CHEM 440/740:
Introduction to Quantum Information and Quantum Control in Physical Systems

Jonathan Baugh

Winter 2010

Contents

References

1. *Quantum Computation and Quantum Information*, M. A. Nielsen and I. L. Chuang. Cambridge Univ. Press (2000).

2. *An Introduction to Quantum Computing*, P. Kaye, R. Laflamme, and M. Mosca. Oxford Univ. Press (2007).

3. *Spin dynamics: Basics of nuclear magnetic resonance*, M. H. Levitt. Wiley (2001).

4. *NMR Based Quantum Information Processing: Achievements and Prospects*, D. G. Cory *et al.* arXiv: quant-ph/0004104 (2000).

5. *Principles of the quantum control of molecular processes*, P. Brumer and M. Shapiro. Wiley (2003).

6. *Simulated Quantum Computation of Molecular Energies*, A. Aspuru-Guzik *et al.* Science **309**,1704 (2005).

Nomenclature and Notation

Identity matrix:

$$\mathbb{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Dirac notation:

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

Pauli matrices:

$$\sigma_x = \mathbf{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \rho_y = \mathbf{Y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_z = \mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\mathbf{X} \otimes \mathbf{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Hadamard matrix:

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

$\hbar = \frac{h}{2\pi}$	Reduced Planck's constant
$\langle \mathbf{Z} \rangle$	Expectation value of $\mathbf{Z}$
$\mathcal{H}$	Hamiltonian
ρ	Density matrix
$\mathbf{M}_m$	Generalized measurement operator

Chapter 1

Introduction to Quantum Information and Quantum Information Processing

Main reference: *Quantum Computation and Quantum Information*, M. A. Nielsen and I. L. Chuang. Cambridge Univ. Press (2000).

1.1 Postulates of Quantum Mechanics

1.1.1 Postulate no. 1:

Associated to any isolated physical system is a complex inner-product vector space (*i.e.* a Hilbert space) called the *state space* of the system.

The closed system is completely described by its *state vector*, a unit vector in it's state space. The quantum bit, or qubit, is an example of such a system.

Example: The qubit

The qubit is the simplest example of a state space in quantum mechanics because it has only two dimensions.
Let $|0\rangle$ and $|1\rangle$ form an orthnormal basis. An arbitrary state vector is represented as

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$$

where α and β are complex numbers satisfying $|\alpha|^2 + |\beta|^2 = 1$ (*i.e.* $\langle\psi|\,\psi\rangle = 1$).

The coefficients α and β are often referred to as *amplitudes*. The probability of finding the system in the associated basis state upon measurement is given by the square of the amplitudes (*e.g.* $|\alpha|^2$, $|\beta|^2$).

1.1.2 Postulate no. 2:

The evolution of a *closed* quantum system is described by a *unitary transformation.*

The state $|\psi\rangle$ at time t_1 is related to $|\psi'\rangle$ at time t_2 by a unitary operator $\mathbf{U}$ that depends *only* on the times t_1 and t_2.

A unitary operator, $\mathbf{U}$, satisfies the condition $\mathbf{U}^\dagger\mathbf{U} = \mathbb{I}$.

Example: Single-qubit unitary operators

X operator: "bit flip" or "NOT" operator

$$\mathbf{X}(\alpha|0\rangle + \beta|1\rangle) = \alpha|1\rangle + \beta|0\rangle$$

Z operator: "phase flip" operator

$$\mathbf{Z}(\alpha|0\rangle + \beta|1\rangle) = \alpha|0\rangle - \beta|1\rangle$$

Hadamard operator

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

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

$$\left.\begin{array}{l}\mathbf{H}^2|0\rangle = |0\rangle\\ \mathbf{H}^2|1\rangle = |1\rangle\end{array}\right\} \Rightarrow \mathbf{H}^2 = \mathbb{I}$$

Postulate no. 2′:

The continuous-time evolution of a closed quantum system is described by the Schrödinger equation:

$$i\hbar\frac{\mathrm{d}|\psi\rangle}{\mathrm{d}t} = \mathcal{H}|\psi\rangle$$

The Hamiltonian, $\mathcal{H}$, is a *Hermitian* operator, which means that $\mathcal{H}^\dagger = \mathcal{H}$. As a result, it has a *spectral decomposition*

$$\mathcal{H} = \sum_E E|E\rangle\langle E|$$

where E are the energy eigenvalues and $|E\rangle$ are the eigenvectors of the system, also referred to as energy eigenstates.

Example: $\mathcal{H} = \hbar\omega\mathbf{X}$

Energy eigenstates of $\mathcal{H}$ are the eigenvectors of $\mathbf{X}$.

$$\mathbf{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The eigenvectors of $\mathbf{X}$ are

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

with corresponding eigenvalues of $\lambda = \pm 1$. As a result, the energy eigenvalues of $\mathcal{H}$ are $E_0 = -\hbar\omega$, which is the ground state, and $E_1 = +\hbar\omega$.

The solution to the Schrödinger equation for a time-independent Hamiltonian is described by a unitary transformation:

$$\begin{aligned} |\psi(t_2)\rangle &= \overbrace{e^{-i\frac{\mathcal{H}}{\hbar}(t_2-t_1)}}^{\text{unitary operator}} |\psi(t_1)\rangle \\ &= \mathbf{U}(t_2, t_1)\,|\psi(t_1)\rangle \end{aligned}$$

1.1.3 Postulate no. 3:

Quantum measurement is described by a set $\{\mathbf{M}_m\}$ of measurement operators.

Quantum measurements act on the state space of the system being measured. The index m refers to the possible measurement outcomes. By applying the measurement operators $\mathbf{M}_m$ to the state $|\psi\rangle$, the probability for outcome m is

$$p(m) = \langle\psi|\,\mathbf{M}_m^\dagger\mathbf{M}_m\,|\psi\rangle$$

and the state after measurement is

$$\frac{\mathbf{M}_m}{\sqrt{\langle\psi|\,\mathbf{M}_m^\dagger\mathbf{M}_m\,|\psi\rangle}}\,|\psi\rangle\,.$$

The operator $\mathbf{M}_m$ satisfies the completeness relation

$$\sum_m \mathbf{M}_m^\dagger\mathbf{M}_m = \mathbb{I}$$

which means that the probabilities sum to unity.

$$\begin{aligned}\sum_m p(m) &= \sum_m \langle\psi|\, \mathbf{M}_m^\dagger \mathbf{M}_m \,|\psi\rangle \\ &= \langle\psi|\, \underbrace{\sum_m \mathbf{M}_m^\dagger \mathbf{M}_m}_{\mathbb{I}} \,|\psi\rangle \\ &= \langle\psi|\,\psi\rangle \\ &= 1\end{aligned}$$

Example: Measurement of a qubit in the computational basis

Let $\mathbf{M}_0 = |0\rangle\langle 0|$ and $\mathbf{M}_1 = |1\rangle\langle 1|$.
Note the following:

$$\begin{aligned}\mathbf{M}_0^2 &= |0\rangle\langle 0|\,0\rangle\langle 0| \\ &= |0\rangle\langle 0| \\ &= \mathbf{M}_0\end{aligned}$$

Similarly:

$$\mathbf{M}_1^2 = \mathbf{M}_1$$

Also:

$$\mathbf{M}_0^\dagger \mathbf{M}_0 + \mathbf{M}_1^\dagger \mathbf{M}_1 = \mathbb{I}$$

For an arbitrary quantum state $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$, the probability of measuring a $|0\rangle$ state is

$$\begin{aligned}p(0) &= \langle\psi|\, \mathbf{M}_0^\dagger \mathbf{M}_0 \,|\psi\rangle \\ &= \langle\psi|\, \mathbf{M}_0 \,|\psi\rangle \\ &= \langle\psi|\,0\rangle\langle 0|\,\psi\rangle \\ &= |\alpha|^2.\end{aligned}$$

Similarly for the $|1\rangle$ state,

$$p(1) = \langle\psi|\, \mathbf{M}_1^\dagger \mathbf{M}_1 \,|\psi\rangle = |\beta|^2.$$

After measurement, the $|0\rangle$ state becomes

$$\frac{\mathbf{M}_0}{|\alpha|}\,|\psi\rangle = \frac{\alpha}{|\alpha|}\,|0\rangle \cong |0\rangle$$

and the $|1\rangle$ state becomes

$$\frac{\mathbf{M}_1}{|\beta|}|\psi\rangle = \frac{\beta}{|\beta|}|1\rangle \cong |1\rangle\,,$$

up to global phase factors, which are irrelevant.

Projective measurements

Projective measurements are a special class of measurements described by an *observable* $\mathbf{M}$, which is a Hermitian operator on the state space being observed. The spectral decomposition of $\mathbf{M}$ is

$$\mathbf{M} = \sum_m m\ \mathbf{P}_m$$

where $\mathbf{P}_m$ is the *projector* onto the eigenspace of $\mathbf{M}$ with eigenvalue m. Upon measuring $|\psi\rangle$, the probability of outcome m is given by

$$p(m) = \langle\psi|\,\mathbf{P}_m\,|\psi\rangle$$

and the state immediately after a measurement outcome m is

$$\frac{\mathbf{P}_m}{\sqrt{p(m)}}|\psi\rangle\,.$$

As long as the general measurement operators $\mathbf{M}_m$ satisfy the additional condition

$$\mathbf{M}_m\mathbf{M}_{m'} = \delta_{m,m'}\mathbf{M}_m$$

(*i.e.* they are *orthogonal projectors*), then they are also *projective measurement operators*.

The average value of a projective measurement is

$$\begin{aligned}\sum_m m\ p(m) &= \sum_m m\,\langle\psi|\,\mathbf{P}_m\,|\psi\rangle \\ &= \langle\psi|\sum_m m\ \mathbf{P}_m\,|\psi\rangle \\ &= \langle\psi|\,\mathbf{M}\,|\psi\rangle\end{aligned}$$

Often this formula is written as an *expectation value*

$$\langle\mathbf{M}\rangle = \langle\psi|\,\mathbf{M}\,|\psi\rangle\,.$$

There is an associated variance for observations of $\mathbf{M}$:

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

Note that the Heisenberg Uncertainty Principle can be stated as

$$\Delta(\mathbf{A})\Delta\mathbf{B} \geq \frac{1}{2}|\langle\psi|\,[\mathbf{A},\mathbf{B}]\,|\psi\rangle|,$$

where $\mathbf{A}$ and $\mathbf{B}$ are Hermitian operators (*i.e.* observables).

Example: Expectation values of operators

Suppose the qubit state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ is measured using the observable $\mathbf{Z}$.

$$\begin{aligned}
\langle\mathbf{Z}\rangle &= \langle\psi|\,\mathbf{Z}\,|\psi\rangle \\
&= \left(\frac{1}{\sqrt{2}}(\langle 0| + \langle 1|)\right)\mathbf{Z}\left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\right) \\
&= \frac{1}{2}\left[\langle 0|\,\mathbf{Z}\,|0\rangle + \langle 1|\,\mathbf{Z}\,|0\rangle + \langle 0|\,\mathbf{Z}\,|1\rangle + \langle 1|\,\mathbf{Z}\,|1\rangle\right] \\
&= \frac{1}{2}[1+0+0-1] \\
&= 0 \\
\langle\mathbf{Z}^2\rangle &= \langle\psi|\,\mathbf{Z}^2\,|\psi\rangle \\
&= \frac{1}{2}\left[\langle 0|\,\mathbf{Z}^2\,|0\rangle + \langle 1|\,\mathbf{Z}^2\,|0\rangle + \langle 0|\,\mathbf{Z}^2\,|1\rangle + \langle 1|\,\mathbf{Z}^2\,|1\rangle\right] \\
&= \frac{1}{2}[1+0+0+1] \\
&= 1
\end{aligned}$$

Hence, the average measurement of the observable $\mathbf{Z}$ is 0, and the variance of the measurement of $\mathbf{Z}$ is 1.
Similarly, suppose the state is $|0\rangle$, then $\langle\mathbf{Z}\rangle = \langle 0|\,\mathbf{Z}\,|0\rangle = 1$ and $\langle\mathbf{Z}\rangle = \langle 0|\,\mathbf{Z}\,|0\rangle = 1$, then the average measurement is 1 and the variance is 0.

