1 = |0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad F_2 = |1\rangle\langle 1| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{4.29}$$

Understanding which sets of operations gives rise to same quantum operations is important for two reasons. (a) Understanding the freedom in representation gives us more insight into how different physical processes give rise to same dynamics (b) Understanding freedom in operator-sum representation is cruicial to a good understanding of quantum error-correction. This is described usin the following theorem

Theorem 9 (Unitary freedom in the operator-sum representation). *Suppose $\{E_1, \cdots, E_m\}$ and $\{F_1, \cdots, F_n\}$ are operation elements giving rise to quantum operations $\mathcal{E}$ and $\mathcal{F}$ respectively. We may ensure $m = n$ by appending 0 operators to the end of shorter operation elements' list. Then $\mathcal{E} = \mathcal{F}$ if and only if there exist complex numbers u_{ij} such that $E_i = \sum_j u_i j F_j$ and u_{ij} is a unitary matrix.*

This is similar to the result when two states are denoted by the same density operator. Using this theorem we can also answer the question, what is the maximum number of dimensions of an environment needed to mock up a quantum operation as described by this theorem.

Theorem 10. *All quantum operations $\mathcal{E}$ on a principal system of Hilbert space dimension d can be generated by an operator sum representation containing atmost d^2 elements, i.e*

$$\mathcal{E}(\rho) = \sum_{k=1}^{M} E_k \rho E_k^{\dagger} \tag{4.30}$$

where $1 \leq M \leq d^2$.

4.3 Examples of quantum noise and quantum operations

4.3.1 Trace and partial trace

Quantum operations formalism can be used to describe a measurement, it's outcome and the state of the system after the measurement. The simplest such operation is related to the map $\rho \longrightarrow \mathrm{tr}(\rho)$. Let H_Q be the input Hilbert space with orthogonal basis $|1\rangle, \cdots, |d\rangle$ and H_Q' is a one dimensional Hilbert space, spanned by $|0\rangle$, define $\mathcal{E}(\rho)$ as

$$\mathcal{E}(\rho) \equiv \sum_{i=1}^{d} |0\rangle \langle i| \rho |i\rangle \langle 0| \tag{4.31}$$

Do note that $\mathcal{E}(\rho) = \mathrm{tr}(\rho) |0\rangle \langle 0|$, thus this quantum operation is identical to the trace operation, ignoring the factor of $|0\rangle \langle 0|$.

Another interesting thing comes when we have a joint system QR and we want to trace out R. Let $|j\rangle$ be the basis of R, let's define a linear operator $E_i : H_{QR} \longrightarrow H_Q$ as

$$E_i \left(\sum_j \lambda_j |q_j\rangle |j\rangle \right) = \lambda_i |q_i\rangle \tag{4.32}$$

when $|q_j\rangle$ are arbitrary vectors in Q and λ_j are just constants. Then we define a quantum operation as

$$\mathcal{E}(\rho) \equiv \sum_k E_k \rho E_k^{\dagger} \tag{4.33}$$

We also notice that

$$\mathcal{E}(\rho \otimes |j\rangle \langle j'|) = \rho \delta_{j,j'} = \rho \mathrm{tr}(|j\rangle \langle j'|) = \mathrm{tr}_R(\rho \otimes |j\rangle \langle j'|). \tag{4.34}$$

where $|j\rangle, |j'\rangle$ are members of orthonormal basis of R. Thus, by linearity of $\mathcal{E}$ and tr_R we have $\mathcal{E} = \mathrm{tr}_R$

4.3.2 Geometric picture of single qubit quantum operations

We've seen before that a single qubit's density operator can be represented by

$$\rho = \frac{I + \vec{r} \cdot \vec{\sigma}}{2} \tag{4.35}$$

where $\vec{r}$ is a three dimensional real unit vector and $\vec{\sigma}$ is a vector of pauli matrices. Thus

$$\rho = \frac{1}{2}\begin{bmatrix} 1+r_z & r_x - ir_y \\ r_x + ir_y & 1-r_z \end{bmatrix} \tag{4.36}$$

Using this representation, it can be shown that an arbitrary trace-preserving quantum operation is equivalent to the map

$$\vec{r} \longrightarrow \vec{r'} = M\vec{r} + \vec{c} \tag{4.37}$$

where M is a 3×3 real matrix, and $\vec{c}$ is a constant vector. This is an *affine map*, mappping Bloch sphere into itself. To see this, suppose the operators E_i generating $\mathcal{E}$ are written in the form

$$E_i = \alpha_i I + \sum_{k=1}^{3} a_i k \sigma_k \tag{4.38}$$

Then it can be checked that

$$M_{jk} = \sum_l \left[a_{lj}a_{lk}^* + a_{lj}^* a_{lk} + \left(|\alpha_l|^2 - \sum_p a_{lp}a_{lp}^* \right) \delta_{jk} + i \sum_p \epsilon_{jkp}(\alpha_l a_{lp}^* - \alpha_l^* a_{lp}) \right] \tag{4.39}$$

$$c_k = 2i \sum_l \sum_{jp} \epsilon_{jpk} a_{lj} a_{lp}^* \tag{4.40}$$

The affine map can be better understood by te polar decomposition of M into $M = U|M|$, where U is unitary. Since M is real, $|M|$ is real and hermitian, i.e a symmetric matrix. Thus, as M is real, we may assume U is real, and an orthogonal matrix i.e $U^T U = I$ Thus, we can write

$$M = OS \tag{4.41}$$

where determinant of O is 1 and it's an orthogonal matrix describing rotation. S is a real symmetric matrix. Thus the above equation can be understood as the deformation of the bloch sphere about axis as specified by S followed by a rotation defined by O, then a displacment due to $\vec{c}$.

4.3.3 Bit flip and phase flip channels

Suppose we take a single qubit and pass it through a noisy *bit-flip* channel, which flips the state of a qubit from $|0\rangle$ to $|1\rangle$ with probability $1-p$. It has operation elements

$$E_0 = \sqrt{p}I = \sqrt{p}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad E_1 = \sqrt{1-p}X = \sqrt{1-p}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{4.42}$$

This affects the bloch vector in the following way:

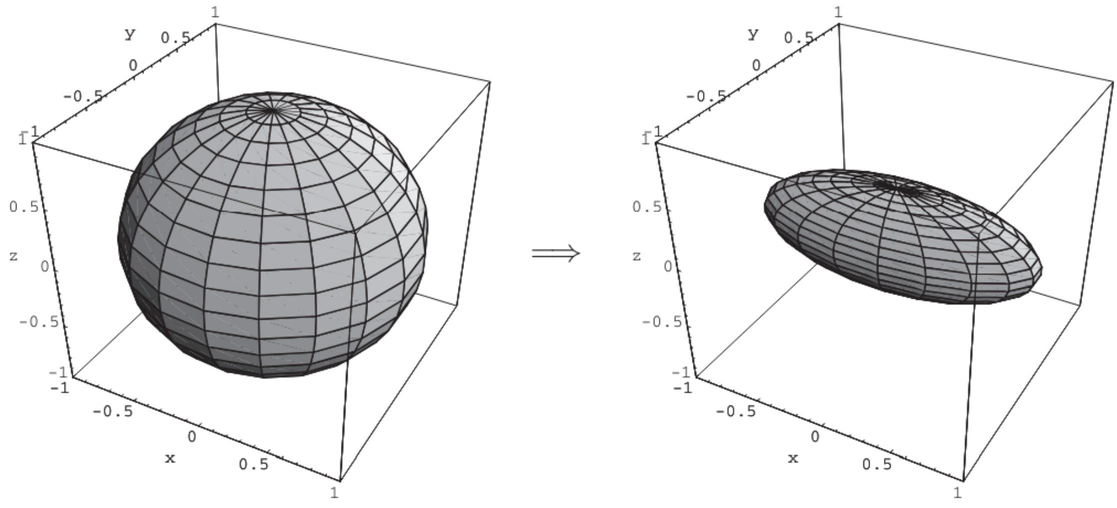


Figure 4.3: Bit flip channl. The operation doesn't affect the x component, but shrinks the y and z component by a factor of $1-2p$. Here $p=0.3$.

It's easy to verify that $\mathrm{tr}(\rho^2) = \frac{1+|\vec{r}|^2}{2}$. Thus, this operation can never increase the length of Bloch vector i.e can never increase $\mathrm{tr}(\rho^2)$. Thus the purity of ρ is always decreased by this. Thus Bloch sphere provides an intuitive way of understanding quantum operations.

Similarly if we consider the *phase-flip* channel which has operation elements

$$E_0 = \sqrt{p}I = \sqrt{p}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} \qquad E_1 = \sqrt{1-p}Z = \sqrt{1-p}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} \tag{4.43}$$

we see that the following transformation is happening to the Bloch sphere.

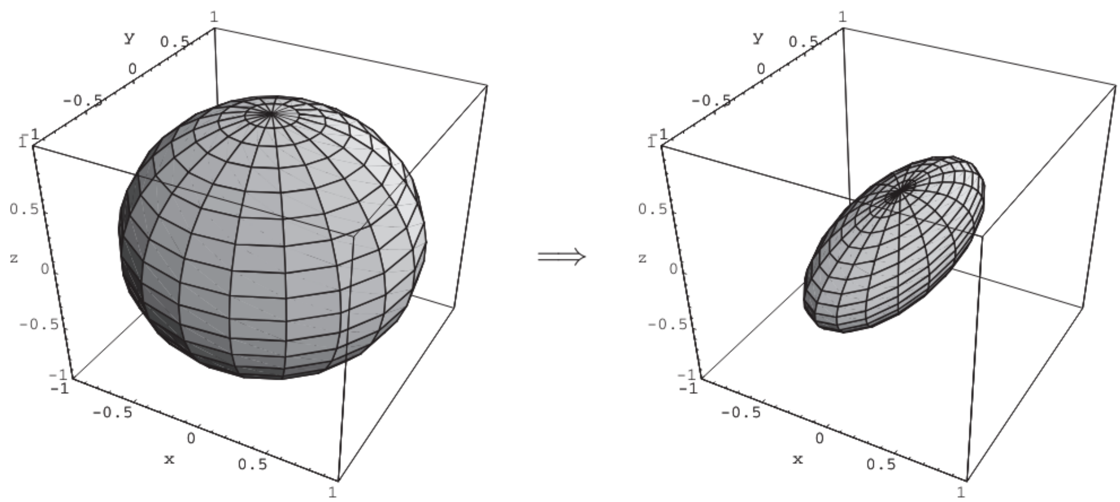


Figure 4.4: Phase flip channel. As we can see, it does nothing to the z component, but shrinks x and y component by a factor of $1-2p$. Here p=0.3.

