Contents

1 Introduction and overview ........................................... 1
2 Introduction to quantum mechanics ............................... 3
3 Introduction to computer science ................................. 9
4 Quantum circuit ..................................................... 13
5 The quantum Fourier transform and its application ............ 15
6 Quantum search algorithms ....................................... 17
7 Quantum computers: physical realization ......................... 19
8 Quantum noise and quantum operations .......................... 21
9 Distance measures for quantum information ..................... 27
10 Quantum error-correction ....................................... 31
11 Entropy and information ....................................... 39
12 Quantum information theory .................................... 45
Chapter 1
Introduction and overview

In general a qubit’s state is a unit vector in a 2-dimensional complex vector space. The typical states $|0\rangle$ and $|1\rangle$ are known as computational basis states, and form an orthonormal basis for this vector space.

One qubit state $|\psi\rangle = \cos \frac{\theta}{2}|0\rangle + e^{i\phi}\sin \frac{\theta}{2}|1\rangle$ is represented by a point on the unit Bloch sphere with parameters $\theta$ and $\phi$, see Fig. 1.1.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{bloch_sphere.png}
\caption{Bloch sphere representation of a qubit}
\end{figure}

**John Bell**: the measurement correlations in the Bell state are stronger than could ever exist between classical systems.

Single qubit gates correspond to rotations and reflections of the Bloch sphere. For example, the Hadamard gate is just a rotation of Bloch sphere about the $\hat{y}$ axis by 90°, followed by a rotation about the $\hat{x}$ axis by 180°.

In classical circuit, NAND(negation AND) gate is universal.

**Any multiple qubit gate may be composed from CNOT and single qubit gates.**
In quantum circuit, “wire” represents the passage of “time”.

**Toffoli gate**: Toffoli gate has three input bits and three output bits. Two of the input bits are control bits that are not affected by the gate. The third one is the target bit that is flipped if both of the control bits are set to 1, and otherwise it is left alone.

Quantum parallelism: Deutsch-Jozsa algorithm

Quantum algorithms:

- Quantum Fourier transform: Deutsch-Jozsa algorithm, Shor’s algorithms for factoring and discrete logarithm (FFT on classical computer $N \log N$ for $N$ numbers, while on quantum computer $\log^2 N$)
- Grover’s quantum search algorithms (for search space size of $N$, $N$ operations on classical computer, but $\sqrt{N}$ on quantum computer)
- Quantum simulation

Quantum corollary of Moore’s law: quantum computers are keeping pace with classical computers provided a single qubit is added to the quantum computer every two years.

**Quantum state tomography** is a method for determining the quantum state of a system. If there is hidden information that is not accessible to measurement, how could we determine the quantum state by tomography?

**Quantum process tomography** is a procedure to completely characterize the dynamics of a quantum system.

**Shannon’s noiseless channel coding theorem** quantifies how many bits are required to store information being emitted by a source of information.

**Shannon’s noisy channel coding theorem** quantifies how much information can be reliably transmitted through a noisy communication channel.

**Schumacher’s noiseless channel coding theorem** quantifies the resources required to do quantum data compression, with the restriction that it be possible to recover the source with fidelity close to 1.
Life is complex - it has both real and imaginary parts.

*Zero vector* denoted by 0, satisfying $|v⟩ + 0 = |v⟩$, $∀|v⟩$ and $z0 = 0$, $∀z ∈ C$.

Any two sets of linearly independent vectors which span a vector space $V$ contain the same number of elements.

In QCQI, Hilbert space $≡$ Inner product space.

The norm of a vector $|v⟩$ is $∥|v⟩∥ = \sqrt{⟨v|v⟩}$. If $∥|v⟩∥ = 1$, $|v⟩$ is a unit vector.

**Gram-Schmidt procedure**......

**Cauchy-Schwarz inequality** $⟨v|v⟩⟨w|w⟩ ≥ |⟨v|w⟩|^2$

**Normal operator** $A$: $[A, A^†] = 0$ or $AA^† = A^†A$. Normal operators are diagonalizable—spectral decomposition, $A = \sum a|a⟩⟨a|$, then $f(A) = \sum f(a)|a⟩⟨a|$. **Hermitian operators** are automatically normal.

**Hermiticity** of positive operator. $∀|v⟩$, $⟨v|A|v⟩ ≥ 0$: A is a positive operator $⇒$ A is Hermitian and has diagonal representation $A = \sum \lambda_i|i⟩⟨i|$, $\lambda_i ≥ 0$. Moreover $∀|v⟩ ≠ 0$, $⟨v|A|v⟩ > 0$: A is positive definite.

Suppose vector spaces $V$ and $W$ have dimension $m$ and $n$ respectively. The tensor product space $V \otimes W$ has dimension $mn$.

Hadamard on $n$ qubits, $H^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{x,y=0}^{2^n-1} (-1)^{x\cdot y}|x⟩⟨y|$

$(A \otimes B)|(v⟩ \otimes |w⟩) = A|v⟩ \otimes B|w⟩$

Trace invariant under unitary similarity transformation $A → UAU^†$ $⇒ tr(UAU^†) = tr(U^†UA) = tr(A)$

$tr(A|ψ⟩⟨ψ|) = ⟨ψ|A|ψ⟩$

**Trick**: $σ_jσ_k = δ_{jk}I + i \sum_{l=1}^{3} ε_{jkl}σ_l$.

The *polar* and *singular value* decompositions are useful ways of breaking linear operators up into simpler parts.

**Polar decomposition** linear operator $A$ on vector space $V \Rightarrow A = UJ = KU$, where $U$ is unitary; $J, K$ are positive and $J = \sqrt{A^†A}, K = \sqrt{AA^†}$. $UJ$ is called left polar decomposition, $KU$ right polar decomposition.
**Singular value decomposition** square matrix $A \Rightarrow A = UDV$, where $U, V$ are unitary; $D$ is diagonal with $\geq 0$ entries.

**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 $|\psi\rangle$, which is a unit vector in the system’s state space.

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

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

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

**Postulate 3:** Quantum measurements are described by a collection of measurement operators $M_m$ acting on the state space of the system being measured. The index $m$ refers to the measurement outcomes. If the initial state is $|\psi\rangle$, then the probability that result $m$ occurs is $p(m) = \frac{\langle \psi|M_m^\dagger M_m|\psi\rangle}{\sqrt{\langle \psi|M_m^\dagger M_m|\psi\rangle}}$, and the final state is $M_m|\psi\rangle \sqrt{p(m)}$.

Non-orthogonal states can’t be reliably distinguished.

**Projective measurement or von Neumann measurement:**

A projective measurement is described by an observable $M$, a Hermitian operator on the state space of the system being observed. The observable has a spectral decomposition $M = \sum_m M_m$, where $P_m$ is the projector onto the eigenspace of $M$ with eigenvalue $m$. Upon measuring the initial state $|\psi\rangle$, the probability of getting result $m$ is $p(m) = \langle \psi|P_m|\psi\rangle$, and the final state is $P_m|\psi\rangle \sqrt{p(m)}$.

Under the condition that $M_m$ are orthogonal projectors, that is $M_m$ are Hermitian and $M_m M_{m'} = \delta_{m,m'} M_m$, Postulate 3 reduces to a projective measurement.

**PVM (Positive Operator-Valued Measure):**

Define $E_m = M_m^\dagger M_m$. $E_m$ is a positive operator such that $\sum_m E_m = 1$ and $p(m) = \langle \psi|E_m|\psi\rangle$. The complete set $E_m$ is known as a PVM. Then it is convenient to define a PVM to be any set of operators $\{E_m\}$ such that (a) each operator $E_m$ is positive, and (b) $\sum_m E_m = 1$, expressing the fact that probabilities sum to one.

**Postulate 4:** The state space of a composite physical system is the tensor product of the state spaces of the component physical systems. If the $i$th system is prepared in the state $|\psi_i\rangle$, $1 \leq i \leq n$, then the joint state of the total system is $|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$.

A quantum system whose state $|\psi\rangle$ is known exactly is said to be in a pure state.

In this case the density matrix is simply $\rho = |\psi\rangle \langle \psi|$.

Otherwise $\rho$ is in a mixed state, it is said to be a mixture of the different pure states in the ensemble for $\rho$. A pure state satisfies $tr(\rho^2) = 1$, while a mixed state $tr(\rho^2) < 1$. 

**Theorem: Characterization of density operators:** An operator $\rho$ is the density operator associated to some ensemble $p_i, |\psi_i\rangle$ iff it satisfies the conditions:

1. (Trace condition) $\text{tr}(\rho) = 1$;
2. (Positivity condition) $\rho$ is a positive operator.

**Postulates reformulated by density operator:**

**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 density operator $\rho$ (positive and $\text{tr}(\rho) = 1$) acting on the state space of the system. If a quantum system is in the state $\rho_i$ 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 a unitary transformation $U$ which depends only on the times $t_1$ and $t_2$. If the state of the system is $\rho$ at time $t_1$ and $\rho'$ at time $t_2$, then $\rho' = U \rho U^\dagger$.

**Postulate 3:** Quantum measurements are described by a collection of measurement operators $M_m$ acting on the state space of the system being measured. The index $m$ refers to the measurement outcomes. If the initial state of the system is $\rho$, then the probability that result $m$ occurs is $p(m) = \text{tr}(M_m^\dagger M_m \rho)$, and the final state is $M_m \rho M_m^\dagger$. Completeness equation $\sum_m M_m^\dagger M_m = 1$.

**Postulate 4:** The state space of a composite physical system is the tensor product of the state spaces of the component physical systems. If the $i$th system is prepared in the state $\rho_i$, $1 \leq i \leq n$, then the joint state of the total system is $\rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n$.

*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.*

**Theorem: Unitary freedom in the ensemble for density matrices**

The sets $|\tilde{\psi}_i\rangle$ and $|\tilde{\phi}_i\rangle$ generate the same density matrix ($\rho = \sum_i |\tilde{\psi}_i\rangle \langle\tilde{\psi}_i| = |\tilde{\phi}_i\rangle \langle\tilde{\phi}_i|$) iff $|\tilde{\psi}_i\rangle = \sum_j u_{ij} |\tilde{\phi}_j\rangle$, where $u_{ij}$ is a unitary matrix of complex numbers, and we ‘pad’ whichever set of vectors $|\tilde{\psi}_i\rangle$ or $|\tilde{\phi}_i\rangle$ is smaller with additional vectors 0 so that the two sets have the same number of elements.

The state of physical systems $A$ and $B$ is described by a density operator $\rho^{AB}$. The reduced density operator for system $A$ is $\rho^A = \text{tr}_B(\rho^{AB})$, where $\text{tr}_B$ is the partial trace over system $B$, defined by $\text{tr}_B(|a_1\rangle \langle a_2| \otimes |b_1\rangle \langle b_2|) \equiv |a_1\rangle \langle a_2| \text{tr}(|b_1\rangle \langle b_2|)$, where $|a_1\rangle$ and $|a_2\rangle$ are any two vectors in the state space of $A$, and $|b_1\rangle$ and $|b_2\rangle$ in the state space of $B$. 
\[ \rho = \left( \begin{array}{cc} \frac{|00\rangle + |11\rangle}{\sqrt{2}} & \frac{|0\rangle + |1\rangle}{\sqrt{2}} \\
\end{array} \right) \]

\[ \rho^1 = \text{tr}_2(\rho) = \frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2} \]

The state of the joint system of two qubits is a pure state, that is, it is known exactly; however the first qubit is in a mixed state, that is, a state about which we do not have maximal knowledge. This strange property, that the joint state of a system can be completely known, yet a subsystem (reduced density operator) be in mixed states, is a hallmark of quantum entanglement.

Measurement op.\( M \), \( m \) measurement outcomes. Before measurement in state \( |\psi\rangle \), probability of result \( m \): \( p(m) = \langle \psi | M_m^\dagger M_m | \psi \rangle \), state after measurement \( \text{Im} \)

Projective measurements (sometimes called “strong” measurement): observable \( M = \sum_m p_m P_m \), \( P_m \) projector onto eigenspace of \( M \) with eigenvalue \( m \). Before measurement in state \( |\psi\rangle \), probability of result \( m \): \( p(m) = \langle \psi | P_m | \psi \rangle \), state after measurement \( \text{Im} \)

\[ \rho = \sum_i p_i |\psi_i\rangle\langle \psi_i| \]

Schmidt decomposition Given a pure state \( |\psi\rangle \) of composite system \( AB \). There are orthonormal basis (Schmidt basis) \( |i_A\rangle \) and \( |i_B\rangle \) for systems \( A \) and \( B \) respectively, such that \( |\psi\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle \), where \( \lambda_i \) is called Schmidt coefficient satisfying \( \lambda_i \geq 0 \) and \( \sum_i \lambda_i^2 = 1 \), and the number of non-zero \( \lambda_i \) is called the Schmidt number for state \( |\psi\rangle \), which in some sense quantifies the ‘amount’ of entanglement between systems \( A \) and \( B \). Moreover \( \rho^A = \sum_i \lambda_i^2 |i_A\rangle\langle i_A| \) and \( \rho^B = \sum_i \lambda_i^2 |i_B\rangle\langle i_B| \), so the eigenvalues of \( \rho^A \) and \( \rho^B \) are identical, namely \( \lambda_i^2 \) for both density matrices.

Purification Given a state \( \rho^A \) of system \( A \). Introducing a reference system \( R \), a fictitious system without a direct physical significance, we could define a pure state \( |AR\rangle \) for the composite system \( AR \) such that \( \rho^A = \text{tr}_R |AR\rangle\langle AR| \). Suppose \( \rho^A \) has orthonormal decomposition \( \rho^A = \sum_i p_i |i_A\rangle\langle i_A| \). Introduce a system \( R \) which has the same state space as system \( A \), with orthonormal basis states \( |R\rangle \). Then \( |AR\rangle \equiv \sum_i (\sqrt{p_i} |i_A\rangle) |i_R\rangle \).

Notice the close relationship of the Schmidt decomposition to purification: the procedure used to purify a mixed system of state \( A \) 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 matrix being purified.

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 \( D \) on others, then the standard deviation \( \Delta(C) \) of the results for \( C \) times the standard deviation \( \Delta(D) \) of the results for \( D \) will satisfy the inequality:
Bell inequality is not obeyed in Quantum Mechanics. It has two assumptions 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 assumptions of local realism.
entscheidungsproblem (David Hilbert): whether or not there existed some algorithm which could be used, in principle, to solve all the problems of mathematics.