Example: The single-qubit observable

Consider the single-qubit observable:

$$\begin{aligned}
\vec{v}\cdot\vec{\sigma} &= v_x\mathbf{X} + v_y\mathbf{Y} + v_z\mathbf{Z} \text{ where } |\vec{v}| = 1 \\
&= v_x\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} + v_y\begin{pmatrix}0 & i\\ -i & 0\end{pmatrix} + v_z\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \\
&= \begin{pmatrix}v_z & v_x + iv_y\\ v_x - iv_y & -v_z\end{pmatrix}
\end{aligned}$$

The eigenvalues of $\vec{v}\cdot\vec{\sigma}$ are $\lambda = \pm 1$, therefore the projectors onto the $\lambda = \pm 1$ subspace are given by

$$\mathbf{P}_{\pm} = \frac{(\mathbb{I} \pm \vec{v}\cdot\vec{\sigma})}{2}.$$

For $\lambda = +1$, since $\vec{v}\cdot\vec{\sigma}\,|\psi_+\rangle = (+1)\,|\psi_+\rangle$ and $(\vec{v}\cdot\vec{\sigma})^2 = \mathbb{I}$, then

$$\begin{aligned}\frac{(\mathbb{I}+\vec{v}\cdot\vec{\sigma})}{2}\,|\psi_+\rangle &= \left(\frac{1+1}{2}\right)|\psi_+\rangle \\ &= |\psi_+\rangle\,.\end{aligned}$$

Similarly, $\frac{(\mathbb{I}-\vec{v}\cdot\vec{\sigma})}{2}\,|\psi_-\rangle = \left(\frac{1+1}{2}\right)|\psi_-\rangle = |\psi_-\rangle$ for the $\lambda = -1$ case.

1.1.4 Postulate no. 4:

The state space of a composite system is the tensor product of the state spaces of the component systems.

The state of a n component system prepared in the states $|\psi_i\rangle$ is given by $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle$.

Example: Calculating $\langle \mathbf{X}_1\mathbf{Z}_2\rangle$

Show that $\langle \mathbf{X}_1\mathbf{Z}_2\rangle$ for a two-qubit system in state $\frac{|00\rangle+|11\rangle}{\sqrt{2}}$ is equal to zero.

$$\begin{aligned}\langle \mathbf{X}_1\mathbf{Z}_2\rangle &= \frac{1}{2}\big(\langle 00| + \langle 11|\big)\mathbf{X}_1\mathbf{Z}_2\big(|00\rangle + |11\rangle\big) \\ &= \frac{1}{2}\big[\langle 0_1|\mathbf{X}_1|0_1\rangle\langle 0_2|\mathbf{Z}_2|0_2\rangle + \langle 1_1|\mathbf{X}_1|0_1\rangle\langle 1_2|\mathbf{Z}_2|0_2\rangle \\ &\quad + \langle 0_1|\mathbf{X}_1|1_1\rangle\langle 0_2|\mathbf{Z}_2|1_2\rangle + \langle 1_1|\mathbf{X}_1|1_1\rangle\langle 1_2|\mathbf{Z}_2|1_2\rangle\big] \\ &= \frac{1}{2}[0+0+0+0] \\ &= 0\end{aligned}$$

Note that $|\psi\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$ is an example of an *entangled* state: there are *no* single qubit states $|a\rangle$, $|b\rangle$ for which $|\psi\rangle = |a\rangle \otimes |b\rangle$.

$$\begin{aligned}\frac{|00\rangle+|11\rangle}{\sqrt{2}} &= (\alpha|0\rangle + \beta|1\rangle)\otimes(\gamma|0\rangle + \delta|1\rangle) \\ &= \alpha\gamma|00\rangle + \beta\gamma|10\rangle + \alpha\delta|01\rangle + \beta\delta|11\rangle\end{aligned}$$

This result implies that $\alpha\gamma = \beta\delta = \frac{1}{\sqrt{2}}$ and $\beta\gamma = \alpha\delta = 0$, which is impossible.

1.2 Density matrices

The density operator language provides a convenient means for describing quantum systems whose states are not completely known (*i.e. mixed* states versus *pure* states). The density matrix of a system is defined by

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

where p_i is the probability that the system is in the quantum state $|\psi_i\rangle$. This is appropriate, for example, for describing an *ensemble* of quantum systems (*e.g.* nuclear spins in NMR).

The evolution of a density matrix ρ under the operation of $\mathbf{U}$:

$$\rho = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \quad \xrightarrow{\mathbf{U}} \quad \sum_i p_i \mathbf{U} \left|\psi_i\right\rangle \left\langle\psi_i\right| \mathbf{U}^\dagger = \mathbf{U}\rho\mathbf{U}^\dagger$$

The probability of a measurement outcome of m for a given state $|\psi_i\rangle$:

$$\begin{aligned} p(m|i) &= \left\langle\psi_i\right| \mathbf{M}_m^\dagger \mathbf{M}_m \left|\psi_i\right\rangle \\ &= \mathrm{Tr}\left(\mathbf{M}_m^\dagger \mathbf{M}_m \left|\psi_i\right\rangle \left\langle\psi_i\right|\right) \\ \text{So } p(m) &= \sum_i p(m|i) p_i \\ &= \sum_i p_i \,\mathrm{Tr}\left(\mathbf{M}_m^\dagger \mathbf{M}_m \left|\psi_i\right\rangle \left\langle\psi_i\right|\right) \\ &= \mathrm{Tr}\left(\mathbf{M}_m^\dagger \mathbf{M}_m \rho\right) \end{aligned}$$

The state after measurement with an outcome m:

$$\rho_m = \frac{\mathbf{M}_m \rho \mathbf{M}_m^\dagger}{\mathrm{Tr}\left(\mathbf{M}_m^\dagger \mathbf{M}_m \rho\right)}$$

Note that a pure state (*i.e.* $\rho = |\psi\rangle\langle\psi|$) satisfies $\mathrm{Tr}(\rho^2) = 1$, whereas a mixed state satisfies $\mathrm{Tr}(\rho^2) < 1$.

1.2.1 General properties of density matrices

1. Trace condition: $\mathrm{Tr}(\rho) = 1$

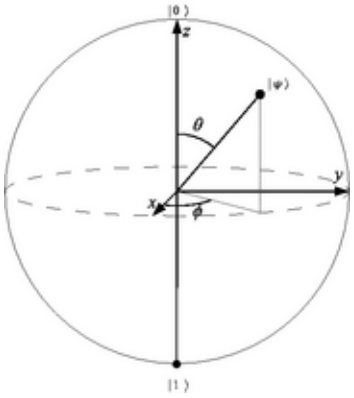
$$\begin{aligned} \mathrm{Tr}(\rho) &= \sum_i p_i \,\mathrm{Tr}(\left|\psi_i\right\rangle \left\langle\psi_i\right|) \\ &= \sum_i p_i \\ &= 1 \end{aligned}$$

2. Positivity: ρ is a *positive* operator

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

Note that Postulate 4 becomes $\rho = \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n$ for n component systems prepared individually in the states ρ_i.

Example: The Bloch sphere for single-qubit mixed states



An arbitrary density matrix for a single-qubit mixed state is expressed as

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

where $\vec{r}$ is a real, three-dimensional vector and satisfies $|\vec{r}| \leq 1$. It is known as the *Bloch vector*. If $|\vec{r}| = 1$, then ρ lies on the surface of the Bloch sphere and is a pure state. If $|\vec{r}| < 1$, then ρ lies inside the Bloch sphere and is a mixed state.
The axes of the Bloch sphere correspond to the Pauli spin matrices $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$. If $\vec{r} = 0$, then $\rho = \frac{\mathbb{I}}{2}$, which is a fully mixed state (*i.e.* we have no information about it).
Note that the statement that ρ is a pure state if $|\vec{r}| = 1$ follows from the earlier observation that $\mathbf{P}_\pm = \frac{\mathbb{I}\pm\vec{v}\cdot\vec{\sigma}}{2}$ is a projector onto a single-qubit pure state and, therefore, has the form $|\psi\rangle\langle\psi|$ just as a pure density matrix does.

1.2.2 Reduced density matrices

The density matrix formalism is extremely useful for analyzing composite quantum systems. Let A and B be two systems described by a joint density matrix ρ^{AB}. The *reduced density matrix* for subsystem A is

$$\rho^A = \mathrm{Tr}_B\left(\rho^{AB}\right)$$

where Tr_B denotes the *partial trace* over subsystem B, that is

$$\mathrm{Tr}_B\left(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|\right) \equiv |a_1\rangle\langle a_2|\,\mathrm{Tr}\left(|b_1\rangle\langle b_2|\right)$$

where $\mathrm{Tr}\left(|b_1\rangle\langle b_2|\right) = \langle b_2|\,b_1\rangle$. Note that the partial trace is also linear in its input.

Example: Partial trace over an unentangled state

Let the system be in an unentangled state $\rho^{AB} = \rho \otimes \sigma$.
Therefore,

$$\begin{aligned}\rho^A &= \mathrm{Tr}_B(\rho \otimes \sigma)\\ &= \rho\ \mathrm{Tr}(\sigma)\\ &= \rho\end{aligned}$$

The density matrix of a Bell state:

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

The reduced density matrix of subsystem A:

$$\begin{aligned}\rho^A &= \mathrm{Tr}_B\left(\rho^{AB}\right)\\ &= \frac{|0\rangle\langle 0|_A\langle 0|\,0\rangle_B + |1\rangle\langle 0|_A\langle 1|\,0\rangle_B + |0\rangle\langle 1|_A\langle 0|\,1\rangle_B}{2}\\ &\quad + \frac{|1\rangle\langle 1|_A\langle 1|\,1\rangle_B}{2}\\ &= \frac{|0\rangle\langle 0|_A + |1\rangle\langle 1|_A}{2}\\ &= \frac{\mathbb{I}_A}{2}\end{aligned}$$

The initial joint state is known exactly, but if all the information about one subsystem was 'discarded' by taking its partial trace, then the other subsystem is in a mixed state. In this case, we have no knowledge about the remaining subsystem as it is in a maximally mixed state.

1.3 Qubits, Gates, and Circuits

A qubit exists in a two-dimensional complex vector space. Let $|0\rangle$ and $|1\rangle$ constitute an orthonormal basis for this space. An arbitrary qubit state is a linear *superposition* state expressed as

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$$

where $|\alpha|^2 + |\beta|^2 = 1$.

Some physical examples of qubits:

Spin-$\frac{1}{2}$ particles	$\|\uparrow\rangle \equiv \|0\rangle$	$\|\downarrow\rangle \equiv \|1\rangle$
Photon polarization	$\|\circlearrowleft\rangle \equiv \|0\rangle$	$\|\circlearrowright\rangle \equiv \|1\rangle$
Ground (g) and first excited (e) atomic states	$\|g\rangle \equiv \|0\rangle$	$\|e\rangle \equiv \|1\rangle$

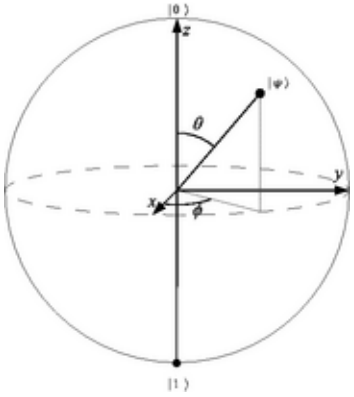
Since $|\alpha|^2 + |\beta|^2 = 1$, the qubit state can be rewritten as

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

where γ, ϕ, and θ are real numbers. The *global phase* $e^{i\gamma}$ has no observable effects, so the qubit state is effectively

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

where θ and ϕ define a point on the Bloch sphere.



Note that, in principle, it takes an infinite amount of information to perfectly describe the continuous variables θ and ϕ (*i.e.* there are an infinite number of points on the sphere), but that does not mean that an infinite amount of information can be stored in one qubit; this is because upon measurement only one of two results (0 or 1) can be found and the state will subsequently collapse. Therefore, only one bit of information is gained from a qubit measurement. Nevertheless, the extra 'hidden' quantum information is the reason that quantum information processing (*i.e.* with many qubits) apparently allows us to solve problems that are classically intractable.