As a special case when $p = 0.5$, using the freedom in the operator-sum representation, this operation can be written as

$$\rho \longrightarrow \mathcal{E}(\rho) = P_0 \rho P_0 + P_1 \rho P_1 \tag{4.44}$$

which is nothing but the measurement with respect to the basis $|0\rangle$, $|1\rangle$, the result unknown. The corresponding map on Bloch sphere here is

$$(r_x, r_y, r_z) \longrightarrow (0, 0, r_z) \tag{4.45}$$

i.e the Bloch vector is projected along z axis loosing x and y components.

The *bit-phase flip* channel has the operation elements

$$E_0 = \sqrt{p}I = \sqrt{p}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad E_1 = \sqrt{1-p}Y = \sqrt{p}\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \tag{4.46}$$

when this operation is done, the y component of the Bloch sphere remain intact. Whereas x, z components are shrinked by a factor of $1 - 2p$ as shown

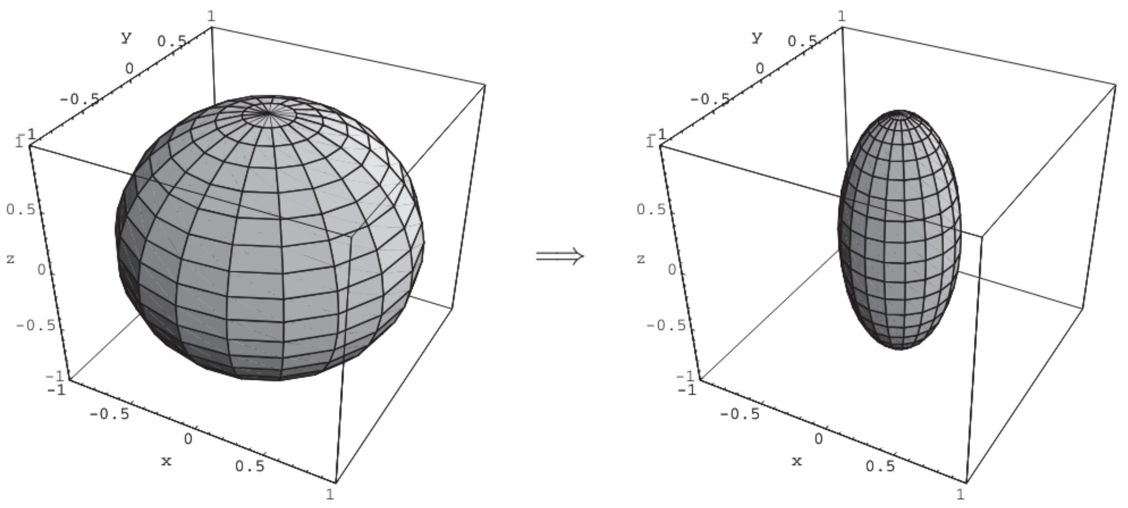


Figure 4.5: Bit-phase flip channel. Here p=0.3, it can e seen that y axis remains intact while x and z axes are shrinked by a factor of $1 - 2p$.

4.3.4 Depolarizing channel

It's an important type of quantum noise channel. We take a single qubit, and with probability p the qubit is *depolarized* i.e replaced by a completely mixed state $\frac{I}{2}$ or left untouched with probability $1 - p$. Thus the state of the qubit after noise is

$$\mathcal{E}(\rho) = \frac{pI}{2} + (1 - p)\rho \tag{4.47}$$

The effect on Bloch sphere is as shown

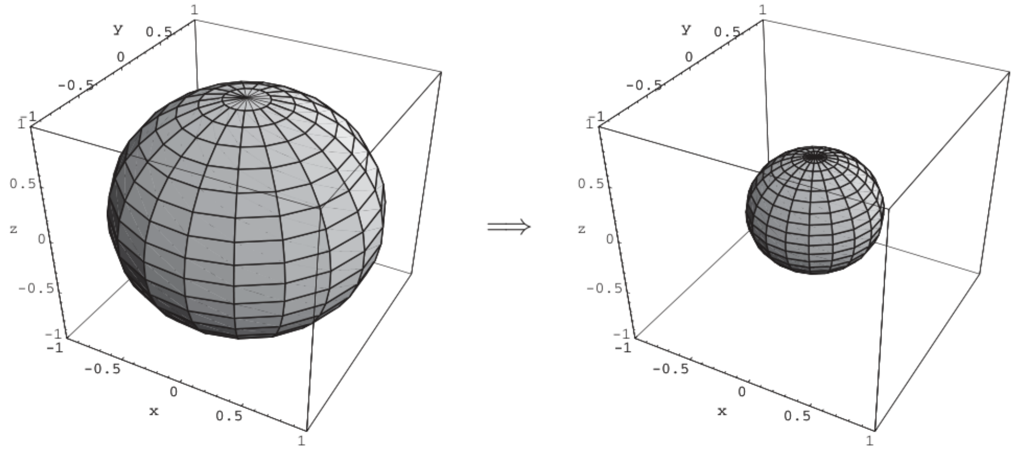


Figure 4.6: Effect of depolarizing channel on the Bloch sphere

The quantum circuit depicting the depolarizing channel is

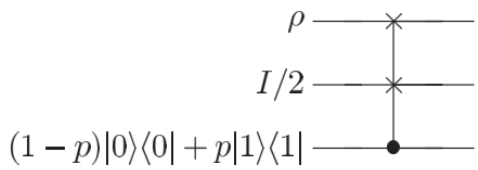


Figure 4.7: Quantum circuit representing the depolarizing channel

The third qubit is qubit is a mixture of state $|0\rangle$ and $|1\rangle$ with probability p and $1-p$ respectively. This is responsible for the second qubit getting swapped into the first qubit. In 4.47, $I/2$ can be represented as

$$\frac{I}{2} = \frac{\rho + X\rho X + Y\rho Y + Z\rho Z}{4} \tag{4.48}$$

Thus 4.47 becomes

$$\mathcal{E}(\rho) = \left(1 - \frac{3p}{4}\right)\rho + \frac{p}{4}\left(X\rho X + Y\rho Y + Z\rho Z\right) \tag{4.49}$$

Here the operation elements will be $\sqrt{1-3p/4}I$, $\sqrt{p}X/2$, $\sqrt{p}Y/2$, $\sqrt{p}Z/2$. Note that it is convenient to parametrize the depolarizing channel in different ways. The operation elements would differ, such as

$$\mathcal{E}(\rho) = (1-p)\rho + \frac{p}{3}\left(X\rho X + Y\rho Y + Z\rho Z\right) \tag{4.50}$$

Here, this does nothing with probability $1-p$ and applies X, Y or Z gates with probability $p/3$ each.

The depolarizing channel can be generalized for systems with dimensionality greater than 2. Suppose a system has dimension d the quantum operation would become

$$\mathcal{E}(\rho) = \frac{pI}{d} + (1-p)\rho \tag{4.51}$$

4.3.5 Amplitude damping

The process of *amplitude damping* helps us understand *energy dissipation* in quantum systems. For example, to describe dynamics of an atom emitting a photon, how a spin system at high temperature reaches equilibrium with it's environment etc. What's common in these examples is described by a quantum operation known as *amplitude damping*, which can be derived as follows. Suppose we have an optical mode at state $a\,|0\rangle + b\,|1\rangle$ which is superposition of zero or one photon. The scattering of a photon can be now described as keep a partially silvered mirror (a beamsplitter) in the path of the photon. It couples with another single optical node (it's environment) according to unitary transform $B = \exp\left[\theta(a^\dagger b - ab^\dagger\right]$. Where $a, a^\dagger$ and $b, b^\dagger$ are annihilation and creation operators for photons in two modes. The output after beamsplitter assuming there's no photon in the environment is $B\,|0\rangle\,(a\,|0\rangle + b\,|1\rangle) = a\,|00\rangle + b(\cos\theta\,|01\rangle + \sin\theta\,|10\rangle$. Tracing out the environment gives us the quantum operation

$$\mathcal{E}_{AD}(\rho) = E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger \tag{4.52}$$

where $E_k = \langle e_k|\, 0\, |e_0\rangle$ are

$$E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix} \tag{4.53}$$

$$E_1 = \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}, \tag{4.54}$$

the operations elements for amplitude damping. $\gamma = \sin^2(\theta)$ can be thought of as the probability of loosing a photon. E_1 operation changes the state $|1\rangle$ to $|0\rangle$, this corresponds to the loss of quantum of energy to the environment. E_0 leaves $|0\rangle$ unchanged but reduces the amplitude of $|1\rangle$, this happens because there's no loss of quantum of energy, thus the environment perceives it to be more likely that the system is in $|0\rangle$ state.

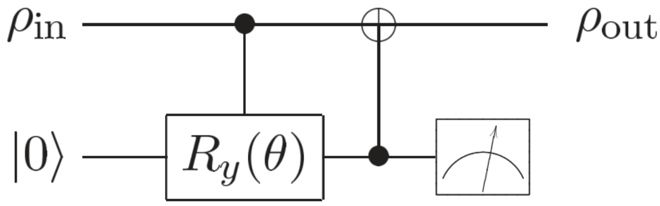


Figure 4.8: This models an amplitude damping circuit with $\gamma = \sin^2(\theta/2)$

A general characteristic of a quantum operation is the set of states that are left invariant under the operation, as we've seen before. Here, $|0\rangle$ is left invariant, but that is a cause of us assuming that the environment is starting at $|0\rangle$ state, as if it were at *zero temperature*.

The quantum operation defining amplitude damping at a finite temperature is known as *generalized amplitude damping*, $\mathcal{E}_{\text{GAD}}$ and is defined for single qubits using operation elements

$$E_0 = \sqrt{p}\begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix} \qquad E_1 = \sqrt{p}\begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix} \tag{4.55}$$

$$E_2 = \sqrt{1-p}\begin{bmatrix} \sqrt{1-\gamma} & 0 \\ 0 & 1 \end{bmatrix} \qquad E_3 = \sqrt{1-p}\begin{bmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{bmatrix} \tag{4.56}$$

and the stationary state ρ_∞ satisfies $\mathcal{E}(\rho_\infty) = \rho_\infty$. This GAD describes '$T_1$' relaxation processes due to coupling of spin to their surrounding lattice, a large system which is in thermal equilibrium at a temperature much higher than the spin temperature. This thing is useful in NMR quantum computation.

The effect of amplitude damping on Bloch sphere is

$$(r_x, r_y, r_z) \longrightarrow \left(r_x\sqrt{1-\gamma}, r_y\sqrt{1-\gamma}, \gamma + r_z(1-\gamma)\right) \tag{4.57}$$

where γ is replaced with a time varying function like $1 - e^{-t/T_1}$ (t is time and T_1 characterizes speed of the process). Using this we can visualize the effects as *flow* on the Bloch sphere, where every point moves towards a fixed point, something like north pole where $|0\rangle$ is located.

