Quantum information and quantum computation; summary? version 0.1

May 24, 2004

Notification

Nielsen & Chuang의 “Quantum Information and Quantum Computation” 요약입니다. 2004년도 겨울 세미나를 시작으로 한학기 정도 진행된 세미나이고, 보충버전이 만들어 질 마음의 계획은 있으나, 확실하지는 않습니다. 영문으로 작성한 이유는 의역에 의한 오류를 막기 위함입니다. 영타 많이 늘었습니다.

오타 수정 및 보강 작업을 한 이후 방학 이후 새로운 판을 내놓을 생각입니다. 연습문제 풀어 놓은 것 있으면, 보충해주셔도 좋습니다. 현재 버전은 0.1입니다. (2004년 5월 24일 현재) 버전은 첫 섹션에 계속 추가하는 방식으로 쓰겠습니다.

– jiinny

Contents

Chapter 1

Introduction and overview

1.1 Global perspectives

1.1.1 History of quantum computation and quantum information

1.1.2 Future directions

1.2 Quantum bits

1.2.1 Multiple qubits

1.3 Quantum computation

1.3.1 Single qubit gates

1.3.2 Multiple qubit gates

1.3.3 Measurements in bases other than the computational basis

1.3.4 Qubit copying circuit?

1.3.5 Example: quantum teleportation

1.4 Quantum algorithms

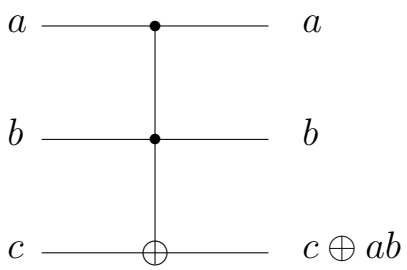
1.4.1 Classical computations on a quantum computer

We can simulate a classical logic circuit using a quantum circuit. ($\because$ classical world $\subset$ quantum world)

Q-circuit cannot be used to directly simulate classical circuit, because unitary Q logic gates are inherently **reversible**, whereas many classical logic gates are inherently **irreversible**. *eg)* NAND gate

∀ Classical circuit can be replaced by an equivalent circuit containing only reversible elements, by making use of a reversible gate like *Toffoli gate*.

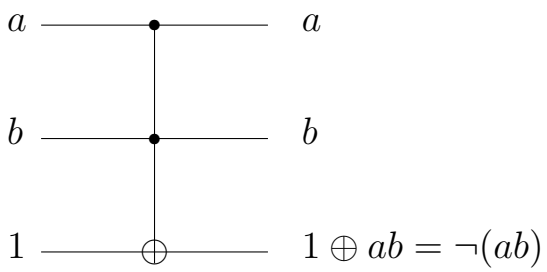
Inputs			Outputs		
a	b	c	a'	b'	c'
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	1	1	0	1
1	1	0	1	1	1
1	1	1	1	1	0



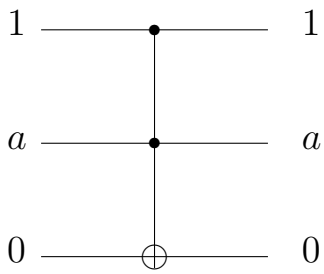
- Toffoli gate has 3 input bits (2 control bits & 1 target bit) and 3 output bits.

$$T(a, b, c) = (a, b, c \oplus ab)$$

- Toffoli gate is reversible.
- NAND gate can be simulated by the Toffoli gate.



- FANOUT gate can be simulated by the Toffoli gate. FANOUT = T (1, a, 0) = (1, a, a)



- ∀ classical gate can be simulated using NAND & FANOUT gates.
- Toffoli gate has been described as a classical gate,
  but can also be implemented as a Q logic gate.
  can be represented as a 8x8 unitary mat U.
  is a legitimate Q gate
- The Q-Toffoli gate can be used to simulate irreversible classical logic gates.
- Q computers are capable of performing any computation which a classical computer may do and has ability to efficiently simulate a non-deterministic classical computer.

1.4.2 Quantum parallelism

Q parallelism is a fundamental feature of many Q algorithms.

Q parallelism allows Q computers to evaluate a function f(x) for many different values of x simultaneously.

Suppose $f(x) : \{0,1\} \rightarrow \{0,1\}$ is a functioin with a 1-bit domain & range.

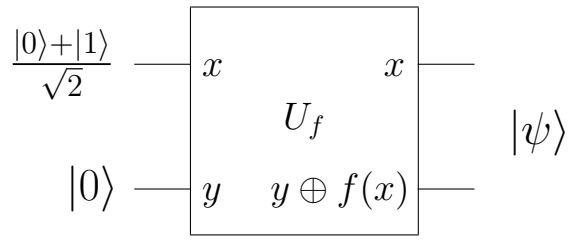
Two qubit Q-computerr computing f(x) : A convenient way of computing $f(x)$ on a Q computer

$$|x, y\rangle \rightarrow |x, y \oplus f(x)\rangle$$

where, x is data register and y is target register. If $y = 0$, then the final state of the second qubit is just the value $f(x)$.

$$U_f|x, y\rangle = |x, y \oplus f(x)\rangle$$

where, U_f is unitary transformation.



eg) Q-cct evaluating $f(0)$ & $f(1)$ simultaneously.

U_f is the Q-circuit which takes input like $|x, y\rangle$ to $|x, y \oplus f(x)\rangle$

If data register is prepared in the superposition $f(0) = (|0\rangle + |1\rangle)/\sqrt{2}$, which can be created with a Hadamard gate acting on $|0\rangle$

$$\begin{aligned} U_f\left[H(|0\rangle), |0\rangle\right] &= U_f\left[\frac{|0\rangle + |1\rangle}{\sqrt{2}}, |0\rangle\right] \\ &= \frac{|0, f(0)\rangle + |1, f(1)\rangle}{\sqrt{2}} \end{aligned}$$

It is almost as if we have evaluated f(x) for two values of x simultaneously. A single f(x) circuit evaluate f(x) for multiple values of x simultaneously by exploiting the ability of a Q computer to be in superposition of different states.

This procedure can be easily generalized to ftns on arbitrary # of bits, by using Hadamard transform : n Hadamard gates acting in parallel on n qubits.

For example, n = 2

$$\begin{aligned} H(|0\rangle)H(|0\rangle) &= \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right) \\ &= \frac{|00\rangle + |01\rangle + |10\rangle + |1\rangle}{2} \end{aligned}$$

More generally,

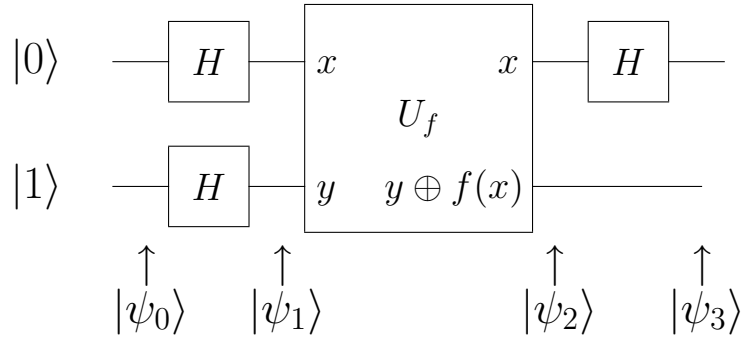
$$H^{\otimes n}|0 \ldots 0\rangle = \frac{1}{\sqrt{2^n}} \sum_x |x\rangle$$

Hadamard transform produces an equal superposition of all computational basis states and does it extremely efficiently, producing a superposition of 2^n states using just n gates.

Q computation requires the ability to extract information about more than one value of f(x) from superposition states like $\sum_x |x, f(x)\rangle$

1.4.3 Deutsch's algorithm

Deutsch's algorithm combines Q-parallelism with interference.



$$
\begin{aligned}
|\psi_0\rangle &= |01\rangle \\
|\psi_1\rangle &= H|\psi_0\rangle \\
&= |H(0), H(1)\rangle \\
&= \left[\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right]\left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right]
\end{aligned}
$$

Applying U_f to the state $|x\rangle(|0\rangle - |1\rangle)/\sqrt{2}$

$$
\begin{aligned}
U_f|x\rangle\frac{|0\rangle - |1\rangle}{\sqrt{2}} &= |x\rangle\frac{|0 \oplus f(x)\rangle - |1 \oplus f(x)\rangle}{\sqrt{2}} \\
&= \begin{cases} \quad |x\rangle\frac{|0\rangle - |1\rangle}{\sqrt{2}} & \text{if } f(x) = 0 \\ - \quad |x\rangle\frac{|0\rangle - |1\rangle}{\sqrt{2}} & \text{if } f(x) = 1 \end{cases} \\
&= (-1)^{f(x)}|x\rangle\frac{|0\rangle - |1\rangle}{\sqrt{2}} \\
|\psi_2\rangle &= U_f|\psi_1\rangle \\
&= U_f\left[\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right]\left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \\
&= \left[\frac{(-1)^{f(0)}|0\rangle + (-1)^{f(1)}|1\rangle}{\sqrt{2}}\right]\left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right] \\
&= \begin{cases} (-1)^{f(0)}\left[\frac{|0\rangle+|1\rangle}{\sqrt{2}}\right]\left[\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right] & \text{if } f(0) = f(1) \\ (-1)^{f(0)}\left[\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right]\left[\frac{|0\rangle-|1\rangle}{\sqrt{2}}\right] & \text{if } f(0) \neq f(1) = f(0) \oplus 1 \end{cases} \\
&= (-1)^{f(0)}\left[\frac{|0\rangle + (-1)^{f(0)\oplus f(1)}|1\rangle}{\sqrt{2}}\right]\left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right]
\end{aligned}
$$

$$
\begin{aligned}
|\psi_3\rangle &= (-1)^{f(0)} H \left[\frac{|0\rangle + (-1)^{f(0)\oplus f(1)}|1\rangle}{\sqrt{2}} \right] \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right] \\
&= \begin{cases} (-1)^{f(0)}|0\rangle \frac{|0\rangle - |1\rangle}{\sqrt{2}} & \text{if } f(0) = f(1) \\ (-1)^{f(0)}|1\rangle \frac{|0\rangle - |1\rangle}{\sqrt{2}} & \text{if } f(0) = f(1) \neq f(0) \oplus 1 \end{cases} \\
&= (-1)^{f(0)} |f(0) \oplus f(1)\rangle \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right]
\end{aligned}
$$

Q-circuit can determine a global property of $f(x)$, namely $f(0) \oplus f(1)$, using only one evaluation of $f(x)$.

Possible for two alternatives to interfere with one another to yield some global property of f by using sth like H-gate to recombine the different alternatives.

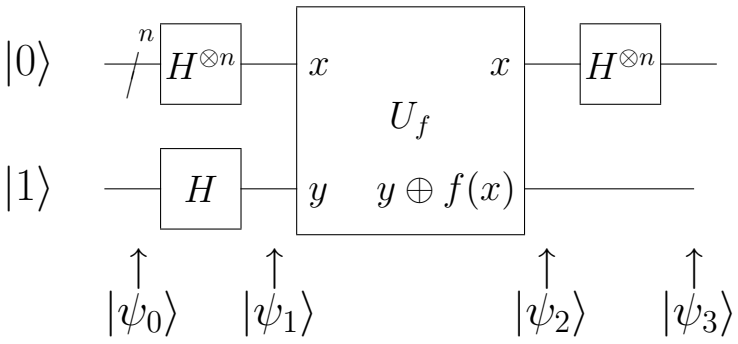
A clever choice of function & final transform allows efficient determination of useful global information about the function.

1.4.4 The Deutsch–Jozsa algorithm

Deutsch's problem

Alice selects a # x from 0 to $2n-1$, & mail it in a letter to Bob.
Bob calculates some ftn f(x) & replies with the result, which is either 0 or 1. Bob has promised to use a function f which is one of two kinds; constant or balanced ftn. Alice's goal is to determine with certainty what ftn Bob has used.

How fast can she succeed? If Alice & Bob were able to exchange qubit, & if B agreed to calculate $f(x)$ using U_f, then A could achieve her goal in just one correspondence with B, using D-J algorithm.



$$
\begin{aligned}
|\psi_0\rangle &= |0\rangle^{\otimes n}|1\rangle \\
|\psi_1\rangle &= \sum_{x\in\{0,1\}^n} \frac{|x\rangle}{\sqrt{2^n}} \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right]
\end{aligned}
$$

The query register is now a superposition of all values

$$
\begin{aligned}
|\psi_2\rangle &= U_f|\psi_1\rangle \\
&= \sum_{[x\in\{0,1\}^n]} (-1)^{f(x)} \frac{|x\rangle}{\sqrt{2^n}} \frac{|0\rangle - |1\rangle}{\sqrt{2}}
\end{aligned}
$$

$$
\begin{aligned}
H|x\rangle &= \frac{|0\rangle \pm |1\rangle}{\sqrt{2}} \qquad x = 0\text{or}1 \\
&= \sum_{[z\in\{0,1\}^n]} (-1)^{xz} \frac{|z\rangle}{\sqrt{2}} \\
H^{\otimes n}|x_1,\ldots,x_n\rangle &= \frac{\sum_{z_1,\ldots,z_n}(-1)^{x_1z_1+\ldots+x_nz_n}|z_1,\ldots,z_n\rangle}{\sqrt{2^n}} \\
H^{\otimes n}|x\rangle &= \frac{\sum_z (-1)^{x\cdot z}|z\rangle}{\sqrt{2^n}} \\
|\psi_3\rangle &= H^{\otimes n}|\psi_2\rangle \\
&= \sum_z \sum_x \frac{(-1)^{x\cdot z+f(x)}|z\rangle}{2^n} \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right]
\end{aligned}
$$

where, $\sum_x (-1)^{f(x)}/2^n$ is the amplitude for the state $|0\rangle^{\otimes n}$

If f is const, $\sum_x (-1)^{f(x)}/2^n$ is ± 1, depending on the const value of $f(x)$.

An observation will yield zeros for all qubit in the query system. ($\because \langle\psi_3|\psi_3\rangle = 1$)

If f is balanced, $\sum_x (-1)^{f(x)}/2^n$ is 0 (+ & - contribution to $\sum_x (-1)^{f(x)}/2^n$ cancel). Alice's measure must yield a result other than 0 on at least one qubit in the query register.

D-J algo contains the seeds for more impressive Q-algorithm.

1.4.5 Quantum algorithms summarized

The Deutsch-Jozsa algorithm suggests that quantum computers may be capable of solving some computational problems much more efficiently than classical computers. Three classes of quantum algorithms which provide an advantage over known classical algorithms

1. Quantum versions of Fourier transformation
2. Quantum search algrithms
3. Quantum simulations

Quantum algorithm based upon the Fourier transform

The descrete Fourier transform is usually described as trasforming a set $x_0,\ldots,x_{N-1}$ of N complex numbers into a set of complex numbers $y_0,\ldots,y_{N-1}$ defined by

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

The Hadamard transformation used in D-J algorithm is an example of this generalized class of F-transformation. Moreover, many of the other important quantum algorithms also involve some type of Fourier trasformation.

eg) Shor's fast algorithms for factoring and discrete logarithm

Imagine that we define a lineartransformation U on n qubits by its actioin on computational basis states $|j\rangle$, where $0 \le j \le 2^n - 1$,

$$U|j\rangle = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} e^{2\pi ijk/2^n}|k\rangle.$$

Moreover, if we write out its action on superpositions,

$$U \sum_{j=0}^{2^n-1} x_j |j\rangle = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} \left[\sum_{j=0}^{2^n-1} e^{2\pi i jk/2^n} x_j \right] |k\rangle = \sum_{k=0}^{2^n-1} y_k |k\rangle,$$

Deutsch-Jozsa algorithm, Shor's algorithms for factoring and discrete logarithm.
Classical(FFT) : $n2^n$ steps
Quantum computer : n^2 steps

However, that is not exactly the case; the Fourier transformation is being performed on the "hidden" in the collapse amplitude of the quantum state.

Quantum search of given size N Classical : N operations Quantum computer : $\sqrt{N}$ operations

Q simulation simulate a Q system

The power of quantum computation

The power of quantum computer is not yet verified. (a few example exits)

Computational complexity theory : the subject of classifying the difficulty of various computational problems

A *complexity class* : a collection of computational problems, all of which share some common feature with respect to the computational resources needed to solve those problems.

P The class of computational problems that can be solved quickly on a classical computer.

NP The class of problems which have *solutions* which can be quickly checked on a classical computer.

NP-complete An important subclass of **NP** problems

- There are many important problems that known to be **NP**-complete.
- Any given **NP**-complete problem is in some sense 'at least as hard' as all other problems in **NP**. So, an algorithm to solve a specific **NP**-complete problem can be adapted to solve any other problem in **NP**, with small overhead.

P is a subset of **NP**. What is not so clear is whether or not there are problems in**NP** that are not in **P**.

If $\mathbf{P} \neq \mathbf{NP}$, then it will follow that no **NP**-complete problem can be efficiently solved on a classical computer

Factoring is not known to be **NP**-complete, and it can be solved using quantum algorithm through quantum parallelism - quantum computation is expected to solve other **NP** problems.

PSPACE Problems which can be solved using resources which are few in spatial size, but not necessarily in time. **PSPACE** is believed to be strictly larger than both **P** and **NP**

BPP The class of problems that can be solved using randomized algorithms in polynomial time, **BPP** is widely regarded as being the class of problems which should be considered efficiently soluble on a classical computer.

What of quantum complexity classes?

BQP : The class of all computational problems which can be solved efficiently on a quantum computer, where a bounded probability of error is allowed.

Exactly where **BQP** fits with respect to **P**, **NP** and **PSPACE** is as yet unknown.

BQP may be somewhere between **P** and **PSPACE**

1.5 Experimental quantum information processing

1.5.1 The Stern–Gerlach experiment

SG–experiment provides the evidence for the existence of qubit structure in Nature, that is, two level Q system.

Conceived by Stern in 1921 & performed by Gerlach in 1922.

Hot silver atoms were beamed from an oven through a B-field : The original exp.
Hot hydrogen atoms are : Abstract schematic exp

A continuous distribution of atoms at all angles from the SG device was expected: the atoms are deflected by an amount that depends on atom's m_z (the orig. exp.)

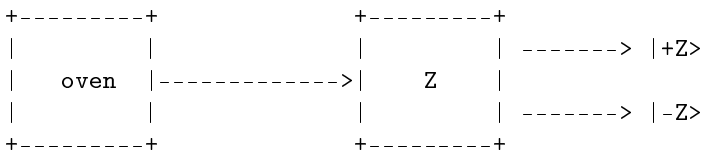
m is quantized : Atoms were emerged to a discrete set of angles.

m_z of hydrogen atom related to orbital angular momentum = 0. Only 1 beam of atoms, No deflection was expected

Instead, Two beams were seen : One is deflected up by the B field, the other is down.

This puzzling doubling was explained by considering *spin*

- Not in any way associated to the usual rot motion of the el around the proton.
- An entirely new quantity to be associated with an electron.
- make an extra contribution to the m of a H-atom.



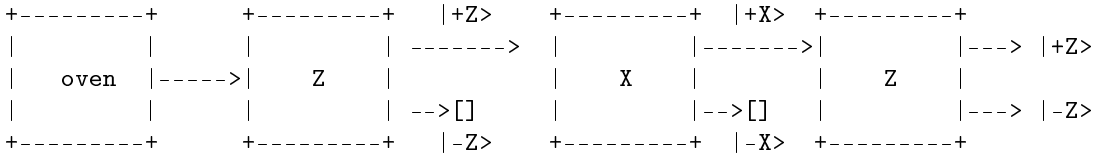


Fig. 1.22 Fig. 1.24.
qubit model provides a simple explanation of this experimentally observed behavior. predicts results from the cascaded SG-exp.

qubit could be a believable way of modeling sys in Nature. turned out to be a correct way of understanding el spin.

1.5.2 Prospects for practical quantum information processing

Will we rise to meet the challenge, building quantum information devices?
It is possible at all?
Is it worth attempting?
Whether there is any point of principle that prohibits us from doing Q-information processing?
Two possibilities:

1. Noise may place a fundamental barrier to useful Q-inf processing.

2. QM may fail to be correct.

For noise, Q error-correction theory strongly suggests that
while Q noise is a practical problem that needs to be addressed,
it does not present a fundamental problem of principle

Threshold Theorem :

- If the level of noise in a Q-computer can be reduced below a certain const threshold value,
  Q error correcting codes can be used to push it down even further,
  essentially *ad infinitum*, for a small overhead in the complexity of the computation
- Threshold Theorem makes some broad assumptions about the nature & magnitude of the noise occuring in a Q computer & the architecture available for performing Q-computation.
- If the assumptions are satisfied, the effect of noise can be made essentially negligible for Q information processing.

We assume that QM is a complete & correct description of the world.

There's no fundamental obstacle to building quantum information processing devices.

However, It is fair to say that a clearer picture of the relative power of quantum and classical information processing is needed in order to assess their relative merits

Many small scale applications on Q-computation & Q-information are known.

- Q state tomography & Q process tomography are two elementary process whose perfection is of great importance to Q-computation & Q-information.
  - Q state tomography : A method for determining the Q state of a sys.
    The hidden nature of the Q-state has to be overcome by performing repeated preparations of the same description of the Q-state.
  - Q process tomography : Intend to completely characterize the dynamics of a Q-system.
    can be used to characterize the performance of an alleged Q gate or Q comm channel, or to determine the types & magnitudes of different noise process in a sys.
- Q cryptography is likely to be useful in practical applications involving the distribution of small amount of key material that needs to be highly secure.
- Q teleportation may be an extremely useful primitive for transmitting Q-states between distant nodes in a network.
  Distributing EPRs between the nodes has been focussed.
  Special entanglement distillation protocols can be used to clean up the EPR pairs which was corrupted during communication enables EPR pairs to be used to teleport Q-sts from one location to another.
- Other applications, like quantum game theory, quantum walk and so on, are considered in terms of quantum communications.

medium scale application of Q-information processing.
The simulation of Q sys

eg) 50 qubit $2^{50} \cong 10^{15}$ complex amplitudes 32×10^{15} bytes need 32 bytes in order to store each amplitude for 128 bit precision. 32 thousand terabytes of information : well beyond the capacity of existing computer.

- elementary properties of materials(bond strength, basic spectroscopic properties)
- A lab for the design & testing of properties of novel molecules.

Large scale application : the factoring of large #s & taking discrete logarithm, Q searching

How can it be achieved in real sys?

- Optical technique : EM radiation
  - Producing a single photon on demand is a major difficulty.
  - Physicists opt to use schemes which produce single photons every 'now and then', at random, & wait for such an event to occur.
  - Major advantage : Photon tend to be highly stable carriers of quantum mechanical informations.
  - Major disadvantage :
    * Photon don't directly interact with one another.
    * Interaction has to be mediated by something else, like atoms, which introduction of additional noise & complications into the exp.
- Ion trap : a small # of charged atoms are trapped in a confined space. Neutral atom traps : trap uncharged atoms in a confined space.
- NMR Nuclear Magnetic Resonance : store Q information in the nuclear spin of atoms in a molecule & manipulate that information using EM radiation.
- other techniques of nano technologies like SQUID(josephson junction), fabricated magnetic materials, quantum well, quantum dot, quantum wire *e*tc.

1.6 Quantum information

The term "*quantum information*":

1. All manner of operations that might be interpreted as related to informatin processing using quantum mechanics.(*e.g. q–computation, q–teleportation, the no-cloning theorem*)
2. The study of elementary quantum information process tasks.
   - q info. theory $\Longleftrightarrow$ (classical) info. theory

A few fundamental goals uniting work on quantum information theory :

1. Identify elementary classes of static resources in quantum mechanics
   - qubits
   - bits
   - bell states shared between two distant parties.
2. Identify elementary classes of dynamical processes in quantum mechanics
   - memory – storing q-st. over some period of time
   - q-information transmission between two parties
   - copying(or trying to copy) a quantum state

- noise protecting quantum communication

3. Quantify resource tradeoffs incurred performing elementary dynamics
    - What are the minimal resources to transfer quantum information between two parties using noisy quantum channel.

The quantum information theory includes all the static and dynamic elements of classical information theory, as well as additional static and dynamic elements ⇒ Broader.

Some information theoretic questions have been studied as samples:
Exploring Semi-classical elements (*sending classical information through q-channel*)

⇓

The new static and dynamic processes present in quantum mechanics (*quantum error-correction, the problem of distinguishing quantum states, and entanglement transformation.*)

1.6.1 Quantum information theory: example problems

Classical information through quantum channels

Classical elements Shannon's noiseless channel coding theorem and Shannon's noisy channel coding theorem : The fundamental results of classical information theory.

noiseless How many bits are required to store information being emitted by a source of information

noisy How much information can be reliably transmitted through a noisy commnunication channel

A classical information source is described by a set of probabilities p_j, $j = 1, 2, 3...d$ (*eg. English text* - the numbers j : the alphabet and punctuation, with the probabilities p_j giving the relative frequencies with which the different letters appear in regular English text)

Regular text includes redundancy, the redundancy have the possibility of compress the text.
→ Shannon's noiseless channel coding theorem quantifies exactly how well such a compression scheme can be made to work.

The noiseless channel coding theorem tells

- A classical source described by probabilities p_j can be compressed so that on average each use of a classical source can be represented using $H(p_j)$ bits of information, where

$$H_(p_j) \equiv -\sum_j p_j \log p_j$$

  is a function of the source probability distribution known as the *Shannon entropy*.

- to attempt to represent the source using fewer bits than this will result in a high probability of error when the information is decompressed.

Shannon's noiseless coding theorem provides a good example where the goals of information theory

goal 1 Two static resources are identified (the bit and the information source)

goal 2 Two-stage dynamic process is identified (compressing, decompressing)

goal 3 A quantitative criterion for determining the resources consumed (an optimal data compression scheme)

The noisy channel coding theorem

- quantifies the amount of information that can be reliably transmitted through a noisy channel.
  In particular, to transfer the information being produced by some source to another location (that location : at another point in space, or at another point in time - storing information in the presence of noise -)
- The idea in both instances is to encode the information being produced using *error–correcting codes*, so that any noise introduced by the channel can be corrected at the end of the channel.
- The error–correcting codes achieve this is by introducing *enough redundancy* into the information sent through the channel, so that even after some of the information has been corrupted it is still possible to recover the original message.
- provieds a general procedure for calculating the capacity of an arbitrary noisy channel. (for one bit information transmission through noisy channel, one must encode using two bits : such a channel has a capacity of half a bit)

Shannon's noisy channel coding theorem also achieves the three goals of information thoery

goal 1 Two static resources are involved (the bit and the information source)

goal 2 Three dynamic processes are involved (noise, encoding and decoding the state in an error–correcting code)

goal 3 How much redundancy must be introduced by an optimal error–correction scheme, for a fixed noise channel

Semi-classical elements A natural question for q-information theory : what happens if the storage medium is canged so that classical information is transmitted using quantum states as the medium.

In Ch. 12 : Using qubits allows a better compression rate than is possible classically and we will prove that qubits do not allow any significant saving in the amount of communication required to transmit information over a noiseless channel.

Transmission classical information through a noisy channel:
Might it be advantageous, for example, to encode the classical information using *entangled* states, which are then transmitted one piece at a time through the noisy channel?
Perhaps it will be advantageous to decode using entangled measurements?

HSW(Holevo-Schumacher-Westmoreland) theorem

- provides a lower bound on the capacity of a channel.
- is believed to provide an exact evaluation of capacity

Encoding using entangled does not help raise the capacity beyond the lower bound by the HSW theorem. (but still open problem)

Semi-classical schemes are concerned with Eavesdropping of orthogonal quantum key distribution

Quantum information through quantum channels - pure quantum elements

def) probabilities p_j and corresponding quantum states $|\psi_j\rangle$

A quantum source with output states $|0\rangle$ with probability p and $|1\rangle$ with probability $1-p$ can be compressed so that $H(p, 1-p)$ qubits are required to store the compressed source.

If the source had been producing the state $|0\rangle$ with probability p, and the state $(|0\rangle + |1\rangle)/\sqrt{2}$ with probability $1-p$, the standard techniques of classical data compression no longer apply. (we can not distingush $|0\rangle$ and $|+\rangle$)

A type of compression is still possible, but the compression may no longer be *error-free* (the quantum state being produced by the source may be slightly distorted by the compression-decompression procedure)

Suppose the distortion is small and ultimately neglibigle in limit of large blocks of source output being compressed.

