Contents

1. Quantum mechanics essentials

1.1. States and wave functions

- Probability of finding particle in (a, b) is

$$P(a, b; t) = \int_a^b |\psi(x, t)|^2 \, \mathrm{d}x$$

  Wave function is normalised so that $P(-\infty, +\infty; t) = 1$.

1.2. Dirac notation

Definition. **Dual** of vector space V is set of linear functionals from V to $\mathbb{C}$:

$$V^* := \{\Phi : V \to \mathbb{C} : \forall a, b \in \mathbb{C}, \forall z, w \in V, \quad \Phi(a\boldsymbol{z} + b\boldsymbol{w}) = a\Phi(\boldsymbol{z}) + b\Phi(\boldsymbol{w})\}$$

We have $\dim(V^*) = \dim(V)$.

Remark. If V has inner product $\langle \cdot, \cdot \rangle$, then an isomorphism is given by $\boldsymbol{z} \mapsto \Phi_{\boldsymbol{z}}(\cdot) = \langle \boldsymbol{z}, \cdot \rangle$.

Definition. **Dual** of $\boldsymbol{z} \in V$ is the corresponding element in V^*, i.e. $\Phi_{\boldsymbol{z}}$.

Remark. If $V = \mathbb{C}^n$, can think of vectors in V as $n \times 1$ matrices and vectors in V^* as $1 \times n$ matrices.

Definition. **Dirac notation** denotes vectors in a Hilbert space or its dual:

- Write $|\psi\rangle$ (a **ket**) for vector in Hilbert space $\mathscr{H}$ corresponding to wave function ψ.
- Write $\langle\varphi|$ (a **bra**) for dual vector in $\mathscr{H}^*$.
- A **bra-ket** denotes an inner product:

$$\langle \varphi | \psi \rangle := \langle \varphi, \psi \rangle = \int_{-\infty}^{\infty} \varphi^*(x, t)\psi(x, t) \, \mathrm{d}x$$

1.3. Hilbert spaces

Definition. **Hilbert space** is real or complex vector space with Hermitian inner product that is also a complete metric space with metric induced by the inner product. In particular, inner product satisfies:

- **Hermiticity**: $\langle \psi | \varphi \rangle = \langle \varphi | \psi \rangle^*$.
- **Sesquilinearity** (linear in the second factor, anti-linear in the first). For $|\varphi\rangle = c_1|\varphi_1\rangle + c_2|\varphi_2\rangle$:

$$\begin{aligned} \langle \psi | \varphi \rangle &= c_1 \langle \psi | \varphi_1 \rangle + c_2 \langle \psi | \varphi_2 \rangle \\ \langle \varphi | \psi \rangle &= c_1^* \langle \varphi_1 | \psi \rangle + c_2^* \langle \varphi_2 | \psi \rangle \end{aligned}$$

- **Positive definiteness**: $\langle \psi | \psi \rangle \geq 0$ and $\langle \psi | \psi \rangle = 0 \Longleftrightarrow |\psi\rangle = 0$ (this corresponds with a **physical state** condition).

Definition. A quantum mechanical system is described by a **state** $|\psi\rangle$ in Hilbert space $\mathscr{H}$.

Remark. States which differ by only a normalisation factor are physically equivalent:

$$\forall c \in \mathbb{C}^*, \quad |\psi\rangle \sim c|\psi\rangle$$

For this reason, pure quantum mechanical states are called **rays** in the Hilbert space, and we normally assume that a state $|\psi\rangle$ has norm 1: $\| |\psi\rangle \| = 1$.

Remark. Note that the state labelled zero, $|0\rangle$, is not equal to the zero state (the 0 vector).

1.4. Operators

Definition. $\hat{A} : \mathcal{H} \to \mathcal{H}$ is **linear operator** if

$$\forall a, b \in \mathbb{C}, \forall |\psi\rangle, |\varphi\rangle \in \mathcal{H}, \quad \hat{A}(a|\psi\rangle + b|\varphi\rangle) = a(\hat{A}|\psi\rangle) + b(\hat{A}|\varphi\rangle)$$

Proposition. Products and linear combinations of linear operators are also linear operators.

Definition. **Adjoint (Hermitian conjugate)** of $\hat{A}$, $\hat{A}^\dagger$, is defined by

$$\langle\psi|\hat{A}^\dagger|\varphi\rangle = \left(\langle\varphi|\hat{A}|\psi\rangle\right)^*$$

for all states $|\psi\rangle$ and $|\varphi\rangle$.

Definition. $\widehat{H}$ is **self-adjoint (Hermitian)** if $\widehat{H}^\dagger = \widehat{H}$. Self-adjoint operators correspond to **observables** (measurable quantities) since they have real eigenvalues.

Definition. $\hat{U}$ is **unitary** if $\hat{U}^\dagger\hat{U} = \hat{I}$. Unitary operators describe time-evolution in quantum mechanics.

Definition. **Commutator** of operators $\hat{A}$ and $\hat{B}$ is

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$$

Definition. **Anti-commutator** of operators $\hat{A}$ and $\hat{B}$ is

$$\{\hat{A}, \hat{B}\} = \hat{A}\hat{B} + \hat{B}\hat{A}$$

Definition. **Expectation value** of observable $\hat{A}$ on state $|\psi\rangle$ is

$$\langle A\rangle_\psi := \langle\psi|\hat{A}|\psi\rangle$$

Interpreted as average outcome of many measurements of $\hat{A}$ on same state $|\psi\rangle$.

1.5. Matrix representation

Definition. **Matrix form** of operator $\hat{A}$ with respect to orthonormal basis $\{|n\rangle\}$ is given by $A_{ij} = \langle i|\hat{A}|j\rangle$.

Proposition. For operator $\hat{A}$ with matrix representation A in basis $\{|n\rangle\}$, matrix representation of $\hat{A}$ in basis $\{|m\rangle\}$ is $B = SAS^{-1}$ where S is change of basis matrix from old basis $\{|n\rangle\}$ to new basis $\{|m\rangle\}$.

1.6. Time-evolution

Theorem. Time-evolution of state is given by **Schrodinger equation**:

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = \widehat{H}|\psi(t)\rangle \Longrightarrow |\psi(t)\rangle = \widehat{U}_t|\psi(0)\rangle$$

where $\widehat{H} = \widehat{K} + \widehat{V}$ is Hamiltonian operator, $\widehat{U}_t$ is unitary operator. If $\widehat{H}$ independent of t, then $\widehat{U}_t = \exp\left(-\frac{i}{\hbar}t\widehat{H}\right)$.

- **Principle of superposition**: Schrodinger equation is linear, so any linear combination of solutions is another solution.

Definition. **Exponential** of operator $\hat{A}$ is

$$\exp(\hat{A}) := \sum_{n\in\mathbb{N}_0}\frac{\hat{A}^n}{n!}$$

2. Measurement and uncertainty

2.1. Observables

Proposition. For Hilbert space of finite dimension N, operator $\hat{A}$ has N eigenvalues (counting multiplicities). Eigenvalues of Hermitian operator $\widehat{M}$ correspond to possible values of the measurable quantity it represents.

Definition. **Spectrum** of operator $\widehat{H}$ is

$$\operatorname{Spec}(\widehat{H}) := \{\lambda \in \mathbb{C} : \widehat{H} - \lambda\hat{I} \text{ non invertible}\}$$

For finite-dimensional Hilbert space, this is equal to the set of eigenvalues of $\widehat{H}$.

Proposition. Eigenstates $|n\rangle$ of Hermitian operator $\widehat{H}$ corresponding to different eigenvalues λ_n are orthogonal. If eigenvalue is degenerate (multiplicity greater than one) then for each eigenspace (vector space spanned by the eigenvectors) with dimension greater than one, we can choose an orthogonal basis of eigenstates.

Definition. Let $\hat{A}$ have orthonormal eigenstates $\{|v_i\rangle : i \in [N]\}$ and corresponding eigenvalues $\{\lambda_i : i \in [N]\}$. **Spectral representation** of $\hat{A}$ is

$$\hat{A} = \sum_{i=1}^{N}\lambda_i|v_i\rangle\langle v_i|$$

In particular, only eigenvalue of $\hat{I}$ is 1 with degeneracy N, so for any orthonormal basis $\{|v_i\rangle : i \in [N]\}$ of $\mathcal{H}$:

$$\hat{I} = \sum_{i=1}^{N}|v_i\rangle\langle v_i|$$

Definition. When measurement is made on state $|\psi\rangle = \sum_{i=1}^{N} c_i|v_i\rangle$, result is λ with probability

$$p = \sum_{i \in [N], \lambda_i = \lambda} |\langle v_i | \psi \rangle|^2 = \sum_{i \in [N], \lambda_i = \lambda} |c_i|^2$$

If result is λ, measuring again immediately after the measurement will yield λ, so state collapses (up to irrelevant phase $e^{i\alpha}$, $\alpha \in \mathbb{R}$) to

$$\frac{1}{\sqrt{p}} \sum_{i \in [N], \lambda_i = \lambda} c_i |v_i\rangle$$

This is **collapse of the wavefunction** and cannot be represented by unitary transformation, so is not reversible.

Definition. Linear operator $\hat{P}$ is **projector** if $\hat{P}^\dagger = \hat{P}$ and $\hat{P}^2 = \hat{P}$.

Definition. For orthonormal eigenstates $\{|v_i\rangle : i \in [N]\}$ of operator $\hat{A}$ and corresponding eigenvalues $\{\lambda_i : i \in [N]\}$, define projection operator

$$\hat{P}_\lambda = \sum_{i \in [N], \lambda_i = \lambda} |v_i\rangle\langle v_i|$$

Proposition. Probability of measurement $\hat{A}$ on state $|\psi\rangle$ yielding λ is $p_\lambda = \langle \psi | \hat{P}_\lambda | \psi \rangle$ and state collapses to $\frac{1}{\sqrt{p_\lambda}} \hat{P}_\lambda |\psi\rangle$.

Definition. $\hat{A}$ and $\hat{B}$ are **compatible** if $[\hat{A}, \hat{B}] = 0$.

Remark. State can only have definite values for observables A and B if it is simultaneous eigenstate of both $\hat{A}$ and $\hat{B}$. There always exist simultaneous eigenstates for compatible operators.

Remark. If $\hat{A}$ and $\hat{B}$ not compatible, measuring A then B then A again will not always give same result for both measurements of A.

2.2. Density matrices

Definition. A state is **pure state** if it is definite, i.e. state of system is completely known, and only uncertainties are due to inherent uncertain nature of quantum mechanics.

Definition. **Density matrix (density operator)** of **pure state** $|\psi\rangle$ is

$$\hat{\rho} := |\psi\rangle\langle\psi|$$

Theorem. There is bijection between density matrices and pure states, and

$$\begin{aligned} \hat{M}|\psi\rangle = \lambda|\psi\rangle \quad &\Longleftrightarrow \quad \hat{M}\hat{\rho} = \lambda\hat{\rho} \\ |\psi\rangle \to \hat{U}|\psi\rangle \quad &\Longleftrightarrow \quad \hat{\rho} \to \hat{U}\hat{\rho}\hat{U}^\dagger \end{aligned}$$

i.e. transforming state $|\psi\rangle$ by unitary operator $\hat{U}$ is equivalent to transforming density matrix $\hat{\rho}$ to $\hat{U}\hat{\rho}\hat{U}^\dagger$.

Definition. For any orthonormal basis states $\{|v_i\rangle : i \in [N]\}$, **trace** of $\hat{A}$ is

$$\operatorname{tr}(\hat{A}) = \sum_{i=1}^{N} \langle v_i | \hat{A} | v_i \rangle$$

Proposition. Trace satisfies **cyclicity**:

$$\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB)$$

Proposition. Density matrix of pure state is a projector.

Proposition. Density matrix $\hat{\rho}$ of pure state satisfies $\operatorname{tr}(\hat{\rho}) = \operatorname{tr}(\hat{\rho}^2) = 1$.

Definition. **Mixed state** is one where state of system is not known. It is ensemble of pure states, each with associated probability of system being in that state: $\{(p_i, |v_i\rangle) : i \in [M]\}$, where each $|v_i\rangle$ is normalised. This is classical uncertainty rather than quantum uncertainty.

Definition. **Density matrix** of **mixed state** is linear combination of density matrices for each pure state weighted by probability:

$$\hat{\rho} := \sum_{i=1}^{M} p_i |v_i\rangle\langle v_i|$$

Can generalise definition to include possibility of ensembles containing mixed states: $\hat{\rho} = \sum_{i=1}^{M} p_i \hat{\rho}_i$ where $\hat{\rho}_i$ are mixed and/or pure density matrices.

Note. One density matrix may be given by multiple mixed states.

Proposition. Let $\hat{A}$ observable, then expected value of measuring $\hat{A}$ on $\hat{\rho}$ is $\langle \hat{A} \rangle = \operatorname{tr}(\hat{\rho}\hat{A})$.

Proposition. $\hat{\rho}$ is a density matrix of a pure/mixed state iff it satisfies:

- **Normalised**: $\operatorname{tr}(\hat{\rho}) = 1$
- **Hermitian**: $\hat{\rho}^\dagger = \hat{\rho}$
- **Semi-positive-definite**: for every state $|\varphi\rangle$, $\langle \varphi | \hat{\rho} | \varphi \rangle \geq 0$ (can be $= 0$ when $|\varphi\rangle \neq 0$). This holds if $\hat{\rho}$ has non-negative eigenvalues, or if $\operatorname{tr}(\hat{\rho}^2) \leq 1$.