Similarly, generalized amplitude damping performs the transformation

$$(r_x, r_y, r_z) \longrightarrow \left(r_x\sqrt{1-\gamma}, r_y\sqrt{1-\gamma}, \gamma(2p-1) + r_z(1-\gamma)\right) \tag{4.58}$$

Generalized ammplitude damping is different than amplitude damping only in the location of fixed point of flow; the final state is along $\hat{z}$ axis, at the point $(2p-1)$, which is a mixed state.

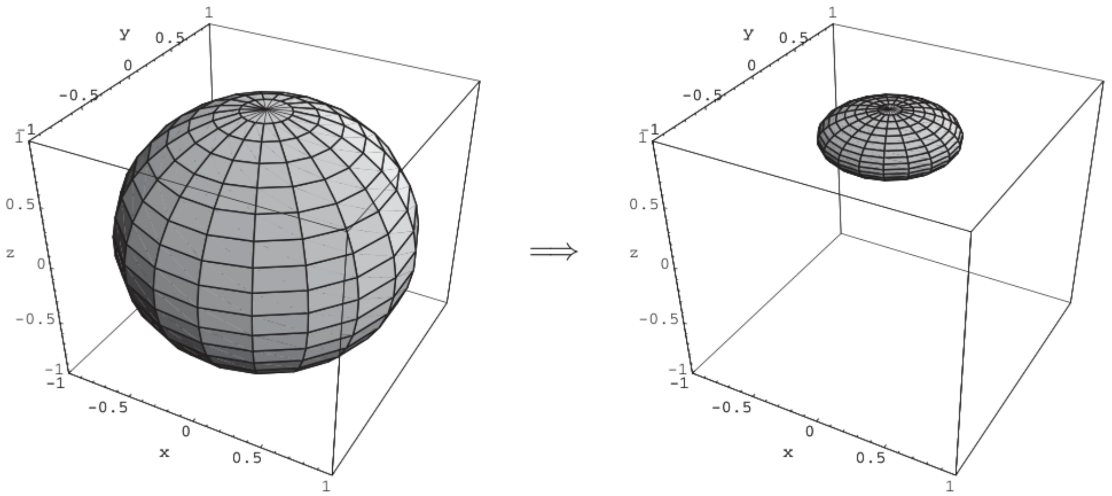


Figure 4.9: The effect of amplitude damping channel on Bloch sphere, for $p = 0.8$. Note how the entire sphere shrinks towards the north pole, the $|0\rangle$ state.

4.3.6 Phase damping

A noise process which is uniquely quantum mechanical, describes *loss of quantum information* without *loss of energy*, is ***phase damping***. For example it describes what happens when a photon scatters randomly as it travels through a waveguide, or how electronic states in an atom are perturbed upon interacting with distant electrical charges. The energy eigenstates don't change as a function of time but accumulate a phase proportional to the eigenvalues. When a system evolves for an amount of time not known, information about this quantum phase - the *relative* phases between the eigenstates is lost.

Part IV

Week 4

Chapter 5

Distance measures for quantum information

We'll try to quantify how two information items are similar, how information is preserved over a process etc. For this, we'll define two classes *distance measures*, *static distance* which is how close two quantum states are, and *dynamic distance* which is how well information is stored over a process. We'll define static distance properly first then build up dynamic distnance with it. Two widely used distance measures currently are *trace distance* and *fidelity*.

5.1 Distance measures for classical information

Hamming distance is defined to quantify distance between two bit strings. It's equal to number of mismatching bits (eg. Hamming$(1011, 0010) = 2$. But this relies on labelling which we don't have in our Hilbert space arena of quantum mechanics.

We can define a classical information source as a random variable over the set of possible outcomes. So we know the probabilities of getting a given output. In this notion, we can compare two information sources with same index set using a distance measure known as *trace distance*, defined as

$$D(p_x, q_x) = \frac{1}{2}\sum_x |p_x - q_x| \tag{5.1}$$

where $\{p_x\}$ and $\{q_x\}$ are probability distributions of both sources. This is also known as the L_1 *distance* or *Kolmogorov distance*. Trace because we use trace when we define this for quantum systems. This is a metric since it satisfies symmetricity ($D(x, y) = D(y, x)$) and triangle inequality ($D(x, y) \leq D(x, z) + D(y, z)$.

Fidelity, a second measure is described by

$$F(p_x, q_x) = \sum \sqrt{p_x q_x} \tag{5.2}$$

This is not a metric as shown by it's physical interpretation in figure 5.1.

A physically motivated operational meaning for trace distance which can be proved is,

$$D(p_x, q_x) = \max_S |p(S) - q(S)| = \max_S \left| \sum_{x\in S} p_x - \sum_{x\in S} q_x \right| \tag{5.3}$$

which is maximum over all subsets over index set. This makes sense because S represents an event which produces maximum difference in outputs. It can also be shown that the absolute

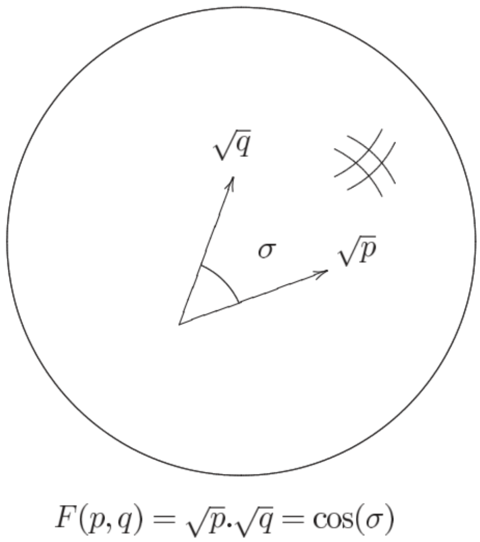


Figure 5.1: $F(p_x, q_x)$ can be seen as dot product between $\sqrt{p}$ and $\sqrt{q}$, each of them lie on a unit sphere since $\sum_x \left(\sqrt{p_x}\right)^2 = 1$ and $\sum_x \left(\sqrt{q_x}\right)^2 = 1$

value is redundant, i.e

$$D(p_x, q_x) = \max_S \left(\sum_{x \in S} p_x - \sum_{x \in S} q_x \right) \tag{5.4}$$

Trace distance and fidelity are static distance measures.

For *dynamic measure* of distance, suppose a random variable X is passed through a noise channel to produce Y shown by Markov process $X \longrightarrow Y$, a dynamic measure showing how much information is preserved is $p(X \neq Y)$. This can be done using trace distance, for which let's construct a copy $\tilde{X}$ of X which is also a random variable. Now X is passed through the channel, leaving output Y. The closeness between initial perfectly correlated pair $(\tilde{X}, X)$ and the pair $(\tilde{X}, Y)$ can be shown to be

$$D((\tilde{X}, X), (X, Y)) = p(X \neq Y) \tag{5.5}$$

This thing we did to calculate $p(X \neq Y)$ is unique to classical mechanics. We can't *directly* calculate quantum analogue of $p(X \neq Y)$ if X and Y exist at different times. We use quantum entanglement to define dynamic measure similar to the construction shown above.

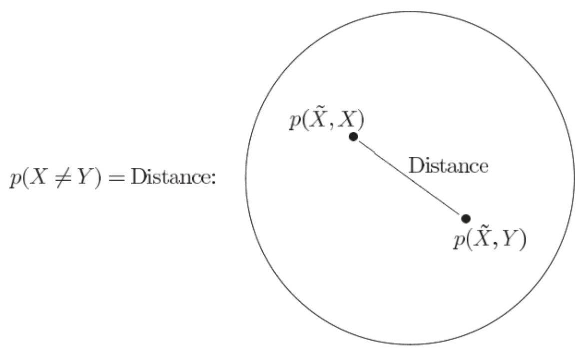


Figure 5.2: The probability of an error in the channel is equal to the trace distance between the probability distributions for $(\tilde{X}, X)$ and $(\tilde{X}, Y)$.

5.2 How close are two quantum states

We'll see quantum generalizations of trace distance and fidelity.

5.2.1 Trace distance

Trace distance between two quantum states ρ and σ is defined as

$$D(\rho,\sigma) = \frac{1}{2}\text{tr}|\rho - \sigma| \tag{5.6}$$

here $|A| = \sqrt{A^\dagger A}$. If ρ and σ commute (matrix multiplication), this reduces to classical trace distance. Suppose ρ and σ commute, then

$$\rho = \sum_i r_i \left|i\right\rangle \left\langle i\right| ; \quad \sigma = \sum_i s_i \left|i\right\rangle \left\langle i\right| \tag{5.7}$$

for orthogonal basis $|i\rangle$. Then trace distance becomes

$$\begin{aligned} D(\rho,\sigma) &= \frac{1}{2}\text{tr}|\rho - \sigma| &(5.8)\\ &= \frac{1}{2}\text{tr}|(r_i - s_i)\left|i\right\rangle \left\langle i\right|| &(5.9)\\ &= D(r_i, s_i) &(5.10)\end{aligned}$$

For qubits in Bloch sphere representation, it becomes nicer. Let

$$\rho = \frac{1 + \vec{r}\vec{\sigma}}{2}; \quad \sigma = \frac{1 + \vec{s}\vec{\sigma}}{2} \tag{5.11}$$

then trace distance between ρ and σ is

$$D(\rho,\sigma) = \frac{|\vec{r} - \vec{s}|}{2} \tag{5.12}$$

using the fact that $(\vec{r} - \vec{s}) \cdot \vec{\sigma}$ has eigen values $\pm|\vec{r} - \vec{s}|$. It can be shown that

$$D(\rho,\sigma) = \max_P \text{tr}(P(\rho - \sigma)) \tag{5.13}$$

where maximum is over all positive operators P, $P \leq I$.

Theorem 11. *Let $\{E_m\}$ me a POVM, with $p_m \equiv tr(\rho E_m)$ and $q_m \equiv tr(\sigma E_m)$ be the probability of getting output labelled by m. Then*

$$D(\rho,\sigma) = \max_{\{E_m\}} D(p_m, q_m) \tag{5.14}$$

where maximization is over all POVMs $\{E_m\}$.

It can be shown that this trace distance is also a metric. Here goes another nice theorem

Theorem 12. *Suppose $\mathcal{E}$ is a trace preserving quantum operation. Let ρ, σ be density operators. Then*

$$D(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq D(\rho,\sigma) \tag{5.15}$$

Thus there exists no physical processes which increases distance between two states.