A *fidelity* measure for the compression scheme

- measures the average distortion introduced by compression scheme.
- The idea of quantum data compression is that the compressed data should be recovered with very good fidelity.
- is thought to be analogues to the probability of doing the decompression correctly (large data block limit $\rightarrow$ no error limit of 1)

Schumacher's noiseless channel coding theorem

- quantifies the resources required to dot quantum data compression, with the restriction that it be possible to recover the source with fidelity close to 1.
- In the case of a source producing orthogonal quantum states $|\psi_j\rangle$ with probabilities p_j the theorem reduces to telling that the source may be compressed down to but beyond the classical limit $H(p_j)$.
- In the more general case of non-orthogonal states being produced by the source, the theorem tells how much a quantum source may be compressed (the new entropic quantity *von Neumann* entropy)

The von Neumann entropy for the sourced p_j, $|\psi_j\rangle$ is in general strictly smaller than the Shannon entropy $H(p_j)$. (*eg.* $|0\rangle$ with p and $(|0\rangle + |1\rangle)/\sqrt{2}$ with probability $1-p$ can be reliably compressed using fewer than $H(p, 1-p)$ qubits per use of the source)

Suppose above sources, and use the large # limit law with high probability the source emits about np copies of $|0\rangle$ and $n(1-p)$ copies of $(|0\rangle + |1\rangle)/\sqrt{2}$. It has the form

$$|0\rangle^{\otimes np} \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \right)^{\otimes n(1-p)}$$

up to re-ordering of the systems involved. Approximately, the state emitted by the source has the form

$$|0\rangle^{\otimes n(1+p)/2} |1\rangle^{\otimes n(1-p)/2}$$

How many states of this form are there? Roughly n choose $n(1+p)/2$, which by Stirling's approximation is equal to $N \equiv 2^{nH[(1+p)/2,(1-p)/2]}$

$$\left(\frac{n!}{(n-n(1+p)/2)!\,(n(1+p)/2)!} = \frac{n!}{(n(1-p)/2)!\,(n(1+p)/2)!}, \right.$$
$$\left. n! \approx n^n e^{-n} \sqrt{2\pi n}, \qquad H(p_i) = -\sum_i p_i \log_2 p_i \right)$$

A simple compression method then is to label all states of the form $|0\rangle^{\otimes n(1+p)/2} |1\rangle^{\otimes n(1-p)/2}$ $|c_1\rangle$ through $|c_N\rangle$.

It is possible to perform a unitary transform on the n qubits emitted from the source that takes $|c_j\rangle$ to $|j\rangle |0\rangle^{\otimes n - nH[(1+p)/2,(1-p)/2]}$ since j is an $nH\left[(1+p)/2, (1-p)/2\right]$ bit number. The compression operation is to perform this unitary trasformation, and then drop the final $n - nH\left[(1+p)/2, (1-p)/2\right]$ qubits, leaving a compressed state of $nH\left[(1+p)/2, (1-p)/2\right]$

qubits. To decompress, we append the state $|0\rangle^{n-nH[(1+p)/2,(1-p)/2]}$ to the copressed state, and perform the inverse unitary transformation.

The Storage requirement : $H\left[(1+p)/2,(1-p)/2\right]$ qubits per use of the source.

Schumacher's noiseless channel coding theorem allows us to do somewhat better even than this (Ch. 12)

For $|0\rangle$ and $(|0\rangle+|1\rangle)/2$, the states contain some redundancy since both have a components in the $|0\rangle$ direction.

Shumacher's noiseless channel coding theorem is an analogue of Shannon's but there's no fully satisfactory analogue of Shannon's on noisy channel. (however, it has been made with quantum error-correcting codes)

Quantum distinguishability

Up to now, compression, decompression, noise, encoding and decoding error correcting codes have been considered in both classical and quantum information theory. The new types of information (such as quantum state) enlarges the class of dynamical processes.

The problem of distinguishing quantum state : it is not always possible to distinguish between arbitrary states (*eg.* $|0\rangle$ and $|+\rangle$ are not distinguishable)

However, this indistinguishability of non-orthogonal quantum states is at the heart of QC and QI.

A quantum state contains **hidden** information that is not accessible to measurement, and thus plays a key role in quantum algorithm and quantum cryptography.

One of the central problems of quantum information theory is to develop measures quantifying how well non-orthogonal quantum state may be distinguished. (Ch.9, Ch.12)

The distinguishability between non-orthogonal quantum states implies the ability to communicate faster than light, using entanglement.

Alice and Bob share an entangled pair of qubits in the state $(|00\rangle+|11\rangle)/\sqrt{2}$. Then if Alice measures in the computational basis, the post-measurement state will be $|00\rangle$ with probability 1/2, and $|11\rangle$ with 1/2. Thus Bob has same probabilities in both states.

If Alice had instead measured in the $|+\rangle$, $|-\rangle$ basis, the state of Bob's system after the measurement will be $|+\rangle$ or $|-\rangle$ with probability 1/2 each.

If Bob had access to a devie that could distinguish the four states $|0\rangle$, $|1\rangle$, $|+\rangle$, $|-\rangle$ from one another, then he could tell whether alice had measured in the computational basis, or in the $|+\rangle$, $|-\rangle$ basis. Moreover he could get that information *instantaneously*, as soon as Alice had made the measurement. (faster-than-light communication)

Quantum money : banknotes imprinted with a serial #, and a sequence of qubits each in either the state $|0\rangle$ or $(|0\rangle+|1\rangle)/\sqrt{2}$. The bank maintains a list matching serial # to embedded states. The note is impossibel to counterfeit exactly, because it is impossible for a would-be counterfeiter to determine with certainty the state of the quvits in the original note, without destroying them.

This idea is the basis for numerous other quantum cryptographic protocols, and demonstrates the utility of the indistinguishability of non-orthogonam quantum states.

Creation and transformation of entanglement

Engtanglement is another elementary static resource of quantum mecanics.
→ Amazingly different; HOWEVER note yet well understood

eg. Two information-theoretic problems related to entanglement Creating entanglement
Transfering entanglement : A, B share Bell state, and wish to transform it into some other type of entanglement.

- What resources do they need to accomplish this task?
- Can they do it without communicating?

- With classical communication only?
-

1.6.2 Quantum information in a wider context

Quantum information theory : PART III

especially
Ch. 11 : deals with fundamental properties of entropy in quantum and classical information theory.
Ch. 12 : focuses on pure quantum information theory

Quantum information theory is the most abstract part of quantum computation and quantum information, yet in some sense it is also the most fundamental.

What is it that separates the quantum and the classical world?
What resources, unavailable in a classical world, are being utilized in a quantum computation?

Existing answers to these questions are foggy and incomplete; ti is our hope that the fog may yet lift in the years to come, and we will obtain a clear appreciation for the possibilities and limitations of quantum information processing.

Chapter 2

Introduction to quantum mechanics

2.1 Linear algebra

2.1.1 Bases and linear independence

2.1.2 Linear operators and matrices

2.1.3 The Pauli matrices

2.1.4 Inner products

2.1.5 Eigenvectors and eigenvalues

2.1.6 Adjoints and Hermitian operators

2.1.7 Tensor products

2.1.8 Operator functions

2.1.9 The commutator and anti-commutator

2.1.10 The polar and singular value decompositions

2.2 The postulates of quantum mechanics

The postulates of quantum mechanics were derived afeter a long process of trial and (mostly) error, which involved a considerable amount of guessing and fumbling by the orginators of the theory.

2.2.1 State space

What the arena in which quantum mechanics take place is.

Postulate 1; Associated to any isolated physical system is a complex vector space with inner product (that is a Hilbert space) known as the *state space*of the system. The system is completely described by its *state vector*, which is a unit vector in the system's state space.

Qauntum mechanics does not tell us what the state space of that system is, nor that the space vector of the system is. Figuring that out for a specific system is a difficult problem. eg) **QED, ISING, HUBBARD model**

A general state can be said as a linear combination of unit vectors, that is

$$|\Psi\rangle = \sum_i \alpha_i |\psi_i\rangle$$

where $|\Psi\rangle$ is a general state, α_i is (probability) amplitude for unit state $|\psi_i\rangle$. We can say that a $|\Psi\rangle$ is the superposition of the states $|\psi_i\rangle$

2.2.2 Evolution

Postulate 2: The evolution of a *closed* quantum system is described by a *unitary transformation*. That is, the state $|\psi\rangle$ of the system at time t_1 is related to the state $|\psi'\rangle$ of the system at time t_2 by a unitary operator U which depends only on the times t_1 and t_2

$$|\psi'(t')\rangle = U(t', t)|\psi(t)\rangle.$$

Just as quantum mechanics does not tell us the state space or quantum state of a $particular$ quantum system, it does not tell us which unitary operators U describe real-world quantum dynamics. Recall unitary operators described in Chapter 1. (the quantum NOT gate X, Z, and Hadamard gate)

Postulate 2 requires that the system being described be closed. (no interaction with other systems) In reality, all systems interact at least somewhat with other systems.Nevertheless, there are interesting systems which can be described as part of a larger closed system. **In principle, every open system can be described as part of a larger closed system which is undergoing unitary evolution.**(Ch.8)

Postulate 2′: The time evolution of the state of a closed quantum system is described by the *Schrödinger equation*,

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

In general figuring out the Hamiltonian needed to describe a particular physical system is very difficult problem which requires substantial input from experiment in order to be answered. Because Hermitonian is a Hermitian operator it has a spectral decomposition

$$H = \sum_E E|E\rangle\langle E|$$

with eigenvalues E(Energy) and corresponding normalized e-vectors $|E\rangle$: energy e-sts., stationary sts. The lowest energy: *ground state energy*
corresponding energy e-st. : *ground state*
The reason the states $|E\rangle$ are sometimes known as stationary states is because their only change in time is to acquire an overall numerical factor,

$$|E\rangle \rightarrow \exp\left(-iEt/\hbar\right)|E\rangle.$$

As an example, suppose a simgle qubit has Hamiltonian

$$H = \hbar\omega X.$$

The energy eigenstates of this Hamiltonian are obviously the same as the eigenstates fo X, $|+\rangle$ and $|-\rangle$, with coressponding energies $\hbar\omega$ and $-\hbar\omega$. The ground state is therefore $|-\rangle$ and ground state energy is $-\hbar\omega$. What is the connection between Postulate 2′, and Postulate 2?

$$|\psi(t2)\rangle = \exp\left[\frac{-iH(t_2 - t_1)}{\hbar}\right]|\psi(t_1)\rangle = U(t_1, t_2)|\psi(t_1)\rangle$$

where we define

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

Any operator U can be realized in the form $U = \exp(iK)$ for some Hermitian operator K. There is therefore a one-to-one correspondence between the discrete-time description of dynamics using unitary operators, and the continuous time description using Hamiltonians.

In quantum computation and quantum information we often speak of *applying* a unitary operator to a particular quantum system. $\Rightarrow$ Doesn't this contradict what we said earlier, about unitary operators describing the evolution of a *closed* quantum system? After all, if we are 'applying' a unitary operator, then that implies that there is an external 'we' who is interacting with the quantum system, and the is not closed.

An example of this occurs when a laser is focused an atom. After a lot of thought and hard work it is possible to write down a Hamiltonian describing the total atom-laser system. The behavior of the state vector of the atom turns out to be almost but not quite prefectly descibed by another Hamiltonian, the *atomic Hamiltonian*. The atomic Hamiltonian contains terms related to laser intensity, and other parameters of the laser. It is *as if* the evolution of the atom were begin described by a Hamiltonian which we can vary at will, despite the atom not being a closed system.

The upshot is that to gein we will often describe the evolution of quantum systems - even systems which aren't closed - using unitary operators. The main exception to this is quantum measurement.

2.2.3 Quantum measurement

The evolution of the closed system is well described unitary operator, but if the other system like experimentalist observes the system to find what is going on inside the system, an interaction makes the system no longer closed, and thus not necessarily subject to unitary evolution. Postulate 3 explain the effects of measurements on quantum systems,

Postulate 3: Quantum measuerments ared described by an collection M_m of *measurement operators*. These are operators acting on the state space of the system being measured. The index m refers to the measurement outcomes that may occur int the experiment. If the state of the quantum system is $|\psi\rangle$ immediatedly before the measurement then the probability that result m occurs is given by

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle.$$

and the state of the system after the measurement is

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

The measuement operators satisfy the *completeness equation*,

$$\sum_m M_m^\dagger M_m = I.$$

The completeness equation expresses the fact that probabilities sum to one:

$$1 = \sum_m p(m) = \sum_m \langle\psi|M_m^\dagger M_m|\psi\rangle$$

The completeness equation is much easier to check directly, to that's why it appears in the statement of the postulate.
The *measurement of a qubit in the computational basis*:

$$\begin{aligned} M_0 &= |0\rangle\langle 0| \\ M_1 &= |1\rangle\langle 1|. \end{aligned}$$

Each measurement operator is Hermitian, and that $M_0^2 = M_0$, $M_1^2 = M_1$. Thus completeness relationis obeyed, $I = M_0^\dagger M_0 + M_1^\dagger M_1 = M_0 + M_1$. Suppose the state being measure is $|\psi\rangle = a|0\rangle + b|1\rangle$. Then the probability of obtaining measurement outcome 0 is

$$p(0) = \langle\psi|M_0^\dagger M_0|\psi\rangle = \langle\psi|M_0|\psi\rangle = |a|^2.$$

Similarly, $p(1) = |b|^2$. The state after measurement in the two cases is therefore

$$\begin{aligned} \frac{M_0|\psi\rangle}{|a|} &= \frac{a}{|a|}|\psi\rangle \\ \frac{M_1|\psi\rangle}{|b|} &= \frac{b}{|b|}|\psi\rangle \end{aligned}$$

The status of Postulate 3 as a fundamental postulate intrgues many people.
→ According to Postualte 2, the evolution of this larger isolated system can be described by a unitary evolution. Might it be possible to derive Postulate 3 as a consequence of this pircture? Despite considerable investigation along these lines there is still disagreement between physicists about whether or not this is possible.

2.2.4 Distinguishing quantum states

An important application of Postulate 3 is to the problem of *distinguishing quantum states*.

Classical world : distinct states of an object are usually distinguishable.
Quantum mechanics : non-orthogonal quantum states cannot be distinguished.

Alice choose $|\psi_i\rangle (1 \leq i \leq n)$ from some fixed set of states known to Alice and Bob. She gives state $|\psi_i\rangle$ to Bob, whose task is to identify the index i of the state Alice has given him.

1. If states $|\psi_i\rangle$ are orthonormal, the bob can do a quantum measurement to *distinguish* these states.

   Def. measurement operators $M_i \equiv |\psi_i\rangle\langle\psi_i|$, one for each index i
   An additional measurement operators M_0 defined as the positive square root of the positive operator $I - \sum_{i\neq 0} |\psi_i\rangle\langle\psi_i|$. These operators satisfy the completeness relation, and if the state $|\psi\rangle$ is prepared then $p(i) = \langle\psi_i|M_i|\psi_i\rangle = 1$, so the result i occurs with certainty, Thus, it is possible to reliably distinguish the orthonormal states $|\psi_i\rangle$

2. If the states $|\psi\rangle$ are not orthonormal then we can prove that there is *no quantum measurement capable of distinguishing the states.* $\rightarrow$ non-orthogonal states can't be reliably distinguished

   Suppose that measurement distinguishing the non-orthogonal states $|\psi_1\rangle$ and $|\psi_2\rangle$ is possible.
   If the state $|\psi_1\rangle(|\psi_2\rangle)$ is prepared then the probability of measuring j such that $f(j) = 1(f(j) = 2)$ must be 1. Defining $E_i \equiv \sum_{j:f(j)=i} M_j^\dagger M_j$, these observations may be written as:

   $$\langle\psi_1|E_1|\psi_1\rangle = 1; \langle\psi_2|E_2|\psi_2\rangle = 1$$

   Since $\sum_i E_i = I$ it follows that $\sum_i \langle\psi_i|E_i|\psi_i\rangle = 1$, and since $\langle\psi_1|E_1|\psi_1\rangle = 1$ we must have $\langle\psi_1|E_2|\psi_1\rangle = 0$, and thus $\sqrt{E_2}|\psi_1\rangle = 0$. Suppose we decompose $|\psi_2\rangle = \alpha|\psi_1\rangle + \beta|\phi\rangle$, where $|\phi\rangle$ is orthonormal to $|\psi_1\rangle$, $|\alpha|^2 + |\beta|^2 = 1$, and $|\beta| < 1$ since $|\psi_1\rangle$and $|\psi_2\rangle$ are not orthogonal. Then $\sqrt{E_2}|\psi_2\rangle = \beta\sqrt{E_2}|\phi\rangle$, which implies a contradiction with above, as

   $$\langle\psi_2|E_2|\psi_2\rangle = |\beta|^2\langle\phi|E_2|\phi\rangle \leq |\beta|^2 < 1,$$

   where the second last inequality follows from the observation that

   $$\langle\phi|E_2|\phi\rangle \leq \sum_i \langle\phi|E_i|\phi\rangle = \langle\phi|\phi\rangle = 1.$$

2.2.5 Projective measurements

Projective measurements is a special case of the general measurement postulate. Projective measurements actually turn out to be *equivalent* to the general measurement postulate, when they are augmented with the ability to perform unitary transformations.

Projective measurements: A projective measurement is described by an *observable*, M, a Hermitian operator on the state space of the system being observed.

$$M = \sum_m mP_m$$

(spectrally decomposed), where P_m is the projector onto the eigenspace of M with eigenvalue m.
The probability of getting result m of the state $|\psi\rangle$ is given by

$$p(m) = \langle\psi|P_m|\psi\rangle.$$

Given that outcome m occurred, the state of the quantum system immediately after the measurement is

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

If $\sum_m M_m^\dagger M_m = I$, the M_m are Hermitian, and $M_m M_{m'} = \delta_{m,m'}$, then Postulate 3 reduces to a projective measurement as just defined.

It is very easy to calculate average values for projective measurements. By definition, the average value of the measurement is

$$\mathbf{E}(M) = \sum_m mp(m)$$

$$
\begin{aligned}
&= \sum_m m\langle\psi|P_m|\psi\rangle \\
&= \langle\psi|\left(\sum_m mP_m\right)|\psi\rangle \\
&= \langle\psi|M|\psi\rangle \\
&= \langle M\rangle
\end{aligned}
$$

The standard deviation associated to observations of M,

$$
\begin{aligned}
[\Delta(M)]^2 &= \langle(M - \langle M\rangle)^2\rangle \\
&= \langle M^2\rangle - \langle M\rangle^2.
\end{aligned}
$$

For large # limit $\Delta(M) = \sqrt{\langle M^2\rangle - \langle M\rangle^2}$. This formulation of measurement and standard deviations in terms of observables gives rise in an elegant way to results such as the *Heisenberg uncertainty principle*.

Two widely used nomenclatures for measurements deserve emphasis. Rather than giving an observable to describe a projective measurement, often people simply list a complete set of orthogonal projectors P_m satisfying the relations $\sum_m P_m = I$ and $P_m P_{m'} = \delta_{m,m'} P_m$. The corresponding observable implicit in this usage is $M = \sum_m mP_m$. Another widely used phrase, to 'measure in a basis $|m\rangle$', simply means to perform the projective measurements with $|m\rangle\langle m|$.

eg. examples of projective measurements on single qubits.

Measurement of Z on the state $|\psi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ gives result +1 with probability $\langle\psi|0\rangle\langle 0|\psi\rangle = 1/2$, similarly result -1 with probability 1/2.

Generally, suppose $\vec{v}$ is any real three-dimensional unit vector. Then we can define and observable:

$$
\vec{v}\cdot\vec{\sigma} \equiv v_1\sigma_1 + v_2\sigma_2 + v_3\sigma_3.
$$

Measurement of this observable is sometimes referred to as a 'measurement of spin along the $\vec{v}$ axis'.

The uncertainty principle
A, B : Hermitian operators.
$|\psi\rangle$: a quantum state.
Suppose $\langle\psi|AB|\psi\rangle = x + iy$
$\langle\psi|[A,B]|\psi\rangle = 2iy$
$\langle\psi|\{A,B\}|\psi\rangle = 2x$
This implies that

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

By the Cauchy–Schwarz inequality

$$|\langle\psi|AB|\psi\rangle|^2 \leq \langle\psi|A^2|\psi\rangle\langle\psi|B^2|\psi\rangle,$$

with above equation and dropping anti-commutation term gives

$$|\langle\psi|[A,B]|\psi\rangle|^2 \leq 4\langle\psi|A^2|\psi\rangle\langle\psi|B^2|\psi\rangle.$$

Suppose C and D are two observables. Substituting $A = C - \langle C\rangle$ and $B = D - \langle D\rangle$ into the last equation, we contain Heisenberg's uncertainty principle as it is usually states:

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

A common misconception about the uncertainty principle:
measuring an observable C to some 'accuracy' $\Delta(C)$ causes the value of D to be 'disturbed' by an amount $\Delta(D)$ in such a way that some sort of inequality similar to above relation is satisfied.While it is true that measurements in quantum mechanics cause disturbance to the system being measured, this is most emphatically *not* the content of the uncertainty principle.
The correct interpretation of the uncertainty principle is that if we prepare a large number of quantum systems in identical states, $|\psi\rangle$, and then perform measurements of C on some of those systems, and of D in others,then the standard deviation $\Delta(C)$ of the C results times the standard deviation $\Delta(D)$ of the results for D will satisfy the inequality relation.
eg. Consider the observables X, Y when measured for the quantum state $|0\rangle$.

$$\Delta(X)\Delta(Y) \geq \frac{|\langle 0|[X,Y|0\rangle|}{2} = \langle 0|Z|0\rangle = 1.$$

One elementary consequence of this is that $\Delta(X)$ and $\Delta(Y)$ must both be strictly greater than 0, as can be verified by direct calculation.

2.2.6 POVM measurement

Review of Postulate 3 :

1. It gives a rule describing the measurement statistics, that is, the respective probabilities of the different possible measurement outcomes.
2. it gives a rule describing the post measurement state of the system.

However, for some application post-measurement state is of little interest, with the main item of interest being the probabilities of the respective measurement outcomes.

This is the case in an experiment where the system is measured only once, upon conclusion of the experiment. In such instances there is a mathematical tool known as the it Positive Operator-valued Measurement (POVM) formalism

Suppose a measurement described by measurement operators M_m is performed upon a quantum system in the state $|\psi\rangle$. Then the probability of outcome m is given by $p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle$. Define

$$E_m \equiv M_m^\dagger M_m$$

E_m is positive operator such that $\sum_m E_m = I$ and $p(m) = \langle\psi|E_m|\psi\rangle$. Thus the set of E_m are sufficient of determine the probabilities of the different measurement outcomes.

E_m : **POVM** elements
$\{E_m\}$: **POVM**

If a projective measurement satisfies $P_m P_{m'} = \delta_{m,m'} P_m$ and $\sum_m P_m = I$, then all the POVM elements are the same as the measurement operators themselves, since $E_m \equiv P_m^\dagger P_m = P_m$.

Suppose that $\{E_m\}$ is some arbitrary set of positive operators such that $\sum_m E_m = I$. Defining $M_m \equiv \sqrt{E_m}$, so that $\sum_m M_m^\dagger M_m = \sum_m E_m = I$, and therefore the set $\{M_m\}$ describes a measurement with POVM $\{E_m\}$.

For this reason it is convenient to *define* a POVM to be any set of operators $\{E_m\}$ such that:

1. each operator E_m is positive
2. the *completeness relation* $\sum_m E_m = I$ is obeyed
3. the probability of outcome m is given by $p(m) = \langle\psi|E_m|\psi\rangle$

The other example of POVM formalism as a guide for our intuition in quantum computation and quantum information.

Suppose Alice gives Bob a qubit prepared in one of two states $|\psi_i\rangle = |0\rangle$ or $|\psi_2\rangle = |+\rangle$. It is important to Bob to determine whether he has been given $|\psi_1\rangle$ or $|\psi_2\rangle$ with perfect reliability. However it is possible for him to perform a measurement which distinguishes the states some of the time, but *never* makes an error of mis-identification. Consider a POVM containing three elements,

$$\begin{aligned} E_1 &\equiv \frac{\sqrt{2}}{1+\sqrt{2}}|1\rangle\langle 1|, \\ E_2 &\equiv \frac{\sqrt{2}}{1+\sqrt{2}}|-\rangle\langle -|, \\ E_3 &\equiv I - E_1 - E_2. \end{aligned}$$

Bob has $E_1 \rightarrow |\psi_2\rangle$
Bob has $E_2 \rightarrow |\psi_1\rangle$
However, the case of $E_3 \rightarrow$ can not identify the state he have been given.
The key point is that Bob *never* makes a mistake identifying the state he has been given. This infallibility comes at the price that sometimes Bob obtains no information about the identity of the state.

General measurements, projective measurements, and POVMs
Projective measurements : useful for the region of coarsely measurements are meaningful.
In quantum computation and quantum information we aim for an exquisite level of control over the measurements that may be done, and consequently it helps to use a more comprehensive formalism for the description of measurements.
Of course projective measurements augmented by unitary operations turn out to be completely *equivalent* to general measurements.
Why we start with the general formalism

1. General measurements are in some sense mathematically simpler than projective measurements, since they involve fewer restrictions on the measurement operators. This simpler structure also gives rise to many useful properties for general measurements that are not possessed by projective measurements.

2. There are important problems in quantum computation and quantum information – such as the optimal way to distinguish a set of quantum states – the answer to which involves a general measurement, rather than a projective measurement.

3. Projective measurements are repeatable in the sense that if we perform a projective measurement once, and obtain the outcome m, repeating the measurement gives the outcome m again and does not change the state. After a measurement, $|\psi_m\rangle = (P_m|\psi\rangle) / \sqrt{\langle\psi|P_m|\psi\rangle}$ and repeating P_m

2.2.7 Phase

$|\psi\rangle$ and $e^{i\theta}|\psi\rangle$: Two states are same, up to the global phase factor $e^{i\theta}$.

$$\because \langle\psi|e^{-i\theta}M_m^\dagger M_m e^{i\theta}|\psi\rangle = \langle\psi|M_m^\dagger M_m|\psi\rangle$$

Therefore observational point of view these two states are identical.

Consider the states

$$\frac{|0\rangle + |1\rangle}{\sqrt{2}} \qquad \text{and} \qquad \frac{|0\rangle - |1\rangle}{\sqrt{2}}$$

In the first state the amplitude of $|0\rangle$ is $1/\sqrt{2}$. For the second state the amplitude is $-1/\sqrt{2}$. In each case the *magnitude* of the amplitudes is the same, but they differ in sign. More generally,

$$|\psi\rangle = a|0\rangle + b|1\rangle \qquad (a = e^{i\theta}b),$$

we say that two amplitudes *differ by a relative phase*. More generally still, two states are said to *differ by a relative phase* in some basis if each of the amplitudes in that basis is related by such a phase factor. $|+\rangle$ and $|-\rangle$ are the same up to a relative phase shift because the $|0\rangle$ amplitudes are identical, and the $|1\rangle$ amplitudes differ only by a relative phase factor of -1

For relative phase the phase factors may vary from amplitude to amplitude. This makes the relative phase a basis–dependent concept unlike global phase. As a result, states which differ only by relative phases in some basis give rise to physically observable differences in measurement statistics, and it is not possible to regard these states as physically equivalent.

2.2.8 Composite systems

Postulate 4: The state space of a composite physical system is the tensor product of the state spaces of the component physical systems. Moreover, if we have systems numbered 1 through n, and system number i is prepared in the state $|\psi_i\rangle$, then the joint state of the total system is $|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots |\psi_n\rangle$.

Why? We certainly expect that there be *some canonical way* of describing composite systems in quantum mechanics. The *superposition principle of quantum mechanics*, so called, could be expected. (One expect the linearity) This is not a derivation, since we are not taking the superposition principle as a fundamental part of our description of quantum mechanics.

The proof of the statement, that is projective measurements together with unitary dynamics are sufficient to implement a general measurement, makes use of composite quantum systems.

Suppose a quantum system with state space Q, measurement operators M_m on Q. Introduce *ancilla system*, with state space M, having a orthonormal basis $|m\rangle$.
The ancilla system can be regarded as merely mathematical device appearing in the construction, or it can be interpreted physically as an extra quantum system introduced into the problem.

$|0\rangle$: any fixed state of M, define an operator U on products $|\psi\rangle|0\rangle$ of states $|\psi\rangle$ from Q with the state $|0\rangle$ by

$$U|\psi\rangle|0\rangle \equiv \sum_m M_m|\psi\rangle|m\rangle.$$

Using the orthonormality of the states $|m\rangle$ and the completeness relation $\sum_m M_m^\dagger M_m = I$, U preserves inner products between states of the form $|\psi\rangle|0\rangle$,

$$\begin{aligned}\langle\phi|\langle 0|U^\dagger U|\psi\rangle|0\rangle &= \sum_{m,m'}\langle\phi|M_m^\dagger M_m|\psi\rangle\langle m|m'\rangle \\ &= \sum_m \langle\phi|M_m^\dagger M_m|\psi\rangle \\ &= \langle\phi|\psi\rangle\end{aligned}$$

Probability $p(m)$ of projectors $P_m \equiv I_Q \otimes |m\rangle\langle m|$ is

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

The joint state of the system QM after measurement, conditional on result m occurring, is given by

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

It follows that the state of system M after the measurement is $|m\rangle$, and the state of system Q is

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

Thus unitary dynamics, projective measurements, and the ability to introduce ancillary systems, together allow any measurement of the form described in Postulate 3 to be realized.