A Turing machine includes

1. **finite state control** consists of internal states $q_1, q_2, \ldots, q_m$, specially $q_s$ means starting state, $q_h$ halting state
2. **tape** 1-D object consisting of an infinite sequence of tape squares (numbered $0, 1, 2, \ldots, \infty$) extending in one direction, each square contains one of the four symbols, $0, 1, b$ (blank), $>$ (left hand edge of the tape)
3. **read-write tape-head** it points to the position on the tape which is currently being accessed
4. **program** (for 1-D tape) a finite ordered list of program lines like $\langle q, x, q', x', s \rangle$

**Operation procedure**: Turing machine starts operation when it is in state $q_s$ and the read-write tape-head is at the left hand edge $>$. In each machine cycle, Turing machine looks through the program lines searching for a line with the form $\langle q, x, q', x', s \rangle$ (initially $q = q_s, x = >$), such that the current state is $q$ and the read-write tape-head points to tape square which has symbol $x$. If there is no such line, the internal state changes to $q_h$ and the machine halts operation. If such a line is found, then the program line is executed. The internal state changes from $q$ to $q'$ and the symbol on the tape changes from $x$ to $x'$. The tape-head moves left, right, or stands still depending on whether $s = -1, +1, 0$. The program lines are executed all the way until the internal state is $q_h$. The output is the current (non-blank) contents of the tape. (This is a loose definition for output)

**Church-Turing thesis**: The class of functions computable by a Turing machine corresponds exactly to the class of functions which we would naturally regard as being computable by an algorithm.

The $O$ (‘big $O$’) notation: upper bounds on the behavior of a function, worst case behavior
If there are constants \( c \) and \( \forall n \geq n_0, f(n) \leq cg(n) \iff f(n) \) is \( O(g(n)) \) or \( f(n) \) is in the class of functions \( O(g(n)) \)

The \( \Omega \) (‘big Omega’) notation: lower bounds on the resources required

If there are constants \( c \) and \( \forall n \geq n_0, f(n) \geq cg(n) \iff f(n) \) is \( \Omega(g(n)) \)

The \( \Theta \) (‘big Theta’) notation: \( f(n) \) behaves the same as \( g(n) \) asymptotically

If \( f(n) \) is both \( O(g(n)) \) and \( \Omega(g(n)) \iff f(n) \) is \( \Theta(g(n)) \)

Computational complexity: study of space and time resources required to solve computational problems. Task: prove lower bounds on the resources required by the best possible algorithm for solving a problem.

easy, tractable or feasible \( \iff \) polynomial resources required algorithm

hard, intractable or infeasible \( \iff \) exponential resources required algorithm

**Strong Church-Turing thesis:** Any model of computation can be simulated on a probabilistic Turing machine with at most a polynomial increase in the \# of elementary operations required.

A language \( L \) over the alphabet \( \Sigma \) is a subset of the set \( \Sigma^* \) of all finite strings of symbols from \( \Sigma \)

A language \( L \) is decided by a Turing machine: an input \( x \) on the machine’s tape, if \( x \in L \), machine halts in state \( q_Y \), the machine accepted \( x \); if \( x \notin L \), machine halts in state \( q_N \), the machine rejected \( x \).

A problem is in \( \text{TIME}(f(n)) \) if there exists a Turing machine which decides whether \( x \) is in the language in time \( O(f(n)) \), \( n \) is length of \( x \). A problem is solvable in polynomial time if it is in \( \text{TIME}(n^k) \) for finite \( k \). The collection of languages which are in \( \text{TIME}(n^k) \) is denoted by \( P \).

**FACTORING** (factoring decision problem NOT in \( P \)): given a composite int. \( m \), does \( m \) have a non-trivial factor less than \( l \), \( l < m \)?

A language \( L \) is in \( NP \) if there is a Turing machine \( M \), when it started in the state \( x - \text{blank} - w \), after a time polynomial in \( |x| \), (1) if \( x \in L \), there exists a witness string \( w \) st. \( M \) halts in state \( q_Y \); (2) if \( x \notin L \), for all witness strings \( w \) st. \( M \) halts in state \( q_N \).

\( P \) is a subset of \( NP \).

The most famous open problem in CS: whether or not there are problems in \( NP \) which are not in \( P \), or \( P \neq NP \) problem
Euler’s theorem: A connected graph contains an Euler cycle iff every vertex has an even # of edges incident upon it.

Cook-Levin theorem: CSAT(circuit satisfiability) is NP-complete.

We don’t know whether $P \neq NP$ nor $P \neq PSPACE$

Landauer’s principle First form: Suppose a computer erases a single bit of information, energy dissipated into the environment at least $k_B T \ln 2$. Second form: Suppose a computer erases a single bit of information, the entropy of the environment increased by at least $k_B \ln 2$.

Fredkin gate: 3 input bits $a, b, c$, 3 output bits $a’, b’, c’$. $c$ is control bit and $c’ \equiv c$; if $c = 0, a’ = a, b’ = b$, if $c = 1, a’ = b, b’ = a (a, b \text{ swapped})$

Toffoli gate: 3 input bits $a, b, c$, $a, b$ control bits(unchanged), $c$ target bit, flipped if $a = b = 1$. $|a, b, c\rangle \rightarrow |a, b, c \oplus ab\rangle$
Chapter 4
Quantum circuit

Quantum algorithm core

The rows and columns of the unitary transforms (gates) are labeled from left to right and top to bottom as 00...0, 00...1, to 11...1 with the bottom-most wire being the least significant bit.

Hadamard gate $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

phase gate $S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$

$\pi/8$ gate (or $T$ gate) $T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$

Ids: $H = (X+Z)/\sqrt{2}$, $S = T^2$, $HXX = Z$, $HYH = -Y$, $HZH = X$

Bloch vector $r = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$

Arbitrary rotation by $\theta$ about $\hat{n}$ axis $R_\theta(\hat{n}) = e^{-i\frac{\theta}{2} \hat{n} \cdot \sigma} = \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} \hat{n} \cdot \sigma$, where real unit vector $\hat{n} = (n_x, n_y, n_z)$ and $\sigma = (X, Y, Z)$.

Z-Y decomposition for a single qubit
Suppose $U$ is a unitary operation on a qubit, there exist $\alpha, \beta, \gamma, \delta \in \mathbb{R}$, st $U = e^{i\alpha} R_z(\beta) R_y(\gamma) R_z(\delta)$. Corollary Suppose $U$ is a unitary gate on a qubit, there exist unitary ops $A, B, C$ on a single qubit, st $ABC = 1$ and $U = e^{i\delta} ABXC$, for $\alpha \in \mathbb{R}$.

CNOT (controlled-NOT) is a quantum gate with 2 input qubits, 1 control qubit and 1 target qubit. $|c\rangle|t\rangle \rightarrow |c\rangle|t \oplus c\rangle$

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

controlled-U operation $|c\rangle|t\rangle \rightarrow |c\rangle U^c|t\rangle$

controlled operation $C^n(U)$: $n$ controlled qubits on $k$ target qubits unitary operator

$C^n(U)|x_1x_2\cdots x_n\rangle|\psi\rangle = |x_1x_2\cdots x_n\rangle U^{x_1x_2\cdots x_n} |\psi\rangle$
open circle indicates conditioning on the qubit being set to 0
closed circle indicates conditioning on the qubit being set to 1

**Principle of deferred measurement:** Measurements can always be moved from an intermediate stage of a quantum circuit to the end; 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.

A unitary matrix $U$ could be decomposed into a product of two-level unitary matrices, which act non-trivially only on two-or-fewer vector comps.

An arbitrary unitary operation on $n$ qubits can be implemented using a circuit containing $O(n^24^n)$ single qubit and CNOT gates.

$U$ target unitary operator, $V$ actually implemented unitary operator, define error

$$E(U, V) \equiv \max_{\psi} \| (U - V)\psi \|, \forall \| \psi \|, \langle \psi | \psi \rangle = 1$$

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

$$E(U_mU_{m-1} \ldots U_1, V_mV_{m-1} \ldots V_1) \leq \sum_{j=1}^m E(U_j, V_j)$$

BQP is a quantum computational complexity class including decision problems that can be solved with bounded probability of error using a polynomial size quantum circuit.

One of the most significant results in quantum computational complexity is $BQP \subseteq PSIZE$

$BPP \subseteq BQP$

**Hubbard model**

$$H = \sum_{k=1}^n V_0n_kn_{\bar{k}} + \sum_{k,j \text{ neighbors}, \sigma} t_{0\sigma\bar{c}_k c_j}$$

**Ising model**

$$H = \sum_{k=1}^n \sigma_k \cdot \sigma_{k+1}$$

**Trotter formula:** $A, B$ Hermitian ops. $\forall t \in \mathbb{R}$, $\lim_{n \to \infty} \left( e^{iAt} / n e^{iBt} / n \right)^n = e^{i(A+B)t}$

$$e^{i(A+B)\Delta t} = e^{iA\Delta t} e^{iB\Delta t} + O(\Delta t^2)$$

$$e^{i(A+B)\Delta t} = e^{iA\Delta t / 2} e^{iB\Delta t / 2} e^{iA\Delta t / 2} + O(\Delta t^3)$$

**Not all unitary ops. can be efficiently approximated:** There are unitary ops. on $n$ qubits which require $\Omega \left( \frac{2^n \log(1/\epsilon)}{\log(n)} \right)$ gates to approximate to within a distance $\epsilon$ using any finite set of gates
Chapter 5
The quantum Fourier transform and its application

discrete Fourier transform \( y_k \equiv \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_j e^{2\pi i j k/N} \)

quantum Fourier transform \( |j\rangle \rightarrow \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{2\pi i j k/N} |k\rangle \leftrightarrow \sum_{k=0}^{N-1} y_k |k\rangle, \)

where \( y_k \) and \( x_j \) satisfy discrete Fourier transform above, and binary representation
\( |j\rangle = |j_1 j_2 \cdots j_n\rangle \), in which \( j = j_1 2^{n-1} + j_2 2^{n-2} + \cdots + j_n 2^0 \)

quantum Fourier transform product representation
\( |j_1, \cdots, j_n\rangle \rightarrow \frac{(|0\rangle + e^{2\pi i 0 j_1} |1\rangle) (|0\rangle + e^{2\pi i 0 j_2} |1\rangle) \cdots (|0\rangle + e^{2\pi i 0 j_n} |1\rangle)}{2^{n/2}} \)

notation \( 0_j j_{j+1} \cdots j_n \) represents binary fraction \( \frac{j_1}{2} + \frac{j_2}{4} + \cdots + \frac{j_n}{2^n} \)

\( R_k = \begin{pmatrix} 1 & 0 \\ 0 & e^{2\pi i/2^k} \end{pmatrix} \)

quantum Fourier transform \( \Theta(n^2) \), Fast Fourier Transform (FFT) \( \Theta(n2^n) \)
Chapter 6
Quantum search algorithms
Chapter 7
Quantum computers: physical realization

Four basic requirements for quantum computation:
1. Robustly represent quantum information
2. Perform a universal family of unitary transformations
3. Prepare a fiducial initial state
4. Measure the output result

\[ H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2x^2 = \hbar\omega(a^\dagger a + \frac{1}{2}), \] in which
\[ a = \frac{1}{\sqrt{2m\hbar\omega}}(m\omega x + ip) \]
\[ a^\dagger = \frac{1}{\sqrt{2m\hbar\omega}}(m\omega x - ip) \]
\[ [a, a^\dagger] = 1 \]

Encode two-qubit logical states, subscript \( L \) means logical states
\[ |00\rangle_L = |0\rangle \]
\[ |01\rangle_L = |2\rangle \]
\[ |10\rangle_L = (|4\rangle + |1\rangle)/\sqrt{2} \]
\[ |00\rangle_L = (|4\rangle - |1\rangle)/\sqrt{2}. \]

dual-rail representation: \( c_0|01\rangle + c_1|10\rangle \), in which \( |01\rangle \) represents photon (with energy \( \hbar\omega \)) in one cavity, and \( |10\rangle \) photon in the other cavity.

A laser outputs a coherent state
\[ |\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \]

Three most experimentally accessible devices for manipulating photon states:
1. mirrors

2. phase shifters: a photon propagating through a phase shifter with index of refraction \( n \) will experience a phase shift of \( e^{i(n-n_0)L\omega/c_0} \) compared to free space, in which \( n_0, c_0 \) are index of refraction and speed of light in free space respectively, and \( L \) the distance light propagated.

3. beamsplitters: reflects a fraction \( R \) of the incident light, and transmits \( 1 - R \), define \( \cos \theta = R \),

\[
\begin{align*}
a_{\text{out}} &= a_{\text{in}} \cos \theta + b_{\text{in}} \sin \theta \\
b_{\text{out}} &= -a_{\text{in}} \sin \theta + b_{\text{in}} \cos \theta
\end{align*}
\]

For nonlinear optical material, optical Kerr effect \( n(I) = n + n_2 I \), in which \( I \) is the total intensity of light through the material.

For any unitary operator \( U \), states \( \rho = 2^{-n}(1 - \varepsilon)I + \varepsilon U|00\cdots0\rangle \langle 00\cdots0|U^\dagger \) are called ‘effective pure states’ or ‘pseudopure’ states.
Chapter 8
Quantum noise and quantum operations

(A single stage) Markov process: \( q = E p \)

\[
\begin{bmatrix}
q_0 \\
q_1
\end{bmatrix} = 
\begin{bmatrix}
1 - p & p \\
p & 1 - p
\end{bmatrix} 
\begin{bmatrix}
p_0 \\
p_1
\end{bmatrix},
\]

where \( E \) is the evolution matrix, whose entries are transition probabilities. Positivity condition: all the entries of \( E \geq 0 \); Completeness condition: all the columns of \( E \) sum to 1.

Classical noise processes involving multiple stages are described as Markov processes, provided the noise is caused by independent environments.

Quantum operation formalism: quantum states transform as \( \rho' = \mathcal{E}(\rho) \), where \( \mathcal{E} \) is a quantum operation.

1. Environments and quantum operations: Assume that the principal system-environment input state is a product state, \( \rho \otimes \rho_{env} \). After transformation \( U \) the system no longer interacts with the environment, the reduced state of the system is \( \mathcal{E}(\rho) = \text{tr}_{env}[U(\rho \otimes \rho_{env})U^\dagger] \).

2. Operator-sum representation: Let \( |e_k\rangle \) be an orthonormal basis for the (finite dimensional) state space of the environment, and let \( \rho_{env} = |e_0\rangle \langle e_0| \) be the initial state of the environment.

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

\[
= \sum_k E_k \rho E_k^\dagger,
\]

where \( E_k \equiv \langle e_k|U|e_0\rangle \), known as operation element, is an operator on the state space of the principal system. This is the operator-sum representation of \( \mathcal{E} \). Completeness relation \( \sum_k E_k^\dagger E_k = 1 \).