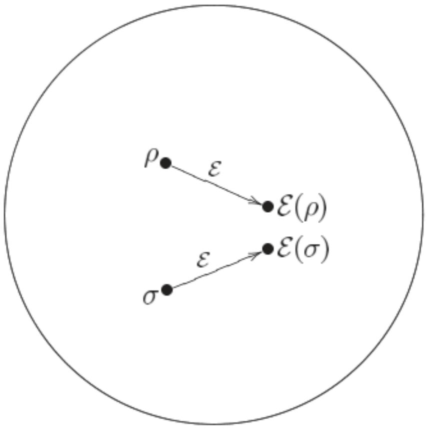


Figure 5.3: Trace-preserving quantum operations cause a contraction on the space of density operators.

It can also be showed that

$$D(\rho^A, \sigma^A) \leq D(\rho^{AB}, \sigma^{AB}) \tag{5.16}$$

intuitively, it's like things are less differentiable if corresponding parts of them are covered.

Theorem 13 (Strong convexity of trace distance). *Let $\{p_i\}$ and $\{q_i\}$ be two probability distributions over the same index set, and ρ_i, σ_i be density operators, also with indices from same index set, then*

$$D\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) \leq D(p_i, q_i) + \sum_i p_i D(\rho_i, \sigma_i) \tag{5.17}$$

where $D(p_i, q_i)$ is the trace distance between the probability distributions p_i and q_i.

using this we can show convexity properties of trace distance,

$$D\left(\sum_i p_i\rho_i, \sigma\right) \leq \sum_i p_i D(\rho_i, \sigma) \tag{5.18}$$

Also, any trace-preserving quantum operation $\mathcal{E}$ has a fixed point ρ such that $\mathcal{E}(\rho) = \rho$.

5.2.2 Fidelity

Fidelity is not a metric on density operators but does give rise to a useful metric. Fidelity of state ρ and σ is defined as

$$F(\rho, \sigma) \equiv \mathrm{tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}} \tag{5.19}$$

when ρ and σ commute, then

$$\rho = \sum_i p_i \left|i\right\rangle \left\langle i\right| ; \quad \sigma = \sum_i s_i \left|i\right\rangle \left\langle i\right| \tag{5.20}$$

for orthoonormal basis $|i\rangle$, we see that

$$F(\rho,\sigma) = \mathrm{tr}\sqrt{\sum_i r_i s_i |i\rangle\langle i|} \tag{5.21}$$

$$= \mathrm{tr}\left(\sum_i \sqrt{r_i s_i} |i\rangle\langle i|\right) \tag{5.22}$$

$$= \sum_i \sqrt{r_i s_i} \tag{5.23}$$

$$= F(r_i, s_i) \tag{5.24}$$

thus, when ρ and σ commute, quantum fidelity reduces to classical fidelity between eigenvalue distributions r_i, s_i of ρ, σ respectively. Also the fidelity between a pure state $|\psi\rangle$ and an arbitrary state ρ is

$$F(|\psi\rangle, \rho) = \mathrm{tr}\sqrt{\langle\psi|\rho|\psi\rangle |\psi\rangle\langle\psi|} \tag{5.25}$$

$$= \sqrt{\langle\psi|\rho|\psi\rangle} \tag{5.26}$$

i.e fidelity is square root of overlap between $|\psi\rangle$ and ρ.

There's no similar Bloch sphere representation for fidelity but it does satisfy similar properties like *invariance under unitary transformation* i.e

$$F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho,\sigma) \tag{5.27}$$

Theorem 14 (Uhlmann's theorem). *Suppose ρ, σ be states of a system Q, Let R be a copy of Q. Then*

$$F(\rho,\sigma) = \max_{|\psi\rangle,|\varphi\rangle} |\langle\psi|\varphi\rangle| \tag{5.28}$$

where the maximization is over all purifications $|\psi\rangle$ of ρ and $|\varphi\rangle$ of σ.

Lemma 1. *Let A be any operator, U be unitary. Then*

$$|tr(AU)| \le |tr(A)| \tag{5.29}$$

equality is when $U = V^\dagger$ when $A = |A|V$ is polar decomposition of A.

By Uhlmann's formula it can be shown that fidelity is *symmetric* in its inputs i.e $F(\rho,\sigma) = F(\sigma,\rho)$ and also that it's bound, $0 \le F(\rho,\sigma) \le 1$. Also $\rho = \sigma \implies F(\rho,\sigma) = 1$ else it's less than 1. Quantum fidelity is related to classical fidelity in the following way:

$$F(\rho,\sigma) = \min_{E_m} F(p_m, q_m) \tag{5.30}$$

where E_m are POVMs, $p_m = \mathrm{tr}(\rho E_m)$, $q_m = \mathrm{tr}(\sigma E_m)$ are probability distributions corresponding to the POVM $\{E_m\}$.

We can define *angle* between states ρ and σ by

$$A(\rho,\sigma) = \arccos F(\rho,\sigma) \tag{5.31}$$

This is non-negative, symmetric in inputs and zero when $\rho = \sigma$. It can be shown to satisfy triangle inequality, and thus is a metric.

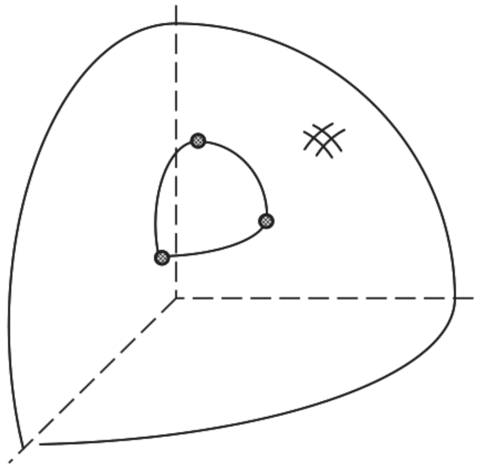


Figure 5.4: Angle between unit vectors is a metric

Qualitatively, fidelity behaves like "upside down" of trace distance. It decreases as two states become more distinguishable and vice versa. Fidelity in *non decreasing* as shown by the theorem

Theorem 15 (Monotonicity of fidelity). *Suppose ρ, σ be density operators, and $\mathcal{E}$ is a trace-preserving operation, then*

$$F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \geq F(\rho, \sigma) \tag{5.32}$$

Angles we've defined above follow **contractivity** which is $A(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq A(\rho, \sigma)$, there's also *strong concavity* of fidelity like trace distance

Theorem 16 (Strong concavity of fidelity). *Let p_i, q_i be probability distributions over the same index set, ρ_i, σ_i also indexed by the same index set. Then*

$$F\left(\sum_i p_i\rho_i, q_i\sigma_i\right) \geq \sum_i \sqrt{p_i q_i} F(\rho_i, \sigma_i) \tag{5.33}$$

5.2.3 Relationships between distance measures

Despite their different forms, trace distance and fidelity are closely related. In pure states, they're completely equivalent to each other, to see this consider two pure state $|a\rangle$ and $|b\rangle$, using Gram-schmidt we can find states $|0\rangle$ and $|1\rangle$ such that $|a\rangle = |0\rangle$ and $|b\rangle = \cos\theta\,|0\rangle + \sin\theta\,|1\rangle$. Then $F(|a\rangle, |b\rangle) = |\cos\theta|$ and

$$D(|a\rangle, |b\rangle) = \frac{1}{2}\mathrm{tr}\left|\begin{bmatrix} 1-\cos^2\theta & -\cos\theta\sin\theta \\ -\cos\theta\sin\theta & -\sin^2\theta \end{bmatrix}\right| \tag{5.34}$$

$$= |\sin\theta| \tag{5.35}$$

$$= \sqrt{1 - F(|a\rangle, |b\rangle)^2} \tag{5.36}$$

For mixed states ρ, σ let $|\psi\rangle$ and $|\varphi\rangle$ be purifications of ρ, σ such that $F(\rho, \sigma) = |\langle\psi|\varphi\rangle| = F(|\psi\rangle, |\varphi\rangle)$. Since trace distance is non-increasing under the partial trace,

$$D(\rho, \sigma) \leq D(|\psi\rangle, |\varphi\rangle) \tag{5.37}$$

$$= \sqrt{1 - F(\rho, \sigma)^2} \tag{5.38}$$

It can be further shown that

$$1 - F(\rho, \sigma) \leq D(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)^2} \tag{5.39}$$

5.3 How well does a quantum channel preserve information?

We'll try to quantify how much that state has "changed" from $|\psi\rangle$ to $\mathcal{E}(|\psi\rangle\langle\psi|)$ due to an operation $\mathcal{E}$ using the distance measures discussed. Let's look at a simple example how a state changes under depolarizing channel using fidelity

$$F(|\psi\rangle, \mathcal{E}(|\psi\rangle\langle\psi|)) = \sqrt{\langle\psi| \left(p\frac{I}{2} + (1-p)|\psi\rangle\langle\psi|\right)|\psi\rangle} \tag{5.40}$$

$$= \sqrt{1 - \frac{p}{2}} \tag{5.41}$$

it can be clearly seen when we reduce p, fidelity get's close to one hence the states becoming more indistinguishable. There's nothing special about fidelity that trace distance doesn't have but we'll stick to this. But in real quantum systems, we don't even know initial state $|\psi\rangle$ in advance so we try to *minimize* fidelity to get worst-case measure

$$F_{\min}(\mathcal{E}) \equiv \min_{|\psi\rangle} F(|\psi\rangle, \mathcal{E}(|\psi\rangle\langle\psi|)) \tag{5.42}$$

for depolarizing channel it doesn't matter and it's just $\sqrt{1-p/2}$. But for phase damping channel

$$\mathcal{E}(\rho) = p\rho + (1-p)Z\rho Z \tag{5.43}$$

The fidelity will be

$$F(|\psi\rangle, \mathcal{E}(|\psi\rangle\langle\psi|)) = \sqrt{\langle\psi|\left(p|\psi\rangle\langle\psi| + (1-p)Z|\psi\rangle\langle\psi|Z\right)|\psi\rangle} \tag{5.44}$$

$$= \sqrt{p + (1-p)\langle\psi|Z|\psi\rangle^2} \tag{5.45}$$

Now this minimizes when $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$. Thus for phase damping, minimum fidelity becomes

$$F_{\min}(\mathcal{E}) = \sqrt{p} \tag{5.46}$$

We haven't considered mixed states, but using joint concavity of fidelity it can be shown that allowing mixed states doesn't change $F_{\min}$. Now, we're not only interesting in saving quantum states while transmitting them but we should also look up on when we're computing for example implementing a quantum gate described by U. Thus we define *gate-fidelity* as

$$F(U, \mathcal{E}) = \min_{|\psi\rangle} F(U|\psi\rangle, \mathcal{E}(|\psi\rangle\langle\psi|)) \tag{5.47}$$

For example, if we're implementing NOT gate, but instead implement a noisy operation $\mathcal{E}(\rho) = (1-p)X\rho X + pZ\rho Z$, then gate fidelity is

$$F(X, \mathcal{E}) = \min_{|\psi\rangle} \sqrt{\langle\psi| X\left((1-p)X|\psi\rangle\langle\psi|X + pZ|\psi\rangle\langle\psi|Z\right)X|\psi\rangle} \tag{5.48}$$

$$= \min_{|\psi\rangle} = \sqrt{(1-p) + p\langle\psi|Y|\psi\rangle^2} \tag{5.49}$$

$$= \sqrt{1-p}. \tag{5.50}$$

5.3.1 Quantum sources of information and the entanglement fidelity

Even though we've talked a lot about quantum sources we haven't actually defined it. One interesting definition is imagining a stream of quantum systems being produced by some physical process, with their states represented as $\rho_{X_1}, \rho_{X_2}, \ldots$, where X_j are identically distributed random variables and ρ_j is some fixed set of operators. For example, a stream of qubits preparing $|0\rangle$ with $1/2$ probability and $(|0\rangle + |1\rangle)/\sqrt{2}$ with another $1/2$ of probability.