Proposition. After taking measurement of pure or mixed state $\hat{\rho}$:

- Result is λ with probability $p_\lambda = \operatorname{tr}(\hat{P}_\lambda \hat{\rho} \hat{P}_\lambda) = \operatorname{tr}(\hat{P}_\lambda \hat{\rho}) = \operatorname{tr}(\hat{\rho} \hat{P}_\lambda)$.
- Density matrix after measuring value of λ is $\frac{1}{p_\lambda} \hat{P}_\lambda \hat{\rho} \hat{P}_\lambda$.

Theorem. Let $\hat{\rho}$ be density matrix, then $\hat{\rho}$ corresponds to pure state iff $\operatorname{tr}(\hat{\rho}^2) = 1$.

3. Qubits and the Bloch sphere

3.1. Qubits

Definition. A **qubit** is state in two-dimensional Hilbert space. Usually **computational basis** $\{|0\rangle, |1\rangle\}$ denotes basis for such a Hilbert space.

Proposition. General pure state in qubit system is of the form

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\varphi}\sin\left(\frac{\theta}{2}\right)|1\rangle, \quad 0 \le \theta \le \pi, 0 \le \varphi < 2\pi$$

So there is bijection between pure qubit states and points on S^2, called the **Bloch sphere**. Any point on Bloch sphere can be labelled by its position vector:

$$\boldsymbol{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad x = \sin(\theta)\cos(\varphi), y = \sin(\theta)\sin(\varphi), z = \cos(\theta)$$

Definition. We define six special states on the Bloch sphere:

$$\begin{aligned}
|+\rangle := \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \leftrightarrow \frac{1}{\sqrt{2}}(1,1)^T &: \quad \boldsymbol{r} = (1,0,0)^T, \quad (\theta,\varphi) = (\pi/2, 0) \\
|-\rangle := \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \leftrightarrow \frac{1}{\sqrt{2}}(1,-1)^T &: \quad \boldsymbol{r} = (-1,0,0)^T, \quad (\theta,\varphi) = (\pi/2, \pi) \\
|L\rangle := \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle) \leftrightarrow \frac{1}{\sqrt{2}}(1,i)^T &: \quad \boldsymbol{r} = (0,1,0)^T, \quad (\theta,\varphi) = (\pi/2, \pi/2) \\
|R\rangle := \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle) \leftrightarrow \frac{1}{\sqrt{2}}(1,-i)^T &: \quad \boldsymbol{r} = (0,-1,0)^T, \quad (\theta,\varphi) = (\pi/2, 3\pi/2) \\
|0\rangle \leftrightarrow (1,0)^T &: \quad \boldsymbol{r} = (0,0,1)^T, \quad (\theta,\varphi) = (0, \cdot) \\
|1\rangle \leftrightarrow (0,1)^T &: \quad \boldsymbol{r} = (0,0,-1)^T, \quad (\theta,\varphi) = (\pi, \cdot)
\end{aligned}$$

3.2. Inside the Bloch sphere

Definition. **Pauli σ-matrices** are

$$\sigma_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 := \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Definition. For pure state $|\psi\rangle$, **Bloch vector** $\boldsymbol{r}$ is corresponding point on Bloch sphere. For mixed state $\{(p_i, |v_i\rangle) : i \in [M]\}$, **Bloch vector** is

$$\boldsymbol{r} := \sum_{i=1}^{M} p_i \boldsymbol{r_i}$$

where $\boldsymbol{r_i}$ is Bloch vector corresponding to pure state $|v_i\rangle$.

Proposition. Density matrix for state with Bloch vector $\boldsymbol{r}$ is

$$\rho = \frac{1}{2}(I_2 + \boldsymbol{r} \cdot \sigma)$$

where $\boldsymbol{r} \cdot \sigma = r_1\sigma_1 + r_2\sigma_2 + r_3\sigma_3 = x\sigma_1 + y\sigma_2 + z\sigma_3$.

Proposition. State is mixed iff its Bloch vector $\boldsymbol{r}$ satisfies $|\boldsymbol{r}| < 1$.

Proposition. For any density matrix ρ defined by Bloch vector $\boldsymbol{r}$,

$$\operatorname{tr}(\rho^2) = \frac{1}{2}(1 + |\boldsymbol{r}|^2)$$

3.3. Time evolution of a qubit

Remark. Unitary transformations of a qubit correspond to rotations of points on/in Bloch sphere about the origin, representing the fact that unitary transformations cannot transform pure states to mixed states.

Remark. Measurements transform any state to a pure state.

Proposition. $\mathrm{tr}(\rho^2)$ is invariant under unitary transformations (time evolution).

- $\mathrm{tr}(\rho^2)$ measures how mixed a state is: $\mathrm{tr}(\rho^2) = 1$ for pure states, $\mathrm{tr}(\rho^2) = \frac{1}{2}$ for the most mixed single qubit state, corresponding to the origin: $\boldsymbol{r} = \boldsymbol{0}$, $\rho = \frac{1}{2}I$.

Proposition. Mixing states can never produce a state further from origin than furthest initial state.

Note. There are an infinite number of ways of writing a mixed state as an ensemble of two pure states: any line passing through the point represented by the mixed states intersects with the Bloch sphere twice - the intersection points give the pure states in the ensemble.

Definition. **Trace distance** between density matrices $\hat{\rho}_1$ and $\hat{\rho}_2$ is

$$D(\hat{\rho}_1, \hat{\rho}_2) := \frac{1}{2}\mathrm{tr}|\hat{\rho}_1 - \hat{\rho}_2| = \frac{1}{4}\mathrm{tr}|(\boldsymbol{r_1} - \boldsymbol{r_2}) \cdot \sigma| = \frac{1}{2}\,|\boldsymbol{r_1} - \boldsymbol{r_2}| = \frac{1}{2}\sum_{i=1}^{N}|\lambda_i|$$

where $|\hat{A}| = \sqrt{\hat{A}^\dagger \hat{A}}$, λ_i are the eigenvalues of $\hat{\rho}_1 - \hat{\rho}_2$ (trace distance is equal to sum of eigenvalues since $\hat{\rho}_1 - \hat{\rho}_2$ is Hermitian).

Remark. Trace distance gives notion of distance between two states.

Proposition. Trace distance defines a **metric** on set of density matrices:

- **Non-negative**: $D(\hat{\rho}_1, \hat{\rho}_2) \geq 0$.
- **Separates points**: $D(\hat{\rho}_1, \hat{\rho}_2) = 0 \iff \hat{\rho}_1 = \hat{\rho}_2$.
- **Symmetric**: $D(\hat{\rho}_1, \hat{\rho}_2) = D(\hat{\rho}_2, \hat{\rho}_1)$.
- **Triangle inequality**: $D(\hat{\rho}_1, \hat{\rho}_3) \leq D(\hat{\rho}_1, \hat{\rho}_2) + D(\hat{\rho}_2, \hat{\rho}_3)$

3.4. Pauli matrices

Definition. **Levi-Cevita** tensor ε_{ijk} is defined for $\{i, j, k\} \subseteq \{1, 2, 3\}$ as:

- $\varepsilon_{123} := \varepsilon_{231} := \varepsilon_{312} := 1$.
- $\varepsilon_{321} := \varepsilon_{132} := \varepsilon_{213} := -1$.
- $\varepsilon_{ijk} := 0$ otherwise.

Proposition. Pauli matrices satisfy following properties:

- **Hermitian**: $\sigma_i^\dagger = \sigma_i$.
- **Traceless**: $\mathrm{tr}(\sigma_i) = 0$.
- $[\sigma_i, \sigma_j] = \sigma_i\sigma_j - \sigma_j\sigma_i = 2i\varepsilon_{ijk}\sigma_k$.
- $\{\sigma_i, \sigma_j\} = \sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij}I_2$.
- $\sigma_i\sigma_j = \delta_{ij}I_2 + i\varepsilon_{ijk}\sigma_k$.
- They form a basis for vector space of 2×2 Hermitian traceless matrices over $\mathbb{R}$.

Proposition. Define measurement operators X, Y, Z as

$$X := \frac{1}{2}(I_2 - \sigma_1), \quad Y := \frac{1}{2}(I_2 - \sigma_2), \quad Z := \frac{1}{2}(I_2 - \sigma_3)$$

X, Y and Z have their eigenvectors as the six special Bloch states, with eigenvalues 0 or 1:

$$\begin{aligned} X|+\rangle &= 0|+\rangle, \quad X|-\rangle = 1|-\rangle, \\ Y|L\rangle &= 0|L\rangle, \quad Y|R\rangle = 1|R\rangle, \\ Z|0\rangle &= 0|0\rangle, \quad Z|1\rangle = 1|1\rangle \end{aligned}$$

Proposition. Exponentials of Pauli matrices are unitary matrices: $\forall \alpha \in \mathbb{R}$,

$$\begin{aligned} \exp(i\alpha\sigma_1) &= \cos(\alpha)I_2 + i\sin(\alpha)\sigma_1, \\ \exp(i\alpha\sigma_2) &= \cos(\alpha)I_2 + i\sin(\alpha)\sigma_2, \\ \exp(i\alpha\sigma_3) &= \cos(\alpha)I_2 + i\sin(\alpha)\sigma_3 \end{aligned}$$

Proposition. For $\alpha \in \mathbb{R}$, $\boldsymbol{n} \in \mathbb{R}^3$, $|\boldsymbol{n}|^2 = 1$,

$$U_\alpha(\boldsymbol{n}) := \exp(i\alpha\boldsymbol{n} \cdot \sigma) = \cos(\alpha)I_2 + i\sin(\alpha)\boldsymbol{n} \cdot \sigma$$

is unitary transformation. If density matrix $\rho = \frac{1}{2}(I_2 + \boldsymbol{r} \cdot \sigma)$ evolves with time according to this operator, then

$$\rho \to U_\alpha(\boldsymbol{n})\rho U_\alpha(\boldsymbol{n})^\dagger = \frac{1}{2}(I_2 + (R_\alpha(\boldsymbol{n})\boldsymbol{r}) \cdot \sigma)$$

where $R_\alpha(\boldsymbol{n})$ is 3×3 orthogonal matrix corresponding to rotation of angle 2α about axis in the direction of $\boldsymbol{n}$.

4. Bipartite systems

4.1. Tensor products

Definition. **Tensor product** $|\varphi\rangle \otimes |\psi\rangle$ in $H_1 \otimes H_2$ satisfies:

- **Scalar multiplication**: $c(|\varphi\rangle \otimes |\psi\rangle) = (c|\varphi\rangle) \otimes |\psi\rangle = |\varphi\rangle \otimes (c|\psi\rangle)$.
- **Linearity**:
  - $a|\psi\rangle \otimes |\varphi_1\rangle + b|\psi\rangle \otimes |\varphi_2\rangle = |\psi\rangle \otimes (a|\varphi_1\rangle + b|\varphi_2\rangle)$.
  - $a|\psi_1\rangle \otimes |\varphi\rangle + b|\psi_2\rangle \otimes |\varphi\rangle = (a|\psi_1\rangle + b|\psi_2\rangle) \otimes |\varphi\rangle$.

Definition. Induced inner product on $H_1 \otimes H_2$ is defined as

$$(\langle\psi_1| \otimes \langle\varphi_1|)(|\psi_2\rangle \otimes |\varphi_2\rangle) = \langle\psi_1|\psi_2\rangle\langle\varphi_1|\varphi_2\rangle$$

Proposition. For bases $\{|v_i\rangle : i \in [N_1]\}$ for H_1 and $\{|w_j\rangle : j \in [N_2]\}$ for H_2, $\{|v_i\rangle \otimes |w_j\rangle, i \in [N_1], j \in [N_2]\}$ is basis for $H_1 \otimes H_2$ and is orthonormal if $\{|v_i\rangle\}$ and $\{|v_j\rangle\}$ are orthonormal.

Definition. Most general vector $|\psi\rangle \in H_1 \otimes H_2$ can be expressed as

$$|\psi\rangle = \sum_{i \in [N_1],\ j \in [N_2]} c_{i,j} |v_i\rangle \otimes |v_j\rangle$$

Generally, this cannot be written as a tensor product $|\psi\rangle \otimes |\varphi\rangle$. If it can be, it is a **separable** state. If not, it is **entangled**.

Definition. Hilbert space of N-qubit system is 2^N-dimensional Hilbert space $H_N = H_q^{\otimes N}$ where H_q is a single qubit Hilbert space.

Example. Let $H_3 = H_q \otimes H_q \otimes H_q$. Operator $\hat{I} \otimes \hat{\sigma_1} \otimes \hat{I}$ acts on the second qubit and leaves the other two invariant.

