CS3222
Quantum Computing

Course Outline

Intro [1]
Linear Algebra [3]
Quantum Mechanics [3]
Entanglement [1]
Reading Week [2 Ex]
Simple quantum algorithms [3 + Ex]
Quantum Search [5 + Ex]
Quantum Factoring [3 + Ex]

Book: Nielsen M., Chuang I. (2000) Quantum Computation and Quantum Information. CUP.
Additional Reading: See Syllabus / Web Page.

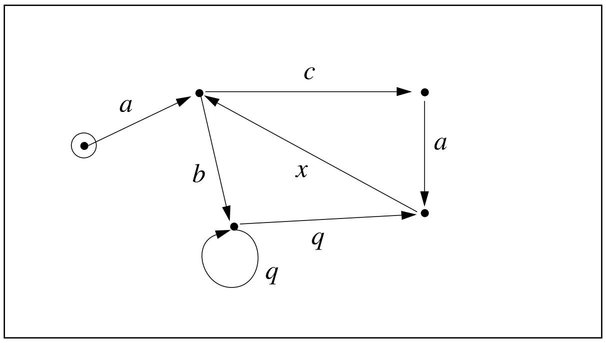
Notes: These notes plus Abbas Edalat's Quantum Computing course notes from Imperial College (used with permission).

Syllabus Web Page > Additional Info:
1. Downloadable Material, eg. the course notes and exercises.
2. Other relevant material, eg. pointers to even more reading, etc. (and subject to updating).

Transition Systems

A ***state space*** is just a set S.
An ***initial state*** is just a fixed element $s_I \in S$.
A ***current state*** is just a variable element $s \in S$.
An ***action*** is an element of a set Act of actions.
A ***transition*** is a triple (s,act,s'), where s, s' are states and act is an action.
A (***labelled***) ***transition system*** $\mathcal{S}$ on (S,Act) is just a set of transitions built from (S,Act).

Picture:



The initial state is usually indicated in some way in a picture, eg. ⊙ . A current state may or may not be indicated.

Action names *may* be used to represent external control by the user over what the TS does next (i.e. to express initiative).

Example 1 (Computing): A 32-bit register.

The state space is the set of values that can be held in the register. The initial state is the value at power-on (usually 0). The current state is the value held in the register at the current time as the computation progresses. The actions are the individual machine instructions that may be invoked. The transitions are the possible computation steps.

Example 2 (Physics): A classical particle in a potential well.

The state space is the set of position/velocity pairs of the particle. The initial state is the position and velocity at the start of the dynamics. The current state is the position/velocity at the current time t. There is only one action, which corresponds to the law of motion. The transitions correlate $(\boldsymbol{x}(t),\boldsymbol{v}(t))$ with $(\boldsymbol{x}(t'),\boldsymbol{v}(t'))$ where $t < t'$, if these are consistent with the dynamics.
(Cf. phase space.)

Example 3 (Mathematics): Group actions.

Let G be a group. Regard G as *both* a state space *and* a set of actions. The initial state is e, the identity. The transitions are triples (a,g,ag) where $a,g \in G$; i.e. the group elements act by right multiplication. (N.B. The current state is not a concept much used in this sort of treatment.)

Issues for Transition Systems

Nondeterministic TS: If for some state s, and action a, we have two states $s' \neq s''$ such that both (s,a,s') and (s,a,s'') are transitions, then the TS is ***nondeterministic***. Otherwise it is deterministic.

Examples 1 and 3 are deterministic. Example 2 is nondeterministic if the particle is stationary on a local maximum of the potential, otherwise it is deterministic.

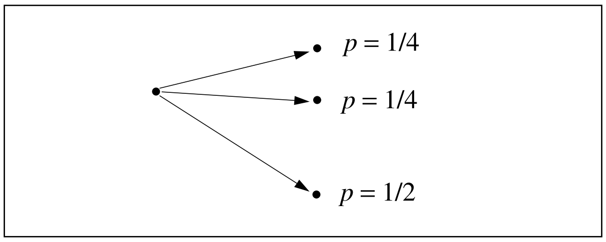
Assume only one action from now on.

Stochastic TS: If each transition (s,s') has an associated ***probability*** $p(s,s')$, then the TS is ***stochastic***.

Since when starting from any state, there must, with certainty, be *some* outcome, we must have for all s:

$$\Sigma_{s'}\, p(s,s') = 1$$

Picture:



Quantum TS: If each transition (s,s') has an associated

probability amplitude $\psi(s,s')$

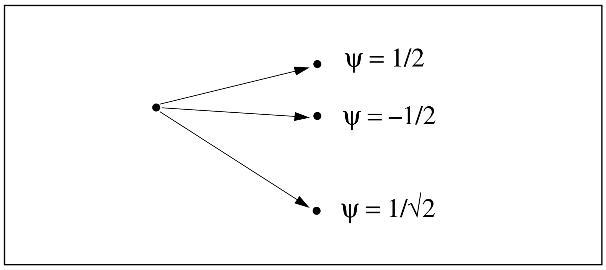
then the TS is a ***quantum TS***.

Probability amplitudes are complex numbers; they generate probabilities by ***squaring the modulus***.

Since when starting from any state, there must, with certainty, be *some* outcome, we must have for all s:

$$\sum_{s'} |\psi(s,s')|^2 = 1$$

Picture:



Why bother with probability amplitudes if all you do is square them to get probabilities?

Probabilities vs. Amplitudes

Sequential composition

Suppose s,s',s'' are states visited in order by the TS. Then:

Stochastic: $p(s,s',s'') = p(s,s') \times p(s',s'')$
Quantum: $\psi(s,s',s'') = \psi(s,s') \times \psi(s',s'')$

The quantum case leads to the same behaviour for the derived probabilities for sequences of transitions.

Composition of alternatives

Suppose a TS can move from state s to state s' via either u_1 or u_2.

Then for a *stochastic* TS we have:

$$p(s,u_1,s') = p(s,u_1) \times p(u_1,s') \quad \text{and}$$
$$p(s,u_2,s') = p(s,u_2) \times p(u_2,s')$$

Altogether:

$$p(s,s') = p(s,u_1)p(u_1,s') + p(s,u_2)p(u_2,s')$$

For a *quantum* TS

$$\psi(s,u_1,s') = \psi(s,u_1) \times \psi(u_1,s') \quad \text{and}$$
$$\psi(s,u_2,s') = \psi(s,u_2) \times \psi(u_2,s')$$

Altogether:

$$\psi(s,s') = \psi(s,u_1)\,\psi(u_1,s') + \psi(s,u_2)\,\psi(u_2,s')$$

However ψ is not a probability. To get the probability of going from s to s' via either u_1 or u_2 we need the modulus squared:

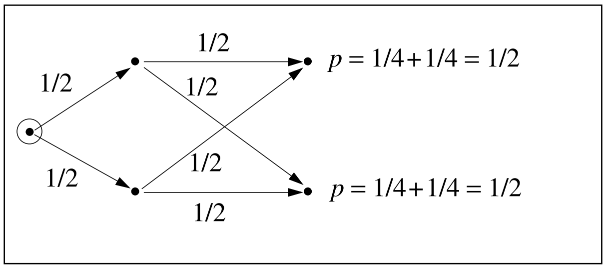
$$p_Q(s,s') = |\psi(s,u_1)\psi(u_1,s') + \psi(s, u_2)\psi(u_2, s')|^2$$

$$= |\psi(s,u_1)\psi(u_1,s')|^2 + |\psi(s,u_2)\psi(u_2,s')|^2 + \psi(s,u_1)\psi(u_1,s')\psi^*(s,u_2)\psi^*(u_2,s') + \psi^*(s,u_1)\psi^*(u_1,s')\psi(s,u_2)\psi(u_2,s')$$

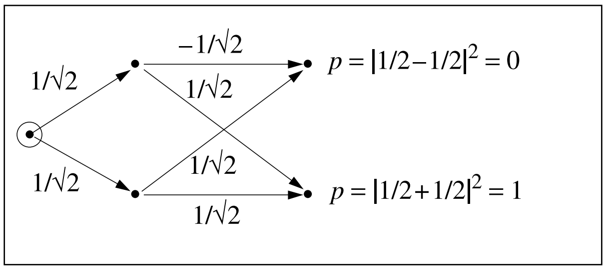
This can lead to possibilities different from the stochastic case, because the interference terms need not be positive.

The interference terms can indeed be negative, leading to ***destructive interference***, which is impossible in stochastic TS.

Stochastic:



Quantum:



Features of Quantum Computing

- exponential stochasticity
- superposition and destructive interference
- measurement
- entanglement

Linear Algebra

A ***complex inner product space*** $\boldsymbol{H}$ is a set H which carries a complex vector space structure:

(1) $|x\rangle, |y\rangle \in H \Rightarrow |x\rangle + |y\rangle = |x + y\rangle = |y\rangle + |x\rangle \in H$

(2) $|x\rangle, |y\rangle, |z\rangle \in H \Rightarrow (|x\rangle + |y\rangle) + |z\rangle = |x\rangle + (|y\rangle + |z\rangle)$

(3) $\underline{0} \in H$ and $\underline{0} + |x\rangle = |x\rangle \in H$

(4) $|x\rangle \in H$, $\lambda \in \mathbb{C} \Rightarrow |\lambda x\rangle = \lambda|x\rangle \in H$

(5) $|x\rangle \in H$, $\lambda, \mu \in \mathbb{C} \Rightarrow |\lambda\mu x\rangle = \lambda|\mu x\rangle = \lambda\mu|x\rangle \in H$

(6) $|x\rangle, |y\rangle \in H$, $\lambda \in \mathbb{C} \Rightarrow \lambda(|x\rangle + |y\rangle) = \lambda|x\rangle + \lambda|y\rangle \in H$

(7) $|x\rangle \in H$, $\lambda, \mu \in \mathbb{C} \Rightarrow (\lambda + \mu)|x\rangle = \lambda|x\rangle + \mu|x\rangle \in H$

together with an inner product $\langle y|x\rangle$, i.e.:

(8) $|x\rangle, |y\rangle \in H \Rightarrow \langle y|x\rangle \in \mathbb{C}$

(9) $\langle x|y\rangle = \langle y|x\rangle^*$

(10) $\langle z|(|x\rangle + |y\rangle) = \langle z|x\rangle + \langle z|y\rangle$

(11) $\mu \in \mathbb{C} \Rightarrow \langle y|\mu x\rangle = \mu\langle y|x\rangle = \langle \mu^* y|x\rangle$

(12) $\langle x|x\rangle \geq 0$ and $\langle x|x\rangle = 0 \Rightarrow |x\rangle = \underline{0}$

N.B. $|...\rangle$ is called a ***ket***, and is standard Dirac notation, used in quantum physics.

We also write $||x|| = \sqrt{\langle x|x\rangle}$, where $||x||$ is the ***norm*** of $|x\rangle$. The norm is like a length, in that: $||x|| \geq 0$, $||\lambda x|| = |\lambda|.||x||$, and $||x||$ is only zero for $|x\rangle = \underline{0}$.

In arbitrary inner product spaces, the inner product is the analogue of the 'dot product' for conventional vectors in 3D space.

If $\langle y|x\rangle = 0$ then $|x\rangle$ and $|y\rangle$ are ***orthogonal***.

Examples

Example 1: Vectors in 3-space. (N.B. This is not an example in the strict sense since the scalars used are the reals).

A typical vector in 3-space $\boldsymbol{a} = a_x\boldsymbol{e}_x + a_y\boldsymbol{e}_y + a_z\boldsymbol{e}_z$, where $\boldsymbol{e}_x$, $\boldsymbol{e}_y$, $\boldsymbol{e}_z$ are the three orthogonal unit vectors along the coordinate axes.

Then $\lambda\boldsymbol{a} = \lambda a_x\boldsymbol{e}_x + \lambda a_y\boldsymbol{e}_y + \lambda a_z\boldsymbol{e}_z$ (with $\lambda \in \mathbb{R}$).

If $\boldsymbol{b} = b_x\boldsymbol{e}_x + b_y\boldsymbol{e}_y + b_z\boldsymbol{e}_z$, then $\boldsymbol{a} + \boldsymbol{b} = (a_x + b_x)\boldsymbol{e}_x + (a_y + b_y)\boldsymbol{e}_y + (a_z + b_z)\boldsymbol{e}_z$.

The inner product is the usual dot product $\boldsymbol{a}.\boldsymbol{b} = a_x b_x + a_y b_y + a_z b_z$.

The norm is given by the length: $||\boldsymbol{a}|| = |\boldsymbol{a}| = \sqrt{(a_x^2 + a_y^2 + a_z^2)} = \sqrt{(\boldsymbol{a}.\boldsymbol{a})}$, giving rise to the definition of the angle θ between $\boldsymbol{a}$ and $\boldsymbol{b}$ via $\boldsymbol{a}.\boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}|\cos(\theta)$.

Example 2: $L_2(K)$. Here K is any space upon which integration is defined, eg. $\mathbb{R}^n$, $[0,1]\times[1,3]$, etc., etc.

Then $L_2(K)$ consists of all the complex valued functions f such that $\int_K |f|^2$ is well defined. If f and g are two such functions then $f+g$ is also in $L_2(K)$ since $\int_K |f+g|^2 \leq \int_K ||f|+|g||^2 \leq \int_K (2\max(|f|,|g|))^2 \leq \int_K (4|f|^2 + 4|g|^2)$ which is finite. Similarly for λf.

Addition and scalar multiplication are just addition and scalar multiplication of functions, and the inner product is given by $\langle f|g\rangle = \int_K f^* g$.

The norm is thus $||f|| = (\int_K |f|^2)^{1/2}$.

The elements of $L_2(K)$ are used in quantum theory for wave functions of particles in a region K.

Example 3: $\mathbb{C}^n$ (Complex Column Vectors). This example will be used extensively in the rest of the course. A typical vector is:

$$|x\rangle = \begin{bmatrix} x_0 \\ x_1 \\ \dots \\ x_{n-1} \end{bmatrix}$$

We will often write this as $[x_0, x_1 \dots x_{n-1}]^T$ to save space. Addition and scalar multiplication are just addition and scalar multiplication of vectors. The inner product is given by:

$$\langle y|x\rangle = y_0^*.x_0 + y_1^*.x_1 + \dots + y_{n-1}^*.x_{n-1}$$

Consequently the norm is given by:

$$||x|| = \sqrt{(|x_0|^2 + |x_1|^2 + \dots + |x_{n-1}|^2)}$$

These vectors give a good account of quantum systems with a small (finite) number of degrees of freedom; such systems exclusively will concern us in this course. In future $\mathbb{C}^2$ will be called $\boldsymbol{Q}$.

Basic Properties

The ***Parallelogram Identity***:

$$||x+y||^2 + ||x-y||^2 = 2(||x||^2 + ||y||^2)$$

Proof. LHS is $\langle x+y|x+y\rangle + \langle x-y|x-y\rangle$. Simplify.

The ***Polarisation Identity***:

$$4\langle x|y\rangle = ||x+y||^2 - ||x-y||^2 - \mathrm{i}||x+\mathrm{i}y||^2 + \mathrm{i}||x-\mathrm{i}y||^2$$

Proof. Expand RHS.

