INTRODUCTION TO QUANTUM INFORMATION SCIENCE

Quantum phenomena provide computing and information handling paradigms that are distinctly different and arguably much more powerful than their classical counterparts. In the past quarter of the century, much progress has been made on the theoretical side, and experiments have been carried out in which quantum computational operations were executed on a very small number of quantum bits. The NSF has declared this general area to be one of the 10 big ideas for future investments. In June 2018, the science committee of the House of Representatives unanimously approved the National Quantum Initiative Act (H.R. 6227), which would create a 10-year federal effort aimed at boosting quantum science.

This new course will provide an introduction to the theory of quantum computing and information. The topics that will be covered include 1) the fundamental elements of quantum information processing (qubits, unitary transformations, density matrices, measurements); 2) entanglement, protocols for teleportation, the Bell inequality, 3) basic quantum algorithms such as Shor’s factoring and Grover’s search, and 4) basic quantum data compression and error correction. The course material will be accessible to undergraduate and graduate students with a variety of backgrounds, e.g., electrical engineers, physicists, mathematicians, and computer scientists.

Learning Objective:
The students will learn the fundamentals of quantum information science, as well as a selected number of more advanced topics of their individual interests.

Instructor: Emina Soljanin emina.soljanin@rutgers.edu, CoRE 511, 848-445-5256.

Office hours: TBD & by appointment,

Class time and place: M&W, 3:20 PM – 4:40 PM, SEC 207

Prerequisites: Calculus, linear algebra, and probability at an undergraduate level as well as familiarity with complex numbers are required. Prior exposure to quantum mechanics and information/coding theory is helpful but not essential.

Grading: homework 20%, 2 midterm exams 20% each, final exam or project 40%.
(Midterm exams will be in class, approximately late September and late October.)

(See attached table of contents. Some other text covering the same topics might be used instead.)

Recommended reading:
L. Susskind and A. Friedman, Quantum Mechanics: The Theoretical Minimum.

Course notes: given per week in separate documents on the class Sakai page.

Registration: Undergrad students can request an SPN through a waiting list at http://www.ece.rutgers.edu → Academics → Undergraduate
Non-ECE graduate students should contact the instructor.

Remarks: The topics outlined above are very common for a quantum information science course at the advanced-undergraduate/graduate level. Such courses have been taught at several universities for many years, e.g., for almost two decades at Cornell based on the proposed textbook. There are plans for a companion course at RU Physics in the Spring 2019, which will look at some qubit technologies (e.g., superconducting, topological, semiconductor) as well as how quantum information can be applied to quantum gravity of black holes and tensor networks. The goal of these two courses is to together teach basic theory and possible realizations of quantum computing as well as some advanced topics and connections with other disciplines.
Introduction to Quantum Information Science\textsuperscript{1}

Prof. Emina Soljanin

Lecture #1, September 5

This lecture\textsuperscript{2} informally introduces several important notions: information, measurements, algorithms. It also talks about some misconceptions we may have about the importance of prior knowledge, and shows that there are surprises even in the classical world.

Information and Measurement

Bit is a unit of information that we get when we ask a yes/no question – yes or no, true or false, on or off, 0 or 1. Suppose you want to find out the position of the black king (that can be equally likely anywhere) on a chessboard. Take a look at Fig. 1. What is the minimum number of yes/no questions you need to ask?

To represent a bit in a computer, we need a physical entity which can exist in two distinguishable physical states. For example, magnetized cells in hard disk drives could be oriented in two different directions: “up” (0) or “down” (1). Flesh memory cells made from floating-gate transistors act as switches that could be open (0) or closed (1). (There are multi-level cell devices that can store more than one bit per cell.)

A physical system with $N = 2^b$ distinguishable physical states can represent $b$ bits of information. To specify an object in a set of $N$, we need $\lceil \log_2 N \rceil$ binary digits; Table 1 show how for $N = 8$.

\begin{center}
\begin{tabular}{|c|c|c|c|}
\hline
Decimal & Binary & mod 2 & Parity \\
\hline
0 & 000 & 0 & 0 \\
1 & 001 & 1 & 1 \\
2 & 010 & 0 & 1 \\
3 & 011 & 1 & 0 \\
4 & 100 & 0 & 1 \\
5 & 101 & 1 & 0 \\
6 & 110 & 0 & 0 \\
7 & 111 & 1 & 1 \\
\hline
\end{tabular}
\end{center}

Note that distinguishable is the crucial word here. Distinguishable how? By the naked eye? By a given measuring apparatus? Is there some fundamental limit to the number of states that can be distinguished by a physical measurement regardless of whether we can build it or not? If your measuring apparatus can only tell you the last digits of the numbers in Table 1, you will get only a single bit of information telling you whether the number is even or odd (see the third

\textsuperscript{1} Rutgers, ECE 579, Fall 2018

\textsuperscript{2} You are likely to find this class more conceptually than technically hard. In that sense, this is the hardest lecture.

Figure 1: What is the minimum number of yes/no questions that have to be asked to locate the king on a chessboard?

Table 1: We can specify each object in a set of 8 by assigning to it a unique label, e.g., a decimal number or a binary string.
column of Table 1). You will get a single but different bit of information if your measuring apparatus can only tell you the parity, namely, the XOR of the digits in the binary string representing the number (see the fourth column of Table 1).

**Algorithms**

A Penny Weighing Problem: You are given a balance scale and 8 pennies, one of which has a different weight. What is the minimum number of measurements that will always let you determine which penny has a different weight? How will you perform the measurements?

The minimum number of measurements that will always let us determine which penny has a different weight is three. Why? A possible way to perform the three measurements is given in Table 2. The three rows starting with M1, M2, and M3 correspond to the three measurements. The table entry at the intersection between a column corresponding to a penny and a row corresponding to a measurement indicates whether the penny is put on the scale in that measurement (0 if it is not) and if yes, whether it is placed on the left platform (L) or on the right platform R.

<table>
<thead>
<tr>
<th>PENNY</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>L</td>
<td>L</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>M2</td>
<td>o</td>
<td>o</td>
<td>L</td>
<td>L</td>
<td>o</td>
<td>o</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>M3</td>
<td>o</td>
<td>L</td>
<td>o</td>
<td>L</td>
<td>o</td>
<td>R</td>
<td>o</td>
<td>R</td>
</tr>
</tbody>
</table>

Suppose that the penny 4 has different weight, then measurement M1 will result in an unbalanced state of the scale and M2 and M3 in the balanced state of the scale, as illustrated in Fig. 3.

Observe that since there is only one penny of different weight, a measurement will result in an unbalanced state of the scale if the penny of different weight is placed on the scale in that measurement. Therefore, the possible measurement outcomes are as given in Table 3.
In each measurement, the scale can be either balanced (0) or unbalanced (1). Not that for each of the 8 “different penny” possibilities, we have a different set of measurement outcomes. Therefore a set of measurement outcomes uniquely identifies a different penny.

<table>
<thead>
<tr>
<th>DIFFERENT PENNY</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>M2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>M3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 3: Scale states corresponding to measurements for each of the 8 “different penny” possibilities. The scale can be either balanced (0) or unbalanced (1).

### Some Observations

1. We have committed to the way we perform the three measurements before the measuring process started. We do not adapt our measuring actions based on the results of the previous measurement, e.g., how we perform M2 does not change based on the outcome of M1.

2. Having some additional information could be helpful in designing a set of measurements, even if it cannot reduce the number of measurements.

3. How we conduct measurements evidently depends on the kind of scale we have. And so does the number of measurements. What would you do if you had a scale which has the unit weight corresponding to a regular penny fixed to the right tray, as in Fig. 4, and you can only use the left tray to place pennies?

### Problems

1. Consider the “king on the chessboard” problem. Why was it important to know that the king can be equally likely anywhere?

2. Suppose you have a balance scale as in Fig. 2. Find a set of 3 measurements that you can use to identify the different penny only if you know that it is heavier (or lighter) than the other seven.  
   Hint: Consider adaptive measurements.

3. Suppose you have a fixed weight scale as in Fig. 4. How many measurements would you need on average to find the single penny that does not have the unit weight?  
   \footnote{We will see such “measuring scales” when we study the Grover’s quantum search algorithm.}

\footnote{Non-adaptive measuring can be as powerful as adaptive.}
This lecture 1) introduces the notion of single qubit, as the quantum computing counterpart to the (classical) bit, and 2) provides mathematical background necessary to define states of multiple qubits and explain how they can be transformed.

**Bits as Mathematical Objects**

In this class, we will treat bits as mathematical objects. For us, bits take values in the set \(\{0, 1\}\) where we can add and multiply as follows:

\[
\begin{array}{c|cc}
& 0 & 1 \\
\hline
0 & 0 & 0 \\
1 & 1 & 0
\end{array}
\quad
\begin{array}{c|cc}
& 0 & 1 \\
\hline
0 & 0 & 0 \\
1 & 0 & 1
\end{array}
\]

Associative and distributive laws for binary addition and multiplication are identical to those for real numbers. Strings of \(n\) bits are mathematical objects that live in the field \(\mathbb{F}_{2^n}\), which is a set of \(2^n\) elements with specially defined addition and multiplication.