4.2. Linear operators and local unitary operations

Definition. Linear operators on $H_1 \otimes H_2$ are linear combinations of $\hat{A} \otimes \hat{B}$, where

$$(\hat{A} \otimes \hat{B})(|\psi\rangle \otimes |\varphi\rangle) := (\hat{A}|\psi\rangle) \otimes (\hat{B}|\varphi\rangle)$$

Proposition. Properties of tensor product of linear operators:

- $\hat{A} \otimes \hat{B} + \hat{C} \otimes \hat{B} = (\hat{A} + \hat{C}) \otimes \hat{B}$.
- $\hat{A} \otimes \hat{B} + \hat{A} \otimes \hat{D} = \hat{A} \otimes (\hat{B} + \hat{D})$.
- $(\hat{A} \otimes \hat{B})^\dagger = \hat{A}^\dagger \otimes \hat{B}^\dagger$.
- $(\hat{A} \otimes \hat{B})(\hat{C} \otimes \hat{D}) = (\hat{A}\hat{C} \otimes \hat{B}\hat{D})$.
- $\mathrm{tr}_{\mathcal{H}_A \otimes \mathcal{H}_B}(\hat{A} \otimes \hat{B}) = \mathrm{tr}_{\mathcal{H}_A}(\hat{A})\,\mathrm{tr}_{\mathcal{H}_B}(\hat{B})$.

In particular, tensor product of linear operators preserves unitarity, Hermiticity, positivity, and tensor product of two projectors is a projector.

Definition. Bipartite system is system described Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ which can be partitioned (separated) into two subsystems A and B, described by Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. Alice has full control over system A, Bob has full control over system B, neither can control the other's system.

Definition. For bipartite system, **local operations (LO)** are of the form $\hat{U}_A \otimes \hat{I}$ (for Alice) or $\hat{I} \otimes \hat{U}_B$ (for Bob) where $\hat{U}_A$ and $\hat{U}_B$ are unitary operators or measurement operators.

Proposition. $\hat{U}_A \otimes \hat{I}$ and $\hat{I} \otimes \hat{U}_B$ commute: $[\hat{U}_A \otimes \hat{I}, \hat{I} \otimes \hat{U}_B] = 0$, and their product is $\hat{U}_A \otimes \hat{U}_B$.

Theorem. Any unitary transformation $\hat{U}_A \otimes \hat{U}_B$ (i.e. using LO) acting on separable state $|\psi\rangle \otimes |\varphi\rangle$ produces another separable state: $\hat{U}_A|\psi\rangle \otimes \hat{U}_B|\varphi\rangle$. In particular, an entangled state cannot be created from a separable state.

Definition. A mixed state is **separable** iff it is an ensemble of separable states, and **entangled** otherwise.

Definition. **Density matrix** of **separable pure state** $|\Psi\rangle = |\psi\rangle \otimes |\varphi\rangle$ is

$$\hat{\rho} := |\Psi\rangle\langle\Psi| = (|\psi\rangle \otimes |\varphi\rangle)(\langle\psi| \otimes \langle\varphi|) = (|\psi\rangle\langle\psi|) \otimes (|\varphi\rangle\langle\varphi|) = \hat{\rho}_A \otimes \hat{\rho}_B$$

where $\hat{\rho}_A = |\psi\rangle\langle\psi|$ and $\hat{\rho}_B = |\varphi\rangle\langle\varphi|$.

Definition. **Density matrix** of **separable mixed state** is

$$\hat{\rho} := \sum_{i=1}^{M} p_i \hat{\rho}_A^{(i)} \otimes \hat{\rho}_B^{(i)}$$

where $\{\hat{\rho}_A^{(i)}\}$ are mixed or pure states of first system, $\{\hat{\rho}_B^{(i)}\}$ are mixed or pure states of second system.

4.3. Matrix representation

Definition. **Tensor product** of two vectors is given by e.g.

$$\begin{bmatrix}1\\2\\3\end{bmatrix} \otimes \begin{bmatrix}4\\5\end{bmatrix} = \begin{bmatrix}1\begin{bmatrix}4\\5\end{bmatrix}\\ 2\begin{bmatrix}4\\5\end{bmatrix}\\ 3\begin{bmatrix}4\\5\end{bmatrix}\end{bmatrix} = \begin{bmatrix}4\\5\\8\\10\\12\\15\end{bmatrix}$$

The expression is similar for matrices:

$$\begin{bmatrix}1&2\\3&4\end{bmatrix} \otimes \begin{bmatrix}5&6\\7&8\end{bmatrix} = \begin{bmatrix}1\begin{bmatrix}5&6\\7&8\end{bmatrix} & 2\begin{bmatrix}5&6\\7&8\end{bmatrix}\\ 3\begin{bmatrix}5&6\\7&8\end{bmatrix} & 4\begin{bmatrix}5&6\\7&8\end{bmatrix}\end{bmatrix} = \begin{bmatrix}5&6&10&12\\7&8&14&16\\15&18&20&24\\21&24&28&36\end{bmatrix}$$

Definition. **Controlled `NOT` (`CNOT`)** operator acts on $H_2 = H_q \otimes H_q$ and is defined as

$$U = \frac{I_2 + \sigma_3}{2} \otimes I_2 + \frac{I_2 - \sigma_3}{2} \otimes \sigma_1$$

We have $U|00\rangle = |00\rangle$, $U|01\rangle = |01\rangle$, $U|10\rangle = |11\rangle$, $U|11\rangle = |10\rangle$.

4.4. Local measurements

Definition. For bipartite system, **local measurements** are Hermitian operators of the form $\hat{F} = \hat{F}_A \otimes \hat{I}$ for Alice and $\hat{G} = \hat{I} \otimes \hat{G}_B$ for Bob.

Notation. Projection operators of $\hat{F}_A$ and $\hat{G}_B$ for eigenvalues λ_i and μ_j are denoted $\hat{F}_{Ai}$ and $\hat{G}_{Bj}$.

Remark. In the full system $H_A \otimes H_B$, $\hat{F}$ and $\hat{G}$ are degenerate, with degeneracy given by dimension of other subsystem, i.e. $\dim(\mathcal{H}_B)$ for Alice's observable and $\dim(\mathcal{H}_A)$ for Bob's. Assuming no degeneracy in their own system, corresponding projection operators in full system are

$$\hat{F}_i = \hat{F}_{Ai} \otimes \hat{I} = \sum_{j=1}^{N_2} |v_i\rangle\langle v_i| \otimes |w_j\rangle\langle w_j|$$

$$\hat{G}_j = \hat{I} \otimes \hat{G}_{Bj} = \sum_{i=1}^{N_1} |v_i\rangle\langle v_i| \otimes |w_j\rangle\langle w_j|$$

Note. Since $[\hat{F}, \hat{G}] = 0$, these measurements are compatible so final state is eigenstate of both $\hat{F}$ and $\hat{G}$. Probability of an outcome occuring is not affected by whether Alice or Bob measures first (or simultaneously).

Example. Let $\{|v_i\rangle\}, \{|w_j\rangle\}$ be orthonormal eigenstates of operators $\hat{F}_A$ and $\hat{G}_B$ with non-degenerate eigenvalues $\{\lambda_i\}$ and $\{\mu_j\}$, $|\Psi\rangle = \sum_{i\in[N_1],\, j\in[N_2]} \gamma_{ij} |v_i\rangle \otimes |w_j\rangle$ be entangled state, define

$$\alpha_i := \left(\sum_{j=1}^{N_2} |\gamma_{ij}|^2\right)^{1/2}, \quad \beta_j := \left(\sum_{i=1}^{N_1} |\gamma_{ij}|^2\right)^{1/2}$$

and define auxiliary states (set $|\psi_j\rangle = \mathbf{0}$ when $\beta_j = 0$ and $|\varphi_i\rangle = \mathbf{0}$ when $\alpha_i = 0$):

$$|\psi_j\rangle := \frac{1}{\beta_j} \sum_{i=1}^{N_1} \gamma_{ij} |v_i\rangle \in \mathcal{H}_A, \quad |\varphi_i\rangle := \frac{1}{\alpha_i} \sum_{j=1}^{N_2} \gamma_{ij} |w_j\rangle \in \mathcal{H}_B$$

$$\Longrightarrow |\Psi\rangle = \sum_{i=1}^{N_1} \alpha_i |v_i\rangle \otimes |\varphi_i\rangle = \sum_{j=1}^{N_2} \beta_j |\psi_j\rangle \otimes |w_j\rangle$$

If Alice measures $\hat{F}$ with result λ_i, entangled state $|\Psi\rangle$ collapses to separable state

$$|\Psi\rangle \to \hat{F}_i |\Psi\rangle = (\hat{F}_{Ai} \otimes \hat{I}) |\Psi\rangle \sim |v_i\rangle \otimes |\varphi_i\rangle$$

So Bob's state depends on the result of Alice's measurement.

4.5. Reduced density matrix

Definition. For operator $\hat{C} \otimes \hat{D} \in \mathrm{End}(\mathcal{H}_A \otimes \mathcal{H}_B)$, **partial trace** over $\mathcal{H}_A$ and $\mathcal{H}_B$, $\mathrm{tr}_A : \mathrm{End}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathrm{End}(\mathcal{H}_B)$ and $\mathrm{tr}_B : \mathrm{End}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathrm{End}(\mathcal{H}_A)$, are

$$\mathrm{tr}_A(\hat{C} \otimes \hat{D}) := \mathrm{tr}(\hat{C})\hat{D}, \quad \mathrm{tr}_B(\hat{C} \otimes \hat{D}) := \mathrm{tr}(\hat{D})\hat{C}$$

Definition. For bipartite system, the **reduced density matrix** of a subsystem is partial trace of density matrix over other subsystem. So for bipartite system,

$$\hat{\rho}_A := \mathrm{tr}_B(\hat{\rho}), \quad \hat{\rho}_B := \mathrm{tr}_A(\hat{\rho})$$

Proposition. We have $\mathrm{tr}(\hat{A} \otimes \hat{B}) = \mathrm{tr}_A \mathrm{tr}_B(\hat{A} \otimes \hat{B}) = \mathrm{tr}_B \mathrm{tr}_A(\hat{A} \otimes \hat{B})$.

Note. A reduced matrix describes one subsystem, assuming no knowledge of the other system.

Proposition.

- $\hat{\rho}_A$ is invariant under all local operations in system B (for measurements, this is provided Alice does not learn about the result of the measurement in system B).
- Under unitary transformations $\hat{U}$ in system A, $\hat{\rho}_A$ transforms as normal: $\hat{\rho}_A \to \hat{U}\hat{\rho}_A\hat{U}^\dagger$.
- Local measurements in system A can be described by $\hat{\rho}_A$ and operators acting on $\mathcal{H}_A$: $\mathrm{tr}_B(\hat{F}_i\hat{\rho}\hat{F}_i) = \hat{F}_{Ai}\hat{\rho}_A\hat{F}_{Ai}$.

Theorem. If $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is pure state, then $\hat{\rho}_A$ is pure iff $|\Psi\rangle$ is separable.

Corollary. If spectrum of $\hat{F}_A$ is non-degenerate then measuring $\hat{F}_A$ in system $\mathcal{H}_A$ produces separable state on system $\mathcal{H}_A \otimes \mathcal{H}_B$, i.e. **measurement destroys entanglement**.

Note. Entanglement does not violate causality (does not allow communication faster than the speed of light). i.e., if Alice makes a local measurement on an entangled system, Bob cannot detect this, even though the reduced density matrix for his system has changed.

4.6. Classical communication

- Alice and Bob can use classical communication (CC) to communicate results of measurements of their own subsystem. If the state was initially entangled, Bob communicating a measurement to Alice would give Alice information about her subsystem.

Definition. **LOCC** is when Alice and Bob can use local operations (LO) and classical communication.

5. Entanglement applications

5.1. Bell states

Proposition. Measurements of entanglement:

- Let $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. If $|\Psi\rangle = a|0\rangle \otimes |\varphi\rangle + b|1\rangle \otimes |\varphi\rangle$ for some $a, b \in \mathbb{C}$, $|\varphi\rangle \in \mathcal{H}_B$, then $|\Psi\rangle$ is separable, otherwise entangled.
- If reduced density matrix of either subsystem gives a pure state $(\operatorname{tr}(\rho^2) = 1)$ then state is separable. If it gives a mixed state $(\operatorname{tr}(\rho^2) < 1)$, state is entangled.
- $\operatorname{tr}(\rho_A^2) = \operatorname{tr}(\rho_B^2)$ gives measure of entanglement, with max value 1 for no entanglement, min value 1/2 (for single qubit subsystem) for maximally entangled states.

Definition. **Bell states** are defined as, for $x, y \in \{0, 1\}$,

$$|\beta_{xy}\rangle := \frac{1}{\sqrt{2}}(|0\rangle \otimes |y\rangle + (-1)^x |1\rangle \otimes |\overline{y}\rangle)$$

Proposition. Bell states are maximally entangled (trace of reduced density matrix of both sides is $\frac{1}{2}$) and form an orthonormal basis.