The ***Cauchy-Schwartz inequality***:

$$|\langle x|y\rangle|^2 \leq \langle x|x\rangle\langle y|y\rangle = ||x||^2||y||^2$$

Proof. If $|x\rangle = \underline{0}$ or $|y\rangle = \underline{0}$ it's easy. Otherwise let $\lambda = -\langle y|x\rangle/\langle y|y\rangle$. Then:

$$0 \leq \langle x+\lambda y|x+\lambda y\rangle = \langle x|x\rangle + \lambda\langle x|y\rangle + \lambda^*\langle y|x\rangle + |\lambda|^2\langle y|y\rangle$$

which when simplified, gives the result.

The ***Triangle inequalities***:

$$\|x\|-\|y\| \le \|x+y\| \le \|x\|+\|y\|$$
$$\|x\|-\|y\| \le \|x-y\| \le \|x\|+\|y\|$$

Proof. $\|x+y\|^2 = \langle x+y|x+y\rangle = \|x\|^2+\langle x|y\rangle+\langle y|x\rangle+\|y\|^2 = \|x\|^2+2\mathrm{Re}\langle x|y\rangle+\|y\|^2 \le \|x\|^2+2|\langle x|y\rangle|+\|y\|^2 \le \|x\|^2+2\|x\|\|y\|+\|y\|^2 = (\|x\|+\|y\|)^2$; now take the square root. For the other part, $(\|x\|-\|y\|)^2 = \|x\|^2-2\|x\|\|y\|+\|y\|^2 \le \|x\|^2-2|\langle x|y\rangle|+\|y\|^2 \le \|x\|^2+2\mathrm{Re}\langle x|y\rangle+\|y\|^2 = \|x+y\|^2$. The second line is similar.

Independence, Bases, (Finite) Dimensionality

A finite collection $|x_0\rangle,|x_1\rangle\ldots|x_d\rangle$ is ***independent*** iff

$$\lambda_0|x_0\rangle+\lambda_1|x_1\rangle\ldots+\lambda_d|x_d\rangle = \underline{0} \Rightarrow \lambda_0 = \lambda_1 = \ldots \lambda_d = 0$$

A maximal (finite) set of independent vectors is a ***basis***. Given a basis $|x_0\rangle,|x_1\rangle\ldots|x_d\rangle$, every $|y\rangle$ has an expansion of the form:

$$|y\rangle = \lambda_0|x_0\rangle+\lambda_1|x_1\rangle\ldots+\lambda_d|x_d\rangle$$

(or else $|x_0\rangle,|x_1\rangle\ldots|x_d\rangle,|y\rangle$ would be a bigger independent set, contradicting maximality). Also, the expansion is unique. (*Exercise*.) We say the basis ***spans*** the whole space.

Moreover, any two bases have the same number of elements, the ***dimension*** of $\boldsymbol{H}$, $\dim(\boldsymbol{H})$. (*Exercise*.)

Here are some bases for $\boldsymbol{Q}$:

$\{[1, 0]^T, [0, 1]^T\}$, $\{[1, 0]^T, [1, 1]^T\}$,
$\{[1, i]^T, [1, -i]^T\}$, $\{[2+3i, -27-4i]^T, [578+675i, 2]^T\}$.

A basis is ***orthogonal*** iff $\langle x_j|x_k\rangle = 0$ for any two distinct elements $|x_j\rangle,|x_k\rangle$. Above, $\{[1, 0]^T, [0, 1]^T\}$ and $\{[1, i]^T, [1, -i]^T\}$ are orthogonal bases.

A basis is ***orthonormal*** iff it is orthogonal and all its elements have norm 1. Above, $\{[1, 0]^T, [0, 1]^T\}$ is orthonormal; $\{[1, i]^T, [1, -i]^T\}$ can be ***normalised*** to give $\{\frac{1}{\sqrt{2}}[1, i]^T, \frac{1}{\sqrt{2}}[1, -i]^T\}$ which *is* orthonormal.

Any maximal finite orthonormal set is a basis. (*Exercise*.)

Henceforth we restrict attention to orthonormal bases. Let $|x_0\rangle,|x_1\rangle\ldots|x_d\rangle$ be an orthonormal basis. Then the expansion of $|y\rangle$ is:

$$|y\rangle = \lambda_0|x_0\rangle+\lambda_1|x_1\rangle\ldots+\lambda_d|x_d\rangle$$

where for each k:

$$\lambda_k = \langle x_k|y\rangle$$

so that:

$$|y\rangle = \Sigma_k \langle x_k|y\rangle|x_k\rangle = \Sigma_k |x_k\rangle\langle x_k|y\rangle$$

Proof. For any k, $\langle x_k|y\rangle = \langle x_k|\Sigma_j\lambda_j x_j\rangle = \lambda_k$.

Linear Operators

A linear operator A from $\boldsymbol{H}_1$ to $\boldsymbol{H}_2$ is a mapping such that for all $|x\rangle,|y\rangle \in H_1$, $\lambda,\mu \in \mathbb{C}$:

$$A(\lambda|x\rangle+\mu|y\rangle) = \lambda A|x\rangle+\mu A|y\rangle$$

where $A|x\rangle, A|y\rangle \in H_2$. Most often, $\boldsymbol{H}_1 = \boldsymbol{H}_2$.

Clearly $A\underline{0} = \underline{0}$, and linear operators can be added and scaled $(\lambda A+\mu B)|x\rangle = \lambda A|x\rangle+\mu B|x\rangle$.

Examples

Example 1: The null operator $\mathbf{0}$; for all $|x\rangle$, $\mathbf{0}|x\rangle = \underline{0}$.

Example 2: The identity operator $\mathbf{I}$; $\mathbf{I}|x\rangle = |x\rangle$.

Example 3: The operator on 3-space that maps $\boldsymbol{a} = a_x\boldsymbol{e}_x + a_y\boldsymbol{e}_y + a_z\boldsymbol{e}_z$ to $\boldsymbol{a}' = 5a_z\boldsymbol{e}_x + (12a_y+6a_x)\boldsymbol{e}_y + a_z\boldsymbol{e}_z$.

Example 4: On $L_2(\mathbb{R})$, the operator *x* that sends $f(x)$ to $x.f(x)$, i.e. it is the operation “multiply by x”.

Example 5: On $L_2(\mathbb{R})$, the operator *p* that sends $f(x)$ to $-i\,df/dx$, i.e. it is the operation “differentiate by x and multiply the result by $-i$”.

Examples 4 and 5 are fundamental to the wave function description of quantum systems in one dimension. Up to factors of $\hbar$, *x* and *p* are the position and momentum operators. Unfortunately

the examples are rather dishonest: there is no reason why for arbitrary $f \in L_2(\mathbb{R})$, that $x.f$ should be (finitely) square integrable, still less that f itself should be differentiable. This shows the problems that arise in the infinite dimensional case.

Example 6: On $\mathbb{C}^n$, operators are matrices.

$$|x\rangle = \begin{bmatrix} x_0 \\ x_1 \\ \ldots \\ x_{n-1} \end{bmatrix} \quad A|x\rangle = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0n-1} \\ a_{10} & a_{11} & \cdots & a_{1n-1} \\ \ldots & \ldots & \cdots & \ldots \\ a_{n-10} & a_{n-11} & \cdots & a_{n-1n-1} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \ldots \\ x_{n-1} \end{bmatrix}$$

Example 7: The Pauli Spin Matrices on $\boldsymbol{Q}$.

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The Pauli Spin Matrices are of enormous practical importance in quantum theory, as we will shortly see. They satisfy a number of useful properties:

$$\sigma_x\sigma_y = i\sigma_z = -\sigma_y\sigma_x$$
$$\sigma_y\sigma_z = i\sigma_x = -\sigma_z\sigma_y$$
$$\sigma_z\sigma_x = i\sigma_y = -\sigma_x\sigma_z \quad \text{(N.B. noncommutativity)}$$

$$\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = \mathbf{I}$$
$$\sigma_x^\dagger = \sigma_x \ ; \ \sigma_y^\dagger = \sigma_y \ ; \ \sigma_z^\dagger = \sigma_z$$

where $A^\dagger = (A^T)^*$ is the adjoint (i.e. the conjugate transpose) of the matrix A.

Operators $A, B : \boldsymbol{H} \to \boldsymbol{H}$ can be composed. Thus $AB : \boldsymbol{H} \to \boldsymbol{H}$ is the operator defined by:

$$AB|x\rangle = A(B|x\rangle)$$

This defines powers of A, eg. $A^2 = AA$, $A^5 = AAAAA$, etc. Furthermore any function defined by a power series that converges, extends to operators (at least in the finite dimensional case). Examples:

$$e^{\lambda A} = \mathbf{I} + \lambda A + \tfrac{1}{2!}\lambda^2 A^2 + \tfrac{1}{3!}\lambda^3 A^3 + \tfrac{1}{4!}\lambda^4 A^4 + \ldots$$
$$\sin(\lambda A) = \lambda A - \tfrac{1}{3!}\lambda^3 A^3 + \tfrac{1}{5!}\lambda^5 A^5 - \ldots$$
etc.

Example 8: Functions of the Pauli Spin Matrices.

$$\begin{aligned}&\exp(i\theta\sigma_z)\\ &= \mathbf{I} + (i\theta\sigma_z) + \tfrac{1}{2!}(i\theta\sigma_z)^2 + \tfrac{1}{3!}(i\theta\sigma_z)^3 + \tfrac{1}{4!}(i\theta\sigma_z)^4 + \ldots\\ &= \mathbf{I} + i\theta\sigma_z - \tfrac{1}{2!}\theta^2\mathbf{I} - \tfrac{1}{3!}i\theta^3\sigma_z + \tfrac{1}{4!}\theta^4\mathbf{I} + \ldots\\ &= \cos(\theta)\mathbf{I} + i\sin(\theta)\sigma_z\end{aligned}$$

Similarly for $\exp(i\theta\sigma_x)$, $\exp(i\theta\sigma_y)$ ... and so on.

$$\begin{aligned}&\exp(i\theta\boldsymbol{n}\cdot\underline{\sigma}) = \exp(i\theta(n_x\sigma_x + n_y\sigma_y + n_z\sigma_z))\\ &= \mathbf{I} + (i\theta(n_x\sigma_x + n_y\sigma_y + n_z\sigma_z)) +\\ &\quad \tfrac{1}{2!}(i\theta(n_x\sigma_x + n_y\sigma_y + n_z\sigma_z))^2 +\\ &\quad \tfrac{1}{3!}(i\theta(n_x\sigma_x + n_y\sigma_y + n_z\sigma_z))^3 +\\ &\quad \tfrac{1}{4!}(i\theta(n_x\sigma_x + n_y\sigma_y + n_z\sigma_z))^4 + \ldots\\ &= \mathbf{I} + i\theta\boldsymbol{n}\cdot\underline{\sigma} - \tfrac{1}{2!}\theta^2\mathbf{I} - \tfrac{1}{3!}i\theta^3\boldsymbol{n}\cdot\underline{\sigma} + \tfrac{1}{4!}\theta^4\mathbf{I} + \ldots\\ &= \cos(\theta)\mathbf{I} + i\sin(\theta)\boldsymbol{n}\cdot\underline{\sigma}\end{aligned}$$

Above, $\boldsymbol{n} = (n_x, n_y, n_z)$ is a unit vector.

Matrix Elements

Suppose in $\mathbb{C}^n$, we have the *standard* orthonormal basis $\{[1, 0, 0 \ldots 0]^T, [0, 1, 0 \ldots 0]^T, [0, 0, 1 \ldots 0]^T, \ldots, [0, 0, 0 \ldots 1]^T\}$. We write this for short as $\{|0\rangle, |1\rangle, |2\rangle, \ldots, |n{-}1\rangle\}$. Suppose we have an operator A given by a matrix (a_{jk}). Then:

$$\langle j|A|k\rangle = a_{jk}$$

Thus a_{jk} is what you get when you apply A to the k'th basis vector and extract the j'th component of the result. Since A is given by the totality of such facts, we can write:

$$A = \textstyle\sum_{jk} |j\rangle a_{jk}\langle k| = \sum_{jk} |j\rangle\langle j|A|k\rangle\langle k|$$

Here $\langle k|$ is a bra vector ($\langle k|$-$|k\rangle$, bra-ket, geddit?).

(Strictly speaking a bra is an element of the dual space, i.e. an operation to linearly map each vector (ket) to a complex number. The inner product $\langle y|x\rangle$ can be seen as the application of the bra $\langle y|$ to the ket $|x\rangle$. For a ket $|x\rangle$ given by a column matrix x, the corresponding bra $\langle x|$, is given by the adjoint $x^\dagger$.)

A notable special case is the identity operator:

$$\mathbf{I} = \textstyle\sum_k |k\rangle\langle k|$$

Suppose we have a subspace Λ spanned by basis $\{|l\rangle\}_{l=0\ldots t}$. Then the projection operator $\boldsymbol{P}_\Lambda = \sum_l |l\rangle\langle l|$ takes a vector and yields the Λ component.

Special Kinds of Operator

The ***adjoint*** $A^\dagger$ of an operator A is defined by:

$$\langle y|Ax\rangle = \langle A^\dagger y|x\rangle$$

If A is $|j\rangle\langle k|$, then it is easy to see that $A^\dagger$ is $|k\rangle\langle j|$, and so our earlier notation is consistent with the new definition. The matrix elements of the adjoint of an operator, are given by the adjoint matrix.

Note that $(\lambda A + \mu B)^\dagger = \lambda^* A^\dagger + \mu^* B^\dagger$.

An operator A is ***normal*** iff:

$$A^\dagger A = AA^\dagger$$

Two kinds of normal operator are noteworthy. An operator A is ***unitary*** iff:

$$A^\dagger A = \mathbf{I} = AA^\dagger$$

An operator A is ***hermitian*** iff:

$$A = A^\dagger$$

If A is hermitian then for real λ, $e^{i\lambda A}$ is unitary.

Proof. First $(e^{i\lambda A})^\dagger = (\mathbf{I} + (i\lambda A) + \frac{1}{2!}(i\lambda A)^2 + \ldots)^\dagger = \mathbf{I} + (-i\lambda^* A^\dagger) + \frac{1}{2!}(-i\lambda^* A^\dagger)^2 + \ldots = \mathbf{I} + (-i\lambda A) + \frac{1}{2!}(-i\lambda A)^2 + \ldots = e^{-i\lambda A}$. Now $e^{-i\lambda A}e^{i\lambda A} = (\mathbf{I} + (-i\lambda A) + \frac{1}{2!}(-i\lambda A)^2 + \ldots) \times (\mathbf{I} + (i\lambda A) + \frac{1}{2!}(i\lambda A)^2 + \ldots) = \ldots = \mathbf{I} = \ldots = e^{i\lambda A}e^{-i\lambda A}$.

We can assume that the converse holds too. For every unitary U, there is a hermitian A (called the generator of U) and a λ, such that $U = e^{i\lambda A}$.

N.B. The Pauli Spin Matrices are both hermitian and unitary.

Unitary operators U are ***invertible***, in that they are operators having an ***inverse*** U^{-1} such that $UU^{-1} = \mathbf{I} = U^{-1}U$. For an invertible U, $U|v\rangle \neq \underline{0}$ unless $|v\rangle = \underline{0}$. For unitary U, the inverse is the adjoint. Note that $(U_1U_2)^{-1} = U_2^{-1}U_1^{-1}$.

Diagonalisation

Let A be an operator. A nonzero vector $|v\rangle$ such that $A|v\rangle = v|v\rangle$ where v is a complex number is called an ***eigenvector*** of A with ***eigenvalue*** v. For an eigenvalue v, $(A - v\mathbf{I})|v\rangle = \underline{0}$, so all eigenvalues are contained in $\{v \mid (A - v\mathbf{I}) \text{ is not invertible}\}$.