Recall our penny weighing problem, and consider the following matrix/vector multiplication, where the arithmetic is done as defined in Fig. 1.

\[
\begin{bmatrix}
0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
0 \\
0 \\
0 \\
1 \\
0 \\
0 \\
0
\end{bmatrix}
= \begin{bmatrix}
0 \\
1 \\
1 \\
1 \\
1 \\
1 \\
1
\end{bmatrix}
\]

Check to see that this classical computing is performing the measurements we did with a balance scale in the penny weighing problem. The result of computing is the binary representation of the number indicating the position (in this case 3) of the different bit. Multiplying (measuring) the vector with all bits equal to 1 except the one at position 3, would give the same result.

After the measuring, in order make all bits identical, we would have to flip the bit at position 3. This flipping can be performed by e.g., 1) a NOT operation on the bit in position 3 or 2) an XOR operation (binary
addition in Fig. 1) on the bit n position 3 together with the bit of the fixed value 1. Note that both these operations are reversible.

**Qubit as a Mathematical Object**

A *qubit* is a quantum information/computing counterpart to a bit. We will treat qubits as mathematical objects as well. What we learn in this class is independent of a particular physical realization.

A qubit is represented by a unit-norm vector in a two dimensional complex vector space. If we denote the basis vectors of this space by |
|\text{0}i\rangle = [\begin{array}{c}1 \\ 0\end{array}] and |
|\text{1}i\rangle = [\begin{array}{c}0 \\ 1\end{array}]

then a single qubit |\psi\rangle is mathematically a linear combination of |\text{0}i\rangle and |\text{1}i\rangle basis vectors:

|\psi\rangle = \alpha |\text{0}i\rangle + \beta |\text{1}i\rangle

where \(\alpha\) and \(\beta\) are complex numbers such that \(|\alpha|^2 + |\beta|^2 = 1\).

In classical computing, we refer to a *bit value* or a binary value. In quantum computing, we refer to a *qubit state* or a *quantum state*. Thus we could say that the quantum state |\psi\rangle above is a superposition of the two basis states.

**Math Interlude**

Quantum theory is a mathematical model of the physical world. We will go over necessary mathematics as the need arises. To understand the terms and the notation we used in describing the qubit, we next review some basic algebraic notions.

**Fundamental Structures in Abstract Algebra**

**Group** \((G, \circ)\) A group is a set \(G\) together with an operation \(\circ: G \times G \rightarrow G\) satisfying:

1. \(\circ\) is associative: \((a \circ b) \circ c = a \circ (b \circ c)\)

2. There is an element \(e\) in \(G\) s.t. \(a \circ e = a\) and \(e \circ a = a\) for every element \(a\) in \(G\). \(e\) is called neutral element.

3. For every element \(a\) in \(G\), there is an element \(a^{-1}\) in \(G\) s.t. \(a \circ a^{-1} = a^{-1} \circ a = e\). \(a^{-1}\) is called the inverse of \(a\).

If \(\circ\) is commutative, we say that \(G\) is commutative or Abelian.

Depending on the context, we will call a group 1) *additive*, its operation \(+\) addition, and its neutral element \(0\), or 2) *multiplicative*, its operation \(*\) or \(\cdot\) multiplication, and its neutral element \(1\) (unity).

Two examples:

1. \(((0, 1), \oplus)\) is an additive group.
2. \(((0, 1), \cdot)\) is not a group.
The most basic of the two-operation structures is called a ring: Ring is a set $A$ with operations called addition $+$ and multiplication $*$ satisfying:

1. $(A, +)$ is an Abelian group.
2. Multiplication is associative.
3. Multiplication is distributive over addition. That is, for all $a, b, c$ in $A$, we have $a(b + c) = ab + ac$

When the multiplication operation is commutative, we say that $A$ is a commutative (Abelian) ring.

A linear space over a field $F$ is an additive Abelian group $V$ together with an operation of multiplication by scalars $F \times V \rightarrow V$. The elements of $V$ are called vectors and the elements of $F$ are called scalars. The product of $\alpha \in F$ and $v \in V$ is denoted by $\alpha v \in V$. In addition, there are requirements connecting $F$ and $V$:

1. $(\alpha \beta) v = \alpha (\beta v)$
2. $\alpha(v + w) = \alpha v + \alpha w$
3. $(\alpha + \beta)v = \alpha v + \beta v$
4. $1v = v$, where $1$ is the unity in $F$

A linear combination of vectors $v_1, \ldots, v_m$ is a vector of the form

$$\alpha_1v_1 + \cdots + \alpha_m v_m.$$ 

The span of vectors $v_1, \ldots, v_m$ is the set of all linear combinations of $v_1, \ldots, v_m$:

$$\text{span}(v_1, \ldots, v_m) = \{ \alpha_1v_1 + \cdots + \alpha_m v_m \mid \alpha_1, \ldots, \alpha_m \in F \}.$$ 

Vectors $v_1, \ldots, v_m$ are linearly independent when

$$\alpha_1v_1 + \cdots + \alpha_m v_m = 0 \text{ only if } \alpha_1 = \cdots = \alpha_m = 0.$$ 

A basis of a vector space $V$ is a set of linearly independent vectors in $V$ that spans $V$. 

Examples of rings:
1. set of integers $\mathbb{Z}$
2. set of $n \times n$ matrices over $\mathbb{Z}$
3. Natural numbers $\mathbb{N}$ is not a ring.

Examples of fields:
1. set of rational numbers $(\mathbb{Q}, +, \cdot)$
2. set of complex numbers $(\mathbb{C}, +, \cdot)$
3. finite field $\{0, 1\}$,

Examples of linear spaces over the field of complex numbers $\mathbb{C}$:
1. $\mathbb{C}^n$ of $n$-tuples over $\mathbb{C}$
2. $\mathbb{C}^{k \times n}$ of $k \times n$ matrices over $\mathbb{C}$
3. All polynomials over $\mathbb{C}$
The dimension of a vector space $v$ is the number of vectors of a basis of $v$ over $\mathbb{F}$.

Let $V$ and $W$ be linear spaces over the same field. Then $f : V \rightarrow W$ is a linear map if for every $v, u \in V$ and $\alpha \in \mathbb{F}$, we have

1. $f(v + u) = f(v) + f(u) \quad \leftarrow \text{additive}$
2. $f(\alpha v) = \alpha f(v) \quad \leftarrow \text{homogeneous}$

**Hilbert Space**

An inner-product space is a vector space equipped with an inner product. An inner product in a complex vector space is a scalar-valued function of the ordered pair of vectors $v$ and $\varphi$, such that

1. $\langle \psi | \varphi \rangle = \langle \varphi | \psi \rangle^*$
2. $(\alpha \psi + \beta \xi | \varphi \rangle = \alpha \langle \psi | \varphi \rangle + \beta \langle \xi | \varphi \rangle$, where $\alpha, \beta \in \mathbb{C}$.
3. $\langle \psi | \psi \rangle \geq 0$ for any $v$ and $\langle \psi | \psi \rangle = 0$ iff $v$ is the 0 vector.

The quantity $\langle \psi | \psi \rangle^{1/2} = ||\psi||$ is often referred to as the norm or the length of the vector $v$.

A complex inner-product space is called a unitary space. Quantum computing deals with vectors and matrices in finite dimensional unitary spaces. The mathematical setting of quantum mechanics is the infinite dimensional generalization of unitary spaces, known as the Hilbert space. Thus we say that a qubit is an element of a two dimensional Hilbert space.

**Dirac’s Notation**

It is important to adopt a notation which let us easily distinguish between scalars and vectors. In mathematics, we usually use lower case letters for scalars and often capitals or bold face for vectors. The notation for vectors used in quantum computing literature (and preferred by physicists in general) is known as the Dirac’s or bra-ket notation.

In the bra-ket notation, a column vector is denoted by $| \varphi \rangle$ and its complex conjugate transpose by $\langle \psi |$. The bra-ket notation is inspired by the standard mathematical notation for the inner product

$$\langle \psi | \varphi \rangle = \langle \psi | \cdot | \varphi \rangle,$$

where $\cdot$ denotes ordinary matrix multiplication. Here a row vector times a column vector gives a number. Bras and kets can be multiplied as matrices also as

$$|\psi\rangle \langle \varphi|$$

Here a column vector times a row vector gives a matrix.
This lecture is concerned with multiple Qubits and reversible actions on single and multiple Qubits.

**Reversible Acting on a Single Qubit**

Recall that single Qubit is a vector in the 2D Hilbert space $\mathcal{H}_2$:

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

where $|0\rangle$ and $|1\rangle$ are the basis vectors of $\mathcal{H}_2$:

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

A quantum state can be transformed to another state only by a physical process consistent with the laws of quantum mechanics. In quantum mechanics, it is possible to act on quantum states by reversible and irreversible operations. The reversible operators will be called *gates* and irreversible will be called *measurements*.

In a closed quantum system, a single-Qubit state $|\psi\rangle \in \mathcal{H}_2$ can be transformed to some other state in $\mathcal{H}_2$, say $|\varphi\rangle$, in a reversible way only by some *unitary* operator $U$, i.e.,

$$|\varphi\rangle = U |\psi\rangle$$

where $U$ is a $2 \times 2$ unitary matrix. Any unitary matrix specifies a valid quantum gate.