This *ensemble* notion of quantum source leads to notion of *ensemble average fidelity* which captures the idea that source is well preserved under the action of a noisy channel described by trace preserving operator $\mathcal{E}$, as

$$F = \sum_j p_j F(\rho_j, \mathcal{E}(\rho_j))^2, \tag{5.51}$$

Here, $F = 1$ if and only if $\mathcal{E}(\rho_j) = \rho_j$ for all j such that $p_j > 0$

The second notion of quantum source is motivated by the fact that a channel preserving information also preserves *entanglement* well. We'll use quantum analogue of what we've considered in 5.5. We'll suppose a quantum system Q entangled with a fictious system R, such that the joint state QR is pure. It turns out the result doesn't depend on how this purification is done. Then the system Q is subjected to dynamics described by $\mathcal{E}$.

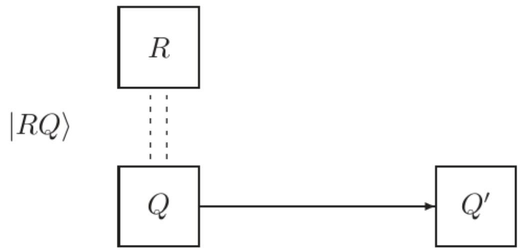


Figure 5.5: The RQ picture of quantum channel. The initial state of RQ is a pure state.

How well is the entanglement preserved by the quantum operation $\mathcal{E}$ is described by *entanglement fidelity* $F(\rho, \mathcal{E})$ which is defined for trace preserving operations $\mathcal{E}$ by

$$F(\rho, \mathcal{E}) \equiv F(RQ, R'Q')^2 \tag{5.52}$$

$$= \langle RQ| \left[(\mathcal{I}_R \otimes \mathcal{E})(|RQ\rangle \langle RQ|)\right] |RQ\rangle \tag{5.53}$$

where prime implies after the operation. Entanglement fidelity close to 1 implies well preserved state and vice versa. One of the important properties of entanglement fidelity is that there's a nice formula to calculate it exactly. Suppose E_i is a set of operation elements for $\mathcal{E}$. Then

$$F(\rho, \mathcal{E}) = \langle RQ| \rho^{R'Q'} |RQ\rangle = \sum_i |\langle RQ|E_i|RQ\rangle|^2. \tag{5.54}$$

Suppose $|RQ\rangle = \sum_j \sqrt{p_j}\, |j\rangle\, |j\rangle$, where $\rho = \sum_j |j\rangle \langle j|$. Then it can be proved that

$$\langle RQ| E_i |RQ\rangle = \mathrm{tr}(E_i \rho) \tag{5.55}$$

Substituting this is equation 5.54 we get a useful formula

$$F(\rho, \mathcal{E}) = \sum_i |\mathrm{tr}(\rho E_i)|^2 \tag{5.56}$$

Surprisingly, these two notions are closely related too. This can be seen in two useful properties. First, the entanglement fidelity is a lower bound on the square of the static fidelity between the output and input to the process,

$$F(\rho, \mathcal{E}) \leq [F(\rho, \mathcal{E}(\rho)]^2 \tag{5.57}$$

This states that it's harder to preserve entanglement than it is compared to pure states. The second property is that it is a *convex* function of ρ. i.e

$$F\left(\sum_j p_j \rho_j, \mathcal{E}\right) \leq \bar{F} \tag{5.58}$$

This shows that any channel $\mathcal{E}$ which does a good job of preserving entanglement between a source system ρ and reference system will automatically preserve the ensemble source described by probabilities p_j and states ρ_j such that $\sum_j p_j \rho_j = \rho$.

I'll try to conclude with most important points yet.

1. $0 \leq F(\rho, \mathcal{E}) \leq 1$. From properties of static fidelity.

2. Entanglement fidelity is linear in quantum operation input. This is immediate too.

3. For pure state inputs, entanglement fidelity is square of static fidelity between input and output,
$$F(|\psi\rangle, \mathcal{E}) = F(|\psi\rangle, \mathcal{E}(|\psi\rangle \langle\psi|))^2. \tag{5.59}$$

4. $F(\rho, \mathcal{E}) = 1$ if and only if for all pure state $|\psi\rangle$ in support of ρ
$$\mathcal{E}(|\psi\rangle \langle\psi|) = |\psi\rangle \langle\psi| . \tag{5.60}$$

5. Suppose that $\langle\psi| \mathcal{E}(|\psi\rangle \langle\psi|) |\psi\rangle \leq 1 - \eta$ for all $|\psi\rangle$ in support of ρ, for some η then $F(\rho, \mathcal{E}) \geq 1 - (3\eta/2)$.

Part V

Week 5

Chapter 6

Entropy and information

We know that *entropy* measures uncertainity of state of a physical system. We'll look into it's definitions and properties deeply both in classical and quantum information theory.

6.1 Shannon entropy

It's the key concept of classical information theory. It can be viewed in two complementary views, first, suppose we have a random variable X, how much information we *gain* on average if we know the value of X. Second, it's the measure of *uncertainity* before we know the value of X.

Information content of a random variable doesn't depend on labels. For example, a random variaible giving 'heads' and 'tails' with probability $1/4$ and $3/4$ contains same information as a random variable giving '0' and '1' with probability $1/4$ and $3/4$ respectively. Thus, entropy also doesn't depend on labels. *Shannon entropy* associated with a probability distribution $p_1, \ldots, p_n$ is given by

$$H(X) \equiv H(p_1, \ldots, p_n) \equiv -\sum_x p_x \log p_x \tag{6.1}$$

Note that by $\log$ we mean base 2, by $\ln$ we mean base e. Since a never occuring event with probability $p_i = 0$ should never affect entropy, thus we agree that $0 \log 0 = 0$ even though we mean $\lim_{x \to 0} x \log x = 0$. There's a nice reason why entropy is defined this way. We have to *quantify the resources needed to store information*. Let's suppose a source producing symbols $X_1, X_2, \ldots$ of independent *identically* distributed random variables, minimal amount of resources required to store the information produced by source such that it can be reconstructed turns out to be just entropy i.e $H(X) \equiv H(X_1) \equiv H(X_2) = \ldots$, this result is known as *Shannon's noiseless coding theorem*.

I suggest looking at an example to understand this. Suppose our source produces $1, 2, 3, 4$ which are symbols but with probability $1/2, 1/4, 1/8$ and $1/8$ respectively. We don't need two bits to store output of each use of source. We can compress this usage, by representing 1 as 0, 2 as 10, 3 as 110 and 4 as 111, with this we can reconstruct the message, for eg. 11010111 as 324 too. We need on an average $\frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{8} \cdot 3 = 7/4$ bits only! It also turns out that this is just the entropy of the source $H(X) = -1/2 \log(1/2) - 1/4 \log(1/4) - 1/8 \log(1/8) - 1/8 \log(1/8) = 7/4$. It's also true that average information gained by an event out of a probability distribution $p_1, \ldots, p_n$ is $k \sum_i p_i \log p_i$ for some constant k.

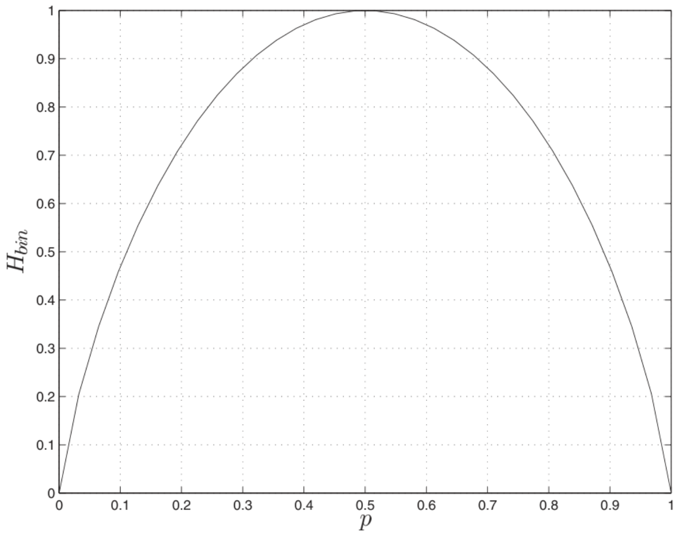


Figure 6.1: Binary entropy function $H(p)$

6.2 Basic properties of entropy

6.2.1 The binary entropy

The entropy of a two-outcome random variable is so special that we name it as *binary entropy*

$$H_{\text{bin}}(X) \equiv -p\log p - (1-p)\log(1-p) \tag{6.2}$$

where p and $1-p$ are probabilities of outcome. It's graph is shown in figure 6.1

A really good intuition comes when we consider this example of Alice having two biased coins one from US and another from Australia. Suppose the US coin gives head with probability p_U and the Australian coin with probability p_A and Alice flips one of those coins with probability q for US coin and $1-q$ for Australian coin and she tells final 'head' or 'tail' to Bob. The information Bob gets is atleast the average information he gets from flipping one coin. Hence

$$H(qp_U + (1-p)p_A) \geq qH(p_U) + (1-q)H(p_A) \tag{6.3}$$

It's greater because Bob gets *more* information about the country sometimes. For eg. if $p_U = 0.1$ and $p_A = 0.9$ and Alice says head, Bob would understand that it's more probable to be an aussie coin. This leads us into a very important property of entropy, *concavity*. Which can be seen in binary case in the figure 6.1.