For normal matrices A the eigenvalues can be found by solving:

$$\det(A - v\mathbf{I}) = 0$$

which is called the ***characteristic equation*** of A. Once the eigenvalues v are found, finding the eigenvectors reduces to solving $(A - v\mathbf{I})|v\rangle = \underline{0}$.

A normal matrix has a ***complete eigenvector decomposition*** in that a normal A can be written:

$$A = \sum_i v_i |v_i\rangle\langle v_i|$$

where the $|v_i\rangle$ are eigenvectors with eigenvalues v_i, and the collection of the $|v_i\rangle$ forms an orthonormal basis for the whole space. (This is not true for arbitrary operators.)

For unitary and hermitian matrices the eigenvector decompositions have specific forms.

For unitary $U = \sum_i v_i |v_i\rangle\langle v_i|$, we have $|v_i| = 1$ for all i.

For hermitian $A = \sum_i v_i |v_i\rangle\langle v_i|$, v_i is real for all i.

This is consistent with remarks above. Consider a hermitian A. Then:

$$\begin{aligned}
&e^{i\lambda A} \\
&= \mathbf{I} + (i\lambda A) + \tfrac{1}{2!}(i\lambda A)^2 + \tfrac{1}{3!}(i\lambda A)^3 + \tfrac{1}{4!}(i\lambda A)^4 + \ldots \\
&= (\textstyle\sum_i |v_i\rangle\langle v_i|) + (i\lambda)(\sum_i v_i |v_i\rangle\langle v_i|) + \\
&\quad \tfrac{1}{2!}(i\lambda)^2(\textstyle\sum_i v_i |v_i\rangle\langle v_i|)^2 + \tfrac{1}{3!}(i\lambda)^3(\sum_i v_i |v_i\rangle\langle v_i|)^3 + \ldots \\
&= (\textstyle\sum_i 1 |v_i\rangle\langle v_i|) + (\sum_i (i\lambda v_i) |v_i\rangle\langle v_i|) + \\
&\quad (\textstyle\sum_i \tfrac{1}{2!}(i\lambda v_i)^2 |v_i\rangle\langle v_i|) + (\sum_i \tfrac{1}{3!}(i\lambda v_i)^3 |v_i\rangle\langle v_i|) + \ldots \\
&= \textstyle\sum_i (1 + (i\lambda v_i) + \tfrac{1}{2!}(i\lambda v_i)^2 + \tfrac{1}{3!}(i\lambda v_i)^3 + \ldots) |v_i\rangle\langle v_i|) \\
&= \textstyle\sum_i \exp(i\lambda v_i) |v_i\rangle\langle v_i|
\end{aligned}$$

If λ and the v_i are real then $|\exp(i\lambda v_i)| = 1$ for all i as required.

Change of Basis

In general eigenvectors will not coincide with basis vectors in the particular orthonormal basis you happen to be working with. So let $\{|u_l\rangle\}_{l=0..n-1}$ and $\{|v_i\rangle\}_{i=0..n-1}$ be two orthonormal bases. Then each $|v_i\rangle$ will be a linear combination of the $|u_l\rangle$:

$$|v_i\rangle = \textstyle\sum_l \langle u_l|v_i\rangle |u_l\rangle = \sum_l |u_l\rangle\langle u_l|v_i\rangle$$

Moreover $\langle u_j|u_i\rangle = \langle v_j|v_i\rangle = \delta_{ij}$ (1 if $i = j$, else 0); i.e.:

$$\begin{aligned}
\langle v_j|v_i\rangle &= \textstyle\sum_{lm} \langle u_m|\langle u_m|v_j\rangle^* \langle u_l|v_i\rangle |u_l\rangle \\
&= \textstyle\sum_{lm} \delta_{lm} \langle u_m|v_j\rangle^* \langle u_l|v_i\rangle = \sum_l \langle v_j|u_l\rangle\langle u_l|v_i\rangle = \delta_{ij}
\end{aligned}$$

So the $\langle u_l|v_i\rangle$ are the matrix elements of a unitary operator U since the last equality is just $U^\dagger U = \mathbf{I}$.

Under a change of basis (rotation) described by a unitary U, vectors and operators transform by:

$$|x\rangle \to U|x\rangle \quad \text{and} \quad A \to UAU^\dagger$$

respectively, since then:

$$(A|x\rangle) \to UAU^\dagger U|x\rangle = U(A|x\rangle)$$

Example 1: Rotation by unitary operator σ_x. This is:

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

It looks like it just swaps components round. The standard basis, $\{|0\rangle, |1\rangle\} = \{[1, 0]^T, [0, 1]^T\}$, i.e.

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

goes to $\{|0'\rangle, |1'\rangle\} = \{\sigma_x|0\rangle, \sigma_x|1\rangle\} = \{[0, 1]^T, [1, 0]^T\}$.

An operator eg. σ_y, goes to $\sigma_x\sigma_y\sigma_x{}^\dagger = \sigma_x\sigma_y\sigma_x = i\sigma_z\sigma_x = ii\sigma_y = -\sigma_y$, which is indeed consistent with the view that σ_x just swaps components round.

Example 2: Rotation by unitary operator *H*. This is:

$$H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

The standard basis, $\{|0\rangle, |1\rangle\} = \{[1, 0]^T, [0, 1]^T\}$, goes to $\{|0_H\rangle, |1_H\rangle\} = \{H|0\rangle, H|1\rangle\}$ which is

$$|0_H\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

$$|1_H\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

H is called the Hadamard rotation, $\{|0_H\rangle, |1_H\rangle\} = \{H|0\rangle, H|1\rangle\} = \{\frac{1}{\sqrt{2}}[1, 1]^T, \frac{1}{\sqrt{2}}[1, -1]^T\}$ is called the Hadamard basis. We will meet these frequently below.

Constructions on Spaces

It is important to be able to make bigger spaces and operators out of smaller ones. The most important constructions for this are the direct sum and the tensor product.

Direct Sums

Let $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ be spaces. We define the ***direct sum*** $\boldsymbol{H}_1 \oplus \boldsymbol{H}_2$ as the space whose vectors are pairs $(|u\rangle, |v\rangle)$ where $|u\rangle$ and $|v\rangle$ are vectors in $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ respectively. In essence $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are orthogonal components of $\boldsymbol{H}_1 \oplus \boldsymbol{H}_2$ just as the X and Y axes are orthogonal components in the plane.

The linear operations are defined as follows:

(1) $(|u_1\rangle, |v_1\rangle), (|u_2\rangle, |v_2\rangle) \in H_1 \oplus H_2 \Rightarrow$
$(|u_1\rangle, |v_1\rangle) + (|u_2\rangle, |v_2\rangle) =$
$(|u_1\rangle + |u_2\rangle, |v_1\rangle + |v_2\rangle) \in H_1 \oplus H_2$

(2) $(|u\rangle, |v\rangle) \in H_1 \oplus H_2$, $\lambda \in \mathbb{C} \Rightarrow$
$\lambda(|u\rangle, |v\rangle) = (\lambda|u\rangle, \lambda|v\rangle) \in H_1 \oplus H_2$

An obvious consequence is that $(\underline{0}, \underline{0})$ is the zero vector in $\boldsymbol{H}_1 \oplus \boldsymbol{H}_2$. For the inner product:

(3) $(|u_1\rangle, |v_1\rangle), (|u_2\rangle, |v_2\rangle) \in H_1 \oplus H_2 \Rightarrow$
$\big\langle (|u_1\rangle, |v_1\rangle) \big| (|u_2\rangle, |v_2\rangle) \big\rangle = \langle u_1|u_2\rangle + \langle v_1|v_2\rangle \in \mathbb{C}$

Operators go as follows. Let A_1 be an operator on $\boldsymbol{H}_1$ and A_2 be an operator on $\boldsymbol{H}_2$. Then $A_1 \oplus A_2$ is the operator on $\boldsymbol{H}_1 \oplus \boldsymbol{H}_2$ defined as follows:

(4) $(|u\rangle, |v\rangle) \in H_1 \oplus H_2 \Rightarrow$
$A_1 \oplus A_2(|u\rangle, |v\rangle) = (A_1|u\rangle, A_2|v\rangle) \in H_1 \oplus H_2$

Note that you don't get an operator on $\boldsymbol{H}_1 \oplus \boldsymbol{H}_2$ from just A_1 alone. If there's no obvious A_2 around, a couple of default choices for A_2 are $\mathbf{I}$ and $\mathbf{0}$ (the latter giving a projection effect).

Note also that operators of the form $A_1 \oplus A_2$ are just a small portion of all operators on $\boldsymbol{H}_1 \oplus \boldsymbol{H}_2$ as they do not mix the $\boldsymbol{H}_1$ part with the $\boldsymbol{H}_2$ part.

Consider eigenvalues and eigenvectors, let $A_1|u\rangle = u|u\rangle$, and $A_2|v\rangle = v|v\rangle$. Then:

$$A_1 \oplus A_2(|u\rangle, \underline{0}) = u(|u\rangle, \underline{0})$$

so $(|u\rangle, \underline{0})$ is an eigenvector of $A_1 \oplus A_2$ with eigenvalue u, and:

$$A_1 \oplus A_2(\underline{0}, |v\rangle) = v(\underline{0}, |v\rangle)$$

so $(\underline{0}, |v\rangle)$ is an eigenvector of $A_1 \oplus A_2$ with eigenvalue v. This covers all the possibilities.

Let $\{|u_j\rangle\}_{j=0..m-1}$ be a basis of $\boldsymbol{H}_1$, and $\{|v_k\rangle\}_{k=0..n-1}$ be a basis of $\boldsymbol{H}_2$. Each $|u_j\rangle$ can be identified with the $\boldsymbol{H}_1 \oplus \boldsymbol{H}_2$ vector $(|u_j\rangle, \underline{0})$; likewise each $|v_k\rangle$ can be identified with the $\boldsymbol{H}_1 \oplus \boldsymbol{H}_2$ vector $(\underline{0}, |v_k\rangle)$. All of

these are orthogonal, so $\dim(\boldsymbol{H}_1 \oplus \boldsymbol{H}_2) = \dim(\boldsymbol{H}_1) + \dim(\boldsymbol{H}_2)$. N.B. For every $\boldsymbol{H}$ and every basis $\{|u_j\rangle\}_j$ of $\boldsymbol{H}$, $\boldsymbol{H}$ is the direct sum of all the subspaces $\boldsymbol{H}_j$ corresponding to the individual basis vectors $|u_j\rangle$.

Tensor Products for Dummies

An enormous amount can be said about tensor products, and from the point of view of quantum theory tensor products are far more important than direct sums. We will keep it as simple as we can.

Let $\boldsymbol{H}_1$ with basis $\{|u_j\rangle\}_{j=0..m-1}$, and $\boldsymbol{H}_2$ with basis $\{|v_k\rangle\}_{k=0..n-1}$ be spaces. Then the ***tensor product*** $\boldsymbol{H}_1 \otimes \boldsymbol{H}_2$ has as one possible basis the collection $\{|u_j\rangle \otimes |v_k\rangle\}_{j=0..m-1,k=0..n-1}$ (and thus $\dim(\boldsymbol{H}_1 \otimes \boldsymbol{H}_2) = \dim(\boldsymbol{H}_1) \times \dim(\boldsymbol{H}_2)$). This means that a vector in $\boldsymbol{H}_1 \otimes \boldsymbol{H}_2$ has the general form:

$$\Sigma_{jk}\, a_{jk}|u_j\rangle \otimes |v_k\rangle$$

The tensor product $\otimes$ extends by linearity to all of $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$. Thus if:

$$|u\rangle = \Sigma_j\, a_j|u_j\rangle \quad ; \quad |v\rangle = \Sigma_k\, b_k|v_k\rangle$$

then:

$$|u\rangle \otimes |v\rangle = \Sigma_{jk}\, a_j.b_k|u_j\rangle \otimes |v_k\rangle$$

The best way to think of an object like $|u\rangle \otimes |v\rangle$ is as a pair of 'siamese twins'. Inextricably tied together

into an indivisible entity, but nevertheless retaining some individuality as vectors in their own right.

The linear operations are defined as follows:

(1) $\Sigma_{jk}\, a_{jk}|u_j\rangle \otimes |v_k\rangle, \Sigma_{jk}\, b_{jk}|u_j\rangle \otimes |v_k\rangle \in H_1 \otimes H_2 \Rightarrow$
$\Sigma_{jk}\, a_{jk}|u_j\rangle \otimes |v_k\rangle + \Sigma_{jk}\, b_{jk}|u_j\rangle \otimes |v_k\rangle =$
$\Sigma_{jk}\, (a_{jk} + b_{jk})|u_j\rangle \otimes |v_k\rangle \in H_1 \otimes H_2$

(2) $\Sigma_{jk}\, a_{jk}|u_j\rangle \otimes |v_k\rangle \in H_1 \otimes H_2$, $\lambda \in \mathbb{C} \Rightarrow$
$\lambda \Sigma_{jk}\, a_{jk}|u_j\rangle \otimes |v_k\rangle = \Sigma_{jk}\, \lambda a_{jk}|u_j\rangle \otimes |v_k\rangle \in H_1 \otimes H_2$

An obvious consequence is that $\underline{0} \otimes \underline{0}$ is the zero vector in $\boldsymbol{H}_1 \otimes \boldsymbol{H}_2$. But that isn't the only way of writing it, since by linearity, for any $|u\rangle$ or $|v\rangle$ we have $|u\rangle \otimes \underline{0} = \underline{0} \otimes \underline{0} = \underline{0} \otimes |v\rangle$.

For the inner product:

(3) $(|\underline{u}\rangle \otimes |\underline{v}\rangle), (|\overline{u}\rangle \otimes |\overline{v}\rangle) \in H_1 \otimes H_2 \Rightarrow$
$\big\langle (|\underline{u}\rangle \otimes |\underline{v}\rangle) \big| (|\overline{u}\rangle \otimes |\overline{v}\rangle) \big\rangle = \langle \underline{u}|\overline{u}\rangle \times \langle \underline{v}|\overline{v}\rangle \in \mathbb{C}$

so $\big\langle (\Sigma_{jk}\, a_{jk}|u_j\rangle|v_k\rangle) \big| (\Sigma_{j'k'}\, b_{j'k'}|u_{j'}\rangle|v_{k'}\rangle) \big\rangle = \Sigma_{jk}\, a_{jk}{}^* b_{jk}$.

Operators go as follows. Let A_1 be an operator on $\boldsymbol{H}_1$ and A_2 be an operator on $\boldsymbol{H}_2$. Then $A_1 \otimes A_2$ is the operator on $\boldsymbol{H}_1 \otimes \boldsymbol{H}_2$ defined as follows:

(4) $\Sigma_{jk}\, a_{jk}|u_j\rangle \otimes |v_k\rangle \in H_1 \otimes H_2 \Rightarrow$
$A_1 \otimes A_2 \Sigma_{jk}\, a_{jk}|u_j\rangle \otimes |v_k\rangle =$
$\Sigma_{jk}\, a_{jk}(A_1|u_j\rangle) \otimes (A_2|v_k\rangle) \in H_1 \otimes H_2$

Note that you don't get an operator on $\boldsymbol{H}_1 \otimes \boldsymbol{H}_2$ from

just A_1 alone. If there's no obvious A_2 around, the default for A_2 is $\mathbf{I}$ (N.B. Choosing $\mathbf{0}$ reduces $A_1 \otimes A_2$ to the $\mathbf{0}$ operator by linearity arguments as above). Also operators of the form $A_1 \otimes A_2$ are just a small portion of all operators on $\boldsymbol{H}_1 \otimes \boldsymbol{H}_2$.