- Bell state basis is related to standard basis by unitary transformation, but Bell states can't be created from the separable standard basis by any LOCC process, since unitary transformations between them are not of form $\hat{U}_A \otimes \hat{U}_B$ (since this preserves separability), and measurements always produce a separable state.
- Alice and Bob can individually transform any Bell state to any other Bell state by the unitary operators $\hat{U}_{xy} \otimes \hat{I}$ and $\hat{I} \otimes \hat{U}_{xy}$ respectively:

$$\left(\hat{U}_{xy} \otimes \hat{I}\right)|\beta_{00}\rangle = \left(\hat{I} \otimes \hat{U}_{xy}\right)|\beta_{00}\rangle = |\beta_{xy}\rangle$$

where

$$U_{00} = I_2, \quad U_{01} = \sigma_1, \quad U_{10} = \sigma_3, \quad U_{11} = i\sigma_2$$

5.2. Superdense coding

- Qubit can be used instead of classical bit: $|0\rangle$ corresponds to the bit 0, $|1\rangle$ corresponds to the bit 1. In this case, the qubit can be measured with probability 1 with the measurement operator $Z = \frac{1}{2}(I_2 - \sigma_3)$, since $Z|0\rangle = 0|0\rangle$, $Z|1\rangle = 1|1\rangle$ so measurement with outcome 0 means state is $|0\rangle$ with probability 1, measurement with outcome 1 means state is $|1\rangle$ with probability 1.
- Alice can prepare the qubit to represent the classical bit to send to Bob: prepare any state $|\psi\rangle$ and measure on it with operator $\frac{1}{2}(I_2 - \sigma_3)$. Outcome is 0 or 1 - if outcome is equal to the bit x she wants to send, $|\psi\rangle$ has been projected to $|x\rangle$, so send this state to Bob. Otherwise, perform unitary transformation $\sigma_1|\overline{x}\rangle = |x\rangle$ and send this state to Bob.
- **Superdense coding**:
  - Superdense coding allows one qubit to transmit two classical bits of information.
  - Alice and Bob share state $|\beta_{00}\rangle$.
  - Alice applies operation $\hat{U}_{xy} \otimes \hat{I}$ to whole system where $(xy)_2$ is the two bit message she wants to send (this just acts on her qubit). Note that this does not transmit any information to Bob, as his reduced density matrix is $\rho_B = \frac{1}{2}I$ before and after the transformation.
  - Alice sends her qubit to Bob. Then Bob has the full Bell state $|\beta_{xy}\rangle$ (he has both qubits). Bob then applies a measurement which has the four Bell states as eigenstates, which gives him the eigenvalue with probability 1, e.g. he measures

$$\hat{B} = 0|\beta_{00}\rangle\langle\beta_{00}| + 1|\beta_{01}\rangle\langle\beta_{01}| + 2|\beta_{10}\rangle\langle\beta_{10}| + 3|\beta_{11}\rangle\langle\beta_{11}|$$

5.3. No-cloning theorem

Theorem (No-cloning theorem). In quantum mechanics, it is impossible to clone an unknown state $|\psi\rangle$. More precisely, it is impossible to perform transformation $|\psi\rangle \otimes |\varphi\rangle \rightarrow |\psi\rangle \otimes |\psi\rangle$ for an arbitrary unknown state $|\psi\rangle$ and fixed initial state $|\varphi\rangle$.

5.4. Teleportation

Definition. **Hadamard gate** is transformation given by operator

$$U_H := \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{\sqrt{2}}(\sigma_1 + \sigma_3)$$

We have $\hat{U}_H|0\rangle = |+\rangle$, $\hat{U}_H|1\rangle = |-\rangle$.

Definition. **Teleportation** is process of transferring quantum state $|\psi\rangle$ without using quantum communication (i.e. only using LOCC). It is as follows:

- Alice has state $|\psi\rangle = a|0\rangle + b|1\rangle$, Alice and Bob share Bell state $|\beta_{00}\rangle$, so full system state is

$$\begin{aligned}|\psi\rangle \otimes|\beta_{00}\rangle &= \frac{1}{\sqrt{2}}|\psi\rangle \otimes|0\rangle \otimes|0\rangle + \frac{1}{\sqrt{2}}|\psi\rangle \otimes|1\rangle \otimes|1\rangle \\ &= \frac{1}{\sqrt{2}}(a|000\rangle + a|011\rangle + b|100\rangle + b|111\rangle)\end{aligned}$$

Alice has first two qubits, Bob has third.

- Alice performs `CNOT` on her two qubits, transforming state to

$$\frac{1}{\sqrt{2}}(a|000\rangle + a|011\rangle + b|110\rangle + b|101\rangle)$$

`CNOT` operator is not of form $A \otimes B$ so it entangles Alice's qubits.

- Alice applies Hadamard gate to her system:

$$\hat{U}_H \otimes \hat{I} \otimes \hat{I}\frac{1}{\sqrt{2}}(a|000\rangle + a|011\rangle + b|100\rangle + b|111\rangle) = \frac{1}{2}\sum_{x,y}|x\rangle \otimes|y\rangle \otimes \hat{U}_{xy}|\psi\rangle$$

- Alice measures with operator Z on both her qubits, giving measurement $(xy)_2$, causing state to collapse to $|x\rangle \otimes|y\rangle \otimes \hat{U}_{xy}|\psi\rangle$.
- Alice uses CC to send $(xy)_2$ to Bob. Bob then performs transformation $\hat{U}_{xy}^{-1} = \hat{U}_{xy}^{\dagger}$ so his state becomes $|\psi\rangle$.

5.5. Quantum key distribution (QKD)

Definition. Let message M and secret key K be n-bit integers, K is shared by Alice and Bob, where each bit of k has value 0 or 1 with equal probability. **One-time pad encryption** is as follows:

- Alice produces encrypted message $C = M \oplus K$, where $\oplus$ is bitwise addition $\bmod 2$ (also bitwise `XOR`).
- Alice transmits C to Bob. Bob decrypts message by calculating

$$C \oplus K = (M \oplus K) \oplus K = M \oplus (K \oplus K) = M \oplus 0 = M$$

- It is important that K is at least as long as M and is never reused.
- Drawback is that K might be very long, and must be transmitted securely prior to communication.

Definition. **BB84** protocol for transmitting secret key is as follows:

- Alice chooses random bit $x \in \{0, 1\}$ with equal probability, makes random choice of X or Z with equal probability, then prepares qubit state according to the outcome:

$$(0, Z) \mapsto|0\rangle, \quad (1, Z) \mapsto|1\rangle, \quad (0, X) \mapsto|+\rangle, \quad (1, X) \mapsto|-\rangle$$

and sends this qubit to Bob using quantum communication.

- Bob randomly chooses X or Z with equal probability, then measures qubit with measurement operator $\frac{1}{2}(I - \sigma_X)$ or $\frac{1}{2}(I - \sigma_Z)$.
- This process is repeated enough to generate a sufficiently long key.
- Alice and Bob publicly reveal their choices of X or Z for each qubit (must be after Bob receives the qubit), discarding all qubits for which same choice was not made.

When same choice is made for qubit, Alice's choice of qubit will match with Bob's measurement.

- **Security of BB84**:
  - If Eve intercepts qubit, she must measure it to obtain information from it. But the four possible states are not all orthogonal, so Eve cannot make measurement which is guaranteed to distinguish them.
  - If Eve measures with Z and Alice chose Z, Eve would correctly measure the qubit. But if Alice chose X, Eve would measure 0 or 1 with equal probability, and forward the same random qubit $|0\rangle$ or $|1\rangle$ to Bob. If Bob measures with X, result is discarded anyway. If Bob measures with Z, measurement is same random result as Eve's measurement, so differs from Alice's key half the time.
  - So for each (non-discarded) bit of key Eve intercepts and measures, probability that Alice and Bob's value differs is $\frac{1}{4}$, so currently Eve expects to know $\frac{3}{4}$ of the key, which is insecure. So Alice and Bob compare random subset of their keys and estimate error rate.
  - If rate too high, they assume interference from Eve, discard the key and repeat entire process again.

5.6. Bell inequalities

Definition. **Local realism** is a property of a system:

- **Locality**: an effect at one point can be detected at another point only if something travels between those two points (no faster than the speed of light).
- **Realism**: measurements must be deterministic, i.e. measurements tell us a property of the system.
- **CHSH Bell-inequality**:
  - Let system have observables Q, R, S, T which takes values ± 1. Realism states that any system state must have specific values for these, (q, r, s, t).
  - Take large number of system states and measure $QS + RS + QT - RT$ for each, calculate mean which gives estimate of expectation $\mathbb{E}(QS + RS + QT - RT)$.
  - Now $Q = \pm R$, so either $(Q + R)S = 0$ and $(Q - R)T = \pm 2$ or $(Q + R)S = \pm 2$ and $(Q - R)T = 0$, hence $QS + RS + QT - RT = \pm 2$, and

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

- Consider following experiment:
  - Charlie is in middle of Alice and Bob, who are separated arbitrarily.
  - Charlie prepares many Bell states $|\beta_{11}\rangle$ and sends one qubit of each simultaneously to Alice and Bob, so they receive them at same time.
  - Alice randomly chooses Q or R and makes that measurement on her qubit, Bob does same for random S or T. Assuming locality, it is impossible that Alice or Bob's measurement affects the other by an influence of finite speed.
  - If quantum mechanics satisfied local realism, Alice's and Bob's results are predetermined by a hidden variable describing Charlie's Bell state.
  - Alice and Bob record measurement operator and result for each qubit, then compute $\mathbb{E}(QS)$, $\mathbb{E}(RS)$, $\mathbb{E}(QT)$, $\mathbb{E}(RT)$.

- Measurement operators are given by

$$Q = \sigma_1 \otimes I_2, R = \sigma_3 \otimes I_2, \quad S = I_2 \otimes \frac{-1}{\sqrt{2}}(\sigma_1 + \sigma_3), T = I_2 \otimes \frac{-1}{\sqrt{2}}(\sigma_1 - \sigma_3)$$

- These give $\mathbb{E}(QS) = \mathbb{E}(RS) = \mathbb{E}(QT) = -\mathbb{E}(RT) = \frac{1}{\sqrt{2}}$, giving $\mathbb{E}(QS) + \mathbb{E}(RS) + \mathbb{E}(QT) - \mathbb{E}(RT) = 2\sqrt{2} > 2$, violating CHSH inequality.
- Experimental data confirms this violation, showing nature isn't described by theory obeying local realism, and nature is consistent with quantum mechanics.

6. Information theory

6.1. Classical information and Shannon entropy

Definition. Let X be random variable representing a message, $p(x) = \mathbb{P}(X = x)$ **Shannon entropy** is

$$H(X) \coloneqq -\sum_x p(x) \log_2(p(x))$$

where conventionally $0 \log 0 = 0$.

Theorem (Shannon's noiseless coding theorem). $H(X)$ gives lower bound on average number of bits needed to encode message X.

Definition. **Joint entropy** is

$$H(X, Y) \coloneqq -\sum_{x,y} p(x, y) \log_2(p(x, y))$$

Proposition. Joint entropy obeys **subadditivity**:

$$H(X, Y) \leq H(X) + H(Y)$$

with equality iff X and Y are independent variables, i.e. when $p(x, y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$.

Definition. **Relative entropy of $p(x)$ to $q(x)$** is defined for two random variables which take same values but with different distributions $p(x)$ and $q(x)$:

$$\begin{aligned} H(p(x) \parallel q(x)) &\coloneqq \sum_x (p(x) \log_2(p(x)) - p(x) \log_2(q(x))) \\ &= -H(X) - \sum_x p(x) \log_2(q(x)) \end{aligned}$$

Proposition. Relative entropy is non-negative and

$$H(p(x) \parallel q(x)) = 0 \Longleftrightarrow \forall x, p(x) = q(x)$$

Remark. Relative entropy can diverge if for some x, $q(x) = 0$ and $p(x) \neq 0$

Definition. **Conditional entropy** is

$$H(X|Y) \coloneqq H(X, Y) - H(Y) \leq H(X)$$

Definition. **Mutual information** of X and Y is

$$H(X : Y) := H(X) + H(Y) - H(X, Y) \geq 0$$

6.2. Quantum entropy

Definition. **Von Neumann entropy** of quantum state with density operator $\hat{\rho}$ is

$$S(\hat{\rho}) := -\operatorname{tr}(\hat{\rho} \log_2(\hat{\rho})) = -\sum_i p_i \log_2(p_i)$$

where $\hat{\rho} = \sum_i p_i |i\rangle\langle i|$, $|i\rangle$ are eigenstates of $\hat{\rho}$, p_i are eigenvalues of $\hat{\rho}$. $S(\hat{\rho})$ is Shannon entropy of ensemble of pure states described by $\hat{\rho}$.

Note. To compute $\log_2(\hat{\rho})$, diagonalise $\hat{\rho}$ (use spectral decomposition) and take $\log_2$ of each diagonal element (use here the convention $\log_2(0) = 0$).

Remark. For pure state, $S(\hat{\rho}) = -1 \log_2(1) = 0$.

Definition. **(quantum) relative entropy** is measure of distance between two states:

$$S(\hat{\rho}_1 \parallel \hat{\rho}_2) := \operatorname{tr}(\hat{\rho}_1 \log_2(\hat{\rho}_1)) - \operatorname{tr}(\hat{\rho}_1 \log_2(\hat{\rho}_2))$$

Proposition. $S(\hat{\rho}_1 \parallel \hat{\rho}_2) \geq 0$ with equality iff $\hat{\rho}_1 = \hat{\rho}_2$.