If we know how $U$ acts on the basis vectors $|0\rangle$ and $|1\rangle$, then we also know how it acts on any vector $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$. To see that, recall that matrix multiplication is a linear operation:

$$U |\psi\rangle = \alpha U |0\rangle + \beta U |1\rangle.$$

**Math Interlude**

A unitary matrix $U$ is a complex *square* matrix whose inverse is equal to its conjugate transpose $U^\dagger$, i.e.,

$$U^\dagger U = U U^\dagger = I.$$ 

$U^\dagger$ is called the *adjoint* of $U$. Real unitary matrices are called *orthogonal*.

If only $U^\dagger U = I$, we say that $U$ is an isometry.
Some Single-Qubit Gates

- Identity: \( I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \)

- Hadamard gate:
  \[
  H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}
  \]

- Pauli matrices:
  \[
  \sigma_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}
  \]

Math Interlude

Let \( A \) be an \( m \times n \) matrix and \( B \) a \( p \times q \) matrix. The Kronecker product \( A \otimes B \) is the \( mp \times nq \) matrix given by

\[
A \otimes B = \begin{bmatrix}
  a_{11}B & a_{12}B & \cdots & a_{1n}B \\
  a_{21}B & a_{22}B & & a_{2n}B \\
  & & \ddots & \\
  a_{m1}B & a_{m2}B & & a_{mn}B
\end{bmatrix}.
\]

Some properties of the Kronecker product:

- Let \( A \) and \( C \) be \( n \times n \) matrices and \( B \) and \( D \) \( m \times m \) matrices. Then
  \[
  (A \otimes B) \cdot (C \otimes D) = AC \otimes BD.
  \]

- (Conjugate) transposition are distributive over the Kronecker product (cf. regular matrix product):
  \[
  (A \otimes B)^\dagger = A^\dagger \otimes B^\dagger
  \]
• A \otimes B \text{ has the inverse iff both } A \text{ and } B \text{ are invertible, and then (cf. regular matrix product):}

\[(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}\]

**Multiple Qubits**

As in classical computing, we mostly operate jointly on multiple Qubits rather than deal with a single Qubit on an individual basis.\(^5\) If we have two Qbits, one in the state \(|\psi_1\rangle = \alpha_1|0\rangle + \beta_1|1\rangle\) and the other \(|\psi_2\rangle = \alpha_2|0\rangle + \beta_2|1\rangle\), then the state of the pair is the Kronecker product of the individual states:

\[|\psi_1\rangle \otimes |\psi_2\rangle = (\alpha_1|0\rangle + \beta_1|1\rangle) \otimes (\alpha_2|0\rangle + \beta_2|1\rangle)\]

\[= \alpha_1\alpha_2|0\rangle \otimes |0\rangle + \alpha_1\beta_2|0\rangle \otimes |1\rangle + \beta_1\alpha_2|1\rangle \otimes |0\rangle + \beta_1\beta_2|1\rangle \otimes |1\rangle\]

where

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

In general, a 2-Qubit state is any superposition of these 4 basis states, and thus cannot always be expressed as a product of single Qubit states. 2-Qubit states that can be written as a Kronecker product of two single-Qubit states are called *separable* and those than cannot are called *entangled*\(^7\) states.

The individual Qubits that make up an entangled state cannot always be characterized as having individual states of their own. To see this, consider the following two-Qubit state

\[|\psi\rangle = \frac{1}{\sqrt{2}} (|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle).\]

This state is known as the Bell state or the EPR pair.\(^8\)

A system of \(n\) Qubits is a vector in \(\mathcal{H}_{2^n}\):

\[
\sum_{i=0}^{2^n-1} \alpha_i |i_0\rangle \otimes |i_1\rangle \otimes \cdots \otimes |i_{n-1}\rangle,
\]

where

• \(\mathcal{H}_{2^n} = \mathcal{H}_2 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_2\)
• \(\alpha_i \in \mathbb{C}\), \(\sum_{i=0}^{2^n-1} |\alpha_i|^2 = 1\)
• \(i_0, i_1, \ldots, i_{n-1}\) is the binary representation of \(i\)

\(^5\) How are \(n\) Qubits mathematically represented?

\(^6\) separable and entangled states

\(^7\) Entangled states are responsible for much of “quantum magic”.

\(^8\) EPR stands for Einstein, Podolsky and Rosen, who were the first to point out the “strange” properties of this state.
Several shorthand notations are used for the $i$-th basis vector of $\mathcal{H}_{2n}$:

$$
|0\rangle \otimes |i_1\rangle \otimes \cdots \otimes |i_{n-1}\rangle \sim |0\rangle |i_1\rangle \cdots |i_{n-1}\rangle \\
\sim |i_0, i_1, \ldots, i_{n-1}\rangle \\
\sim |i_0 i_1 \cdots i_{n-1}\rangle
$$

If an $n$-Qubit state can be expressed as a Kronecker product of $n$ single-Qubit states, we say that it is separable. Otherwise, we say that it is entangled.

There is a notion of Qudit, as a basic quantum state corresponding to a $d$-level physical systems. A single Qudit state is a vector in the $d$-dimensional Hilbert space $\mathcal{H}_d$, and an $n$-Qudit state is a vector in $\mathcal{H}_d^n$. Generalization from Qubit to Qudit systems is mathematically straightforward. Infinite dimensional systems will be left for later studies.

In general, if quantum (multi-Qubit) system $A$ is in the state $|\Psi\rangle_A$ in the Hilbert space $\mathcal{H}_A$ and if quantum (multi-Qubit) system $B$ is in the state $|\Phi\rangle_B$ in the Hilbert space $\mathcal{H}_B$, then the composite system’s state is the product $|\Psi\rangle_A \otimes |\Phi\rangle_B$ in the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$.

**Reversible Acting on $n$ Qubits**

In a closed quantum system, an $n$-Qubit state $|\Psi\rangle \in \mathcal{H}_{2^n}$ can be transformed to some other state in $\mathcal{H}_{2^n}$, say $|\Phi\rangle$, in a reversible way only by some unitary operator $U$, i.e.,

$$
|\Phi\rangle = U |\Psi\rangle
$$

where $U$ is a $2^n \times 2^n$ unitary matrix. As in the single-Qubit case, a unitary action on any $n$-Qubit state is completely described by its actions on the basis states of $\mathcal{H}_{2^n}$.

The following 2-Qubit gate is known as the controlled NOT (CNOT) or quantum XOR:

$$
\text{CNOT} : |x, y\rangle \rightarrow |x, x \oplus y\rangle \quad |x\rangle \quad |x\rangle \\
\quad |y\rangle \quad |x \oplus y\rangle
$$

A $2^n \times 2^n$ unitary matrix $U$ can be a Kronecker product of matrices of smaller dimensions (or not). When $U = U_0 \otimes U_1 \otimes \cdots \otimes U_{n-1}$, then its action on the basis vector of $\mathcal{H}_{2^n}$: $|i_0\rangle \otimes |i_1\rangle \otimes \cdots \otimes |i_{n-1}\rangle \in \mathcal{H}_{2^n}$ is given by

$$
U|i_0i_1\cdots i_{n-1}\rangle = [U_0|i_0\rangle \otimes U_1|i_1\rangle \otimes \cdots \otimes U_{n-1}|i_{n-1}\rangle]
$$

By restricting attention to collections of 2-state systems (or even $d$-state systems for finite $d$) one can avoid much suffering. Of course one also loses much wisdom, but hardly any of it — at least at this stage of the art — is relevant to the basic theory of quantum computation.

David Mermin

*Quantum Computer Science: An Introduction.* Cambridge Univ. Press.
Math Interlude

Projection Matrices

Recall $|\varphi\rangle\langle \varphi|$ is a matrix. Vector $|\varphi\rangle\langle \varphi| \cdot |\psi\rangle$ is the orthogonal projection of vector $|\psi\rangle$ on vector $|\varphi\rangle$. We say that $|\psi\rangle\langle \varphi|$ is a rank-1 projection matrix. (Higher rank projection matrices project vectors onto subspaces.) We have

$$|0\rangle\langle 0| + |1\rangle\langle 1| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = I$$

also known as a resolution of the identity. Let $|u_1\rangle, \ldots, |u_m\rangle$ be the columns of the unitary matrix $U$. Then $U^\dagger U = UU^\dagger = I_m$ implies

1. $\langle u_i|u_j\rangle = \delta_{ij}$
2. $|u_1\rangle\langle u_1| + |u_2\rangle\langle u_2| + \cdots + |u_m\rangle\langle u_m| = I_m$

Eigenvalues and Eigenvalues

An eigenvector of a complex $m \times m$ matrix $M$ is a vector $|v\rangle$ such that

$$M|u\rangle = \lambda_u |u\rangle, \quad |u\rangle \neq 0, \quad \lambda_u \in \mathbb{C}$$

where $\lambda_u$ is known as the eigenvalue of $M$ corresponding to $|u\rangle$.

Hermitian Matrices

A Hermitian matrix $M$ (or self-adjoint matrix) is a complex square matrix that is equal to its own conjugate transpose $M^\dagger$, i.e., the element in the $i$-th row and $j$-th column is equal to the complex conjugate of the element in the $j$-th row and $i$-th column, for all indices $i$ and $j$:

$$h_{ij} = h_{ji}^*.$$ 

We call real Hermitian matrices symmetric.