6.2.2 The relative entropy

The *relative entropy* is useful entropy-like measure describing closeness of two probability distributions $p(x)$ and $q(x)$ over the same index set x, defined by

$$H(p(x) \parallel q(x)) \equiv \sum_x p(x)\log\frac{p(x)}{q(x)} \equiv -H(X) - \sum_x p(x)\log q(x) \tag{6.4}$$

The following theorem motivates us to understand it as a distance measure

Theorem 17 (Non-negativity of relative entropy). *The relative entropy is non-negative, $H(p(x) \parallel q(x)) \geq 0$ with equality if and only if $p(x) = q(x)$ for all x.*

The proof is simple using the definition of relative entropy and $\log x \ln 2 \leq x - 1$ with equality only when $x = 1$. Relative entropy itself is not very useful but used to study entropy, for example let $q(x)$ be uniform probability distribution over index set X of size d then for a probability distribution $p(x)$ over X, we have

$$H(p(x) \parallel q(x)) = H(p(x) \parallel 1/d) = -H(X) - \sum_x p(x) \log(1/d) = \log d - H(X) \quad (6.5)$$

and hence by theorem 17 we have $\log d - H(X) \geq 0$ thus $H(X) \leq \log d$, a nice property hence a theorem!

Theorem 18. *Suppose X is a random variable with d outcomes. Then $H(X) \leq \log d$, with equality if and only if X is uniformly distributed over those d outcomes.*

6.2.3 Conditional entropy and mutual information

For two random variables X, Y we'll try to understand how information of X is related to that of Y. First we define *joint entropy* of X and Y as

$$H(X, Y) \equiv -\sum_{x,y} p(x, y) \log p(x, y) \quad (6.6)$$

The *entropy of X conditional on knowing Y* is similar to that in probability, it's the uncertainity that we have in the value of X given that we know the value of Y. It's defined by

$$H(X|Y) \equiv H(X, Y) - H(Y) \quad (6.7)$$

A second quantity *mutual information content of X and Y* measures how much information X and Y have in common. It's very similar to intersection, given we take $H(X, Y)$ as information content of both X and Y which is counted twice. Thus, mutual information content is

$$H(X : Y) = H(X) + H(Y) - H(X, Y) \quad (6.8)$$

The relation $H(X : Y) = H(X) - H(X|Y)$ also comes in handy. Let's see some more simple relationships between different entropies.

Theorem 1 (Basic properties of Shannon entropy)

1. $H(X, Y) = H(Y, X), H(X : Y) = H(Y : X)$
2. $H(Y|X) \geq 0$ and thus $H(X : Y) \leq H(Y)$ with equality if and only if Y is a function of X, $Y = f(X)$.
3. $H(X) \leq H(X, Y)$ with equality if and only if Y is a function of X.
4. **Subadditivity:** $H(X, Y) \leq H(X) + H(Y)$ with equality if and only if X and Y are independent variables.
5. $H(Y|X) \leq H(Y)$ and thus $H(X : Y) \geq 0$ with equality in each if and only if X and Y are independent distributions.

6. **Strong subadditivity:** $H(X,Y,Z)+H(Y)\leq H(X,Y)+H(Y,Z)$, with equality if and only if $Z\to Y\to X$ forms a markov chain.
7. **Conditioning reduces entropy:** $H(X|Y,Z)\leq H(X|Y)$.

All this can be easily understood by a Venn diagram for entropy.

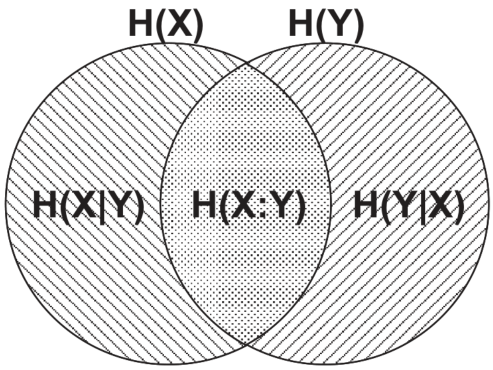


Figure 6.2: Relationships between different entropies

Let's finally consider a simple and useful chaining rule for conditional entropies.

Theorem 19. *Let $X_1,\ldots,X_n$ and Y be any set of random variables. Then*

$$H(X_1,\ldots,X_n|Y)=\sum_{i=1}^{n}H(X_i|Y,X_1,\ldots,X_{i-1}) \tag{6.9}$$

Proof. Use induction with base case $n=2$. □

6.2.4 The data processing inequality

This states that information from the output of a source can only *decrease with time*; once information is lost, it's gone forever. The intuitive notion of *information processing* is captured by *Markov chanis* of random variables. A Markov chain is a sequence $X_1\to X_2\to\cdots$ such that X_{n+1} is independent of $X_1,X_2,\ldots,X_{n-1}$ if X_n is known. In other words

$$p(X_{n+1}=x_{n+1}|X_n=x_n,\ldots,X_1=x_1)=p(X_{n+1}=x_{n+1}|X_n=x_n) \tag{6.10}$$

Theorem 20. *Suppose $X\to Y\to Z$ is a Markov chain. Then*

$$H(X)\geq H(X:Y)\geq H(X:Z). \tag{6.11}$$

Moreover, this first inequality becomes equal if and only if, given Y, it's possible to reconstruct X.

This theorem is intuitive, as it says that if there's some noise introduced in X (which leads to Markov chain $X\to Y$) then the mutual information between X and Y (which is what we want) is irretrievably lost, since it's not possible to reconstruct X from Y (noise).

If $X\to Y\to Z$ is a Markov chain, then so is $Z\to Y\to X$[1]. Thus with above conditions, we see that

$$H(Z:Y)\geq H(Z:X) \tag{6.12}$$

This thing is also known as *data pipelining inequality* i.e any information Z shares with X is also shared by Y. The information is *pipelined* from X to Y to Z.

[1]Can be proved by using $P(X|Y)=P(X,Y)/P(Y)$

6.3 Von Neumann entropy

Shannon entropy is for classical probability distributions, Von Neumann defined this notion of *entropy* for quantum state with density operator ρ as

$$S(\rho) \equiv -\text{tr}(\rho \log \rho) \tag{6.13}$$

where $\log$ still means base 2 and $0 \log 0 \equiv 0$. This can be simplified if we know eigenvalues λ_x of ρ as

$$S(\rho) \equiv -\sum_x \lambda_x \log \lambda_x \tag{6.14}$$

This is very useful for calculations. We will refer *entropy* to either Shannon or Von Neumann entropy based on context.

6.3.1 Quantum relative entropy

Similar to Shannon entropy, *relative entropy* between two state ρ and σ is defined by

$$S(\rho \parallel \sigma) \equiv \text{tr}(\rho \log \rho) - \text{tr}(\rho \log \sigma) \tag{6.15}$$

as in classical case, this can sometimes shoot upto infinity. Particularly, relative entropy is defined to be $+\infty$ if the kernel of σ (vector space spanned by eigenvectors of σ with eigenvalue 0) has non-trivial intersection with the support of ρ (vector space spanned by eigenvectors of ρ with non-zero eigenvalues). A familiar inequality as a theorem is given.

Theorem 21 (Klein's inequality). *The quantum relative entroy is non-negative. i.e*

$$S(\rho \parallel \sigma) \geq 0, \tag{6.16}$$

with equality if and only if $\rho = \sigma$.

6.3.2 Basic properties of entropy

These are few interesting properties of von Neumann entropy:

Theorem 2 (Basic properties of von Neumann entropy)

1. The entropy is non-negative. The entropy is zero if and only if the state is pure.
2. In a d-dimensional Hilbert space, entropy is atmost $\log d$. It's equal if and only if the system is completely mixed state I/d.
3. If a composite system AB is pure, then $S(A) = S(B)$.
4. Suppose p_i are probabilities and ρ_i have support on orthogonal subspaces. Then

$$S\left(\sum_i p_i \rho_i\right) = H(p_i) + \sum_i p_i S(\rho_i) \tag{6.17}$$

5. **Joint entropy theorem:** Suppose p_i are probabilities and $|i\rangle$ are orthogonal states for a system A, and ρ_i is any set of density operators for another system, B. Then

$$S\left(\sum_i p_i |i\rangle \langle i| \otimes \rho_i\right) = H(p_i) + \sum_i p_i S(\rho_i). \tag{6.18}$$

Again similarly, the *joint entropy* $S(A,B)$ for a composite system of two components A and B is defined obviously as $S(A,B) \equiv -\text{tr}(\rho^{AB}\log\rho^{AB})$, we also define conditional entropy and mutual information as

$$S(A|B) = S(A,B) - S(B) \tag{6.19}$$
$$S(A:B) = S(A) + S(B) - S(A,B) \tag{6.20}$$
$$= S(A) - S(A|B) = S(B) - S(B|A). \tag{6.21}$$

But there are some properties which aren't carried over from Shannon to von Neumann entropy. One such is $H(X,Y) \geq H(X)$ which is intuitive since we have more information if we know about X and Y compared to if we only know about X. But this isn't true for von Neumann entropy, consider a system AB of two qubits in entangled state $(|00\rangle + |11\rangle)/\sqrt{2}$. Since this is pure, $S(A,B) = 0$. But it's component A is $I/2$, and has entropy 1 defying what we've seen above. Another way of seeing this is $S(B|A) = S(A,B) - S(A)$ is negative.

6.3.3 Measurements and entropy

We'll try to understand how measurements change the entropy of a system, if they do. This change ofcourse depends on the type of measurement we make. Suppose we do a projective measurement on system ρ described by projectors P_i, the state of the system is

$$\rho' = \sum_i P_i \rho P_i \tag{6.22}$$

The following theorem shows the entropy never decreases

Theorem 22 (**Projective measurements increase entropy**). *Suppose P_i is a complete set of orthogonal projectors and ρ is a density operator. Then the entropy of the state $\rho' \equiv \sum_i P_i \rho P_i$ of the system after measurement is atleast as great as the original entropy,*

$$S(\rho') \geq S(\rho) \tag{6.23}$$

with equality if and only if $\rho' = \rho$.

6.3.4 Subadditivity

If two system A and B have a joint state ρ^{AB}. Then the joint entropy for the two systems satisfy

$$S(A,B) \leq S(A) + S(B) \tag{6.24}$$
$$S(A,B) \geq |S(A) - S(B)| \tag{6.25}$$

The first relation is known as *subadditivity* inequality[2] for von Neumann entropy, with equality if and only if A and B are correlated, i.e $\rho^{AB} = \rho^A \otimes \rho^B$. Second one is known as *triangle* inequality or *Araki-Lieb* inequality. It's the quantum analogue of $H(X,Y) \geq H(X)$.

[2] This can be proved simply using Klein's inequality.

6.3.5 Concavity of the entropy

The entropy is a *concave* function in its inputs. That is, given the probabilities p_i, and corresponding density operators ρ_i, entropy satisfies the inequality[3]

$$S\left(\sum_i p_i\rho_i\right) \geq \sum_i p_i S(\rho_i). \tag{6.26}$$

A good intuition is that $\sum_i p_i\rho_i$ expresses a quantum state which is in an uncertain state ρ_i with probability p_i, this uncertainity of index i of state adds to the average uncertainity we have of each state ρ_i which is $\sum_i p_i S(\rho_i)$. Equality holds if and only if all the states ρ_i for which $p_i > 0$ are identical; i.e, entropy is strictly concave in its inputs!