Definition. For bipartite system $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ described by density matrix $\hat{\rho}$ and reduced density matrices $\hat{\rho}_A$ and $\hat{\rho}_B$, define

$$S(A) := S(\hat{\rho}_A), \quad S(B) := S(\hat{\rho}_B), \quad S(A, B) := S(\hat{\rho})$$

where $S(A, B)$ is **(quantum) joint entropy** of A and B.

Definition. **(quantum) conditional entropy** of A and B is

$$S(A \mid B) := S(A, B) - S(B)$$

Remark. Unlike classical conditional entropy, quantum conditional entropy can be negative, e.g. if $\hat{\rho}$ describes pure state, $S(A, B) = 0$ but if entangled, $\hat{\rho}_B$ is not pure state so $S(B) > 0$.

Definition. **(Quantum) mutual information** is

$$I(A : B) = S(A : B) := S(A) + S(B) - S(A, B)$$

Remark. When $\hat{\rho}$ is pure state, $S(A) = S(B)$ so $I(A : B) = 2S(A)$. So entanglement can be interpreted as mutual information: information shared by A and B and not in either one alone.

Definition. **Entanglement entropy** is $S(A) = S(B)$ (these are equal since both reduced density matrices have same non-zero eigenvalues).

7. Classical computing

7.1. Basic gates

Notation. Input for circuit diagrams has most significant bit at the top, circuits are read left to right, with last operation on the right.

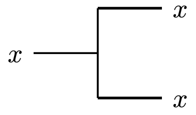
Definition. **(logical) gate** is function mapping bits to bits.

Definition. Simplest gates are $f : \{0,1\} \to \{0,1\}$:

- **Identity gate**: $\mathrm{id}(x) := x$.
- $c_0(x) := 0$.
- $c_1(x) := 1$.
- NOT **gate**: $\mathrm{NOT}(x) = \overline{x}$.

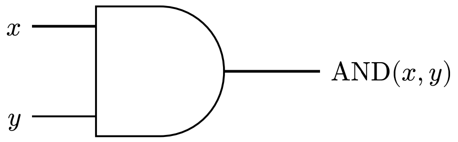
Definition. FANOUT **gate** is defined as

$$\mathrm{FANOUT} : \{0,1\} \to \{0,1\}^2, \quad \mathrm{FANOUT}(x) := (x, x)$$



Definition. AND **gate** is given by its **truth table**:

	0	1
0	0	0
1	0	1



Definition. **OR gate** is given by its **truth table**:

	0	1
0	0	1
1	1	1

Remark. AND and OR are not reversible (invertible) so cannot be implemented by unitary operators.

- **Landauer's principle**: energy E required to erase one bit satisfies

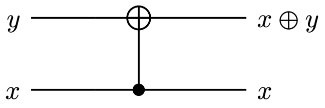
$$E \geq k_B T \log(2)$$

where k_B is Boltzmann's constant, T is temperature at which system operates.

Definition. **Controlled** NOT **(CNOT) gate**, $\mathrm{CNOT} : \{0,1\}^2 \to \{0,1\}^2$, is

$$\mathrm{CNOT}(x, y) := \begin{cases} (x, y) & \text{if } x = 0 \\ (x, \mathrm{NOT}(y)) & \text{if } x = 1 \end{cases} = (x, x \oplus y) = (x, x + y \bmod 2)$$

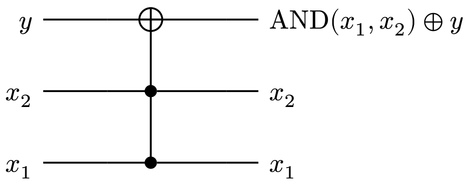
Inverse of CNOT is CNOT. x is **control bit**, y is **target bit**.



Definition. C^nNOT **gate** is defined as

$$C^n\text{NOT}(x_1, ..., x_n, y) := (x_1, ..., x_n, y \oplus \text{AND}(x_1, ..., x_n))$$

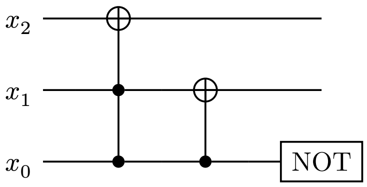
C^nNOT is reversible for all $n \in \mathbb{N}$ and $(C^n\text{NOT})^{-1} = C^n\text{NOT}$. For $n = 2$, CCNOT gate is called a **Toffoli gate**.



Definition. NAND **gate** is defined as

$$\text{NAND}(x, y) := \text{NOT}(\text{AND}(x, y))$$

Example. Circuit diagram for $f : \mathbb{Z}/8 \to \mathbb{Z}/8$, $f(x) = x + \overline{1}$.

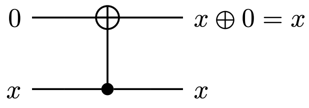


7.2. Universal gate sets

Notation. For $f : \{0, 1\}^m \to \{0, 1\}^n$, can write as

$$f(x_{m-1}, ..., x_0) = (f_{n-1}(x_{m-1}, ..., x_0), ..., f_0(x_{m-1}, ..., x_0))$$

Remark. Can "copy" bits by introducing extra "ancillary" bits and using CNOT gates:



Definition. A **universal gate set (UGS)** is finite set of gates which can construct an arbitrary function $f : \{0, 1\}^n \to \{0, 1\}^m$.

Proposition. $\{\text{NOT}, \text{AND}, \text{OR}, \text{CNOT}\}$ is a universal gate set.

Corollary. $\{\text{CNOT}, \text{AND}\}$ is a universal gate set.

Proposition. {CCNOT} is a minimal (1-gate) UGS for reversible classical computation.

Remark. There is an infinite number of UGSs.

7.3. Computational resources and complexity

Definition. An **algorithm** is a set of instructions (systematic procedure) for computing some output for a given input.

Definition. Resources considered in complexity:

- **Time**: corresponds to numbers of gates in any UGS needed for implementing the circuit.
- **Space**: corresponds to number of bits (lines) in the circuit.
- n denotes size in bits of input.

Example. Computing $\gcd(a, b)$ (assuming WLOG $a \geq b$, $2^{n-1} \leq b < 2^n$ so b has n bits).

- Brute-force algorithm: try all $1 \leq c \leq b$, check if $c \mid a$ and $c \mid b$, return largest such c. Time complexity: $O(2^n)$.
- Euclid's algorithm has time complexity $O(n^3)$ (assuming division and remainder algorithm is $O(n^2)$) (since $r_{i+2} < r_i/2$).

Definition.

- **P** is complexity class of algorithms whose run time is at most polynomial time in n.
- **EXP** is complexity class of algorithms whose run time is at most exponential time in n. $P \subset \text{EXP}$.
- **PSPACE** is class of algorithms which require space at most polynomial in n. $P \subseteq \text{PSPACE}$ (e.g. each line in circuit diagram is assumed to involve at least one gate).
- **NP** is complexity class of algorithms whose output can be verified to be correct in polynomial time, e.g. integer factorisation. Clearly $P \subseteq \text{NP}$.
- **NP-hard** problem is one such that, if you have an oracle for solving them, you can solve any NP problem in polynomial time (NP problems reduce polynomially to NP-hard problems).
- **NP-complete** is complexity class of problems which are NP-hard, e.g. travelling salesman.
- **PP** is class of algorithms which require time at most polynomial in n to return correct answer with probability $> 1/2$.
- **BPP** is class of algorithms which require time at most polynomial in n to return correct answer with probability $> c > 1/2$. $P \subseteq \text{BPP}$.

8. Quantum circuits

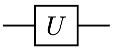
Definition. A **qubit** is a quantum system whose Hilbert space H_1 is 2-dimensional, with basis $\{|0\rangle, |1\rangle\}$. An n-qubit system has 2^n-dimensional Hilbert space $H_n = H_1 \otimes \cdots \otimes H_1$. The **computational basis** for H_n is

$$\{|0\rangle, ..., |2^n - 1\rangle\}$$

where $|k\rangle = |(k_{n-1}...k_0)_2\rangle$ corresponds to $|k_{n-1}\rangle \otimes \cdots \otimes |k_0\rangle$.

Definition. **Quantum gate** is unitary map from H_n to H_n.

Notation. Let X, Y, Z denote Pauli matrices $\sigma_1, \sigma_2, \sigma_3$ respectively.

Notation. A unitary $U : H_1 \to H_1$ is denoted 

Definition. Define the gates

$$S := \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}, \quad T := \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{bmatrix}, \quad H := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

where H is **Hadamard gate**. $S^2 = Z$, $T^2 = S$, $H^2 = I$. $H|0\rangle = |+\rangle$, $H|1\rangle = |-\rangle$.

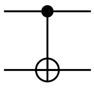
Example. Hadamard gate is useful when constructing uniform superpositions of all basis states:

$$(H|0\rangle) \otimes (H|0\rangle) = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$$

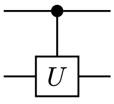
and in general,

$$(H|0\rangle) \otimes \cdots \otimes (H|0\rangle) = \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} |k\rangle$$

Definition. CNOT gate is $\begin{bmatrix} I & 0 \\ 0 & X \end{bmatrix}$. Most significant bit is control bit, least significant bit is target bit.

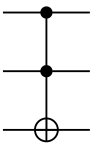


Definition. **Controlled-U** gate, C-U maps $|0\rangle \otimes |\psi\rangle = |0\rangle \otimes |\psi\rangle$ and $|1\rangle \otimes |\psi\rangle = |1\rangle \otimes (U|\psi\rangle)$.



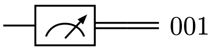
Definition. **CCNOT (Toffoli) gate** is

$$\begin{bmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & X \end{bmatrix}$$



Note: Toffoli gate maps computational basis elements to computational basis elements, and computational basis elements are orthonormal.

Notation. Measurement and classical bits are shown as e.g.



8.1. Universal quantum computation

Proposition. Every $N \times N$ unitary can be written in terms of U_{ij}: “elementary” unitaries acting on (i,i), (i,j), (j,i) and (j,j) entries only, i.e. they are non-trivial in only one 2×2 block (they act non-trivially on a two-dimensional subspace of the Hilbert space, spanned by two basis states $|i-1\rangle$ and $|j-1\rangle$).

Proposition. U is unitary iff its rows are orthonormal iff its columns are orthonormal (with respect to Hermitian inner product).

Example. For unitary

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

we can find unitaries U_1, U_2, U_3 with $U_3 U_2 U_1 U = I$. Choose U_1 to have upper left 2×2 block non-trivial and such that

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

If $b = 0$, set $U_1 = I$. If $b \neq 0$, set

$$U_1 = \begin{bmatrix} \alpha^* & \beta^* & 0 \\ \beta & -\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \alpha := \frac{a}{\sqrt{|a|^2 + |b|^2}}, \beta = \frac{b}{\sqrt{|a|^2 + |b|^2}} \Longrightarrow \beta a - \alpha b = 0$$

Then set

$$\gamma = \frac{a'}{\sqrt{|a'|^2 + |c'|^2}}, \delta = \frac{c'}{\sqrt{|a'|^2 + |c'|^2}}, \quad U_2 = \begin{bmatrix} \gamma^* & 0 & \delta^* \\ 0 & 1 & 0 \\ \delta & 0 & -\gamma \end{bmatrix} \Longrightarrow U_2 U_1 U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e'' & h'' \\ 0 & f'' & j'' \end{bmatrix} =: U_3^\dagger$$

If $U \in U(N)$ is unitary, then can find $N-1$ unitaries $U_1, ..., U_{N-1}$ where U_i is non-trivial in first and $(i+1)$th row such that $U_{N-1} \cdots U_1 U$ has first row and first column

$(1,0,...,0)$ and non-trivial bottom-right $(N-1)\times(N-1)$ block. So it can be reduced entirely by induction, to $\frac{1}{2}N(N-1)$ unitaries.

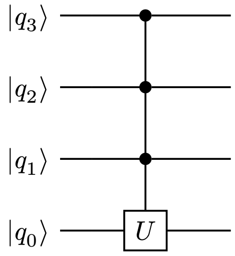
Remark. U acts on n qubits so $N=2^n$, so we need $\approx 4^n$ elementary matrices, so complexity is exponential in number of qubits.

Example. Any 4×4 unitary can be written as product of 6 elementary unitaries:

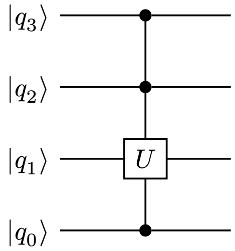
$$U=\begin{bmatrix} * & * & 0 & 0\\ * & * & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix} * & 0 & * & 0\\ 0 & 1 & 0 & 0\\ * & 0 & * & 0\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix} * & 0 & 0 & *\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ * & 0 & 0 & *\end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & 0\\ 0 & * & * & 0\\ 0 & * & * & 0\\ 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & 0\\ 0 & * & 0 & *\\ 0 & 0 & 1 & 0\\ 0 & * & 0 & *\end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & * & *\\ 0 & 0 & * & *\end{bmatrix}$$

Definition. A **multiply-controlled unitary** is an $N\times N$ unitary acting on subspace span$\{|1...10\rangle,|1...11\rangle\}$. It applies a 2×2 unitary to last qubit if all other qubits are 1 and the identity otherwise.