Claim: A matrix $M$ is Hermitian if and only if $\langle x|M|x\rangle$ is real for all $|x\rangle$.

If follows that the eigenvalues of a Hermitian operator are real.
Hermitian and unitary matrices are normal. If $A$ is normal, then its eigenvectors corresponding to distinct eigenvalues are orthogonal. For a Hermitian matrix $M$, there exists a unitary matrix $U$ such that $U^\dagger MU$ is a diagonal matrix:

$$U^\dagger MU = \begin{bmatrix}
\lambda_1 & & \\
& \lambda_2 & \\
& & \ddots \\
& & & \lambda_{m-1} \\
& & & & \lambda_m
\end{bmatrix}$$

Let $|u_1\rangle, \ldots, |u_m\rangle$ be the columns of $U$, and multiply the above equation by $U$ from the left. \Rightarrow

$$[M |u_1\rangle \ldots M |u_m\rangle] = [\lambda_1 |u_1\rangle \ldots \lambda_m |u_m\rangle]$$

and thus $|u_1\rangle, \ldots, |u_m\rangle$ are eigenvectors of $M$ and $\lambda_1, \ldots, \lambda_m$ are the corresponding eigenvalues. Recall that $|u_1\rangle, \ldots, |u_m\rangle$ form a basis of $\mathbb{C}^m$ and thus a resolution of the identity.$^6$

$$|u_1\rangle\langle u_1| + |u_2\rangle\langle u_2| + \cdots + |u_m\rangle\langle u_m| = I_m$$

$^6$There can be more or less projections than $m$ in a resolution of $I_m$

Pauli Matrices

These matrices were introduced in the early days of quantum mechanics by Wolfgang Pauli, to describe the angular momentum associated with the spin of an electron. They often appear in both physics and mathematics for various purposes. It is interesting that they are both unitary and Hermitian, and thus can serve as quantum gates, and (as we will soon see) to define quantum measurements.

<table>
<thead>
<tr>
<th>matrix</th>
<th>action</th>
<th>eigenvalue/eigenvector</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_X = \begin{bmatrix} 0 &amp; 1 \ 1 &amp; 0 \end{bmatrix}$</td>
<td>$</td>
<td>0\rangle \rightarrow [X]</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>1\rangle \rightarrow [X]</td>
</tr>
<tr>
<td>$\sigma_Y = \begin{bmatrix} 0 &amp; -i \ i &amp; 0 \end{bmatrix}$</td>
<td>$</td>
<td>0\rangle \rightarrow [Y] i</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>1\rangle \rightarrow [Y] -i</td>
</tr>
<tr>
<td>$\sigma_Z = \begin{bmatrix} 1 &amp; 0 \ 0 &amp; -1 \end{bmatrix}$</td>
<td>$</td>
<td>0\rangle \rightarrow [Z]</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>1\rangle \rightarrow [Z] -</td>
</tr>
</tbody>
</table>
Quantum Measurements

To every physical observable, there corresponds an operator defined by a Hermitian matrix. The only possible results of measuring an observable are the eigenvalues of its corresponding Hermitian matrix. The only possible states after measuring an observable are the eigenvectors of its Hermitian matrix.

\[
\begin{align*}
|\psi\rangle & \rightarrow \lambda_i |\psi\rangle \langle \psi | u_i \rangle^2 \\
|u_1\rangle, \ldots, |u_m\rangle
\end{align*}
\]

Let \( |u_1\rangle, \ldots, |u_m\rangle \) be the eigenvectors of the Hermitian matrix corresponding to the observable. We also refer to \( |u_1\rangle, \ldots, |u_m\rangle \) as the measurement basis. Then, after the measurement is performed on state \(|\psi\rangle\), it gets projected (collapses) to state \(|u_i\rangle\) with probability \((wp) |\langle \psi | u_i \rangle|^2\), \(1 \leq i \leq m\).

Example: What can we get if we measure Qubit \(\frac{1}{\sqrt{2}} (|0\rangle + |1\rangle)\) in the computational basis\(^7\) \(|0\rangle, |1\rangle\)?

Remark: The number of outcomes of a quantum measurement is finite. How many bits of information does a measurement provide?

How Much Classical Information is in a Qubit?

To describe a Qubit, say \(|\psi\rangle = \alpha |0\rangle + \beta |1\rangle\), in a given basis, one needs to specify two complex numbers \(\alpha\) and \(\beta\). That may require a very large number of bits (depending on the chosen precision), in general, infinite.

Suppose you have acquired a Qubit. Do you possess an infinite amount of information? You would if you could read out the values of \(\alpha\) and/or \(\beta\). Is there a quantum measurement that would allow you to do that? The answer is no. Can quantum computers be more powerful than classical computers?

Error Correcting Codes

Consider the following scaled down version of the penny weighing problem we solved in our first class: How would you use a balance scale to determine which of the 3 pennies has a different weight if any? We will next use some of that reasoning to discuss the basic

\(\text{Figure 1: Quantum Measurement: When we "see" } \lambda_i \text{ (which happens } wp |\langle \psi | u_i \rangle|^2\text{), we know that state } |\psi\rangle \text{ has collapsed to } |u_i\rangle\).

\(\text{Figure 2: If in a 20-faced die, we can only discern if the number has one or two digits, then rolling the die is equivalent to tossing a slightly biased coin.}\)
principles in classical and quantum error correction. In both systems we will be concerned with *bit flips*.

Error correcting codes add redundancy to data in order to be less sensitive to errors. The most basic form of redundancy is simple replication, known as repetition coding. For example, if each bit is replicated 3 times, any single bit flip among the 3 replicas can be corrected by turning it to the value of the other two replicas.

We will next formally describe the process of introducing redundancy (encoding) and correcting errors (decoding) for a 1-to-3 bits repetition code, which will allow us to introduce and understand its quantum 1-to-3 Qubit counterpart, and (later in the course) study more general quantum error correction.

### Classical Error Correction

- **Encoding** is a map that introduces redundancy. In our 1-to-3 bits repetition code example, encoding is the following map:

  \[
  0 \rightarrow 000 \quad \text{and} \quad 1 \rightarrow 111
  \]

  Therefore, each bit \( x \) is mapped to a 3 bit string \( xxx \).

- **Error Model**: In our model, at most one of the bits \( xxx \) gets flipped. Such flipping is equivalent to adding (component-wise) a string in the set \{000, 100, 010, 001\} to \( xxx \) and getting \( y_0 y_1 y_2 \):

  \[
  \begin{array}{c|ccc}
  \text{error} & y_0 & y_1 & y_2 \\
  \hline
  000 & x & x & x \\
  100 & x \oplus 1 & x & x \\
  010 & x & x \oplus 1 & x \\
  001 & x & x & x \oplus 1 \\
  \end{array}
  \]

- **Measurements**: In a classical system, we perform the following matrix vector multiplication\(^6\):

  \[
  \begin{bmatrix}
  1 & 1 & 0 \\
  1 & 0 & 1 \\
  \end{bmatrix}
  \begin{bmatrix}
  y_0 \\
  y_1 \\
  y_2 \\
  \end{bmatrix}
  =
  \begin{bmatrix}
  y_0 \oplus y_1 \\
  y_0 \oplus y_2 \\
  \end{bmatrix}
  \]

- **Error Correction**: The results of the two measurements (2 bits) instruct us how to correct errors, as follows:

  \[
  \begin{array}{ccc|cc|c}
  y_0 & y_1 & y_2 & y_0 \oplus y_1 & y_0 \oplus y_2 & \text{add} \\
  \hline
  x & x & x & 0 & 0 & 000 \\
  x \oplus 1 & x & x & 1 & 1 & 100 \\
  x & x \oplus 1 & x & 1 & 0 & 010 \\
  x & x & x \oplus 1 & 0 & 1 & 001 \\
  \end{array}
  \]

  \(^6\)cf. Lesson # 1, penny weighing

**Correcting errors might sound like a dreary practical problem, of little aesthetic or conceptual interest. But aside from being of crucial importance for the feasibility of quantum computation, it is also one of the most beautiful and surprising parts of the subject.**

David Mermin

Quantum Computer Science: An Introduction. Cambridge Univ. Press.
The No-Cloning Theorem

There is no reversible quantum operator that takes any state $|\psi\rangle$ to $|\psi\rangle \otimes |\psi\rangle$.

**Proof.** Suppose there is a unitary $U_c$ such that

$$
U_c(|\psi\rangle \otimes |\omega\rangle) = |\psi\rangle \otimes |\psi\rangle
$$

$$
U_c(|\varphi\rangle \otimes |\omega\rangle) = |\varphi\rangle \otimes |\varphi\rangle
$$

where $\omega$ is some fixed quantum state. Note the following identities:

1. By the properties of the Kronecker product, we have

$$
\langle \psi \otimes \langle \omega \rangle \cdot (|\varphi\rangle \otimes |\omega\rangle) = \langle \psi | \varphi \rangle
$$

2. Since $U_c$ is unitary, by the properties of the Kronecker product, we have

$$
\langle \psi | \varphi \rangle = (\langle \psi | \langle \omega \rangle \cdot (|\varphi\rangle \otimes |\omega\rangle))
$$

$$
= (\langle \psi | \langle \omega \rangle \cdot |\varphi\rangle) \cdot U_c \cdot \langle \omega | \varphi \rangle
$$

$$
= \langle \psi | \varphi \rangle \otimes \langle \psi | \varphi \rangle
$$

$$
= (\langle \psi | \varphi \rangle)^2
$$

Therefore $\langle \psi | \varphi \rangle$ is either equal to 0 or to 1. Thus if $U_c$ can clone some state $|\psi\rangle$, then the only other state $U_c$ can clone has to be orthogonal to $|\psi\rangle$.