Postulate 4 also enables us to define one of the most interesting and puzzling ideas associated with composite quantum systems - *entanglement*.

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

This state hase the remarkable property that there are no single qubit states $|a\rangle$ and $|b\rangle$ such that $|\psi\rangle = |a\rangle|b\rangle$.

We say that a state of a coposite system having this property (that is can't be written as a product of states of its component systems) is an *entangled* state. For reasons which nobody fully understands, entangled states play a crucial role in quantum computation and quantum information. (Teleportation, dense coding, Bell's inequality)

2.2.9 Quantum mechanics: a global view

Review :

- Postulate 1 sets the arena for quantum mechanics, by specifying how the state of an isolated quantum system is to be described.
- Postulate 2 tells us that the dynamics of evolution.
- Postulate 3 tells us how to extract information from our quantum system by giving a prescription for the desription of measurement.
- Postulate 4 tells us how the state spaces of different quantum systems may be combined to give a description of the composition system.

Odd stuff :

- We cannot directly observe the state vector.
- Classical physics tells us that the fundamental properties of an object, like energy, position, and velocity, are directly accessible to observation. → are not fundamental quantities in quantum mechanics.
  It is as thought there is a *hidden world* in quantum mechanics, which we can only indirectly and imperfectly access.
- Observation in quantum mechanics is an invasive procedure that typically changes the state of the system.

2.3 Application: superdense coding

Two parties, conventionally known as Alice and Bob, want to transmit some classical information. Suppose Alice has two classical bits of information which she wishes to send Bob, but is only allowed to send a single qubit to Bob.

Suppose Alice and Bob initially share a pair of qubits in the entangled state

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

Alice is initially in possession of the first qubit, while Bob has possessionof the second qubit. Note that $|\psi\rangle$ is a fixed state(from Alice or third party).

She wishes to send the bits string;

- '00' : she does noting at all to her quibit.
- '01' : she applies the phase flip Z to her qubit.
- '10' : she applies the quantum NOT gate X.
- '11' : the she applies the iY gate to her qubit.

The four resulting states are easily seen to be:

$$\begin{aligned} 00 : |\psi\rangle &\rightarrow \frac{|00\rangle + |11\rangle}{\sqrt{2}} \\ 01 : |\psi\rangle &\rightarrow \frac{|00\rangle - |11\rangle}{\sqrt{2}} \end{aligned}$$

$$
\begin{aligned}
10 : |\psi\rangle &\rightarrow \frac{|01\rangle + |10\rangle}{\sqrt{2}} \\
11 : |\psi\rangle &\rightarrow \frac{|01\rangle - |10\rangle}{\sqrt{2}}.
\end{aligned}
$$

Then Alice sends her qubit to Bob, giving Bob possession of both qubits, then by doing a measurement in the Bell basis Bob can dertime which of the four possible bit string Alice sent. It means Alice, interacting with only a single qubit i, is able to transmit two bits of informtion to Bob.

Of course, two qubits are involved in the protocol, but Alice never need interact with the second qubit. Classically, the task Alice accomplishes would have been impossible had she only transmitted a single classical bit.

2.4 The density operator

The density operator : The alternate mathematical formulation which is equivalent to the state vector formalism and provides a much more convenient language in quantum mechanics.

2.4.1 Ensembles of quantum states

The density operator formalism provides a convenient means for describing quantum systems whose state is not completely known.

Suppose a quantum system is in one of a number of states $|\psi_i\rangle$, where i is an index, with respective probabilities p_i.

$\{p_i, |\psi_i\rangle\}$: an *ensemble of pure states*

The density operator(density matrix) for the system is defined by

$$\rho \equiv \sum_i p_i |\psi_i\rangle\langle\psi_i|$$

All postulates of quantum mechanics can be reformulated in terms of the density operator language.

The state vector formalism and the density operator language give the same results; however it is sometimes much easier to approach problems from one point of view rather than the other.

An unitary evolution for sys. initiated by $|\psi_i\rangle$ with p_i can be described

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \rightarrow \sum_i p_i U|\psi_i\rangle\langle\psi_i|U^\dagger = U\rho U^\dagger$$

For a measurement operator M_m, when the initial state was $|\psi_i\rangle$, the probability of getting result m is

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

The probability of obtaining result m is

$$
\begin{aligned}
p(m) &= \sum_i p(m|i)p_i && (2.1)\\
&= \sum_i p_i \mathrm{tr}(M_m^\dagger M_m|\psi_i\rangle\langle\psi_i|) && (2.2)\\
&= tr(M_m^\dagger M_m \rho). && (2.3)
\end{aligned}
$$

If the initial state was $|\psi_i\rangle$ then the state after obtainging the result m is

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

Thus, after a measurement which yields the result m we have an ensemble of states $|\psi_i^m\rangle$ with respective probabilities $p(i|m)$. The coresponding density operator ρ_m is

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

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

$$\begin{aligned}\rho_m &= \sum_i p_i \frac{M_m|\psi_i\rangle\langle\psi_i|M_m^\dagger}{\mathrm{tr}(M_m^\dagger M_m \rho)} \\ &= \frac{M_m \rho M_m^\dagger}{\mathrm{tr}(M_m^\dagger M_m \rho)}\end{aligned}$$

We will complete rephrasing of quantum postulates in next section.

A pure state : if a quantum system whose stat $|\psi\rangle$ is known exactly then we said the system is in a pure state. In this case the density operator is simply $\rho = |\psi\rangle\langle\psi|$.
Otherwise, ρ is in a *mixed state*; it is said to be a *mixture* of the different pure states in the ensemble for ρ.

A pure state stisfies $\mathrm{tr}(\rho^2) = 1$ ($\rho^2 = \rho$), while a mixed state satisfiles $\mathrm{tr}(\rho^2) < 1$ ($\rho^2 \neq \rho$).

1. Sometime people use the term 'mixed state' as a catch-all to include both pure and mixed quantum states. (The origin for this usage seems to be that it implies that the writer is not necessarily *assuming* that a state is pure)
2. The term 'pure state' is often used in reference to a state vector $|\psi\rangle$, to distinguish it from a density operator ρ.

Imagine a quantum system is prepared in the state ρ_i with probability p_i. For pure states, suppose ρ_i from some ensemble $\{p_{ij}, |\psi_{ij}\rangle\}$, so the density matrix of a subsystem i is $\rho_i = \sum_j p_{ij}|\psi_{ij}\langle\psi_{ij}|\rangle$ and the probability for being in the state $|\psi_{ij}\rangle$ is $p_i p_{ij}$. The density matrix for the system is thus

$$\begin{aligned}\rho &= \sum_{ij} p_i p_{ij} |\psi_{ij}\rangle\langle\psi_{ij}| \\ &= \sum_i p_i \rho_i\end{aligned}$$

We say that ρ is a *mixture* of the states ρ_i with probabilities p_i. This concept of a mixture comes up repeatedly in the analysis of problems like quantum noise, there the effect of the noise is to introduce ignorance into our knowledge of the quantum state.

eg) Imagine that our record of the result m of the measuremnt was lost. We would have a quantum system in the state ρ_m with probability $p(m)$, bout would no longer know the actual value of m. The state of such a quantum system would therefore be described by the density operator

$$\begin{aligned}\rho &= \sum_m p(m)\rho_m \\ &= \sum_m \mathrm{tr}(M_m^\dagger M_m \rho)\frac{M_m \rho M_m^\dagger}{\mathrm{tr}(M_m^\dagger M_m \rho)} \\ &= \sum_m M_m \rho M_m^\dagger,\end{aligned}$$

a nice compact formula which may be used as the staring point for analysis of further operations on the system.

2.4.2 General properties of the density operator

The density operator : describing ensembles of quantum states.

Theorem 2.5: **(Characterization of density operators)** An operator ρ is the density operator associated to some ensemble $\{p_i, |\psi_i\rangle\}$ iff it satisfies the conditions:

1. **(Trace dondition)** ρ has trace equal to one
2. **(Positivity condition)** ρ is a positive operator.

Pf.
Suppose $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ is a density operator. Then

$$\mathrm{tr}(\rho) = \sum_i p_i \mathrm{tr}(|\psi_i\rangle\langle\psi_i|) = \sum_i p_i = 1,$$

so the trace condition $\mathrm{tr}(\rho) = 1$ is satisfied. Suppose $|\phi\rangle$ is an arbitrary vector in state space. Then

$$\begin{aligned} \langle\phi|\rho|\phi\rangle &= \sum_i p_i \langle\phi|\psi_i\rangle\langle\psi_i|\phi\rangle \\ &= \sum_i p_i |\langle\phi|\psi_i\rangle|^2 \\ &\geq 0, \end{aligned}$$

so the positivity condition is satisfied.

Conversely, suppose ρ is any operator satisfying the trace and positivity conditions. Since ρ is positive, it must have a spectral decomposition

$$\rho = \sum_j \lambda_j |j\rangle\langle j|,$$

where the vectors $|j\rangle$ are orthogonal, and λ_j are real, non-negative eigenvalues of ρ. From the trace condition we see that $\sum_j \lambda_i = 1$. Therefore, a system in state $|j\rangle$ with probability λ_j will have density operator ρ. That is, the ensemble $\lambda_j, |j\rangle$ is an ensemble of states giving rise to the density operator ρ.

We can define a density operator to be a positive operator ρ which has trace equal to one. Making this definition allows us to reformulate the postulates of quantum mechanics in the density operator picture.

Postulate 1 : Associated to any isolated physical system is a complex vector space with inner product (that is, a Hilbert space) known as *state space* of the system. The system is completely described by its *density operator*, which is a positive operator ρ with trace one, acting on the state space of the system. If a quantum system is in the state ρ_i with probability p_i, then the density operator for the system is $\sum_i p_i \rho_i$.

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

$$\rho' = U\rho U^\dagger$$

Postulate 3 : Quantum measurements are described by a collection M_m of *measurement operators*. These are operators acting on the state apace of the system being measured. The index m refersto the measurement outcomes that may occur in the experiment. If the state of the quantum system is ρ immediately before the measurement then the probability that result m occurs is given bye

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

and the state of the system after the measurement is

$$\frac{M_m \rho M_m^\dagger}{\mathrm{tr}(M_m^\dagger M_m \rho)}$$

The measurement operators satisfy the completeness relation,

$$\sum_m M_m^\dagger M_m = I.$$

Postulate 4 : The state space of a composite physical system is the tensor product of the state spaces of the component physical systems. Moreover, if we have systems numberd 1 through n, and system number i is prepared in the state ρ_i, then the joint state of the total system is $\rho_i \otimes \rho_2 \otimes \ldots \rho_n$.

Above postulates are mathematically equivalent to the description in terms of the state vector.

The density operator approach really shines for two applications:

- The description of quantum systems whose state is not known
- The description of subsystems of a composite quantum system

It is tempting fallacy to suppose that the eigenvlues and eigenvectors of a density matrix have some special significance with regard to the ensemble of quantum states represented by that density matrix.

eg) Suppose that a quantum system with density matrix

$$\rho = \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1|.$$

And suppose we define other vectors

$$\begin{aligned} |a\rangle &\equiv \sqrt{\frac{3}{4}}|0\rangle + \sqrt{\frac{1}{4}}|1\rangle \\ |b\rangle &\equiv \sqrt{\frac{3}{4}}|0\rangle - \sqrt{\frac{1}{4}}|1\rangle \end{aligned}$$

and the quantum system is prepared in the state $|a\rangle$ with probability 1/2 and in the state $|b\rangle$ with probability 1/2. Then

$$\rho = \frac{1}{2}|a\rangle\langle a| + \frac{1}{2}|b\rangle\langle b| = \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1|$$

That is, these two different ensembles of quantum states give rise to the same density matrix. In general, the eigenvectors and eigenvalues of a density matrix just indicate *one* of many possible ensembles that may give rise to a specific density matrix, and there is no reason to suppose it is an especially privileged ensemble.

What classes of ensembles does give rise to a particular density matrix? The solution to this problem has suprisingly many applications in quantum computation and quantum information, notably in the understanding of quantum noise and quantum error-correction. For the solution it is convenient to make use of vectors $|\tilde{\psi}_i\rangle$ which may not be normalized to unit length. We say the set $|\tilde{\psi}_i\rangle$ *generates* the operator $\rho \equiv \sum_i |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|$, and thus the connection to the usual ensemble pircture of density operators is expressed by the equation $|\tilde{\psi}_i\rangle = \sqrt{p_i}|\psi_i\rangle$. When do two sets of vectors, $|\tilde{\psi}_i\rangle$ and $|\tilde{\phi}_j\rangle$ generate the same operator ρ? The solution to this problem will enable us to answer the question of what ensembles give rise to a given density matrix.

Theorem 2.6: **Unitary freedom in the ensemble for density matrices** The set $|\tilde{\psi}_i\rangle$ and $|\tilde{\phi}_j\rangle$ generate the same density matrix iff

$$|\tilde{\psi}_i\rangle = \sum_j u_{ij}|\tilde{\phi}_j\rangle$$

where u_{ij} is a unitary matrix of complex numbers, with indices i and j, and we 'pad' whichever set of vectors $|\tilde{\psi}_i\rangle$ or $|\tilde{\phi}_j\rangle$ is smaller with additional vectors 0 so that the two sets have the same number of elements.

As a consequence of the theorem, note that $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$ for normalized state $|\psi_i\rangle, |\phi_i\rangle$ and probability distributions p_i and q_i iff

$$\sqrt{p_i}|\psi_i\rangle = \sum_j u_{ij}\sqrt{q_i}|\phi_j\rangle,$$

for some unitary matrix u_{ij}, and we may pad the smaller ensemble with entries having probability zero in order to make the two ensembles the same size. Thus, Theorem 2.6 characterizes the freedom in ensembles $\{p_i, |\psi_i\rangle\}$ giving rise to a given density matrix ρ.

Pf.
Suppose $|\tilde{\psi}_i\rangle = \sum_i u_{ij}|\tilde{\phi}_j\rangle$ for some unitary u_{ij}. Then

$$\begin{aligned}
\sum_i |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i| &= \sum_{ijk} u_{ij}u_{ik}^*|\tilde{\phi}_j\rangle\langle\tilde{\phi}_k| \\
&= \sum_{jk}\left(\sum_i u_{ki}^\dagger u_{ij}\right)|\tilde{\phi}_j\rangle\langle\tilde{\phi}_k| \\
&= \sum_{jk}\delta_{kj}|\tilde{\phi}_j\rangle\langle\tilde{\phi}_k| \\
&= \sum_j |\tilde{\phi}_j\rangle\langle\tilde{\phi}_j|
\end{aligned}$$

which shows that $|\tilde{\psi}_i\rangle$ and $|\tilde{\phi}_j\rangle$ generate the same operator.

Conversely, suppose

$$A = \sum_i |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i| = \sum_j |\tilde{\phi}_j\rangle\langle\tilde{\phi}_j|.$$

Let $A = \sum_k \lambda_k|k\rangle\langle k|$ be a decomposition for A such that $|k\rangle$ are orthonormal, and the λ_k are strictly positive. Let $|\psi\rangle$ be any vector orthonormal to the space spanned by $|\tilde{k}\rangle$ ($|\tilde{k}\rangle = \sqrt{\lambda_i}|k\rangle$), so $\langle\psi|\tilde{k}\rangle\langle\tilde{k}|\psi\rangle = 0$ for all k, and

$$0 = \langle\psi|A|\psi\rangle = \sum_i \langle\psi|\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|\psi\rangle = \sum_i |\langle\psi|\tilde{\psi}_i\rangle|^2.$$

Thus $\langle\psi|\tilde{\psi}_i\rangle = 0$ for all i and all $|\psi\rangle$ orthonormal to the space spanned by the $|\tilde{k}\rangle$. It follows that each $|\tilde{\psi}_i\rangle$ can be expressed as a linear combination of the $|\tilde{k}\rangle$, $|\tilde{\psi}_i\rangle = \sum_k c_{ik}|\tilde{k}\rangle$. Since $A = \sum_k |\tilde{k}\rangle\langle\tilde{k}| = \sum_i |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|$ we see that

$$\sum_k |\tilde{k}\rangle\langle\tilde{k}| = \sum_{kl}\left(\sum_i c_{ik}c_{il}^*\right)|\tilde{k}\rangle\langle\tilde{l}|.$$

The operators $|\tilde{k}\rangle\langle\tilde{l}|$ are easily seen to be *linearly independent*, and thus it must be that $\sum_i c_{ik}c_{il}^* = \delta_{kl}$. This ensures that we may append extra columns to c to obtain a unitary matrix v such that $|\tilde{\psi}_i\rangle = \sum_k v_{ik}|\tilde{k}\rangle$, where we have appended zero vectors to the list of $|\tilde{k}\rangle$. Similarly, we can find a unitary matrix w such that $|\tilde{\phi}_j\rangle = \sum_k w_{jk}|\tilde{k}\rangle$. Thus $|\tilde{\psi}_i\rangle = \sum_j u_{ij}|\tilde{\phi}_j\rangle$, where $u = vw^\dagger$ is unitary.

2.4.3 The reduced density operator

The deepest application of the density operator may be as a descriptive tool for *subsystems* of a composite quantum system. Such a description is provided by *reduced density operator*.

Suppose we have physical systems A and B, whose state is described by a density operator ρ^{AB}. The reduced density operator for system A is defined by

$$\rho^A \equiv tr_B(\rho^{AB}),$$

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

$$\begin{aligned}\mathrm{tr}_B(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|) &\equiv |a_1\rangle\langle a_2|\mathrm{tr}(|b_1\rangle\langle b_2|) \\ &= \langle b_2|b_1\rangle|a_1\rangle\langle a_2|\end{aligned}$$

where $|a_1\rangle$ and $|a_2\rangle$ are any two vectors in the state space A, and $|b_1\rangle$, $|b_2\rangle$ are of B.

The physical justification for making this identification is that the reduced density operator provides the correct measrement statistics for measurements made on system A.

*eg)*Suppose a quantum system is in the product state $\rho^{AB} = \rho \otimes \sigma$, where ρ is a density operator for system A, and σ is a density operator for system B. Then

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

which is the result we intuitively expect. Similarly, $\rho^B = \sigma$ for this state.

eg) For the Bell state, the density operator

$$\begin{aligned}\rho &= \left(\frac{|00\rangle + |11\rangle}{\sqrt{2}}\right)\left(\frac{\langle 00| + \langle 11|}{\sqrt{2}}\right) \\ &= \frac{|00\rangle\langle 00| + |11\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 11|}{2}.\end{aligned}$$

Tracing out the second qubit, we find the reduced density operator of the first qubit,

$$\begin{aligned}\rho^1 &= \mathrm{tr}_2(\rho) \\ &= \frac{\mathrm{tr}_2(|00\rangle\langle 00|) + \mathrm{tr}_2(|11\rangle\langle 00|) + \mathrm{tr}_2(|00\rangle\langle 11|) + \mathrm{tr}_2(|11\rangle\langle 11|)}{2} \\ &= \frac{|0\rangle\langle 0|\langle 0|0\rangle + |1\rangle\langle 0|\langle 1|0\rangle + |0\rangle\langle 1|\langle 0|1\rangle + |1\rangle\langle 1|\langle 1|1\rangle}{2} \\ &= \frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2} \\ &= \frac{I}{2}\end{aligned}$$

Notice that this state is a *mixed state*, since $\mathrm{tr}((I/2)^2) = 1/2 < 1$. This is quite a remarkable result. The state of the joint system of two qubits is a pure state; however, the first qubit is in a mixed state, a state about which we apparently do not have maximal knowledge. This strange property, that the joint state of a system can be completely known, yet a sybsystem be in mixed states, another hallmark of quantum entanglement.

Quantum teleportatin and the reduced density operator

A useful application of the reduced density operator is to the analysis of quantum teleportation.

Before Alice makes her measurement the quantum state of the three qubits is:

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

Measuring in Alice's computational basis, the state of the system after the measurement is:

$$\begin{aligned}
&|00\rangle\left[\alpha|0\rangle + \beta|1\rangle\right] \text{ with probability } \frac{1}{4}\\
&|01\rangle\left[\alpha|1\rangle + \beta|0\rangle\right] \text{ with probability } \frac{1}{4}\\
&|10\rangle\left[\alpha|0\rangle - \beta|1\rangle\right] \text{ with probability } \frac{1}{4}\\
&|11\rangle\left[\alpha|1\rangle - \beta|0\rangle\right] \text{ with probability } \frac{1}{4}
\end{aligned}$$

The density operator of the system is thus

$$\begin{aligned}
\rho = \frac{1}{4}[\quad & |00\rangle\langle 00|\left(\alpha|0\rangle + \beta|1\rangle\right)\left(\alpha^*\langle 0| + \beta^*\langle 1|\right) + |01\rangle\langle 01|\left(\alpha|1\rangle + \beta|0\rangle\right)\left(\alpha^*\langle 1| + \beta^*\langle 0|\right)\\
& + |10\rangle\langle 10|\left(\alpha|0\rangle - \beta|1\rangle\right)\left(\alpha^*\langle 0| - \beta^*\langle 1|\right) + |11\rangle\langle 11|\left(\alpha|1\rangle - \beta|0\rangle\right)\left(\alpha^*\langle 1| - \beta^*\langle 0|\right)].
\end{aligned}$$

Tracing out Alice's system, we see that the reduced density operator of Bob's system is

$$\begin{aligned}
\rho^B &= \frac{1}{4}\left[\left(\alpha|0\rangle + \beta|1\rangle\right)\left(\alpha^*\langle 0| + \beta^*\langle 1|\right) + \left(\alpha|1\rangle + \beta|0\rangle\right)\left(\alpha^*\langle 1| + \beta^*\langle 0|\right)\right.\\
&\quad \left. + \left(\alpha|0\rangle - \beta|1\rangle\right)\left(\alpha^*\langle 0| - \beta^*\langle 1|\right) + \left(\alpha|1\rangle - \beta|0\rangle\right)\left(\alpha^*\langle 1| - \beta^*\langle 0|\right)\right]\\
&= \frac{2\left(|\alpha|^2 + |\beta|^2\right)|0\rangle\langle 0| + 2\left(|\alpha|^2 + |\beta|^2\right)|1\rangle\langle 1|}{4}\\
&= \frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2}\\
&= \frac{I}{2},
\end{aligned}$$

Thus, the state of Bob's system after Alice has performed the measurement but before Bob has learned the measurement result is $I/2$. This state has no dependence upon the state $|\psi\rangle$ being teleported, and thus any measurements performed by Bob will contain no informatin about $|\psi\rangle$, thus preventing Alice from using teleportation to transmit information to Bob faster than light.

2.5 The Schmidt decomposition and purifications

The study of composite quantum systems are at the heart of quantum computation and quantum information. Two additional tools of great value are the *Schmidt decomposition* and *purifications*.

Theorem 2.7 (Schmidt decomposition). *$|\psi\rangle$: pure state of a composite system, AB. $\exists|i_A\rangle$ and $|i_B\rangle$: orthogonal states. of A and B, respectively such that*

$$|\psi\rangle = \sum_i \lambda_i |i_A\rangle|i_B\rangle$$

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

This result is very useful.

If $|\psi\rangle$ is a pure st. of a composite system, AB, then by Schmidt decomposition

$$\rho^A = \sum_i \lambda^2 |i_A\rangle\langle i_A|$$

$$\rho^B = \sum_i \lambda^2 |i_B\rangle\langle i_B|$$

so, the eigenvalues of ρ^A and ρ^B are identical (λ^2) for both density operators. Many important properties of quantum system are completely determined by the eigenvalues of the reduced density operator of the system, so for a pure state of a composite system such properties will be the same for both sytems.

eg) two qubits, $(|00\rangle + |01\rangle + |11\rangle)/\sqrt{3}$

$$\begin{aligned}
\rho^A &= \frac{1}{3}\left(2|0\rangle\langle 0| + |1\rangle\langle 1|\right) \\
\rho^B &= \frac{1}{3}\left(|0\rangle\langle 0| + 2|1\rangle\langle 1|\right) \\
\mathrm{tr}\left(\left(\rho^A\right)^2\right) &= \frac{1}{9}\mathrm{tr}\left(4|0\rangle\langle 0| + |1\rangle\langle 1|\right) \\
&= \frac{1}{9}\mathrm{tr}\left(|0\rangle\langle 0| + 4|1\rangle\langle 1|\right) \\
&= \frac{5}{9}
\end{aligned}$$

Proof
System A, B have state spaces of the same dimension.
$|j\rangle$, $|k\rangle$: any fixed orthonormal bases for systems A and B

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

By the singular value decomposition, $a = udv$, where d is a diagonal matrix with non-negative elements, and u and v are unitary matrices

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

Defining $|i_A\rangle \equiv \sum_j u_{ji}|j\rangle$, $|i_B\rangle \equiv \sum_k v_{ik}|k\rangle$, and $\lambda = d_{ii}$

$$|\psi\rangle = \sum_i \lambda_i |i_A\rangle|i_B\rangle$$

$|i_A\rangle$ and $|i_B\rangle$: *Schmidt bases* for A and B
λ_i : *Schmidt number* for the state $|\psi\rangle$. In some sense, Schmidt number is the 'amount' of entanglement between systems A and B. The Schmidt number is preserved under unitary transformations on system A or system B alone.

Purification
Suppose a system ρ^A and introduce another system, R, and define a pure state $|AR\rangle$ for the joint system AR such that $\rho^A = \mathrm{tr}_R\left(|AR\rangle\langle AR|\right)$. That is, the pure state $|AR\rangle$ reduces to ρ^A when we look at system A alone. Purification allows us to associate pure states with mixed states.

To prove that purification can be done for *any* state, we explain how to construct a system R and purification $|AR\rangle$ for ρ^A.

A system ρ^A has orthonormal decomposition $\rho^A = \sum_i p_i |i^A\rangle\langle i^A|$. To purify ρ^A, introduced R which has the same state space as system A, with orthonormal basis states $|i^R\rangle$, and define a pure state for the combined system

$$|AR\rangle \equiv \sum_i \sqrt{p_i}|i^A\rangle|i^R\rangle$$

We now calculate the reduced density operator for system A corresponding to the state $|AR\rangle$

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

Thus $|AR\rangle$is a purification of ρ^A.

The procedure used to purify a mixed state of system A is to define a pure state whose Schmidt basis for system A is just the basis in which the mixed state is diagonal, with the Schmidt coefficients being the square root of the eigenvalues of the density operator being purified.

2.6 EPR and the Bell inequality

A compelling example of an essential difference between quantum and classical physics.

For an object, we assume that the physical properties of that object have an existence independent of observation. That is, measurements merely act to reveal such physical properties.

QM: an unobserved particle does not posses physical properties that exist independent of observation. Rather, such physical properties arise as a consequence of measurements performed upon the system.

EPR were interested in what they termed 'elements of reality'. Their belief was that any such element of reality must be represented in any complete physical theory. The goal of the argument was to show that quantum mechanics is not a complete phsical theory, by identifying elements of reality that were not included in quantum mechanics.

The way they attempted to do was by introducing what they claimed was a sufficient condition for a physical property to be an element of reality, namely, that it be possible to predict with certainty the value that property will have, immediately before measurement.

Box 2.7: Anti-correlation in the EPR experiment The prepared two qubit state

$$|\psi\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$$

The measurement of observable $\vec{v} \cdot \vec{\sigma}$ on each qubit will get $+1$ or -1 for each qubit. It turns out that no matter what choice of $\vec{v}$ we make, the results of the two measuerments are always opposite to one another.
Suppose $|a\rangle$ and $|b\rangle$ are the eigenstates of $\vec{v} \cdot \vec{\sigma}$. Then

$$\begin{aligned} |0\rangle &= \alpha|a\rangle + \beta|b\rangle \\ |1\rangle &= \gamma|a\rangle + \delta|b\rangle. \end{aligned}$$

Substituting we obtain

$$\frac{|01\rangle - |10\rangle}{\sqrt{2}} = (\alpha\delta - \beta\gamma)\frac{|ab\rangle - |ba\rangle}{\sqrt{2}}$$

But $\alpha\delta - \beta\gamma$ is the determinant of the unitary matrix, and thus is equal to a phase factor $e^{i\theta}$ for some real θ. Thus

$$\frac{|01\rangle - |10\rangle}{\sqrt{2}} = \frac{|ab\rangle - |ba\rangle}{\sqrt{2}}$$

As a result, if a measuerment of $\vec{v}\cdot\vec{\sigma}$ is performed on both qubits, then we can see that a result of $+1(-1)$ on the first qubit implies a result of $-1(+1)$ on the second qubit.

eg) Consider, an entangled pair of qubits belonging to Alice and Bob, respectively:

$$\frac{|01\rangle - |10\rangle}{\sqrt{2}}.$$

Alice performs a measurement of spin along the $\vec{v}$ axis. Suppose Alice recives the result $+1$. Bob will measuer -1 on his qubit if he also measures sping along the $\vec{v}$ axis.