The action of quantum operation is equivalent to taking the state \( \rho \) and randomly replacing it by \( \frac{E_k \rho E_k^\dagger}{tr(E_k \rho E_k^\dagger)} \), with probability \( tr(E_k \rho E_k^\dagger) \).

Operation elements appearing in an operator-sum representation for a quantum operation are not unique.
Generally if the principal system is initially in state \( \rho \) and the environment is initially in state \( \sigma = \sum_j q_j |j\rangle \langle j | \) and if a projective measurement \( P_m \) is performed on the joint system after the unitary interaction, the operator-sum representation for the principal system is:

\[
\mathcal{E}(\rho) = \text{tr}_E(P_m U (\rho \otimes \sigma) U^\dagger P_m) \\
= \sum_{jk} q_j \text{tr}_E(|e_k\rangle \langle e_k | P_m U (\rho \otimes |j\rangle \langle j |) U^\dagger P_m |e_k\rangle \langle e_k |)
\]

\[
= \sum_{jk} E_{jk} \rho E_{jk}^\dagger
\]

where \( |e_k\rangle \) is the orthogonal basis for the environment and \( E_{jk} = \sqrt{q_j} |e_k\rangle \langle e_k | P_m U |j\rangle \) is the operation element.

Given a set of operation elements \( \{E_k\} \), and the orthonormal basis set \( \{e_k\} \) for the environment, in one-to-one correspondence with the index \( k \) for \( E_k \). Define operator \( U |\psi\rangle |e_0\rangle = \sum_k E_k |\psi\rangle |e_k\rangle \). It is easy to verify that \( \text{tr}_E(U (\rho \otimes |e_0\rangle \langle e_0 |) U^\dagger) = \sum_k E_k \rho E_k^\dagger \).

3) Axiomatic approach to quantum operation: Define a quantum operation \( \mathcal{E} \) as a map from the set of density operators of the input space \( Q_1 \) to the set of density operators for the output space \( Q_2 \).

Axiom 1: \( \text{tr}[\mathcal{E}(\rho)] \) is the probability that the process represented by \( \mathcal{E} \) occurs, when \( \rho \) is the initial state. Thus, \( 0 \leq \text{tr}[\mathcal{E}(\rho)] \leq 1 \) for any state \( \rho \).

Axiom 2: \( \mathcal{E} \) is a convex-linear map on the set of density matrices, that is, for probabilities \( \{p_i\}, \mathcal{E}(\sum_i p_i \rho_i) = \sum_i p_i \mathcal{E}(\rho_i) \).

Axiom 3: \( \mathcal{E} \) 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 \( A \). If we introduce an extra system \( R \) of arbitrary dimensionality, \( (1 \otimes \mathcal{E})(A) \) is positive for any positive operator \( A \) on the combined system \( R Q_1 \), where \( 1 \) denotes the identity map on system \( R \).

\( \mathcal{E} \) does not necessarily preserve the trace property of density matrices, \( \text{tr}(\rho) = 1 \). Rather we define \( \text{tr}[\mathcal{E}(\rho)] \) to be the probability of measurement outcome described by \( \mathcal{E} \) occurs.

In the case when \( \forall \rho, \text{tr}[\mathcal{E}(\rho)] = 1 = \text{tr}(\rho) \), the quantum operation is a trace-preserving quantum operation, since on its own \( \mathcal{E} \) provides a complete description of the quantum process. On the other hand, if there is a \( \rho \), such that \( \text{tr}[\mathcal{E}(\rho)] < 1 \), then the quantum operation is non-trace-preserving, since on its own \( \mathcal{E} \) does not provide a complete description of the process that may occur in the system.

**Theorem 8.1:** The map \( \mathcal{E} \) satisfies axioms 1, 2 and 3 iff \( \mathcal{E}(\rho) = \sum E_i \rho E_i^\dagger \), for some set of operators \( \{E_i\} \) which map the input Hilbert space to the output Hilbert space, and \( \sum E_i^\dagger E_i \leq 1 \).

Operation elements appearing in an operator-sum representation for a quantum operation are NOT unique.

**Theorem 8.2:** (Unitary freedom in the operator-sum representation)

Suppose \( \{E_1, \cdots, E_m\} \) and \( \{F_1, \cdots, F_n\} \) are operation elements giving rise to quantum operations \( \mathcal{E} \) and \( \mathcal{F} \) respectively. By appending zero operators to the shorter
list of operation elements we may ensure that \( m = n \). Then \( \mathcal{E} = \mathcal{F} \) iff there exist an \( m \times m \) unitary matrix \( U_{ij} \) of complex number such that \( E_i = \sum_j U_{ij} F_j \).

**Theorem 8.3:** All quantum operations \( \mathcal{E} \) on a system of Hilbert space dimension \( d \) can be generated by an operator-sum representation containing at most \( d^2 \) elements, \( \mathcal{E}(\rho) = \sum_{k=1}^{M} E_k \rho E_k^\dagger \), where \( 1 \leq M \leq d^2 \).

An arbitrary state of a single qubit can always be written in the Bloch representation \( \rho = \frac{1}{2} + r \cdot \sigma \). Then arbitrary trace-preserving quantum operation is equivalent to a map of the form \( r \rightarrow r' = Mr + c \), where \( c \) is a constant vector and \( M \) is a 3 by 3 real matrix, which is a deformation of the Bloch sphere along some principal axes followed by a proper rotation, followed by a displacement \( c \).

**Bit flip channel:** with probability \( 1 - p \) the state of a qubit is flipped from \( |0 \rangle \) to \( |1 \rangle \) vice versa, \( E_0 = \sqrt{p} I \) and \( E_1 = \sqrt{1-p} X \).

**Phase flip channel:** \( E_0 = \sqrt{p} I \) and \( E_1 = \sqrt{1-p} Z \).

**Depolarizing channel:** Take a single qubit, with probability \( p \) it is depolarized, that is it is replaced by the completely mixed state, \( 1/2 \). With probability \( 1 - p \) the qubit is left untouched. \( \mathcal{E}(\rho) = p \frac{1}{2} + (1-p) \rho \).

**Amplitude damping channel:** \( \mathcal{E}_{AD}(\rho) = E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger \), where

\[
E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix},
E_1 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}.
\]

For **generalized amplitude damping**, \( \mathcal{E}_{GAD} = \sum_{i=0}^{3} E_i \rho E_i^\dagger \), where

\[
E_0 = \sqrt{p} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix},
E_1 = \sqrt{p} \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix},
E_2 = \sqrt{1-p} \begin{pmatrix} \sqrt{1-\gamma} & 0 \\ 0 & 1 \end{pmatrix},
E_3 = \sqrt{1-p} \begin{pmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{pmatrix}.
\]

Stationary state \( \rho_\infty = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix} \), satisfies \( \mathcal{E}_{GAD}(\rho_\infty) = \rho_\infty \).

**Phase damping channel:**

\[
E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\chi} \end{pmatrix},
E_1 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\chi} \end{pmatrix}.
\]
Upon unitary freedom, phase damping quantum operation is exactly the same as the phase flip channel,

\[
\tilde{E}_0 = \sqrt{\alpha} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},
\]

\[
\tilde{E}_1 = \sqrt{1 - \alpha} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},
\]

where \( \alpha = \frac{(1 + \sqrt{1 - \lambda})}{2} \).

**Master equation Lindblad form**

\[
\frac{d\rho}{dt} = \frac{1}{i\hbar} [H, \rho] + \sum_j (2L_j \rho L_j^\dagger - \{L_j^\dagger L_j, \rho\}),
\]

where \( H \) 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.

Quantum state tomography: experimentally determine unknown states.
Limitation of quantum operations formalism: 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.

Derivation of phase damping quantum operations (8.127) and (8.128). The interaction Hamiltonian is $H = \chi a^\dagger a (b + b^\dagger)$, therefore evolution unitary is $U = e^{-iHt}$. Consider oscillator $a$ as our system which could only be in state $|0\rangle$ or $|1\rangle$ and oscillator $b$ as the environment which is initially in state $|0\rangle$. Therefore the operation elements are $E_k = \langle k_a | U | 0_b \rangle$, which are $E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\chi} \end{pmatrix}$ and $E_1 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\chi} \end{pmatrix}$.

Proof. $U = e^{-i\chi t a^\dagger a (b + b^\dagger)} = e^{\theta a^\dagger a (b + b^\dagger)}$, where $\theta = -i\chi t$. Since $E_k = \langle k_b | U | 0_b \rangle$, $(E_k)_{mn} = \langle m_a k_b | U | n_a 0_b \rangle$.

- $k = 0$, $(E_0)_{mn} = \langle m_a 0_b | U | n_a 0_b \rangle$. From the properties of creation operators, $|n_a 0_b \rangle = \frac{(a^\dagger)^n}{\sqrt{n!}} |0_a 0_b \rangle$ and $|m_a 0_b \rangle = \frac{(a)^m}{\sqrt{m!}} |0_a 0_b \rangle$. We know that $U |0_a 0_b \rangle = |0_a 0_b \rangle$.

$$(E_0)_{mn} = \langle m_a 0_b | U | n_a 0_b \rangle$$

$$= \langle 0_a 0_b | \frac{a^m}{\sqrt{m!}} U \frac{(a^\dagger)^n}{\sqrt{n!}} |0_a 0_b \rangle$$

$$= \langle 0_a 0_b | \frac{a^m}{\sqrt{m!}} U (a^\dagger)^n U^\dagger |0_a 0_b \rangle$$

$$= \langle 0_a 0_b | \frac{a^m}{\sqrt{m!}} (U a^\dagger U^\dagger)^n |0_a 0_b \rangle$$

so the problem becomes what is $U a^\dagger U^\dagger$? Using Baker–Campbell–Hausdorff formula (7.27 in Nielsen and Chuang), $e^{\theta G} a^\dagger e^{-\theta G} = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} C_n$, where $G = a^\dagger a (b + b^\dagger)$. It is easy to find that $C_0 = a^\dagger$, $C_1 = [G, C_0] = a^\dagger (b + b^\dagger)$, $C_2 = [G, C_1] = a^\dagger (b + b^\dagger)^2$, $C_n = a^\dagger (b + b^\dagger)^n$.

$$U a^\dagger U^\dagger = e^{\theta G} a^\dagger e^{-\theta G} = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} a^\dagger (b + b^\dagger)^n = a^\dagger e^{\theta (b + b^\dagger)}.$$ 

So $(E_0)_{mn} = \langle 0_a 0_b | \frac{a^m}{\sqrt{m!}} e^{\theta (b + b^\dagger)} |0_a 0_b \rangle$. From Ex 8.21, we know $a^m (a^\dagger)^n |0_a \rangle = \delta_{mn} n! |0_a \rangle$. Therefore $(E_0)_{mn} = \delta_{mn} \langle 0_b | e^{\theta (b + b^\dagger)} |0_b \rangle$. So $E_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Where

$$(E_0)_{11} = \langle 0_b | e^{\theta (b + b^\dagger)} |0_b \rangle$$

$$= \langle 0_b | \sum_{n=0}^{\infty} \frac{\theta^n}{n!} (b + b^\dagger)^n |0_b \rangle$$

$$= \sum_{n=0}^{\infty} \frac{\theta^n}{n!} \langle 0_b | (b + b^\dagger)^n |0_b \rangle$$
The problem is reduced to what is $\langle 0|(b + b^\dagger)^n|0\rangle$, where the subscript of state is dropped? First of all, $\langle 0|(b + b^\dagger)^n|0\rangle = \langle 0|(b + b^\dagger)^{n/2}|(b + b^\dagger)^{-n/2}(b + b^\dagger)^{n/2}|0\rangle$, in which $n - 2\lfloor n/2 \rfloor = \begin{cases} 1 & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases}$.

Let $k = \lfloor n/2 \rfloor$, then

$$|k\rangle \equiv (b + b^\dagger)^k|0\rangle = \begin{cases} c_0|0\rangle + c_2|2\rangle + \cdots + c_{2m}|2m\rangle & \text{if } k = 2m \\ c_1|1\rangle + c_3|3\rangle + \cdots + c_{2m+1}|2m+1\rangle & \text{if } k = 2m + 1 \end{cases} \quad (8.1)$$

It is trivial to prove that $\langle k|b + b^\dagger|k\rangle = 0$. So $\langle 0|(b + b^\dagger)^n|0\rangle = \begin{cases} 0 & \text{if } n \text{ is odd} \\ 1 & \text{if } n \text{ is even} \end{cases}$.

Finally

$$(E_0)_{11} = \sum_{n \text{ even}} \frac{\theta^n}{n!} = 1 - \frac{(\chi \Delta t)^2}{2!} + \frac{(\chi \Delta t)^4}{4!} \cdots = \cos(\chi \Delta t)$$

So $E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1 - \lambda} \end{pmatrix}$, where $\lambda = 1 - \cos^2(\chi \Delta t)$.

- $k = 1$, $(E_1)_{mn} = \langle m_b|U|n_a0_b\rangle$. Following the same line as for the case $k = 0$,

$$(E_1)_{mn} = \langle m_b|U|n_a1_b\rangle = \frac{a^n}{\sqrt{m!} \sqrt{n!}} \langle Ua^\dagger U^\dagger \rangle^n|0_a0_b\rangle$$

$$= \langle 0_a1_b|a^n \rangle \frac{a^n}{\sqrt{m!} \sqrt{n!}} e^{\sqrt{\theta}(b + b^\dagger)}|0_a0_b\rangle$$

$$= \delta_{mn} \langle 1_b|e^{\sqrt{\theta}(b + b^\dagger)}|0_b\rangle$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & (E_1)_{11} \end{pmatrix}$$

where $(E_1)_{11} = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} \langle 1|(b + b^\dagger)^n|0\rangle$, where the subscript of states is dropped. Notice that $|1\rangle = (b + b^\dagger)|0\rangle$, we have

$$(E_1)_{11} = \sum_{n=0}^{\infty} \frac{\theta^n}{n!} \langle 0|(b + b^\dagger)^{n+1}|0\rangle$$

$$= \sum_{n=0}^{\infty} \frac{\theta^n}{n!}$$

$$= -i(\chi \Delta t - \frac{(\chi \Delta t)^3}{3!} + \frac{(\chi \Delta t)^5}{5!} \cdots)$$

$$= -i \sin(\chi \Delta t)$$

Upon unitary freedom, $E_1 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\lambda} \end{pmatrix}$. \qed
Chapter 9
Distance measures for quantum information

Static measures quantify how close two quantum states are, while dynamic measures quantify how well information has been preserved during a dynamic process.