1.4 Multiple Qubits and the Exponential Growth of the Hilbert Space

Consider a two-qubit system. An arbitrary quantum state can be written in the *computational basis* as:

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

The probability of measuring the system in, say, the state $|10\rangle$ is $|\alpha_{10}|^2$. Note that there are $2^2 = 4$ *amplitudes*. If only the first qubit is measured, then the probability of finding it in the state $|0\rangle$ would be $|\alpha_{00}|^2 + |\alpha_{01}|^2$, and the post-measurement state would be

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

More generally, for n qubits, computational basis states are of the form $|x_1, x_2, \cdots x_n\rangle$ where $x \,\epsilon\, \{0, 1\}$. A general superposition state, therefore, requires 2^n amplitudes. What if $n = 500$? Then there are 2^{500} amplitudes, which is more than the estimated number of atoms in the universe! No conceivable *classical computer* could keep track of all of this data, but nature can. This is what quantum information processing (quantum computing) seeks to somehow exploit!

1.5 Single-qubit Gates

Logic gates perform manipulations on information stored in bits. Classically, for example, the NOT gate reverses a bit state (*i.e.* $0 \to 1$ and $1 \to 0$), which is called a *bit flip*. What would a quantum NOT gate do?

$$\alpha|0\rangle + \beta|1\rangle \to \alpha|1\rangle + \beta|0\rangle$$

This is accomplished by $\mathbf{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, since $\mathbf{X}|0\rangle = |1\rangle$ and $\mathbf{X}|1\rangle = |0\rangle$. For any valid gate operation, the qubit normalization condition must be satisfied both before *and* after the gate; this implies that any gate U must be *unitary* (*i.e.* $\mathbf{U}^\dagger\mathbf{U} = \mathbb{I}$). This is the only condition!

Other useful single-qubit gates:

Phase flip $\quad \mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \begin{cases} \mathbf{Z}|0\rangle = |0\rangle \\ \mathbf{Z}|1\rangle = -|1\rangle \end{cases}$

Bit flip and phase flip

$$\mathbf{Y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \mathbf{Y}|0\rangle = i|1\rangle \qquad \mathbf{Y}|1\rangle = -i|0\rangle$$
$$\mathbf{XZ} = i\mathbf{Y} = -\mathbf{ZX}$$

$$\mathbf{H} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} \qquad \mathbf{H}|0\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}} \qquad \mathbf{H}|1\rangle = \frac{|0\rangle-|1\rangle}{\sqrt{2}}$$
$$\mathbf{H}^2 = \mathbb{I}$$
$$\mathbf{Z}^2 = \mathbf{X}^2 = \mathbf{Z}^2 = \mathbb{I}$$

There are an infinite number of single-qubit unitary operators, and thus quantum gates, but only the properties of a subset need to be understood to understand those of the full set. An arbitrary 2×2 unitary matrix can be decomposed as

$$\mathbf{U} = e^{i\alpha}\begin{pmatrix}e^{-i\frac{\beta}{2}} & 0\\ 0 & e^{i\frac{\beta}{2}}\end{pmatrix}\begin{pmatrix}\cos\frac{\gamma}{2} & -\sin\frac{\gamma}{2}\\ \sin\frac{\gamma}{2} & \cos\frac{\gamma}{2}\end{pmatrix}\begin{pmatrix}e^{-i\frac{\delta}{2}} & 0\\ 0 & e^{i\frac{\delta}{2}}\end{pmatrix}$$

where $e^{i\alpha}$ is called a *global phase shift*. This is the *Z-Y decomposition* since

$$\mathbf{U} = e^{i\alpha}\mathbf{R}_z(\beta)\ \mathbf{R}_y(\gamma)\ \mathbf{R}_z(\delta)$$

where $\mathbf{R}_z(\beta)$ is a rotation about the $\hat{z}$ axis by an angle β (*i.e.* $\mathbf{R}_z(\beta) = e^{-i\frac{\beta}{2}z}$).

An arbitrary rotation about the $\hat{n}$ axis is written as

$$\begin{aligned}\mathbf{R}_n(\theta) &= e^{-i\theta\frac{\vec{\sigma}\cdot\hat{n}}{2}}\\ &= \cos\left(\frac{\theta}{2}\right)\mathbb{I} - i\sin\left(\frac{\theta}{2}\right)(n_x\mathbf{X} + n_y\mathbf{Y} + n_z\mathbf{Z})\end{aligned}$$

where $\vec{\sigma}\cdot\vec{x} = \mathbf{X}$, $\vec{\sigma}\cdot\vec{y} = \mathbf{Y}$, and $\vec{\sigma}\cdot\vec{z} = \mathbf{Z}$.

Example: Generalized rotation

Show that $e^{-i\frac{\theta}{2}\mathbf{X}} = \cos\left(\frac{\theta}{2}\right)\mathbb{I} - i\sin\left(\frac{\theta}{2}\right)\mathbf{X}$.

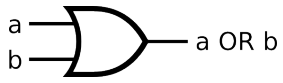
$$\begin{aligned}e^{-i\frac{\theta}{2}\mathbf{X}} &= \mathbb{I} + \left(-i\frac{\theta}{2}\mathbf{X}\right) + \frac{\left(-i\frac{\theta}{2}\mathbf{X}\right)^2}{2} + \frac{\left(-i\frac{\theta}{2}\mathbf{X}\right)^3}{6} + \cdots\\ &= \left(\mathbb{I} - \frac{\left(\frac{\theta}{2}\right)^2}{2}\mathbf{X}^2 + \cdots\right) + \mathbf{X}\left(-i\left(\frac{\theta}{2}\right) + i\frac{\left(\frac{\theta}{2}\right)^3}{6}\mathbf{X}^2 - \cdots\right)\\ &= \mathbb{I}\left(1 - \frac{1}{2}\left(\frac{\theta}{2}\right)^2 + \cdots\right) - i\mathbf{X}\left(\left(\frac{\theta}{2}\right) - \frac{1}{6}\left(\frac{\theta}{2}\right)^3 + \cdots\right)\\ &= \mathbb{I}\cos\left(\frac{\theta}{2}\right) - i\mathbf{X}\sin\left(\frac{\theta}{2}\right)\end{aligned}$$

As a result, if Z and Y rotations of arbitrary angles are possible, then any possible 2×2 unitary (and thus any single-qubit gate) can be generated.

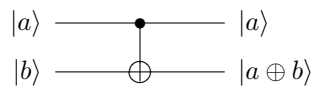
In addition, two-qubit logic (*i.e. conditional*) gates are needed. An appropriate finite set of one- and two-qubit gates is *universal*, meaning that the set can be used to implement any possible unitary on the entire system of many qubits.

1.6 Conditional Logic Gates

Classical two-bit gates are irreversible.



However, quantum gates are unitary and, therefore, reversible. For example, the controlled-NOT gate:

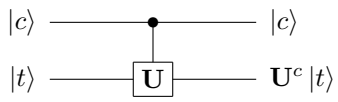


Note that the number of output qubits is always equal to the number of input qubits.

$$\mathbf{U}_{\text{CNOT}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = |0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes \mathbf{X}$$

In other words, if qubit a is 0, then apply $\mathbb{I}$ to qubit b; if qubit a is 1, then apply $\mathbf{X}$ to qubit b.

Note that $\mathbf{U}^2_{\text{CNOT}} = \mathbb{I} \otimes \mathbb{I}$, so it is reversible. The first qubit, a, is called the *control qubit*, and the second qubit, b, is the *target qubit*. More generally, there can be a controlled-U gate.

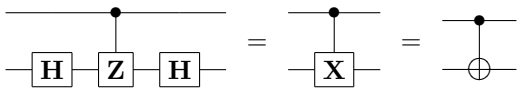


$$|c\rangle |t\rangle \rightarrow |c\rangle \mathbf{U}^c |t\rangle$$

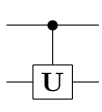
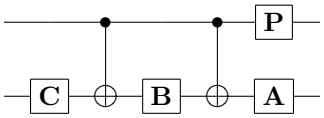
Example: Building a CNOT from a controlled-Z gate

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{c}-\mathbf{Z} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Since $\mathbf{HZH} = \mathbf{X}$:



A controlled-U gate, for an arbitrary U, can be decomposed as

$\mathbf{G} =$  $=$ 

where $\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\alpha} \end{pmatrix}$ and $\mathbf{U} = e^{i\alpha}\mathbf{A} \times \mathbf{B} \times \mathbf{C}$ and $\mathbf{ABC} = \mathbb{I}$.

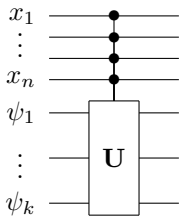
Example: Decomposing the controlled-Z gate

Verify that $\mathbf{A} = \mathbb{I},\ \mathbf{B} = \mathbf{R}_z\left(-\frac{\pi}{2}\right),\ \mathbf{C} = \mathbf{R}_z\left(\frac{\pi}{2}\right)$, and $\mathbf{P} = \mathbb{I}$.

$$\begin{aligned}
\mathbf{G} &= \mathbf{R}_z\left(-\frac{\pi}{2}\right)\mathbf{U}_{\text{CNOT}}\ \mathbf{R}_z\left(\frac{\pi}{2}\right)\mathbf{U}_{\text{CNOT}}\ \mathbb{I} \\
&= \left(\frac{\mathbb{I}_2 - i\mathbf{Z}_2}{\sqrt{2}}\right)\left(|0\rangle\langle 0|_1\,\mathbb{I}_2 + |1\rangle\langle 1|_1\,\mathbf{X}_2\right)\left(\frac{\mathbb{I}_2 + i\mathbf{Z}_2}{\sqrt{2}}\right) \\
&\quad \cdot \left(|0\rangle\langle 0|_1\,\mathbb{I}2 + |1\rangle\langle 1|_1\,\mathbf{X}_2\right) \\
&= \left(|0\rangle\langle 0|_1\,\mathbb{I}_2 + |1\rangle\langle 1|_1\left(\frac{\mathbb{I} - i\mathbf{Z}_2}{\sqrt{2}}\right)\mathbf{X}_2\left(\frac{\mathbb{I} + i\mathbf{Z}_2}{\sqrt{2}}\right)\right) \\
&\quad \cdot \left(|0\rangle\langle 0|_1\,\mathbb{I}_2 + |1\rangle\langle 1|_1\,\mathbf{X}_2\right) \\
&= \left(|0\rangle\langle 0|_1\,\mathbb{I}_2 + |1\rangle\langle 1|_1\,\mathbf{Y}_2\right)\left(|0\rangle\langle 0|_1\,\mathbb{I}_2 + |1\rangle\langle 1|_1\,\mathbf{X}_2\right) \\
&= \left(|0\rangle\langle 0|_1\,\mathbb{I}_2 + |1\rangle\langle 1|_1\,\mathbf{Y}_2\mathbf{X}_2\right) \\
&= |0\rangle\langle 0|_1\,\mathbb{I}_2 + |1\rangle\langle 1|_1\,i\mathbf{Z}
\end{aligned}$$

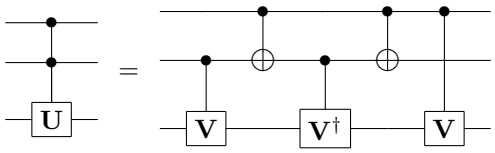
A general controlled operation on multiple qubits is defined by

$$C^n(\mathbf{U})\overbrace{|x_1, x_2, \cdots x_n\rangle}^{n-\text{qubits}}\overbrace{|\psi\rangle}^{k-\text{qubits}} = |x_1, x_2, \cdots x_n\rangle\,\mathbf{U}^{x_1,x_2,\cdots x_n}\,|\psi\rangle$$



where qubits x_1 through x_n are control qubits, and the k qubits of $|\psi\rangle$ are target qubits.

The important property to know is that these operations can be decomposed into a series of one- and two-qubit gates. For example,



where $\mathbf{V}$ satisfies $\mathbf{V}^2 = \mathbf{U}$. The *Toffoli gate* (controlled-controlled-NOT) is given by $\mathbf{V} = \frac{1-i}{2}(\mathbb{I} + i\mathbf{X})$. In fact, the general result will be that *any* unitary on an arbitrary number of qubits can be decomposed into a *universal set* of one- and two-qubit gates!

1.7 Universality and Universal Gate Sets

It is beyond the scope of this course to prove the following theorems so they will simply be stated.