Consider eigenvalues and eigenvectors, let $A_1|u\rangle = u|u\rangle$, and $A_2|v\rangle = v|v\rangle$. Then:

$$A_1 \otimes A_2|u\rangle \otimes |v\rangle = u.v|u\rangle \otimes |v\rangle$$

so that $|u\rangle \otimes |v\rangle$ is an eigenvector of $A_1 \otimes A_2$ with eigenvalue $u.v$. This covers all the possibilities.

Notes

The tensor product is a bit like a 'multiplication that doesn't go all the way'. Tensor products are quite a bit trickier than direct sums. In a direct sum, once you have an element of $\boldsymbol{H}_1 \oplus \boldsymbol{H}_2$ in the form $(|u\rangle, |v\rangle)$ that's it. That form is canonical: essentially, you cannot rewrite $(|u\rangle, |v\rangle) = (|u\rangle, \underline{0}) + (\underline{0}, |v\rangle)$ in any different way, while still retaining the $(|\text{-}\rangle, |\text{-}\rangle)$ shape.

Tensor products are different. Writing an element of $\boldsymbol{H}_1 \otimes \boldsymbol{H}_2$ in the canonical way with respect to a basis $\{|u_j\rangle \otimes |v_k\rangle\}_{j=0..m-1,k=0..n-1}$, i.e. in the form:

$$|\psi\rangle = \Sigma_{jk}\, a_{jk}|u_j\rangle \otimes |v_k\rangle$$

gives no clue about whether $|\psi\rangle$ can be written in the form $|u\rangle \otimes |v\rangle$, or in the form $\Sigma_q\, \lambda_q|u_q\rangle \otimes |v_q\rangle$ with q ranging over fewer than $m.n$ elements, and if it can, whether the representation is unique or not.

Example: Consider the standard and Hadamard bases, $\{|0\rangle, |1\rangle\}$ and $\{|0_H\rangle, |1_H\rangle\} = \{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$. It is easy to check directly that:

$$(|0\rangle\otimes|0\rangle + |1\rangle\otimes|1\rangle) = (|0_H\rangle\otimes|0_H\rangle + |1_H\rangle\otimes|1_H\rangle)$$

(I tend to call such surprising but true equalities between terms of the form $\sum_q \lambda_q|u_q\rangle\otimes|v_q\rangle$ 'tensor product sleight of hand'; we will see more of this in due course.) This just gets worse with multiple tensor products, i.e. $\boldsymbol{H}_1\otimes\boldsymbol{H}_2\otimes \ldots \otimes\boldsymbol{H}_n$.

We will often write $|u\rangle\otimes|v\rangle$ as $|u\otimes v\rangle$ or as $|u\rangle|v\rangle$ or as $|u,v\rangle$ or just as $|uv\rangle$ when the tensor product structure can be infered.

We will need tensor powers of $\boldsymbol{Q}$, i.e. $\boldsymbol{Q}^{\otimes n}$.

Starting with the standard basis $\{|0\rangle, |1\rangle\}$, for the tensor square we get a basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. For the tensor cube we get $\{|000\rangle, |001\rangle, |010\rangle, |011\rangle, |100\rangle, |101\rangle, |110\rangle, |111\rangle\}$. Etc.

In general the standard basis vectors of the n'th tensor power of $\boldsymbol{Q}$ can be labelled by the binary strings of length n.

Now apply the Hadamard rotation to the basis vector $|0\rangle$ in each factor of an n-fold tensor power of $\boldsymbol{Q}$.

$$\begin{aligned} &H^{\otimes n}|0\ldots0\rangle = H|0\rangle\otimes H|0\rangle\otimes \ldots \otimes H|0\rangle \\ &= [\tfrac{1}{\sqrt{2}}(|0\rangle+|1\rangle)]\otimes[\tfrac{1}{\sqrt{2}}(|0\rangle+|1\rangle)]\otimes \ldots \otimes[\tfrac{1}{\sqrt{2}}(|0\rangle+|1\rangle)] \\ &= \tfrac{1}{\sqrt{2}}^n(|00\ldots00\rangle+|00\ldots01\rangle+|00\ldots10\rangle+\ldots) \\ &= \tfrac{1}{\sqrt{2}}^n \textstyle\sum_z |z\rangle \end{aligned}$$

where the sum over z is over all binary strings of length n. So just by a simultaneous change of basis in each of the factors of the n-fold tensor product, we have a uniformly weighted sum of vectors representing all binary strings of length n.

Some more: Apply the Hadamard rotation to an arbitrary standard basis vector $|w\rangle = |011\ldots0\rangle$ say.

$$\begin{aligned} &H^{\otimes n}|w\rangle = H^{\otimes n}|011\ldots0\rangle \\ &= H|0\rangle\otimes H|1\rangle\otimes H|1\rangle\otimes \ldots \otimes H|0\rangle \\ &= [\tfrac{1}{\sqrt{2}}(|0\rangle+|1\rangle)] && (+ \textit{ for } |0\rangle \textit{ factors}) \\ &\otimes[\tfrac{1}{\sqrt{2}}(|0\rangle-|1\rangle)] && (- \textit{ for } |1\rangle \textit{ factors}) \\ &\otimes[\tfrac{1}{\sqrt{2}}(|0\rangle-|1\rangle)] && (- \textit{ for } |1\rangle \textit{ factors}) \\ &\otimes\ldots\otimes \\ &\otimes[\tfrac{1}{\sqrt{2}}(|0\rangle+|1\rangle)] && (+ \textit{ for } |0\rangle \textit{ factors}) \\ &= \tfrac{1}{\sqrt{2}}^n \textstyle\sum_z (-)^{w.z}|z\rangle \end{aligned}$$

Here $w.z$ is the scalar product between the two bit vectors w = 011...0 and z = (whatever), and is the total number of bit positions in which both w and z have a 1. Thus:

$$\boxed{H^{\otimes n}|w\rangle = \tfrac{1}{\sqrt{2}}^n \textstyle\sum_z (-)^{w.z}|z\rangle}$$

Learn to love this formula.

Quantum Theory

Anybody who is not shocked by quantum theory has not understood it.
Niels Bohr

It requires a certain degree of sophistication ... to grasp the existence of quantum mechanics, I would say there's much more difference from this point of view, between a human being who knows quantum mechanics and one who doesn't, than between one that doesn't and the other great apes. The big divide is between people who know quantum mechanics and people who don't. The ones who don't, for this purpose, are goldfish.
Murray Gell-Mann

Quantum theory as we now understand it endured a long, painful and erratic birth process, from the oversimplified attempts of the 'old quantum theory', via subtle adaptations of the most sophisticated formulations of classical mechanics and classical electrodynamics ever conceived[1], to the enduring framework of today. A framework that moreover continues to disdainfully rebut all experimental challenges, and the understanding of which is still maturing.

Sadly, we don't have time to cover any of this fascinating stuff in any detail.

We just present quantum theory as an axiomatic system in linear algebra. This together with the rules for relating the linear algebra to observable phenomena is enough to get quantum theory to do the job that we want.

1. In physics, it is palpably the case that the search for abstraction in physical theory has led to some of the most profound advances, a fact that even the most recalcitrant experimentalist would find hard to deny. In computer science it is palpably *not* the case that the search for abstraction in computation theory has led to deep advances; increasing elegance often leads to increasing irrelevance; *vede* Church's Thesis.

Principles of Quantum Mechanics

Different accounts give the principles in different ways. However it all amounts to the same thing. Here's one way of doing it.

I The (pure) states of a quantum system are described by the unit vectors up to phase (rays) of a Hilbert Space.

So the phase of a state doesn't matter. The vectors $|\psi\rangle$ and $e^{i\alpha}|\psi\rangle$ describe the same physical state. Note that *every* unit vector in the state space is a potential configuration of the system.

II The state space of a composite system (of distinguishable components) is the tensor product of the component state spaces.

So the state space of an electron and a proton is the tensor product of the electron state space and the proton state space.

III The time evolution from time t_0 to time t_1 of the state of a (closed) quantum system is given by a unitary operator $U(t_0,t_1)$.

So if the state at time t_0 is $|\psi(t_0)\rangle$ then the state at time $t_1 > t_0$ is $|\psi(t_1)\rangle = U(t_0,t_1)|\psi(t_0)\rangle$.

[Aside: The unitary $U(t_0,t_1)$ can be written $U(t_0,t_1) = e^{-itH}$, where $t = t_1 - t_0$ is real, and H is a hermitian operator, as we know. Up to a factor of $\hbar$, this H is the Hamiltonian, the generator of time evolution in quantum theory. For physics, knowing the correct H for any given system is ***the*** crucial question. The dynamics, (given eg. by the Schrödinger equation) relates the time derivative of the state to H. For us, H is of no concern.]

IV Every physical observable *A* corresponds to a hermitian operator A on the state space. The eigenvalues of A are the only values that can be obtained for *A* by experiment.

This is radically different to the classical idea of a physical observable.

For a composite system, whose state space is $\boldsymbol{H}_1 \otimes \boldsymbol{H}_2$, observable A_1 of system $\boldsymbol{H}_1$ becomes the observable $A_1 \otimes \mathbf{I}_2$ of $\boldsymbol{H}_1 \otimes \boldsymbol{H}_2$ and observable A_2 of system $\boldsymbol{H}_2$ becomes the observable $\mathbf{I}_1 \otimes A_2$ of $\boldsymbol{H}_1 \otimes \boldsymbol{H}_2$. These are both different from observable $A_1 \otimes A_2$ which is also an observable of $\boldsymbol{H}_1 \otimes \boldsymbol{H}_2$.

Va A measurement M_B is a set of projections $M = \{\boldsymbol{P}_{\Lambda_1} = \Sigma_{l_1} |l_1\rangle\langle l_1|, \ldots, \boldsymbol{P}_{\Lambda_m} = \Sigma_{l_m} |l_m\rangle\langle l_m|\}$ corresponding to a partition of a basis $B = \{\{|l_1\rangle\}_{l_1}, \ldots, \{|l_m\rangle\}_{l_m}\} = \{\{|l_{1,0}\rangle \ldots |l_{1,n_1-1}\rangle\}, \ldots, \{|l_{m,0}\rangle \ldots |l_{m,n_m-1}\rangle\}\}$ of the state space.

Thus $\boldsymbol{P}_{\Lambda_1} + \ldots + \boldsymbol{P}_{\Lambda_m} = \mathbf{I} = \Sigma_k |k\rangle\langle k|$ (the identity).

Vb Performing a measurement M_B on a state $|\psi\rangle$ nondeterministically sends $|\psi\rangle$ to one of the following outcomes:

$$N_1^{-1}\boldsymbol{P}_{\Lambda_1}|\psi\rangle = N_1^{-1}\Sigma_{l_1}|l_1\rangle\langle l_1|\psi\rangle \quad \text{where}$$
$$N_1 = \|\boldsymbol{P}_{\Lambda_1}|\psi\rangle\|, \text{ with probability } \|\boldsymbol{P}_{\Lambda_1}|\psi\rangle\|^2$$
$$\ldots \qquad \ldots$$
$$N_m^{-1}\boldsymbol{P}_{\Lambda_m}|\psi\rangle = N_m^{-1}\Sigma_{l_m}|l_m\rangle\langle l_m|\psi\rangle \quad \text{where}$$
$$N_m = \|\boldsymbol{P}_{\Lambda_m}|\psi\rangle\|, \text{ with probability } \|\boldsymbol{P}_{\Lambda_m}|\psi\rangle\|^2$$

Each outcome is a 'projection and renormalization' of $|\psi\rangle$ by one of the $\boldsymbol{P}_{\Lambda_j}$ with probability $\|\boldsymbol{P}_{\Lambda_j}|\psi\rangle\|^2$. [This is often called the projection postulate.]

Vc Performing an observation of an observable *A* (experiment to measure *A*), means performing the measurement M_B, where *B* is a basis of eigenvectors of A partitioned according to the eigenspaces of A. For each possible outcome of the measurement, $N_j^{-1}\boldsymbol{P}_{\Lambda_j}|\psi\rangle$, the value obtained for *A* is the eigenvalue v_j, where Λ_j is the eigenspace of eigenvectors for v_j.

Since the eigenvectors of *A* form a basis of the state space $\{|l_{1,0}\rangle \ldots |l_{1,n_1-1}\rangle, |l_{2,0}\rangle \ldots |l_{2,n_2-1}\rangle, \ldots, |l_{m,0}\rangle \ldots |l_{m,n_m-1}\rangle\}$, the total probability for obtaining *some* outcome is:

$$\|\boldsymbol{P}_{\Lambda_1}|\psi\rangle\|^2 + \ldots + \|\boldsymbol{P}_{\Lambda_m}|\psi\rangle\|^2 = \Sigma_k |\langle k|\psi\rangle|^2$$
$$= |\langle\psi|\psi\rangle|^2 = 1$$

(because $|\psi\rangle$ is a unit vector). Thus the probabilities add up to 1, as they should. [Moreover, the fact that all observable probabilities look like $\|\boldsymbol{P}_{\Lambda_j}|\psi\rangle\|^2$, explains why the the phase of a state does not matter.]

When a large number of observations of *A* on state $|\psi\rangle$ are performed, the average (or ***expectation value*** (of *A* in state $|\psi\rangle$) that is obtained is:

$$\Sigma_j v_j \|\boldsymbol{P}_{\Lambda_j}|\psi\rangle\|^2$$
$$= (\Sigma_{l_1}\langle\psi|l_1\rangle\langle l_1|) v_1 (\Sigma_{l_1}|l_1\rangle\langle l_1|\psi\rangle)$$
$$+ \qquad \ldots \qquad \ldots$$
$$+ (\Sigma_{l_m}\langle\psi|l_m\rangle\langle l_m|) v_m (\Sigma_{l_m}|l_m\rangle\langle l_m|\psi\rangle)$$
$$= \langle\psi|(\Sigma_{l_j}|l_j\rangle v_j \langle l_j|)|\psi\rangle$$
$$= \langle\psi|A|\psi\rangle$$

[N.B. For any projection operator $\boldsymbol{P}_{\Lambda_j}$, $\boldsymbol{P}_{\Lambda_j}^{\,2} = \boldsymbol{P}_{\Lambda_j}$.]

Example 1: An observable *A* corresponding to hermitian operator $A = |4\rangle 4\langle 4| + |5\rangle 5\langle 5| + |6\rangle 6\langle 6|$. The collection of projections for this is $\{\boldsymbol{P}_4 = |4\rangle\langle 4|, \boldsymbol{P}_5 = |5\rangle\langle 5|, \boldsymbol{P}_6 = |6\rangle\langle 6|\}$, corresponding to the partitioned basis $\{\{|4\rangle\}, \{|5\rangle\}, \{|6\rangle\}\}$. In state $|\psi\rangle = \frac{1}{3}|4\rangle + \frac{2\sqrt{2}}{3}|6\rangle$, the probabilities of observing the various outcomes is: $|4\rangle$: $\frac{1}{9}$, $|5\rangle$: 0, $|6\rangle$: $\frac{8}{9}$; while the expectation value of *A* in state $|\psi\rangle$ is $\frac{1}{9}4 + \frac{8}{9}6 = \frac{52}{9}$.