**Hamming distance**: the number of places at which two bit strings are not equal.

**trace distance**: Given two probability distributions \(\{p_x\}\) and \(\{q_x\}\) over the same index set \(x\), the trace distance is defined as \(D(p_x, q_x) \equiv \frac{1}{2} \sum_x |p_x - q_x|\), aka \(L_1\) distance or Kolmogorov distance.

A metric \(D(x, y)\) must be symmetric \(D(x, y) = D(y, x)\) and satisfy the triangle inequality \(D(x, z) \leq D(x, y) + D(y, z)\).

**fidelity**: Given two probability distributions \(\{p_x\}\) and \(\{q_x\}\), the fidelity is defined as \(F(p_x, q_x) \equiv \sum_x \sqrt{p_x q_x}\).

When two two probability distributions \(\{p_x\}\) and \(\{q_x\}\) are identical, \(D(p_x, q_x) = 0\) and \(F(p_x, q_x) = 1\).

useful alternatives:

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

where the maximization is over all subsets \(S\) of the index set \(\{x\}\).

**trace distance** between quantum states \(\rho\) and \(\sigma\), \(D(\rho, \sigma) \equiv \frac{1}{2} \text{tr} |\rho - \sigma|\), where \(|A| \equiv \sqrt{A^\dagger A}\) (for positive operators two definitions are equivalent). If \(\rho\) and \(\sigma\) commute then the quantum trace distance between \(\rho\) and \(\sigma\) is equal to the classical trace distance between the eigenvalues of \(\rho\) and \(\sigma\), i.e. if \(\rho = \sum_i r_i |i\rangle \langle i|; \sigma = \sum_i s_i |i\rangle \langle i|\), thus \(D(\rho, \sigma) = D(r_i, s_i)\).

For single qubit states, \(\rho = \frac{1+\rho}{2}; \sigma = \frac{1+\sigma}{2}\), then \(D(\rho, \sigma) = \frac{|r-s|}{2}\), that is the distance between two single qubit states is equal to one half the Euclidean distance between them on the Bloch sphere.

Reasonably trace distance is preserved under unitary transformations \(D(U\rho U^\dagger, U\sigma U^\dagger) = D(\rho, \sigma)\).


\[ D(\rho, \sigma) = \max_P \text{tr}(P(\rho - \sigma)), \]

where the maximization may be taken alternatively over all projectors \( P \) or over all positive operators \( P \leq I \).

**Theorem 9.1:** Let \( \{E_m\} \) be a POVM with \( p_m = \text{tr}(\rho E_m) \) and \( q_m = \text{tr}(\sigma E_m) \) as the probabilities of obtaining a measurement outcome labeled by \( m \). Then \( D(\rho, \sigma) = \max_{\{E_m\}} D(p_m, q_m) \), where the maximization is over all POVMs \( \{E_m\} \).

**Theorem 9.2:** (Trace-preserving quantum operations are contractive) Suppose \( \mathcal{D} \) is a trace-preserving quantum operation. Let \( \rho \) and \( \sigma \) be density operators. Then \( D(\mathcal{D}(\rho), \mathcal{D}(\sigma)) \leq D(\rho, \sigma) \).

Notice that partial trace is trace-preserving operation, hence \( D(\rho^A, \sigma^A) \leq D(\rho^{AB}, \sigma^{AB}) \).

**Theorem 9.3:** (Strong convexity of the trace distance) Let \( \{p_i\} \) and \( \{q_i\} \) be probability distributions over the same index set, and \( \rho_i \) and \( \sigma_i \) be density operators also with the same indices from the same index set. Then

\[
D(\sum_i p_i \rho_i, \sum_i q_i \sigma_i) \leq D(p_i, q_i) + \sum_i p_i D(\rho_i, \sigma_i),
\]

where \( D(p_i, q_i) \) is the classical trace distance between the probability distributions \( \{p_i\} \) and \( \{q_i\} \).

**convexity of the trace distance:**

\[
D(\sum_i p_i \rho_i, \sigma) \leq \sum_i p_i D(\rho_i, \sigma)
\]

jointly convex:

\[
D(\sum_i p_i \rho_i, \sum_i p_i \sigma_i) \leq \sum_i p_i D(\rho_i, \sigma_i)
\]

Notice that any trace-preserving quantum operation \( \mathcal{D} \) has a fixed point \( \rho \) such that \( \mathcal{D}(\rho) = \rho \).

**Schauder’s fixed point theorem** Any trace-preserving quantum operation \( \mathcal{D} \) has a fixed point \( \rho \) (a density matrix), such that \( \mathcal{D}(\rho) = \rho \).

\( \mathcal{D} \) is a strictly contractive trace-preserving quantum operation such that \( D(\mathcal{D}(\rho), \mathcal{D}(\sigma)) < D(\rho, \sigma) \), for all \( \rho \) and \( \sigma \). \( \mathcal{D} \) has a unique fixed point. Suppose \( \mathcal{D} \) is a trace-preserving quantum operation and there exists a constant density operator \( \rho_0 \) and another trace-preserving quantum operation \( \mathcal{D}' \) such that \( \mathcal{D}(\rho) = p \rho_0 + (1 - p) \mathcal{D}'(\rho) \) for \( 0 < p \leq 1 \). Then \( \mathcal{D} \) is strictly contractive.

**trace distance** of states \( \rho \) and \( \sigma \) is defined to be \( F(\rho, \sigma) \equiv \text{tr}\sqrt{\rho^{1/2} \sigma \rho^{1/2}}. \) If \( \rho \) and \( \sigma \) commute then the quantum fidelity \( F(\rho, \sigma) \) reduces to the classical fidelity \( F(r, s) \) between the eigenvalue distributions \( r \) and \( s \) of \( \rho \) and \( \sigma \). That is when \( \rho \) and \( \sigma \) commute, they are simultaneously diagonal in some basis \( |i\rangle \), i.e. \( \rho = \sum_r |r_i\rangle \langle r_i|, \sigma = \sum_s |s_i\rangle \langle s_i| \), then \( F(\rho, \sigma) = \text{tr}\sqrt{\sum_r r_i s_i |i\rangle \langle i|} = \sum_i \sqrt{r_i s_i} = F(r, s) \).

The fidelity between a pure state \( |\psi\rangle \) and an arbitrary state \( \rho \) is \( F(|\psi\rangle, \rho) = \sqrt{\langle \psi | \rho | \psi \rangle}, \) that is equal to the square root of the overlap between \( |\psi\rangle \) and \( \rho \).

Invariance of fidelity under unitary transforms: \( F(U \rho U^\dagger, U \sigma U^\dagger) = F(\rho, \sigma) \), using the fact that \( \sqrt{UAU^\dagger} = U \sqrt{A} U^\dagger \), for all \( A > 0 \).

**Uhlmann’s theorem:** Suppose \( \rho \) and \( \sigma \) are states of a quantum system \( Q \). Introduce a second quantum system \( R \) which is a copy of \( Q \). Then
where the maximization is over all purifications $|\psi\rangle$ of $\rho$ and $|\phi\rangle$ of $\sigma$ into $RQ$.

**Lemma:** Let $A$ be any operator, and $U$ unitary. Then $|\text{tr}(AU)| \leq |\text{tr}|A|$, with equality being attained by choosing $U = V^\dagger$, where $A = |A|V$ is the polar decomposition of $A$.

From Uhlmann’s theorem, the fidelity is symmetric in its inputs $F(\rho, \sigma) = F(\sigma, \rho)$, and $0 \leq F(\rho, \sigma) \leq 1$, with equality in the first inequality iff $\rho$ and $\sigma$ have orthogonal support, and equality in the second inequality iff $\rho = \sigma$.

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

where the minimum is over all POVMs $\{E_m\}$ and $p_m = \text{tr}(\rho E_m), q_m = \text{tr}(\sigma E_m)$. Turing fidelity into a metric, define the angle between states $\rho$ and $\sigma$ to be $A(\rho, \sigma) = \arccos F(\rho, \sigma)$, which satisfies the triangle inequality $A(\rho, \tau) \leq A(\rho, \sigma) + A(\sigma, \tau)$. *Contractivity of the angle:* let $\mathcal{E}$ be a trace-preserving quantum operation, then $A(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq A(\rho, \sigma)$

*Monotonicity of the fidelity:* Suppose $\mathcal{E}$ is a trace-preserving quantum operation. Let $\rho$ and $\sigma$ be density operators. Then

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

**Strong concavity of the fidelity:** Let $p_i$ and $q_i$ be probability distributions over the same index set, and $\rho_i$ and $\sigma_i$ density operations also indexed by the same index set. Then

$$F(\sum_i p_i \rho_i, \sum_i q_i \sigma_i) \geq \sum_i \sqrt{p_i q_i} F(\rho_i, \sigma_i)$$

*Joint concavity of fidelity:*

$$F(\sum_i p_i \rho_i, \sum_i p_i \sigma_i) \geq \sum_i p_i F(\rho_i, \sigma_i)$$

*Concavity of fidelity:*

$$F(\sum_i p_i \rho_i, \sigma) \geq \sum_i p_i F(\rho_i, \sigma)$$

For any two pure states $|a\rangle$ and $|b\rangle$, $D(|a\rangle, |b\rangle) = \sqrt{1 - F^2(|a\rangle, |b\rangle)}$.

Let $\rho$ and $\sigma$ be any two quantum states, $1 - F(\rho, \sigma) \leq D(\rho, \sigma) \leq \sqrt{1 - F^2(\rho, \sigma)}$, which implies that trace distance and the fidelity are qualitatively equivalent measures of closeness for quantum states.

**How well does a quantum channel preserve information?**

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

minimizing over all possible initial states which we don’t know beforehand.
To implement a unitary quantum gate $U$, using a trace-preserving quantum operation $E$ which will include the effect of noise, gate fidelity

$$F(U, E) = \min_\psi \langle \psi | U \psi, E(\psi \langle \psi |)$$

Two possible quantum definitions for the notion of an information source:

1. *ensemble average fidelity*: $\bar{F} = \sum_j p_j F^2(\rho_j, E(\rho_j))$, where $p_j$ are the respective probabilities for the different possible preparations of the system $\rho_j$.

2. *entanglement fidelity*: A quantum system $Q$ is prepared in a state $\rho$ which is entangled with the external world. We represent the entanglement by introducing a fictitious system $R$ such that the joint state of $RQ$ is a pure state. $F(\rho, E) \equiv F^2(RQ, RQ') = \langle RQ | (I_R \otimes \rho') (RQ) | RQ \rangle$, where prime indicates the state of a system after quantum operation has been applied. **Note that the entanglement fidelity depends only upon $\rho$ and $E$ and not upon the details of the purification $|RQ\rangle$**. The entanglement fidelity provides a measure of how well the entanglement between $R$ and $Q$ is preserved by the process $E$ with values close to 1 indicating that the entanglement has been well preserved, and values close to 0 indicating that most of the entanglement has been destroyed.

Suppose $E_i$ is a set of operation elements for a quantum operation $E$.

$$F(\rho, E) = \sum_i |\langle RQ | E_i | RQ \rangle|^2 = \sum_i |\text{tr}(\rho E_i)|^2$$

And $0 \leq F(\rho, E) \leq 1$

Convexity of entanglement fidelity $F(\sum_j p_j \rho_j, E) \leq \sum_j p_j F(\rho_j, E) \leq \sum_j p_j F^2(\rho_j, E(\rho_j))$, that is

$$F \left( \sum_j p_j \rho_j, E \right) \leq \bar{F}$$
Chapter 10
Quantum error-correction

The key idea of classical error-correcting codes is always to encode messages by adding enough redundancy.

![Binary symmetric channel](image)

Fig. 10.1: *binary symmetric channel*

Potential difficulties from classical error-correcting codes to quantum error-correcting codes (all solved):

- No cloning
- Errors are continuous
- Measurement destroys quantum information

Two-stage quantum error-correction procedure:

1. error-detection or syndrome diagnosis
2. recovery

**bit flip code**

\[ a|0\rangle + b|1\rangle \rightarrow a|000\rangle + b|111\rangle \rightarrow \text{bit flip channel} \rightarrow \text{error syndrome} \]

\[
P_0 = |000\rangle \langle 000| + |111\rangle \langle 111| \text{ no error} \\
P_1 = |100\rangle \langle 100| + |011\rangle \langle 011| \text{ bit flip on qubit one} \\
P_2 = |010\rangle \langle 010| + |101\rangle \langle 101| \text{ bit flip on qubit two} \\
P_3 = |001\rangle \langle 001| + |110\rangle \langle 110| \text{ bit flip on qubit three} \\
\]

31
→ recovery using bit flip operator X on corresponding qubit

syndrome measurement $Z_1Z_2$ and $Z_2Z_3$

<table>
<thead>
<tr>
<th>$Z_1Z_2$</th>
<th>post state of $Z_1Z_2$</th>
<th>$Z_2Z_3$</th>
<th>post state of $Z_2Z_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a(000) + b(111)$</td>
<td>+1</td>
<td>$a(000) + b(111)$</td>
<td>+1</td>
</tr>
<tr>
<td>$a(100) + b(011)$</td>
<td>-1</td>
<td>$-(a(100) + b(011))$</td>
<td>+1</td>
</tr>
<tr>
<td>$a(010) + b(101)$</td>
<td>-1</td>
<td>$-(a(010) + b(101))$</td>
<td>-1</td>
</tr>
<tr>
<td>$a(001) + b(110)$</td>
<td>+1</td>
<td>$a(001) + b(110)$</td>
<td>-1</td>
</tr>
<tr>
<td>$a(110) + b(001)$</td>
<td>+1</td>
<td>$a(110) + b(001)$</td>
<td>-1</td>
</tr>
<tr>
<td>$a(101) + b(010)$</td>
<td>-1</td>
<td>$-(a(101) + b(010))$</td>
<td>-1</td>
</tr>
<tr>
<td>$a(011) + b(100)$</td>
<td>-1</td>
<td>$-(a(011) + b(100))$</td>
<td>+1</td>
</tr>
<tr>
<td>$a(111) + b(000)$</td>
<td>+1</td>
<td>$a(111) + b(000)$</td>
<td>+1</td>
</tr>
</tbody>
</table>

**phase flip code**

Notice $H|0\rangle = |+\rangle$ and $H|1\rangle = |-\rangle$

$a(0) + b(1) \rightarrow a|++\rangle + b|--\rangle \rightarrow \text{phase flip channel} \rightarrow \text{error syndrome} P_j^f = H \otimes P_j H \otimes 3$

- $P_0^f \equiv |000\rangle \langle 000| + |111\rangle \langle 111|$ no error
- $P_1^f \equiv |100\rangle \langle 100| + |011\rangle \langle 011|$ bit flip on qubit one
- $P_2^f \equiv |010\rangle \langle 010| + |101\rangle \langle 101|$ bit flip on qubit two
- $P_3^f \equiv |001\rangle \langle 001| + |110\rangle \langle 110|$ bit flip on qubit three

→ recovery using phase flip operator $HXH = Z_i$ on corresponding qubit

syndrome measurement $H \otimes Z_1Z_2H \otimes 3 = X_1X_2$ and $H \otimes Z_2Z_3H \otimes 3 = X_2X_3$

bit flip channel and phase flip channel are unitarily equivalent.