1. *Two-level unitary gates are universal.*
   A unitary $\mathbf{U}$ acting on a d-dimensional Hilbert space, where $d = 2^n$, can be decomposed into a product of *two-level matrices* (*i.e.* those that act non-trivially only on two or fewer vector components).
   e.g. $\mathbf{U} = \mathbf{V}_1 \cdots \mathbf{V}_k$ where $\mathbf{V}_i$ are two-level matrices and $k \leq \frac{d(d-1)}{2}$
2. *Single-qubit and CNOT gates are universal.*
   Any two-level unitary $\mathbf{V_i}$ can be decomposed into a series of arbitrary single-qubit rotations and CNOT gates.

While these theorems are sufficient for proving the universality of one- and two-qubit gates, it turns out that we do not know how to construct *fault tolerant* quantum circuits containing arbitrary single-qubit rotations. Fortunately, we do know how to do this for certain finite gate sets (*i.e.* universal sets consisting of a finite number of discrete gates). One such set is

$$\{\mathbf{H},\ \mathbf{S},\ \text{CNOT},\ \mathbf{T}\}$$

where $\mathbf{H} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is a Hadamard gate, $\mathbf{S} = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$ is a phase gate, and $\mathbf{T} = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{pmatrix}$ is a $\frac{\pi}{8}$-rotation gate.

1.8 The Deutsch-Jozsa Algorithm

The game:

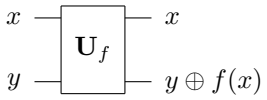
Alice picks a number between 0 and $2^n - 1$ (*i.e.* a binary number using n bits) and sends it to Bob, who calculates a function $f(x)$ and sends the result back to Alice, which is either 0 or 1. The function $f(x)$ is either *constant* (*i.e.* independent of x) or *balanced* (*i.e.* gives a result of 1 for half of all x, and 0 for the other half). Alice wants to determine with certainty whether $f(x)$ is constant or balanced with as little communication with Bob as possible.

Classical solution:

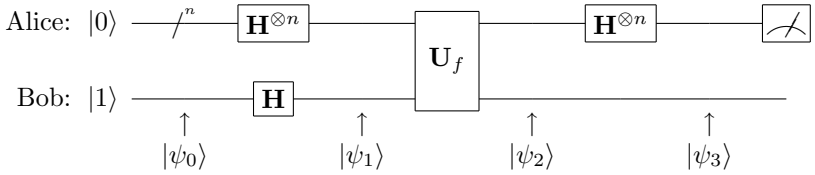
Alice must query Bob at least $\frac{2^n}{2} + 1$ times, each time sending n bits.

Quantum solution:

If Bob agrees to calculate $f(x)$ using a unitary transform $\mathbf{U}_f$ using the Deutsch-Jozsa algorithm, then Alice only requires *one* query to find the result with certainty!



How it works:



1. Alice and Bob initialize their qubits:

$$|\psi_0\rangle = |0\rangle^{\otimes n} |1\rangle = |00\cdots0\rangle |1\rangle$$

2. Alice applies Hadamard gates to $|0\rangle^{\otimes n}$ to create a superposition over all basis states of her n qubits, and Bob applies a Hadamard gate to $|1\rangle$ to create a superposition of his qubit:

$$\begin{aligned}
\mathbf{H}|0\rangle^{\otimes n} &= \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\cdots\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right) \\
&= \sum_{x\epsilon\{0,1\}^n} \frac{|x\rangle}{\sqrt{2^n}} \\
\mathbf{H}|1\rangle &= \frac{|0\rangle - |1\rangle}{\sqrt{2}} \\
\therefore |\psi_1\rangle &= \sum_{x\epsilon\{0,1\}^n} \frac{|x\rangle}{\sqrt{2^n}}\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right)
\end{aligned}$$

3. Bob applies the unitary $\mathbf{U}_f$ to evaluate $f(x)$:

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

$$\because |x, 0\rangle \rightarrow \begin{cases} |x, 0\rangle, f = 0 \\ |x, 1\rangle, f = 1 \end{cases}$$

$$|x, 1\rangle \rightarrow \begin{cases} |x, 1\rangle, & f = 0 \\ |x, 0\rangle, & f = 1 \end{cases}$$

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

4. Alice applies Hadamard gates to her register of n qubits:

$$\begin{aligned}
\mathbf{H}\left|x\right\rangle &= \sum_{z} \frac{(-1)^{x\cdot z}\left|z\right\rangle}{\sqrt{2^n}} \\
\therefore \mathbf{H}^{\otimes n}\left|x_1, x_2, \cdots, x_n\right\rangle &= \sum_{z_1,\cdots,z_n} \frac{(-1)^{x_1 z_1+\cdots+x_n z_n}}{\sqrt{2^n}}\left|z_1, \cdots, z_n\right\rangle \\
&= \sum_{z} \frac{(-1)^{x\cdot z}}{\sqrt{2^n}}\left|z\right\rangle \\
\therefore \left|\psi_3\right\rangle &= \sum_{x}\sum_{z} \frac{(-1)^{x\cdot z+f(x)}}{2^n}\left|z\right\rangle \left(\frac{\left|0\right\rangle-\left|1\right\rangle}{\sqrt{2}}\right)
\end{aligned}$$

   where $x \,\epsilon\, \{0,1\}^n,\ z \,\epsilon\, \{0,1\}^n$, and $x \cdot z$ is the bitwise product of x and z modulo 2. Note that for a single qubit, the action of the Hadamard can be written as $x \,\epsilon\, \{0,1\}$.

5. Alice measures her register.

For the state $\left|z\right\rangle = \left|0\right\rangle^{\otimes n}$, the measured amplitude is $\sum_x \frac{(-1)^{f(x)}}{2^n}$. If $f(x)$ is constant, then the amplitude will be +1 or -1 depending on the constant value that $f(x)$ takes. For example, in the case that $f(x)$ evaluates to either 0 or 1:

$$\sum_{x} \frac{(-1)^{f(x)}}{2^n} = \begin{cases} +1, & f(x) = 0 \\ -1, & f(x) = 1 \end{cases}$$

Since $\left\langle\psi_3\right|\left.\psi_3\right\rangle = 1$, then the output state will be $\left|0\right\rangle^{\otimes n}$ with certainty in these cases. However, if $f(x)$ is balanced, then

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

because the positive and negative contributions to the amplitude will cancel out. So there will be zero amplitude for measuring the output state $\left|0\right\rangle^{\otimes n}$ if $f(x)$ is balanced. In other words, another output state will be observed.

The Deutsch-Jozsa algorithm makes clever use of the quantum mechanical effects of *superposition* and *interference* in order to very efficiently extract *global* information about the function $f(x)$ in one query and to encode this information in the *classical* output of the algorithm. The last Hadamard gates $\mathbf{H}^{\otimes n}$ *interfere* the amplitudes from $\left|\psi_2\right\rangle$ such that they constructively or destructively interfere for the output $\left|0\right\rangle^{\otimes n}$ if $f(x)$ is constant or balanced, respectively. To learn the same global information about $f(x)$ classically and deterministically would require many more $(\frac{2^n}{2}+1)$ queries by Alice.

Caveats:

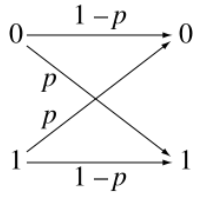
While this is an impressive quantum algorithm, the problem it solves is not really useful and there are classical probabilistic methods that would also allow

Alice to very rapidly discern the answer (not with perfect certainty, but with very high probability). However, it illustrates the power of quantum information and falls within a general category of algorithms based on the *quantum Fourier transform*. This category also includes Shor's famous algorithm that gives exponential speed-up for factoring numbers; a quantum algorithm that certainly *does* solve a useful problem with incredible efficiency!

Chapter 2

Quantum Error Correction and Fault Tolerance

Quantum states are extremely fragile, which means that they are very susceptible to *noise*. 'Noise' is a catch-all term that refers to *uncontrolled interactions with an environment*, where it is implied that the environment is also not under the control of the observer. For example, noise acting on a classical bit will flip it with some probability p:



Classically, one strategy for protecting against the loss of information is *redundancy*. In other words, *logical* bits are defined as multiple *physical* bits. For example, three physical bits can be used to represent one logical bit:

$$0_L = 000$$
$$1_L = 111$$

For example, with probability p, noise will transform 000 to 001. However, taking a simple 3-bit majority recovers the logical bit. The probability of 000 transforming into 011 is p^2, a smaller probability, but in this case the majority technique fails. As a result, the logical encoding reduces the bit error rate from one proportional to p to one proportional to p^2.

A similar idea can be applied to qubits, but not through simple redudancy. It is impossible to copy qubits. This is known as the *no-cloning theorem*.

Proof: No-cloning theorem

Consider the initial state $|\psi\rangle \otimes |s\rangle$, where $|\psi\rangle$ is to be copied to replace the pure state $|s\rangle$:

$$|\psi\rangle\,|s\rangle \xrightarrow{\mathbf{U}_{\text{copy}}} \mathbf{U}\,|\psi\rangle\,|s\rangle = |\psi\rangle\,|\psi\rangle$$

Suppose that this works for two particular states, $|\psi\rangle$ and $|\phi\rangle$.

$$\mathbf{U}\,|\psi\rangle\,|s\rangle = |\psi\rangle\,|\psi\rangle$$
$$\mathbf{U}\,|\phi\rangle\,|s\rangle = |\phi\rangle\,|\phi\rangle$$

Taking the inner product of two equations gives

$$\because \langle s|\,\langle\phi|\,\mathbf{U}^\dagger = \langle\phi|\,\langle\phi|$$
$$\therefore \langle\phi|\,\psi\rangle = (\langle\phi|\,\psi\rangle)^2$$
$$\therefore x = x^2$$

However, $x = x^2$ has only two solutions: $x = 0$ and $x = 1$. Therefore, $|\psi\rangle$ and $|\phi\rangle$ must be the same state or orthogonal states. A general cloning device is not possible!

2.1 3-qubit Code

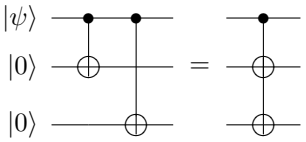
However, it is possible to 'spread' the desired quantum information among multiple physical qubits to protect against noise. The *3-qubit code* is the simplest example: it protects against individual bit flip errors (*i.e.* **X** errors). The encoding scheme is:

$$|0_L\rangle = |000\rangle$$
$$|1_L\rangle = |111\rangle$$

In other words, a general qubit state is given by:

$$\alpha\,|0_L\rangle + \beta\,|1_L\rangle = \alpha\,|000\rangle + \beta\,|111\rangle$$

A quantum circuit to perform this encoding would be:



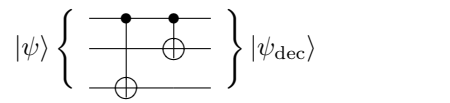
$$(\alpha\,|0\rangle + \beta\,|1\rangle) \otimes |00\rangle = \alpha\,|000\rangle + \beta\,|100\rangle \xrightarrow{\text{CNOTs}} \alpha 000 + \beta 111$$

The initial $|0\rangle$ qubits are known as the *ancilla* qubits.

Suppose that a bit-flip error (**X** operator) occurs on one of the three physical qubits:

$$\begin{array}{rl} \text{no error} & |\psi_0\rangle = \alpha\,|000\rangle + \beta\,|111\rangle \\ \text{flip } 1^{\text{st}} \text{ qubit} & |\psi_1\rangle = \alpha\,|100\rangle + \beta\,|011\rangle \\ \text{flip } 2^{\text{nd}} \text{ qubit} & |\psi_2\rangle = \alpha\,|010\rangle + \beta\,|101\rangle \\ \text{flip } 3^{\text{rd}} \text{ qubit} & |\psi_3\rangle = \alpha\,|001\rangle + \beta\,|110\rangle \end{array}$$

Now perform the *decoding* operation:



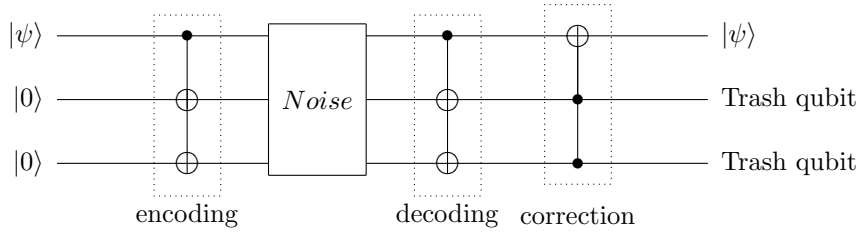
$$\begin{array}{rl} \text{no error} & (\alpha\,|0\rangle + \beta\,|1\rangle) \otimes |00\rangle = |\psi\rangle\,|00\rangle \\ \text{flip } 1^{\text{st}} \text{ qubit} & (\alpha\,|1\rangle + \beta\,|1\rangle) \otimes |11\rangle = |\psi\rangle\,|11\rangle \\ \text{flip } 2^{\text{nd}} \text{ qubit} & (\alpha\,|0\rangle + \beta\,|1\rangle) \otimes |10\rangle = |\psi\rangle\,|10\rangle \\ \text{flip } 3^{\text{rd}} \text{ qubit} & (\alpha\,|0\rangle + \beta\,|1\rangle) \otimes |01\rangle = |\psi\rangle\,|01\rangle \end{array}$$

Note that after decoding:

- Only in one instance is the quantum state corrupted: a bit-flip on the first qubit.
- The ancilla states corresponding to each instance are mapped to *orthogonal* subspaces.