Above result seems to correspond to an element of reality, by the EPR criterion, and should be represented in any complete physical theory. **However** standard quantum mechanics merely tells one how to caculate the probabilities of the respective measurement outcomes if $\vec{v} \cdot \vec{\sigma}$ is measured. Standard quantum meachanics certainly does not include any fundamental element intended to represent the value of $\vec{v} \cdot \vec{\sigma}$, for all unit vector $\vec{v}$.

The goal of EPR was to show that quantum mechanics is incomplete, by demonstrating that quantum mechanics lacked some essential 'element of reality', by their criterion. They hoped to force a return to a more classical view of the world, one in which systems could be ascribed properties which existed independently of measurements performed on those systems.

EPR defeated nearly 30 years after its publishing, by invalidation of *Bell's Inequality*. Bell's inequality is *not* a result about quantum mechanics. To obtain Bell's inequality, we're going to do a thought(gedanken) experiment, which we will analyze using our common sense notions of how the world works(classically) – the sort of notions Einstein and his collaborators thought Nature ought to obey. After we have done the common sense analysis, we will perform a quantum mechanical analysis which we can show *is not consistent with the common sense analysis*

Thought Experiment:

- Charlie prepares to particles.
- Charlie sends one particle to Alice, and the second particle to Bob.
- Once Alice receives her particle, she performs a measurement on it using randomly one of two different measurements. These measurements are of physical properties which we shall labe P_Q and P_R, respectively. For simplicity, suppose that the measurements can each have one of two outcomes, $+1$ or -1. $P_Q \to Q$ and $P_R \to R$

- Similarly, $P_S \to S$ and $P_T \to T$, each taking value $+1$ or -1.
- Alice and Bob do their measurements *at the same time*. Therefore the measurement which performs cannot disturb the result of Bob's measurement, since physical influences cannot propagate faster than light.

We are going to do some simple algebra with the quantity $QS + RS + RT - QT$.

$$QS + RS + RT - QT = (Q + R)S + (R - Q)T.$$

Because $R, Q = \pm 1$ it follows that either $(Q+R)S = 0$ or $(R-Q)T = 0$. In either case, it is easy to see from above equation that $QS + RS + RT - QT = \pm 2$.

Suppse that $p(q, r, s, t)$ is the probability that, before the emasurements are performed, the system is in a state where $Q = q, R = r, S = s, T = t$. These probabilities may depend on how Charlie performs his preparation, and on experimental noise. Letting $\mathrm{E}(\cdot)$ denote the mean value of a quantity, we have

$$\begin{aligned} \mathrm{E}(QS + RS + RT - QT) &= \sum_{qrst} p(q, r, s, t)(qs + rs + rt - qt) \\ &\leq \sum_{qrst} p(q, r, s, t) \times 2 \\ &= 2. \end{aligned}$$

Also,

$$\begin{aligned} \mathrm{E}(QS + RS + RT - QT) &= \sum_{qrst} p(q, r, s, t)qs + \sum_{qrst} p(q, r, s, t)rs + \sum_{qrst} p(q, r, s, t)rt - \sum_{qrst} p(q, r, s, t)qt \\ &= \mathrm{E}(QS) + \mathrm{E}(RS) + \mathrm{E}(RT) - \mathrm{E}(QT). \end{aligned}$$

We obtain th *Bell inequality*,

$$\mathrm{E}(QS) + \mathrm{E}(RS) + \mathrm{E}(RT) - \mathrm{E}(QT) \leq 2$$

This result is also often known as the *CHSH inequality* after the initials of its four discoverers. It is part of a larger set of inequalities known generically as Bell inequalities.

Quantum mechanical analysis:
Charlie prepares a quantum system of two qubits in the state

$$|\psi\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$$

Alice and Bob perform measurements of the following observables:

$$\begin{aligned} Q &= Z_1 \\ R &= X_1 \\ S &= \frac{-Z_2 - X_2}{\sqrt{2}} \\ T &= \frac{Z_2 - X_2}{\sqrt{2}}. \end{aligned}$$

Simple calculations show that the average values for these observables are:

$$\langle QS\rangle = \frac{1}{\sqrt{2}}; \langle RS\rangle = \frac{1}{\sqrt{2}}; \langle RT\rangle = \frac{1}{\sqrt{2}}; \langle QT\rangle = \frac{1}{\sqrt{2}};$$

Thus,

$$\langle QS\rangle + \langle RS\rangle + \langle RT\rangle - \langle QT\rangle = 2\sqrt{2}.$$

Above result violates the Bell inequality. And the experimental results denoted that the Bell inequality is not obeyed by Nature.

What does it mean? I means that one or more of the assumptions that went into the derivation of the Bell inequality must be incorrect.

There are two assumptions made in the proof of Bell inequality which are questionable:

1. The assumption that the physical properties P_Q, P_R, P_S, P_T have definite values Q, R, S, T which exist independent of observation. This is sometimes known as the assumption of *realism*.
2. The assumption that Alice performing her measurement does not influence the result of Bob's measurement. This is sometimes known as the assumption of *locality*.

These two assumptions together are known as the assumption of *local realism*. They are certainly intuitively plausible assumptions about how the world works, and they fit our everyday experience.

The world is *not* locally realistic. Bell's inequality points out to the conclusion that either or both of locality and realism must be dropped from our view of the world if are to develop a good intuitive understading of quantum mechanics.

Bell's inequality teaches us that entanglement is a fundamentally new resource in the world that goes essentially beyond classical resources.

Chapter 4

Quantum circuit

Quantum circuit model → universal quantum gates

4.1 Quantum algorithms

Many interesting problems are impossible to solve on a classical computer → requirement of astronomial resources.

new algorithms of quantum computation

- Shor's *quantum Fourier transformation*
- Grover's algorithm for performing *quantum searching*

The quantum search algorithm has many potential applications,

- can be used to extract statistics from an unordered data set
- can be used to speed up algorithms for some problems in **NP**
- can be used to speed up the search for keys to cryptosystems such as DES

The quantum Fourier transform also has many interesting applications

- to solve the discrete logarithm and factoring problems.
- finding hidden subgroup (a generation of finding the period of a periodic function)

Why are there so few quantum algorithms known which are better than their classical counterparts? The answer is that coming up with good quantum algorithms seems to be a difficult problem.

1. Algorithm design is not an easy business
2. Our intuitions are much better adapted to the classical world than they are to the quantum world

This chapter will provide an efficient and powerful language for describing quantum algorithms, the language of quantum circuits – assemblies of discrete sets of components which describe computational procedures.

4.2 Single qubit operations

A single qubit is a vector $|\psi\rangle = a|0\rangle + b|1\rangle$ parameterized by two complex numbers satisfying $|a|^2 + |b|^2 = 1$. Operations on a qubit must preserve this norm, and thus are described by 2×2 unitary matrices. Of these, some of the most important are the Pauli matrices;

$$X \equiv \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; Y \equiv \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}; Z \equiv \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Three other quantum gates will play a large part in what follows, the Hadamard gate(H), phase gate (S), and $\pi/8$ gate (T):

$$H \equiv \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}; S \equiv \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}; T \equiv \begin{bmatrix} 1 & 0 \\ 0 & \exp(i\pi/4) \end{bmatrix}.$$

A couple of useful algebraic facts to keep in mind are that $H = (X + Z)/\sqrt{2}$ and $S = T^2$. The reason that T gate is called the $\pi/8$ gate is for unimportant global phase T is equal to a gate which has $\exp(\mp i\pi/8)$ apearing on its diagonals.

A single qubit in the state $a|0\rangle + b|1\rangle$ can be visualized as a point (θ, ϕ) on the unit sphere, where $a = \cos(\theta/2), b = e^{i\phi}\sin(\theta/2)$, and a can be taken to be real because the overall phase of the state is unobservable. This is called Bloch sphere representation, and the vector $(\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)$ is called the Bloch vector.

The Pauli matrices give rise to three useful classes of unitary matrices when they are exponentiated, the *rotation operators* about the $\hat{x}$, $\hat{y}$, and $\hat{z}$ axes, defined by the equations:

$$\begin{aligned} R_x(\theta) &\equiv e^{-i\theta X/2} = \cos\frac{\theta}{2} I - i\sin\frac{\theta}{2} X = \begin{bmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} \\ R_y(\theta) &\equiv e^{-i\theta Y/2} = \cos\frac{\theta}{2} I - i\sin\frac{\theta}{2} Y = \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} \\ R_z(\theta) &\equiv e^{-i\theta Z/2} = \cos\frac{\theta}{2} I - i\sin\frac{\theta}{2} Z = \begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix} \end{aligned}$$

If $\hat{n} = (n_x, n_y, n_z)$ is a real unit vector in three dimensions then we generalize the previous definitions by defining a rotation by θ about the $\hat{n}$ axis by the equation

$$R_{\hat{n}} \equiv \exp(-i\theta\hat{n}\cdot\vec{\sigma}/2) = \cos\left(\frac{\theta}{2}\right) I - i\sin\left(\frac{\theta}{2}\right)(n_x X + n_y Y + n_z Z),$$

An arbitrary unitary operator on a single qubit can be written in may ways as a combination of rotations, together with global phase shifts on the qubit.

Theorem 4.1 : (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 and δ such that

$$U = e^{i\alpha} R_z(\beta) R_y(\gamma) R_z(\delta).$$

Corollary 4.2: Suppose U is a unitary gate on a single qubit. Then there exist unitary operators A, B, C on a single qubit such that $ABC = I$ and $U = e^{i\alpha} AXBXC$, where α is some overall phase factor.

Proof
In the notation of Theorem 4.1, set $A \equiv R_z(\beta)R_y(\gamma/2)$, $B \equiv R_y(-\gamma/2)R_z(-(\delta+\beta)/2)$ and $C \equiv R_z((\delta-\beta)/2)$. Note that

$$ABC = R_z(\beta) R_y\left(\frac{\gamma}{2}\right) R_y\left(-\frac{\gamma}{2}\right) R_z\left(-\frac{\delta+\beta}{2}\right) R_z\left(\frac{\delta-\beta}{2}\right) = I.$$

Since $X^2 = I$, $XYX = -Y$, and $XR_y(\theta)X = R_y(-\theta)$, we see that

$$XBX = XR_y\left(-\frac{\gamma}{2}\right)XXR_z\left(-\frac{\delta+\beta}{2}\right)X = R_y\left(\frac{\gamma}{2}\right)R_z\left(\frac{\delta+\beta}{2}\right).$$

Thus

$$\begin{aligned} AXBXC &= R_z(\beta)R_y\left(\frac{\gamma}{2}\right)R_y\left(\frac{\gamma}{2}\right)R_z\left(\frac{\delta+\beta}{2}\right)R_z\left(\frac{\delta-\beta}{2}\right) \\ &= R_z(\beta)R_y(\gamma)R_z(\delta). \end{aligned}$$

Thus $U = e^{i\alpha}AXBXC$ and $ABC = I$, as required. □

4.3 Controlled operations

'If A is true, then do B' : *Controlled operation* – one of the most useful in computing, both classical and quantum.

CNOT : $|c\rangle|t\rangle \rightarrow |c\rangle|t \oplus c\rangle$
The matrix representation of CNOT is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

More generally, suppose U is an arbitrary single qubit unitary operation. A *controlled-U* is a two qubit operation with a control and a target qubit. If the control qubit is set then U isapplied to the target qubit, otherwise the target qubit is left alone; $|c\rangle|t\rangle \rightarrow |c\rangle U^C|t\rangle$.

Our immediate goal is to understand how to implement the controlled-U operation for arbitrary single qubit U, using only single qubit operation and the CNOT gate.

Applying the phase shift $\exp(i\alpha)$ on the target qubit, controlled by the control qubit. If the control qubit is $|0\rangle$, then the target qubit is left alone, while if the control qubit is $|1\rangle$, a phase shift $\exp(i\alpha)$ is applied to the target.

$$|00\rangle \rightarrow |00\rangle,\ \ |01\rangle \rightarrow |01\rangle\ \ |10\rangle \rightarrow e^{i\alpha}|10\rangle,\ \ |11\rangle \rightarrow e^{i\alpha}|11\rangle$$

The construction of the controlled-U operation (Figure 4.6): recalling from Corollary 4.2 that U may be written in the form $U = e^{i\alpha}AXBXC$, where A, B and C are single qubit operations such that $ABC = I$.

Suppose that the control qubit is set. Then the operation $e^{i\alpha}AXBXC = U$ is applied to the second qubit. If on the other hand, the control qubit is not set, then the operation $ABC = I$ is applied to the second qubit; that is, no change is made. That is , this circuit implements the controlled-U operation.

What about conditioning on multiple qubits?
eg) Toffoli gate, which flips the third qubit, the target qubit, conditioned on the first two qubits, the control qubits, being set to one. More generaly, suppose we have $n + k$ qubits, and U is a k qubit unitary operator. Then we define the controlled operation $C^n(U)$ by the equation

$$C^n(U)|x_1x_2\ldots x_n\rangle|\psi\rangle = |x_1x_2\ldots x_n\rangle U^{x_1x_2\ldots x_n}|\psi\rangle$$

where $x_1x_2\ldots x_n$ in the exponent of U means the product of the bits $x_1, x_2, \ldots, x_n$. That is, the operator U is applied to the last k qubits if the first n qubits are all equal to one, and otherwise, nothing is done.

(Fig 4.7)

Larger k can be dealt with using essentially the same methods, however for $k > 2$ there is the added complication that we don't (yet) know how to perform arbitrary operations on k qubits.

Suppose U is a single qubit unitary operator, and V is a unitary operator chosen so that $V^2 = U$. Then the operation $C^2(H)$ may be implemented using the circuit shown in Figure 4.8.

The familiar Toffoli gate is an especialy important special case of the $C^2(U)$ operation, the $C^2(X)$. Defining $V \equiv (1-i)(I+iX)/2$ and noting that $V^2 = X$, we see that Figure 4.8 gives an implementation of the Toffoli gate in terms of one and two qubit operations. (Classical reversible gates are not sufficient to implement the Toffoli gate, or universal computation. But, in the quantum case we see that one and two qubit reversible gates are sufficient to implement the Toffoli gate, and will eventually prove that they suffice for universal computation)

Ultimately we will show that any unitary operation can be composed to an arbitrarily good approximation from just the Hadamard, phase, controlled-NOT and $\pi/8$ gates. Figure 4.9 illustrates a simple circuit for the Toffoli gate made up of just Hadamard, phase, CNOT and π/8 gates. (Figure 4.9)

How may we implement $C^n(U)$ gates using our existing repertoire of gates, where U is an arbitrary single qubit unitary operation? A particularly siple circuit for achieving this task is illustrated in Figure 4.10. The circuit divides up into three stages, and makes use of a small number $(n-1)$ of working qubits, which all start and end in the state $|0\rangle$. Suppose the control qubits are in the computational basis state $|c_1, c_2, \ldots, c_n\rangle$. The first stage of the circuit is to revesibly AND all the control bits $c_1, \ldots, c_n$ together to produce the product $c_1 \cdot c_2 \ldots c_n$. To do this, the first gate int the circuit ANDs c_1 and c_2 together, using a Toffoli gate, changing the state of the first work qubit to $|c_1 \cdot c_2\rangle$. The next Toffoli gate ANDs c_3 with the product $c_1 \cdot c_2$, changing the state of the second work qubit to $|c_1 \cdot c_2 \cdot c_3\rangle$. We continue applying Toffoli gates in this fashion, until the final work qubit is in the state $|c_1 \cdot c_2 \ldots c_n\rangle$. Next, a U operation on the target qubit is performed, conditional on the final work qubit being set to one. That is, U is applied iff all of c_1 through c_n are set. Finally the last part of the circuit just reverses the steps of the first stage, returning all the work qubits to their initial state, $|0\rangle$.

In the controlled gates we have been considering, conditional dynamics on the target qubit occurs if the control bits are set to *one*. Of course, there is nothing special about one, and it is often useful to consider dynamics which occur conditional on the control bit being set to zero.

Figure 4.11

Figure 4.12

4.4 Measurement

A final element used in quntum circuits, almost implicitly sometimes, is measurement. In our circuits, we shall denote a projective measurement in the computational basis using a 'meter' symbol.

It is conventinal to not use any special symbols to denote more general measurements, because they can always be represented by unitary transforms with ancilla qubits followed by prejective measurements.

Two important principles:

Principle of deferred measurement: Measurements can always be moved from an intermediate stage of a quantum circuit to the end of the circuit; if the measurement results are used at any stage of the circuit then the classically controlled 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.

1 – Often, quantum measurements are performed as an intermediate step in a quantum circuit, and the measurement results are used to conditionally control subsequent quantum gates. However, such measurements can *always* be moved to the end of the circuit.

2 – Imagine you have a quantum circuit containing just two qubits, and only the first qubit is measured at the end of the circuit. Then the measurement statistics observed at this time are completely determined by the reduced density

matrix of the first qubit. The measurement of the second qubit does not change the result of the measurement of the first qubit.

4.5 Universal quantum gates

A small set of gates (AND, OR, NOT) is *universal* for classical computation. In fact, since the Toffoli gate is universal for classical computation, quantum circuits subsume classical circuits. A similar universality result is true for quantum computation, where a set of gates is said to be *universal for quantum computation* if any unnitary operation may be approximated to arbitrary accuracy bye a quantum circuit involving only those gates.

Any unitary operation can be approximated to arbitrary accuracy using Hadamard, phase, CNOT, and $\pi/8$ gates.

1. An arbitrary operator may be expressed *exactly* as a product of unitary operators that each acts non-trivially only on a subspace spanned by two computational basis states.
2. The second construction combines the first construction with the results of the previous section to show that an arbitrary unitary operator may be expressed *exactly* using single qubit and CNOT gates.
3. Single qubit operation may be approximated to arbitrary accuracy using the Hadamard, phase, and $\pi/8$ gates. This in turn implies that any unitary operation can be approximated to arbitrary accuracy using Hadamard, phase, CNOT, and $\pi/8$ gates.

4.5.1 Two-level unitary gates are universal

Consider a unitary matrix U which acts on a d-dimensional Hilbert space. We explain how U may be decompsed into a product of *two-level unitary matrices*; unitary matrices which act non-trivially only on two-or-fewer vector components.

Suppose that U has the form

$$U = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & j \end{bmatrix}.$$

We will find two-level unitary matrices $U_1, \ldots, U_3$ such that

$$U_3 U_2 U_1 U = I.$$

It follows that

$$U = U_1^\dagger U_2^\dagger U_3^\dagger.$$

U_1, U_2 and U_3 are all two-level unitary matrices, and it is easy to see that their inverse, $U_1^\dagger$, $U_2^\dagger$ and $U_3^\dagger$ are also two-level unitary matrices.

Use the following procedure to construct U_1: if $b = 0$ then set

$$U_1 \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

if $b \neq 0$ then set

$$U_1 \equiv \begin{bmatrix} \frac{a^*}{\sqrt{|a|^2+|b|^2}} & \frac{b^*}{\sqrt{|a|^2+|b|^2}} & 0 \\ \frac{b}{\sqrt{|a|^2+|b|^2}} & \frac{-a}{\sqrt{|a|^2+|b|^2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus

$$U_1 U = \begin{bmatrix} a' & d' & g' \\ 0 & e' & h' \\ c' & f' & j' \end{bmatrix}.$$

The key point to note is that the middle entry in the left had comlumn is zero.

Now apply a similar procedure to find a two-level matrix U_2 such that $U_2 U_1 U$ has no entry in the *bottom left* corner. That is, if $c' = 0$ we set

$$U_2 \equiv \begin{bmatrix} a'^* & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

while if $c' \neq 0$ then we set

$$U_2 \equiv \begin{bmatrix} \frac{a'^*}{\sqrt{|a'|^2+|c'|^2}} & 0 & \frac{c'^*}{\sqrt{|a'|^2+|c'|^2}} \\ 0 & 1 & 0 \\ \frac{c'}{\sqrt{|a'|^2+|c'|^2}} & 0 & \frac{-a'}{\sqrt{|a'|^2+|c'|^2}} \end{bmatrix}.$$

In either case, when we carry out the matrix multiplication we find that

$$U_2 U_1 U = \begin{bmatrix} 1 & d'' & g'' \\ 0 & e'' & h'' \\ 0 & f'' & j'' \end{bmatrix}.$$

Since U, U_1 and U_2 are unitary, it follows that $U_2 U_1 U$ is unitary, and thus $d'' = g'' = 0$ since the first row must have norm 1. Finally, set

$$U_3 \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & e''^* & f''^* \\ 0 & h''^* & j''^* \end{bmatrix}.$$

It is now easy to verify that $U_3 U_2 U_1 U = I$, and thus $U = U_1^\dagger U_2^\dagger U_3^\dagger$, which is a decomposition of U into two-level unitaries.

More generally, suppose U acts on a d-dimensional space. Then, in a similar fashion to the 3×3 case, we can find two-level unitary matrices $U_1, \ldots, U_{d-1}$ such that the matrix $U_{d-1} U_{d-2} \ldots U_1 U$ has a one in the top left hand corner, and all zeroes elsewhere in the first row and column. We then repeat this procedure for the $d-1$ by $d-1$ unitary submatrix in the lower right hand corner of $U_{d-1} U_{d-2} \ldots U_1 U$, ans so on, with the end result that an arbitrary $d \times d$ unitary matrix may be written

$$U = V_1 \ldots V_k$$

where the matrices V_i are two-level unitary matrices, and $k \leq (d-1) + (d-2) + \ldots + 1 = d(d-1)/2$.

For specific unitry matrices, it may be possible to find much more efficient decompositions, but as you will now show there exist matrices which cannot be docomposed as a product of fewer than d-1 two-level unitary matrices!

4.5.2 Single qubit and CNOT gates are universal

Now we show that single qubit and CNOT gates together can be used to implement an arbitrary two-level unitary operation on the state space of n qubits.

Suppose U is a two-level unitary matrix on an n qubit quantum computer. Suppose in particular that U acts non-trivially on the space spanned by the computational basis state $|s\rangle$ and $|t\rangle$, where $s = s_1 \ldots s_n$ and $t = t_1 \ldots t_n$ are the binary expansions for s and t. Let $\tilde{U}$ be the non-trivial 2×2 unitary submatrix of U; $\tilde{U}$ can be thought of as a unitary operator on a single qubit.

Our immediate goal is to construct a circuit implementing U, built from single qubit and CNOT gates. To do this, we need to make use of *Gray codes*. Suppose we have distinct binary numbers, s and t. A *Gray code* connecting s and t is a sequence of binary numbers, starting with s and concluding with t, such that adjacent mumbers of the list differ in exactly one bit. For instance, with $s = 101001$ and $t = 110011$ we have the Gray code

$$\begin{array}{cccccc} 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{array}$$

Let g_1 through g_m be the elements of a Gray code connecting s and t, with $g_1 = s$ and $g_m = t$. Note that we can always find a Gray code such that $m \leq n + 1$ since s and t can differ in at most n locations.

The basic idea of the quantum circuit implementing U is to perform a sequence of gates effecting the state changes $|g_1\rangle \rightarrow |g_2\rangle \rightarrow \ldots \rightarrow |g_{m-1}\rangle$, then to perform a controlled-$\tilde{U}$ operation, with the target qubit located at the single bit where g_{m-1} and g_m differ, and then to undo the first stage, transforming $|g_{m-1}\rangle \rightarrow |g_{m-2}\rangle \rightarrow \ldots \rightarrow |g_1\rangle$.

A more precies description of the implementation is as follows. The first step is to swap the state $|g_1\rangle$ and $|g_2\rangle$. Suppose g_1 and g_2 differ at the ith qubit, conditional on the values of the other qubits being identical to those in both g_1 and g_2. Next we use a controlled operation to swap $|g_2\rangle$ and $|g_3\rangle$. We continue in this fashion until we swap $|g_{m-2}\rangle$ with $|g_{m-1}\rangle$.

$$\begin{array}{rcl} |g_1\rangle & \rightarrow & |g_{m-1}\rangle \\ |g_2\rangle & \rightarrow & |g_1\rangle \\ |g_{m-1}\rangle & \rightarrow & |g_{m-2}\rangle \end{array}$$

Suppose g_{m-1} and g_m differ in the jth bit. We apply a controlled-$\tilde{U}$ operation with the jth qubit as target, conditional on the other qubits having the same values as appear in both g_m and g_{m-1}. Finally, we complete the U operation by undoing the swap operations.

eg)

Suppose we wish to implement the two-level unitary transformation

$$U = \begin{bmatrix} a & 0 & 0 & 0 & 0 & 0 & 0 & c \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ b & 0 & 0 & 0 & 0 & 0 & 0 & d \end{bmatrix}$$

a, b, c and d are any complex numbers such that $\tilde{U} \equiv \begin{bmatrix} a & c \\ b & d \end{bmatrix}$ is a unitary matrix. We write a Gray code connecting 000 and 111:

$$\begin{array}{ccc} A & B & C \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{array}$$

From this we read off the required circuit, shown in Figure 4.16. The first two gates shuffle the states so that $|000\rangle$ gets swapped with $|011\rangle$. Next the operation $\tilde{U}$ is applied to the first qubit of the states $|011\rangle$ and $|111\rangle$, conditional on the second and third qubits being in the state $|11\rangle$. Finally, we unshuffle the states, ensuring that $|011\rangle$ gets swapped back with the state $|000\rangle$.

Returning to the general case, we see that implementing the two-level unitary operation U requires at most $2(n - 1)$ controlled operations to swap $|g_1\rangle$ with $|g_{m-1}\rangle$ and then back again. Each of thes controlled operations can

be realized using $O(n)$ single qubit and CNOT gates; the controlled-$\tilde{U}$ operation also requires $O(n)$ gates. Thus, implementing U requires $O(n^2)$ single qubit and CNOT gates.

2^n-dimensional state space of n qubits may be written as a product of $O(2^{2n}) = O(4^n)$ two-level unitary operations. Combining these results, we see that an arbitrary unitary operation on n qubits can be implemented using a circuit containing $O(n^2 4^n)$ single qubit and CNOT gates. Obviously, this construction does not provide terribly efficient quantum circuits. However, the construction is close to optimal in the sense that there are unitary operations that require an exponential number of gates to implement. Thus, to find fast quantum algorithms we will clearly need a different approach that is taken in the universality construction.

4.5.3 A discrete set of universal operations

No straightforward method is known to implement all these gates in a fashion thich is resistant to errors. Fortunately, in this section we'll find a discrete set of gates which can be used to perform universal quantum computation, and in Chapter 10 we'll show how to perform these gates in an error-resistant fashion, using quantum error-correcting codes.

Approximating unitary operators

A discrete set of gates can't be used to implement an arbitrary unitary operation *exactly*, since the set of unitary operations is continuous. Rather, it turns out that a discrete set can be used to *approximate* any unitary operation.

Suppose U and V are two unitary operators on the same state space. U is the target unitary operator that we wish to implement, and V is the unitary operator that is actually implemented in practice. We define the *error* when V is implemented instead of U by

$$E(U, V) \equiv \max_{|\psi\rangle} \|(U - V)|\psi\rangle\|,$$

where the maximum is over all normalized quantum states $|\psi\rangle$ in the state space.