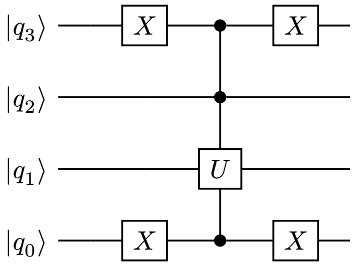
Example. Unitary acting on subspace span$\{|1110\rangle,|1111\rangle\}$ is implemented as



Example. If $i-1$ and $j-1$ differ in single bit, with all other bits 1, this is multiply-controlled unitary with that bit as target, e.g. unitary acting on subspace span$\{|1101\rangle,|1111\rangle\}$ is implemented as



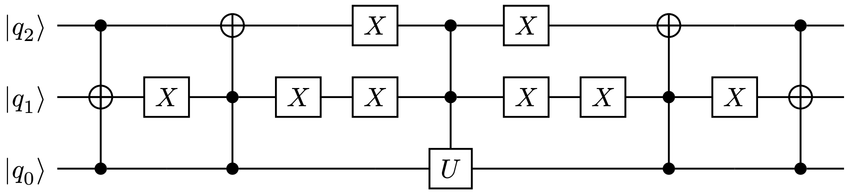
Example. If $i-1$ and $j-1$ differ in single bit but others are not all 1, use NOT gates to reverse the control bits which are 0, e.g. unitary acting on span$\{|0100\rangle,|0110\rangle\}$ is implemented as



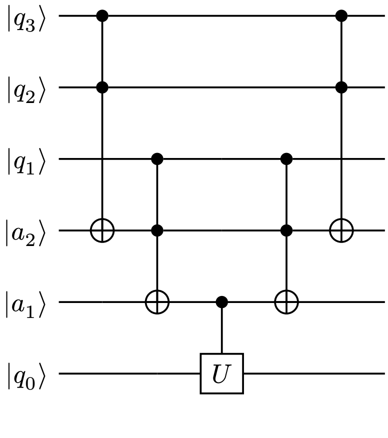
Definition. **Gray code** between $(p_{n-1}...p_0)$ and $(q_{n-1}...q_0)$ is sequence of single bit flips that maps from $(p_{n-1}...p_0)$ to $(q_{n-1}...q_0)$, e.g. a Gray code for 111 and 000 is $111, 101, 001, 000$.

Remark. Gray codes are not unique. (For practical reasons, it is easier to preserve the ordering between first and last, and penultimate and last items in the code.)

Example. If $i-1$ and $j-1$ differ in multiple bits (e.g. $U_{i,j} = U_{8,1}$, $i-1 = 7 = (111)_2$, $j-1 = 1 = (000)_2$), then use a Gray code to flip bits so that all apart from one are the same as $j-1$. First bit flip $111 \to 101$ is implemented as CCNOT. Second bit flip $101 \to 001$ is implemented as CCNOT but if second qubit is 0 instead of 1. Then act with U on subspace $\text{span}\{|001\rangle, |000\rangle\}$ (i.e. on third qubit), then "undo" these CCNOT in reverse order:

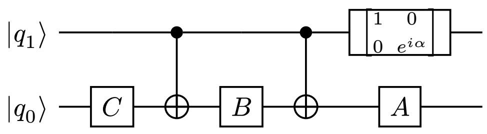


Example. We can implement any multiply-controlled unitary with controlled-unitary (single control qubit) and CCNOT gates, by introducing ancillary bits. e.g. to implemented the multiply-controlled unitary acting on q_0 if $q_1 = q_2 = q_3 = 1$, use ancillary qubits $|a_1\rangle$, $|a_2\rangle$ (initially set to 0):



Proposition. CCNOT can be implemented with H (Hadamard) and T gates (and their Hermitian conjugates).

Lemma. Any single qubit unitary U can be written as $U = e^{i\alpha}AXBXC$ with A, B, C single-qubit (2×2) unitaries, $ABC = 1$, $\alpha \in \mathbb{R}$. In particular, C-U can be implemented as

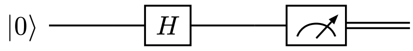


Corollary. Any unitary can be implemented with single-qubit unitaries and CNOT.

Remark. Number of elementary unitaries U_i needed is $O(2^{2n})$. Gray code requires $O(n)$ C^nNOT gates, and representing these multiply-controlled unitaries as controlled-unitaries requires $O(n)$ CCNOT gates, so overall U is represented as $O(n^2 2^{2n})$ operations.

Definition. **BQP (bounded-error quantum polynomial)** decision problems are those which a unitary operation solves with probability of success $p > c$, with $c > \frac{1}{2}$ a fixed constant (conventionally, $c = \frac{2}{3}$), with polynomial growth in resources (i.e. number of CNOT and single-qubit unitary gates) as n (number of qubits) is increased.

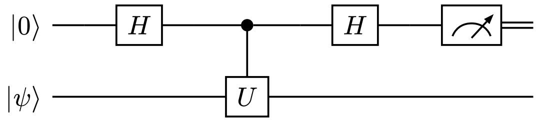
Note. BPP $\subseteq$ BQP, since any classical computation can be written in terms of CCNOT and CCNOT has fixed quantum cost. A source of randomness is the following circuit, giving 0 and 1 each with probability 1/2:



8.2. Measurement

Note. We can always measure using the Pauli Z operator (so measure in the computational basis). To measure in different basis, act with a unitary to transform desired basis into computational basis, then measure in computational basis, then transform back to desired basis.

Example. Let U be single-qubit operator, with eigenvalues ± 1, so it is Hermitian and unitary. Measuring U can be achieved with the following circuit:



Acting with H maps $|0\rangle \otimes |\psi\rangle \mapsto \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |\psi\rangle$. Acting with controlled-U gives $\frac{1}{\sqrt{2}}(|0\rangle \otimes |\psi\rangle + |1\rangle \otimes U|\psi\rangle)$. Acting with H again gives output

$$\frac{1}{2}((|0\rangle + |1\rangle) \otimes |\psi\rangle + (|0\rangle - |1\rangle) \otimes U|\psi\rangle) = \frac{1}{2}|0\rangle \otimes (I + U)|\psi\rangle + \frac{1}{2}|1\rangle \otimes (I - U)|\psi\rangle$$

But $\frac{1}{2}(I + U)$ is projector to $+1$ eigenspace of U, $\frac{1}{2}(1 - U)$ is projector to -1 eigenspace of U, so if $|\psi\rangle = \alpha|U_+\rangle + \beta|U_-\rangle$, with $U|U_\pm\rangle = \pm U_\pm$ then output is

$$\alpha|0\rangle \otimes|U_+\rangle + \beta|1\rangle \otimes|U_-\rangle$$

So result of measurement is 0 with probability $|\alpha|^2$, which collapses state to $|0\rangle \otimes |U_+\rangle$, and 1 with probability $|\beta|^2$, which collapses state to $|1\rangle \otimes|U_-\rangle$.

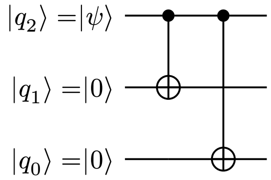
9. Quantum error correction

Note. We assume that an error only affects a single qubit.

9.1. Correcting single bit flips

Definition. A **code subspace** is a two-dimensional subspace of an n-qubit Hilbert space, in which the logical qubits live, such that each possible error (being considered) maps states in the code subspaces into a distinct two-dimensional subspace, and all of these error subspaces and the codespace are orthogonal.

Example. Assume only error that can occur is flip of single qubit (same as classical case), i.e. each qubit has probability p of X gate being applied. We encode the state in a **code subspace**. Each qubit is encoded as 3 qubits: the **logical qubit** $|\overline{0}\rangle$ is encoded as the *physical* state $|000\rangle$, $|\overline{1}\rangle$ is encoded as $|111\rangle$. So $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is mapped to $\alpha|000\rangle + \beta|111\rangle$, in the subspace span$\{|000\rangle,|111\rangle\}$ of the Hilbert space of 3 qubits. The embedding is implemented as



Single bit flip can map this state to

$$\alpha|001\rangle + \beta|110\rangle, \quad \alpha|010\rangle + \beta|101\rangle, \quad \alpha|100\rangle + \beta|011\rangle$$

which are all orthogonal to original state and each other. So different errors map to different orthogonal subspaces, hence we can make measurement to determine which subspace it is without affecting the α, β coefficients.

Error syndromes are operators with eigenspaces as the different subspaces, each with distinct eigenvalue. In this case, choose syndromes formed from Z operator (this has eigenvalue 1 for $|0\rangle$, -1 for $|1\rangle$). Let $Z_0 = I \otimes I \otimes Z$, $Z_1 = I \otimes Z \otimes I$, $Z_2 = Z \otimes I \otimes I$, then

$$\begin{gathered}
Z_0Z_1|000\rangle =|000\rangle, \quad Z_0Z_1|111\rangle =|111\rangle, \quad Z_0Z_2|000\rangle =|000\rangle, \quad Z_0Z_2|111\rangle =|111\rangle \\
Z_0Z_1|001\rangle = -|001\rangle, \quad Z_0Z_1|110\rangle = -|110\rangle, \quad Z_0Z_2|001\rangle = -|001\rangle, \quad Z_0Z_2|110\rangle = -|110\rangle \\
Z_0Z_1|010\rangle = -|010\rangle, \quad Z_0Z_1|101\rangle = -|101\rangle, \quad Z_0Z_2|010\rangle =|010\rangle, \quad Z_0Z_2|101\rangle =|101\rangle \\
Z_0Z_1|100\rangle =|100\rangle, \quad Z_0Z_1|011\rangle =|011\rangle, \quad Z_0Z_2|100\rangle = -|100\rangle, \quad Z_0Z_2|011\rangle = -|011\rangle
\end{gathered}$$

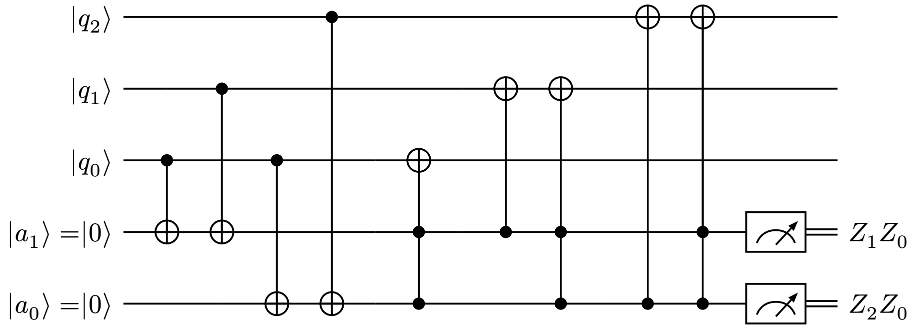
So span$\{|000\rangle,|111\rangle\}$ is $(1,1)$ eigenspace, span$\{|001\rangle,|110\rangle\}$ is $(-1,-1)$ eigenspace, span$\{|010\rangle,|101\rangle\}$ is $(-1,1)$ eigenspace, span$\{|100\rangle,|011\rangle\}$ is $(1,-1)$ eigenspace. So if $|\psi\rangle$ is mapped to

$$(1-\varepsilon)|\psi\rangle + \delta_2 X_2|\psi\rangle + \delta_1 X_1|\psi\rangle + \delta_0 X_0|\psi\rangle$$

then we measure $Z_0 Z_1$ and $Z_0 Z_2$, which collapses state to either

$$|\psi\rangle, \quad X_2|\psi\rangle, \quad X_1|\psi\rangle, \quad X_0|\psi\rangle$$

Since the eigenvalues for this combination of measurements are distinct, they tell us which state $|\psi\rangle$ has been projected to. So can apply I, X_2, X_1 or X_0 to map back to $|\psi\rangle$. This can be implemented as



where the measurements are to reset the ancilla so they can be reused.

Note. We cannot use less than 3 qubits, since to encode with n qubits, we need $n+1$ orthogonal two-dimensional subspaces, which is possible in 2^n-dimensional n qubit Hilbert space iff $2^n \geq 2(n+1)$.

9.2. Correcting general single qubit errors

Remark. General error consists of acting with unitary operation U_i on single physical qubit. Can use Bloch sphere rotation representation to write

$$U_i = e_i I + a_i X_i + b_i Y_i + c_i Z_i$$

So if state $|\psi\rangle$ is single logical qubit encoded in n-qubit Hilbert space, action of single qubit error on qubit i transforms $|\psi\rangle$ to

$$(1-\varepsilon)|\psi\rangle + a_i X_i|\psi\rangle + b_i Y_i|\psi\rangle + c_i Z_i|\psi\rangle \quad \text{for some } i$$

If error depends of state of environment, state after errors occurs is entangled:

$$|e_1\rangle \otimes |\psi\rangle + \sum_{i=1}^{n} |e_{2i}\rangle \otimes X_i|\psi\rangle + |e_{3i}\rangle \otimes Y_i|\psi\rangle + |e_{4i}\rangle \otimes Z_i|\psi\rangle$$

(This is linear superpositon of single qubit errors). Measuring chosen error syndromes projects qubits to one of the subspaces, so state becomes one of

$$|\psi\rangle, \quad X_i|\psi\rangle, \quad Y_i|\psi\rangle, \quad Z_i|\psi\rangle$$

$3n+1$ 2d subspaces are needed (corresponding to $3n$ single-qubit errors and original state), so we require

$$2^n \geq 2(3n+1)$$

which is saturated by $n=5$.