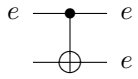
Now the ancilla qubits can be measured—without disturbing the quantum information—and the result gives exactly *which* error has occurred. In this case, an error only occurred if the ancilla state is $|11\rangle$, and the corresponding error operator is $\mathbf{X}$, so simply applying $\mathbf{X}$ again will correct the quantum state. This correction is contingent on observing $|11\rangle$.

The 3-qubit code illustrates the basic idea of quantum error correction. Note that instead of measuring the ancilla bits, an alternative approach could be to employ a controlled quantum gate so that the entire circuit would look like:

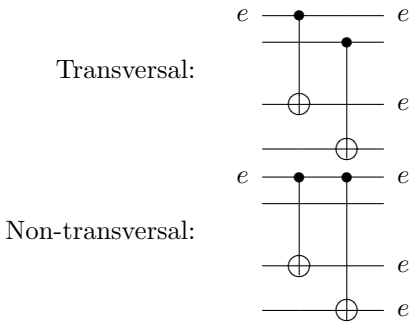


2.2 Fault Tolerance

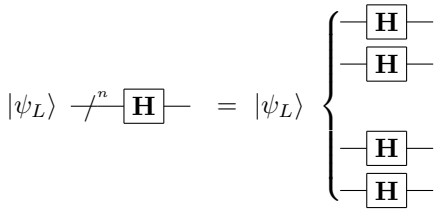
Fault tolerance is the concept that all operations on encoded bits must be performed in a way that minimizes the spread of errors. Every quantum operation, including state preparation, manipulation, correction, and readout, must be done in a fault tolerant way. For example, errors propagate through the **CNOT** gate in the following manner:



where e represents an error. When performing two-qubit logic on *encoded* qubits, an error on one physical qubit of one block cannot be allowed to affect more than one physical qubit of another block. A transversal application of a gate prevents this from occurring.

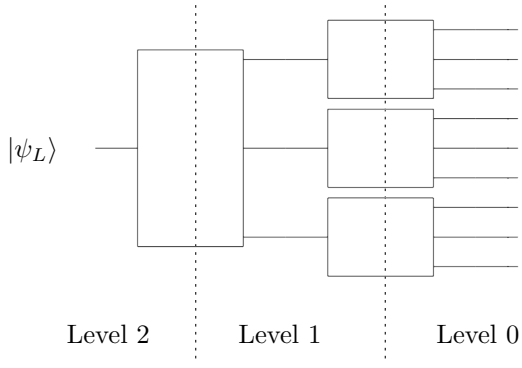


If a single gate on a logical qubit can be implemented as single physical qubit gates, then clearly it is *transversal* and, thus, fault tolerant.



2.3 Fault Tolerant Threshold

Provided that the noise in individual quantum gates is below a certain constant threshold, then it is possible to efficiently perform an arbitrarily large quantum computation. Concatenated error correction can be used to reduce the error rate.



	Level 0	Level 1	Level 2
Error probability	p	cp^2	$c(cp^2)^2$

For k levels of concatenation, the failure probability of a gate at the highest level is $\frac{(cp)^{2^k}}{c}$, while the size of the concatenated circuit goes as d^k times the size of the original circuit, where d is the maximum number of operations used in a fault tolerant procedure for performing an encoded gate of error correction.

For an algorithm with N gates with a desired accuracy of $1-\epsilon$, where ϵ is some small error, each gate must have an error on the order of $\frac{\epsilon}{N}$. Therefore, k levels of concatenation must be used to satisfy:

$$\frac{(cp)^{2^k}}{c} \leq \frac{\epsilon}{N}$$

Provided that $p \leq \frac{1}{c}$, such a k can be found. The limit $P_{th} = \frac{1}{c}$ is known as the *threshold error probability*. The aim is to reduce the threshold so that $P_{sys} \leq P_{th}$. The best current estimates for the threshold error probability is in the range of $10^{-3} \leq P_{th} \leq 10^{-6}$, but an exact P_{th} is not known for any specific system or implementation. Reaching error levels for real gates in real systems as small as these estimates for P_{th} is the *primary challenge* for experimental implementation of quantum information processing.

2.4 Simulating Quantum Dynamics

The dynamical behaviour of many simple quantum systems is governed by Schrödinger's equation:

$$i\hbar \frac{\mathrm{d}\,|\psi\rangle}{\mathrm{d}t} = \mathcal{H}\,|\psi\rangle\,.$$

For example,

$$i\frac{\partial \psi(x)}{\partial t} = \underbrace{\left[-\frac{1}{2m}\frac{\partial^2}{\partial x^2} + V(x)\right]}_{\mathcal{H}} \psi(x)$$

where $\psi(x) = \langle x|\,\psi\rangle$. This is the position representation and the $\hbar$ has been absorbed into $\mathcal{H}$ (*i.e.* set $\hbar = 1$). This is simply a differential equation, and it would appear that classical simulation of it is well established. However, the challenge is due to the number of differential equations; the number that must be simulated grows exponentially with the degrees of freedom of the system (*i.e.* the number of basis states). For one qubit evolving under the Schrödinger equation, a system of two differential equations must be solved. For two qubits, four equations, and so on. In other words, 2^n differential equations must be solved for n qubits. For many physical systems of interest, no approximations are known that will effectively reduce this number so that simulations become classically feasible; in general, classical simulation of large quantum systems is totally hopeless!

How would a one-dimensional Schrödinger's equation be simulated on a quantum computer?

$$\mathcal{H} = \frac{p^2}{2m} + V(x)$$

where p is the momentum operator and x is the position operator.

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

Hence, $|\psi\rangle$ resides in an infinite-dimensional Hilbert space and the eigenvalues of x are continuous. However, the focus can be restricted to a region of interest $-d \leq x \leq d$, where d discretizes the problem by choosing a finite step size Δx so that

$$\left|\tilde{\psi}\right\rangle = \sum_{k=\frac{-d}{\Delta x}}^{+\frac{d}{\Delta x}} a_k |k\Delta x\rangle$$

is a good physical approximation of $|\psi\rangle$.

Hence, there are $\frac{2d}{\Delta x} + 1$ basis states which can be represented by n qubits provided $2^n = \frac{2d}{\Delta x} + 1$, *i.e.* there need to be $n = \log_2(\frac{2d}{\Delta x} + 1)$ qubits. Simply replace the basis states $|k\Delta x\rangle$ (eigenstates of x) with $|k\rangle$ to create a computational basis state of n qubits. Here, only n qubits are needed, whereas 2^n complex numbers would be necessary classically.

Next, we must compute $\left|\tilde{\psi}(t)\right\rangle = e^{-i\mathcal{H}t}\left|\tilde{\psi}(0)\right\rangle$. In general, $\left[\frac{p^2}{2m}, V(x)\right] \neq 0$, that is $\mathcal{H}_0 = \frac{p^2}{2m}$ and $\mathcal{H}_1 = V(x)$ do not necessarily *commute*. This means that $e^{-i\mathcal{H}t} \neq e^{-i\mathcal{H}_0 t}e^{-i\mathcal{H}_1 t}$. Therefore, an approximate method for computing $e^{-i\mathcal{H}t}$ must be found. An exact way of doing this is called the Trotter formula:

$$e^{i(A+B)t} = \lim_{n\to\infty}\left(e^{i\frac{At}{n}}e^{i\frac{Bt}{n}}\right)^n$$

Of course, in a real computation we cannot take $n \to \infty$, so an approximation must be used. One such approximation:

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

This approximation is good up to order $(\Delta t)^3$. So, to compute (approximately) $e^{-i\mathcal{H}t}$, it must be possible to compute $e^{-i\mathcal{H}_0 t} = e^{-i\frac{p^2}{2m}t}$ and $e^{-i\mathcal{H}_1 t} = e^{-iV(x)t}$.

For this problem, note that $\left|\tilde{\psi}\right\rangle$ is expressed in the eigenbasis of $\mathcal{H}_1$, so $e^{-i\mathcal{H}_1 t}$ is a diagonal transform

$$|k\rangle \longrightarrow e^{-iV(k\Delta x)\Delta t}|k\rangle$$

so this is straightforward to compute.

$e^{-i\mathcal{H}_0 t}$ can be computed by realizing that x and p are conjugate variables and are thus related by the *quantum Fourier transform*:

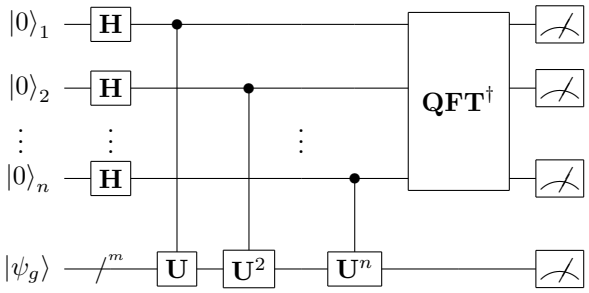
$$\mathbf{U}_{\mathrm{FFT}} \times \mathbf{U}_{\mathrm{FFT}}^{\dagger} = p.$$

Therefore, $e^{-i\mathcal{H}_0 t} = \mathbf{U}_{\mathrm{FFT}} e^{-i\frac{x^2}{2m}\Delta t} \mathbf{U}_{\mathrm{FFT}}^{\dagger}$. Note that the quantum Fourier transform is defined as

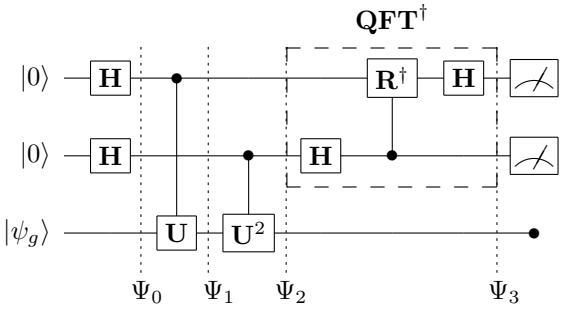
$$|j\rangle \longrightarrow \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{i2\pi\frac{jk}{N}} |k\rangle$$

Example: ***Simulated Quantum Computation of Molecular Energies****. A. Aspuru-Guzik* et al. *Science 309, 1704–1707 (2005).*

The calculation time for the energies of atoms and molecules scales exponentially with system size on a classical computer, but polynomially using quantum algorithms. The following algorithm can be applied to problems of chemical interest using modest numbers of qubits; it is based on the phase-estimation algorithm (PEA).
The first n qubits are the readout register, and the m qubits represented by $|\psi\rangle$ are the state register. The input state of the state register, $|\psi_y\rangle$ is an eigenstate of the Hamiltonian, $\mathcal{H}$, *e.g.* the ground state.



Measurement of the readout register gives $\phi = \omega_g t$.
To understand the function of this circuit, here are some simple examples worked through.
Say that there is a 2-qubit readout register and a 1-qubit state register. Say the Hamiltonian to be simulated is $\mathcal{H} = \frac{\omega_z}{2}$, whose eigenvalues are $\pm\frac{\omega}{2}$. The ground state is $|1\rangle$ with an eigenvalue of $-\frac{\omega}{2}$.



Note that $\mathbf{R}^\dagger = \begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix}$.