**Problems – Homework due on September 19**

1. Show that if $U$ and $V$ are unitary matrices, then $U \otimes V$ is also a unitary matrix.

2. Show that the map $|x, y\rangle \rightarrow |x, x \otimes y\rangle$ for $x, y \in \{0, 1\}$ can be achieved by the following unitary matrix:

$$
U_{\text{CNOT}} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0
\end{bmatrix}
$$

3. Construct a quantum gate that performs the following map:

$$
\underbrace{\alpha|\psi\rangle + \beta|1\rangle}_{\in \mathcal{H}_2} \rightarrow \underbrace{\alpha|000\rangle + \beta|111\rangle}_{\in \mathcal{H}_{23}}
$$

You are allowed to use additional fixed-state quantum systems.
Error Correcting Codes – Review

We start by reviewing our classical error correcting code example and Pauli matrices.

Classical Error Correction

- **Encoding** is a map that introduces redundancy. In our 1-to-3 bits repetition code example, encoding is the following map:

\[
0 \rightarrow 000 \quad \text{and} \quad 1 \rightarrow 111
\]

Therefore, each bit \(x\) is mapped to a 3 bit string \(xxx\).

- **Error Model**: In our model, at most one of the bits \(xxx\) gets flipped. Such flipping is equivalent to adding (component-wise) a string in the set \(\{000, 100, 010, 001\}\) to \(xxx\) and getting \(y_0 y_1 y_2\):

<table>
<thead>
<tr>
<th>error</th>
<th>(y_0)</th>
<th>(y_1)</th>
<th>(y_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
</tr>
<tr>
<td>100</td>
<td>(x \oplus 1)</td>
<td>(x)</td>
<td>(x)</td>
</tr>
<tr>
<td>010</td>
<td>(x)</td>
<td>(x \oplus 1)</td>
<td>(x)</td>
</tr>
<tr>
<td>001</td>
<td>(x)</td>
<td>(x)</td>
<td>(x \oplus 1)</td>
</tr>
</tbody>
</table>

- **Measurements**: In a classical system, we perform the following matrix vector multiplication:

\[
\begin{bmatrix}
1 & 1 & 0 \\
1 & 0 & 1 \\
\end{bmatrix} \cdot \begin{bmatrix}
y_0 \\
y_1 \\
y_2 \\
\end{bmatrix} = \begin{bmatrix}
y_0 \oplus y_1 \\
y_0 \oplus y_2 \\
\end{bmatrix}
\]

We refer to the vector \(\begin{bmatrix} y_0 \oplus y_1 \\ y_0 \oplus y_2 \end{bmatrix}\) as the *error syndrome*.

- **Error Correction**: The results of the two measurement (2 bits) instruct us how to correct errors, as follows:

<table>
<thead>
<tr>
<th>(y_0)</th>
<th>(y_1)</th>
<th>(y_2)</th>
<th>(y_0 \oplus y_1)</th>
<th>(y_0 \oplus y_2)</th>
<th>add</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x)</td>
<td>(x)</td>
<td>(x)</td>
<td>0</td>
<td>0</td>
<td>000</td>
</tr>
<tr>
<td>(x \oplus 1)</td>
<td>(x)</td>
<td>(x)</td>
<td>1</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>(x)</td>
<td>(x \oplus 1)</td>
<td>(x)</td>
<td>1</td>
<td>0</td>
<td>010</td>
</tr>
<tr>
<td>(x)</td>
<td>(x)</td>
<td>(x \oplus 1)</td>
<td>0</td>
<td>1</td>
<td>001</td>
</tr>
</tbody>
</table>
Pauli Matrices – Review

These matrices were introduced in the early days of quantum mechanics by Wolfgang Pauli, to describe the angular momentum associated with the spin of an electron. They often appear in both physics and mathematics for various purposes. It is interesting that they are both unitary and Hermitian, and thus can serve as quantum gates, and (as we will soon see) to define quantum measurements.

\[
\begin{array}{c|c|c}
\text{matrix} & \text{action} & \text{eigenvalue/eigenvector} \\
\hline
\sigma_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & |0\rangle -|X\rangle |1\rangle & +1/(|0\rangle + |1\rangle) \\
& |1\rangle -|X\rangle |0\rangle & -1/(|0\rangle - |1\rangle) \\
\sigma_Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} & |0\rangle -|Y\rangle i|1\rangle & +1/(|0\rangle + i|1\rangle) \\
& |1\rangle -|Y\rangle -i|0\rangle & -1/(|0\rangle - i|1\rangle) \\
\sigma_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} & |0\rangle -|Z\rangle |0\rangle & +1/|0\rangle \\
& |1\rangle -|Z\rangle |1\rangle & -1/|1\rangle \\
\end{array}
\]

A Little More General Quantum Measurements

A simple generalization of the quantum measurement we defined in the previous lecture is known as the von Neumann’s measurement. \(^3\) Also known as the projective measurement. Mathematically, this type of measurement is defined by a set of \(m \times m\) matrices \(\{\Pi_i\}_{i=1}^\ell\) such that

1. \(\{\Pi_i\}\) are pairwise orthogonal projection operators
2. \(\{\Pi_i\}\) form a complete resolution of the identity, that is,
   \[\Pi_1 + \Pi_2 + \cdots + \Pi_\ell = I_m.\]
3. The measured state \(|\psi\rangle\) gets projected (collapses) to state \(\Pi_i |\psi\rangle\) with probability \(\langle\psi|\Pi_i |\psi\rangle\).

Note that the only difference between this and our previous definition is that here we do not require that \(\Pi_i\) be rank-1 projections. This generalization comes about when we remove the requirement that all eigenvalues of the Hermitian matrix corresponding to the measured observable be distinct.
Suppose we know that a quantum system is in one of \( \ell \) possible states, say \( |\psi_1\rangle, \ldots, |\psi_\ell\rangle \). Only when the possible states are orthogonal can a quantum measurement be designed to give an unambiguous answer about the state of the system. Therefore the set of errors we can tell apart are those which take the original state to a set of orthogonal states.

**A Quantum Error Correcting Code**

Quantum error correction\(^4\) has to follow the laws of quantum mechanics. Therefore each action on Qubits has to be either unitary or a measurement.

- **Encoding**: As in the classical case, encoding is a map that introduces redundancy. In our example, a single Qubit state is mapped into a 3-Qubit state as follows:

\[
\alpha |0\rangle + \beta |1\rangle \rightarrow \alpha |000\rangle + \beta |111\rangle
\]

The following unitary circuit can serve as a quantum mechanically valid encoder for our code. It uses two CNOT gates and two ancillary Qubits, each initially in the state \( |0\rangle \).

The result is an entangled 3-Qubit state.

- **Error Model**: In our model, at most one Qubit experiences the basis flip. This flipping errors result from unitary error operators as follows:

<table>
<thead>
<tr>
<th>error operators</th>
<th>resulting state</th>
</tr>
</thead>
<tbody>
<tr>
<td>I \otimes I \otimes I</td>
<td>\alpha</td>
</tr>
<tr>
<td>\sigma_X \otimes I \otimes I</td>
<td>\alpha</td>
</tr>
<tr>
<td>I \otimes \sigma_X \otimes I</td>
<td>\alpha</td>
</tr>
</tbody>
</table>

- **Measurements**: We perform the following two measurements:

\[ M_1 : \text{Defined by the Hermitian operator } \sigma_Z \otimes \sigma_Z \otimes I, \text{ i.e., the following two orthogonal projection operators} : \]

\[
\Pi_1 = |000\rangle \langle 000| + |111\rangle \langle 111| + |001\rangle \langle 001| + |110\rangle \langle 110|
\]

\[
\Pi_2 = |010\rangle \langle 010| + |101\rangle \langle 101| + |011\rangle \langle 011| + |100\rangle \langle 100|
\]

\(^4\text{As significant as Shor’s factoring algorithm may prove to be, there is another recently discovered feature of quantum information that may be just as important: the discovery of quantum error correction. Indeed, were it not for this development, the prospects for quantum computing technology would not seem bright.}\]

John Preskill
Quantum Computation Lecture Notes.
Chapter 1, 1997/98.
\( \Pi_1 \) projects on the eigenspace of \( \sigma_Z \otimes \sigma_Z \otimes \mathbb{I} \) with eigenvalue 1, and \( \Pi_2 \) projects on the eigenspace of \( \sigma_Z \otimes \sigma_Z \otimes \mathbb{I} \) with eigenvalue -1.