**Shor code** can protect against the effects of an arbitrary error on a single qubit!

A fundamental fact about quantum error-correction is that by correcting just a discrete subset of errors, bit flip, phase flip, and combined bit-phase flip, a quantum error-correcting code is able to automatically correct an apparently much larger (continuous!) class of errors.

Quantum states are encoded by a unitary operation into a quantum error-correcting code space, a subspace $C$ of some larger Hilbert space. The projector onto the code space is denoted by $P$. Assumptions for theory of quantum error-correction:

1. the noise is described by a quantum operation $\mathcal{E}$
2. the complete error-correction procedure is effected by a trace-preserving error-correction operation $\mathcal{R}$

We require for any state $\rho$ whose support lies in the code $C$, $(\mathcal{R} \circ \mathcal{E})(\rho) \propto \rho$
Quantum error-correction conditions: Let $C$ be a quantum code, and let $P$ be the projector onto $C$. Suppose $\mathcal{E}$ is a quantum operation with operation elements $\{E_i\}$. A necessary and sufficient condition for the existence of an error-correction operation $\mathcal{R}$ correcting $\mathcal{E}$ on $C$ is that $PE_i^\dagger E_j^\dagger P = \alpha_{ij}P$, for some Hermitian matrix $\alpha$ of complex numbers.

$$d = u^\dagger \alpha u \rightarrow F_j \equiv u_j E_k \rightarrow PF_k^\dagger F_j P = d_{kj}P \rightarrow F_k P = U_k \sqrt{PF_k^\dagger F_k P} = \sqrt{u_{kk}}U_k P$$

Theorem 10.2: Suppose $C$ is a quantum code and $\mathcal{R}$ is the error-correction operation to recover from a noise process $\mathcal{E}$ with operation elements $\{E_i\}$. Suppose $\{F_j\}$ is a quantum operation with operation elements $\{F_j\}$ which are linear combinations of the $E_i$, that is $F_j = \sum_i m_{ij} E_i$ for some matrix $m_{ij}$ of complex numbers. Then the error-correction operation $\mathcal{R}$ also corrects for the effects of the noise process $\mathcal{F}$ on the code $C$.

Quantum Hamming bound (only applies to non-degenerate codes)

$$\sum_{j=0}^t \binom{n}{j} 3^j 2^k \leq 2^n,$$

where a non-degenerate code is used to encode $k$ qubits in $n$ qubits, such that it can correct errors on any subset of $t$ or fewer qubits. Suppose $j$ errors occur, $j \leq t$.

Classical linear code

A linear code $C$ encoding $k$ bits of information into an $n$ bit code space is specified by an $n$ by $k$ generator matrix $G$ whose entries are all elements of $\mathbb{Z}_2$, zeroes and ones. $y = Gx$, where $x$ is the message and $y$ is the encoded message or the codeword, and the columns of $G$ are linearly independent. Using Gaussian elimination and swapping of bits, generator matrix has the standard form $G = [I_k - A]$, where $A$ is some $(n-k) \times k$ matrix.

This code using $n$ bits to encode $k$ bits of information is called an $[n,k]$ code, simply multiplying the $k$-bit message by the $n \times k$ generator matrix $G$ to obtain the $n$-bit encoded message, a procedure can be done using $O(nk)$ operations.

In an $[n,k]$ code, for all $n$-element vectors $y$ over $\mathbb{Z}_2$, $Hy = 0$ such that $H$ is an $n-k$ by $n$ matrix known as the parity check matrix, with entries all zeroes and ones. In other words, the code $y$ is the kernel of $H$. $H$ is equivalent to $G$ and $HG = 0$. 

...
Correspondingly $H$ has linearly independent rows. Again using Gaussian elimination and swapping of bits, parity check matrix has the standard form $H = [A | I_{n-k}]$, where $A$ is the same $(n-k) \times k$ matrix as in the standard form of $G$.

To go from the parity check matrix to the generator matrix, pick $k$ linearly independent vectors $y_1, \ldots, y_k$ spanning the kernel of $H$, and set $G$ to have columns $y_1$ through $y_k$. To go from the generator matrix to the parity check matrix, pick $n-k$ linearly independent vectors $y_1, \ldots, y_{n-k}$ orthogonal to the columns of $G$, and set the rows of $H$ to be $y_1^T, \ldots, y_{n-k}^T$.

Suppose message $x$ is encoded as $y = Gx$. But some error $e$ due to noise corrupts $y$ to the corrupted codeword $y' = y + e$. Since $Hy = 0$ for all codeword $y$, $Hy' = He$ is called the error syndrome.

The (Hamming) distance $d(x,y)$ between $x$ and $y$ is defined to be the number of places at which $x$ and $y$ differ. The (Hamming) weight of a word $x$ is defined to be the distance between $x$ and the all zero string, $wt(x) = d(x,0)$. Note that $d(x,y) = wt(x+y)$.

Define the distance of a code $C$ to be the minimum distance between any two codewords $d(C) \equiv \min_{x,y \in C, x \neq y} d(x,y)$. Use $d(x,y) = wt(x+y)$ and note that for linear code $C$, if $x$ and $y$ are code words then $x+y$ is also a code word. Hence $d(C) = \min_{x \in C, x \neq 0} wt(x)$. Setting $d \equiv d(C)$, $C$ is an $[n,k,d]$ code.

The importance of distance is that a code with distance $2t+1$ for some integer $t$ is able to correct errors on up to $t$ bits.

**Singleton bound**: An $[n,k,d]$ code must satisfy $n-k \geq d-1$.

**Hamming code** has parity check matrix $H$ whose columns are all $2^r-1$ bit strings of length $r$ which are not identically 0, where $r \geq 2$. All Hamming codes have distance 3, hence it is a $[2^r-1, 2^r-r-1, 3]$ linear code.

**Gilbert-Varshamov bound**: For large $n$ there exists an $[n,k]$ error-correcting code protecting against errors on $t$ bits for some $k$ s.t.

$$\frac{k}{n} \geq 1 - H\left(\frac{2t}{n}\right)$$

where $H(x) \equiv -x \log(x) - (1-x) \log(1-x)$ is the binary Shannon entropy.

**Quantum Gilbert-Varshamov bound**: In the limit as $n$ becomes large, an $[n,k]$ quantum code protecting against errors on up to $t$ qubits exists for some $k$ s.t.

$$\frac{k}{n} \geq 1 - 2H\left(\frac{2t}{n}\right).$$
Suppose $C$ is an $[n, k]$ code with generator $G$ and parity check matrix $H$. Then we define the dual of $C$, denoted $C^\perp$, to be the code with generator matrix $H^T$ and parity check matrix $G^T$. $C^\perp$ consists of all codewords $y$ such that $y$ is orthogonal to all the codewords in $C$. A code is weakly self-dual if $C \subseteq C^\perp$ and strictly self-dual if $C = C^\perp$.

**Calderbank-Shor-Steane** CSS code: Suppose $C_1$ and $C_2$ are $[n, k_1]$ and $[n, k_2]$ classical linear codes, respectively, such that $C_2 \subset C_1$, and both $C_1$ and $C_2^\perp$ can correct on up to $t$ bits. Then $\text{CSS}(C_1, C_2)$ (pronounced as CSS code of $C_1$ over $C_2$) is an $[n, k_1 - k_2]$ quantum error-correcting code which can correct arbitrary errors on up to $t$ qubits. Furthermore, the error-detection and correction steps require only the application of Hadamard and controlled-NOT gates, in each case a number linear in the size of the code. Encoding and decoding can also be performed using a number of gates linear in the size of the code.

Suppose $x \in C_1$, define quantum state $|x + C_2\rangle \equiv \frac{1}{\sqrt{|C_2|}} \sum_{y \in C_2} |x + y\rangle$. Only when $x$ is in the coset of $C_2$, $|x + C_2\rangle$ is different. Hence the dimension of $\text{CSS}(C_1, C_2)$ is $|C_1|/[C_2] = 2^{k_1-k_2}$.

**Stabilizer** codes, sometimes known as additive quantum codes, is analogous to classical linear codes.

The general Pauli group on $n$ qubits $G_n$ is defined to consist of all $n$-fold tensor products of Pauli matrices with multiplicative factors $\pm 1, \pm i$. Suppose $S$ is a subgroup of $G_n$ and define $V_S$ to be the set of $n$ qubit states which are fixed by every element of $S$, such that $\forall |\psi\rangle \in V_S$, $s \in S$ satisfies $s|\psi\rangle = |\psi\rangle$. $V_S$ is the vector space stabilized by $S$ and $S$ is the stabilizer of the space $V_S$. $S$ satisfies two conditions: (a) the elements of $S$ commute and (b) $-\mathbf{1} \notin S$.

A set of elements $g_1, \cdots, g_l$ in a group $G$ generates the group if every element in $G$ can be written as a product of elements from the set. $G \equiv \langle g_1, \cdots, g_l \rangle$. A group $G$ of size $|G|$ has a set of at most $\log |G|$ generators. Check matrix, a $l \times 2n$ matrix, nicely represents the $l$ generators. If we use $r(g)$ to denote the $2n$-dimensional row vector representation of an element $g$ of the Pauli group and define a $2n \times 2n$ matrix $\Lambda = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$, elements $g$ and $g'$ of the Pauli group commute if and only if $r(g)\Lambda r(g')^T = 0$.

**Proposition 10.3:**

Let $S = \langle g_1, \cdots, g_l \rangle$ be s.t. $-\mathbf{1}$ is not an element of $S$. The generators $g_1, \cdots, g_l$ are independent iff the rows of the corresponding check matrix are linearly independent.
**Proposition 10.4:**
Let $S = \langle g_1, \cdots, g_l \rangle$ be generated by $l$ independent generators and satisfy $-1 \notin S$. Fix $i$ in the range $1, \cdots, l$. Then there exists $g \in G_n$ s.t. $gg_i g^\dagger = -g_i$ and $gg_j g^\dagger = g_j, \forall j \neq i$.

**Proposition 10.5:**
Let $S = \langle g_1, \cdots, g_{n-k} \rangle$ be generated by $n-k$ independent and commuting elements from $G_n$ s.t. $-1 \notin S$. Then $V_S$ is a $2^k$-dimensional vector space.

If the vector space $V_S$ is stabilized by group $S = \langle g_1, \cdots, g_l \rangle$. For some unitary operation $U$, the vector space $UV_S$ is stabilized by the group $USU^\dagger \equiv \{ UgU^\dagger | g \in S \}$ and $USU^\dagger = \langle Ug_1 U^\dagger, \cdots, Ug_l U^\dagger \rangle$.

CNOT gate and Hadamard gate is capable of generating entanglement.

The set of unitary operators $U$ such that $UG_n U^\dagger = G_n$ is the normalizer of $G_n$, and denote it by $N(G_n)$. The normalizer of $G_n$ is generated by the Hadamard, phase, and CNOT gates. That is any unitary operation taking element of $G_n$ to elements of $G_n$ under conjugation can be composed from the Hadamard, phase, and CNOT gates, which are sometimes referred to as the normalizer gates. CHP, Clifford group.

The $\pi/8$ and Toffoli gates are NOT in the normalizer.

**Theorem 10.6:** Suppose $U$ is any unitary operator on $n$ qubits with the property that if $g \in G_n$, then $UgU^\dagger \in G_n$. Then up to a global phase $U$ may be composed from $O(n^2)$ Hadamard, phase and CNOT gates.

**Measurement in stabilizer code:** Some measurement operator $g$ is a tensor product of Pauli matrices without multiplicative factors,

1. If $g$ commutes with all the generators $\langle g_1, \cdots, g_n \rangle$ of the stabilizer, the measurement leaves stabilizers invariant.
2. If $g$ anti-commutes with one or more of the generators, we could always make $g$ only anti-commute with one generators and commute with all others. Say $g$ anti-commutes with $g_1$ and $g_2$, then by replacing $g_2$ with $g_1 g_2$, $g$ will only anti-commute with $g_1$ and commute with all other generators. Then if the measurement result for the projectors $\frac{I \pm g_2}{\sqrt{2}}$ is $+1$, the stabilizer becomes $\langle g, g_2, \cdots, g_n \rangle$; if the measurement result is $-1$, the stabilizer becomes $\langle g, g_2, \cdots, g_n \rangle$.  

**Gottesman-Knill theorem:** Suppose a quantum computation is performed only involving: state preparations in the computational basis, $H$ gates, phase gates, CNOT gates, Pauli gates, and measurements of observables in the Pauli group (which includes measurements in the computational basis as a special case), together with the possibility of classical control conditioned on the outcome of such measurements. Such a computation may be efficiently simulated on a classical computer.
Quantum teleportation and superdense coding can be simulated efficiently on a classical computer.

An \([n,k]\) stabilizer code is defined to be the vector space \(V_S\) stabilized by a subgroup \(S\) of \(G_n\) such that \(-I \notin S\) and \(S\) has \(n-k\) independent and commuting generators, \(S = \langle g_1, \cdots, g_{n-k} \rangle\). This code is denoted as \(C(S)\). We choose operators \(\bar{Z}_1, \cdots, \bar{Z}_k \in G_n\) such that \(g_1, \cdots, g_{n-k}, \bar{Z}_1, \cdots, \bar{Z}_k\) forms an independent and commuting set. The operators \(\bar{Z}_1, \cdots, \bar{Z}_k\) play the role of logical Pauli \(z\) operators. We could also choose operators \(\bar{X}_1, \cdots, \bar{X}_k \in G_n\) such that \(\bar{X}_i\) commutes with all \(g_j\) and \(\bar{Z}_k\) except \(\bar{Z}_i\). The operators \(\bar{X}_1, \cdots, \bar{X}_k\) play the role of logical Pauli \(x\) operators.

The centralizer \(Z(S)\) of stabilizer \(S\) in \(G_n\) is the set of \(e \in G_n\) such that \(eg = ge\) for all \(g \in S\). For stabilizer groups \(S\) of concern to us, normalizer of \(S\) is identical to centralizer of \(S\), that is \(N(S) = Z(S)\).