Remark. In terms of errors, X is a single bit flip, Z is a phase flip ($\alpha|0\rangle + \beta|1\rangle \to \alpha|0\rangle - \beta|1\rangle$), $Y = iXZ$ is composition of both.

Definition. We define a **coding** $c : H_1 \to H_n$, $|\overline{0}\rangle = c(|0\rangle)$, $|\overline{1}\rangle = c(|1\rangle)$.

Definition. **Steane code** is coding using 7 qubits, which uses the syndromes

$$M_0 := X_0X_4X_5X_6, \quad M_1 := X_1X_3X_5X_6, \quad M_2 := X_2X_3X_4X_6,$$
$$N_0 := Z_0Z_4Z_5Z_6, \quad N_1 := Z_1Z_3Z_5Z_6, \quad N_2 := Z_2Z_3Z_4Z_6$$

which all commute, so have simultaneous eigenstates. Code subspace is spanned by

$$|\overline{0}\rangle = \frac{1}{2^{3/2}}(1+M_0)(1+M_1)(1+M_2)|0000000\rangle,$$
$$|\overline{1}\rangle = \frac{1}{2^{3/2}}(1+M_0)(1+M_1)(1+M_2)|1111111\rangle$$

Remark. $M_j^2 = I$ so $M_j(1+M_j) = 1+M_j$ so $|\overline{0}\rangle$,$|\overline{1}\rangle$ are eigenstates of each M_j with eigenvalue 1. $|\overline{0}\rangle$,$|\overline{1}\rangle$ are also eigenstates of each N_k with eigenvalue 1. Each M_j commutes with each X_i, and

$$X_iZ_j = \begin{cases} Z_jX_i & \text{if } i \neq j \\ -Z_jX_i & \text{if } i = j \end{cases} \implies X_iN_j = \begin{cases} N_jX_i & \text{if } N_j \text{ does not contain } Z_i \\ -N_jX_i & \text{if } N_j \text{ contains } Z_i \end{cases}$$

Hence $M_jX_i(\alpha|\overline{0}\rangle + \beta|\overline{1}\rangle) = X_iM_j(\alpha|\overline{0}\rangle + \beta|\overline{1}\rangle) = X_i(\alpha|\overline{0}\rangle + \beta|\overline{1}\rangle)$ so this has eigenvalue 1 for all M_j, and

$$N_jX_i(\alpha|\overline{0}\rangle + \beta|\overline{1}\rangle) = \begin{cases} X_iN_j(\alpha|\overline{0}\rangle + \beta|\overline{1}\rangle) = X_i(\alpha|\overline{0}\rangle + \beta|\overline{1}\rangle) & \text{if } Z_i \notin N_j \implies \text{eigenvalue } 1 \\ -X_iN_j(\alpha|\overline{0}\rangle + \beta|\overline{1}\rangle) = -X_i(\alpha|\overline{0}\rangle + \beta|\overline{1}\rangle) & \text{if } Z_i \in N_j \implies \text{eigenvalue} - 1 \end{cases}$$

For bit flips X_i:

i	0	1	2	3	4	5	6
(M_0, M_1, M_2)	$(1,1,1)$	$(1,1,1)$	$(1,1,1)$	$(1,1,1)$	$(1,1,1)$	$(1,1,1)$	$(1,1,1)$
(N_0, N_1, N_2)	$(-1,1,1)$	$(1,-1,1)$	$(1,1,-1)$	$(1,-1,-1)$	$(-1,1,-1)$	$(-1,-1,1)$	$(-1,-1,-1)$

For phase flips (sign errors) Z_i:

i	0	1	2	3	4	5	6
(M_0, M_1, M_2)	$(-1,1,1)$	$(1,-1,1)$	$(1,1,-1)$	$(1,-1,-1)$	$(-1,1,-1)$	$(-1,-1,1)$	$(-1,-1,-1)$
(N_0, N_1, N_2)	$(1,1,1)$	$(1,1,1)$	$(1,1,1)$	$(1,1,1)$	$(1,1,1)$	$(1,1,1)$	$(1,1,1)$

Since $XY = -YX$, $ZY = -YZ$,

$$M_jY_i = \begin{cases} Y_iM_j & \text{if } X_i \notin M_j \\ -Y_iM_j & \text{if } X_i \in M_j \end{cases}, \quad \begin{cases} Y_iN_j & \text{if } Z_i \notin M_j \\ -Y_iM_j & \text{if } Z_i \in M_j \end{cases}$$

For errors Y_i:

i	0	1	2	3	4	5	6
(M_0, M_1, M_2)	$(-1,1,1)$	$(1,-1,1)$	$(1,1,-1)$	$(1,-1,-1)$	$(-1,1,-1)$	$(-1,-1,1)$	$(-1,-1,-1)$
(N_0, N_1, N_2)	$(-1,1,1)$	$(1,-1,1)$	$(1,1,-1)$	$(1,-1,-1)$	$(-1,1,-1)$	$(-1,-1,1)$	$(-1,-1,-1)$

Example. If $\left(M_j, N_j\right)$ measured and eigenvalues are $(1,1,1)$, $(1,-1,-1)$ then error is X_3, and we correct it by applying $X_3^{-1} = X_3$.

9.3. Fault-tolerant gates

Definition. A gate $\overline{U}$ is **fault-tolerant** if, when there is error U_i on single physical qubit before the unitary operation, acting with the unitary produces state which differs from desired state only by a single-qubit error V_j, i.e.

$$\overline{U}U_i|\psi\rangle = V_j\overline{U}|\psi\rangle$$

This equivalent to $\overline{U}$ mapping each eigenspace of the error syndromes to some eigenspace of the error syndromes.

Definition. A logical gate $\overline{G}$ is **transversal** if it is a tensor product of single qubit gates.

Proposition. Every transversal gate is fault tolerant.

Example. The operation $\overline{X}$, acting on the logical Hilbert space H_7, acts as the NOT operator on the code subspace span$\{|\overline{0}\rangle,|\overline{1}\rangle\}$:

$$\overline{X} = X_6X_5X_4X_3X_2X_1X_0, \quad \Longrightarrow \overline{X}|\overline{0}\rangle = |\overline{1}\rangle, \quad \overline{X}|\overline{1}\rangle = |\overline{0}\rangle$$

Example. The operation $\overline{Z} = Z_6Z_5Z_4Z_3Z_2Z_1Z_0$ commutes with each M_i and leaves $|0000000\rangle$ invariant so leaves $|\overline{0}\rangle$ invariant. $\overline{Z}$ anti-commutes with $\overline{X}$ so acts within the code subspace and $\overline{Z}|\overline{0}\rangle = |\overline{1}\rangle$, $\overline{Z}|\overline{1}\rangle = -|\overline{1}\rangle$, so acts as Pauli Z on logical qubits.

Example. $\overline{H} = H_6H_5H_4H_3H_2H_1H_0$ realises the Hadamard gate on logical qubits:

$$\overline{H}|\overline{0}\rangle = \frac{1}{\sqrt{2}}(|\overline{0}\rangle + |\overline{1}\rangle), \quad \overline{H}|\overline{1}\rangle = \frac{1}{\sqrt{2}}(|\overline{0}\rangle - |\overline{1}\rangle)$$

We have $HXH = Z$ so $H_i X_i = Z_i H_i$, thus

$$M_j \overline{H}|\psi\rangle = \overline{H} N_j |\psi\rangle, \quad N_j \overline{H}|\psi\rangle = \overline{H} M_j |\psi\rangle$$

Hence if $|\psi\rangle$ is in an eigenspace of M_j and N_j, $\overline{H}|\psi\rangle$ also lies in an eigenspace of M_j and N_j but with the eigenvalues of M_j and N_j swapped. This means $\overline{H}$ preserves the code subspace, so $\overline{H}|\overline{0}\rangle$ and $\overline{H}|\overline{1}\rangle$ lie in the code subspace. Now

$$\overline{H}|\overline{0}\rangle = \overline{H}\frac{1}{2^{3/2}}(1 + M_0)(1 + M_1)(1 + M_2)|0000000\rangle = \frac{1}{2^{3/2}}(1 + N_0)(1 + N_1)(1 + N_2)\overline{H}|0000000\rangle$$

$\overline{H}$ maps $|0000000\rangle$ to uniform superposition of all computational basis states, and $1 + N_j$ is projector onto $+1$ eigenspace of N_j, so we have the component of the uniform superposition which lies in the code subspace, i.e.

$$\overline{H}|\overline{0}\rangle = \frac{1}{\sqrt{2}}(|\overline{0}\rangle + |\overline{1}\rangle)$$

Similarly,

$$\overline{H}|\overline{1}\rangle = \overline{H}\frac{1}{2^{3/2}}(1 + M_0)(1 + M_1)(1 + M_2)|1111111\rangle = \frac{1}{2^{3/2}}(1 + N_0)(1 + N_1)(1 + N_2)\overline{H}|1111111\rangle$$

$\overline{H}$ maps $|1111111\rangle$ to uniform superposition of all computational basis states, with each state with an odd number of 1's negated, hence

$$\overline{H}|\overline{1}\rangle = \frac{1}{\sqrt{2}}(|\overline{0}\rangle - |\overline{1}\rangle)$$

as all computational basis states in $|\overline{1}\rangle$ have odd number of 1's.

Example. If two logical qubits are encoded with 14 physical qubits using Steane code, a logical CNOT can be implemented as

$$\overline{\mathrm{CNOT}} = \prod_{i=1}^{7} C_i \mathrm{NOT}_i$$

where $C_i\mathrm{NOT}_i$ is CNOT with ith qubit in first logical qubit as control and ith qubit in second logical qubit as target.

Theorem (Eastin, Knill). Not all gates in a UGS can be transversal.

10. Quantum algorithms

10.1. Simon's algorithm

Definition. **Bitwise addition** of a and b is $a \oplus b = c$ where $c_i = a_i + b_i \bmod 2$.

Definition. **Simon's problem** is: given an n-bit function $f : \{0,1\}^n \to \{0,1\}^n$, with $f(x \oplus a) = f(x)$ for all x $(a \neq 0)$ and $f(x) \neq f(y)$ otherwise, determine the period a.

Example. Let $f : \{0,1\}^n \to \{0,1\}^n$ be n-bit function with period $a \neq 0$, so $f(x \oplus a) = f(x)$ and $f(x) \neq f(y)$ otherwise. To determine a classically, we compute $f(x_i)$ until we find two values with $f(x_i) = f\left(x_j\right)$, then $a = x_i \oplus x_j$. After m values are computed, we know $a \neq x_i \oplus x_j$ for all $i, j \leq m$, so at most $\frac{1}{2}m(m-1)$ values are eliminated. There are $2^n - 1$ values for a, so this has complexity $O\left(2^{n/2}\right)$.

Definition. **Bitwise product** of $x = (x_{n-1} ... x_0)_2$ and $y = (y_{n-1} ... y_0)_2$ is

$$x \cdot y = x_{n-1}y_{n-1} \cdots + x_0 y_0 \mod 2$$

Proposition.

$$H^{\otimes n}|0\rangle = \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} |k\rangle \quad \text{and}$$

$$H^{\otimes n}|x\rangle = \bigotimes_{i=0}^{n-1} \frac{1}{\sqrt{2}}\left((-1)^{0 \cdot x_i}|0\rangle + (-1)^{1 \cdot x_i}|1\rangle\right) = \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} (-1)^{k \cdot x}|k\rangle$$

Algorithm (Simon's algorithm). Define the unitary operator U_f acting on n input qubits $|x\rangle$ and n output qubits $|m\rangle$:

$$U_f|x\rangle|m\rangle = |x\rangle|m \oplus f(x)\rangle$$

1. Start with system in state $|0\rangle_n \otimes |0\rangle_n$ where $|0\rangle_n = |00...0\rangle$.
2. Apply $H^{\otimes n} \otimes I$ (i.e. acting on input qubits) to give

$$\frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} |k\rangle \otimes |0\rangle_n$$

3. Apply U_f to give

$$\frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} |k\rangle \otimes |0 \oplus f(k)\rangle = \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} |k\rangle \otimes |f(k)\rangle$$

4. Measure the ancillary bits (the $|f(k)\rangle$) in the computational basis, yielding a random $f(x_0)$. The state collapses to

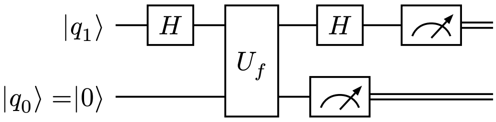
$$\frac{1}{\sqrt{2}}(|x_0\rangle + |x_0 \oplus a\rangle) \otimes |f(x_0)\rangle$$

5. Discard ancillary bits and apply $H^{\otimes n}$ to input bits $\frac{1}{\sqrt{2}}(|x_0\rangle + |x_0 \oplus a\rangle)$ to give

$$\begin{aligned} H^{\otimes n}\frac{1}{\sqrt{2}}(|x_0\rangle + |x_0 \oplus a\rangle) &= \frac{1}{\sqrt{2}}(H^{\otimes n}|x_0\rangle + H^{\otimes n}|x_0 \oplus a\rangle) \\ &= \frac{1}{\sqrt{2}}\frac{1}{2^{n/2}}\left(\sum_{k=0}^{2^n-1}(-1)^{k\cdot x_0}|k\rangle + \sum_{k=0}^{2^n-1}(-1)^{k\cdot(x_0\oplus a)}|k\rangle\right) \\ &= \frac{1}{2^{(n+1)/2}}\sum_{k=0}^{2^n-1}(-1)^{k\cdot x_0}\left(1+(-1)^{k\cdot a}\right)|k\rangle \\ &= \frac{1}{2^{(n-1)/2}}\sum_{\substack{k=0:\\ k\cdot a=0}}^{2^n-1}(-1)^{k\cdot x_0}|k\rangle \end{aligned}$$

6. Measure the state in the computational basis, which gives $k \in \{0, ..., 2^n - 1\}$ such that $k \cdot a = 0 \bmod 2$.
7. a satisfies n linearly independent equations of the form $k \cdot a = 0 \bmod 2$, so $O(n)$ measurements ($O(n)$ values of k) are needed to obtain all bits of a.