**M_2**: Defined by the Hermitian operator \( \sigma_Z \otimes \mathbb{I} \otimes \sigma_Z \), i.e., the following two orthogonal projection operators:

\[
\Pi_1 = |000\rangle\langle 000| + |111\rangle\langle 111| + |010\rangle\langle 010| + |101\rangle\langle 101|
\]
\[
\Pi_2 = |001\rangle\langle 001| + |110\rangle\langle 110| + |011\rangle\langle 011| + |100\rangle\langle 100|
\]

\( \Pi_1 \) projects on the eigenspace of \( \sigma_Z \otimes \mathbb{I} \otimes \sigma_Z \) with eigenvalue 1, and \( \Pi_2 \) projects on the eigenspace of \( \sigma_Z \otimes \mathbb{I} \otimes \sigma_Z \) with eigenvalue -1.

- **Error Correction**: The results of the two measurements are two eigenvalues (2 bits). As in the classical case, we refer to this result as the error syndrome, which instructs us how to correct errors, as follows:

<table>
<thead>
<tr>
<th>corrupted state</th>
<th>( M_1 )</th>
<th>( M_2 )</th>
<th>apply</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha</td>
<td>000\rangle + \beta</td>
<td>111\rangle )</td>
<td>+1</td>
</tr>
<tr>
<td>( \alpha</td>
<td>100\rangle + \beta</td>
<td>011\rangle )</td>
<td>-1</td>
</tr>
<tr>
<td>( \alpha</td>
<td>010\rangle + \beta</td>
<td>101\rangle )</td>
<td>-1</td>
</tr>
<tr>
<td>( \alpha</td>
<td>001\rangle + \beta</td>
<td>110\rangle )</td>
<td>+1</td>
</tr>
</tbody>
</table>

**Remark**: The error detecting procedure we used (1) follows directly from classical error correction and (2) it is useful in generalizing to other quantum codes with more Qubits. However, \( M_1 \) and \( M_2 \) are not the only measurements we can use to obtain the error syndrome that can uniquely identify the error. To see that consider the following set of projections:

\( \Pi_1 = |000\rangle\langle 000| + |111\rangle\langle 111| \) no error
\( \Pi_2 = |100\rangle\langle 100| + |011\rangle\langle 011| \) bit flip on Qubit one
\( \Pi_3 = |010\rangle\langle 010| + |101\rangle\langle 101| \) bit flip on Qubit two
\( \Pi_4 = |001\rangle\langle 001| + |110\rangle\langle 110| \) bit flip on Qubit three

Note that the (no)-error states belong to orthogonal subspaces, and therefore a von Neumann measurement defined by projectors to those subspaces can (1) unambiguously identify the error state and (2) will not disturb the measured state.
Problems – Homework due on September 26

1. Prove that orthogonal projection operators are Hermitian.

2. You are given a set of vectors $|u_i\rangle \in \mathcal{H}_m$, $i = 1, 2, \ldots, m$ that form a resolution of the identity $\langle u_i|u_j\rangle = \delta_{ij}$. Construct a Hermitian matrix which has $|u_i\rangle$ as its eigenvectors corresponding to different eigenvalues.

3. You are given a set of $\ell$ orthogonal projection $m \times m$ matrices that form a resolution of the identity and are pairwise orthogonal on $\mathcal{H}_m$, where $\ell \leq m$. Construct a Hermitian matrix whose eigenspaces corresponding to different eigenvalues are the image spaces of these projection matrices.

4. Suppose you used the 1-to-3 Qubit code and there were bit flips on Qubit 1 and on Qubit 2. Which state would you decode if you followed the measurement and correction procedure we defined in class?
This lecture considers several important multi-Qubit gates, and the Deutsch-Jozsa algorithm.

Some Multi-Qubit Gates

Single-Qubit Hadamard Gate

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

$$|0\rangle \rightarrow \frac{H |0\rangle + |1\rangle}{\sqrt{2}}$$

$$|1\rangle \rightarrow \frac{H |1\rangle - |0\rangle}{\sqrt{2}}$$

n-Qubit Separable Hadamard Gate

$$H^\otimes n = H \otimes H \otimes \cdots \otimes H.$$  

How does $$H^\otimes n$$ act on the basis state $$|0\rangle^\otimes n$$? It is easy to see that $$H^\otimes n$$ creates a uniform superposition of all basis states.

We next look into how $$H^\otimes n$$ acts on an arbitrary base state $$|x\rangle$$ where $$|x\rangle = |x_0\rangle \otimes |x_1\rangle \otimes \cdots \otimes |x_{n-1}\rangle$$ and $$x_0x_1 \cdots x_{n-1}$$ is the binary representation of $$x$$:

$$H^\otimes n |x\rangle = H |x_0\rangle \otimes \cdots \otimes H |x_{n-1}\rangle$$

$$= \frac{1}{\sqrt{2^n}} (|0\rangle + (-1)^{x_0} |1\rangle) \otimes \cdots \otimes (|0\rangle + (-1)^{x_{n-1}} |1\rangle)$$

$$= \frac{1}{\sqrt{2^n}} \sum_{y=0}^{2^n-1} (-1)^{x \cdot y} |y\rangle$$

where $$x \cdot y = x_0y_0 \oplus x_1y_1 \oplus \cdots \oplus x_{n-1}y_{n-1}$$ is the mod 2 sum of the bitwise product.

2-Qubit Controlled NOT Gate

The following 2-Qubit gate is known as the controlled NOT (CNOT) or quantum XOR:

$$\text{CNOT} : |x, y\rangle \rightarrow |x, x \oplus y\rangle$$

$$|x\rangle \rightarrow |x\rangle$$

$$|y\rangle \rightarrow |x \oplus y\rangle$$
**Function Evaluation**

We can evaluate an m-bit valued function $f$ of an n-bit string $x$:

$$ f : \{0, 1\}^n \to \{0, 1\}^m $$

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

Why is $U_f$ a valid quantum gate?

**Quantum Parallelism**

If we first create a superposition of all basis states by applying the $H^\otimes n$ gate to state $|0\rangle^\otimes n$, and then apply the $U_f$ gate, we can simultaneously compute the value of $f$ on its entire domain:

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

**The Deutsch Problem**

**Problem Statement**

In the Deutsch Problem, we are concerned with a binary function

$$ f : \{0, 1\} \to \{0, 1\} $$

and know that it is either constant (0 on the entire domain or 1 on the entire domain) or balanced (1 for half of the domain and 0 for the other half). The goal is to tell whether $f$ is constant by performing only one evaluation of the function.

**An Algorithm**

We begin with the two-qubit state $|0\rangle|1\rangle$ and apply a Hadamard transform to each qubit. This yields

$$ \frac{1}{2}(|0\rangle + |1\rangle)(|0\rangle - |1\rangle) = |0\rangle(|0\rangle - |1\rangle) + |1\rangle(|0\rangle - |1\rangle) $$

We are given a quantum implementation of the function $f$ that maps $|x\rangle|y\rangle$ to $|x\rangle|f(x) \oplus y\rangle$. Applying this function to our current state we obtain

$$ U_f \frac{1}{2}(|0\rangle + |1\rangle)(|0\rangle - |1\rangle) = $$

$$ = \frac{1}{2}(|0\rangle(|f(0) \oplus 0) - |f(0) \oplus 1\rangle) + |1\rangle(|f(1) \oplus 0) - |f(1) \oplus 1\rangle) $$

There are only 4 such functions. What are they?
Let \( \tilde{f}(x) = 1 \oplus f(x) \) denote the complement (NOT) of \( f(x) \). We further have

\[
U_{f_{1/2}}(\ket{0} + \ket{1})(\ket{0} - \ket{1}) = \begin{cases} \frac{1}{2}(\ket{0} + \ket{1})(\ket{f(0)} - \ket{\tilde{f}(0)}) & \text{if } f(0) = f(1) \\ \frac{1}{2}(\ket{0} - \ket{1})(\ket{f(0)} - \ket{\tilde{f}(0)}) & \text{if } f(0) \neq f(1) \end{cases}
\]

We can now perform a measurement of this 2-Qubit state according to the observable \( \sigma_X \otimes I \). Recall that \( \ket{0} + \ket{1} \) and \( \ket{0} - \ket{1} \) are eigenvectors of \( \sigma_X \) with respective eigenvalues 1 and -1.