Following the state evolution through the circuit setting $|\psi_g\rangle = |1\rangle$.

$$\Psi_0 = \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)|1\rangle$$

$$\Psi_1 = \frac{1}{2}\Big[\,|0\rangle\,\big(\,|0\rangle + |1\rangle\,\big)\,|1\rangle + |1\rangle\,\big(\,|0\rangle + |1\rangle\,\big)\mathbf{U}\,|1\rangle\,\Big]$$

$$\Psi_2 = \frac{1}{2}\big[\,|00\rangle\,|1\rangle + |10\rangle\,\mathbf{U}\,|1\rangle + |01\rangle\,\mathbf{U}^2\,|1\rangle + |11\rangle\,\mathbf{U}^3\,|1\rangle\,\big]$$

$$\because \mathbf{U}(t) = e^{-i\frac{\omega t}{2}\mathbf{Z}} = \begin{pmatrix} e^{-i\frac{\omega t}{2}} & 0 \\ 0 & e^{i\frac{\omega t}{2}} \end{pmatrix}$$

$$\therefore \Psi_2 = \frac{1}{2}\big[\,|00\rangle\,|1\rangle + |10\rangle\,e^{i\frac{\omega t}{2}}\,|1\rangle + |01\rangle\,e^{i\omega t}\,|1\rangle + |11\rangle\,e^{i\frac{3\omega t}{2}}\,|1\rangle\,\big]$$

Applying $\mathbf{QFT}^\dagger$:

$$|00\rangle \xrightarrow{\mathbf{H}_2} |0\rangle\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right) \xrightarrow{\text{c-}\mathbf{R}_1^\dagger} \frac{|00\rangle + \mathbf{R}_1^\dagger\,|01\rangle}{\sqrt{2}}$$

$$\xrightarrow{\mathbf{H}_1} \frac{\big(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\big)\,|0\rangle + \big(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\big)\,|1\rangle}{\sqrt{2}} = \frac{1}{2}\big(\,|00\rangle + |01\rangle + |10\rangle + |11\rangle\,\big)$$

$$|01\rangle \xrightarrow{\mathbf{H}_2} |0\rangle\left(\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right) \xrightarrow{\text{c-}\mathbf{R}_1^\dagger} \frac{|00\rangle - \mathbf{R}_1^\dagger\,|01\rangle}{\sqrt{2}}$$

$$\xrightarrow{\mathbf{H}_1} \frac{\big(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\big)\,|0\rangle - \big(\frac{|0\rangle+|1\rangle}{\sqrt{2}}\big)\,|1\rangle}{\sqrt{2}} = \frac{1}{2}\big(\,|00\rangle - |01\rangle + |10\rangle - |11\rangle\,\big)$$

$$|10\rangle \xrightarrow{\mathbf{H}_2} |1\rangle\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right) \xrightarrow{\text{c-}\mathbf{R}_1^\dagger} \frac{|10\rangle + \mathbf{R}_1^\dagger\,|11\rangle}{\sqrt{2}}$$

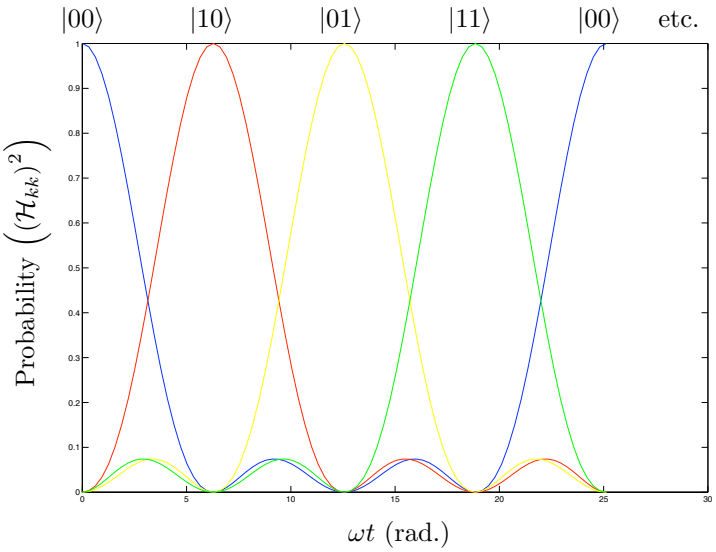
$$\xrightarrow{\mathbf{H}_1} \frac{\big(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\big)\,|0\rangle - i\big(\frac{|0\rangle-|1\rangle}{\sqrt{2}}\big)\,|1\rangle}{\sqrt{2}} = \frac{1}{2}\big(\,|00\rangle - i\,|01\rangle - |10\rangle + i\,|11\rangle\,\big)$$

Similarly, $|11\rangle \longrightarrow \dfrac{|00\rangle + i\,|01\rangle - |10\rangle - i\,|11\rangle}{2}$

Therefore,

$$\Psi_3 = \frac{1}{4}\Big[\big(\,|00\rangle + |01\rangle + |10\rangle + |11\rangle\,\big)\otimes|1\rangle$$
$$+ e^{i\frac{\omega t}{2}}\big(\,|00\rangle - i\,|01\rangle - |10\rangle + i\,|11\rangle\,\big)\otimes|1\rangle$$
$$+ e^{i\omega t}\big(\,|00\rangle - |01\rangle + |10\rangle - |11\rangle\,\big)\otimes|1\rangle$$
$$+ e^{i3\frac{\omega t}{2}}\big(\,|00\rangle + i\,|01\rangle - |10\rangle - i\,|11\rangle\,\big)\otimes|1\rangle\Big]$$

$$= \frac{1}{4}\Big[\,|00\rangle\left(1 + e^{i\frac{\omega t}{2}} + e^{i\omega t + e^{i3\frac{\omega t}{2}}}\right) + |01\rangle\left(1 - ie^{i\frac{\omega t}{2}} - e^{i\omega t} + ie^{i3\frac{\omega t}{2}}\right)$$
$$+ |10\rangle\left(1 - e^{i\frac{\omega t}{2}} + e^{i\omega t} - e^{i3\frac{\omega t}{2}}\right) + |11\rangle\left(1 + ie^{i\frac{\omega t}{2}} - e^{i\omega t} - ie^{i3\frac{\omega t}{2}}\right)\Big] \otimes |1\rangle$$
$$= \frac{1}{4}\big[f_{00}(\omega t)\,|00\rangle + f_{01}(\omega t)\,|01\rangle + f_{10}(\omega t)\,|10\rangle + f_{11}(\omega t)\,|11\rangle\,\big] \otimes |1\rangle$$



Hence, evolving under its Hamiltonian, $\frac{1}{2}\omega\mathbf{Z}$, the ground state $|1\rangle$ picks up a phase, $\phi = \omega t$, which the algorithm estimates by measuring the probabilities of the computational basis states of the readout register. In the Aspuru-Guzik paper, the authors simulate a quantum computer solving for the ground state electronic energies ω_g of some small molecules, such as H_2 and H_2O, using a classical computer. That is, they work out how to represent $\mathbf{U} = e^{-i\mathcal{H}t}$ using a number of quantum gates that scales polynomially with the system size.
A direct mapping was used in which each qubit represents the fermionic occupation state of a particular atomic orbital (*i.e.* $|0\rangle$ represents *occupied*, and $|1\rangle$ represents *not occupied*).

$$\text{Molecular system Fock space} \overset{\text{mapping}}{\Longrightarrow} \text{Hilbert space of qubits}$$

For the most compact mapping (minimal basis set), only eight qubits are required in the state register to simulate H_2O to a high accuracy.

Chapter 3

Quantum Control

Reference: *A. E. Bryson Jr. and Y. Ho,* Applied Optimal Control: Optimization, Estimation and Control. *Wiley (1975).*

Define a *performance index* $L(\vec{u})$, where $\vec{u}$ is a vector (set) of n control parameters $\vec{u} = (u_1, u_2, \ldots, u_n)$. The optimization problem is simply to minimize (or maximize) L and find $\vec{u}$ such that L is the global minimum (or maximum). Of course, finding the *global* minimum is not easy and, in general, one can only guarantee that the *local* minimum has been found.

The simplest optimization problem to consider is that which has no constraints on the control parameters, $\vec{u}$. The necessary conditions for a local minimum of L are then:

$$\frac{\partial L}{\partial \vec{u}} = 0, \textit{ i.e. } \frac{\partial L}{\partial u_i} = 0 \text{ for all } i$$

$$\frac{\partial^2 L}{\partial u^2} \geq 0, \textit{ i.e. the matrix } \frac{\partial^2 L}{\partial u_i \partial u_j} \text{ is positive semi-definite}$$

$$\text{(all eigenvalues } \geq 0)$$

If $\frac{\partial L}{\partial \vec{u}} = 0$ and $\frac{\partial^2 L}{\partial u^2} > 0$, (*i.e.* all eigenvalues are greater than 0,) then it is guaranteed to be a local minimum. If $\frac{\partial^2 L}{\partial u^2} = 0$, (*i.e.* the determinant of the matrix is equal to 0,) then it is a singular point and more information is needed to determine if it is a minimum.

Example: Finding a local minimum

$$L = u_1^2 - 2u_1u_2 + 4u_2^2$$

$$\begin{array}{lll} \frac{\partial L}{\partial u_1} & = 2u_1 - 2u_2 = 0 \quad \Rightarrow & u_1 = u_2 \\ \frac{\partial L}{\partial u_2} & = 8u_2 - 2u_1 = 0 \quad \Rightarrow & u_1 = u_2 = 0 \end{array}$$

The condition $u_1 = u_2 = 0$ identifies a stationary point at $\vec{u} = (0, 0)$.

$$\left.\begin{array}{ll} \frac{\partial^2 L}{\partial u_1^2} = 2 & \frac{\partial^2 L}{\partial u_2 \partial u_1} = -2 \\ \frac{\partial^2 L}{\partial u_1 \partial u_2} = -2 & \frac{\partial^2 L}{\partial u_2^2} = 8 \end{array}\right\} \Rightarrow \begin{pmatrix} 2 & -2 \\ -2 & 8 \end{pmatrix} = \frac{\partial^2 L}{\partial u_i \partial u_j}$$

$$\begin{vmatrix} 2-\lambda & 2 \\ -2 & 8-\lambda \end{vmatrix} = 0 \quad \Rightarrow \quad \lambda^2 - 10\lambda + 12 = 0$$

$$\therefore \lambda = 5 \pm \frac{\sqrt{52}}{2}$$
$$= 5 \pm \sqrt{13}$$

Both λ are greater than 0, and hence $\vec{u} = (0, 0)$ is a minimum.

Example: Identifying a saddle point

$$L = 3u_2^2 - u_1^2 + 2u_1 u_2$$

$$\begin{array}{lll} \frac{\partial L}{\partial u_1} & = 2u_2 - 2u_1 = 0 & \Rightarrow \quad u_1 = u_2 \\ \frac{\partial L}{\partial u_2} & = 6u_2 + 2u_1 = 0 & \Rightarrow \quad u_1 = -3u_2 = 0 \end{array}$$

The conditions $u_1 = u_2 = 0$ and $u_1 = -3u_2 = 0$ identify a stationary point at $\vec{u} = (0, 0)$.

$$\left.\begin{array}{ll} \frac{\partial^2 L}{\partial u_1^2} = -2 & \frac{\partial^2 L}{\partial u_1 \partial u_2} = 2 \\ \frac{\partial^2 L}{\partial u_2 \partial u_1} = 2 & \frac{\partial^2 L}{\partial u_2^2} = 6 \end{array}\right\} \Rightarrow \begin{vmatrix} -2-\lambda & 2 \\ 2 & 6-\lambda \end{vmatrix} = 0$$

$$\Rightarrow \lambda = 2 \pm 2\sqrt{5}$$

$$\therefore \lambda_1 > 0 \text{ and } \lambda_2 < 0$$

Therefore, this is a saddle point. Note that if $\lambda_1 > 0$ and $\lambda_2 = 0$, then this relation gives a singular point. (*e.g.* $L = (u_1 - u_2^2)(u_1 - 3u_2^2)$) It could be a minimum with respect to a certain $\mathrm{d}\vec{u}$ and a maximum with respect to another.