Example 2: An observable *A* corresponding to hermitian operator $A = |4\rangle 4\langle 4| + |5a\rangle 5\langle 5a| + |5b\rangle 5\langle 5b|$. The collection of projections for this is $\{\boldsymbol{P}_4 = |4\rangle\langle 4|, \boldsymbol{P}_5 = |5a\rangle\langle 5a| + |5b\rangle\langle 5b|\}$, corresponding to the partitioned basis $\{\{|4\rangle\}, \{|5a\rangle, |5b\rangle\}\}$. The eigenvalue 5 is ***degenerate***. In state $|\psi\rangle = \frac{1}{3}|4\rangle + \frac{2}{3}|5a\rangle + \frac{2}{3}|5b\rangle$, the probabilities of observing the various outcomes is: $|4\rangle$: $\frac{1}{9}$, $\frac{1}{\sqrt{2}}|5a\rangle + \frac{1}{\sqrt{2}}|5b\rangle$: $\frac{8}{9}$; while the expectation value of *A* in state $|\psi\rangle$ is $\frac{1}{9}4 + \frac{8}{9}5 = \frac{44}{9}$. N.B. Unless the ratio between the $|5a\rangle$ and $|5b\rangle$ components in $|\psi\rangle$ is known to be 1:1 precisely, if 5 is observed, then it will not be known that the resulting state is precisely $\frac{1}{\sqrt{2}}|5a\rangle + \frac{1}{\sqrt{2}}|5b\rangle$.

Example 3: The measurement corresponding to the basis partition $\{\{|0\rangle\}, \{|1\rangle\}\}$ (on the space $\boldsymbol{Q}$). The projections are $\{\boldsymbol{P}_0 = |0\rangle\langle 0|, \boldsymbol{P}_1 = |1\rangle\langle 1|\}$.

Example 4: The measurement corresponding to the basis partition $\{\{|0\rangle, |1\rangle\}\}$ (on the space $\boldsymbol{Q}$). The projection is $\mathbf{I} = |0\rangle\langle 0| + |1\rangle\langle 1|$, i.e. the identity. This is the null measurement, since every state on $\boldsymbol{Q}$ remains unchanged under the projection $\mathbf{I}$.

Note the differences between Examples 1 and 2, and between Examples 3 and 4, even though each pair is constructed from (effectively) the same basis. The granularity of the partition of the basis is what reveals the information yielded by a measurement. Note the difference in terminology:

- A ***measurement*** projects the state (via the projections for a partition of the basis).
- An ***observation*** associates numbers with the projections.

N.B. A projection (i.e. an operator of the form $\boldsymbol{P}_\Lambda = \sum_l |l\rangle\langle l|$ where the summation is over some subset of a basis) is itself a hermitian operator. So it yields an observable, with eigenvalues 1 and 0 (or just 1 if the projection is $\mathbf{I}$, or just 0 if the projection is $\mathbf{0}$). The associated projections are $\boldsymbol{P}_\Lambda = \sum_l |l\rangle\langle l|$ (for eigenvalue 1), and $\boldsymbol{P}_{\Lambda^\perp} = \mathbf{I} - \boldsymbol{P}_\Lambda$ (for eigenvalue 0). Note the dual role of projections.

So quantum systems either evolve ***smoothly***, via a continuous unitary $U(t_0,t_1)$, or ***discontinuously***, via a measurement, when the state jumps to an eigenstate of a suitable projection.

The balance between unitary evolution and measurement projections in the dynamics of a physical system, determines the extent to which the system behaves either like a quantum transition system (unitary evolution) or a stochastic transition system (unitary evolution followed by a measurement).

Those were the principles ...

The principles however don't tell you anything at all about what state space corresponds to what physical system, what hermitian operator corresponds to what physical observable, etc. That remains an issue for patient exploration in the lab.

Conventional presentations of quantum theory invariably discuss issues like:

- Schrödinger's equation and wave functions.
- The Canonical Commutation Relations.
- Heisenberg's Uncertainty Principle.
- Bohr Complementarity.
- The nature of measurement.
- The classical limit of quantum theory.

Contemporary theoretical and experimental work shows that most of these things are at the very least, quite a bit more subtle than was assumed in the early days. We will not discuss them beyond what has been said already.

Basic Stuff

<u>Electrons for Dummies</u>

Electrons are the spin-$\frac{1}{2}$ particles of choice. There are of course many others. They all enjoy a wealth of fascinating properties. We will take a drastically oversimplified view: for us the state of an electron can be described using a unit vector in $\boldsymbol{Q}$.

This is indeed a drastic oversimplification; we can't even speak about where the electron might *be*; we can only discuss the electron's internal state, i.e. its spin.[2] (Such vectors in $\boldsymbol{Q}$ are called spinors.)

A unit vector is $|\psi\rangle = [a, b]^{\mathrm{T}}$ where $|a|^2+|b|^2 = 1$. So there is a θ such that $|a| = \cos(\theta/2)$, $|b| = \sin(\theta/2)$. So $a = e^{i\gamma}\cos(\theta/2)$, $b = e^{i(\gamma+\phi)}\sin(\theta/2)$. Removing the overall (and irrelevant) phase factor $e^{i\gamma}$, we get:

$$|\psi\rangle = [\cos(\theta/2), e^{i\phi}\sin(\theta/2)]^{\mathrm{T}}$$

The total range of possibilities is given by letting $0 \le \theta \le \pi$, $0 \le \phi \le 2\pi$. This parameterisation of the unit vectors (with the first component's phase factor fixed at 1), neatly corresponds to the surface of a sphere, with θ as a "latitude" parameter, and ϕ as longitude. This is the Bloch sphere.

2. Even this is not strictly true. The representation given only works for *nonrelativistic* electrons. You can't handle Lorenz boosts, even if you were to include configuration space information in the description; space inversion would screw up.

The fundamental observables for spin-$\frac{1}{2}$ particles are the ***spin polarisations***, which we take to be given (up to factors of $\hbar$) by:

$$\boldsymbol{n}\cdot\underline{\sigma} = (n_x\sigma_x+n_y\sigma_y+n_z\sigma_z)$$

where $\boldsymbol{n} = (n_x, n_y, n_z)$ is a unit vector (in 3-D space), and $\underline{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is the spin operator vector.

[Aside: In fact, more fundamental physically, is the operator vector $\boldsymbol{S} = (S_x, S_y, S_z) = \frac{1}{2}\underline{\sigma}$. Why the factor $\frac{1}{2}$ in particular? Because S_x, S_y, S_z, must satisfy the angular momentum commutation relations. These are basic physical requirements for any system that carries a representation of 3-D space rotation (as does spin), and are given (again up to factors of $\hbar$) by:

$$[S_x, S_y]_- = S_xS_y - S_yS_x = iS_z$$
$$[S_y, S_z]_- = S_yS_z - S_zS_y = iS_x$$
$$[S_z, S_x]_- = S_zS_x - S_xS_z = iS_y$$

Only the factor of $\frac{1}{2}$ makes this come out right.]

Recall that the only values of spin that an experiment on an individual electron can observe are the eigenvalues of the spin polarisation observable. On the other hand, given an ensemble of spins prepared in the same state, the ensemble average is given by the expectation value of the spin polarisation observable.

To start with we look at the special cases where $\boldsymbol{n}$ is a unit vector parallel to the coordinate axes, i.e. at σ_x, σ_y, σ_z. The eigenvalues and eigenvectors of these are easy enough to find. It is enough to look at σ_z:

$$\sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

It is easy to check that this has two (normalised) eigenvectors. These are $|0\rangle = [1,0]^T$ with eigenvalue $+1$, or 'spin up'; and $|1\rangle = [0,1]^T$ with eigenvalue -1 or 'spin down'. We will often refer to these as $|+_z\rangle$ and $|-_z\rangle$. (Why?)

By rotational symmetry (or by explicit calculation) the eigenvalues of σ_x and σ_y are also $+1$ and -1. It is easy to find the relevant eigenvectors, so:

σ_x: for $+1$, $\frac{1}{\sqrt{2}}[1,1]^T = |+_x\rangle$; for -1, $\frac{1}{\sqrt{2}}[1,-1]^T = |-_x\rangle$
σ_y: for $+1$, $\frac{1}{\sqrt{2}}[1,i]^T = |+_y\rangle$; for -1, $\frac{1}{\sqrt{2}}[1,-i]^T = |-_y\rangle$
σ_z: for $+1$, $[1,0]^T = |+_z\rangle$; for -1, $[0,1]^T = |-_z\rangle$

Note that $\boldsymbol{n}\cdot\underline{\sigma}$ is not only hermitian, but unitary (because the $\sigma_x, \sigma_y, \sigma_z$, are). So $\boldsymbol{n}\cdot\underline{\sigma}$ can generate both smooth evolution (i.e. no measurement, and the state evolves according to the unitary $\boldsymbol{n}\cdot\underline{\sigma}$), and also discontinuous evolution (i.e. a measurement is made of the hermitian observable $\boldsymbol{n}\cdot\underline{\sigma}$).

Suppose we start with a $|+_z\rangle$ eigenvector of σ_z, and now we consider various scenarios.

(I) A measurement of σ_z is made: $+1$ is obtained with certainty. The state remains as $[1,0]^T$. Next σ_x is measured. Then outcomes $+1$ and -1 are equally probable, because:

$$|\langle \tfrac{1}{\sqrt{2}}[1,1]^T | [1,0]^T \rangle|^2 = \tfrac{1}{2} = |\langle \tfrac{1}{\sqrt{2}}[1,-1]^T | [1,0]^T \rangle|^2$$

So the state ends up as either $\frac{1}{\sqrt{2}}[1,1]^T$ or $\frac{1}{\sqrt{2}}[1,-1]^T$.

(II) First σ_x is measured. Outcomes $+1$ and -1 are equally probable as before, and the state becomes either $\frac{1}{\sqrt{2}}[1,1]^T$ or $\frac{1}{\sqrt{2}}[1,-1]^T$. Focus on $+1$ and $\frac{1}{\sqrt{2}}[1,1]^T$. Next σ_z is measured, then outcomes $+1$ and -1 are equally probable, by a similar calculation. Also the -1 and $\frac{1}{\sqrt{2}}[1,-1]^T$ case is similar. In any event the state ends up as either $[1,0]^T$ or $[0,1]^T$ with equal probability. (Cf. the stochastic TS of lecture 1.)

(III) First the state is allowed to evolve according to the unitary σ_x. No measurement is made. Since the evolution is unitary, the state becomes $\sigma_x|+_z\rangle = |-_z\rangle = \frac{1}{\sqrt{2}}(\frac{1}{\sqrt{2}}[1,1]^T - \frac{1}{\sqrt{2}}[1,-1]^T)$. Note that there is no ambiguity. Now the state is allowed to evolve according to the unitary σ_z. No measurement is made. The state becomes:

$$\begin{aligned} &\sigma_z \tfrac{1}{\sqrt{2}}(\tfrac{1}{\sqrt{2}}[1,1]^T - \tfrac{1}{\sqrt{2}}[1,-1]^T) \\ &= \tfrac{1}{\sqrt{2}}(\tfrac{1}{\sqrt{2}}[1,-1]^T - \tfrac{1}{\sqrt{2}}[1,1]^T) \\ &= -|-_z\rangle \end{aligned}$$

(or more simply $\sigma_z[0,1]^T = [0,-1]^T$). Note there is no ambiguity; the state is $[0,-1]^T$ without doubt. Now σ_z is measured. The outcome -1 is certain. (Cf. the quantum TS of lecture 1.)

These sequences of events are quite different, and show two important things which it is vital to appreciate:

- The order in which you do things makes a very big difference in quantum mechanics (regardless of whether the 'things' are periods of unitary evolution, or measurements).
- The points at which you perform measurements and the measurements you perform make a big difference in quantum mechanics.

Now we return to the general case.

What is the expectation value of $\boldsymbol{n}\cdot\underline{\sigma}$ in state $|\psi\rangle = [\cos(\theta/2), e^{i\phi}\sin(\theta/2)]^T$? Well:

$$\boldsymbol{n}\cdot\underline{\sigma} = \begin{bmatrix} n_z & n_x - i n_y \\ n_x + i n_y & -n_z \end{bmatrix}$$

and so:

$$\langle\psi|\boldsymbol{n}\cdot\underline{\sigma}|\psi\rangle =$$

$$[\cos(\theta/2), e^{-i\phi}\sin(\theta/2)] \begin{bmatrix} n_z & n_x - i n_y \\ n_x + i n_y & -n_z \end{bmatrix} \begin{bmatrix} \cos(\theta/2) \\ e^{i\phi}\sin(\theta/2) \end{bmatrix}$$

After a bit of simplification this gives:

$$\begin{aligned} \langle\psi|\boldsymbol{n}\cdot\underline{\sigma}|\psi\rangle = (&n_x\sin(\theta)\cos(\phi) + \\ &n_y\sin(\theta)\sin(\phi) + \\ &n_z\cos(\theta)) \end{aligned}$$

Now $(\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta))$ are the x, y, z, components in spherical polar coordinates of a unit vector built from $|\psi\rangle$, $\underline{\psi} = (\psi_x, \psi_y, \psi_z)$, in 3-D space (just check $\psi_x^2 + \psi_y^2 + \psi_z^2$), so we get finally:

$$\boxed{\langle\psi|\boldsymbol{n}\cdot\underline{\sigma}|\psi\rangle = \boldsymbol{n}\cdot\underline{\psi} = \cos(\angle(\boldsymbol{n},\underline{\psi}))}$$

where $\angle(\boldsymbol{n},\underline{\psi})$ denotes the angle between $\boldsymbol{n}$ and $\underline{\psi}$ (in 3-D space).

N.B.1. An easier way of obtaining this is to assume that the coordinate axes are such that the z axis is aligned with $\boldsymbol{n}$. Then $n_x = n_y = 0$, and $n_z = 1$, so we get just $\cos(\theta) = \cos(\angle(\boldsymbol{n}, \underline{\psi}))$. The generalisation follows from the rotational symmetry of the shape $\boldsymbol{n}\cdot\underline{\sigma}$; it is an inner product, and rotations are *defined* to preserve inner products.[3]

N.B.2. Do not confuse the various unit vectors here. $|\psi\rangle$ is a unit vector in $\boldsymbol{Q}$; while $\boldsymbol{n}$ and $\underline{\psi}$ are unit vectors in real 3-D space. Moreover $\underline{\sigma}$ is more than just a vector in 3-D space, it is a *vector operator* acting in $\boldsymbol{Q}$.