### The Deutsch-Jozsa Problem

**Problem Statement**

The Deutsch-Jozsa problem is a generalization of the Deutsch problem in the sense that we are again asked to tell whether a binary function \( f \) is constant or balanced by performing only one evaluation of the function. But here, the domain of \( f \) is \( \{0, 1\}^n \):

\[
f : \{0, 1\}^n \rightarrow \{0, 1\}.
\]

The algorithm 1) begins with the \( (n + 1) \)-Qubit state \( \ket{0}^\otimes |1\rangle \), 2) creates a superposition by applying the Hadamard transform to each of the \( n + 1 \) Qubits, 3) Calculates \( f \) by passing the resulting \( n + 1 \)-Qubit state through the the \( U_f \) gate, 4) performs the Hadamard transform on the first \( n \) Qubits, the and 5) measures the final state in the computational basis to get an unambiguous answer whether the function is balanced or constant. Before we look into these steps in more detail, we make the following two observations:

1. For a binary function \( f \), we have

\[
|f(x)| - |1 \oplus f(x)| = (-1)^{f(x)}|x\rangle \begin{cases} \ket{0} - \ket{1} & \text{if } f(x) = 0 \\ \ket{1} - \ket{0} & \text{if } f(x) = 1 \end{cases} = (-1)^{f(x)}(\ket{0} - \ket{1})
\]

2. For a binary function \( f : \{0, 1\}^n \rightarrow \{0, 1\} \), we have

\[
\frac{1}{2^n} \sum_{x=0}^{2^n-1} (-1)^{f(x)} = \begin{cases} (-1)^{f(0)} & \text{if } f \text{ is constant} \\ 0 & \text{if } f \text{ is balanced} \end{cases}
\]
**The Algorithm**

1. Set up the initial $n + 1$ Qubit state to $|\psi_0\rangle = |0\rangle^\otimes n |1\rangle$

2. Create a superposition by using Hadamard gates to $|\psi_0\rangle$ obtain the state

$$|\psi_1\rangle = \left( H^\otimes n \otimes H \right) |0\rangle^\otimes n |1\rangle = \left( |0\rangle + |1\rangle \right)^\otimes n (|0\rangle - |1\rangle).$$

$$= \frac{1}{\sqrt{2^{n+1}}} \sum_{x=0}^{2^n-1} |x\rangle (|0\rangle - |1\rangle).$$

3. Calculate function $f$ using $U_f$ that maps $|x\rangle |y\rangle$ to $|x\rangle |y \oplus f(x)\rangle$:

$$U_f |\psi_1\rangle = \frac{1}{\sqrt{2^{n+1}}} \sum_{x=0}^{2^n-1} |x\rangle (|f(x)\rangle - |1 \oplus f(x)\rangle)$$

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

4. At this point the last Qubit may be ignored. We apply a Hadamard transform to each Qubit of $|\psi_2\rangle$ and obtain

$$|\psi_3\rangle = H^\otimes n |\psi_2\rangle = \frac{1}{2^n} \sum_{x=0}^{2^n-1} (-1)^f(x) \sum_{y=0}^{2^n-1} (-1)^{x \cdot y} |y\rangle$$

$$= \frac{1}{2^n} \sum_{y=0}^{2^n-1} \left( \sum_{x=0}^{2^n-1} (-1)^f(x) (-1)^{x \cdot y} \right) |y\rangle$$

$$= \left[ \frac{1}{2^n} \sum_{x=0}^{2^n-1} (-1)^f(x) \right] |0\rangle + \frac{1}{2^n} \sum_{y=1}^{2^n-1} \left( \sum_{x=0}^{2^n-1} (-1)^f(x) (-1)^{x \cdot y} \right) |y\rangle$$

5. We now can measure $|\psi_3\rangle$ in the computational basis. Note that the probability of measuring $|0\rangle^\otimes n$ is

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

which, as we have shown above, is equal to 1 if $f(x)$ is constant or to 0 if $f(x)$ is balanced. Therefore, if the output of the measurement is $|0\rangle^\otimes n$, then $f$ is constant; otherwise $f$ is balanced.
This lecture 1) explains how 2-Qubit entanglement can be created by elementary gates, and 2) describes two communication protocols, dense coding and teleportation, which exploit entanglement.

**Bell States**

Recall that 2-Qubit states that can be written as a Kronecker product of 2 single-Qubit states are called *separable* and those that cannot are called *entangled* states.

An entangled pair of states can be created by applying a unitary transform to separable states, e.g., as shown in Fig. 1.

\[
\begin{align*}
|0\rangle & \quad \begin{array}{c} \text{H} \\ \frac{1}{\sqrt{2}} \end{array} \left( |00\rangle + |11\rangle \right) \\
|0\rangle & \quad \begin{array}{c} \text{H} \\ \frac{1}{\sqrt{2}} \end{array} \left( |00\rangle - |11\rangle \right) \\
|1\rangle & \quad \begin{array}{c} \text{H} \\ \frac{1}{\sqrt{2}} \end{array} \left( |01\rangle + |10\rangle \right) \\
|1\rangle & \quad \begin{array}{c} \text{H} \\ \frac{1}{\sqrt{2}} \end{array} \left( |01\rangle - |10\rangle \right)
\end{align*}
\]

The 4 entangled states in Fig. 1 are known as Bell states. Notice that they are orthogonal, which should not be a surprise since they are created by a unitary transform from the 4 computational basis states. Therefore, Bell states can be used to define a measurement, which is often referred to as the Bell measurement.

Entangled states have some “surprising” properties. To see that, we consider the EPR pair:

\[
\frac{1}{\sqrt{2}} \left( |00\rangle + |11\rangle \right)
\]

and observe the following:

1. The individual Qubits that make up an entangled state cannot always be characterized as having individual states of their own. Consider, for example, the first Qubit, and observe that it cannot be represented in the form \( \alpha |0\rangle + \beta |1\rangle \).

2. There seems to be *spooky action at a distance*: What happens if we measure only the first Qubit in the computational basis? Two outcomes are possible: |0\rangle with probability \( 1/2 \), giving the post-measurement 2-Qubit state |00\rangle, and |1\rangle with probability \( 1/2 \), giving...
the post-measurement 2-Qubit state $|11\rangle$. What happens if we subsequently measure the other Qubit? Only one outcome is possible: the one that gives the same result as the measurement of the first Qubit. This behavior has been confirmed by experiment.

**Dense Coding**

If Alice sends a Qubit, say $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, to Bob, how many bits does he get? Recall that Bob cannot read the values of complex numbers $\alpha$ and/or $\beta$. He can only possibly apply some unitary transformation to $|\psi\rangle$ and then perform a measurement, which would give him at most one bit.

Suppose Alice and Bob had prepared together and entangled pair of Qubits in the state

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

and then Alice took Qubit A and Bob took Qubit B. How does the state $|\psi\rangle$ evolve if only Alice applies a unitary transformation to her Qubit? Consider the following 4 local unitary actions on the first Qubit:

$$(I \otimes I) |\psi\rangle = \frac{1}{\sqrt{2}} (|0_A 0_B\rangle + |1_A 1_B\rangle)$$

$$(\sigma_X \otimes I) |\psi\rangle = \frac{1}{\sqrt{2}} (|1_A 0_B\rangle + |0_A 1_B\rangle)$$

$$(\sigma_X \otimes I) |\psi\rangle = \frac{1}{\sqrt{2}} (|0_A 0_B\rangle - |1_A 1_B\rangle)$$

$$(\sigma_Z \otimes I) |\psi\rangle = \frac{1}{\sqrt{2}} (-|1_A 0_B\rangle + |0_A 1_B\rangle)$$

Note that Alice is able to create 4 orthogonal states. If after performing her local action, Alice sends her Qubit to Bob, he can unambiguously identify which of the 4 orthogonal Bell states the EPR pair assumed as a result of Alice’s action. He can therefore get two bits of information. Alice and Bob have to have agreed on how to label Alice’s actions, e.g.,

- 00 : $(I \otimes I)$
- 01 : $(\sigma_X \otimes I)$
- 10 : $(\sigma_Z \otimes I)$
- 11 : $(\sigma_Z \sigma_X \otimes I)$

For example, if Alice wants to send two classical bits 10 to Bob, she will apply $\sigma_Z$ to her Qubit before sending it to Bob. That would create the global state in Bob’s possession $\frac{1}{\sqrt{2}} (|0_A 0_B\rangle - |1_A 1_B\rangle)$, which he will learn after performing the Bell measurement.

$^6$Would Alice be able to create 4 orthogonal global states by local actions if the qubits were not entangled?
Teleportation

Suppose Alice and Bob had prepared together an entangled pair of Qubits in the state

$$|\Psi\rangle = \frac{1}{\sqrt{2}} (|0_A 0_B\rangle + |1_A 1_B\rangle)$$

and then Alice took Qubit A and Bob took Qubit B. Now, Alice has another Qubit in the state

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

which she would like to send to Bob. However, there is only a classical communications channel between her and Bob. Can Alice send her Qubit to Bob by sending only classical bits of information? How many classical bits does she need to send?

To answer that question, consider the joint state of Alice’s new Qubit and the entangled pair:

$$|\psi\rangle |\Psi\rangle = (\alpha |0\rangle_A + \beta |1\rangle_A) \frac{1}{\sqrt{2}} (|0_A 0_B\rangle + |1_A 1_B\rangle)$$

$$= \alpha |0\rangle_A \frac{1}{\sqrt{2}} (|0_A 0_B\rangle + |1_A 1_B\rangle) + \beta |1\rangle_A \frac{1}{\sqrt{2}} (|0_A 0_B\rangle + |1_A 1_B\rangle)$$

The following protocol, known as teleportation, results in Bob’s Qubit (member of the entangled pair) assuming the state $|\psi\rangle$.