Box 4.1: Approximating quantum circuits

Suppose a quantum system starts in the state $|\psi\rangle$, and we perform either the unitary operation U, or V. Following this, we perform a measurement. Let M be a POVM element associated with the measurement, and let P_U (or P_V) be the probability of obtaining the corresponding measurement outcome if the operation U (or V) was performed. Then

$$|P_U - P_V| = |\langle\psi|U^\dagger M U|\psi\rangle - \langle\psi|V^\dagger M V|\psi\rangle|.$$

Let $|\Delta\rangle \equiv (U - V)|\psi\rangle$.

$$\begin{aligned} |P_U - P_V| &= |\langle\psi|U^\dagger M|\Delta\rangle - \langle\Delta|MV|\psi\rangle| \\ &\leq |\langle\psi|U^\dagger M|\Delta\rangle| + |\langle\Delta|MV|\psi\rangle| \quad (\because \pm\langle\psi|U^\dagger MV|\psi\rangle) \\ &\leq \||\Delta\rangle\| + \||\Delta\rangle\| \quad (\because \||\Delta\rangle\| = \langle\psi|\Delta\rangle) \\ &\leq 2E(U,V). \end{aligned}$$

The inequality $|P_U - P_V| \leq 2E(U,V)$ gives quantitative expression to the idea that when the error $E(U,V)$ is small, the difference in probabilities between measurement outcomes is also small.
Suppose we perform a sequance $V_1, V_2, \ldots, V_m$ of gates intended to approximate some other sequence of gates, $U_1, U_2, \ldots, U_m$. Then it turns out that the error caused by the entire sequence of imperfect gate is at most the sum of the errors in the individual gates,

$$E(U_m U_{m-1} \ldots U_1, V_m V_{m-1} \ldots V_1) \leq \sum_{j=1}^{m} E(U_j, V_j).$$

To prove this we start with the case $m = 2$. Note that for some state $|\psi\rangle$ we have

$$\begin{aligned} E(U_2U_1, V_2V_1) &= \|(U_2U_1 - V_2V_1|\psi\rangle)\| \\ &= \|(U_2U_1 - V_2U_1)|\psi\rangle + (V_2U_1 - V_2V_1)|\psi\rangle\|. \end{aligned}$$

Using the triangle inequality $\||a\rangle + |b\rangle\| \leq \||a\rangle\| + \||b\rangle\|$, we obtain

$$\begin{aligned} E(U_2U_1, V_2V_1) &= \|(U_2 - V_2)U_1|\psi\rangle\| + \|V_2(U_1 - V_1)|\psi\rangle\| \\ &= E(U_2, V_2) + E(U_1, V_1), \end{aligned}$$

which was the desired result. The result for general m follows by induction.

This measure of error has the interpretation that if $E(U,V)$ is small, then any measurement performed on the state $V|\psi\rangle$ will give approximately the same measurement statistis as a measurement of $U|\psi\rangle$, for any initial state $|\psi\rangle$. More precisely, we show that if M is a POVM element in an arbitrary POVM, and P_U (or P_V) is the probability of obtaining this outcome if U (or V) were performed with a starting state $|\psi\rangle$, then

$$|P_U - P_V| \leq 2E(U,V).$$

Thus, if $E(U,V)$ is small, then measurement outcomes occur with similar probabilities, regardless of whether U or V were performed. If we perform a sequence of gates $V_1, \ldots, V_m$ intended to approximate some other sequence of gates $U_1, \ldots, U_m$, then the errors add at most linearly,

$$E(U_m U_{m-1} \ldots U_1, V_m V_{m-1} \ldots V_1) \leq \sum_{j=1}^{m} E(U_j, V_j).$$

Above approximation results are extremely useful. Suppose we wish to perform a quantum circuit containing m gates, U_1 through U_m. Unfortunately, we are only able to approximate the gate U_j by the gate V_j. In order that the probabilities tolerance $\Delta > 0$ of the correct probabilities, it suffices that $E(U_j, V_j) \leq \Delta/(2m)$, by above results.

We're going to consider two different discrete sets of gates, both of which are universal. The first set, the *standard set* of universal gates, consists of the Hardamard, phase, controlled-NOT and $\pi/8$ gates. The second set of gates we consider consists of the Hadamard gates, phase gate, the controlled-NOT gate, and the Toffoli gate.

We begin the universality proof by showing that the hadamard and $\pi/8$ gates can be used to approximate any singel qubit unitary operation to arbitrary accuracy. Consider the gates T and HTH. T is, up to an unimportant global phase, a rotation by $\pi/4$ radians around the $\hat{z}$ axis on the Bloch sphere, while HTH is a rotation by $\pi/4$ radians around the $\hat{x}$ axis on the Bloch sphere. Composing these two operations gives, up to a global phase,

$$\begin{aligned} \exp\left(-i\frac{\pi}{8}Z\right)\exp\left(-i\frac{\pi}{8}X\right) &= \left[\cos\frac{\pi}{8}I - i\sin\frac{\pi}{8}Z\right]\left[\cos\frac{\pi}{8} - i\sin\frac{\pi}{8}X\right] \\ &= \cos^2\frac{\pi}{8}I - i\left[\cos\frac{\pi}{8}(X+Z) + \sin\frac{\pi}{8}Y\right]\sin\frac{\pi}{8} \end{aligned}$$

This is a rotation of the Bloch sphere about an axis along $\vec{n} = (\cos\frac{\pi}{8}, \sin\frac{\pi}{8}, \cos\frac{\pi}{8})$ with corresponding unit vector $\hat{n}$, and through an angle θ defined by $\cos(\theta/2) \equiv \cos^2\frac{\pi}{8}$.That is, using only the Hadamard and $\pi/8$ gates we can construct $R_{\hat{n}}(\theta)$. Moreover, this θ can be shown to be an irrational multiple of 2π. Proving this latter fact is a little beyond our scope.

Next, we show that repeated iteration of $R_{\hat{n}}(\theta)$ can be used to approximate to arbitrary accuracy any rotation $R_{\hat{n}}(\alpha)$. To see this, let $\delta > 0$ be the desired accuracy, and let N be an integer larger thatn $2\pi/\delta$. Define θ_k so that $\theta_k \in [0, 2\pi)$ and $\theta_k = (k\theta) \bmod 2\pi$. Then the pigeonhole principle implies that there are distinct j and k in the range $1, \ldots, N$ such that $|\theta_k - \theta_j| \leq 2\pi/N < \delta$. Assume that $k > j$, so we have $\theta_{k-j} < \delta$. Since $j \neq k$ and θ is an irrational multiple of 2π we must have $\theta_{k-j} \neq 0$. It follows that the sequence $\theta_{l(k-j)}$ fills up the interval $[0, 2\pi)$ as l is varied, so that adjacent members of the sequence are no more than δ apart. It follows that for any $\epsilon > 0$ there exists an n such that

$$E(R_{\hat{n}}(\alpha), R_{\hat{n}}(\theta)^n) < \frac{\epsilon}{3}.$$

We are now in position to verify that any single qubit operation can be approximated to arbitrary accuracy using the Hadamard and $\pi/8$ gates. For any α

$$HR_{\hat{n}}(\alpha)H = R_{\hat{m}}(\alpha),$$

where $\hat{m}$ is a unit vector in the direction $(\cos\frac{\pi}{8}, -\sin\frac{\pi}{8}, \cos\frac{\pi}{8})$, from which it follows that

$$E(R_{\hat{m}}(\alpha), R_{\hat{m}}(\theta)^n) < \frac{\epsilon}{3}.$$

But an arbitrary unitary U on a single qubit may be written as

$$U = R_{\hat{n}}(\beta)R_{\hat{m}}(\gamma)R_{\hat{n}}(\delta),$$

upto an unimportant global phase shift. Thus

$$E(U, R_{\hat{n}}(\theta)^{n_1}HR_{\hat{n}}(\theta)^{n_2}HR_{\hat{n}}(\theta)^{n_3}) < \epsilon.$$

That is, given any single qubit unitary operator U and any $\epsilon > 0$ it is possible to approximate U to within ϵ using a circuit composed of Hadamard gates and $\pi/8$ gates alone.

Since the $\pi/8$ and Hadamard gates allow us to approximate any single qubit unitary operator, it follows from the arguments of Section 4.5.2 that we can approximate any m gate quantum circuit, as follows. Given a quantum circuit containing m gates, we may approximate it using Hadamard, controlled-NOT and $\pi/8$ gates. If we desire an accuracy of ϵ for the entire circuit, then this may be achieved by approximating each single qubit unitary using the above procedure to within ϵ/m and applying the chaining inequality to obtain an accuracy of ϵ for the entire circuit.

How efficient is this procedure for approximating quantum circuits using a discrete set of gates?

eg) Suppose that approximating an arbitrary single qubit unitary to within a distance ϵ were to require $\Omega(2^{1/\epsilon})$ gates from the discrete set. Then to approximate the m gate quantum circuit considered in the previous paragraph would

require $\Omega(m2^{m/\epsilon})$ gates, an exponential increase over the original circuit size! Fortunately, the rate of convergence is much better than this. Intuitively, it is plausible that the sequence of angles θ_k 'fills in' the interval $[0, 2\pi)$ in a more or less uniform fashion, so that to approximate an arbitrary single qubit gate ought to take roughly $\Theta(1/\epsilon)$ gates from the discrete set.

# of gates to approximate an m gate circuit to accuracy $\epsilon \rightarrow \Theta(m^2/\epsilon)$.

However, a much faster rate of convergence can be proved. The *Solovay-Kitaev theorem* implies that an arbitrary single qubit gate may be approximated to an accuracy ϵ using $O(\log^c(1/\epsilon))$ gates from our discrete set, where c is a constant approximately equal to 2. $\rightarrow O(m \log^c(m/\epsilon))$

To sum up, we have shown that the Hadamard, phase, controlled-NOT and $\pi/8$ gates are universal for quantum computation in the sense that given a circuit containing CNOTs and arbitrary single qubit unitaries it is possible to simulate this circuit to good accuracy using only this discrete set of gates.

4.5.4 Approximating arbitrary unitary gates is generically hard

We've seen that any unitary transformation on n qubits can be built up out of a small set of elementary gates. Is it always possible to do this efficiently? In fact, most unitary transformations can only be implemented very inefficiently.

How myny gates does it take to generate an arbitrary state of n qubits? A simple counting argument shows that this requires exponentially many operations, in general.

Suppose we have g different types of gates available, and each gate works on at most f input qubits. f and g may be considered to be constants. Suppose we have a quantum circuit containg m gates, starting from the computational basis state $|0\rangle^{\otimes n}$. For any particular gate in the circuit there are therefore at most $\left[\begin{array}{c} n \\ g \end{array} \right]^g = O(n^{fg})$ possible choices. It follows that at most $O(n^{fgm})$ different states may be computed using m gates.

Suppose we wish to approximate a particular state, $|\psi\rangle$, to within a distance $epsilon$. The space of state vectors of n qubits can be regarded as just the unit $(2^{n+1} - 1)$-sphere. Suppose the n qubit state has amplitudes $\psi_j = X_j + Y_j$ of the jth amplitude. The normalization condition for quantum states can be written $\sum_j (X_j^2 + Y_j^2) = 1$, which is just the condition for a point to be on the unit sphere in 2^{n+1} real dimensions, that is, the unit $(2^{n+1} - 1)$-sphere.

Similarly, the surface area of radius ϵ near $|\psi\rangle$ is approximately the same as the volume of a $(2^{n+1} - 2)$-sphere of radius ϵ. Using the formular $S_k(r) = 2\pi^{(k+1)/2} r^k / \Gamma((k+1)/2)$ for the surface area of a k-sphere of radius r, and $V_k(r) = 2\pi^{(k+1)/2} r^{k+1} / [(k+1)\Gamma((k+1)/2)]$ for the volume of a k-sphere of radius r, we see that the number of pathes needed to cover the state space goes like

$$\frac{S_{2^{n+1}-1}(1)}{V_{2^{n+1}-2}(\epsilon)} = \frac{\sqrt{\pi}\Gamma(2^n - \frac{1}{2})(2^{n+1} - 1)}{\Gamma(2^n)\epsilon^{2^{2n+1}-1}},$$

where Γ is the usual generalization of the factorial function. But $\Gamma(2^n - 1/2) \geq \Gamma(2^n)/2^n$, so the number of patches required to cover the space is at least

$$\Omega\left(\frac{1}{\epsilon^{2^{n+1}-1}}\right).$$

Recall that the number of pathces which can be reached in m gates is $O(n^{fgm})$, so in order to reach all the ϵ-patches we must have

$$O(n^{fgm}) \geq \Omega\left(\frac{1}{\epsilon^{2^{n+1}-1}}\right)$$

which gives us

$$m = \Omega\left(\frac{2^n \log(1/\epsilon)}{log(n)}\right).$$

That is, there are states of n qubits which take $\Omega(2^n \log(1/\epsilon)/\log(n))$ operations to approximate to within a distance ϵ. This is exponential in n, and thus is 'difficult', in the sense of computational complexity. Furthermore, this immediately implies that there are unitary transformations U on n qubits which take $\Omega(2^n \log(1/\epsilon)/\log(n))$ operations to approximate by a quantum circuit implementing an operation V such that $E(U,V) \leq \epsilon$. By contrast, using our universlity constructions and the Solovay–Kitaev theorem it follows that an arbitrary unitary operation U on n qubits may be approximated to within a distance ϵ using $O(n^2 4^n \log^c(n^2 4^n/\epsilon))$ gates.

Thus, to within a polynomial factor the construction for uniersality we have given is optimal; unfortunately, it does not address the problem of determining which families of unitary operations can be computed efficiently in the quantum circuit model.

4.5.5 Quantum computational complexity

We present on result about quantum complexity classes, relating the quantum complexity class **BQP** to the classical complexity class **PSPACE**. Our discussion of this result is rather informal.

BQP is an essentially quantum complexity class consisting of those decision problems that can be solved with bouded probability of error using a polynomial size quantum circuit.

One of the most significant results in quantum computational complexity is that **BQP** $\subseteq$ **PSPACE**. It is clear that **BPP** $\subseteq$ **BQP**, where **BPP** is the classical complexity class of decision problems which can be solved with bounded probability of error using polynomial time on a classical Turing machine. Thus we have the cahin of inclusions **BPP** $\subseteq$ **BQP** $\subseteq$ **PSPACE**. Proving that **BQP** $\neq$ **BPP** – intuitively the statement that quantum computers are more powerful than classical computers – will therefore imply that **BPP** $\neq$ **PSPACE**. However, it is not presently known, and proving this would represent a major breakthrough in classical computer science.

Why is it that **BQP** $\subseteq$ **PSPACE**? Here is an intuitive outline of the proof. Suppose we have an n qubit quantum computer, and do a computation involving a sequence of $p(n)$ gates. Supposing the quantum circuit starts in the state $|0\rangle$ we will explain how to evaluate in polynomial space on a classical computer the probability that it ends up in the state $|y\rangle$. Suppose the gates that are executed on the quantum computer are, in order, $U_1, U_2, \ldots, U_{p(n)}$. Then the probability of ending up in the state $|y\rangle$ is the modulus squared of

$$\langle y|U_{p(n)} \ldots U_2 U_1|0\rangle.$$

This quantity may be estimated in polynomial space on a classical computer. The basic idea is to insert the complete relation $\sum_x |x\rangle\langle x| = I$ between each term, obtaining

$$\langle y|U_{p(n)} \ldots U_2 U_1|0\rangle = \sum_{x_1,\ldots,x_{p(n)-1}} \langle y|U_{p(n)}|x_{p(n)-1}\rangle\langle x_{p(n)-1}|U_{p(n)-1} \ldots U_2|x_1\rangle\langle x_1|U_1|0\rangle.$$

It is clear that each term in the sum can be calculated to high accuracy using only polynomial space on a classical computer, and thus the sum as a whole can be calculated using polynomial space,

4.6 Summary of the quantum circuit model of computation

'quantum computer' $\equiv$ 'quantum circuit model of computation'

The key elements of the quantum circuit mdel of computation:

Classical resources:

A suitable state space:

Ability to prepare states in the computational basis

Ability to perform quantum gates:

Ability to perform measurements in the computaional basis:

Chapter 8

Quantum noise and quantum operations

Until now we have dealt almost solely with dynamics of closed quantum systems, that is, with quantum system that do not suffer any unwanted interactions with the outside world. However, real systems suffer from unwanted interactions with the outside world. These unwanted interactions show up as *noise* in quantum information processing systems. We need to understand and control such noise processes in order to build useful quantum information processing systems.

The third part begins with the description of the *quantum operation formalism*, a powerful set of tools enabling us to describe quantum noise and the behavior of *open* quantum systems.

The quantum operation formalism is very powerful, in that it simultaneously addresses a wide range of physical scenarios. It can be used to describe:

- nearly closed systems which are weakly coupled to their environment.
- systems which strongly coupled to their environment.
- closed systems that are opened suddenly and subject to measurement.

Another advantage of quantum operations in applications to quantum computation and quantum information is that they are especially well adapted to describe *discrete state changes*, that is, transformations between an initial state ρ and final state ρ', without explicit reference to the passage of time. This discrete-time analysis is rather different to the tools traditionally used by physicists for the description of open quantum systems (such as 'master equations', 'Langevin equations', and 'stochastic differential equations'), which tend to be continuous-time descriptions.

8.1 Classical noise and Markov processes

To understand noise in quantum systems it is helpful to build some intuition by understanding noise in classical systems.

Suppose p_0 and p_1 are the initial probabilities that the bit is in the states 0 and 1, respectively. Let q_0 and q_1 be the corresponding probabilities after the noise has occurred. Let X be the initial state of the bit, and Y the final state of the bit. Then the law of total probability states that

$$p(Y = y) = \sum_x p(Y = y|X = x)p(X = x). \tag{8.1}$$

The conditional probabilities $p(Y = y|X = x)$ are called *transition probabilities*, since they summarize the changes that may occur in the system.

$$\begin{bmatrix} q_0 \\ q_1 \end{bmatrix} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \end{bmatrix}. \tag{8.2}$$

Imagine that we are trying to build a classical circuit to perform some computational task : a *Markov Process* $X \to Y \to Z$ (*independent* process *cf.* Markovian an non-Markovian). Physically, this assumption of Marvovicity corresponds to assuming that the environment causing the noise in the first NOT gate, a reasonable assumption given that the gates are likely to be located a considerable distance apart in space.

Noise in classical systems can be described using the theory of stochastic processes. Often, in the analysis of multi-stage processes it is a good assumption to use Markov processes. For a single stage process the output probabilities $\vec{q}$ are related to the input probabilities $\vec{p}$ by the equation

$$\vec{q} = E\vec{p}, \tag{8.3}$$

where E is a matrix of transition probabilities which we shall refer to as the *evolution matrix*. Thus, the final state of the system is linearly related to the starting state. This feature of linearity is echoed in the description of quantum noise, with density matrices replacing probability distributions.

If $\vec{p}$ is a valid probability distribution, then $E\vec{p}$ must also be a valid probability distribution. Thus, E must be satisfied

1. the positivity requirement (non-negative E)
2. completeness requirement (unity sum of E)

8.2 Quantum operations

8.2.1 Overview

The quantum operations formalism is a general tool for describing the evolution of quantum systems in a wide variety of circumstances, including stochastic changes to quantum states, much as Markov processes describe stochastic changes to classical states. Just as a classical state is described by a vector of probabilities, we shall describe quantum states in terms of the density operator ρ. Similar to (8.3), quantum states transform as

$$\rho' = \mathcal{E}(\rho) \tag{8.4}$$

The map $\mathcal{E}$ in this equation is a *quantum operation*.

eg)

- Unitary transformation $\mathcal{E}(\rho) = U\rho U^\dagger$
- Measurement $\mathcal{E}_m(\rho) = M_m \rho M_m^\dagger$

The quantum operation captures the dynamic change to a state which occurs as the result of some physical process; ρ is the initial state before the process, and $\mathcal{E}(\rho)$ is the final state after the process occurs, possibly up to some normalization factor.

We shall develop *three* separate ways of understanding quantum operations.

1. The fist method is based on the idea of studying dynamics as the result of an interaction between a system and an environment, much as classical noise was described in Section 8.1. This method is concrete and easy to relate to the real world. However, it suffers from the drawback of not being mathematically convenient.

2. The second method of understanding quantum operations, completely equivalent to the first, overcomes this mathematical inconvenience by providing a powerful mathematical representation for quantum operations known as *operator-sum representation*. This method is rather abstract, but is very useful for calculations and theoretical work.

3. The third approach to quantum operations, equivalent to the other two, is via a set of physically motivated axioms that we would expect a dynamical map in quantum mechanics to satisfy. The advantage of this approach is that it is exceedingly general. However, it does not offer the calculational convenience of the second approach, nor the concrete nature of the first.

8.2.2 Environments and quantum operations

Consider the unitary transform as a box into which the input state enters and from which the output exits. For our purpose, the interior workings of the box are not of concern to us; it could be implemented by a quantum circuit, or by some Hamiltonian system, or anything else.

A natural way to describe the dynamics of an open system : principal system $\leftrightarrow$ environment (which together form a closed quantum system)

Suppose we have a system in state ρ, which is sent into a box which is coupled to an environment. In general the final state of the system, $\mathcal{E}(\rho)$, may not be related by a unitary transformation to the initial state ρ. We assume that the system–environment input state is a product the environment, and thus we perform a partial trace over the environment to obtain the reduced state of the system alone:

$$\mathcal{E}(\rho) = \mathrm{tr}_{\mathrm{env}}\left[U(\rho \otimes \rho_{\mathrm{env}})U^{\dagger}\right] \tag{8.6}$$

Of course, if U does not involve any interaction with the environment, the $\mathcal{E}(\rho) = \tilde{U}\rho\tilde{U}^{\dagger}$ where $\tilde{U}$ is the part of U which acts on the system alone.

An important assumption is made in this definition – the system and the environment start in a product state. In general, of course, this is not true. Quantum systems interact constantly with their environments, building up correlations.($\rightarrow$ Heat exchange)

However, in many cases of practical interest it is reasonable to assume that the system and its environment start out a product state. Ideally, the correlations will be completely destroyed, leaving the system in a pure state. Even if this is not the case, we shall see later that the quantum operations formalism can even describe quantum dynamics when the system and environment do not start out n a product state.

How can U be specified if the environment has nearly infinite degrees of freedom? It turns out that in order for this model to properly describe any possible transformation $\rho \rightarrow \mathcal{E}(\rho)$, if the principal system has a Hilbert space of d dimensions, then it suffices to model the environment as being in a Hilbert space of no more than d^2 dimensions. It turns out not to be necessary for the environment to start out in a mixed state.

eg) The two qubit quantum circuit in which U is a controlled-NOT gate, with the principal system the control qubit, and the environment initially in the state $\rho_{\mathrm{env}} = |0\rangle\langle 0|$ as the target qubit. Inserting into Equation (8.6), it is easily seen that

$$\mathcal{E}(\rho) = P_0\rho P_0 + P_1\rho P_1, \tag{8.7}$$

where $P_0 = |0\rangle\langle 0|$ and $P_1 = |1\rangle\langle 1|$ are projection operators.Intuitively, this dynamics occurs because the environment stays in the $|0\rangle$ state only when the system is $|0\rangle$; otherwise the environment is flipped to the state $|1\rangle$.

It is convenient to generalize the definition somewhat to allow different input and output spaces.

eg) A single qubit(A) is prepared in an unknown state ρ. A three-level quantum system ('qutrit') labelled B is prepared in some standard state $|0\rangle$, and then interacts with system A via a unitary interaction U, causing the joint system to evolve into the state $U(\rho \otimes |0\rangle\langle 0|)U^{\dagger}$. We then discard system A, leaving system B in some final state ρ'. The quantum operation $\mathcal{E}$ describing this process is

$$\mathcal{E}(\rho) = \rho' = \mathrm{tr}_A(U(\rho \otimes |0\rangle\langle 0|)U^{\dagger}) \tag{8.8}$$

Notice that $\mathcal{E}$ maps density operators of the input system, A, to density operators of the output system, B. Most of our discussion of quantum operations below is concerned with quantum operations 'on' some system A, that is, they map density operators of system A to density operators of system A. The quantum operation $\mathcal{E}$ defining this process simply maps ρ to ρ'.

8.2.3 Operator-sum representation

The operator-sum representation : essentially re-statement of Equation (8.6) explicitly in terms of operators on the principal system's Hilbert space alone.

Let $|e_k\rangle$ be an orthonormal basis for the state space of the environment, and let $\rho_{\text{env}} = |e_0\rangle\langle e_0|$ be the initial state of the environment. There's no loss of generality in assuming that the environment starts in a pure state. (Section 2.5). Equation (8.6) can be rewritten as

$$\begin{aligned} \mathcal{E}(\rho) &= \sum_k \langle e_k|U\left[\rho \otimes |e_0\rangle\langle e_0|\right]U^\dagger|e_k\rangle & (8.9)\\ &= \sum_k E_k \rho E_k^\dagger, & (8.10) \end{aligned}$$

where $E_k \equiv \langle e_k|U|e_0\rangle$ is an operator n the state space of the principal system. Equation (8.10) is known as the operator-sum representation of $\mathcal{E}$. The operators $\{E_k\}$ are known as *operator elements* for the quantum operation $\mathcal{E}$. The operator-sum representation is important.

The operation elements satisfy an important constraint known as the *completeness relation*, analogous to the completeness relation for evolution matrices in the description of classical noise.

Classical case The completeness relation arises from the requirement that probability distributions be normalized to one.

Quantum case The completeness relation arises from the analogous requirement that the trace of $\mathcal{E}(\rho)$ be equal to one

$$\begin{aligned} 1 &= \text{tr}\left(\mathcal{E}(\rho)\right) & (8.11)\\ &= \text{tr}\left(\sum_k E_k \rho E_k^\dagger\right) & (8.12)\\ &= \text{tr}\left(\sum_k E_k^\dagger E_k \rho\right). & (8.13) \end{aligned}$$

Since this relationship is true for all ρ it follows that we must have

$$\sum_k E_k^\dagger E_k = I. \tag{8.14}$$

This equation is satisfied by quantum operations which are *trace-preserving*. There are also non-trace-preserving quantum operations, for which $\sum_k E_k^\dagger E_k \leq I$, but they describe processes in which extra information about what occurred in the process is obtained by measurement.

Maps $\mathcal{E}$ of the form of (8.10) for which $\sum_k E_k^\dagger E_k \leq I$ provide our second definition of a quantum operation. (including cases for non-trace-preserving operations)

The operator-sum representation is important:

- it gives us an intrinsic means of characterizing the dynamics of the principal system. It describes the dynamics of the principal system without having to explicitly consider properties of the environment; all that we need to know is bundled up into the operators E_k, which act on the principal system alone.
- This simplifies calculations and often provides considerable theoretical insight.
- Many different environmental interactions may give rise to the same dynamics on the principal system.

The properties of the operator-sum representation:

1. What's a physical interpretation, in terms of the operation elements E_k.
2. How an operator-sum representation can be determined for any open quantum system (measurement).
3. How to construct a model open quantum system for any operator-sum representation(modelling).

Physical interpretation of the operator-sum representation

Imagine that a measurement of the environment is performed in the basis $|e_k\rangle$ after the unitary transformation U has been applied. Applying the principle of implicit measurement, we see that such a measurement affects only the state of the environment, and does not change the state of the principal system.

Let ρ_k be the state of the principal system given that outcome k occurs, so

$$\begin{aligned} \rho_k \propto \mathrm{tr_E}(|e_k\rangle\langle e_k|U(\rho\otimes|e_0\rangle\langle e_0|)U^\dagger|e_k\rangle\langle e_k|) &= \langle e_k|U(\rho\otimes|e_0\rangle\langle e_0|)U^\dagger|e_k\rangle & (8.18)\\ &= E_k\rho E_k^\dagger. & (8.19)\end{aligned}$$

Normalizing ρ_k,

$$\rho_k \frac{E_k\rho E_k^\dagger}{\mathrm{tr}(E_k\rho E_k^\dagger)} \tag{8.20}$$

we find the probability of outcome k is given by

$$\begin{aligned} p(k) &= \mathrm{tr}(|e_k\rangle\langle e_k|U(\rho\otimes|e_0\rangle\langle e_0|)U^\dagger|e_k\rangle\langle e_k|) & (8.21)\\ &= \mathrm{tr}(E_k\rho E_k^\dagger) & (8.22)\end{aligned}$$

Thus

$$\mathcal{E}(\rho) = \sum_k p(k)\rho_k = \sum_k E_k\rho E_k^\dagger. \tag{8.23}$$

This gives us a beautiful physical *interpretation* of what is going on in a quantum operation with operation elements $\{E_k\}$. The action of the quantum operation is equivalent to taking the state ρ and randomly replacing it by $E_k\rho E_k^\dagger/\mathrm{tr}\left[E_k\rho E_k^\dagger\right]$, with probability $\mathrm{tr}\left[E_k\rho {E_k}^\dagger\right]$. In this sense, it is very similar to the concept of noisy communication channels used in classical information theory; in this vein, we shall sometimes refer to certain quantum operations which describe quantum noise processes as being noisy quantum channels.

eg) Figure 8.4

Suppose we choose the state $|e_k\rangle = |0_E\rangle$ and $|1_E\rangle$. This can be interpreted as doing a measurement in the computational basis of the environment qubit. Doing such a measurement does not change the state of the principal system. The controlled-NOT may be expanded as

$$U = |0_P0_E\rangle\langle 0_P0_E| + |0_P1_E\rangle 0_P\bar{1}_E + |1_P1_E\rangle\langle 1_P0_E| + |1_P0_E\rangle\langle 1_P1_E|. \tag{8.24}$$

$$\begin{aligned} E_0 &= \langle 0_E|U|0_E\rangle = |0_P\rangle\langle 0_P| & (8.25)\\ E_1 &= \langle 1_E|U|0_E\rangle = |1_P\rangle\langle 1_P| & (8.26)\end{aligned}$$

and therefore

$$\mathcal{E}(\rho) = P_0\rho P_0 + P_1\rho P_1, \tag{8.27}$$

Measurement and the operator-sum representation

How to determine an operator-sum representation for an open quantum system → the unitary system-environment transformation operation U and a basis of state $|e_k\rangle$ for the environment, the operation elements are

$$E_k \equiv \langle e_k|U|e_0\rangle. \tag{8.28}$$

It is possible to extend this result even further by allowing the possibility that a measurement is performed on the combined system-environment after the unitary interaction, allowing the acquisition of information about the quantum state. → **non-trace-preserving quantum operations**, that is, maps $\mathcal{E}(\rho)\sum_k E_k\rho {E_k}^\dagger$ such that $\sum_k {E_k}^\dagger E_k \leq I$.