In general the eigenvectors of $\boldsymbol{n}\cdot\underline{\sigma}$ are:

For $+1$, $1/\sqrt{2(1+n_z)}\,[1+n_z,\ n_x+\mathrm{i}n_y]^{\mathrm{T}} = |+_{\boldsymbol{n}}\rangle$

For -1, $1/\sqrt{2(1-n_z)}\,[n_z-1,\ n_x+\mathrm{i}n_y]^{\mathrm{T}} = |-_{\boldsymbol{n}}\rangle$

For a general state $|\psi\rangle$, the probability that an observation of $\boldsymbol{n}\cdot\underline{\sigma}$ will detect $+1$ is:

$$|\langle +_{\boldsymbol{n}}|\psi\rangle|^2 = \tfrac{1}{2}(1 + \cos(\angle(\boldsymbol{n}, \underline{\psi})))$$

and for -1 it is:

$$|\langle -_{\boldsymbol{n}}|\psi\rangle|^2 = \tfrac{1}{2}(1 - \cos(\angle(\boldsymbol{n}, \underline{\psi})))$$

This agrees with $\langle\psi|\boldsymbol{n}\cdot\underline{\sigma}|\psi\rangle$ since:

$$\langle\psi|\boldsymbol{n}\cdot\underline{\sigma}|\psi\rangle = +\langle\psi|+_{\boldsymbol{n}}\rangle\langle +_{\boldsymbol{n}}|\psi\rangle - \langle\psi|-_{\boldsymbol{n}}\rangle\langle -_{\boldsymbol{n}}|\psi\rangle$$

3. N.B. This is the tip of a huge group theoretic iceberg.

More than one Electron

We use the same techniques to get probability amplitudes involving more than one electron. Consider a two-electron system with state space $\boldsymbol{Q}_1 \otimes \boldsymbol{Q}_2$. Let us calculate the expectation value of the observable:

$$(\boldsymbol{n}_1\cdot\underline{\sigma}_1) \otimes (\boldsymbol{n}_2\cdot\underline{\sigma}_2)$$

in the so-called singlet state:

$$|sing\rangle = \frac{1}{\sqrt{2}}(|+-\rangle - |-+\rangle)$$

This can be done simply enough.

$$\begin{aligned}
&\langle sing|(\boldsymbol{n}_1\cdot\underline{\sigma}_1) \otimes (\boldsymbol{n}_2\cdot\underline{\sigma}_2)|sing\rangle \\
&= \tfrac{1}{2}(\langle +-| - \langle -+|)(\boldsymbol{n}_1\cdot\underline{\sigma}_1) \otimes (\boldsymbol{n}_2\cdot\underline{\sigma}_2)(|+-\rangle - |-+\rangle) \\
&= \tfrac{1}{2}(\langle +-|(\boldsymbol{n}_1\cdot\underline{\sigma}_1) \otimes (\boldsymbol{n}_2\cdot\underline{\sigma}_2)|+-\rangle \\
&\quad - \langle +-|(\boldsymbol{n}_1\cdot\underline{\sigma}_1) \otimes (\boldsymbol{n}_2\cdot\underline{\sigma}_2)|-+\rangle \\
&\quad - \langle -+|(\boldsymbol{n}_1\cdot\underline{\sigma}_1) \otimes (\boldsymbol{n}_2\cdot\underline{\sigma}_2)|+-\rangle \\
&\quad + \langle -+|(\boldsymbol{n}_1\cdot\underline{\sigma}_1) \otimes (\boldsymbol{n}_2\cdot\underline{\sigma}_2)|-+\rangle) \\
&= \tfrac{1}{2}(\langle +|\boldsymbol{n}_1\cdot\underline{\sigma}_1|+\rangle \times \langle -|\boldsymbol{n}_2\cdot\underline{\sigma}_2|-\rangle \\
&\quad - \langle +|\boldsymbol{n}_1\cdot\underline{\sigma}_1|-\rangle \times \langle -|\boldsymbol{n}_2\cdot\underline{\sigma}_2|+\rangle \\
&\quad - \langle -|\boldsymbol{n}_1\cdot\underline{\sigma}_1|+\rangle \times \langle +|\boldsymbol{n}_2\cdot\underline{\sigma}_2|-\rangle \\
&\quad + \langle -|\boldsymbol{n}_1\cdot\underline{\sigma}_1|-\rangle \times \langle +|\boldsymbol{n}_2\cdot\underline{\sigma}_2|+\rangle)
\end{aligned}$$

This is built up out of expectation values calculated above, $\langle +|\boldsymbol{n}\cdot\underline{\sigma}|+\rangle = n_z$ and $\langle -|\boldsymbol{n}\cdot\underline{\sigma}|-\rangle = -n_z$; together with other things which are easily calculated, viz. $\langle +|\boldsymbol{n}\cdot\underline{\sigma}|-\rangle = n_x - \mathrm{i}n_y$ and $\langle -|\boldsymbol{n}\cdot\underline{\sigma}|+\rangle = n_x + \mathrm{i}n_y$. We get:

$$\begin{aligned}
&\langle sing|(\boldsymbol{n}_1\cdot\underline{\sigma}_1) \otimes (\boldsymbol{n}_2\cdot\underline{\sigma}_2)|sing\rangle \\
&= -(n_{1,x}n_{2,x} + n_{1,y}n_{2,y} + n_{1,z}n_{2,z}) \\
&= -\boldsymbol{n}_1\cdot\boldsymbol{n}_2 \\
&= -\cos(\angle(\boldsymbol{n}_1, \boldsymbol{n}_2))
\end{aligned}$$

This is a nice simple result. It corresponds to the value observed when the spin polarisations of both electrons are measured simultaneously at various angles, and thus represents the average correlation observed between the two spins (in the singlet state, and at those angles). It is along the lines of what we expect, because in each summand of $|sing\rangle$, the spins are oppositely aligned, and so when $\boldsymbol{n}_1 = \boldsymbol{n}_2$, we get perfect anticorrelation.

Qubits

Given the prevalence of quantum systems with an internal state space $\boldsymbol{Q}$, with two basis vectors $|0\rangle$ and $|1\rangle$, and the convention in computation theory of coding things using the two symbols '0' and '1', being the two possible values of a single 'bit' of information, we can regard an abstract quantum system with state space $\boldsymbol{Q}$, as a ***qubit***, a quantum analogue of the classical 'bit', a carrier of quantum information.

A classical bit can only be in one of the two states, 0 or 1. A qubit can be in a continuum of states given by $\lambda|0\rangle + \mu|1\rangle$, where $|\lambda|^2 + |\mu|^2 = 1$. This might imply that one can embed an arbitrary amount of classical information in a single qubit (eg. by encoding it in the infinite string of decimal places of the decimal representation of λ or μ). However this is misleading. You cannot extract this level of detail using a quantum measurement since measurement causes a discontinuous jump in the state. In practice you can only hope to get one bit of information out of a qubit. What's the point of using qubits then?

Ans: You can manipulate qubits in a different way from classical bits. Thus a classical bit evolves in a discrete step, flipping between 0 and 1, or not, according to an instruction whose outcome may be conditional on the values of some neighbouring bits. A qubit evolves according to quantum theory;

either smoothly during unitary evolution, or discontinuously via a quantum measurement. The trick is to exploit these novel mechanisms to try and gain an edge computationally.

[Aside: There's an issue to do with qubits which we point out now and then immediately ignore for the rest of this course. If you wish to perform computations on collections of qubits that encode binary information, then you have to make the qubits individually identifiable. The equivalent problem doesn't exist classically, you can always tell things apart — you just look! In the quantum regime it's not so easy. If you use a collection of identical systems for the qubits (eg. a collection of electrons), then the quantum statistics of collections of indistinguishable physical systems comes into play, which forces all physical states of the whole system to be either completely symmetric or completely antisymmetric under permutations of the identical subsystems. (Actually we have been surreptitiously exposed to this already. The singlet state $|sing\rangle$ we saw above, and will revisit below, is antisymmetric under interchange of the two electrons.) This issue can potentially cause some problems :– processing binary information where you can't tell the bits apart isn't a lot of use really. The way this is overcome is either to use distinguishable systems for different bits, or if using identical systems, to couple them to the environment sufficiently strongly that the degeneracy due to the symmetry of interchange is lifted. Below we will just assume that the problem has been overcome in one way or another.]

The No-Cloning Theorem

In conventional computing there is no obstacle to making a copy of some information — all systems offer various kinds of *Copy* command and you just use them as required. In the quantum computing world, information is held in the quantum state, and the analogue of copying would be a unitary operator U that could achieve for an arbitrary $|\psi\rangle$:

$$U\,|\psi\rangle|s\rangle = |\psi\rangle|\psi\rangle$$

where $|s\rangle$ is some standard state. However this is impossible: $U|\psi,s\rangle = |\psi,\psi\rangle$ and $U|\phi,s\rangle = |\phi,\phi\rangle$; now taking inner products of LHSs and RHSs gives:

$$\langle\psi|\phi\rangle = \langle\psi,s|U^\dagger U|\phi,s\rangle = \langle\psi,\psi|\phi,\phi\rangle = \langle\psi|\phi\rangle^2$$

so that $\langle\psi|\phi\rangle$ must be either 0 or 1, which is untrue in general.

The no-cloning theorem colludes with quantum measurement to conceal information, so you can't:

1. do some quantum processing to get some unknown state $|\psi\rangle$,
2. replicate it many times, ...

and then make various measurements on distinct copies to pin down the precise details of what the state $|\psi\rangle$ was.

Basis Copying vs. Cloning

Although cloning of arbitrary states is impossible, copying quantum encoded classical bit strings is OK. A classical n-bit string eg. 100...111 will correspond to a basis vector eg. $|100...111\rangle$ in $\boldsymbol{Q}^{\otimes n}$.

Copying these basis vectors to eg. the $|0...0\rangle$ basis vector corresponds to a unitary operator U that acts as:

$$\begin{aligned}
&U|0...00\rangle|0...00\rangle = |0...00\rangle|0...00\rangle\\
&U|0...01\rangle|0...00\rangle = |0...01\rangle|0...01\rangle\\
&U|0...10\rangle|0...00\rangle = |0...10\rangle|0...10\rangle\\
&\quad ... \qquad ... \qquad ...\\
&U|1...11\rangle|0...00\rangle = |1...11\rangle|1...11\rangle
\end{aligned}$$

This partially specifies U as a partial permutation of the basis vectors of $\boldsymbol{Q}^{\otimes 2n}$. This can always be extended to a complete permutation of the basis vectors of $\boldsymbol{Q}^{\otimes 2n}$, and therefore to a valid unitary U by linearity. However the action of U will not agree with that of cloning for non-basis states.

The 'copying instead of cloning' trick is used a lot in quantum computing.

Entanglement

Consider a 2-qubit system on $\boldsymbol{Q}\otimes\boldsymbol{Q}$ which contains the state:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

This state cannot be written as a tensor product of two individual qubits such as $|\phi_1\rangle\otimes|\phi_2\rangle$ in any way whatsoever (i.e. it is not ***separable***). Suppose on the contrary that it was, then:

$$|\phi_1\rangle = a_1|0\rangle + b_1|1\rangle \text{ and } |\phi_2\rangle = a_2|0\rangle + b_2|1\rangle$$

and therefore

$$|\psi\rangle = a_1a_2|00\rangle + a_1b_2|01\rangle + b_1a_2|10\rangle + b_1b_2|11\rangle$$

So $a_1b_2 = b_1a_2 = 0$. It is now becomes impossible to have $a_1a_2 = b_1b_2 \neq 0$.

States that cannot thus be decomposed as tensor products are called ***entangled***.

Consider a similar construction in a larger system $\boldsymbol{Q}\otimes\boldsymbol{H}_{\mathbf{BIG}}$, where $\boldsymbol{Q}$ is a 1-qubit atomic system, and $\boldsymbol{H}_{\mathbf{BIG}}$ is a macroscopic system, eg. the state space of Schrödinger's cat. There is now a state:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle|alive\rangle + |1\rangle|dead\rangle)$$

in which the cat is in a superposition of alive and dead states (such a state can arise by coupling the cat to an unstable atomic particle; if the particle decays, it sets off a device that kills the cat).

Now we make a measurement on the atomic system, forcing it into state $|0\rangle$ or $|1\rangle$. The whole entangled state now jumps discontinuously into either $|0\rangle|\mathit{alive}\rangle$ or $|1\rangle|\mathit{dead}\rangle$. The cat moves from a half alive and half dead superposition to one or the other. This appears bizzare to say the least.

Insofar as macroscopic systems can be prepared in superposed states at all, experiments confirm the prediction of quantum mechanics.

Note that the subsystems in an entangled state *need not be spatially coincident*. When they are widely separated, measurement can give rise to ***action at a distance***. Einstein, Podolsky and Rosen were the first to point out this troublesome possibility, and so entangled states are also called ***EPR states***. Entangled states with spatially separated components illustrate dramatically the difference between classical and quantum phenomena.

In Quantum Computing, entanglement is seen not as a problem but as a ***resource***.

Bell and CHSH Inequalities

Consider ***local realistic theories*** of physics. In these, physical phenomena separated by large distances have no effect on one another; they have *independent existence*. Facts about the world are true or not independent of whether we choose to observe them at any particular point. This is a basic tenet of everyday existence.

Suppose we have four physical systems *A*, *B*, *C*, *D* that are (a) widely separated, (b) can only produce values ± 1 when measured. Let the values obtained when *A*, *B*, *C*, *D* are measured be A, B, C, D. Consider:

$$M = (A+C)B + (C-A)D$$

Now **either** A and C both have the same sign and so $C-A = 0$ and $M = (A+C)B = \pm 2$, **or** A and C have opposite signs so $A+C = 0$ and $M = (C-A)D = \pm 2$. Either way:

$$M = (A+C)B + (C-A)D = \pm 2$$

so that if we take the average over many identically prepared systems we get:

$$|\langle M\rangle| = |\langle AB\rangle + \langle BC\rangle + \langle CD\rangle - \langle DA\rangle| \le 2$$

This is the Clauser-Horne-Shimony-Holt inequality (Bell effectively considered the case *B* = *C*).

Now consider an atomic species that decays and produces a pair of electrons the spin part of whose state is the singlet state:

$$|\mathit{sing}\rangle = \frac{1}{\sqrt{2}}(|{+}{-}\rangle - |{-}{+}\rangle)$$

where the + and – refer (as usual) to the z axis, and which fly apart in opposite directions:

$$A, C \quad \longleftarrow \bullet \longrightarrow \quad B, D$$

Let *A* and *C* be *two different* measurements of the spin polarisation (transverse to the direction of flight) on the leftgoing electron, and *B* and *D* be *two different* measurements on the rightgoing electron. In such a scheme the four sample averages $\langle AB\rangle$, $\langle BC\rangle$, $\langle CD\rangle$, $\langle DA\rangle$ can be experimentally obtained. Note that four sample measurements cannot be obtained *simultaneously* in any individual experiment, but local reality says that the measurements *not performed* in any particular case do not conflict with the value that would have been obtained if it had been (i.e. we depend here on *counterfactual reasoning*).

Now quantum mechanics (via our previous calculations) says that measurements of the correlation of the polarisations of the two electrons in the singlet state in directions A and B yield:

$$\langle AB\rangle = -\cos(\angle(A,B))$$

Inserting this in the CHSH inequality, and choosing the directions to be in the plane perpendicular to the direction of flight of the electrons, gives:

$$|\cos(\angle(A,B)) + \cos(\angle(B,C)) + \cos(\angle(C,D)) - \cos(\angle(D,A))| \le 2$$

Choosing $\angle(A,B) = \angle(B,C) = \angle(C,D) = \angle(A,D)/3 = \pi/4$, the positive $\cos$s each give $1/\sqrt{2}$ and the negative $\cos$ gives $-1/\sqrt{2}$. So we derive:

$$2\sqrt{2} \le 2$$

which is absurd.

Quantum mechanics denies local reality.

Experiments by the Aspect group (on polarised photons[4]) have confirmed the quantum mechanical predictions in situations where the *A* vs *C* and *B* vs *D* choices are made at spacelike separations, confirming that quantum mechanics can act nonlocally.