More generally, some constraints are placed on $\vec{u}$. An example is a set of equality constraints:

$$\left.\begin{array}{l} f_1(\vec{x_1}, \vec{u_1}) = 0 \\ f_2(\vec{x_1}, \vec{u_1}) = 0 \\ \quad \vdots \\ f_m(\vec{x_1}, \vec{u_1}) = 0 \end{array}\right\} \text{define } \vec{f} \text{ as a constraint vector}$$

where $\vec{x} = (x_1, x_2, \cdots, x_m)$. Now, $L = L(\vec{x}, \vec{u})$. A stationary point has $\mathrm{d}L = 0$ for arbitrary $\mathrm{d}\vec{u}$, and while holding $\mathrm{d}\vec{f} = 0$.

One method for finding $\vec{u}$ that yield stationary points is based on the use of *Lagrange multipliers*. A new function $H(\vec{u}, x, \lambda)$ is defined that *adjoins* the performance index and the constraints:

$$H(\vec{u}, \vec{x}, \lambda) = L(\vec{x}, \vec{u}) + \sum_{i=1}^{m} \lambda_i f_i(\vec{x}, \vec{u})$$

where λ_i are undetermined multipliers. The summation above can be written as:

$$\sum_{i=1}^{m} \lambda_i f_i(\vec{x}, \vec{u}) = \lambda^{\mathrm{T}} \vec{f}(\vec{x}, \vec{u})$$

i.e. as the scalar product $(\lambda_1, \lambda_2, \cdots, \lambda_m) \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix}$.

Suppose that some $\vec{u}$ is chosen and then $\vec{x}$ is determined via $\vec{f}(\vec{u}, \vec{x}) = 0$, so that $L = H$. Differential changes in H can be written as:

$$\mathrm{d}H = \frac{\partial H}{\partial x}\,\mathrm{d}x + \frac{\partial H}{\partial u}\,\mathrm{d}u$$

where the notation has been simplified so that $u \equiv \vec{u}$ and $x \equiv \vec{x}$.

The vector λ can be chosen so that

$$\frac{\partial H}{\partial x} = \frac{\partial L}{\partial x} + \lambda^{\mathrm{T}} \frac{\partial f}{\partial x} = 0 \quad \Rightarrow \quad \lambda^{\mathrm{T}} = -\frac{\partial L}{\partial x} \left(\frac{\partial f}{\partial x} \right)^{-1}$$

It follows that $\mathrm{d}H = \mathrm{d}L = \frac{\partial H}{\partial u}\,\mathrm{d}u$, or $\frac{\partial H}{\partial u}$ is the *gradient* of L with respect to u while holding $f(x, u) = 0$.

In summary, the *necessary conditions* for a stationary value of $L(x, u)$ are:

1. $f(x, u) = 0$ (m equations)
2. $\frac{\partial H}{\partial x} = 0$ (m equations)
3. $\frac{\delta H}{\delta u} = 0$ (n equations)

As a result, there are $2m + n$ equations to solve for $2m + n$ parameters.

Example: Finding a stationary value

Find a scalar u to yield a stationary value of $L = \frac{1}{2}\left(\frac{x^2}{a^2} + \frac{u^2}{b^2}\right)$ subject to the linear constraint:

$$f(x, u) = x + mu - c = 0$$

where a, b, c, and m are constants, and x and u are scalars.

$$H = \frac{1}{2}\left(\frac{x^2}{a^2} + \frac{u^2}{b^2}\right) + \lambda\,(x + mu - c)$$

Conditions:

$$\left.\begin{array}{r} x + mu - c = 0 \\ \frac{\partial H}{\partial x} = \frac{x}{a^2} + \lambda = 0 \\ \frac{\partial H}{\partial u} = \frac{u}{b^2} + \lambda m = 0 \end{array}\right\} \text{3 equations for 3 unknowns}$$

Solution:

$$\begin{aligned} x &= \frac{a^2 c}{a^2 + m^2 b^2} \\ u &= \frac{b^2 m c}{a^2 + m^2 b^2} \\ \lambda &= \frac{-c}{a^2 + m^2 b^2} \\ L_{\min} &= \frac{c^2}{2(a^2 + m^2 b^2)} \end{aligned}$$

The conditions above are sufficient for a stationary point; for a *local minimum*, sufficiency requires the additional condition: $\left(\frac{\partial^2 L}{\partial u^2}\right)_{f=0}$ must be a positive matrix.

3.1 Numerical solutions found using a first-order gradient method

Usually, the relations for $L(x, u)$ and $f(x, u)$ are quite complex, and therefore numerical methods must be used to determine the u that minimize H. (Usually, this is true for *quantum* optimal control as well!) One method is using a *steepest descent* method for finding minima (or *steepest ascent* for finding maxima). Such methods are also called *gradient methods*; they are iterative algorithms for improving estimates of control parameters u to satisfy $\frac{\partial H}{\partial u} = 0$.

Steps for finding a numerical solution using a first-order gradient method:

1. Guess an initial u.
2. Determine values of x from $f(x, u) = 0$.
3. Determine λ from $\lambda^T = -\left(\frac{\partial L}{\partial x}\right)\left(\frac{\partial f}{\partial x}\right)^{-1}$.
4. Determine the values of $\frac{\partial H}{\partial u} = \frac{\partial L}{\partial u} + \lambda^T \frac{\partial f}{\partial u}$, which generally does not equal 0.
5. Interpret $\frac{\partial H}{\partial u}$ as a gradient vector, update u by $\Delta u = -k\left(\frac{\partial H}{\partial u}\right)^T$ where

$k > 0$ and k is a scalar constant. The predicted change in H is therefore:

$$\Delta H \approx \frac{\partial H}{\partial u}\Delta u = \frac{\partial H}{\partial u}\left(-k\left(\frac{\partial H}{\partial u}\right)^T\right) = -k\left(\frac{\partial H}{\partial u}\right)\left(\frac{\partial H}{\partial u}\right)^T$$

6. Repeat steps 1 to 5 using revised u estimates until the quantity $\left(\frac{\partial H}{\partial u}\right)\left(\frac{\partial H}{\partial u}\right)^T$ is very small.

This method can be thought of as a hill-climbing method in u-space (if a maximum is being sought). Often, k is varied adaptively from step to step in order to ensure that the linear approximation to predicted change $\Delta H \approx \frac{\partial H}{\partial u}\Delta u$ is accurate. Nonetheless, overshooting is a common problem for this first-order gradient method when $\vec{u}$ is in close vicinity to the true minimum (or maximum).

3.2 Application of gradient methods to quantum optimal control: GRAPE algorithm

Reference: *N. Khaneja* et al. *Journal of Magnetic Resonance 172, 296–305 (2005).*

The GRAPE algorithm is used in NMR, superconducting qubits, ion traps, and other systems. In a system consisting of, for example, spins (or qubits) described by the internal Hamiltonian, $\mathcal{H}_o$, and a set of external control terms, $\mathcal{H}_k$, with amplitudes $u_k(t)$ that can be controlled, the total Hamiltonian is:

$$\mathcal{H}(t) = \mathcal{H}_o + \sum_{k=1}^{m} u_k(t)\mathcal{H}_k$$

where m is the total number of control terms acting on the qubits. In the case of NMR, the external control terms are the RF fields applied to the spins.

The first problem to consider is that of steering some initial quantum state, $\psi(0)$, to a desired final state at time T, $\psi(T) = C$.

$$\psi(0) \rightarrow \psi(T) = C$$

In density matrix notation, this above transformation is $\rho(0) \rightarrow \rho(T) = \rho_c$.

The overlap of $\psi(T)$ with C can be measured as the scalar product $\Phi_o = \langle C|\,\psi(T)\rangle$, or in density matrix notation:

$$\Phi_o = \mathrm{Tr}\left\{\rho_c^\dagger \rho(T)\right\}.$$

The time evolution of the system is given by the Liouvill-von Neumann equation:

$$\dot{\rho}(t) = -i[\mathcal{H}, \rho(t)] = -i(\mathcal{H}\rho - \rho\mathcal{H})$$

Assuming that the total evolution time T is discretized in N equal steps of duration $\Delta t = T/N$. The control amplitudes u_k are constant during each step,

and in the j-th step it is $u_k(j)$ for the k-th control Hamiltonian. During step j, the propagator is:

$$U_j = e^{-i\Delta t\left(\mathcal{H}_o + \sum_{k=1}^{m} u_k(j)\mathcal{H}_k\right)} = e^{-i\Delta t\ \mathcal{H}(j)}$$

The final state density matrix is:

$$\rho(T) = U_N \cdots U_1 \rho_o U_1^\dagger \cdots U_N^\dagger$$

The performance function is:

$$\Phi_o = \mathrm{Tr}\left\{\rho_c^\dagger \left(U_N \cdots U_1 \rho_o U_1^\dagger \cdots U_N^\dagger\right)\right\}$$

Since the trace of a product is invariant under cyclic permutations of factors (*c.f.* first problem set), then

$$\Phi_o = \mathrm{Tr}\left\{\lambda_j^\dagger \rho_j\right\}$$

where $\lambda_j = U_{j+1}^\dagger \cdots U_N^\dagger \rho_c U_N \cdots U_{j+1}$ and $\rho_j = U_j \cdots U_1 \rho_o U_1^\dagger \cdots U_j^\dagger$. ρ_j is the density operator at time $t = j\Delta t$, and λ is the "backward propagated" target state ρ_c at the same time $t = j\Delta t$.

How does the performance index Φ_o change when the control amplitude $u_k(j)$ is perturbed at time step j to $u_k(j) + \delta u_k(j)$? From the equation for u_j, the change to U_j can be calculated to first-order in $\delta u_k(j)$:

$$U_j = e^{-i\Delta t\left(\mathcal{H}_o + \sum_{k=1}^{m} u_k(j)\mathcal{H}_k\right)} = e^{-i\Delta t\ \mathcal{H}(j)}.$$

Note the standard formula:

$$\left.\frac{\mathrm{d}}{\mathrm{d}x} e^{(A+xB)}\right|_{x=0} = e^A \int_0^1 e^{A\tau} \beta e^{-A\tau}\, \mathrm{d}\tau$$

Let $A = -i\Delta t\ \mathcal{H}(j)$, $B = -i\Delta t\ \mathcal{H}_k$, and $x = \delta u_k(j)$, then

$$\begin{aligned}
\delta U_j &= \delta u_k(j) e^{-i\Delta t\ \mathcal{H}(j)} \int_0^1 e^{-i\Delta t\ \mathcal{H}(j)\tau} (-i\Delta t\ \mathcal{H}_k) e^{i\Delta t\ \mathcal{H}(j)\tau} \frac{\mathrm{d}(\,\tau\Delta t)}{\Delta t} \\
&= U_j \delta u_k(j)\,(-i\Delta t) \int_0^{\Delta t} U_j(\eta) \mathcal{H}_k U_j(-\eta) \frac{\mathrm{d}\eta}{\Delta t} \\
&= -i\Delta t\ \delta u_k(j)\ \bar{\mathcal{H}}_k U_j
\end{aligned}$$

where $\eta \equiv \tau\Delta t$ and $\bar{\mathcal{H}}_k = \int_0^{\Delta t} U_j(\eta)\ \mathcal{H}_k U_j(-\eta)\, \mathrm{d}\eta\ /\Delta t$ when $\Delta t \ll \left\|\mathcal{H}_o^\dagger \sum u_k(j)\mathcal{H}_k\right\|^{-1}$ and Δt is small. Since $\bar{\mathcal{H}}_k \approx \mathcal{H}_k$, therefore $\delta U_j \approx -i\Delta t\ \delta u_k(j)\ \mathcal{H}_k U_j$. Recall that $\Phi_o = \mathrm{Tr}\left\{\lambda_j^\dagger \rho_j\right\}$ to calculate $\frac{\delta\Phi_o}{\delta u_k(j)}$ to first-order in Δt.

$$\delta\Phi_o = \mathrm{Tr}\left\{\lambda_j^\dagger \delta\rho_j\right\}$$

Recall that $\rho_j = U_j \cdots U_1 \rho_o U_1^\dagger \cdots U_j^\dagger$. If $\delta u_k(j)$ is varied, then $U_j \to U_j^\dagger + \delta U_j$. So:

$$\begin{aligned}\rho_j + \delta\rho_j &= (U_j + \delta U_j)U_{j-1}\cdots U_1\rho_o U_1^\dagger \cdots U_{j-1}^\dagger(U_j^\dagger + \delta U_j^\dagger) \\ &= \rho_j + \delta U_j \cdots U_1\rho_o U_1^\dagger \cdots U_j^\dagger + U_j \cdots U_1\rho_o U_1^\dagger \cdots \delta U_j^\dagger + \underbrace{O\left((\delta U_j)^2\right)}_{O(\Delta t^2)}\end{aligned}$$

to first order in Δt,

$$\delta\rho_j = \delta U_j \cdots U_1\rho_o U_1^\dagger \cdots U_j^\dagger + U_j \cdots U_1\rho_o U_1^\dagger \cdots \delta U_j^\dagger.$$