Suppose:
$\rho \rightarrow$ initial state (Q: principal system, E: environment)
The joint state of the system is initially

$$\rho^{QE} = \rho \otimes \sigma. \tag{8.29}$$

After some U (: unitary interaction) a projective measurement (P_m) is performed n the joint system.

The final state of QE is given by

$$\frac{P_m U(\rho\otimes\sigma)U^\dagger P_m}{\mathrm{tr}\left[P_m U(\rho\otimes\sigma)U^\dagger P_m\right]}, \tag{8.30}$$

given that measurement outcome m occurred. Tracing out E we see that the final state of Q alone is

$$\frac{\mathrm{tr_E}\left[P_m U(\rho\otimes\sigma)U^\dagger P_m\right]}{\mathrm{tr}\left[P_m U(\rho\otimes\sigma)U^\dagger P_m\right]}. \tag{8.31}$$

This representation of the final state involves the initial state σ of the environment, the interaction U and the measurement operators P_m. Define a map

$$\mathcal{E}_m(\rho) \equiv \mathrm{tr_E}\left[P_m U(\rho\otimes\sigma)U^\dagger P_m\right], \tag{8.32}$$

so the final state of Q alone is $\mathcal{E}_m(\rho)/\mathrm{tr}\left[\mathcal{E}_m(\rho)\right]$. Note that $\mathrm{tr}\left[\mathcal{E}_m(\rho)\right]$ is the probability of outcome m of the measurement occurring.

Let $\sigma = \sum_j q_j|j\rangle\bar{j}$ be an ensemble decomposition for σ. Introduce an orthonormal basis $|e_k\rangle$ for the system E. Note that

$$\mathcal{E}_m(\rho) = \sum_{jk} q_j \mathrm{tr_E}\left[|e_k\rangle\langle e_k P_m U(\rho\otimes|j\rangle\langle j|)U^\dagger P_m|e_k\rangle\langle e_k||\right] \tag{8.33}$$

$$= \sum_{jk} E_{jk}\rho {E_{jk}}^\dagger, \tag{8.34}$$

where

$$E_{jk} \equiv \sqrt{q_j}\langle e_k|P_m U|j\rangle. \tag{8.35}$$

This equation is a generalization of Equation (8.10), and gives an explicit means for calculating the operators appearing in an operator-sum representation for $\mathcal{E}_m$. The quantum operation $\mathcal{E}_m$ can be thought of as defining a kind of measurement process generalizing the description of measurements (Ch. 2)

System-environment models for any operator-sum representation

Given a set of operators $\{E_k\}$ is there some reasonable *model environmental system and dynamics* which give rise to a quantum operation with those operation element?

For any trace-preserving or non-trace-preserving quantum operation, $\mathcal{E}$, with operation elements $\{E_k\}$, there exists a model environment, E, starting in a pure state $|e_0\rangle$, and model dynamics specified by a unitary operator U and projector P onto E such that

$$\mathcal{E}(\rho) = \mathrm{tr_E}\left[PU(\rho\times|e_0\rangle\langle e_0|)U^\dagger P\right]. \tag{8.36}$$

Suppose first that $\mathcal{E} \rightarrow$ a trace-preserving quantum operation (elements $\{E_k\}$) satisfying the completeness relation $\sum_k {E_k}^\dagger E_k = I$, U?

Let $|e_k\rangle$ be an orthonormal basis set for E, in one-to-one correspondence with the index k for the operators E_k. We are trying to find a *model* environment giving rise to a dynamics described by the operation elements $\{E_k\}$. Define an operator U which has the following action on states of the form $|\psi\rangle|e_0\rangle$,

$$U|\psi\rangle|e_0\rangle \equiv \sum_k E_k|\psi\rangle|e_k\rangle, \tag{8.37}$$

where $|e_0\rangle$ is just some standard state of the model environment. For arbitrary states $|\psi\rangle$ and $|\phi\rangle$ of the principal system,

$$\langle\psi|\langle e_0|U^\dagger U|\phi\rangle|e_0\rangle = \sum_k \langle\psi|{E_k}^\dagger E_k|\phi\rangle = \langle\psi|\phi\rangle \tag{8.38}$$

by the completeness relation. Thus the operator U can be extended to a unitary operator acting on the entire state space of the joint system.

$$\mathrm{tr_E}\left[U(\rho \times |e_0\rangle\langle e_0|)U^\dagger\right] = \sum_k E_k \rho {E_k}^\dagger, \tag{8.39}$$

so this model provides a realization of the quantum operation $\mathcal{E}$ with operation elements $\{E_k\}$.

Non-trace-preserving quantum operation $\rightarrow$ A more interesting generalization of this construction is the case of a set of quantum operations $\{\mathcal{E}_m\}$ corresponding to possible outcomes from a measurement, so the quantum operation $\sum_m \mathcal{E}_m$ is trace-preserving, since the probabilities of the distinct outcomes sum to one, $1 = \sum_m p(m) = \mathrm{tr}\left[(\sum_m \mathcal{E}_m)(\rho)\right]$ for all possible inputs ρ.

Box 8.1: Mocking up a quantum operation

8.2.4 Axiomatic approach to quantum operations

We are going to switch to a different viewpoint now, where we try to write down physically motivated axioms which we expect quantum operation to obey. $\rightarrow$ more abstract.

First, we're going to forget everything we've learned about quantum operations, and start over by defining quantum operations according to a set of axioms, which wel" justify on physical grounds. That done, we'll prove that a map $\mathcal{E}$ satisfies these axioms iff it has an operator-sum representation, thus providing the missing link between the abstract axiomatic formulation, and our earlier discussion.

We *define* a quantum operator $\mathcal{E}$ 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, with the following three axiomatic properties:

- **A1**: First, $\mathrm{tr}\left[\mathcal{E}(\rho)\right]$ is the probability that the process represented by $\mathcal{E}$ occurs, when ρ is the initial state. Thus, $0 \leq \mathrm{tr}\left[\mathcal{E}(\rho)\right] \leq 1$ for any state ρ.
- **A2**: Second, $\mathcal{E}$ is a *convex-linear map* on the set of density matrices, that is, for probabilities $\{p_i\}$,

$$\mathcal{E}\left(\sum_i p_i \rho_i\right) = \sum_i p_i \mathcal{E}(\rho_i) \tag{8.43}$$

- **A3**: Third, $\mathcal{E}$ is a completely positive map. That is, if $\mathcal{E}$ maps density operators of system Q_1 to density operators of system Q_2, then $\mathcal{E}(A)$ must be positive for any positive operator A. Furthermore, if we introduce an extra system R of arbitrary dimensionality, it must be true that $(\mathcal{I} \otimes \mathcal{E})(A)$ is positive for any positive operator A on the combined system RQ_1, where $\mathcal{I}$ denotes the identity map on system R.

It is convenient to make the convention that $\mathcal{E}$ does not necessarily preserve the trace property of density matrices, that $\mathrm{tr}\,[\rho] = 1$. Rather, we make the convention that $\mathcal{E}$ is to be defined in such a way that $\mathrm{tr}\,[\mathcal{E}(\rho)]$ is equal to the probability of the measurement outcome described by $\mathcal{E}$ occurring.

eg) Suppose that we are doing a projective measurement in the computational basis of a single qubit. Then two quantum operations are used to describe this process, defined by $\mathcal{E}_0(\rho) \equiv |0\rangle\langle 0|\rho|0\rangle\langle 0|$ and $\mathcal{E}_1(\rho) \equiv |1\rangle\langle 1|\rho|1\rangle\langle 1|$. Notice that the probabilities of the respective outcomes are correctly given by $\mathrm{tr}\,[\mathcal{E}_0(\rho)]$ and $\mathrm{tr}\,[\mathcal{E}_1(\rho)]$. With this convention the correctly normalized final quantum state is therefore

$$\frac{\mathcal{E}(\rho)}{\mathrm{tr}\,[\mathcal{E}(\rho)]} \tag{8.44}$$

The first property is one of mathematical convenience. In the case where the process is deterministic, that is, no measurement is taking place, this reduces to the requirement that $\mathrm{tr}\,[\mathcal{E}(\rho)] = 1 = \mathrm{tr}(\rho)$, for all ρ. ($\rightarrow$ non-trace-preserving) A physical quantum operation is one that satisfies the requirement that probabilities never exceed 1, $\mathrm{tr}\,[\mathcal{E}(\rho)] \leq 1$.

The second property stems from a physical requirement on quantum operations.

$$\rho = \sum_i p_i \rho_i$$

Then we would expect that the resulting state,

$$\frac{\mathcal{E}(\rho)}{\mathrm{tr}\,[\mathcal{E}(\rho)]} = \frac{\mathcal{E}(\rho)}{p(\mathcal{E})}$$

corresponds to a random select from the ensemble $\{p(i|\mathcal{E}), \mathcal{E}(\rho_i)/\mathrm{tr}\,[\mathcal{E}(\rho_i)]\}$, where $p(i|\mathcal{E})$ is the probability that the state prepared was ρ_i, given that the process represented by $\mathcal{E}$ occurred. Thus, we demand that

$$\mathcal{E}(\rho) = p(\mathcal{E}) \sum_i p(i|\mathcal{E}) \frac{\mathcal{E}(\rho_i)}{\mathrm{tr}\,[\mathcal{E}(\rho_i)]}, \tag{8.45}$$

where $p(\mathcal{E}) = \mathrm{tr}\,[\mathcal{E}(\rho)]$ is the probability that the process described by $\mathcal{E}$ occurs on input of ρ. By Bayes' rule

$$(i|\mathcal{E}) = p(\mathcal{E}|i) \frac{p_i}{p(\mathcal{E})} = \frac{\mathrm{tr}\,[\mathcal{E}(\rho_i)]\, p_i}{p(\mathcal{E})} \tag{8.46}$$

so (8.45) reduces (8.43).

The third property also originates from an important physical requirement, that not only must $\mathcal{E}(\rho)$ be a valid density matrix (up to normalization) so long as ρ is valid, but furthermore, if $\rho = \rho_{RQ}$ is the density matrix of some joint system of R and Q, if $\mathcal{E}$ acts only on Q, then $\mathcal{E}(\rho_{RQ})$ must still result in a valid density matrix of the joint system. Formally, suppose we introduce a second system R. Let $\mathcal{I}$ denote the identity map n system R. Then the map $\mathcal{I} \otimes \mathcal{E}$ must take positive operators to positive operators.

Box 8.2: Complete positivity versus positivity

The transpose operation on a single qubit $\rightarrow$ preserves positivity of a single qubit.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \xrightarrow{T} \begin{bmatrix} a & c \\ b & d \end{bmatrix}. \tag{8.47}$$

This map preserves positivity of a single qubit. However, suppose that qubit is part of a two qubit system initially in the entangled state

$$\frac{|00\rangle + |11\rangle}{\sqrt{2}} \tag{8.48}$$

and the transpose operation is applied to the first of these two qubits. Then the density operator of the system after the dynamics has been applied is

$$\frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{8.49}$$

This operator has eigenvalues $1/2$, $1/2$, $1/2$, and $-1/2$, so this is not a valid density operator. Thus the transpose operation is an example of a positive map which is not completely positive, that is, it preserves the positivity of operators on the principal system, but does not continue to preserve positivity when applied to systems which contain the principal system as a subsystem.

Theorem 8.1. *The map $\mathcal{E}$ satisfies axioms **A1, A2,** and **A3** iff*

$$\mathcal{E}(\rho) = \sum_i E_i \rho {E_i}^\dagger, \tag{8.50}$$

for some set of operators $\{E_i\}$ which map the input Hilbert space to the output Hilbert space, and $\sum_i {E_i}^\dagger E_i \leq I$.

Proof
Suppose $\mathcal{E}(\rho) = \sum_i E_i \rho {E_i}^\dagger$.
$\mathcal{E} \rightarrow$ Linear, so to check that $\mathcal{E}$ is a quantum operation we need only prove that it is completely positive.

$A \rightarrow$ a positive operator acting on RQ.
$|\psi\rangle \rightarrow$ some state of RQ.
def. $|\phi_i\rangle \equiv (I_R \otimes {E_i}^\dagger)|\psi\rangle$, we have

$$\begin{aligned} \langle\psi|(I_R \otimes E_i)A(I_R \otimes {E_i}^\dagger)|\psi\rangle &= \langle\phi_i|A|\phi_i\rangle & (8.51)\\ &\geq 0, & (8.52)\end{aligned}$$

by the positivity of the operator A. It follows that

$$\langle\psi|(\mathcal{I} \otimes \mathcal{E})(A)|\psi\rangle = \sum_i \langle\phi_i|A|\phi_i\rangle \geq 0, \tag{8.53}$$

and thus for any positive operator A, the operator $(\mathcal{I} \otimes \mathcal{E})(A)$ is also positive. The requirement $\sum_i {E_i}^\dagger E_i \leq I$ ensures that probabilities are less than or equal to 1.

Suppose next that $\mathcal{E}$ satisfies axioms **A1**, **A2**, and **A3**. Our aim will be to find an operator-sum representation for $\mathcal{E}$.

R : the same dimension as the original quantum system Q.
$|i_R\rangle$, $|i_Q\rangle$: orthonormal bases for R and Q.

Define a joint state $|\alpha\rangle$ of RQ by

$$|\alpha\rangle \equiv= \sum_i |i_R\rangle|i_Q\rangle. \tag{8.54}$$

The state $|\alpha\rangle$ is a maximally entangled state of the systems R and Q. Next, we define an operator σ on the state space of RQ by

$$\sigma \equiv (\mathcal{I}_R \otimes \mathcal{E})(|\alpha\rangle\langle\alpha|). \tag{8.55}$$

We may think of this as the result of applying the quantum operation $\mathcal{E}$ to one half of a maximally entangled state of the system RQ. $\rightarrow$ the operator σ completely specifies the quantum operator $\mathcal{E}$. That is, to know how $\mathcal{E}$ acts on an arbitrary state of Q, it is sufficient to know how it acts on a single maximally entangled state of Q with another system.

Let $|\psi\rangle = \sum_j c_j |j_Q\rangle$ be any state of Q. Define a corresponding state $|\tilde{\psi}\rangle$ of system R by the equation

$$|\tilde{\psi}\rangle = \sum_j c_j^* |j_R\rangle. \tag{8.56}$$

Notice that

$$\begin{aligned}
\langle\tilde{\psi}|\sigma|\tilde{\psi}\rangle &= \langle\tilde{\psi}| \left(\sum_{ij} |i_R\rangle\langle i_R| \otimes \mathcal{E}(|i_Q\rangle\langle i_Q|) \right) |\tilde{\psi}\rangle & (8.57)\\
&= \sum_{ij} c_i c_j^* \mathcal{E}(|i_Q\rangle\langle i_Q|) & (8.58)\\
&= \mathcal{E}(|\psi\rangle\langle\psi|). & (8.59)
\end{aligned}$$

Let $\sigma = \sum_i |s_i\rangle\langle s_i|$ be some decomposition of σ, where the vector $|s_i\rangle$ need not be normalized. Define a map

$$E_i(|\psi\rangle) \equiv \langle\tilde{\psi}|s_i\rangle. \tag{8.60}$$

This map is a linear map, so E_i is a linear operator on the state space Q. Furthermore we have

$$\begin{aligned}
\sum_i E_i|\psi\rangle\langle\psi|{E_i}^\dagger &= \sum_i \langle\tilde{\psi}|s_i\rangle\langle s_i|\tilde{\psi}\rangle & (8.61)\\
&= \langle\tilde{\psi}|\sigma|\tilde{\psi}\rangle & (8.62)\\
&= \mathcal{E}(|\psi\rangle\langle\psi|). & (8.63)
\end{aligned}$$

Thus

$$\mathcal{E}(|\psi\rangle\langle\psi|) = \sum_i E_i|\psi\rangle\langle\psi|{E_i}^\dagger \tag{8.64}$$

for all pure states, $|\psi\rangle$, of Q. By convex-linearity it follows that

$$\mathcal{E}\rho = \sum_i E_i \rho {E_i}^\dagger \tag{8.65}$$

in general The condition $\sum_i {E_i}^\dagger E_i \leq I$ follows immediately from axiom **A1** identifying the trace of $\mathcal{E}(\rho)$ with a probability. □

Freedom in the operator-sum representation

We have seen that the operator-sum representation provides a very general description of the dynamics of an open quantum system. Is it a *unique* description?

Consider quantum operations $\mathcal{E}$ and $\mathcal{F}$ acting on a single qubit with the operator-sum representations $\mathcal{E}(\rho) = \sum_k E_k \rho {E_k}^\dagger$ and $\mathcal{F}(\rho) = \sum F_k \rho {F_k}^\dagger$, where the operation elements for $\mathcal{E}$ and $\mathcal{F}$ are defined by

$$E_1 = \frac{I}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad E_2 = \frac{Z}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{8.66}$$

and

$$F_1 = |0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad F_2 = |1\rangle\langle 1| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{8.67}$$

These appear to be very different quantum operations. What is interesting is that $\mathcal{E}$ and $\mathcal{F}$ are *actually the same quantum operation*. ($\because F_1 = (E_1 + E_2)/\sqrt{2}$, $F_2 = (E_1 - E_2)/\sqrt{2}$) Thus,

$$\mathcal{F}(\rho) = \mathcal{E}(\rho) \tag{8.70}$$

The operation elements appearing in an operator-sum representation for a quantum operation are *not* unique. A random applying either I or Z has the same dynamics of a projective measurement in the $\{|0\rangle, |1\rangle\}$ basis. → Two very different physical process give rise to exactly the same dynamics for the principal system.

Understanding the freedom in operator-sum representation:

- gives us more insight into how different physical processes can give rise to the same system dynamics
- is crucial to a good understanding of quantum error-correction.

From (8.10),

$$\begin{aligned} \mathcal{E}(\rho) &= \sum_k \langle e_k|U\left[\rho \otimes |e_0\rangle\langle e_0|\right] U^\dagger |e_k\rangle \\ &= \sum_k E_k \rho E_k^\dagger, \end{aligned}$$

where $E_k = \langle e_k|U|e_0\rangle$. Suppose that we supplement the interaction U with an additional unitary action U' on the environment alone. Clearly this does not change the state of the principal system. [even they are entangled] We obtain:

$$F_k = \langle e_k|(I \otimes U')|e_0\rangle \tag{8.71}$$

$$= \sum_j \left[I \otimes \langle e_k|U'|e_j\rangle\right] \langle e_j|U|e_0\rangle \tag{8.72}$$

$$= \sum_j U'_{kj} E_j, \tag{8.73}$$

U'_{kj} are the matrix elements of U' with respect to the basis $|e_k\rangle$

Theorem 8.2 (Unitary freedom in the operator-sum representation). *Suppose $\{E_1, \ldots, E_m\}$ and $\{F_1, \ldots, E_n\}$ are operation elements giving rise to quantum operations $\mathcal{E}$ and $\mathcal{F}$, respectively. By appending zero operators to the shorter list of operation elements we may ensure that $m = n$. The $\mathcal{E} = \mathcal{F}$ iff there exist complex numbers u_{ij} such that $E_i = \sum_j u_{ij} F_j$, and u_{ij} is an m by m unitary matrix.*

proof
Recall the theorem 2.6 *Unitary freedom in the ensemble for density matrices.* Two sets of vectors $|\psi_i\rangle$ and $|\phi_i\rangle$ generate the same operator iff

$$|\psi_i\rangle = \sum_j u_{ij} |\phi_i\rangle, \tag{8.74}$$

where u_{ij} is a unitary matrix of complex numbers, and $|\psi_i\rangle$ and $|\phi_i\rangle$ have the same number of elements. Suppose $\{E_i\}$ and $\{F_i\}$ are two sets of operation elements for the same quantum operation,$\sum_i E_i \rho {E_i}^\dagger = \sum_j F_j \rho {F_j}^\dagger$ for all ρ. Define

$$|e_i\rangle \equiv \sum_k |k_R\rangle (E_i|k_Q\rangle) \tag{8.75}$$

$$|f_j\rangle \equiv \sum_k |k_R\rangle (F_j|k_Q\rangle) \tag{8.76}$$

Recall $\sigma \equiv (\mathcal{I}_R \otimes \mathcal{E})(|\alpha\rangle\langle\alpha|)$, from which it follows that $\sigma = \sum_i |e_i\rangle\langle e_i| = \sum_j |f_j\rangle\langle f_j|$, and thus there exits unitary u_{ij} such that

$$|e_i\rangle = \sum_j u_{ij}|f_j\rangle. \tag{8.77}$$

But for arbitrary $|\psi\rangle$ we have

$$\begin{aligned} E_i|\psi\rangle &= \langle\tilde{\psi}|e_i\rangle & (8.78)\\ &= \sum_j u_{ij}\langle\tilde{\psi}|f_j\rangle & (8.79)\\ &= \sum_j u_{ij}F_j|\psi\rangle & (8.80) \end{aligned}$$

Thus

$$E_i = \sum_j u_{ij}F_j. \tag{8.81}$$

Conversely, supposing E_i and F_j are related by a unitary transformation of the form $E_i = \sum_j u_{ij}F_j \rightarrow$ the quantum operation with operation elements $\{E_i\}$ is the same as the quantum operation with operation elements $\{F_i\}$. □

Theorem 8.3. *All quantum operation $\mathcal{E}$ on a system of Hilbert space dimension d can be generated by an operator-sum representation containing at most d^2 elements,*

$$\mathcal{E}(\rho) = \sum_{k=1}^{M} E_k \rho {E_k}^\dagger, \tag{8.82}$$

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

8.3 Examples of quantum noise and quantum operations

8.3.1 Trace and partial trace

One of the main uses of the quantum operations formalism is to describe the effects of measurement.

Quantum operations formalism can be used to describe both the *probability of getting a particular outcome from a measurement on a quantum system*, and also *the state change n the system effected by the measurement*.

The trace map $\rho \rightarrow \mathrm{tr}\,[\rho]$– the simplest operation related to measurement.

H_Q : any input Hilbert space, spanned by an orthonormal basis $|1\rangle \ldots |d\rangle$
H'_Q : any dimensional ouput space, spanned by the state $|0\rangle$

$$\mathcal{E}(\rho) \equiv \sum_{i=1}^{d} |0\rangle\langle i|\rho|i\rangle\langle 0|, \tag{8.83}$$

so that $\mathcal{E}$ is a quantum operation, by Theorem 8.1. ($E_i = |0\rangle\langle i|$). The quantum operation is identical to the trace function ($|0\rangle$: unimportant)

The partial trace is a quantum operation.

QR: a joint system, and we wish to trace out system R.
$|j\rangle$: a basis for system R.

Define a linear operator $E_i : H_{QR} \rightarrow H_Q$ by

$$E_i \left(\sum_j \lambda_j |q_j\rangle |j\rangle \right) \equiv \lambda_i |q_i\rangle, \tag{8.84}$$

where λ_j are complex numbers, and $|q_j\rangle$ are arbitrary states of system Q. Define $\mathcal{E}$ to be the quantum operation with operation elements $\{E_i\}$, that is,

$$\mathcal{E}(\rho) \equiv \sum_i E_i \rho {E_i}^\dagger. \tag{8.85}$$

Bye Theorem 8.1, this is a quantum operation from system QR to system Q. Notice that

$$\mathcal{E}(\rho \otimes |j\rangle\langle j'|) = \rho \delta_{j,j'} = \mathrm{tr}\left[\rho \otimes |j\rangle\langle j'|\right] \tag{8.86}$$

where ρ is any Hermitian operator on the state space of system Q, and $|j\rangle$ and $|j'\rangle$ are members of the orthonormal basis for system R. By linearity of $\mathcal{E}$ and $\mathrm{tr_R}$, it follows that $\mathcal{E} = \mathrm{tr_R}$.

8.3.2 Geometric picture of single qubit quantum operations

The state of a single qubit can always be written in the Bloch epresentation,

$$\rho = \frac{I + \vec{r} \cdot \vec{\sigma}}{2}, \tag{8.87}$$

where $\vec{r}$ is a three component real vector. Explicitly, this gives us

$$\rho = \frac{1}{2} \begin{bmatrix} 1 + r_z & r_x - i r_y \\ r_x + i r_y & 1 - r_z \end{bmatrix}. \tag{8.88}$$

An arbitrary trace-preserving quantum operation is equivalent to a map of the form

$$\vec{r} \xrightarrow{\mathcal{E}} \vec{r}' = M\vec{r} + \vec{c}, \tag{8.89}$$

where M is a 3×3 real matrix, and $\vec{c}$ is a constant vector. This is an *affine map*, mapping the Bloch sphere into itself.

Affine map

Any transformation preserving collinearity and ratios of distances. A affine map may also be thought of as a shearing transformation

$$\begin{aligned} F &: \mathbb{R}^n \rightarrow \mathbb{R}^n \\ F(\vec{p}) &= A\vec{p} + \vec{q} \end{aligned}$$

To see this, suppose the operators E_i generating the operator-sum representation for $\mathcal{E}$ are written in the form

$$E_i = \alpha_i I + \sum_{k=1}^{3} a_{ik} \sigma_k \tag{8.90}$$

Then

$$M_{jk} = \sum_l \left[a_{lj} a_{lk}^* + a_{lj} * a_{lk} + \left(|\alpha_l|^2 - \sum_p a_{lp} a_{lp}^* \right) \delta_{jk} + i \sum_p \epsilon_{jkp} (\alpha_l a_{lp}^* - \alpha_l^* a_{lp}) \right] \tag{8.91}$$

$$c_k = 2i \sum_l \sum_{jp} \epsilon_{jpk} a_{lj} a_{lp}^*, \tag{8.92}$$

The meaning of the affine map is made clearer by considering the polar decomposition of the matrix M, $M = U|M|$, where U is unitary.
$M \to$ real
$|M| \to$ real and Hermitian (symmetric)
$U \to$ real (orthogonal), $U^T U = I$
Thus we may write

$$M = OS, \tag{8.93}$$

where O is a real orthogonal matrix with determinant 1, representing a proper rotation, and S is a real symmetric matrix. Equation(8.89) is just a deformation of the Bloch sphere along principal axes determined by S, followed by a proper rotation due to O, followed by a displacement due to $\vec{c}$.

8.3.3 Bit flip and phase flip channels

The *bit flip* channel flips the state of a qubit from $|0\rangle$ to $|1\rangle$ (and vice versa) with probability $1-p$. It has operation elements

$$E_0 = \sqrt{p} I = \sqrt{p} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad E_1 = \sqrt{1-p} X = \sqrt{1-p} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{8.94}$$

$$\mathcal{E}(\rho) = \frac{1}{2}\left[I + r_x X + (2p-1) r_y Y + (2p-1) r_z Z\right]$$

The contraction of the Bloch sphere illustrated in figure 8.8 cannot increase the norm of the Bloch vector, and therefore we can immediately conclude that $\mathrm{tr}\left[\rho^2\right]$ can only ever decrease for the bit flip channel.

The *phase flip* channel has operation elements

$$E_0 = \sqrt{p} I = \sqrt{p} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad E_1 = \sqrt{1-p} Z = \sqrt{1-p} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{8.95}$$

$$\mathcal{E}(\rho) = \frac{1}{2}\left[I + (2p-1) r_x X + (2p-1) r_y Y + r_z Z\right]$$

If $p = \frac{1}{2}$?