6.3.6 The entropy of a mixture of quantum states

The flip side of concavity is that it provides an upper bound of entropy of a mixture of quantum states too. It's shown by the following theorem.

Theorem 23. *Suppose $\rho = \sum_i p_i\rho_i$, where p_i are some set of probabilities, and ρ_i are density operators. Then*

$$S(\rho) \leq \sum_i p_i S(\rho_i) + H(p_i), \tag{6.27}$$

with equality if and only if the states ρ_i have support on orthogonal subspaces.

This along with concavity completes our bound on entropy of a mixture of quantum states,

$$\sum_i p_i S(\rho_i) \leq S\left(\sum_i p_i\rho_i\right) \leq \sum_i p_i S(\rho_i) + H(p_i). \tag{6.28}$$

The intuition for this comes from the fact that we know too much information if we know on average about ρ_i and p_i also.

6.4 Strong subadditivity

This is an extension of subadditivity and triangle inequalities to three systems. The inequality states that for three quantum systems A, B, C,

$$S(A, B, C) + S(B) \leq S(A, B) + S(B, C) \tag{6.29}$$

6.4.1 Proving strong subadditivity

The proof is based on an important result, the *Lieb's theorem*. For which, let's define few things. A function $f(A, B)$ which is a real-valued function of two matrices, is said to be *jointly concave* in A, B if for all $0 \leq \lambda \leq 1$,

$$f(\lambda A_1 + (1-\lambda)A_2, \lambda B_1 + (1-\lambda)B_2) \geq \lambda f(A_1, B_1) + (1-\lambda)f(A_2, B_2). \tag{6.30}$$

Let's state Lieb's theorem now,

[3]To prove this we introduce an auxillary system B.

Theorem 24 (Lieb's theorem). *Let X be a matrix, and $0 \leq t \leq 1$. Then the function*

$$f(A, B) \equiv tr(X^\dagger A^t X B^{1-t}) \tag{6.31}$$

is jointly concave in positive matrices A and B.

This theorem implies another theorem, convexity of the relative entropy.

Theorem 25 (Convexity of the relative entropy). *The relative entropy $S(\rho \parallel \sigma)$ is jointly convex in its arguments.*

Corollary 25.1 (Concavity of the quantum conditional entropy). *Let AB be a composite quantum system with components A and B. Then the conditional entropy $S(A|B)$ is concave in the state ρ^{AB} of AB.*

This finally leads us to strong subadditivity.

Theorem 26. *For any trio of quantum systems A, B, C the inequalities*

$$S(A) + S(B) \leq S(A, C) + S(B, C) \tag{6.32}$$

$$S(A, B, C) + S(B) \leq S(A, B) + S(B, C) \tag{6.33}$$

hold.

6.4.2 Elementary applications of strong subadditivity

Strong subadditivity and related results have many useful applications in quantum information theory. Take the inequality $S(A) + S(B) \leq S(A, C) + S(B, C)$ which is true, it also holds for Shannon entropy too, but for different reasons. For practical applications strong subadditivity is most easily applied as a rephrasing in terms of conditional and mutual entropies. Here are three major reformulations

Theorem 3 (Reformulations of strong subadditivity)

1. **Conditioning reduces entropy:** Suppose ABC is a composite system. Then $S(A|B, C) \leq S(A|B)$.

2. **Discarding quantum systems never increases mutual information:** Suppose ABC is a composite quantum system. Then $S(A : B) \leq S(A : B, C)$.

3. **Quantum operations never increase mutual information:** Suppose AB be a composite system and $\mathcal{E}$ is a trace preserving quantum operation on B. Let $S(A : B)$ be before operation, $S(A' : B')$ be after operation, then $S(A' : B') \leq S(A : B)$.

We saw before that Shannon mutual information is not subadditive, thus quantum mutual information isn't subadditive either. But it turns out that conditional entropy is subadditive, i.e

$$S(A_1, A_2|B_1, B_2) \leq S(A_1|B_1) + S(A_2|B_2) \tag{6.34}$$

What's more, it's subadditive in the first and second entries.

Theorem 27 (**Subadditivity of conditional entropy**). *Let $ABCD$ be a composite of four systems. Then the conditonal entropy is jointy subadditive in the first two entries:*

$$S(A,B|C,D) \leq S(A|C) + S(B|D) \tag{6.35}$$

Let ABC be a composite of three quantum systems. Then conditional entropy is subadditive in each of first two entries:

$$S(A,B|C) \leq S(A|C) + S(B|C) \tag{6.36}$$
$$S(A|B,C) \leq S(A|B) + S(A|C) \tag{6.37}$$

We got introduced to relative entropy like a measure between density operators. It seems intuitive that this relative entropy decreases when we ignore a part of system is ignored, i.e it makes it harder to distinguish two states of that system and thus decrease any reasonable measure of distance between them. This following theorem states it:

Theorem 28. *Let ρ^{AB} and σ^{AB} be any two density matrices of a composite system AB. Then*

$$S(\rho^A \parallel \rho^A) \leq S(\rho^{AB} \parallel \rho^{AB}). \tag{6.38}$$

Part VI

Week 6

Chapter 7

Quantum circuits

7.1 Quantum algorithms

Many interesting problems are impossible to solve on a classical computer, not because they're in principle insoluble, but because of the astronomical amount of resources they'd need. Quantum computing promises to enable feasible algorithms which would require unreasonably high resources for a solution on a classical computer.

As we've discussed before, two broad classes of quantum algorithms exist now, first is based on Shor's *Quantum fourier transform* which provides *exponential* speedup. Second is based on Grover's algorithm for performing *quantum searching* which provides *quadratic* speedup. Few main applications are listed in this figure

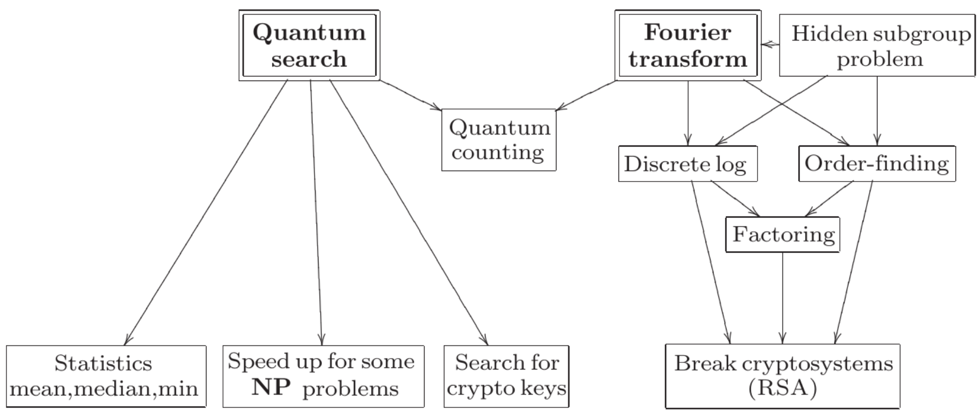


Figure 7.1: Main quantum algorithms with their applications and relationships.

7.2 Single qubit operations

A single qubit is the simplest quantum system. We already know few important things like norm of the state is 1, operations on it are defined by unitary matrices etc. Few important gates/ operations are Pauli matrices (I, X, Y, Z) which we know. Three other important quantum gates are Hadamard gate (H), phase gate (S), and $\pi/8$ gate (T):

$$H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} ; \quad S = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} ; \quad T = \begin{bmatrix} 1 & 0 \\ 0 & \exp{(i\pi/4)} \end{bmatrix} . \tag{7.1}$$

Few useful facts are $H = (X+Z)/\sqrt{2}$ and $S = T^2$, T is known as $\pi/8$ gate instead of $\pi/4$ because it's equal to a gate with diagonal $\exp(\pm\pi/8)$ upto a global phase factor:

$$T = \exp(i\pi/8)\begin{bmatrix}\exp(-i\pi/8) & 0 \\ 0 & \exp(i\pi/8)\end{bmatrix} \tag{7.2}$$

Also, in *Bloch sphere* representation, the state $a|0\rangle + b|1\rangle$ can be represented as a point (θ, φ) where $a = \cos(\theta/2)$ and $b = e^{i\varphi}\sin(\theta/2)$, and the Bloch vector is $(\cos\varphi\sin\theta, \sin\varphi\sin\theta, \cos\theta)$.

Pauli matrices give rise to useful matrices when exponentiated, the *rotation operators* about $\hat{x}$, $\hat{y}$ and $\hat{z}$ axes, defined by equations

$$R_x(\theta) \equiv e^{-i\theta X/2} = \cos\frac{\theta}{2}I - i\sin\frac{\theta}{2}X = \begin{bmatrix}\cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2}\end{bmatrix} \tag{7.3}$$

$$R_y(\theta) \equiv e^{-i\theta Y/2} = \cos\frac{\theta}{2}I - i\sin\frac{\theta}{2}Y = \begin{bmatrix}\cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2}\end{bmatrix} \tag{7.4}$$

$$R_z(\theta) \equiv e^{-i\theta Z/2} = \cos\frac{\theta}{2}I - i\sin\frac{\theta}{2}Z = \begin{bmatrix}e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2}\end{bmatrix}. \tag{7.5}$$

this is because if $x \in \mathbb{R}$ and $A^2 = I$, then $e^{iAx} = \cos xI + i\sin xA$. If $\hat{n} = (n_x, n_y, n_z)$ is real unit vector in three dimensions then generalize the rotation by θ about $\hat{n}$ by the equation

$$R_{\hat{n}}(\theta) = e^{-i\theta\hat{n}\cdot\vec{\sigma}/2} = \cos\left(\frac{\theta}{2}\right)I - \sin\left(\frac{\theta}{2}\right)(n_xX + n_yY + n_zZ), \tag{7.6}$$

where $\vec{\sigma}$ denotes the three component vector (X, Y, Z) of Pauli matrices. A nice fact is that any unitary operator on a single qubit can be written as $U = e^{i\alpha}R_{\hat{n}}(\theta)$ for some α, $\hat{n}$ and θ. More generally

Theorem 29 (Z-Y decomposition for a single qubit). *Suppose U is a unitary operation on a single qubit. Then there exist real numbers α, β, γ, δ such that*

$$U = e^{i\alpha}R_z(\beta)R_y(\gamma)R_z(\delta) \tag{7.7}$$

In general z and y can be replaced by any non-parallel real unit vectors $\hat{m}$ and $\hat{n}$. Also, this leads to a mysterious corollary, which'll be useful later.