**Error-correction conditions for stabilizer codes:** Let \(S\) be the stabilizer for a stabilizer code \(C(S)\). Suppose \(\{E_j\}\) is a set of operators in \(G_n\) st. \(E_j^TE_k \notin N(S) - S\) for all \(j\) and \(k\). Then \(\{E_j\}\) is a correctable set of errors for the code \(C(S)\).

The weight of an operator in \(G_n\) is defined as the number of terms in the tensor product which are not equal to identity. The distance of a stabilizer code \(C(S)\) is defined to be the minimum weight of an element in \(N(S) - S\). If \(C(S)\) is an \([n,k]\) code with distance \(d\), then \(C(S)\) is an \([n,k,d]\) stabilizer code.

<table>
<thead>
<tr>
<th>Name</th>
<th>Operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>(g_1)</td>
<td>XZZXZ</td>
</tr>
<tr>
<td>(g_2)</td>
<td>IXZZX</td>
</tr>
<tr>
<td>(g_3)</td>
<td>XXZZZ</td>
</tr>
<tr>
<td>(g_4)</td>
<td>ZXIXZ</td>
</tr>
<tr>
<td>(\bar{Z})</td>
<td>ZZZZZ</td>
</tr>
<tr>
<td>(\bar{X})</td>
<td>XXXXX</td>
</tr>
</tbody>
</table>

Table 10.1: The four generators for the five qubit code, and the logical \(Z\) and logical \(X\) operators.

The basic idea of fault-tolerant quantum computation is to compute directly on encoded quantum states in such a manner that decoding is never required.

A universal set of logical operations: Hadamard, phase, CNOT and \(\pi/8\) gates.

We define the fault-tolerance of a procedure to be the property that if only one component in the procedure fails then the failure causes at most one error in each encoded block of qubits output from the procedure.
Table 10.2: The six generators for the Steane seven qubit code, and the logical $Z$ and logical $X$ operators.

<table>
<thead>
<tr>
<th>Name</th>
<th>Operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_1$</td>
<td>IIIXXX</td>
</tr>
<tr>
<td>$g_2$</td>
<td>IXXIXX</td>
</tr>
<tr>
<td>$g_3$</td>
<td>XIXIXX</td>
</tr>
<tr>
<td>$g_4$</td>
<td>IIIZZZ</td>
</tr>
<tr>
<td>$g_5$</td>
<td>IZZIIIZ</td>
</tr>
<tr>
<td>$g_6$</td>
<td>ZIZIZIZ</td>
</tr>
<tr>
<td>$\bar{Z}$</td>
<td>ZZZZZZZ</td>
</tr>
<tr>
<td>$\bar{X}$</td>
<td>XXXXXXX</td>
</tr>
</tbody>
</table>

Table 10.3: The eight generators for the Shor nine qubit code, and the logical $Z$ and logical $X$ operators.

<table>
<thead>
<tr>
<th>Name</th>
<th>Operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_1$</td>
<td>ZZIIIIIII</td>
</tr>
<tr>
<td>$g_2$</td>
<td>IZZIIIIII</td>
</tr>
<tr>
<td>$g_3$</td>
<td>IIIZZIIII</td>
</tr>
<tr>
<td>$g_4$</td>
<td>IIIZZZIIII</td>
</tr>
<tr>
<td>$g_5$</td>
<td>IIIIZZZI</td>
</tr>
<tr>
<td>$g_6$</td>
<td>IIIIIIZZ</td>
</tr>
<tr>
<td>$g_7$</td>
<td>XXXXXXXIII</td>
</tr>
<tr>
<td>$g_8$</td>
<td>IIIXXXXXX</td>
</tr>
<tr>
<td>$\bar{Z}$</td>
<td>XXXXXXXXXXX</td>
</tr>
<tr>
<td>$\bar{X}$</td>
<td>ZZZZZZZZZ</td>
</tr>
</tbody>
</table>

Threshold theorem for quantum computation:

A quantum circuit containing $p(n)$ gates may be simulated with probability of error at most $\varepsilon$ using

$$O(poly(\log p(n)/\varepsilon)p(n))$$

gates on hardware whose components fail with probability at most $p$, provided $p$ is below some constant threshold, $p < p_{th}$, and given reasonable assumptions about the noise in the underlying hardware. $p_{th} \equiv 1/c$ s.t. there exists $k$ to satisfy

$$\frac{(cp)^k}{c} \leq \frac{\varepsilon}{p(n)}.$$
Chapter 11
Entropy and information

Entropy is a key concept of quantum information theory. It measures how much uncertainty there is in the state of a physical system. Suppose we learn the value of a random variable $X$. We can view the entropy either as a measure of our uncertainty before we learn the value of $X$, or as a measure of how much information we have gained after we learn the value of $X$.

The Shannon entropy associated with probability distribution $p_1, p_2, \ldots, p_n$ is defined by

$$H(X) \equiv H(p_1, \ldots, p_n) \equiv -\sum_x p_x \log p_x.$$ 

By convention $0\log 0 \equiv 0$. Note ‘$\log$’ indicates logarithm of base 2, while ‘$\ln$’ indicates a natural logarithm. With this convention for the logarithm entropies are measured in ‘bits’.

The entropy is used to quantify the resources needed to store information. For Huffman code, the entropy quantifies the optimal compression that may be achieved. Fundamental measures of information arise as the answers to fundamental questions about the physical resources required to solve some information processing problem.

**Binary entropy** is defined as $H_{\text{bin}}(p) \equiv -p \log p - (1-p) \log (1-p)$, where $p$ and $1-p$ are the probabilities of the two outcomes. Notice that $H(p) = H(1-p)$ and that $H(p)$ attains its maximum value of 1 at $p = 1/2$.

A real-valued function $f$ is concave if for any $p$ in the range 0 to 1, $f(px + (1-p)y) \geq pf(x) + (1-p)f(y)$.

Suppose the quantum system is prepared in a state $|\psi\rangle$, and let $p(c)$ be the probability distribution associated with a measurement of $C = \sum_c c\langle c|c\rangle$, with associated entropy $H(C)$, and $q(d)$ the probability distribution associated with a measurement of $D = \sum_d |d\rangle\langle d|$, with associated entropy $H(D)$. The entropic uncertainty princi-
ple states that
\[ H(C) + H(D) \geq 2 \log \frac{1}{f(C, D)}, \]
where \( f(C, D) \equiv \max_{c, d} |\langle c | d \rangle| \). A weaker form is
\[ H(C) + H(D) \geq -2 \log \frac{1 + f(C, D)}{2}. \]

**Relative entropy** of \( p(x) \) to \( q(x) \) is defined as
\[ H(p(x) \parallel q(x)) \equiv \sum_x p(x) \log \frac{p(x)}{q(x)} = -H(X) - \sum_x p(x) \log p(x), \]
where \( p(x) \) and \( q(x) \) are two probability distributions over the same index set \( x \). We define \(-0 \log 0 \equiv 0\), and \(-p(x) \log 0 \equiv +\infty\) if \( p(x) > 0\).

A very useful inequality \( \ln x \leq \ln 1 + x - 1 \).

**Non-negativity of the relative entropy**: The relative entropy is non-negative, \( H(p(x) \parallel q(x)) \geq 0 \), with equality iff \( p(x) = q(x) \), \( \forall x \).

**Theorem**: Suppose \( X \) is a random variable with \( d \) outcomes. Then \( H(X) \leq \log d \), with equality iff \( X \) is uniformly distributed over those \( d \) outcomes.

Finding expressions for entropic quantities in terms of the relative entropy is often used in the study of both classical and quantum entropies.

The **joint entropy** of a pair of random variables \( X \) and \( Y \) is defined as
\[ H(X, Y) \equiv -\sum_{x, y} p(x, y) \log p(x, y). \]

The **entropy of \( X \) conditional on knowing \( Y \)** is defined as
\[ H(X | Y) \equiv H(X, Y) - H(Y). \]
The conditional entropy is a measure of how uncertain we are, on average, about the value of \( X \), given that we know the value of \( Y \).

The **mutual information** of \( X \) and \( Y \) is defined as
\[ H(X : Y) \equiv H(X) + H(Y) - H(X, Y). \]
Or \( H(X : Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \), measures how much information \( X \) and \( Y \) have in common.

**Basic properties of Shannon entropy:**
1. \( H(X, Y) = H(Y, X) = H(Y : X) \).
2. \( H(Y|X) \geq 0 \) and thus \( H(X : Y) \leq H(Y) \), with equality iff \( Y \) is a function of \( X \), \( Y = f(X) \).
3. \( H(X) \leq H(X : Y) \leq H(Y) \), with equality iff \( Y \) is a function of \( X \).
4. **Subadditivity:** \( H(X, Y) \leq H(X) + H(Y) \) with equality iff \( X \) and \( Y \) are independent random variables.
5. \( H(Y|X) \leq H(Y) \) and thus \( H(X : Y) \geq 0 \), with equality iff \( X \) and \( Y \) are independent random variables.
6. **Strong subadditivity:** \( H(X, Y, Z) + H(Y) \leq H(X, Y) + H(Y, Z) \), with equality iff \( Z \rightarrow Y \rightarrow X \) forms a Markov chain (defined below).
7. **Conditioning reduces entropy:** \( H(X|Y, Z) \leq H(X|Y) \).

Notice that \( H(X : Y) = H(p(x,y) \parallel p(x)p(y)) \) and \( H(Y|X) = - \sum_{x,y} p(x, y) \log p(y|x) \).

![Entropy Venn diagram](image)

**Chaining rule for conditional entropies:** Let \( X_1, \ldots, X_n \) and \( Y \) be any set of random variables. Then
\[
H(X_1, \ldots, X_n|Y) = \sum_{i=1}^{n} H(X_i|Y, X_{i-1}, \ldots, X_1).
\]

**Markov chain** is a sequence of random variables \( X_1 \rightarrow X_2 \rightarrow \cdots \) such that \( X_{n+1} \) is independent of \( X_1, \ldots, X_{n-1} \), given \( X_n \). In other words,
\[ p(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_1 = x_1) = p(X_{n+1} = x_{n+1} | X_n = x_n). \]

**Data processing inequality:** Suppose \( X \rightarrow Y \rightarrow Z \) is a Markov chain, then

\[ H(X) \geq H(X : Y) \geq H(X : Z). \]

The data processing inequality states that information about the output of a source can only decrease with time, that is once information has been lost, it is gone forever.

If \( X \rightarrow Y \rightarrow Z \) is a Markov chain, then so is \( Z \rightarrow Y \rightarrow X \). **Data pipelining inequality:**

\[ H(Z : Y) \geq H(Z : X). \]

**Von Neumann entropy** of a quantum state \( \rho \) is defined as \( S(\rho) \equiv - \text{tr}(\rho \log \rho) \).

If \( \lambda_x \) are the eigenvalues of \( \rho \) then \( S(\rho) = - \sum_x \lambda_x \log \lambda_x \). Again \( 0 \log 0 = 0 \).

The **relative entropy** of \( \rho \) to \( \sigma \) is defined by \( S(\rho \| \sigma) \equiv \text{tr}(\rho \log \rho) - \text{tr}(\rho \log \sigma) \).

**Fannes’ inequality:** Suppose \( \rho \) and \( \sigma \) are density matrices such that the trace distance between them satisfies \( T(\rho, \sigma) \leq 1/e \). Then

\[ |S(\rho) - S(\sigma)| \leq T(\rho, \sigma) \log d + \eta(T(\rho, \sigma)), \]

where \( d \) is the dimension of the Hilbert space, and \( \eta(x) \equiv -x \log x \). Removing the restriction that \( T(\rho, \sigma) \leq 1/e \) we can prove the weaker inequality

\[ |S(\rho) - S(\sigma)| \leq T(\rho, \sigma) \log d + 1/e. \]

**Klein’s inequality:** The quantum relative entropy is non-negative, \( S(\rho \| \sigma) \geq 0 \), with equality iff \( \rho = \sigma \).

**Basic properties of Von Neumann entropy**:

1. \( S(\rho) \geq 0 \) with equality iff the state is pure.
2. In a \( d \)-dimensional Hilbert space \( S(\rho) \leq \log d \), with equality iff the system is in the completely mixed state \( 1/d \).
3. Suppose a composite system \( AB \) is in a pure state, then \( S(A) = S(B) \).
4. Suppose \( p_i \) are probabilities and the states \( \rho_i \) have support on orthogonal subspaces, then \( S(\sum_i p_i \rho_i) = H(p_i) + \sum_i p_i S(\rho_i) \).
5. **Joint entropy theorem:** Suppose \( p_i \) are probabilities, \( |i\) are orthogonal states for a system \( A \), and \( \rho_i \) is any set of density operators for another system \( B \), then \( S(\sum_i p_i |i\rangle \langle i| \otimes \rho_i) = H(p_i) + \sum_i p_i S(\rho_i) \).

\[ S(\rho \otimes \sigma) = S(\rho) + S(\sigma) \]
The joint entropy $S(A, B)$ for a composite system with two components $A$ and $B$ is defined as $S(A, B) \equiv -\text{tr}(\rho^{AB} \log \rho^{AB})$.

The conditional entropy and mutual information are defined as, respectively

$$S(A|B) \equiv S(A, B) - S(B),$$

$$S(A : B) \equiv S(A) + S(B) - S(A, B) = S(A) - S(A|B) = S(B) - S(B|A).$$

Suppose that a projective measurement described by projectors $P_i$ is performed on a quantum system in state $\rho$, but we never learn the result of the measurement. Then after measurement the system is in state $\rho' = \sum_i P_i \rho P_i$.

**Projective measurements increase entropy**: Suppose $P_i$ is a complete set of orthogonal projectors and $\rho$ is a density operator. Then the entropy of the state $\rho' \equiv \sum_i P_i \rho P_i$ of the system after the measurement is at least as great as the original entropy,

$$S(\rho') \geq S(\rho),$$

with equality iff $\rho = \rho'$.

**Subadditivity inequality** for Von Neumann entropy $S(A, B) \leq S(A) + S(B)$ with equality iff systems $A$ and $B$ are uncorrelated $\rho^{AB} = \rho^A \otimes \rho^B$.

**Triangle inequality (Araki-Lieb inequality)** $S(A, B) \geq |S(A) - S(B)|$.

The entropy is a *concave* function of its inputs. Given probabilities $p_i \geq 0$ such that $\sum_i p_i = 1$ and corresponding density operators $\rho_i$, $S(\sum_i p_i \rho_i) \geq \sum_i p_i S(\rho_i)$, with equality iff all the states $\rho_i$ for which $p_i > 0$ are identical.