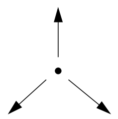
4. The only difference between photons and electrons in these circumstances is that the correlation $\langle AB\rangle = -\cos(\angle(A,B))$ for electrons becomes $\langle AB\rangle = \cos(2\angle(A,B))$ for photons. Armed with this fact, the analogous calculations are easy, though we will not go into details here. Of course, nothing in these notes explains ***why*** the only difference is the one stated.

The GHZ Argument

Greenberger Horne Zeilinger and later Mermin, considered the decay of some atomic species that produces 3 spin-$\frac{1}{2}$ particles[5] for which the spin part of the state is:

$$|GHZ\rangle = \frac{1}{\sqrt{2}}(|{+}{+}{+}\rangle - |{-}{-}{-}\rangle)$$

where the + and – refer (for definiteness) to the z axis, and which fly apart at 120˚ to each other.



Now consider the three operators:

$$O_1 = \sigma_{1x} \otimes \sigma_{2y} \otimes \sigma_{3y} \ ;$$
$$O_2 = \sigma_{1y} \otimes \sigma_{2x} \otimes \sigma_{3y} \ ;$$
$$O_3 = \sigma_{1y} \otimes \sigma_{2y} \otimes \sigma_{3x}$$

These satisfy for $i = 1 \ldots 3$:

$$O_i |GHZ\rangle = |GHZ\rangle$$

Proof: We can check that:

$$\sigma_x|+\rangle = |-\rangle \ ; \ \sigma_y|+\rangle = \mathrm{i}|-\rangle \ ; \ \sigma_z|+\rangle = |+\rangle \ ;$$
$$\sigma_x|-\rangle = |+\rangle \ ; \ \sigma_y|-\rangle = -\mathrm{i}|+\rangle \ ; \ \sigma_z|-\rangle = -|-\rangle$$

5. Little white lie here ...

so that:

$$O_1|GHZ\rangle = \tfrac{1}{\sqrt{2}}\sigma_{1x}\sigma_{2y}\sigma_{3y}(|+\rangle|+\rangle|+\rangle - |-\rangle|-\rangle|-\rangle)$$
$$= \tfrac{1}{\sqrt{2}}(\sigma_{1x}|+\rangle\sigma_{2y}|+\rangle\sigma_{3y}|+\rangle - \sigma_{1x}|-\rangle\sigma_{2y}|-\rangle\sigma_{3y}|-\rangle)$$
$$= \tfrac{1}{\sqrt{2}}(|-\rangle\mathrm{i}|-\rangle\mathrm{i}|-\rangle - |+\rangle{-\mathrm{i}}|+\rangle{-\mathrm{i}}|+\rangle)$$
$$= \tfrac{1}{\sqrt{2}}(|+\rangle|+\rangle|+\rangle - |-\rangle|-\rangle|-\rangle)$$

Similarly, defining:

$$X = \sigma_{1x} \otimes \sigma_{2x} \otimes \sigma_{3x}$$

we have:

$$X|GHZ\rangle = -|GHZ\rangle$$

and remembering that $\sigma_x\sigma_y = \mathrm{i}\sigma_z$ (and cyclically), and that spin operators belonging to different spins commute, we easily get:

$$O_1 O_2 O_3 = -X$$

Now, given that the three components $1 \ldots 3$ are spatially far apart, we can measure say the σ_{1x} and σ_{2y} values of $|GHZ\rangle$, and knowing that $|GHZ\rangle$ is an eigenstate of O_1, determines the value of a measurement of σ_{3y} on the resulting state. This works for all the observables O_1, O_2, O_3, X, and all possible choices of two spin variables to measure first; and can lead us to suspect that all the values obtained for these various spin measurements, say m_{1x}, m_{1y}, m_{2x}, m_{2y}, m_{3x}, m_{3y}, have an *independent existence*.

Now since $O_i|GHZ\rangle = |GHZ\rangle$:

$$m_{1x}.m_{2y}.m_{3y} = 1$$
$$m_{1y}.m_{2x}.m_{3y} = 1$$
$$m_{1y}.m_{2y}.m_{3x} = 1$$

and since $X|GHZ\rangle = -|GHZ\rangle$:

$$m_{1x}.m_{2x}.m_{3x} = -1$$

whereupon multiplying all four LHSs we get:

$$m_{1x}{}^2.m_{1y}{}^2.m_{2x}{}^2.m_{2y}{}^2.m_{3x}{}^2.m_{3y}{}^2 = 1$$

whereas multiplying all four RHSs we get:

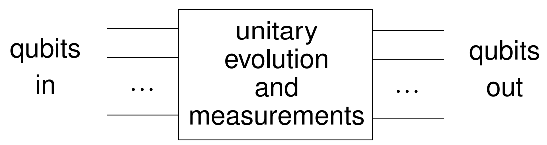
$$1.1.1.-1 = -1$$

So local reality leads to $1 = -1$. Note that this is a sharper contradiction than CHSH since it applies to an *individual* *experiment*, not just statistically.

As for the CHSH inequalities, GHZ type situations have been observed experimentally (in particular by Zeilinger himself, though as ever, he was using photons).

Quantum Computing Basics

Now we start processing qubits quantum mechanically, to see what computations can be done. The model of computation is the quantum circuit model:



Basic facts:

- No. of qubits in = No. of qubits out.[6]
- Classical binary input information is coded in the state of the input qubits by:

  $011\ldots0 \leftrightarrow |011\ldots0\rangle$

- Classical output information must be extracted by measurement.

Of particular note is this:

> Since measurement destroys the state, if the computation is to be useful, the answer sought must be present in the output qubits state vector with (at least) high probability amplitude.

The series of qubits is called a quantum register.

6. Neither unitary evolution nor measurement can alter the shape of the state space.

Quantum Gate Basics

As in classical hardware design, quantum circuits are built out of a relatively small selection of basic gates.[7] There are two basic kinds of gate:

- Unitary gates: these are ***reversible***, because all unitary operators U have inverses $U^\dagger$.[8]
- Measurement gates: these are ***irreversible***, (because of projection) and have to be specified statically; the measurement cannot depend dynamically on the state to be measured.

Here are the most important gates we work with:

One Qubit Gates

Unitary gates:

I the IDENTITY, not normally mentioned, but a valid unitary transformation just the same.

X the NOT transformation (swaps basis vectors $|0\rangle$ and $|1\rangle$ in $\boldsymbol{Q}$); the transformation formerly known as Pauli spin matrix σ_x.

Y the transformation formerly known as Pauli spin matrix σ_y.

Z the transformation formerly known as Pauli spin matrix σ_z.

7. It's no coincidence that qubits are drawn as 'wires' in quantum circuit designs.
8. There are classical models of reversible computation, but we will ignore them.

H the Hadamard transformation.

Φ_ϕ the transformation that maps $|0\rangle$ to $|0\rangle$, and $|1\rangle$ to $e^{i\phi}|1\rangle$ in $\boldsymbol{Q}$.

Here is a reminder of their matrices:

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad \Phi = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{bmatrix} \qquad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Note that $I = \Phi_0$ and $Z = \Phi_\pi$.

Now consider the circuit:

$$—\boxed{H}—\boxed{\Phi_\theta}—\boxed{H}—\boxed{\Phi_{\phi+\pi/2}}—$$

It maps $|0\rangle$ to $\frac{1}{\sqrt{2}}[1,1]^T$, then to $\frac{1}{\sqrt{2}}[1,e^{i\theta}]^T$, then to $\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}[1+e^{i\theta}, 1-e^{i\theta}]^T = e^{i\theta/2}[\cos(\theta/2), -i\sin(\theta/2)]^T$, and finally to $e^{i\theta/2}[\cos(\theta/2), e^{i\phi}\sin(\theta/2)]^T$.

Up to the irrelevant $e^{i\theta/2}$, this is the most general qubit state $|\theta,\phi\rangle = [\cos(\theta/2), e^{i\phi}\sin(\theta/2)]^T$. Since $\boldsymbol{Q}$ is two dimensional, $|\theta,\phi\rangle$ has an orthogonal complement, $|\theta,\phi\rangle^\perp$, unique up to an arbitrary phase, thus: $[\sin(\theta/2), -e^{i\phi}\cos(\theta/2)]^T$. And so the transformation $\Phi_{\phi+\pi/2}H\Phi_\theta H$ (note the order of factors) rotates the basis $\{|0\rangle, |1\rangle\}$ to the basis $\{|\theta,\phi\rangle, |\theta,\phi\rangle^\perp\}$ in effect. This is the most general basis rotation since $|\theta,\phi\rangle$

is arbitrary and $|\theta,\phi\rangle^\perp$ is determined up to phase by $|\theta,\phi\rangle$. So the *H* and Φ gates are sufficient to implement the most general unitary 1-qubit gate (all up to phases). For the record:

$$\Phi_{\phi+\pi/2}H\Phi_\theta H = e^{i\theta/2}\begin{bmatrix} \cos(\theta/2) & -i\sin(\theta/2) \\ e^{i\phi}\sin(\theta/2) & i\,e^{i\phi}\cos(\theta/2) \end{bmatrix}$$

Measurement gates:

M_B projection onto basis *B*; the state vector jumps to one or other element of *B* with the usual probabilities. Normally *B* is $\{|0\rangle, |1\rangle\}$.

There is nothing special about the standard basis *B* of $\boldsymbol{Q}$, though it is the most commonly used one. Equally good is the ***Hadamard basis*** $\{|0_H\rangle, |1_H\rangle\}$ which we have seen:

$$|0_H\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

$$|1_H\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

One qubit is unfortunately rather limited — most interesting computational problems require more than one bit to code their instances! So we need two or more qubit gates so that the qubits can be coupled together.

Two Qubit Gates

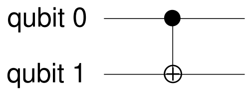
Unitary gates; besides the identity gate there are:

CNOT the CONTROLLED NOT gate; the transformation that maps

$$|00\rangle \to |00\rangle,\ |01\rangle \to |01\rangle, \qquad \text{(identity)}$$
$$|10\rangle \to |11\rangle,\ |11\rangle \to |10\rangle. \qquad \text{(swap)}$$

If *a* and *b* are two classical bits, corresponding to qubit basis vectors $|a\rangle$ and $|b\rangle$, then *CNOT* maps $|a,b\rangle \to |a, a\oplus b\rangle$ where $\oplus$ is EXCLUSIVE OR on *a* and *b*. To see its matrix representation we have to choose a specific order to list the basis vectors of $\boldsymbol{Q}\otimes\boldsymbol{Q}$. Predictably we choose $|00\rangle, |01\rangle, |10\rangle, |11\rangle$, i.e. the lexicographical order on the corresponding bit strings.[9] As a matrix and as a circuit:

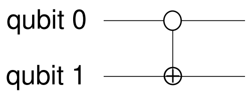
$$CNOT = \begin{bmatrix} \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} & \mathbf{0} \\ \mathbf{0} & \begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix} \end{bmatrix}$$



In the circuit, the $\oplus$ represents the EXCLUSIVE OR whose result is in qubit 1, while the ● represents the controlling effect of qubit 0, with a full blob meaning 'switch on the XOR when this bit is 1' (and there is an analogous empty blob ○ possibil-

9. Thus far we have largely been able to avoid such choices of ordering of basis elements by splitting calculations involving tensor products into smaller ones on the factors, because the operators we have dealt with (eg. $(\boldsymbol{n}_1{\cdot}\boldsymbol{\sigma}_1)\otimes(\boldsymbol{n}_2{\cdot}\boldsymbol{\sigma}_2)$) have been separable. Now that ceases to be the case and we must list all the component explicitly.

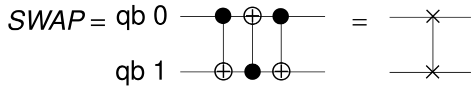
ity meaning ‘switch on the XOR when this bit is 0’ and drawn as follows:)



CNOT is a very important gate. Here’s another gate that you can implement with the *CNOT* gate.

SWAP the SWAPPING gate;
the transformation that maps

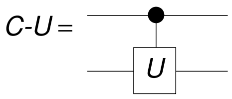
$$|00\rangle \to |00\rangle,\ |01\rangle \to |10\rangle,$$
$$|10\rangle \to |01\rangle,\ |11\rangle \to |11\rangle.$$



[Aside: Why on earth do you need an actual *circuit* to swap two qubits around? Why not just swap the wires over? Well, the lines in our circuits are time lines rather than physical ‘wires’. In reality, insofar as qubits can be implemented at all, individual qubits are either implemented as distinguishable systems and dispersed ‘everywhere’, or implemented as indistinguishable systems and sufficiently nailed down within the environment to enable them to be told apart; i.e. they can’t move. So you need a circuit. Maybe one day someone will invent a technology in which qubits are restrained by purposely engineered conduits, along which they travel to be brought together to interact as required. But for now that’s just a pipe dream.]

C-U the CONTROLLED-*U* gate, where *U* is a 1-qubit unitary gate; the transformation that maps

$$|00\rangle \to |00\rangle,\ |01\rangle \to |01\rangle, \qquad \text{(identity)}$$
$$|1\rangle|\psi\rangle \to |1\rangle \otimes U|\psi\rangle. \qquad \text{(apply } U\text{)}$$



C-U has various useful special cases: *CNOT* is just *C-X*, and another special case of some importance is $C\text{-}\Phi_\phi$, the controlled phase gate. The ‘controlling’ idea is a significant one, as we will see.

Measurement gates:

M_B projection onto basis *B* of subspace *S*.

Taking the standard basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ of $\boldsymbol{Q}\otimes\boldsymbol{Q}$, the subspace *S* may be the whole of $\boldsymbol{Q}\otimes\boldsymbol{Q}$; so if the state is $a_{00}|00\rangle+a_{01}|01\rangle+a_{10}|10\rangle+a_{11}|11\rangle$, then any basis vector $|00\rangle, |01\rangle, |10\rangle, |11\rangle$, is a possible outcome, with probabilities $|a_{00}|^2, |a_{01}|^2, |a_{10}|^2, |a_{11}|^2$, respectively.

Alternatively *S* is the subspace corresponding to one of the qubits, say qubit 0, in which case the relevant projections are:

$\boldsymbol{P}_0 = |00\rangle\langle 00|+|01\rangle\langle 01|$ for obtaining answer 0, with probability $|a_{00}|^2+|a_{01}|^2$, and if this is non-zero, yielding the state:

$$\frac{a_{00}|00\rangle+a_{01}|01\rangle}{\sqrt{|a_{00}|^2+|a_{01}|^2}}$$

$\boldsymbol{P}_1 = |10\rangle\langle 10|+|11\rangle\langle 11|$ for obtaining answer 1, with probability $|a_{10}|^2+|a_{11}|^2$, and if this is non-zero, yielding the state:

$$\frac{a_{10}|10\rangle+a_{11}|11\rangle}{\sqrt{|a_{10}|^2+|a_{11}|^2}}$$

There is nothing special about the standard basis of $\boldsymbol{Q}\otimes\boldsymbol{Q}$, though it is the most commonly used one. Equally good is the ***Bell basis*** which consists exclusively of entangled states:

$$|00_B\rangle = \tfrac{1}{\sqrt{2}}(|00\rangle+|11\rangle)$$
$$|01_B\rangle = \tfrac{1}{\sqrt{2}}(|00\rangle-|11\rangle)$$
$$|10_B\rangle = \tfrac{1}{\sqrt{2}}(|01\rangle+|10\rangle)$$
$$|11_B\rangle = \tfrac{1}{\sqrt{2}}(|01\rangle-|10\rangle)$$

This is just as good as any other basis of $\boldsymbol{Q}\otimes\boldsymbol{Q}$. In particular, measurements can be done using the Bell basis or any other basis. The mechanics of these is just the same ... eg. ‘just add subscript ‘$_B$’.