Corollary 29.1. *Suppose U is a unitary gate on a single qubit. Then there exist unitary operators A, B, and C on a single qubit such that $ABC = I$ and $U = e^{i\alpha}AXBXC$, where α is some overall phase factor.*

Here are few useful circuit identities:

$$HXH = Z;\ HYH = -Y;\ HZH = X. \tag{7.8}$$

also, $HTH = R_x(\pi/4)$. To recap things here are few quantum circuit symbols

Hadamard	$-\boxed{H}-$	$\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$
Pauli-X	$-\boxed{X}-$	$\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$
Pauli-Y	$-\boxed{Y}-$	$\begin{bmatrix}0 & -i\\ i & 0\end{bmatrix}$
Pauli-Z	$-\boxed{Z}-$	$\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}$
Phase	$-\boxed{S}-$	$\begin{bmatrix}1 & 0\\ 0 & i\end{bmatrix}$
$\pi/8$	$-\boxed{T}-$	$\begin{bmatrix}1 & 0\\ 0 & e^{i\pi/4}\end{bmatrix}$

Figure 7.2: Names, symbols and unitary matrices for commonly used quantum gates.

7.3 Controlled operations

'If A is true, then do B' is the most useful and basic controlled operation in both classical and quantum computation. An example is CNOT gate in quantum computation, which does $|c\rangle\,|t\rangle \to |c\rangle\,|t \oplus c\rangle$ where $|c\rangle$ is control qubit and $|t\rangle$ the target qubit. If $|c\rangle$ is in state $|0\rangle$ nothing happens to target qubits, else if it's in $|1\rangle$ $|t\rangle$ is flipped. A general thing for this is *controlled-U* gate, which does $|c\rangle\,|t\rangle \to |c\rangle\,U^c\,|t\rangle$, where U is a unitary operation. In the computational basis |control, target⟩, the matrix representation of CNOT gate is

$$\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{bmatrix}. \tag{7.9}$$

To develop *controlled-U* gate for arbitrary U, we'll use $U = e^{i\alpha}AXBXC$. To apply the controlled phase shift $\exp(i\alpha)$, we use a circuit containing a single qubit gate as shown

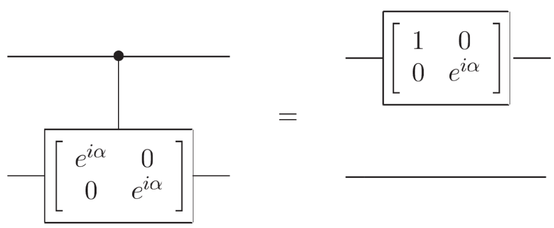


Figure 7.3: Controlled phase shift gate and an equivalent circuit for two qubits.

To complete the gate we use the circuit 7.4. This works because if control qubit is set to $|0\rangle$ then nothing happens at CNOT gates and $ABC = I$ is applied which does nothing, if control qubit is $|1\rangle$ then $e^{i\alpha}AXBXC$ is applied, thus U is applied.

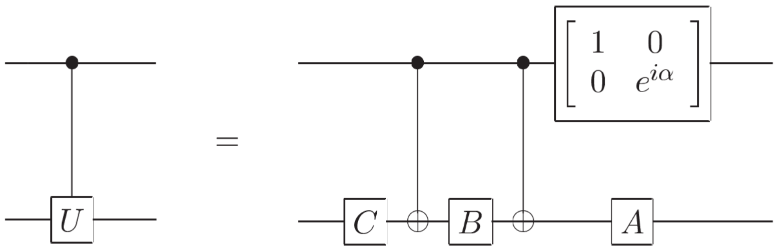


Figure 7.4: *controlled-U* circuit, α, A, B, C satisfy $U = e^{i\alpha}AXBXC$ and $ABC = I$

To use conditioning on multiple-qubits, suppose we have $n+k$ qubits, and U is a k bit operator. Then the controlled operatioin $C^n(U)$ by the equation

$$C^n(U)\left|x_1x_2\ldots x_n\right\rangle\left|\psi\right\rangle = \left|x_1x_2\ldots x_n\right\rangle U^{x_1x_2\ldots x_n}\left|\psi\right\rangle \tag{7.10}$$

where $x_1x_2\ldots x_n$ means product of those bits. It means operation U is applied on last k qubits if all of the first n qubits are $|1\rangle$. We'd assume $k=1$, for $k \geq 2$ we don't yet know how to perform k arbitrary operations at once.

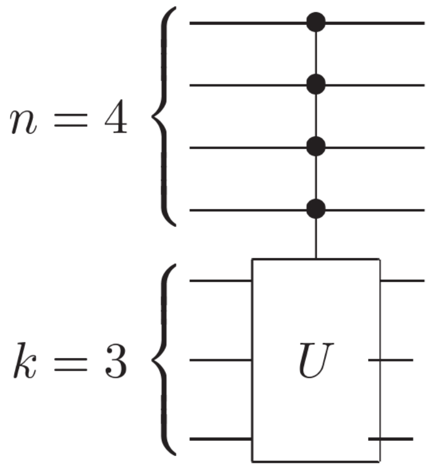


Figure 7.5: $C^n(U)$ operation for $n=4$ and $k=3$.

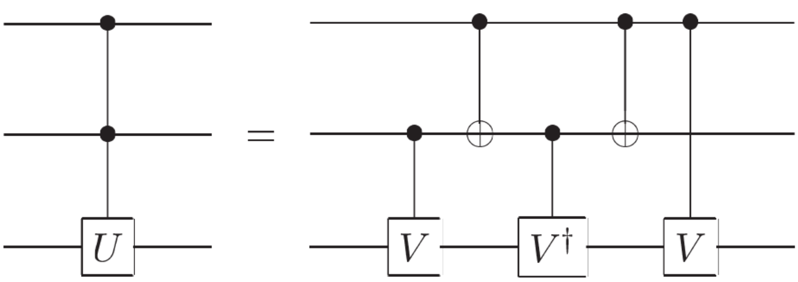


Figure 7.6: $C^2(U)$ gate. V is a unitary operator satisfying $V^2 = U$. When $V \equiv (1-i)(1+iX)/2$ represents the Toffoli gate.

Remember that when $U = X$, $C^2(U)$ is known as the *Toffoli gate*.

How $C^n(U)$ is implemented is interesting, first we take the first n qubits as $|c_1\rangle, |c_2\rangle \dots |c_n\rangle$, we use $(n-1)$ intermediate working qubits which are all initially and finally $|0\rangle$, which store $|c_1c_2\rangle, |c_1c_2c_3\rangle \dots |c_1c_2 \dots c_n\rangle$. This is done using toffoli gates which is applied between $|c_1\rangle$ and $|c_2\rangle$ with the output to first working qubit, the next toffoli gate is applied between first working qubit and $|c_3\rangle$ which gives $|c_1c_2c_3\rangle$ and so on, thus giving $|c_1c_2 \dots c_n\rangle$.

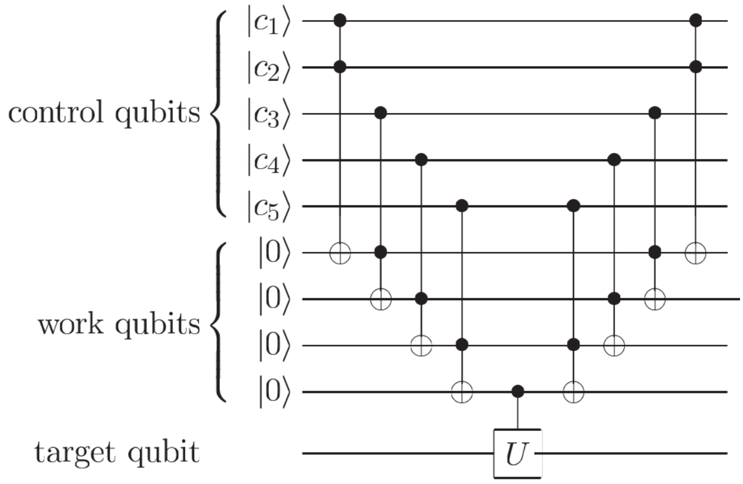


Figure 7.7: Implementation of $C^n(U)$ circuit.

We are considering conditionals being 1 till now, but sometimes it's useful to keep the conditional as 0 for which we introduce a notation, we keep an open circle (without darkening it) whereas we used to have a darkened circle till now.

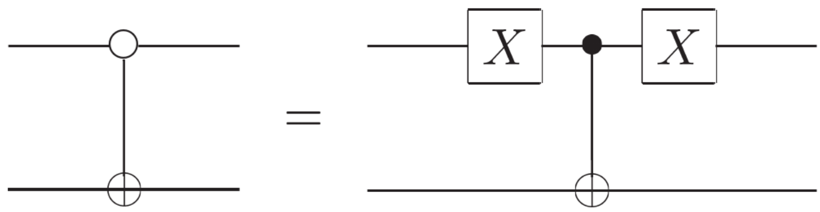


Figure 7.8: CNOT GATE, WITH THE CONDITIONAL BEING THAT FIRST QUBIT IS SET TO 0.

7.4 Measurement

As we've shown before measurements will be shown by a 'meter' symbol, especially projective measurements. There are two major principles worth emphasizing early, first is that classically conditioned operations can be replaced by quantum controlled operations:

> **Principle of deferred measurement:** Measurements can always be moved from an intermediate state of a quantum circuit to the end of the circuit; if the measurement results are used at any stage, then the classically conditioned operations can be replaced by conditional quantum operations.
>
> **Principle of implicit measurement:** Without loss of generality, any unterminated quantum wires (qubits which are not measured) at the end of a quantum circuit may be assumed to be measured.

7.5 Universal quantum gates

If a set of gates can compute any arbitrary classical function, they are said to bee *universal* for classical computation. Toffoli gate is universal for classical computation, hence quantum computation subsumes classical conputation.

7.5.1 Two level unitary gates are universal

We'll see how to decompose a d-dimensional unitary matrix U into a product of *two-level unitary matrices*, which act non-trivially on only two or fewer components. We'll understand what're these by going forward. Let's take an example U where it's 3×3,

$$U = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}. \tag{7.11}$$

We can find two-level unitary matrices such that

$$U_3 U_2 U_1 U = I \tag{7.12}$$

or, $U = U_1^\dagger U_2^\dagger U_3^\dagger$. More generally, for d-dimensional space, U can be written as

$$U = V_1 \cdots V_k \tag{7.13}$$

where the matrices V_i are unitary matrices and $k \leq (d-1)+(d-2)+\cdots+1 = d(d-1)/2$. This follows that any unitary operator on n qubits can be written as a product of $2^{n-1}(2^n-1)$ two-level unitary matrices.

7.5.2 Single qubit and CNOT gates are universal