**Theorem**: Suppose $\rho = \sum_i p_i \rho_i$, where $p_i$ are some set of probabilities and $\rho_i$ are density operators. Then $S(\rho) \leq \sum_i p_i S(\rho_i) + H(p_i)$, with equality iff the states $\rho_i$ have support on orthogonal subspaces.

$$\sum_i p_i S(\rho_i) \leq S(\sum_i p_i \rho_i) \leq H(p_i) + \sum_i p_i S(\rho_i)$$

Suppose $f(A, B)$ is a real-valued function of two matrices, $A$ and $B$. Then $f$ is *jointly concave* in $A$ and $B$ if $\forall 0 \leq \lambda \leq 1$,

$$f(\lambda A_1 + (1 - \lambda) A_2, \lambda B_1 + (1 - \lambda) B_2) \geq \lambda f(A_1, B_1) + (1 - \lambda) f(A_2, B_2).$$

**Lieb’s theorem**: Let $X$ be a matrix, and $0 \leq t \leq 1$. Then the function $f(A, B) \equiv \text{tr}(X^t A^t B^{1-t})$ is jointly concave in positive matrices $A$ and $B$.

**Convexity of the relative entropy**: The relative entropy $S(\rho \| \sigma)$ is jointly convex in its arguments. That is $S(\lambda \rho_1 + (1 - \lambda) \rho_2 \| \lambda \sigma_1 + (1 - \lambda) \sigma_2) \leq \lambda S(\rho_1 \| \sigma_1) + (1 - \lambda) S(\rho_2 \| \sigma_2)$. 

**Concavity of the quantum conditional entropy:** Let $AB$ be a composite quantum system with components $A$ and $B$. Then the conditional entropy $S(A|B)$ is concave in the state $\rho^{AB}$ of $AB$.

**Strong subadditivity:** For any trio of quantum systems $A$, $B$, $C$, the inequalities hold

$$S(A) + S(B) \leq S(A,C) + S(B,C),$$

$$S(A,B,C) + S(B) \leq S(A,B) + S(B,C).$$

Though impossible in classical case, in the quantum case it is possible to have either $S(A) > S(A,C)$ or $S(B) > S(B,C)$, yet somehow Nature conspires in such a way that both of these possibilities are not true simultaneously.

Rephrasing in terms of conditional entropies and mutual information,

$$0 \leq S(C|A) + S(C|B),$$

$$S(A : B) + S(A : C) \leq 2S(A).$$

Similarly though it is possible that either $S(C|A) < 0$ or $S(C|B) < 0$, both of them can’t hold simultaneously. Though it is possible that either $S(A : B) > S(A)$ or $S(A : C) > S(A)$, both of these relations can’t be true simultaneously.

**Theorem:**

1. **Conditioning reduces entropy:** Suppose $ABC$ is a composite quantum system. Then $S(A|B,C) \leq S(A|B)$.
2. **Discarding quantum systems never increases mutual information:** Suppose $ABC$ is a composite quantum system. Then $S(A : B) \leq S(A : B, C)$.
3. **Quantum operations never increase mutual information:** Suppose $AB$ is a composite quantum system and $\mathcal{E}$ is a trace-preserving quantum operation on system $B$. Let $S(A : B)$ denote the mutual information between systems $A$ and $B$ before $\mathcal{E}$ is applied to system $B$, and $S(A' : B')$ the mutual information after $\mathcal{E}$ is applied to system $B$. Then $S(A' : B') \leq S(A : B)$.

**Subadditivity of the conditional entropy:** Let $ABCD$ be a composite of 4 quantum systems. Then the conditional entropy is jointly subadditive in the first and second entries:

$$S(A,B|C,D) \leq S(A|C) + S(B|D).$$

Let $ABC$ be a composite of 3 quantum systems. Then the conditional entropy is subadditive in each of the first and second entries:

$$S(A,B|C) \leq S(A|C) + S(B|C)$$

$$S(A|B,C) \leq S(A|B) + S(A|C).$$

**Monotonicity of the relative entropy:** Let $\rho^{AB}$ and $\sigma^{AB}$ be any 2 density matrices of a composite system $AB$. Then $S(\rho^{A} \parallel \sigma^{A}) \leq S(\rho^{AB} \parallel \sigma^{AB})$. 

Chapter 12
Quantum information theory

Accessible information is the maximum of mutual information over all possible measurement schemes.

No-cloning theorem: No machine could copy an unknown quantum states. We couldn’t copy non-orthogonal quantum states, but we could copy orthogonal states. Different states of classical information can be thought of as orthogonal quantum states.

The no-cloning theorem is a consequence of the fact that the accessible information is less than \( H(p) \); the fact that the accessible information is less than \( H(p) \) is a consequence of the no-cloning theorem. We can view the no-cloning theorem as being equivalent to the statement that in quantum mechanics the accessible information for non-orthogonal states is in general less than the entropy of preparation.

**Theorem 12.1 (The Holevo bound):**
Suppose Alice prepare a state \( \rho_X \) where \( X = 0, \ldots, n \) with probabilities \( p_0, \ldots, p_n \). Bob performs a measurement described by POVM elements \( \{E_y\} = \{E_0, \ldots, E_m\} \) on that state given by Alice, with measurement outcome \( Y \). The Holevo bound states that for any such measurement Bob may do: \( H(X:Y) \leq S(\rho) - \sum_x p_x S(\rho_x) \), where \( \rho = \sum_x p_x \rho_x \) and the rhs is Holevo \( \chi \) quantity, \( \chi = S(\rho) - \sum_x p_x S(\rho_x) \).

states \( \rho_x \) have orthogonal support \( \iff \) states are orthogonal

Fano’s inequality:
Suppose \( \tilde{X} = f(Y) \) is some function of \( Y \) which we are using as our best guess for \( X \). Let \( p_e = p(X \neq \tilde{X}) \) be the probability that this guess is incorrect. Then \( H(p_e) + p_e \log(|X| - 1) \geq H(X|Y) \), where \( H(\cdot) \) is the binary entropy and \( |X| \) is the number of values \( X \) may assume.

NOTE: \( n \) qubits can not be used to transmit more than \( n \) bits of classical information.
Suppose an i.i.d. information source is producing bits $X_1, X_2, X_3, \ldots$, each being equal to zero with probability $p$, and equal to one with probability $1 - p$. The sequences $x_1, \ldots, x_n$ for large $n$ are known as \textit{typical sequences} with high probability a fraction $p$ of the symbols output form the source will be equal to zero and a fraction $1 - p$ will be equal to one.

For typical sequences
\[ p(x_1, \ldots, x_n) = p(x_1)p(x_2) \ldots p(x_n) \approx p^n p(1 - p)^{(1-p)n} \]
\[ -\log p(x_1, \ldots, x_n) \approx -np \log p - n(1 - p) \log (1 - p) = nH(X), \text{ where } H(X) = -p \log p - (1 - p) \log (1 - p) \text{ is the entropy rate of the source}. \]

Given $\varepsilon > 0$ a string of source symbols $x_1 x_2 \ldots x_n$ is $\varepsilon$-typical if $2^{-n(H(X)+\varepsilon)} \leq p(x_1, \ldots, x_n) \leq 2^{-n(H(X)-\varepsilon)} \text{ or } \left| \frac{1}{n} \log \frac{1}{p(x_1, \ldots, x_n)} - H(X) \right| \leq \varepsilon$. Denote the set of all such $\varepsilon$-typical sequences of length $n$ by $T(n, \varepsilon)$.

\textbf{Theorem 12.3 (Law of large numbers):}
Suppose $X_1, X_2, \ldots$ are independent random variables all having the same distribution as a random variable $X$ with finite first and second moments, $|E(X)| < \infty$ and $|E(X^2)| < \infty$. Then for any $\varepsilon > 0$, $p(|S_n - E(X)| > \varepsilon) \to 0$ as $n \to \infty$, where $S_n \equiv \sum_{i=1}^{n} X_i / n$.

\textbf{Theorem 12.2 (Theorem of typical sequences):}
(1) Fix $\varepsilon > 0$. Then for any $\delta > 0$, for sufficiently large $n$, the probability that a sequence is $\varepsilon$-typical is at least $1 - \delta$.
(2) For any fixed $\varepsilon > 0$ and $\delta > 0$, for sufficiently large $n$, the number $|T(n, \varepsilon)|$ of $\varepsilon$-typical sequences satisfies $(1 - \delta)2^{n(H(X)-\varepsilon)} \leq |T(n, \varepsilon)| \leq 2^{n(H(X)+\varepsilon)}$.
(3) Let $S(n)$ be a collection of size at most $2^{nH}$, of length $n$ sequences from the source, where $R < H(X)$ is fixed. Then for any $\delta > 0$ and for sufficiently large $n$, $\sum_{x \in S(n)} p(x) \leq \delta$.

\textbf{Theorem 12.4 (Shannon's noiseless channel coding theorem):}
Suppose $\{X_i\}$ is an i.i.d. information source with entropy rate $H(X)$. Suppose $R > H(X)$, then there exists a reliable compression scheme of rate $R$ for the source. Conversely if $R < H(X)$ then any compression scheme will not be reliable.

Shannon’s noiseless channel coding theorem specifies for what values of the rate $R$ a reliable compression scheme exists, revealing a remarkable operational interpretation for the entropy rate $H(X)$: it is the minimal physical resources necessary and sufficient to reliably store the output from the source.

An i.i.d. quantum source will be described by a Hilbert space $H$, and a density matrix $\rho$ on that Hilbert space. We imagine that the state $\rho$ of the system is merely part of a larger system which is in a pure state, and the mixed nature of $\rho$ is
due to entanglement between $H$ and the remainder of the system. Analogous to the classical case, a compression scheme of rate $R$ for this source: $\mathcal{C}^n$ is the compression operation taking states in $H^\otimes n$ to states in a $2^{nR}$-dimensional state space, the compressed space which represents $nR$ qubits; $\mathcal{D}^n$ is the decompression operation, taking states in the compressed space to states in the original state space. The combined compression-decompression operation is $\mathcal{D}^n \circ \mathcal{C}^n$. The criterion for reliability is that in the limit of large $n$ the entanglement fidelity $F(\rho \otimes^n, \mathcal{D}^n \circ \mathcal{C}^n)$ should tend towards one.

Suppose a quantum source is $\rho = \sum_x p(x) |x\rangle \langle x|$, where $|x\rangle$ is an orthonormal set, and $p(x)$ are the eigenvalues of $\rho$. An $\varepsilon$-typical sequence $x_1, \ldots, x_n$ satisfies

$$\left| \frac{1}{n} \log \left( \prod_{i=1}^{n} p(x_i)^2 \right) - S(\rho) \right| \leq \varepsilon.$$

An $\varepsilon$-typical state is a state $|x_1\rangle |x_2\rangle \cdots |x_n\rangle$ for which the sequence $x_1, x_2, \ldots, x_n$ is $\varepsilon$-typical. Define the $\varepsilon$-typical subspace to be the subspace spanned by all $\varepsilon$-typical states, $|x_1\rangle \cdots |x_n\rangle$. Denote the $\varepsilon$-typical subspace by $T(n, \varepsilon)$, and the projector onto the $\varepsilon$-typical subspace by $P(n, \varepsilon)$.

$$P(n, \varepsilon) = \sum_{x \in \text{typical}} |x_1\rangle \langle x_1| \otimes |x_2\rangle \langle x_2| \otimes \cdots \otimes |x_n\rangle \langle x_n|.$$

**Theorem 12.5 (Typical subspace theorem):**

1. Fix $\varepsilon > 0$. Then for any $\delta > 0$, for sufficiently large $n$, $\text{tr}(P(n, \varepsilon) \rho \otimes^n) \geq 1 - \delta$.
2. For any fixed $\varepsilon > 0$ and $\delta > 0$, for sufficiently large $n$, the dimension $|T(n, \varepsilon)| = \text{tr}(P(n, \varepsilon))$ of $T(n, \varepsilon)$ satisfies $(1 - \delta)2^{nS(\rho) - \varepsilon} \leq |T(n, \varepsilon)| \leq 2^{nS(\rho) + \varepsilon}$.
3. Let $S(n)$ be a projector onto any subspace of $H^\otimes n$ of dimension at most $2^{nR}$, where $R < S(\rho)$ is fixed. Then for any $\delta > 0$ and for sufficiently large $n$, $\text{tr}(S(n) \rho \otimes^n) \leq \delta$.

**Theorem 12.6 (Schumacher’s noiseless channel coding theorem)**

Let $\{H, \rho\}$ be an i.i.d. quantum source. If $R > S(\rho)$ then there exists a reliable compression scheme of rate $R$ for the source $\{H, \rho\}$. If $R < S(\rho)$ then any compression scheme of rate $R$ is not reliable.

The maximum rate for reliable communication through the channel $\mathcal{N}$ is the capacity of the channel.

The maximum rate at which information can be reliably transmitted through a binary symmetric channel is $1 - H(p)$, where $H(\cdot)$ is the Shannon entropy.
**Theorem 12.7 (Shannon’s noisy channel coding theorem)**

For a noisy channel \( \mathcal{N} \) the capacity is given by

\[
C(\mathcal{N}) = \max_{p(x)} H(X : Y),
\]

where the maximum is taken over all input distributions \( p(x) \) for \( X \), for one use of the channel, and \( Y \) is the corresponding induced random variable at the output of the channel.

Restriction: using *product states* of the form \( \rho_1 \otimes \rho_2 \otimes \ldots \), to encode messages. Each of the \( \rho_1, \rho_2, \ldots \) are potential inputs for one use of the channel \( \mathcal{E} \).