Using the freedom in the operator-sum representation this operation may be written [*cf.* (8.66), (8.67)]

$$\rho \to \mathcal{E}(\rho) = P_0 \rho P_0 + P_1 \rho P_1, \tag{8.96}$$

where $P_0 = |0\rangle\langle 0|$, $P_1 = |1\rangle\langle 1|$, which corresponds to a measurement of the qubit in the $|0\rangle$, $|1\rangle$ basis, with the result of the measurement unknown. Using the above presentation it is easy to see that the corresponding map on the Bloch sphere is given by

$$(r_x, r_y, r_z) \to (0, 0, r_z) \tag{8.97}$$

The *bit-phase flip* channel has operation elements

$$E_0 = \sqrt{p} I = \sqrt{p} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad E_1 = \sqrt{1-p} Y = \sqrt{1-p} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{8.98}$$

$$\mathcal{E}(\rho) = \frac{1}{2}\left[I + (2p-1) r_x X + r_y Y + (2p-1) r_z Z\right]$$

8.3.4 Depolarizing channel

The *depolarizing channel* is an important type of quantum noise. Imagine we take a single qubit, and with probability p that qubit is *depolarized*. That is, it is replaced by the completely mixed state, $I/2$. With probability $1-p$ the qubit is left untouched. The state of the quantum system after this noise is

$$\mathcal{E}(\rho) = \frac{pI}{2} + (1-p)\rho \tag{8.99}$$

$$\mathcal{E}(\rho) = \frac{1}{2}I + \frac{1}{2}(1-p)\vec{r}\cdot\vec{\sigma}$$

Figure (8.12) : The top line of the circuit is the input to the depolarizing channel, while the bottom two lines are an 'environment' to simulate the channel.

The idea is that the third qubit, initially a mixture of the state $|0\rangle$ with probability $1-p$ and state $|1\rangle$ with probability p acts as a control for whether or not the completely mixed state $I/2$ stored in the second qubit is swapped into the first qubit.

The form (8.99) is not in the operator-sum representation. For arbitrary ρ,

$$\frac{I}{2} = \frac{\rho + X\rho X + Y\rho Y + Z\rho Z}{4} \tag{8.100}$$

and then substitute for $I/2$ into (8.99) we arrive at the equation

$$\mathcal{E}(\rho) = \left(1 - \frac{3p}{4}\right)\rho + \frac{p}{4}\left(X\rho X + Y\rho Y + Z\rho Z\right), \tag{8.101}$$

showing that the depolarizing channel has operation elements $\{\sqrt{1-3p/4}I, \sqrt{p}X/2, \sqrt{p}Y/2, \sqrt{p}Z/2\}$. Note, incidentally, that it is frequently convenient to parameterize the depolarizing channel in different ways, such as

$$\mathcal{E}(\rho) = (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Z). \tag{8.102}$$

8.3.5 Amplitude damping

An important application of quantum operations is the description of *energy dissipation* – effects due to loss of energy from a quantum system.

- The dynamics of an atom which is spontaneously emitting a photon.
- A spin system at high temperature approach equilibrium with its environment.
- The state of a photon in an interferometer or cavity when it is subject to scattering and attenuation.

$\rightarrow$ the general behavior of all of them is well characterized by a quantum operation known as *amplitude damping*.

We have a single optical mode containing the quantum state $a|0\rangle + b|1\rangle$. The scattering of a photon from this mode can be modeled by thinking of inserting a partially silvered mirror, a beamsplitter, in the path of the photon. This beamsplitter allows the photon to couple to another single optical mode, according to the unitary transformation $B = \exp[\theta(a^\dagger b - ab^\dagger)]$. The output after the beamsplitter, assuming the environment starts out with no photons, is simply $B|0\rangle(a|0\rangle + b|1\rangle) = a|00\rangle + b(\cos\theta|01\rangle + \sin\theta|10\rangle)$. Tracing over the environment gives us the quantum operation

$$\mathcal{E}_{AD}(\rho) = E_0\rho E_0 + E_1\rho E_1, \tag{8.103}$$

where $E_k = \langle k|B|0\rangle$ are

$$\begin{aligned} E_0 &= \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix} \\ E_1 &= \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}, \end{aligned} \tag{8.104}$$

The operation elements for amplitude damping. $\gamma = \sin^2\theta$ can be thought of as the probability of losing a photon.

The E_1 operation changes a $|1\rangle$ stat into a $|0\rangle$ state, corresponding to the physical process of losing a quantum of energy to the environment. E_0 leaves $|0\rangle$ unchanged, but reduces the amplitude of a $|1\rangle$ state; physically, this happens because a quantum of energy was not lost to the environment, and thus the environment now perceives it to be more likely that the system is in the $|0\rangle$, rather than the $|1\rangle$ state.

A general characteristic of a quantum operation is the set of state that are left invariant under the operation. (*eg.* the phase flip channel : $\hat{z}$ axis invariance) In the case of amplitude damping, only the ground state $|0\rangle$ is left invariant. That is a natural consequence of our modeling the environment as starting in the $|0\rangle$ state, as if it were at zero temperature.

What quantum operation describe the effect of dissipation to an environment at finite temperature? $\mathcal{E}_{GAD}$: *generalized amplitude damping*, is defined for single qubits by the operation elements

$$E_0 = \sqrt{p}\begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix} \tag{8.116}$$

$$E_1 = \sqrt{p}\begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix} \tag{8.117}$$

$$E_2 = \sqrt{1-p}\begin{bmatrix} \sqrt{1-\gamma} & 0 \\ 0 & 1 \end{bmatrix} \tag{8.118}$$

$$E_3 = \sqrt{1-p}\begin{bmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{bmatrix} \tag{8.119}$$

where the stationary state

$$\rho_\infty = \begin{bmatrix} p & 0 \\ 0 & 1-p \end{bmatrix} \tag{8.120}$$

satisfies $\mathcal{E}_{GAD} = \rho_\infty$. Generalized amplitude damping describes the 'T_1' relaxation processes due to coupling of spins to their surrounding lattice, a large system which is in thermal equilibrium at a temperature often much higher than the spin temperature.

The effect of amplitude damping in the Bloch representation as the Bloch vector transformation

$$(r_x, r_y, r_z) \to \left(r_x\sqrt{1-\gamma}, r_y\sqrt{1-\gamma}, \gamma + r_z(1-\gamma)\right). \tag{8.122}$$

When γ is replaced with a time-varying function like $1 - e^{-t/T_1}$, as is often the case for real physical processes.

Similarly, generalized amplitude damping performs the trasformation

$$(r_x, r_y, r_z) \to \left(r_x\sqrt{1-\gamma}, r_y\sqrt{1-\gamma}, \gamma(2p-1) + r_z(1-\gamma)\right). \tag{8.123}$$

Amplitude damping and generalized amplitude damping differ only in the location of the fixed point of the flow; the final state is along the $\hat{z}$ axis, at the point $(2p - 1$, which is a mixed state.

8.3.6 Phase damping

A noise process that is uniquely quantum mechanical, which is describes the loss of quantum information without loss of energy, is *phase damping*.

- A photon scatters randomly as it travels through a waveguide
- Electronic states in an atom are perturbed upon interacting with distant electrical charges.

The energy eigenstates of a quantum system do not change as a function of time, but do accumulate a phase which is not precisely known, partial information about this quantum phase – the *relative* phases between the energy eigenstates - is lost.

A very simple model:
Suppose we have a qubit $|\psi\rangle = a|0\rangle + b|1\rangle$ upon which the rotation operation $R_z(\theta)$ is applied, where the angle of rotation θ is random. (a *phase kick*) Let us assume that the phase kick angle θ is well represented as a random variable which has a Gaussian distribution with mean 0 and variance 2λ.

The output state from this process is given by the density matrix obtained from averaging over θ,

$$\rho = \frac{1}{\sqrt{4\pi\lambda}} \int_{-\infty}^{\infty} R_z(\theta)|\psi\rangle\langle\psi|{R_z}^\dagger(\theta) e^{-\theta^2/4\lambda} d\theta \tag{8.124}$$

$$= \begin{bmatrix} |a|^2 & ab^* e^{-\lambda} \\ a^* b e^{-\lambda} & |b|^2 \end{bmatrix}. \tag{8.125}$$

The random phase kicking causes the expected value of the off-diagonal elements of the density matrix to decay exponentially to zero with time. That is a characteristic result of phase damping.

Consider an interaction between two harmonic oscillators with interaction Hamiltonian

$$H = \chi a^\dagger a (b + b^\dagger). \tag{8.126}$$

Letting $U = \exp(-iH\Delta t)$, considering only the $|0\rangle$ and $|1\rangle$ states of the a oscillator as our system, and taking the environment oscillator to initially be $|0\rangle$, we find that tracing over the environment gives the operation elements $E_k = \langle k_b|U|0_b\rangle$, which are

$$E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\lambda} \end{bmatrix} \tag{8.127}$$

$$E_1 = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{\lambda} \end{bmatrix} \tag{8.128}$$

where $\lambda = 1 - \cos^2(\chi\Delta)$ can be interpreted as the probability that a photon from the system has been scattered (without loss of energy). As was the case for amplitude damping, E_0 leaves $|0\rangle$ unchanged, but reduces the amplitude of a $|1\rangle$ state; unlike amplitude damping, however, the E_1 operation destroys $|0\rangle$ and reduces the amplitude of the $|1\rangle$ state, and does not change it into a $|0\rangle$.

By applying the unitary freedom of quantum operations,

$$\tilde{E_0} = \sqrt{\alpha} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{8.129}$$

$$\tilde{E_1} = \sqrt{1-\alpha} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{8.130}$$

where $\alpha = (1 + \sqrt{1-\lambda})/2$. Thus the phase damping quantum operation is ***exactly*** the same as the phase flip channel.

Phase damping is often referred to as a 'T_2' (or 'spin-spin') relaxation process, where $e^{-t/2T_2} = \sqrt{1-\lambda}$.

Historically, phase damping was a process that was almost always thought of as resulting from a random phase kick or scattering process. It was not until the connection tot the phase flip channel was discovered that quantum error-correction was developed, since it was thought that phase errors were continuous and couldn't be described as ad discrete process.

8.4 Applications of quantum operations

The quantum operations formalism has numerous applications

8.4.1 Master equations

Open quantum systems occur in a wide range of disciplines, and many tools other than other than quantum operations can be employed in their study.

The dynamics of open quantum systems have been studied extensively in the field of quantum optics. The main objective in this context is to describe the time evolution of an open system with a differential equation which properly describes non-unitary behavior.

This description is provided by the master equation, which can be written most generally in the *Lindblad form* as

$$\frac{d\rho}{dt} = -\frac{i}{\hbar}\left[H, \rho\right] + \sum_j \left[2L_j \rho {L_j}^\dagger - \{{L_j}^\dagger L_j, \rho\}\right], \tag{8.134}$$

where $\{x, y\} = xy + yx$ denotes an anticommutator, H is the system Hamiltonian, a Hermitian operator representing the coherent part of the dynamics, and L_j are the *Lindblad operators*, representing the coupling of the system to its environment. $\rightarrow$

- The differential equation takes on the above from in order that the process be completely positive.
- It is also generally assumed that the system and environment begin in a product state.
- To derive a master equation for a process, one usually begins with a system-environment model Hamiltonian, and them makes the Born and Markov approximation in order to determine L_j.
- In the master equation approach, $\mathrm{tr}\left[\rho(t)\right] = 1$ at all times.

8.4.2 Quantum process tomography

How do quantum operations relate to experimentally measurable quantities?
What measurements should an experimentalist do if they wish to characterize the dynamics of a quantum system?
Classical system $\rightarrow$ system identification
Quantum open system $\rightarrow$ quantum process tomography (can be performed for finite dimensional quantum systems)

Quantum state tomography : the procedure of experimentally determining an unknown quantum state.

A single copy of ρ : it is impossible to characterize ρ (no cloning theorem; indistinguishability)
A large number of copies of ρ : it is possible to estimate ρ

Quantum process tomography : The experimental procedure which can determine the ensemble of output state (with d-dimension) using d^2-dimensional pure input states. $\rightarrow$ experimental procedure for understanding and controlling noisy quantum system.

Schrödinger's cat

When I hear about Schrödinger's cat, I reach for my gun
Schrödinger's infamous cat faces life or death contingent upon an automatic device which breaks a vial of poison and kills the cat if an excited atomic state is observed to decay: Schrödinger asked what happens when the atom is in a superposition state? **Is the cat alive or dead?** Why do superposition states such as this apparently not occur in the everyday world?

$$|\psi\rangle = [|\text{dead}\rangle|0\rangle + n|\text{alive}\rangle|1\rangle] / \sqrt{2},$$

in which the cat's state has become entangled with that of the atom. The density matrix of this state,

$$\rho = |\psi\rangle\langle\psi|$$

Now, in practice it is impossible to perfectly isolate the cat and the atom in their box, and thus information about this superposition state will leak into the external world.
eg) heat from the cat's body could permeate the wall and give some indication of its state to the outside. Such effects may be modeled as phase damping, which exponentially damps out the final two (off-diagonal) terms in ρ. To a first approximation, we may model the cat-atom system as a simple harmonic oscillator. An important result about the decoherence of such a system is that coherence between states of high energy difference decays faster than between states with a lower energy difference. Thus ρ will quickly be transformed into a nearly diagonal state, which represents an ensemble of cat-atom states which correspond to eigher dead or alive, and are not in a superposition of the two states.

8.4 Limitations of quantum operations formalism

Are there interesting quantum systems whose dynamics are not described by quantum operations?

Suppose a single qubit is prepared in some unknown quantum state, which we denote ρ. The preparation of this qubit involves certain procedure to be carried out in the laboratory in which the qubit is prepared.

Suppose

$$\rho \otimes |0\rangle\langle 0| \otimes \text{other degrees of freedom} \tag{8.186}$$

if ρ is a state on the bottom half of the Bloch sphere, and

$$\rho \otimes |1\rangle\langle 1| \otimes \text{other degrees of freedom} \tag{8.187}$$

if ρ is a state on the top half of the Bloch sphere.

Once the state preparation is done, the system begins to interact with the environment, in this case all the laboratory degrees of freedom. Suppose the interaction is such that a c-not is performed between the principal system and the extra qubit in the laboratory system. Thus, if the system's Bloch vector was initially in the bottom half of the Bloch sphere it is left invariant by the process, while if it was initially in the top half of the Bloch sphere it is rotated into the bottom half of the Bloch sphere.

A quantum system which interacts with the degrees of freedom used to prepare that system after the preparation is complete will in general suffer a dynamics which is not adequately described within the quantum operations formalism.

It is an interesting problem for further research to study quantum information processing beyond the quantum operations formalism.

Chapter 9

Distance measures for quantum information

What does it mean to say that two items of information are similar?
What does it mean to say that information is preserved by some process?

- static measure : concerned with two broad classes of distance measure
- dynamic measure : quantify how well information has been preserved during a dynamic process

→ trace distance & fidelity

9.1 Distance measures for classical information

The *Hamming distance* : a way of quantifying the distance between classical codes. Unfortunately, the Humming distance between two objects is simply a matter of labeling, and *a priori* there aren't any labels in the Hilbert space arena of quantum mechanics.

A much better place to launch the study of distance measures for quantum information is with the comparison of classical probability distributions.

The relative frequency difference of information sources as probability distributions over some alphabet encourages us to concentrate on the comparison of probability distributions in our search for measures of distance.

What does it mean to say that tow probability distributions $\{p_x\}$ and $\{q_x\}$ over the same index set, x, are similar to one another? (there's no unique correct answer)The widely used two measures are the trace distance and the fidelity.

The *trace distance* is defined by the equation:

$$D(p_x, q_x) \equiv \frac{1}{2} \sum_x |p_x - q_x|. \tag{9.1}$$

This quantity is sometimes known as the L_1 *distance* or *Kolmogorov distance*.

The trace distance is a metric on probability distributions.

1. $D(x, x) = 0$
2. $D(x, y) = D(y, x)$
3. $D(x, z) \leq D(x, y) + D(y, z)$

The second measure of distance between probability distributions, the *fidelity* of the probability distributions $\{p_x\}$ and $\{q_x\}$, is defined by

$$F(p_x, q_x) \equiv \sum_x \sqrt{p_x q_x}. \tag{9.2}$$

The fidelity is a very different way of measuring distance between probability distributions that is the trace distance. It is not a metric ($F(p_x, p_x) = \sum_x p_x = 1$) A better geometric understanding of the fidelity is $F(p, q) = \sqrt{p} \cdot \sqrt{q} = \cos \sigma$.

The trace distance and fidelity are mathematically useful means of defining the notion of a distance between two probability distributions.

Do these measures have physically motivated operational meaning? In the case of trace distance, the answer to this question is yes.

$$D(p_x, q_x) = \max_S |p(S) - q(S)| = \max_S \left| \sum_{x \in S} p_x - \sum_{x \in S} q_x \right|, \tag{9.3}$$

where the maximization is over all subsets S of the index set $\{x\}$. The quantity being maximized is the difference between the probability that the event S occurs, according to the distribution $\{p_x\}$, and the probability that the event S occurs, according to the distribution $\{q_x\}$. The event S is thus in some sense the optimal event to examine when trying to distinguish the distance $\{p_x\}$ and $\{q_x\}$, with the trace distance governing how well it is possible to make this distribution.

Unfortunately, a similarly clear interpretation for the fidelity is not known. (open questions?)

It turns out that there are close connections between the fidelity and the trace distance, so properties of one quantity can often be used to deduce properties of the other.

The trace distance and fidelity are static measures of distance for comparing two fixed probability distributions.

A *dynamic measure of distance* : measures how well information is preserved by some physical process.

X : a input
Y : a output
(to form a Markov process : $X \to Y$.

This dynamic measure of distance can be understood as a special case of the static trace distance!

1. A copy of the input X : $\tilde{X} = X$

2. X now passes through the noisy channel, leaving as output the random variable Y.

How close is the initial perfectly correlated pair, $(\tilde{X}, X)$, to the final pair, $(\tilde{X}, Y)$? Using the trace distance as our measure of 'closeness', we obtain with some simple algebra,

$$\begin{aligned}
D((\tilde{X}, X), (X, Y)) &= \frac{1}{2} \sum_{xx'} |\delta_{xx'} p(X = x) - p(\tilde{X} = x, Y = x')| && (9.4)\\
&= \frac{1}{2} \sum_{x \neq x'} p(\tilde{X} = x, Y = x') + \frac{1}{2} \sum_x |p(X = x) - p(\tilde{X} = x, Y = x')| && (9.5)\\
&= \frac{1}{2} \sum_{x \neq x'} p(\tilde{X} = x, Y = x') + \frac{1}{2} \sum_x (p(X = x) - p(\tilde{X} = x, Y = x')) && (9.6)\\
&= \frac{p(\tilde{X} \neq Y) + 1 - p(\tilde{X} = Y}{2} && (9.7)\\
&= \frac{p(X \neq Y) + p(\tilde{X} \neq Y)}{2} && (9.8)\\
&= p(X \neq Y). && (9.9)
\end{aligned}$$

Thus the probability of an error in the channel is equal to the trace distance between the probability distribution for $(\tilde{X}, X)$ and $(\tilde{X}, Y)$. This is an important construction, since it will be the basis for analogous quantum constructions. This is necessary because there is no *direct* quantum analogue of the probability $p(X \neq Y)$, since there is no notion in quantum mechanics analogous to the joint probability distribution for variables X and Y that *exist at different times*

9.2 How close are two quantum state?

9.2.1 Trace distance

We begin by defining the *trace distance* between quantum states ρ and σ,

$$D(\rho, \sigma) \equiv \frac{1}{2}\mathrm{tr}|\rho - \sigma|, \tag{9.11}$$

where as per usual we define $|A| \equiv \sqrt{A^\dagger A}$.

IF ρ and σ commute, quantum trace distance between ρ and σ is equal to the classical distance between the eigenvalues of ρ and σ.

$$\rho = \sum_i r_i |i\rangle\langle i|; \;\; \sigma \sum_i s_i |i\rangle\langle i|. \tag{9.12}$$

Thus

$$\begin{aligned} D(\rho, \sigma) &= \frac{1}{2}\mathrm{tr}\left|\sum_i (r_i - s_i)|i\rangle\langle i|\right| && (9.13)\\ &= D(r_i, s_i) && (9.14)\end{aligned}$$

A good way of getting a feel for the trace distance is to understand it for the special case of a qubit, in the Bloch sphere representation. Suppose ρ and σ have respective Bloch vectors $\vec{r}$ and $\vec{s}$,

$$\rho = \frac{I + \vec{r}\cdot\vec{\sigma}}{2}; \;\; \sigma = \frac{I + \vec{s}\cdot\vec{\sigma}}{2}. \tag{9.17}$$

The trace distance between ρ and σ is easily calculated:

$$\begin{aligned} D(\rho, \sigma) &= \frac{1}{2}\mathrm{tr}|\rho - \sigma| && (9.18)\\ &= \frac{1}{4}\mathrm{tr}|(\vec{r} - \vec{s})\cdot\vec{\sigma}|. && (9.19)\end{aligned}$$

$(\vec{r} - \vec{s})\cdot\vec{\sigma}$ has eigenvalues $\pm|\vec{r} - \vec{s}|$, so the trace of $|(\vec{r} - \vec{s})\cdot\vec{\sigma}|$ is $|\vec{r} - \vec{s}|$, and we see that

$$D(\rho, \sigma) = \frac{|\vec{r} - \vec{s}|}{2}. \tag{9.20}$$

That is, the distance between two single qubit states is equal to one half the ordinary Euclidean distance between them on the Bloch sphere.

eg) rotations of the Bloch sphere leave the Euclidean distance invariant

$$D(U\rho U^\dagger, U\sigma U^\dagger) = D(\rho, \sigma). \tag{9.21}$$

The trace distance generalization is

$$D(\rho, \sigma) = \max_P \mathrm{tr}\,[2]\,(P(\rho - \sigma)), \tag{9.22}$$

where the maximization may be taken alternately over all projectors, P, or over all positive operators $P \leq I$.

$$\rho - \sigma = Q - S,$$

where Q and S are positive operators with orthogonal support. It implies that

$$|\rho - \sigma| = Q + S,$$

so

$$D(\rho, \sigma) = \frac{1}{2}(\mathrm{tr}(Q) + \mathrm{tr}(S)).$$

But

$$\mathrm{tr}(Q - S) = \mathrm{tr}(\rho - \sigma) = 0,$$

so

$$\mathrm{tr}(Q) = \mathrm{tr}(S)$$

and therefore

$$D(\rho, \sigma) = \mathrm{tr}(Q).$$

Let P be the projector onto the support of Q. Then

$$\begin{aligned} \mathrm{tr}(P(\rho - \sigma)) &= \mathrm{tr}(P(Q - S)) \\ &= \mathrm{tr}(Q) \\ &= D(\rho, \sigma). \end{aligned}$$

Conversely, let P be any projector. Then $\mathrm{tr}\left[P(\rho - \sigma)\right] = \mathrm{tr}P(Q - S) \leq \mathrm{tr}PQ \leq \mathrm{tr}(Q) = D(\rho, \sigma)$. There is a closely related way of viewing the quantum trace distance which relates it more closely to the classical trace distance.

Theorem 9.1. *Let $\{E_m\}$ be a POVM, with $p_m \equiv \mathrm{tr}\rho E_m$ and $q_m \equiv \mathrm{tr}\sigma E_m$ as the probabilities of obtaining a measurement outcome labeled by m. Then*

$$D(\rho, \sigma) = \max_{\{E_m\}} D(p_m, q_m), \tag{9.23}$$

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

Proof
Note that

$$D(p_m, q_m) = \frac{1}{2} \sum_m |\mathrm{tr}(E_m(\rho - \sigma))|. \tag{9.24}$$

$$\left(\because (9.1)\ D(p_x, q_x) \equiv \frac{1}{2} \sum_x |p_x - q_x| \right)$$

We may write

$$\begin{aligned} |\mathrm{tr}(E_m(\rho - \sigma))| &= |\mathrm{tr}(E_m(Q - S))| && (9.25) \\ &\leq \mathrm{tr}(E_m(Q + S)) && (9.26) \\ &\leq \mathrm{tr}(E_m|\rho - \sigma|). && (9.27) \end{aligned}$$

Thus

$$\begin{aligned} D(p_m, q_m) &\leq \frac{1}{2} \sum_m \mathrm{tr}(E_m|\rho - \sigma|) && (9.28) \\ &= \frac{1}{2}\mathrm{tr}|\rho - \sigma| \left(\because \sum_m E_m = I \right) && (9.29) \\ &= D(\rho, \sigma). && (9.30) \end{aligned}$$

Conversely, by choosing a measurement whose POVM elements include projetors onto the support for Q and S, we see that there exist measurements which give rise to probability distributions such that $D(p_m, q_m) = D(\rho, \sigma)$. □

If two density operators are close in trace distance, then any measurement performed on those quantum states will give rise to probability distributions which are close together in the classical sense of trace distance, giving a second interpretation of the trace distance between probability distributions arising from measurements performed on those quantum states.

The trace distance is a metric on the space of density operators.

$$D(\rho, \sigma) = 0$$

iff $\rho = \sigma$, and $D(.,.)$ is symmetric:

$$D(\rho, \sigma) = D(\sigma, \rho).$$

Moreover, the triangle inequality holds,

$$D(\rho, \tau) \leq D(\rho, \sigma) + D(\sigma, \tau). \tag{9.31}$$

From Equation (9.22) that there exists a project P such that

$$\begin{aligned} D(\rho, \tau) &= \mathrm{tr}(P(\rho - \tau)) && (9.32)\\ &= \mathrm{tr}(P(\rho - \sigma)) + \mathrm{tr}(P(\sigma - \tau)) && (9.33)\\ &\leq D(\rho, \sigma) + D(\rho + \tau). && (9.34)\\ &\quad (\ \because\ \mathrm{tr}(P(\rho - \sigma)) \leq \mathrm{tr}(P_{\rho-\sigma}(\rho - \sigma))\ ) && (9.35) \end{aligned}$$

The most interesting result is that no physical process ever increase the distance between two quantum states.

Theorem 9.2 (Trace-preserving quantum operations are contractive). *Suppose $\mathcal{E}$ is a trace-preserving quantum operation. Let ρ and σ be density oprators. Then*

$$D(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq D(\rho, \sigma). \tag{9.36}$$

Proof
We see that

$$\begin{aligned} D(\rho, \sigma) &= \frac{1}{2}\mathrm{tr}|\rho - \sigma| && (9.37)\\ &= \frac{1}{2}\mathrm{tr}|Q - S|\ (\ \because \rho - \sigma = Q - S,\ |\rho - \sigma| = Q + S\ ) && (9.38)\\ &= \frac{1}{2}\mathrm{tr}(Q) + \frac{1}{2}\mathrm{tr}(S) && (9.39)\\ &= \frac{1}{2}\mathrm{tr}(\mathcal{E}(Q)) + \frac{1}{2}\mathrm{tr}(\mathcal{E}(S))\ (\because \text{ trace preserving }) && (9.40)\\ &= \mathrm{tr}(\mathcal{E}(Q))\ (\ \because \mathrm{tr}(\rho - \sigma) = \mathrm{tr}(Q - S) = 0\ ) && (9.41)\\ &\geq \mathrm{tr}(P\mathcal{E}(Q)) && (9.42)\\ &\geq \mathrm{tr}(P(\mathcal{E}(Q) - \mathcal{E}(S))) && (9.43)\\ &= \mathrm{tr}(P(\mathcal{E}(\rho) - \mathcal{E}(\sigma))) && (9.44)\\ &= D(\mathcal{E}(\rho), \mathcal{E}(\sigma)), && (9.45) \end{aligned}$$

which completes the proof. □

eg 1) Gallery
eg 2) If we 'cover up' parts of two quantum states then we can show that the distance between those two states is never increased. If we take quantum state ρ^{AB} and σ^{AB} of a composite system AB then the distance between $\rho^A = \mathrm{tr}_\mathrm{B}\left[\rho^{AB}\right]$ and $\sigma^A = \mathrm{tr}_\mathrm{B}\left[\sigma^{AB}\right]$ is never more than the distance between ρ^{AB} and σ^{AB},

$$D(\rho^A, \sigma^A) \leq D(\rho^{AB}, \sigma^{AB}). \tag{9.46}$$

In many applications we want to estimate the trace distance for a mixture of inputs. Such estimates are greatly aided by the following theorem:

Theorem 9.3 (Strong convexity of the trace distance). *Let $\{p_i\}$ and $\{q_i\}$ be probability distributions over the same index set, and ρ_i and σ_i be density operators, also with indices from the same index set. Then*

$$D\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) \leq D(p_i, q_i) + \sum_i p_i D(\rho_i, \sigma_i), \tag{9.47}$$

where $D(p_i, q_i)$ is the classical trace distance between the probability distribution $\{p_i\}$ and $\{q_i\}$.