N.B. To go from the standard basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ to the Bell basis $\{|00_B\rangle, |01_B\rangle, |10_B\rangle, |11_B\rangle\}$ (both listed in the order given), you need the unitary matrix:

$$U_{SB} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

Superdense Coding

With 1-qubit gates and entanglement we can send two classical bits using one qubit as a channel.

Alice wants to send two bits of classical information to Bob, but can only pay the shipping costs of one qubit on the Space Shuttle. Luckily, Alice and Bob share an entangled state $\frac{1}{\sqrt{2}}(|00\rangle+|11\rangle) = |00_B\rangle$ that they prepared earlier; Alice has qubit 0 of this and Bob has qubit 1. Fortunately they are also both pretty good at quantum theory.

Alice does one of four things to qubit 0 of the entangled pair, according to the value of the two classical bits she wants to send:

To send 00, Alice applies $\mathbf{I}$ to qubit 0; the total state remains:

$$\frac{1}{\sqrt{2}}(|00\rangle+|11\rangle) = |00_B\rangle$$

To send 01, Alice applies Z to qubit 0; the total state becomes:

$$Z\otimes\mathbf{I}\frac{1}{\sqrt{2}}(|00\rangle+|11\rangle) = \frac{1}{\sqrt{2}}(|00\rangle-|11\rangle) = |01_B\rangle$$

To send 10, Alice applies X to qubit 0; the total state becomes:

$$X\otimes\mathbf{I}\frac{1}{\sqrt{2}}(|00\rangle+|11\rangle) = \frac{1}{\sqrt{2}}(|10\rangle+|01\rangle) = |10_B\rangle$$

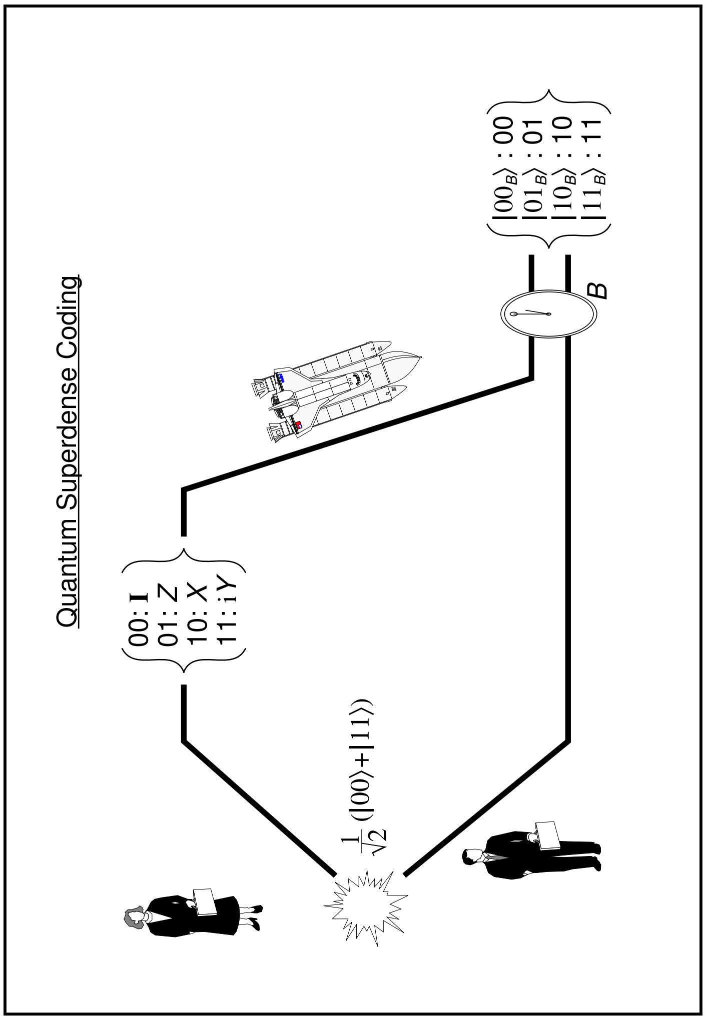
To send 11, Alice applies $\mathrm{i}Y$ to qubit 0; the total state becomes:

$$\mathrm{i}Y\otimes\mathbf{I}\frac{1}{\sqrt{2}}(|00\rangle+|11\rangle) = \frac{1}{\sqrt{2}}(-|10\rangle+|01\rangle) = |11_B\rangle$$

Note that these outcomes are orthogonal.

Alice now puts qubit 0 on the Shuttle to Bob. When Bob receives it, he has both qubits of the pair and can perform a maximal measurement (on $\boldsymbol{Q}\otimes\boldsymbol{Q}$) corresponding to the Bell basis partition $\{\{|00_B\rangle\}, \{|01_B\rangle\}, \{|10_B\rangle\}, \{|11_B\rangle\}\}$. Because he is measuring one of four basis states in the basis which they constitute, he is certain to discover which basis state it is. Hence he is certain to discover the two classical bits Alice had in mind.

Using photons, quantum superdense coding has been observed experimentally (by Zeilinger).



Quantum Teleportation

With 2-qubit gates and entanglement we can do something fairly startling, namely teleportation.

Alice wants to send a qubit $|\psi\rangle = a|0\rangle+b|1\rangle$ to Bob who is far away. She doesn't know what the state $|\psi\rangle$ is (i.e. what a and b are), and can't copy it to experiment on it to find out, without destroying it (by no-cloning). Luckily, Alice and Bob share an entangled state $\frac{1}{\sqrt{2}}(|00\rangle+|11\rangle)$ that they prepared earlier; Alice has qubit 0 of this and Bob has qubit 1. The total state is the tensor product of $|\psi\rangle$ and the entangled state (remember Alice has the first two qubits of this and Bob the last qubit):

$$\frac{1}{\sqrt{2}}[a|0\rangle(|00\rangle+|11\rangle) + b|1\rangle(|00\rangle+|11\rangle)]$$

Alice applies *CNOT* to her two qubits, amounting to applying *CNOT*$\otimes\mathbf{I}$ to the whole state:

$$\frac{1}{\sqrt{2}}[a|0\rangle(|00\rangle+|11\rangle) + b|1\rangle(|10\rangle+|01\rangle)]$$

Alice next applies H to her first qubit, i.e. $H\otimes\mathbf{I}\otimes\mathbf{I}$ to the whole state:

$$\frac{1}{2}[a(|0\rangle+|1\rangle)(|00\rangle+|11\rangle) + b(|0\rangle-|1\rangle)(|10\rangle+|01\rangle)]$$

Now comes the tensor product sleight of hand. The above is the same as:

$$\frac{1}{2}[\,|00\rangle(a|0\rangle+b|1\rangle) + |01\rangle(a|1\rangle+b|0\rangle) + |10\rangle(a|0\rangle-b|1\rangle) + |11\rangle(a|1\rangle-b|0\rangle)\,]$$

Now Alice measures both her qubits, projecting onto the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. The linear superposition of Bob's qubit states collapses to one of four possibilities:

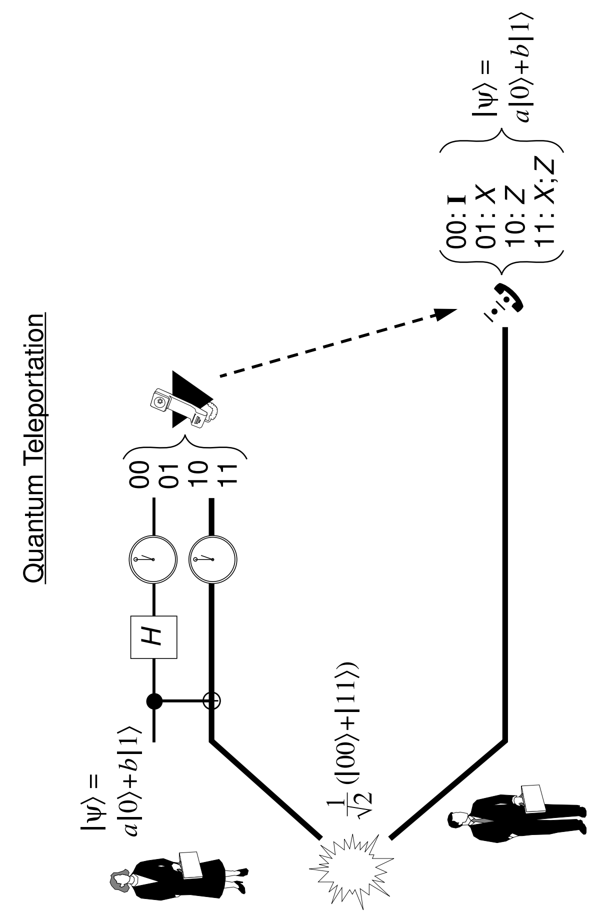
Alice got $|00\rangle$, then Bob's qubit is $(a|0\rangle+b|1\rangle)$
Alice got $|01\rangle$, then Bob's qubit is $(a|1\rangle+b|0\rangle)$
Alice got $|10\rangle$, then Bob's qubit is $(a|0\rangle-b|1\rangle)$
Alice got $|11\rangle$, then Bob's qubit is $(a|1\rangle-b|0\rangle)$

Bob doesn't yet know which of these happened, so Alice has to tell him using classical communication (no signalling faster than light then). However once he knows the classical value Alice obtained, i.e. 00, 01, 10, 11, he knows what to do to recover an exact replica of the original $|\psi\rangle$:

Alice got 00, Bob applies $\mathbf{I}$ to recover $|\psi\rangle$
Alice got 01, Bob applies X to recover $|\psi\rangle$
Alice got 10, Bob applies Z to recover $|\psi\rangle$
Alice got 11, Bob applies ZX to recover $|\psi\rangle$

Note that Alice's original qubit got projected away, destroying its value, so teleportation is not a way around no-cloning. [Bob must have needed the qubit more than she did.]

Using (inevitably) photons, quantum teleportation has been observed experimentally (by Zeilinger).

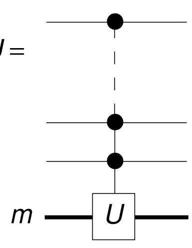


Gates with more than Two Qubits

The more qubits in a gate, the harder it is to imagine how to implement it. However some n-qubit gates are conceptually useful.

C^n-U the (multi) CONTROLLED-*U* gate, where *U* is an m-qubit unitary gate; writing the state in the form $|\xi\psi\rangle$ where $|\xi\rangle$ is the n-qubit control part and $|\psi\rangle$ is the m-qubit target, *C^n-U* acts by:

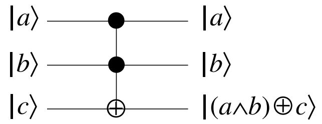
$|\xi\rangle$ a basis vector of $\boldsymbol{Q}^n$ other than $|1^n\rangle$:
$|\xi\psi\rangle \rightarrow |\xi\psi\rangle$, (identity)
$|1^n\rangle|\psi\rangle \rightarrow |1^n\rangle \otimes U|\psi\rangle$. (apply *U*)

C^n-U =


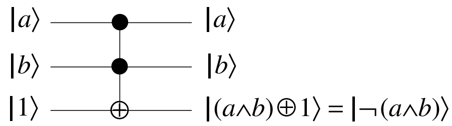
As for *C-U*, *C^n-U* has variants where the control is concentrated on some other basis vector of $\boldsymbol{Q}^n$ than $|1^n\rangle$; diagramatically the 0 qubits are represented by open circles as previously.

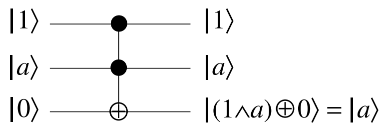
As for *C-U*, *C^n-U* has various useful special cases. The most useful of these is the Toffoli gate, C^2-$\oplus$.

Toffoli or C^2-$\oplus$; in pictures:

Toffoli =


Toffoli is useful because, *acting on basis vectors*, it can simulate classical ***nand*** and ***fanout*** as follows. Note the basis copying (not cloning) in fanout.

nand =


fanout =


In turn *nand* and *fanout* are sufficient to implement any classical circuit computation, a well known fact from digital logic design. As a result we conclude:[10]

In principle, quantum computation is able to subsume classical computation.

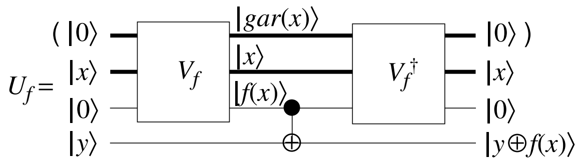
10.To properly quantify this statement, we have to clarify how the size of the circuit varies with the size of the inputs etc etc. This is beyond the scope of this course.

In detail, suppose we want to quantum compute a function f with several binary inputs and one binary output whose classical circuit implementation we know. We construct a quantum circuit that corresponds to our classical circuit by replacing each *nand* and each binary *fanout* with a suitably tuned Toffoli. This produces a quantum circuit with more qubits than just the input and output qubits; we have ancilla qubits, and in general the computation will produce some garbage qubits too. The circuit so far implements the transformation:

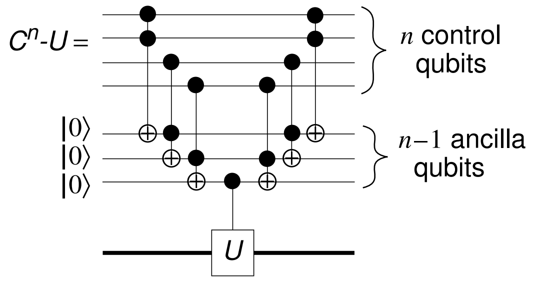
$$V_f|0,x,0\rangle = |gar(x),x,f(x)\rangle$$

where the garbage qubits are written first. To get this into a standard form, we want to eliminate the garbage by getting it back to $|0\rangle$. We add an extra qubit $|y\rangle$ and follow the application of V_f with a *CNOT*, controlling the flip of $|y\rangle$ with the $|f(x)\rangle$ qubit. Finally we apply $V_f^\dagger$ to undo both the computation of $|f(x)\rangle$ and the garbage, getting the standard quantum implementation of f, U_f:

$$U_f|0,x,0,y\rangle = (V_f^\dagger \otimes \mathbf{I})(\mathbf{I} \otimes \mathbf{I} \otimes \mathit{CNOT})(V_f \otimes \mathbf{I})$$
$$|0,x,0,y\rangle \rightarrow |0,x,0,y \oplus f(x)\rangle$$

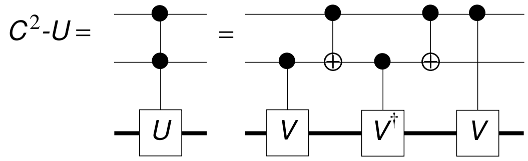


With Toffoli and a C-U we can emulate a C^n-U:



Note the use of ancilla (workspace) qubits again, each initially in state $|0\rangle$, for storing intermediate results of the $(((q0 \wedge q1) \wedge q2) \wedge q3)$ calculation.

The Toffoli gate itself may be built from 1- and 2-qubit gates. The solution is not unique. A simple solution follows as a special case of a special purpose C^2-U design, which depends on the availability of a unitary operator V such that $V^2 = U$.



Setting $V = (1-\mathrm{i})(\mathbf{I} + \mathrm{i}X)/2$ gives Toffoli. In general quite a bit is now known about building larger gates out of smaller ones.