This can be implemented as



Example. Let $f : \{0,1\}^3 \to \{0,1\}^3$, $a = 010$, and

$$\begin{aligned} f(000) = f(010) = x, \quad f(001) = f(011) = y, \\ f(100) = f(110) = z, \quad f(101) = f(111) = w \end{aligned}$$

Using Simon's algorithm:

- Applying $H^{\otimes 3}$ to $|000\rangle \otimes |000\rangle$ gives

$$\frac{1}{2\sqrt{2}}(|000\rangle + |001\rangle + |010\rangle + |011\rangle + |100\rangle + |101\rangle + |110\rangle + |111\rangle) \otimes |000\rangle$$

- Applying U_f gives

$$\frac{1}{2\sqrt{2}}((|000\rangle + |010\rangle) \otimes |x\rangle + (|001\rangle + |011\rangle) \otimes |y\rangle + (|100\rangle + |110\rangle) \otimes |z\rangle + (|101\rangle + |111\rangle) \otimes |w\rangle)$$

- Measure the ancillary bits, assuming it yields value corresponding to $|x\rangle$, so state has collapsed to

$$\frac{1}{\sqrt{2}}(|000\rangle + |010\rangle) \otimes |x\rangle$$

- Apply $H^{\otimes 3}$ to the input bits, giving

$$\frac{1}{2^{(3-1)/2}}\sum_{\substack{k=0:\\ k\cdot a=0}}^{2^3-1}|k\rangle = \frac{1}{2}(|000\rangle + |001\rangle + |101\rangle + |100\rangle)$$

- Measuring 000 gives no information. Measuring 001 implies that $a_0 = 0$. Measuring 010 implies that $a_1 = 0$. Measuring 101 implies that $a_0 + a_2 = 0$. So measuring the last three imply $a = 010$ (since $a \neq 000$).

10.2. Quantum Fourier transform

Definition. **Quantum Fourier transform** is unitary operation U_{FT} acting on the n qubit space H_n, given by action on computational basis states:

$$U_{\text{FT}}|x\rangle = \frac{1}{2^{n/2}} \sum_{y=0}^{2^n-1} e^{2\pi ixy/2^n} |y\rangle$$

It is quantum version of the discrete Fourier transform: by linearity, if $|\psi\rangle = \sum_{x=0}^{2^n-1} \psi_x |x\rangle$ and $|\varphi\rangle = U_{\text{FT}}|\psi\rangle = \sum_{y=0}^{2^n-1} \varphi_y |y\rangle$, then

$$\varphi_y = \frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} e^{2\pi ixy/2^n} \psi_x$$

Note this is precisely the discrete Fourier transform on the vector ψ_x.

Note. Can check U_{FT} is unitary by checking $U_{\text{FT}}|x\rangle$ has norm 1 and $U_{\text{FT}}|x\rangle$ orthogonal to $U_{\text{FT}}|x'\rangle$ for $x \neq x'$ (i.e. it preserves the inner product).

Example. Note that classically, computing φ_y requires 2^n additions. If $y = y_{n-1} \cdots y_0$, i.e. $y = y_{n-1}2^{n-1} + \cdots + y_0$, then

$$e^{2\pi ixy/2^n} = \prod_{l=0}^{n-1} e^{2\pi ixy_l/2^{n-l}}$$

which gives

$$U_{\text{FT}}|x\rangle = \frac{1}{2^{n/2}} \sum_{y=0}^{2^n-1} e^{2\pi ixy/2^n} |y\rangle = \frac{1}{2^{n/2}} \bigotimes_{l=0}^{n-1} \left(|0\rangle + e^{2\pi ix/2^{n-l}} |1\rangle\right)$$

Note this is similar to

$$H^{\otimes n}|x\rangle = \bigotimes_{i=0}^{n-1} \frac{1}{\sqrt{2}} \left(|0\rangle_i + (-1)^{x_i} |1\rangle_i\right) = \sum_{y=0}^{2^n-1} (-1)^{x \cdot y} |y\rangle$$

However, for U_{FT}, phases in individual qubit states depend on x, not just x_l, so U_{FT} cannot be realised only by single-qubit operations. Now also

$$e^{2\pi x/2^{n-l}} = e^{2\pi i(x_{n-1}2^{l-1} + \cdots + x_0 2^{l-n})} = \prod_{m=0}^{n-1} e^{2\pi i x_m/2^{n-l-m}} = \prod_{m=0}^{n-l-1} e^{2\pi i x_m/2^{n-l-m}}$$

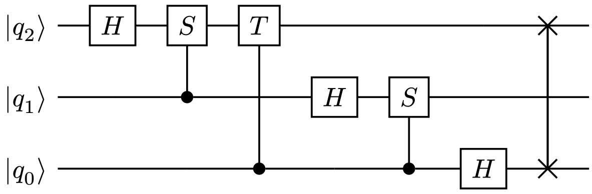
since $e^{2\pi ir} = 1$ for $r \in \mathbb{Z}$. So phase for $l = n - 1$ only depends on x_0, phase for $l = n - 2$ only depends on x_0 and x_1:

$$U_{\text{FT}}|x\rangle = \frac{1}{2^{n/2}} (|0\rangle + e^{i\pi x_0} |1\rangle) \otimes \left(|0\rangle + e^{i\pi x_1} e^{i\pi x_0/2}\right) \otimes \cdots$$

When U_{FT} is realised, order of qubits is reversed. QFT can be implemented of controlled-phase gates, where we apply unitaries

$$R_k = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/2^k} \end{bmatrix}$$

Each qubit i has controlled-R_k applied, controlled by each qubit $j < i$, where $k = i - j$. E.g. for $n = 3$,



where × indicates swapping qubits. One gate is applied to qubit 0, two gates applied to qubit 1, ..., n gates applied to qubit $n - 1$, so total number of gates required to implemented QFT is $O(n^2)$

10.3. Shor's algorithm

Example. Given $N \in \mathbb{N}$, pick random $1 < y < N$. If $\gcd(y, N) \neq 1$, we can find a divisor of N. If $\gcd(y, N) = 1$, define

$$f_y : \mathbb{Z} \to \mathbb{Z}/N, \quad f_y(a) = y^a \mod N$$

Period of f_y is smallest $r \in \mathbb{N}$ such that $f_y(r) = 1$. We have $f_y(a) = f_y(b)$ iff $a - b = 0 \bmod r$. Let r be even (if r odd, start again with different y). Now

$$y^r - 1 = 0 \bmod N \Longrightarrow \left(y^{r/2} - 1\right)\left(y^{r/2} + 1\right) = 0 \bmod N$$

If either factor on LHS is multiple of N, start again with different y. Otherwise, we know $y^{r/2} - 1$ and N have common factor $< N$, and so use Euclid's algorithm to find $\gcd\left(y^{r/2} - 1, N\right)$.

Algorithm (Shor's algorithm).

- Shor's algorithm finds the smallest $r > 0$ such that $y^r \equiv 1 \bmod N$.
- Start with state $|0\rangle_n |0\rangle_{n_0}$ where $n_0 = \lceil \log_2(N) \rceil$, $n = 2n_0$.
- Act with $H^{\otimes n}$ on input bits, giving

$$\frac{1}{2^{n/2}} \sum_{x=0}^{2^n - 1} |x\rangle \otimes |0\rangle$$

- Act with U_f (where $U_f |x\rangle |m\rangle = |x\rangle |m \oplus f(x)\rangle$), giving

$$\frac{1}{2^{n/2}} \sum_{x=0}^{2^n - 1} |x\rangle \otimes |f(x)\rangle$$

- Measure the output bits, yielding a random value $f(x_0)$ (assume WLOG that $x_0 < r$), which projects the state to

$$\frac{1}{\sqrt{Q+1}} \sum_{m=0}^{Q} |x_0 + mr\rangle |f(x_0)\rangle$$

where $Q = |\{i \in \{0, ..., 2^n - 1\} : f(i) = f(x_0)\}|$ which is approximately the largest integer strictly less than $2^n/r$. Shift by random x_0 means we can't learn anything about r by measuring input bits. Discard output bits.

- Perform QFT on input bits, giving

$$\begin{aligned} U_{\mathrm{FT}} \frac{1}{\sqrt{Q+1}} \sum_{m=0}^{Q} |x_0 + mr\rangle &= \frac{1}{\sqrt{Q+1}} \sum_{m=0}^{Q} \frac{1}{2^{n/2}} \sum_{y=0}^{2^n-1} e^{2\pi i (x_0 + mr) y/2^n} |y\rangle \\ &= \frac{1}{2^{n/2}} \sum_{y=0}^{2^n-1} e^{2\pi i x_0 y/2^n} \left(\frac{1}{\sqrt{Q+1}} \sum_{m=0}^{Q} e^{2\pi i m r y/2^n} \right) |y\rangle \end{aligned}$$

- Measure input bits in the computational basis. Probability that this yields value y is

$$\begin{aligned} p(y) &= \left| \frac{1}{\sqrt{Q+1}} \frac{1}{2^{n/2}} e^{2\pi i x_0 y/2^n} \left(\sum_{m=0}^{Q} e^{2\pi i m r y/2^n} \right) \right|^2 \\ &= \frac{1}{2^n (Q+1)} \left| \sum_{m=0}^{Q} e^{2\pi i m r y/2^n} \right|^2 \\ &= \frac{1}{2^n (Q+1)} \left| \frac{e^{2\pi i r y (Q+1)/2^n} - 1}{e^{2\pi i r y/2^n} - 1} \right|^2 \\ &= \frac{1}{2^n (Q+1)} \left| \frac{e^{\pi i r y (Q+1)/2^n} \left(e^{\pi i r y (Q+1)/2^n} - e^{-\pi i r y (Q+1)/2^n} \right)}{e^{\pi i r y/2^n} \left(e^{\pi i r y/2^n} - e^{-\pi i r y/2^n} \right)} \right|^2 \\ &= \frac{1}{2^n (Q+1)} \frac{\sin^2(\pi r y (Q+1)/2^n)}{\sin^2(\pi r y/2^n)} \end{aligned}$$

- When $ry/2^n \in \mathbb{Z}$, we have $p(y) = \frac{1}{2^n(Q+1)} \left| \sum_{m=0}^{Q} 1 \right|^2 = (Q+1)/2^n$. Now $Q + 1 \approx 2^n/r$ so $p(y) \approx 1/r$.
- If $ry/2^n \notin \mathbb{Z}$ (and not close to being an integer), then $\sum_{m=0}^{Q} e^{2\pi i m r y/2^n} < 1$ (typically a small value since phases do not add coherently) and $p(y) \approx 1/(2^n(Q+1)) \approx r/4^n$. Note $r \leq N < 2^{n_0} \ll 2^n$ implies that summing over all the approximately 2^n possibly values of y gives

$$\sum_{y : ry/2^n \notin \mathbb{Z}} p(y) \approx 2^n r/4^n \approx r/2^n \ll 1$$

- Hence it is likely to measure y such that $ry/2^n$ is approximately an integer. Equivalently, $y/2^n = j/r$ for some $j \in \mathbb{Z}$.
- With high probability, y will be the nearest integer to a multiple of $2^n/r$, i.e. within $1/2$ of $j2^n/r$, so

$$\left|\frac{y}{2^n}-\frac{j}{r}\right| \leq \frac{1}{2^{n+1}} \leq \frac{1}{2N^2}$$

since $N \leq 2^{n_0} = 2^{n/2}$. There is unique fraction j/r with $r < N$ satisfying this (by triangle inequality), as

$$\left|\frac{j_1}{r_1}-\frac{j_2}{r_2}\right| \geq \frac{1}{r_1 r_2} > \frac{1}{N^2}$$

unless $\frac{j_1}{r_1} = \frac{j_2}{r_2}$ (this is why $n = 2n_0$ is chosen). $\frac{j}{r}$ can be obtained from $y/2^n$ via continued fractions.

- If j and r have common divisor, we obtain $r_0 = r/\gcd(j,r)$ instead of r. But given the guess r_0, we can check if r_0 is the period by checking if $f(r_0) = 1$, and if not try $f(2r_0), f(3r_0), \ldots$. If these fail, run algorithm again. Probability of j and r having common divisor is $< 1/2$.