Proof
By Equation(9.22) there exists a projector P such that

$$\begin{aligned}
& D\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) && (9.48)\\
&= \sum_i p_i \mathrm{tr}(P\rho_i) - \sum_i q_i \mathrm{tr}(P\sigma_i) && (9.49)\\
&= \sum_i \mathrm{tr}(P(\rho_i - \sigma_i)) + \sum_i (p_i - q_i)\mathrm{tr}(P\sigma_i) && (9.50)\\
&\quad (\ \because \pm \sum_i \mathrm{tr}(P\sigma_i)\ )\\
&\leq \sum_i p_i D(\rho_i, \sigma_i) + D(p_i, q_i) && (9.51)\\
&\quad (\ \because \mathrm{tr}(P(\rho_i - \sigma_i) \leq \mathrm{tr}(P_{\rho_i - \sigma_i}(\rho_i - \sigma_i) = \sum_i p_i D(\rho_i, \sigma_i)\ )\\
&\quad (\ \because \sum_i (p_i - q_i)\mathrm{tr}(P\sigma_i) = \sum_{i\in S}(p_i - q_i) \leq \max_S \sum_{i\in S}(p_i - q_i) = D(p_i, q_i)\ ) && (9.52)
\end{aligned}$$

9.2.2 Fidelity

A second measure of distance between quantum states is the *fidelity*. The dielity is not a metric on density operators, but we will see that it does give rise to a useful metric. The fidelity of states ρ and σ is defined to be

$$F(\rho, \sigma) \equiv \mathrm{tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}} \tag{9.53}$$

eg1) When $[\rho, \sigma] = 0$,

$$\rho = \sum_i r_i |i\rangle\langle i|; \ \ \sigma = \sum_i s_i |i\rangle\langle i|. \tag{9.54}$$

$$\begin{aligned}
F(\rho, \sigma) &= \mathrm{tr}\sqrt{\sum_i r_i s_i |1\rangle\langle 1|} && (9.55)\\
&= \mathrm{tr}\left(\sum_i \sqrt{r_i s_i}|i\rangle\langle i|\right) && (9.56)\\
&= \sqrt{r_i s_i} && (9.57)\\
&= F(r_i, s_i). && (9.58)
\end{aligned}$$

That is, when ρ and σ commute the quantum fidelity $F(\rho, \sigma)$ reduces to the classical fidelity $F(r_i, s_i)$ between the eigenvalues distributions r_i and s_i of ρ and σ.

eg2) The fidelity between a pure state $|\psi\rangle$ and an arbitrary state ρ. We see that

$$F(|\psi\rangle, \rho) = \mathrm{tr}\sqrt{\langle\psi|\rho|\psi\rangle|\psi\rangle\langle\psi|} \tag{9.59}$$
$$= \sqrt{\langle\psi|\rho|\psi\rangle}. \tag{9.60}$$

That is, the fidelity is equal to the square root of the overlap between $|\psi\rangle$ and ρ. This is an important result which we will make use of often.

There's no clear geometric interpretation for the fidelity between two states of a qubit.

However, the fidelity does satisfy many of the same properties as the trace distance. For example, it is invariant under unitary transformations:

$$F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho, \sigma). \tag{9.61}$$

There is also a useful characterization of the fidelity analogous to the characterization (**??**9.22)) for the trace distance.

Theorem 9.4 (Uhlmann's theorem). *Suppose ρ and σ are states of a quantum system Q. Introduce a second quantum system R which is a copy of Q. Then*

$$F(\rho, \sigma) = \max_{|\psi\rangle, |\phi\rangle} |\langle\psi|\phi\rangle|, \tag{9.62}$$

where the maximization is over all purification $|\psi\rangle$ of ρ and $|\phi\rangle$ of σ into RQ.

Lemma 9.5. *Let A be any operator, and U unitary. Then*

$$|\mathrm{tr}(AU)| \leq \mathrm{tr}|A|, \tag{9.63}$$

with equality being attained by choosing $U = V^\dagger$, where $A = |A|V$ is the polar docomposition of A.

Proof
Equality is clearly attained under the conditins stated. Observe that

$$|\mathrm{tr}(AU)| = |\mathrm{tr}(|A|VU)| = |\mathrm{tr}|A|^{1/2}|A|^{1/2}VU|. \tag{9.64}$$

The Cauchy–Schwarz inequality for the Hilbert-Schmidt inner product gives:

$$|\mathrm{tr}(AU)| \leq \sqrt{\mathrm{tr}|A|\mathrm{tr}(U^\dagger V^\dagger |A| VU)} = \mathrm{tr}|A|, \tag{9.65}$$

which is completes the proof. *Proof*
Uhlmann's theorem
Fix orthonormal bases $|i_R\rangle$ and $|i_Q\rangle$ in systems R and Q. Because R and Q are of the same dimension, the index i may be assumed to run over the same set of values. Define $|m\rangle = \sum_i |i_R\rangle|i_Q\rangle$. Let $|\psi\rangle$ be any purification of ρ. Then the Schmidt decomposition and a moment's thought should convince you that

$$|\psi\rangle = (U_R \otimes \sqrt{\rho} U_Q)|m\rangle, \tag{9.66}$$

for some unitary operators U_R and U_Q. Similarly, if $|\phi\rangle$ is any purification of σ then there exist unitary operators V_R and V_Q such that

$$|\phi\rangle = (V_R \otimes \sqrt{\sigma} V_Q)|m\rangle. \tag{9.67}$$

Then, the inner product is

$$|\langle\psi|\phi\rangle| = \langle m|\Big({U_R}^\dagger V_R \otimes {U_Q}^\dagger \sqrt{\rho}\sqrt{\sigma} V_Q\Big)|m\rangle|. \tag{9.68}$$

Using

$$\mathrm{tr}(A^\dagger B) = \langle m|(A \otimes B)|m\rangle,$$
$$|\langle\psi|\phi\rangle| = |\mathrm{tr}\Big({V_R}^\dagger U_R {U_Q}^\dagger \sqrt{\rho}\sqrt{\sigma} V_Q\Big)|. \tag{9.69}$$

Setting $U \equiv V_Q {V_R}^\dagger U_R {U_Q}^\dagger$ we see that

$$|\langle\psi|\phi\rangle| = |\mathrm{tr}(\sqrt{\rho}\sqrt{\sigma}U)|. \tag{9.70}$$

By Lemma 9.5,

$$|\langle\psi|\phi\rangle| \leq \mathrm{tr}|\sqrt{\rho}\sqrt{\sigma}| = \mathrm{tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}. \tag{9.71}$$

To see that equality may be attained, suppose $\sqrt{\rho}\sqrt{\sigma} = |\sqrt{\rho}\sqrt{\sigma}|V$ is the polar decomposition of $\sqrt{\rho}\sqrt{\sigma}$.

Properties of the fidelity are more easily proved using Uhlmann's formula in many instances.

*eg)*Uhlmann's formula shows that the fidelity is

- symmetric in its inputs, $F(\rho,\sigma) = F(\sigma,\rho)$
- bounded between 0 and 1, $0 \leq F(\rho,\sigma) \leq 1$.

On the other hand original formula of the fidelity is sometimes useful as a means for understanding properties of the fidelity. For instance, we see that $F(\rho,\sigma) = 0$ iff ρ and σ have support on orthogonal subspaces. $\rightarrow$ perfectly distinguishable. The relation between the quantum fidelity and the classical fidelity.

$$F(\rho,\sigma) = \min_{E_m} F(p_m, q_m), \tag{9.74}$$

where the minimum is over all POVMs $\{E_m\}$, and $p_m \equiv \mathrm{tr}(\rho E_m)$, $q_m \equiv \mathrm{tr}(\sigma E_m)$ are the probability distributions for ρ and σ corresponding to the POVMs $\{E_m\}$. To see that this is true, note that

$$\begin{aligned} F(\rho,\sigma) &= \mathrm{tr}(\sqrt{\rho}\sqrt{\sigma}U) && (9.75)\\ &= \sum_m \mathrm{tr}(\sqrt{\rho}\sqrt{E_m}\sqrt{E_m}\sqrt{\sigma}U). \ (\ \because \sum_m E_m = I\ ) && (9.76)\end{aligned}$$

The Cauchy-Schwarz inequality and some simple algebra gives

$$\begin{aligned} F(\rho,\sigma) &\leq \sqrt{\mathrm{tr}(\rho E_m)\mathrm{tr}(\sigma E_m)} && (9.77)\\ &= F(p_m, q_m), && (9.78)\end{aligned}$$

which establishes that

$$F(\rho,\sigma) \leq F(p_m, q_m). \tag{9.79}$$

To see that equality can be obtained in this inequality, we need to find a POVM $\{E_m\}$ such that the Cauchy–Schwarz inequality is satisfied with equality for each term in the sum, that is,

$$\sqrt{E_m}\sqrt{\rho} = \alpha_m\sqrt{E_m}\sqrt{\sigma}U$$

for some complex number α_m. But

$$\sqrt{\rho}\sqrt{\sigma} = \sqrt{\rho^{1/2}\sigma\rho^{1/2}}$$

, so for invertible ρ,

$$\sqrt{\sigma}U = \rho^{-1/2}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}. \tag{9.80}$$

We find the equality conditions are that

$$\sqrt{E_m}(I - \alpha_m M) = 0 \tag{9.81}$$

$$\begin{aligned} (\ \because\ \sqrt{E_m}\sqrt{\rho} &= \alpha_m\sqrt{E_m}\rho^{-1/2}\sqrt{\rho^{1/2}\sigma\rho^{1/2}}\\ \sqrt{E_m} &= \alpha\sqrt{E_m}M\ ) \end{aligned}$$

Three important properties of the trace distance

- the metric property
- contractivity

- strong convexity

Analogous properties all hold for the fidelity.

The fidelity is not a metric; however, there is a simple way of turning the fidelity into a metric. Uhlmann's theorem tells us that the fidelity between two states is equal to the maximum inner product between purifications of those states. We define the *angle* between states ρ and σ by

$$A(\rho, \sigma) \equiv \arccos F(\rho, \sigma). \tag{9.82}$$

Obviously the angle is non-negative, symmetric in its inputs, and is equal to zero iff $\rho = \sigma$. $\rightarrow$ a metric.

We prove the triangle inequality using Uhlmann's theorem and some obvious facts about vectors in three dimensions. Let $othercat$, $|\psi\rangle$, $|\gamma\rangle$ be purifications of σ, ρ, and τ respectively such that

$$F(\rho, \sigma) = \langle\psi|\phi\rangle \tag{9.83}$$
$$F(\sigma, \tau) = \langle\phi|\gamma\rangle, \tag{9.84}$$

and $\langle\psi|\gamma\rangle$ is real and positive. ($\rightarrow$ **imagine Schmidt decomposition**) It is clear that (**image generic pure states on a sphere**)

$$\arccos(\langle\psi|\gamma\rangle) \leq A(\rho, \sigma) + A(\sigma, \tau). \tag{9.85}$$

But by Uhlmann's theorem, $F(\rho, \tau) \geq \langle\psi|\gamma\rangle$, so $A(\rho, \tau) \leq \arccos(\langle\psi|\gamma\rangle)$. Combining with the previous inequality gives the triangle inequality,

$$A(\rho, \tau) \leq A(\rho, \sigma) + A(\sigma, \tau). \tag{9.86}$$

Qualitively, the fidelity behaves like an 'upside-down' version of the trace distance, decreasing as two states become more distinguishable, and increasing as they become less distinguishable. Therefore

- trace distance: *contractivity* $\rightarrow$ fidelity: *monotonicity*
- trace distance: *strong concavity* $\rightarrow$ fidelity: *strong convexity*

Theorem 9.6 (Monotonicity of the fidelity). *Suppose $\mathcal{E}$ is a trace-preserving quantum operation. Let ρ and σ be density operators. Show that*

$$F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \geq F(\rho, \sigma) \tag{9.87}$$

Proof
Let $|\psi\rangle$ and $|\phi\rangle$ be purification of ρ and σ into a joint system RQ such that $F(\rho, \sigma) = |\langle\psi|phi\rangle|$. Introduce a model environment E for the quantum operation, $\mathcal{E}$, which starts in a pure state $|0\rangle$, and interacts with the quantum system Q via a unitary interaction U. Note that $U|\psi\rangle|0\rangle$ is a purification of $\mathcal{E}(\rho)$, and $U|\phi\rangle|0\rangle$ is a purification of $\mathcal{E}(\sigma)$. By Ulhmann's theorem it follows that

$$F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \geq |\langle\psi|\langle 0|U^\dagger U|\phi\rangle|0\rangle| \tag{9.88}$$
$$= |\langle\psi|\phi\rangle| \tag{9.89}$$
$$= F(\rho, \sigma), \tag{9.90}$$

establishing the property we set out to prove. □

Theorem 9.7 (Strong concavity of the fidelity). *Let p_i and q_i be probability distributions over the same index set, and ρ_i and σ_i density operators also indexed by the same index set. Then*

$$F\left(\sum_i p_i \rho_i, \sum_i q_i \sigma_i\right) \geq \sum_i \sqrt{p_i q_i} F(\rho_i, \sigma_i). \tag{9.91}$$

Proof
Let $|\psi_i\rangle$ and $|\phi_i\rangle$ be purification of ρ_i and σ_i chosen such that $F(\rho_i, \sigma_i) = \langle\psi_i|\phi_i\rangle$. Introduce an ancillary system which has orthonormal basis states $|i\rangle$ corresponding to the index set i for the probability distributions. Define

$$|\psi\rangle \equiv \sum_i \sqrt{p_i}|\psi_i\rangle|i\rangle; \;\; |\phi\rangle \equiv \sum_i \sqrt{q_i}|\phi_i\rangle|i\rangle. \tag{9.92}$$

Note that $|\psi\rangle$ is a purification of $\sum_i p_i\rho_i$ and $|\phi\rangle$ is a purification of $\sum_i q_i\sigma_i$,

$$\left(cf. \quad \begin{array}{ll} \mathrm{tr}_{\mathrm{R}}(|\psi\rangle\langle\psi|) & = \sum_i p_i|\psi_i\rangle\langle\psi_i| \\ & = \sum_i p_i\rho_i \end{array} \right)$$

$$=$$

so by Uhlmann's formula,

$$F\left(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\right) \geq |\langle\psi|\phi\rangle| = \sum_i \sqrt{p_iq_i}\langle\psi_i|\phi_i\rangle = \sum_i \sqrt{p_iq_i}F(\rho_i, \sigma_i). \tag{9.93}$$

which establishes the result we set out to prove. □

9.2.3 Relationships between distance measures

The trace distance and the fidelity may be considered to be *equivalent* measures of distance for many applications.

In the case of pure states, the trace distance and the fidelity are completely equivalent to one another. Consider the trace distance between two pure states $|a\rangle$ and $|b\rangle$. Using Gram-Schmidt procedure we may find orthonormal states $|0\rangle$ and $|1\rangle$ such that

$$|a\rangle = |0\rangle; \;\; |b\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle.$$

Note that $F(|a\rangle, |b\rangle) = |\cos\theta|$. Furthermore,

$$\begin{aligned} D(|a\rangle, |b\rangle) &= \frac{1}{2}\mathrm{tr}\left|\begin{bmatrix} 1-\cos^2\theta & -\cos\theta\sin\theta \\ -\cos\theta\sin\theta & -\sin^2\theta \end{bmatrix}\right| && (9.94)\\ &= |\sin\theta| && (9.95)\\ &= \sqrt{1 - F(|a\rangle, |b\rangle)^2}. && (9.96)\\ &= (\;\; F^2 + D^2 = 1 \;\; \textit{cf.} \text{ PRA } \mathbf{50}\ 1047) && (9.97) \end{aligned}$$

This relationship at the level of pure states can be used to deduce a relationship at the level of mixed states. Let ρ and σ be any two quantum states, and let $|\psi\rangle$ and $|\phi\rangle$ be purifications chosen such that $F(\rho, \sigma) = |\langle\psi|\phi\rangle| = F(|\psi\rangle, |\phi\rangle)$. We see that

$$\begin{aligned} D(\rho, \sigma) &\leq D(|\psi\rangle, |\phi\rangle) && (9.98)\\ &= \sqrt{1 - F(\rho, \sigma)^2}. && (9.99) \end{aligned}$$

Thus, if the fidelity between two states is close to one, it follows that the states are also close in trace distance. The converse is also true. To see this, let $\{E_m\}$ be a POVM such that

$$F(\rho, \sigma) = \sum_m \sqrt{p_m q_m}, \tag{9.100}$$

where $p_m \equiv \mathrm{tr}(\rho E_m)$, $q_m \equiv \mathrm{tr}(\sigma E_m)$ are the probabilities for obtaining outcome m for the states rho and σ, respectively. Observe first that

$$\begin{aligned} \sum_m (\sqrt{p_m} - \sqrt{q_m})^2 &= \sum_m p_m + \sum_m q_m - 2F(\rho, \sigma) \\ &= 2(1 - F(\rho, \sigma)). \end{aligned}$$

However, it is also true that $|\sqrt{p_m} - \sqrt{q_m}| \leq |\sqrt{p_m} + \sqrt{q_m}|$, so

$$\begin{aligned}
\sum_m (\sqrt{p_m} - \sqrt{q_m})^2 &= \sum_m |\sqrt{p_m} - \sqrt{q_m}||\sqrt{p_m} + \sqrt{q_m}| && (9.101)\\
&= \sum_m |p_m - q_m| && (9.102)\\
&= 2D(p_m, q_m) && (9.103)\\
&\leq 2D(\rho, \sigma). \ (\ \because \text{(9.23)}) && (9.104)
\end{aligned}$$

Thus

$$1 - F(\rho, \sigma) \leq D(\rho, \sigma). \tag{9.105}$$

Summarizing, we have

$$1 - F(\rho, \sigma) \leq D(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)^2}. \tag{9.106}$$

The implication is that the trace distance and the fidelity are qualitatively equivalent measures of closeness for quantum states. Indeed, for many purposes it does not matter whether the trace distance or the fidelity is used to quantify distance, since results about one may be used to deduce equivalent results about the other.

9.3 How well does a quantum channel preserve information?

– for dynamics distance measurements.
$|\psi\rangle \rightarrow \mathcal{E}(|\psi\rangle\langle\psi|)$

This type of scenario occurs often in quantum computation and quantum information.

eg) we can compute the fidelity between the starting state $|\psi\rangle$ and the ending state $\mathcal{E}(|\psi\rangle\langle\psi|)$. For the case of the depolarizing channel(page 378, (8.100)), we obtain

$$\begin{aligned}
F(|\psi\rangle, \mathcal{E}(|\psi\rangle\langle\psi|)) &= \sqrt{\langle\psi| \left(p\frac{I}{2} + (1-p)|\psi\rangle\langle\psi| \right) |\psi\rangle} && (9.112)\\
&= \sqrt{1 - \frac{p}{2}}. && (9.113)
\end{aligned}$$

This result agrees well with our intuition - the higher the probability p of depolarizing, the lower the fidelity of the final state with the initial state. ($\rightarrow$ We could equally well have used the trace distance)

In a real quantum memory or quantum communications channel, we don't know in advance what the initial state $|\psi\rangle$ of the system will be. However, we can quantify the worst-case behavior of the system by minimizing over all possible initial states,

$$F_{\min}(\mathcal{E}) \equiv \min_{|\psi\rangle} F(|\psi\rangle, \mathcal{E}(|\psi\rangle, |\psi\rangle\langle\psi|)). \tag{9.114}$$

For the phase damping channel,

$$\mathcal{E}(\rho) = p\rho + (1-p)Z\rho Z, \tag{9.115}$$

the fidelity is given by

$$\begin{aligned}
F(|\psi\rangle, \mathcal{E}(|\psi\rangle\langle\psi|)) &= \sqrt{\langle\psi|(p|\psi\rangle\langle\psi| + (1-p)Z|\psi\rangle\langle\psi|Z)|\psi\rangle} && (9.116)\\
&= \sqrt{p + (1-p)\langle\psi|Z|\psi\rangle^2}. && (9.117)
\end{aligned}$$

The second term is

$$0 \leq \langle\psi|Z|\psi\rangle^2 \leq 1. \ (\ |\psi\rangle = |+\rangle \rightarrow \langle Z\rangle = 0\ )$$

Thus for the phase damping channel the minimum fidelity is

$$F_{\min}(\mathcal{E}) = \sqrt{p}. \tag{9.118}$$

You might wonder why we minimized over *pure states* in the definition of $F_{\min}$. Might not the quantum system of interest start in a mixed state ρ? Fortunately, the joint concavity of the fidelity can be used to show that allowing mixed states can be used to show that allowing mixed states does not change $F_{\min}$.

Suppose that $\rho = \sum_i \lambda_i |i\rangle\langle i|$ is the initial state of the quantum system. Then we have

$$F(\rho, \mathcal{E}(\rho)) = F\left(\sum_i \lambda_i |i\rangle\langle i|, \sum_i \lambda_i \mathcal{E}(|i\rangle\langle i|)\right) \tag{9.119}$$

$$\geq \sum_i \lambda_i F(|i\rangle, \mathcal{E}(|i\rangle\langle i|)). \tag{9.120}$$

It follows that

$$F(\rho, \mathcal{E}(\rho)) \geq F(|i\rangle, \mathcal{E}(|i\rangle\langle i|)) \tag{9.121}$$

for at least one of the states $|i\rangle$, and thus $F(\rho, \mathcal{E}(\rho)) \geq F_{\min}$

Suppose that as part of a quantum computation we attempt to implement a quantum gate described by the unitary operator U. (the operation will have inevitably some noise) The correct description of the gate is using a trace-preserving quantum operation $\mathcal{E}$.A natural measure of how successful our gate has been is the *gate fidelity*,

$$F(U, \mathcal{E}) \equiv \min_{|\psi\rangle} F(U|\psi\rangle, \mathcal{E}(|\psi\rangle\langle\psi|)). \tag{9.122}$$

eg) A NOT gate on a single qubit, but instead implement the noisy operation

$$\mathcal{E}(\rho) = (1-p)X\rho X + pZ\rho Z$$

. Then the gate fidelity is given by

$$F(X, \mathcal{E}) = min_{|\psi\rangle} \sqrt{\langle\psi|(X(1-p)X|\psi\rangle\langle\psi|X + pZ|\psi\rangle\langle\psi|Z)X)|\psi\rangle} \tag{9.123}$$

$$= min_{|\psi\rangle} \sqrt{(1-p) + p\langle\psi|Y|\psi\rangle^2} \tag{9.124}$$

$$= \sqrt{1-p} \tag{9.125}$$

Quantum sources of information and the entanglement fidelity

Quantum sources of information;

- a ensemble : a stream of identical quantum systems being produce by some physical process
- entanglement : motivated by the idea that a channel which preserves information well is a channel that preserves entanglement

Thus those two sources lead to notions of *ensemble average fidelity*, and *entanglement fidelity*.

The ensemble fidelity captures the idea that the source is well-reserved under the action of a noisy channel described by a trace-preserving quantum operation $\mathcal{E}$, namely

$$\bar{F} = \sum_j p_j F(\rho_j, \mathcal{E}(\rho_j))^2, \tag{9.126}$$

where p_j are the respective probabilities for the different possible preparations of the system ρ_j. ($0 \leq \bar{F} \leq 1$) $F(\rho_j, \mathcal{E}(\rho_j))^2 \rightarrow$ it is not entirely clear what the 'correct'definitions for notions such as information preservation are.

The entanglement fidelity is from the discussion of the classical probability of error in Section 9.1. ($\rightarrow$ there is no direct quantum analogue of a probability distribution defined at two times.

A quantum system, Q, is prepared in a state ρ. The state of Q is assumed to be *entangled* in some way with the external world. This entanglement replaces the (quantum) correlation between X and $\tilde{X}$ in the classical model. We represent the entanglement by introduction a fictitious system R, such that the joint state of RQ is a pure state. It turns out that all results that we prove do not depend in any way on how this purification is performed, so we may as well suppose that this is an arbitrary entanglement with the outside world. The system Q is then subjected to a dynamics described by a quantum operation, $\mathcal{E}$.

The entanglement fidelity $F(\rho, \mathcal{E})$ is

$$\begin{aligned} F(\rho, \mathcal{E}) &\equiv F(RQ, R'Q')^2 && (9.127)\\ &= \langle RQ|[(\mathcal{I}_R \otimes \mathcal{E})(|RQ\rangle\langle RQ|)]|RQ\rangle, && (9.128)\end{aligned}$$

where the use of a prime indicates the state of a system after the quantum operation has been applied, and the absence of a prime indicates the state of a system before the quantum operation has been applied. **The use of the square of the static fidelity is purely a convenience**. The entanglement fidelity depends only upon ρ and $\mathcal{E}$. To see this, use $|R_2Q_2\rangle = U|R_1Q_1\rangle$. Thus

$$F(|R_2Q_2\rangle, \rho^{R_2'Q_2'}) = F(|R_1Q_1\rangle, \rho^{R_1'Q_1'}). \tag{9.129}$$

The entanglement fidelity provides a measure of how well the entanglement between R and Q is preserved by the process $\mathcal{E}$.

Suppose E_i is a set of operation elements for a quantum operation $\mathcal{E}$. Then

$$F(\rho, \mathcal{E}) = \langle RQ|\rho^{R'Q'}|RQ\rangle = \sum_i |\langle RQ|E_i|RQ\rangle|^2. \tag{9.130}$$

Suppose we write $|RQ\rangle = \sum_j \sqrt{p_j}|j\rangle|j\rangle$, where $\rho = \sum_j p_j|j\rangle\langle j|$. Then

$$\begin{aligned} \langle RQ|E_i|RQ\rangle &= \sum_{jk} \sqrt{p_j p_k}\langle j|k\rangle\langle j|E_i|k\rangle && (9.131)\\ &= \sum_j p_j\langle j|E_i|j\rangle && (9.132)\\ &= \mathrm{tr}(E_i\rho). && (9.133)\end{aligned}$$

Thus

$$F(\rho, \mathcal{E}) = \sum_i |\mathrm{tr}(\rho E_i)|^2. \tag{9.134}$$

eg) The entanglement fidelity for the phase damping channel (phase flip)

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

is given by

$$F(\rho, \mathcal{E}) = (1-p)|\mathrm{tr}(\rho)|^2 + p|\mathrm{tr}(\rho Z)|^2 = (1-p) + p\mathrm{tr}(\rho Z)^2, \tag{9.135}$$

so we see that as p increases, the entanglement fidelity decreases.

These two entanglements, that is, ensemble fidelity and entanglement fidelity are closely related. The entanglement fidelity is a lower bound on the square of the static fidelity between the input and output to the process,

$$F(\rho, \mathcal{E}) \leq [f(\rho, \mathcal{E}(\rho))]^2. \tag{9.136}$$

Intuitively, this result states that it is harder to preserve a state plus entanglement with the outside world than it is to merely preserve the state alone.

The second property of entanglement fidelity we need to relate it to the ensemble average definition is that it is a *convex* function of ρ. Define $f(x) \equiv F(x\rho_1 + (1-x)\rho_2, \mathcal{E})$, and using (9.135)

$$f''(x) = \sum_i |\mathrm{tr}((\rho_1 - \rho_2)E_i)|^2, \tag{9.137}$$

so that $f''(x) \geq 0$.

Combing these two results we see that

$$F\left(\sum_j p_i \rho_j, \mathcal{E}\right) \leq \sum_j p_j F(\rho_j, \mathcal{E}) \tag{9.138}$$

$$\leq \sum_j p_j F(\rho_j, \mathcal{E}(\rho_j))^2, \tag{9.139}$$

and thus

$$F\left(\sum_j p_i \rho_j, \mathcal{E}\right) \leq \bar{F}. \tag{9.140}$$

Thus, any quantum channel $\mathcal{E}$ which does a good job of preserving the entanglement between a source described by a density operator ρ and a reference system will automatically do a good job of preserving an ensemble source described by probabilities p_j and stats ρ_j such that $\rho = \sum_j p_j \rho_j$. $\rightarrow$ a quantum source based on entanglement fidelity is a more stringent notion than the ensemble definition, and for this reason **the entanglement fidelity based definition in our study of quantum information theory**

Properties of the entanglement fidelity

1. $0 \leq F(\rho, \mathcal{E}) \leq 1$.
2. The entanglement fidelity is linear in the quantum operation input ($F = \langle\psi|\rho|\psi\rangle$)
3. For pure state inputs, the entanglement fidelity is equal to the static fidelity squared between input and output

$$F(|\psi\rangle, \mathcal{E}) = F(|\psi\rangle, \mathcal{E}(|\psi\rangle\langle\psi|))^2. \tag{9.141}$$

4. $F(\rho, \mathcal{E}) = 1$ iff for all pure states $|\psi\rangle$ in the support of ρ,

$$\mathcal{E}(|\psi\rangle\langle\psi|) = |\psi\rangle\langle\psi|. \tag{9.142}$$

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