**Theorem 12.8 (Holevo-Schumacher-Westmoreland (HSW) theorem)**

**Theorem 12.9 (Quantum Fano inequality)**

**Theorem 12.10 (Quantum data processing inequality)**

**Quantum Singleton bound**: An \([n,k,d]\) code uses \( n \) qubits to encode \( k \) qubits and is able to correct located errors on up to \( d - 1 \) of the qubits, then we must have

\[
2(d - 1) \leq n - k.
\]
<table>
<thead>
<tr>
<th><strong>Information Theory</strong></th>
<th><strong>Quantum</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classical</strong></td>
<td><strong>von Neumann entropy</strong></td>
</tr>
<tr>
<td>Shannon entropy</td>
<td>$S(\rho) = -tr(\rho \log \rho)$</td>
</tr>
<tr>
<td>$H(X) = -\sum x p(x) \log p(x)$</td>
<td></td>
</tr>
</tbody>
</table>

| **Distinguishability and accessible information** |  |
| Letters always distinguishable |  |
| $N = |X|$ |  |

| **Noiseless channel coding** |  |
| Shannon’s theorem | Schumacher’s theorem |
| $n_{\text{bits}} = H(X)$ | $r_{\text{qubits}} = S(\sum x p_x \rho_x)$ |

| **Capacity of noisy channels for classical information** |  |
| Shannon’s noisy coding theorem | Holevo-Schumacher-Westmoreland theorem |
| $C(\mathcal{N}) = \max p(x)$ | $C^{(1)}(\mathcal{E}) = \max_{\{p, \rho\}} [S(\rho') - \sum p x S(\rho'_x)]$ |
| $\rho'_x = \mathcal{E}(\rho_x)$, $\rho' = \sum x p_x \rho'_x$ |  |

| **Information-theoretic relations** |  |
| Fano inequality | Quantum Fano inequality |
| $H(p_x) + p_x \log(|X| - 1) \geq H(X|Y)$ | $H(F(\rho, \mathcal{E}')) + (1 - F(\rho, \mathcal{E}')) \log (d^2 - 1) \geq S(\rho, \mathcal{E}')$ |
| Mutual information | Coherent information |
| $H(X : Y) = H(Y) - H(Y|X)$ | $I(\rho, \mathcal{E}') = S(\mathcal{E}'(\rho)) - S(\rho, \mathcal{E}')$ |
| Data processing inequality | Quantum data processing inequality |
| $X \to Y \to Z$ | $\rho \to \mathcal{E}_1(\rho) \to (\mathcal{E}_2 \circ \mathcal{E}_1)(\rho)$ |
| $H(X) \geq H(X : Y) \geq H(X : Z)$ | $S(\rho) \geq I(\rho, \mathcal{E}_1) \geq I(\rho, \mathcal{E}_2 \circ \mathcal{E}_1)$ |
LOCC–local operations and classical communication

**Proposition 12.11**

**Theorem 12.12 (Birkhoff’s theorem)**

A $d \times d$ matrix $D$ is doubly stochastic (that is, has non-negative entries and each row and column sums to 1) if and only if $D$ can be written as a convex combination of permutation matrices, $D = \sum_j p_j P_j$.

Suppose $H$ and $K$ are Hermitian operators. Then we say $H \prec K$ if $\lambda(H) \prec \lambda(K)$, where we use $\lambda(H)$ to denote the vector of eigenvalues of a Hermitian operator $H$.

**Theorem 12.13:** Let $H$ and $K$ be Hermitian operators. Then $H \prec K$ iff there is a probability distribution $p_j$ and unitary matrices $U_j$ st. $H = \sum_j p_j U_j K U_j^\dagger$.

**Proposition 12.14:**

**Theorem 12.15:**

**Theorem 12.16:**

**Theorem 12.17:**

**Proposition 12.18 (Information gain implies disturbance):** In any attempt to distinguish between two non-orthogonal quantum states, information gain is only possible at the expense of introducing disturbance to the signal.

QKD protocols:

- BB84 protocol
- B92 protocol
- EPR protocol

QKD security criterion: A QKD protocol is defined as being secure if, for any security parameters $s > 0$ and $l > 0$ chosen by Alice and Bob, and for any eavesdropping strategy, either the scheme aborts, or it succeeds with probability at least $1 - \Theta(2^{-s})$, and guarantees that Eve’s mutual information with the final key is less than $2^{-l}$. The key string must also be essentially random.

**Proposition 12.19 (High fidelity implies low entropy):**
Appendix 1: Basic probability theory
Appendix 2: Group theory

A group $G$ is finite if the number of elements in $G$ is finite. The order of a finite group $G$ is the number of elements it contains, denoted as $|G|$.

The order of an element $g \in G$ is the smallest positive integer $r$ such that $g^r(g$ multiplied with itself $r$ times) equals the identity element $e$. Otherwise if there is no such $r$, $g$ has infinite order.

A subgroup $H$ of $G$ is a subset of $G$ which forms a group under the same group multiplication operation as $G$.

Lagrange's theorem If $H$ is a subgroup of a finite group $G$ then $|H|$ divides $|G|$.

The order of an element $g \in G$ divides $|G|$.

If $g_1$ and $g_2$ are elements of $G$, then the conjugate of $g_2$ wrt $g_1$ is the element $g_1^{-1}g_2g_1$. If $H$ is a subgroup of $G$ and if $g^{-1}Hg = H$, for all $g \in G$, then $H$ is a normal subgroup. The conjugacy class $G_x$ of an element $x \in G$ is defined by $G_x = \{g^{-1}xg | g \in G\}$.

A set of elements $\{g_1, \ldots, g_l\} \subseteq G$ generate the group $G$ if every element of $G$ can be written as a product of elements from the set. The group is denoted by $G = <g_1, \ldots, g_l>$. $G$ must have a set of generators containing at most $\log |G|$ elements.

A cyclic group $G$ has an element $a$ such that any element $g \in G$ can be expressed as $a^n$ for some integer $n$. $a$ is a generator, $G = <a>$.

If $H$ is a subgroup of $G$, the left coset of $H$ in $G$ determined by $g \in G$ is the set $gH = \{gh | h \in H\}$.

Let $M_n$ be the set of $n \times n$ complex matrices. The (matrix) representation $\rho$ of a group $G$ is defined as a function which maps $G$ to a matrix group in $M_n$, preserving group multiplication. And $\rho$ has dimension $d_\rho = n$. Specifically $g \in G$ is mapped to $\rho(g) \in M_n$ such that $g_1g_2 = g_3$ implies $\rho(g_1)\rho(g_2) = \rho(g_3)$. If the map is many to one, it is homomorphism; if it is one to one, it is isomorphism.

The character of a matrix group $G \subset M_n$ is a function on the group defined by $\chi(g) = \text{tr}(g)$, for $g \in G$. Properties:

- $\chi(I) = n$,
- $|\chi(g)| \leq n$,
- $|\chi(g)| = n$ implies $g = e^{i\theta}I$
- $\chi$ is constant on any given conjugacy class of $G$
- $\chi(g^{-1}) = \chi^*(g)$
- $\chi(g)$ is an algebraic number for all $g$

Two matrix groups are equivalent if they are isomorphic and corresponding elements under the isomorphism have the same character.

A matrix group $G$ in $M_n$ is completely reducible if it is equivalent to another matrix group $H$ which is of block diagonal form. If no such equivalence exists, the matrix group is irreducible.

Schur's lemma Let $G \subset M_n$ and $H \subset M_k$ be two matrix groups of the same size $|G| = |H|$. If there exists a $k \times n$ matrix $S$ such that $Sg_i = h_iS$ for some ordering of all elements $g_i \in G$ and $h_i \in H$ then either $S$ is the zero matrix or $n = k$ and $S$ is a square nonsingular matrix.
Theorem A matrix group is irreducible iff 
\[ \frac{1}{|G|} \sum_{g \in G} |\chi(g)|^2 = 1. \]

Fundamental theorem Every group has exactly \( r \) inequivalent irreducible representations, where \( r \) is the number of conjugacy classes of \( G \). If \( \rho^p \) and \( \rho^q \) in \( M_d \) are any two of these, then the matrix elements satisfy the orthogonality relations

\[ \sum_{g \in G} [\rho^p(g)]_{ij}^{-1} [\rho^q(g)]_{kl} = \frac{|G|}{d_p} \delta_{il} \delta_{jk} \delta_{pq} \]

where \( \delta_{pq} = 1 \) if \( \rho^p = \rho^q \) and is zero otherwise.

A representation is faithful if the matrix group of the representation is isomorphic to the original group. The regular representation is a faithful representation which exists for any group and is constructed as the following (permutation matrices).......

Theorem If \( \rho \) is an arbitrary representation of \( G \) with character \( \chi \) and \( \rho^p \) are the inequivalent irreducible representations of \( G \) with characters \( \chi^p \), then \( \rho = \bigoplus p c_p \rho^p \), where \( \bigoplus \) denotes a direct sum and \( c_p \) are the numbers determined by

\[ c_p = \frac{1}{|G|} \sum_{t=1}^r r_t (\chi^p) \ast \chi_t \]

where \( r \) is the number of inequivalent irreducible representations.

Let \( G \) be a finite group of order \( N \) and \( f \) be a function which maps elements of \( G \) to complex numbers. For an irreducible representation \( \rho \) of \( G \), of dimension \( d_\rho \), define the Fourier transform of \( f \) to be \( \hat{f} \),

\[ \hat{f}(\rho) \equiv \sqrt{\frac{d_\rho}{N}} \sum_{g \in G} f(g) \rho(g). \]

Let \( \hat{G} \) be a complete set of inequivalent irreducible representations of \( G \). The inverse Fourier transform of \( f \) is defined as

\[ f(g) = \frac{1}{\sqrt{N}} \sum_{\rho \in \hat{G}} \sqrt{d_\rho} \psi(\hat{f}(\rho) \rho(g^{-1})) \]
Appendix 3: The Solovay-Kitaev theorem
Appendix 4: Number theory

set of int. $Z = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$

natural #s: pos. int.

An int. $d$ divides $n$, denoted by $d | n$, if $\exists k \in \text{int. st. } n = dk \iff d$ is a factor/divisor of $n$

a prime # is an int. $> 1$ which has only 1 and itself as factors

**Fundamental theorem of arithmetic:** int. $a > 1$, $a$ has a prime factorization of the form, $a = p_1^{a_1} p_2^{a_2} \cdots p_n^{a_n}$, where $p_1, \cdots, p_n$ are distinct prime #s, $a_1, \cdots, a_n$ pos. int. This is unique, up to order of factors.

$\forall x, n \in \text{pos.int., } x = kn + r, k = \text{pos.int., } r \text{ remainder } 0 \leq r \leq n - 1$(indicating $x > n$)

greatest common divisor of int $a,b$: $\gcd(a,b)$

**Representation theorem for gcd:** $\gcd(a,b)$ is the least pos. int. that can be written in the form $ax + by, x, y \in \text{int.}$

c divides $a$ and $b \iff c$ divides $\gcd(a,b)$

multiplicative inverse: given $a$ and $n$, there exists $b \text{ st. } ab = 1 (\text{mod } n)$

$a,b$ co-prime $\iff \gcd(a,b) = 1$

$n$ int. $> 1$, int. $a$ has a multiplicative inverse modulo $n$ iff $\gcd(a,n) = 1$

$a, b \in \text{int. } r$ is remainder when $a$ is divided by $b$. If $r \neq 0, \gcd(a,b) = \gcd(b,r)$

**Chinese remainder theorem:** $m_1, \cdots, m_n$ are pos. int. st. any pair $m_i$ and $m_j (i \neq j)$ are co-prime. The equations,

$$x = a_1 (\text{mod } m_1)$$
$$x = a_2 (\text{mod } m_2)$$
$$\cdots$$
$$x = a_n (\text{mod } m_n)$$

has a solution. Any two solutions are equal modulo $M \equiv m_1 m_2 \cdots m_n$

$p$ is prime and $k$ is int. $1 \leq k \leq p - 1 \iff p$ divides $\binom{p}{k}$

**Fermat’s little theorem:** $p$ is prime, and $a$ is any int. then $a^p = a (\text{mod } p)$. If $a$ is not divisible by $p$, then $a^{p-1} = 1(\text{mod } p)$

Euler function $\varphi(n) \equiv \# \text{ of pos. ints. less than } n \text{ and co-prime to } n$

If $a$ and $b$ are co-prime, $\varphi(ab) = \varphi(a)\varphi(b)$

Based on prime factorization $n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}, \varphi(n) = \prod_{j=1}^{k} p_j^{a_j - 1} (p_j - 1)$

$a$ is co-prime to $n$, then $a^{\varphi(n)} = 1(\text{mod } n)$

Define $Z_n^*$ to be the set of all elements in $Z_n$ which have inverse modulo $n$, that is, the set of all elements in $Z_n$ which are co-prime to $n$

$p$ odd prime, $\alpha$ pos. int. then $Z_{p^n}^*$ is cyclic

$N$ pos. int. $x$ is co-prime to $N$, $1 \leq x < N$. The least pos. int. $r$, st. $x^r = 1(\text{mod } N)$, is the order of $x$ modulo $N$

$N$ is a composite # $L$ bits long. $x$ is a non-trivial (neither $x = 1(\text{mod } N)$ nor $x = N - 1 = -1(\text{mod } N)$) solution to $x^2 = 1(\text{mod } N)$ in range $1 < x < N$. Then at least one of $\gcd(x - 1, N)$
and \( \gcd(x + 1, N) \) is a non-trivial factor of \( N \) that can be computed using \( O(L^3) \) operations.

\( p \) is an odd prime. \( 2^d \) is the largest power of 2 dividing \( \varphi(p^a) \). Then with probability exactly \( \frac{1}{2} \), \( 2^d \) divides the order of a randomly chosen element of \( \mathbb{Z}_{p^a}^* \) modulo \( p^a \).

\( N = p_1^{a_1} \cdots p_m^{a_m} \) is the prime factorization of an odd composite positive integer. Then

\[
p(r \text{ is even and } x^{r/2} \not\equiv -1 \pmod{N}) \geq 1 - \frac{1}{2^m}.
\]

A finite simple continued fraction is defined by a finite collection \( a_0, \ldots, a_N \) of positive integers.

\[
[a_0, \ldots, a_N] \equiv a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\ddots + \frac{1}{a_N}}}}.
\]

The \( n \)th convergent (\( 0 \leq n \leq N \)) to this continued fraction is defined to be \( [a_0, \ldots, a_n] \).

\( x \) is a rational \# \( \geq 1 \). Then \( x \) has a representation as a continued fraction \( x = [a_0, \ldots, a_N] \), which may be found by the continued fractions algorithm (split then invert).

\( a_0, \ldots, a_N \) is a sequence of positive numbers. Then \( [a_0, \ldots, a_n] = \frac{p_n}{q_n} \), where \( p_n \) and \( q_n \) are real numbers defined inductively by \( p_0 \equiv a_0, q_0 \equiv 1 \) and \( p_1 \equiv 1 + a_0a_1, q_1 \equiv a_1, \) and for \( 2 \leq n \leq N, \)

\[
p_n \equiv a_n p_{n-1} + p_{n-2},
\]

\[
q_n \equiv a_n q_{n-1} + q_{n-2}.
\]

In the case where \( a_j \) are positive integers, so too are the \( p_j \) and \( q_j \).

\( x \) is a rational \# and \( \frac{p}{q} \) is a rational \# such that \( \left| \frac{p}{q} - x \right| \leq \frac{1}{2q^2} \). Then \( \frac{p}{q} \) is a convergent of the continued fraction for \( x \).