Since $\delta U_j \approx -i\Delta t \delta u_k(j)\mathcal{H}_k U_j$, then $\delta\rho_j = -i\Delta t[\mathcal{H}_k, \rho_j] = -i\Delta t(\mathcal{H}_k\rho_j - \rho_j\mathcal{H}_k)$. Therefore,

$$\frac{\delta\Phi_o}{\delta u_k(j)} = -i\Delta t \,\mathrm{Tr}\left\{\lambda_j^\dagger\left[\mathcal{H}_F, \rho_j\right]\right\}$$

From the classical gradient method discussed above, the performance function can be increased by choosing an updated $u_k(j)$ such that

$$u_k(j) \longrightarrow u_k(j) + \epsilon\frac{\delta\Phi_o}{\delta u_k(j)}$$

which indicates the correct direction in which to improve Φ_o, where ϵ is a small constant. This idea forms the basic tenet of the GRAPE algorithm:

1. Guess initial control amplitudes $u_k(j)$.
2. Calculate $\rho_j = U_j \cdots U_1\rho_o U_1^\dagger \cdots U_j^\dagger$ for all $j \leq N$.
3. Calculate $\lambda_j = U_{j+1}^\dagger \cdots U_N^\dagger \rho_c U_N \cdots U_{j+1}$ for all $j \leq N$.
4. Evaluate $\frac{\delta\Phi_o}{\delta u_k(j)}$ and update the $m \times N$ control amplitudes
$$u_k(j) \longrightarrow u_k(j) + \epsilon\frac{\delta\Phi_o}{\delta u_k(j)}.$$
5. Repeat steps 2 to 4 until the change in Φ_o is smaller than a chosen threshold value.

3.3 Efficiency of GRAPE algorithm compared to "brute force" difference methods

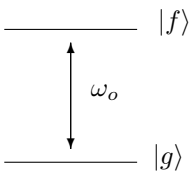
The GRAPE algorithm gives an *analytical* expression for finding the gradient (*i.e.* the direction in which to move in u-space in order to improve Φ). How would it be done otherwise? A *brute force* method in which each u_k is randomly varied and the resulting $\delta\Phi$ is calculated for each instance. On average, this method would require $m \times N + 1$ evaluations of the time evolution of the density

matrix from $t = 0$ to $t = T$ simply in order to calculate *one* set of gradients $\frac{\delta\Phi}{\delta u_k(j)}$; a large number of iterations is additionally required to reach the set of u_k that optimize Φ. With GRAPE, calculating the gradient $\frac{\delta\Phi}{\delta u_k(j)}$ only requires two full time evolutions to be evaluated (*i.e.* $\rho_o \longrightarrow \rho(t)$ and $\lambda_N(T) \longrightarrow \lambda_N(0)$). Therefore, GRAPE is $O(m \times N)$ more efficient.

While considerably better than the brute force method, note that GRAPE still requires computational resources that scale *exponentially* with the number of qubits (or coupled spins, etc.) If it were somehow more efficient than this, then arbitrary quantum dynamics can be efficiently simulated on a classical computer — but all indications are that this is not possible!

3.4 Single-photon absorption

Single-photon absorption:



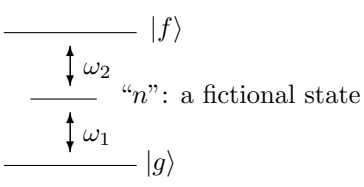
Amplitude of the excited state starting from $|g\rangle$:

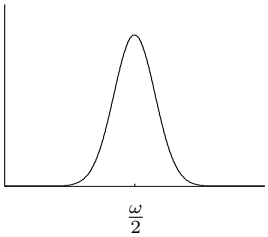
$$a_f(t) = \frac{\mu_{fg}}{i\hbar} \int_{-\infty}^{t} \epsilon(t_1) e^{i\omega_o t_1} \, \mathrm{d}t_1$$

where μ_{fg} is the dipole moment matrix element and $\hbar\omega_o = E_f - E_g$. The above is simply the Fourier component of the pulse, $\epsilon(t)$, at the resonance frequency, ω_o. The probability of a single-photon transition only depends on the energy content of the on-resonance frequency component — the phases or amplitudes of other components does not matter!

3.5 Two-photon absorption

Two-photon absorption:





The two-photon absorption peak is centered at $\omega_o/2$, but contains a distribution of frequency components as well.

Amplitude of the excited state:

$$a_f(t) = -\frac{1}{\hbar^2}\sum_n \mu_{fn}\mu_{ng}\int_{-\infty}^{t_2}\int_{-\infty}^{t_1}\epsilon(t_1)\epsilon(t_2)\overbrace{e^{i\omega_{fn}t_1}}^{(E_n-E_f)/\hbar}\, e^{i\omega_{ng}t_2}\,\mathrm{d}t_1\,\mathrm{d}t_2.$$

With some approximation, the transition probability is

$$P_{g\to f}^{\text{2-photon}} \approx \frac{1}{\hbar^4}\frac{\langle f|\,\mu^2\,|g\rangle}{\bar{\omega}}\left|\int_{-\infty}^{\infty}\epsilon^2(t)e^{i\omega_o t}\,\mathrm{d}t\right|^2.$$

The key property is:

$$P_{f\to g} \propto \left|\int_{-\infty}^{\infty}\epsilon^2(t)e^{i\omega_o t}\,\mathrm{d}t\right|^2.$$

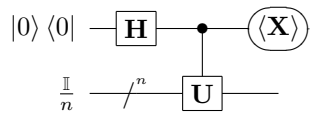
This integral can be rewritten in the spectral (frequency) domain, using $\tilde{\epsilon}(\omega) = A(\omega)e^{i\Phi(\omega)}$, which is the Fourier transform of $\epsilon(t)$, $A(\omega)$, which is the spectral amplitude, and $\Phi(\omega)$, which is the spectral phase.

$$\begin{aligned}\left|\int_{-\infty}^{\infty}\epsilon^2(t)e^{i\omega_o t}\,\mathrm{d}t\right|^2 &= \left|\int_{-\infty}^{\infty}\tilde{\epsilon}\left(\frac{\omega_o}{2}+\Omega\right)\tilde{\epsilon}\left(\frac{\omega_o}{2}-\Omega\right)\,\mathrm{d}\Omega\right|^2\\ &= \left|\int_{-\infty}^{\infty}A\left(\frac{\omega_o}{2}+\Omega\right)A\left(\frac{\omega_o}{2}-\Omega\right)e^{i\left(\Phi\left(\frac{\omega_o}{2}+\Omega\right)+\Phi\left(\frac{\omega_o}{2}-\Omega\right)\right)}\,\mathrm{d}\Omega\right|^2\end{aligned}$$

Two-photon transitions occur for all pairs of photons with frequencies ω_i and ω_j such that $\omega_i + \omega_j = \omega_o$ when ω_i and ω_j lie within the spectrum of the pulse. By manipulating the amplitude and phase information of the pulse, the two-photon absorption probability can be maximized or minimized. Similar ideas can be extended to exciting target molecular excitations — constructively interfering a certain excitation (*e.g.* motional mode) while destructively interfering with others.

3.6 Deterministic quantum computing with one bit

Deterministic quantum computing with one bit (DQC1) uses one *pure* bit and n *mixed* bits.



$$\begin{aligned}\rho_{in} &= |0\rangle\langle 0| \otimes \frac{\mathbb{I}}{n} \\ &\to \frac{1}{2}\big(|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| + |1\rangle\langle 1|\big) \otimes \frac{\mathbb{I}}{n} \\ &\to \frac{1}{2}\big(|0\rangle\langle 0|\mathbb{I} + |0\rangle\langle 1|\mathbb{I}\mathbf{U}^\dagger + |1\rangle\langle 0|\mathbf{U}\mathbb{I} + |1\rangle\langle 1|\mathbf{U}\mathbb{I}\mathbf{U}^\dagger\big)/n = \rho_f\end{aligned}$$

The measurement is performed on the pure qubit, which *traces over* all the mixed qubits.

$$\begin{aligned}\rho &= \mathrm{Tr}_n\{\rho_f\} \\ &= \frac{1}{2}\big(|0\rangle\langle 0| + |0\rangle\langle 1|\,\mathrm{Tr}(\mathbf{U}^\dagger) + |1\rangle\langle 0|\,\mathrm{Tr}(\mathbf{U}) + |1\rangle\langle 1|\big) \\ &= \frac{1}{2}\begin{pmatrix} 1 & \mathrm{Tr}(\mathbf{U}^\dagger) \\ \mathrm{Tr}(\mathbf{U}) & 1 \end{pmatrix}\end{aligned}$$

Let $\mathrm{Tr}(\mathbf{U}) = a + ib$ where $a = \mathrm{Re}\big(\mathrm{Tr}(\mathbf{U})\big)$
$\mathrm{Tr}(\mathbf{U}^\dagger) = (\mathrm{Tr}(\mathbf{U}))^\dagger = a - ib$ $b = \mathrm{Im}\big(\mathrm{Tr}(\mathbf{U})\big)$

$$\rho = \frac{1}{2}\begin{pmatrix} 1 & a-ib \\ a+ib & 1 \end{pmatrix} = \frac{1}{2}(\mathbb{I} + a\mathbf{X} + b\mathbf{Y})$$

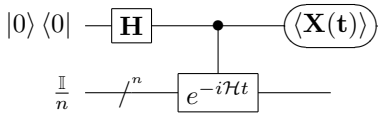
The measured expectation value of $\mathbf{X}$ on a pure qubit:

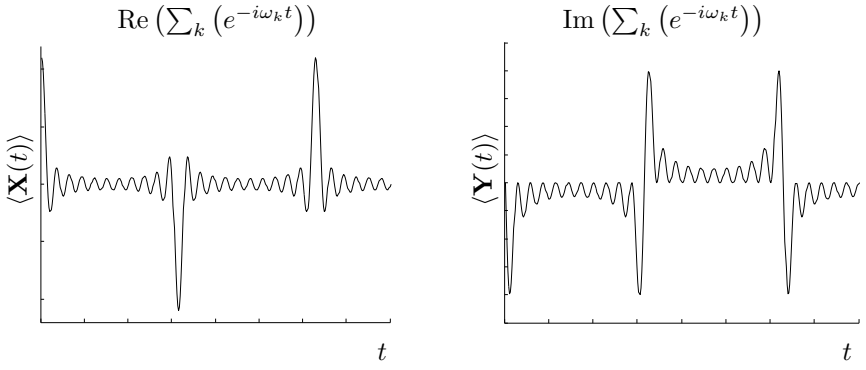
$$\langle \mathbf{X}\rangle = \mathrm{Tr}(\rho\mathbf{X}) = a = \mathrm{Re}\big(\mathrm{Tr}(\mathbf{U})\big)$$

$$\langle \mathbf{Y}\rangle = \mathrm{Tr}(\rho\mathbf{Y}) = b = \mathrm{Im}\big(\mathrm{Tr}(\mathbf{U})\big)$$

DQC1 can be applied to obtain the spectrum of a Hamiltonian. Let $\mathbf{U}(t) = e^{-i\mathcal{H}t}$. Assume that $\mathcal{H}$ is diagonal, so that $\mathbf{U} = \begin{pmatrix} e^{-i\omega_1 t} & & & \\ & e^{-i\omega_2 t} & & \\ & & \ddots & \\ & & & e^{-i\omega_N t} \end{pmatrix}$ where ω_k are the energy eigenvalues of $\mathcal{H}$. (Recall the spectral decomposition theorem, where $\mathcal{H} = \sum_k \alpha_k |k\rangle\langle k|$ in some basis $|k\rangle$.)

Now, the DQC1 algorithm can be run as a function of t:





Fourier transform of the energy spectrum of $\mathcal{H}$