1. Alice first applies a CNOT gate to her two Qubits

$$|\chi\rangle_A |x_A\rangle \rightarrow |x_a\rangle |x_a \oplus x_A\rangle$$

and the 3-Qubit state becomes

$$|\Phi\rangle = \alpha |0\rangle_A \frac{1}{\sqrt{2}} (|0_A 0_B\rangle + |1_A 1_B\rangle) + \beta |1\rangle_A \frac{1}{\sqrt{2}} (|1_A 0_B\rangle + |0_A 1_B\rangle)$$

2. Alice then applies a Hadamard transformation $H$ to her Qubit $a$, and the joint state becomes

$$\left( H \otimes I \otimes I \right) |\Phi\rangle = \frac{1}{\sqrt{2}} |0\rangle_A \left( \alpha |0\rangle_B + \beta |1\rangle_B \right) \frac{1}{\sqrt{2}} (|0_A 0_B\rangle + |1_A 1_B\rangle)$$

$$+ \frac{1}{\sqrt{2}} |1\rangle_A \left( \alpha |0\rangle_B - \beta |1\rangle_B \right) \frac{1}{\sqrt{2}} (|0_A 0_B\rangle + |1_A 1_B\rangle)$$

$$= \frac{1}{2} |00\rangle_A A \left( \alpha |0\rangle_B + \beta |1\rangle_B \right) + \frac{1}{2} |01\rangle_A A \left( \alpha |1\rangle_B + \beta |0\rangle_B \right)$$

$$+ \frac{1}{2} |10\rangle_A A \left( \alpha |0\rangle_B - \beta |1\rangle_B \right) + \frac{1}{2} |11\rangle_A A \left( \alpha |1\rangle_B - \beta |0\rangle_B \right)$$

Observe the following:

(a) The 4 states in the above sum are orthogonal.
(b) For each of the 4 basis states on Alice’s side, we have a corresponding state on Bob’s side that can be obtained from $|\psi\rangle$ by a unitary action:

\[
\begin{align*}
\alpha |0\rangle_B + \beta |1\rangle_B &= I |\psi\rangle \\
\alpha |1\rangle_B + \beta |0\rangle_B &= \sigma_X |\psi\rangle \\
\alpha |0\rangle_B - \beta |1\rangle_B &= \sigma_Z |\psi\rangle \\
\alpha |1\rangle_B - \beta |0\rangle_B &= \sigma_Z \sigma_X |\psi\rangle
\end{align*}
\]

3. Alice performs a joint measurement of her two Qubits in the computational basis. Her pair of Qubits will collapse to one of the basis states and Bob’s Qubit will assume its corresponding state. After the measurement, Alice knows which state she is left with and thus which state Bob’s Qubit is in. Bob can turn that state to $|\psi\rangle$ by applying the appropriate unitary operator. Whether that operator should be $I$, or $\sigma_X$ or $\sigma_Z$ or $\sigma_Z \sigma_X$ can be communicated to him by Alice with 2 bits of classical information. They have to have agreed on how to label the 4 operators.

Observe that there is only one copy of state $|\psi\rangle$ at the end of the protocol, the one that Bob has. Alice’s 2-Qubit state collapsed to a basis state after her measurement.

*by the entanglement magic

*Teleportation is not cloning.
This lecture introduces classical information sources.

Math Interlude

A sequence of random variables (RVs) $X_1, X_2, \ldots, X_n$ that are mutually independent and have the same distribution is called a sequence of independent trials or an independent trials process.

Example:

A Bernoulli trials process is a sequence of chance experiments s.t.

1. Each experiment has two possible outcomes, which we may call $\text{H} & \text{T}$ or in general, success & failure
2. The outcome of each experiment is independent of other outcomes.

$\Rightarrow$

The probability of success $p$ is

- the same for each experiment
- not affected by any knowledge of previous outcomes

Let $X_1, X_2, \ldots, X_n$ be an independent trials process with $E(X_i) = \mu < \infty$ and $V(X_i) = \sigma^2 < \infty$ (finite mean and variance).

Let $S_n = X_1 + X_2 + \cdots + X_n$.

The Weak Law of Large Numbers (WLLN) says that, for any $\epsilon > 0$,

$$\Pr \left( \left| \frac{S_n}{n} - \mu \right| \geq \epsilon \right) \to 0 \text{ as } n \to \infty$$

Jensen’s Inequality:\ For an RV $X$ with expectation $E(X)$ and a convex function $f$, we have

$$f \left( E[X] \right) \leq E \left[ f(X) \right].$$

A Classical Source of Information

A discrete memoryless source (DMS) of information produces a sequence of independent, identically distributed discrete random variables taking values in a finite set called the source alphabet. A DMS is therefore an independent trials process characterized by an RV $X$ and its probability distribution $P_X$. Each RV in the i.i.d. source sequence $X_1, \ldots, X_n$ is distributed as $X$. $^\dagger$
We also say that a DMS produces sequences of letters where each letter is drawn from the set $\mathcal{X}$ (source alphabet, domain of $X$) independently according to the probability distribution $P_X$. Thus a source sequence $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$ occurs with probability

$$P_X(x) = P_X(x_1) \cdots P_X(x_n),$$

where $P_X(x_i) = \Pr(X = x_i)$. Note that $x = (x_1, \ldots, x_n)$ is a particular realization of the sequence of random variables $X_1, \ldots, X_n$.

**Information Content and Shannon Entropy**

Suppose a fair coin is tossed and we are told that the head turned up. How much information did we get? Would the answer be the same if the coin was biased? What if an 8-faced fair die is rolled, and 4 turned up? What if our 8-faced die was so biased that it shows an even number wp $1/4 - \delta$ and an even number wp $\delta$ for some very small $\delta$.

To uniquely identify each face on such a die, we need $3$ bits. However, information contained in an event is defined so that it measures our surprise on learning that that event has happened.

We say that event $A$ with probability $\Pr(A)$ contains

$$I(A) := \log \frac{1}{\Pr(A)}$$

units of information. If the base of the logarithm is 2, the unit is the bit.

If the base of the logarithm is $e$, the unit is called the nat (for natural).

The information content is additive: If $E$ and $F$ are two independent events, then

$$I(A, B) = I(A) + I(B)$$

For an information source with alphabet $\mathcal{X}$ and the associated RV $X$ whose domain (sample space) is $\mathcal{X}$ and probability distribution is $P_X$. The information content associated with the letter $x$ in $\mathcal{X}$ is $-\log P_X(x)$.

The Shannon entropy of the information source (or equivalently RV $X$ or probability distribution $P_X$) is the expected information content of its letters:

$$H(X) = \sum_{x \in \mathcal{X}} -P_X(x) \log P_X(x)$$

When the source alphabet is binary with the probability of the two letters $p$ and $1 - p$, the Shannon entropy is known as the binary entropy, shown in Fig. 2. The binary entropy function attains its maximum value at $p = \frac{1}{2}$ (cf. unbiased coin flip).

Entropy can be seen as a measure of the expected uncertainty associated with an RV or a probability distribution. It should then be maximized by the uniform distribution among all distributions with equal domain sizes.
Information Theory and Other Fields

Figure 3: Relationship of information theory to other fields. Taken from Cover & Thomas Elements of Information Theory, Wiley 1991.
This lecture discusses (asymptotically) lossless classical source coding (data compression).

Math Interlude

The Weak Law of Large Numbers

Let $X_1, X_2, \ldots, X_n$ be an independent trials process with $E(X_j) = \mu < \infty$ and $V(X_j) = \sigma^2 < \infty$ (finite mean and variance).

Let $S_n = X_1 + X_2 + \cdots + X_n$.

The Weak Law of Large Numbers (WLLN) says that, for any $\varepsilon > 0$,

$$\Pr \left( \left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) \to 0 \text{ as } n \to \infty$$

A Function of a Random variable

If $X$ is an RV with the range $\Omega_X$ and $Y = g(X)$, then $Y$ is a random variable. We have the following:

- The range of $Y$ is $\Omega_Y = \{ g(x) \mid x \in \Omega_X \}$.
- The probability distribution of $Y$ is given by

$$P(Y = y) = P(g(X) = y) = \sum_{x : g(x) = y} P(X = x)$$

- The mean of $Y$ is given by

$$E[Y] = \sum_{y \in \Omega_Y} yP(Y = y) = \sum_{y \in \Omega_Y} y \cdot \sum_{x : g(x) = y} g(x)P(X = x)$$

$$= \sum_{x \in \Omega_X} g(x)P(X = x)$$

Example: Let $X$ be the RV corresponding to rolling a die, and $P^1_X$ and $P^2_X$ two probability distributions for $X$ given by

<table>
<thead>
<tr>
<th>$X$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P^1_X$</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
</tr>
<tr>
<td>$P^2_X$</td>
<td>1/3</td>
<td>1/6</td>
<td>1/12</td>
<td>1/6</td>
<td>1/6</td>
<td>1/12</td>
</tr>
</tbody>
</table>

Then $P^1_X$ as an RV that takes value $1/6$ wp 1, and $P^2_X$ is an RV that take values $1/3$ wp $1/3$, $1/6$ wp $1/2$ and $1/12$ wp $1/6$. \( \text{\textsuperscript{3}} \) An unusual example but we need to understand it.
Asymptotic Equipartition

Consider a sequence of RVs $X_1, \ldots, X_n$ produced by a DMS characterized by the RV $X$ whose probability distribution is $P_X$ and entropy is $H(X)$. The asymptotic equipartition property (AEP) is a theorem that states that the following is true:

$$\lim_{n \to \infty} \Pr \left[ \left| -\frac{1}{n} \log P_X(X_1, X_2, \ldots, X_n) - H(X) \right| > \varepsilon \right] = 0 \quad \forall \varepsilon > 0.$$  

The AEP is a direct consequence of the weak law of large numbers. To see that, consider the sequence of random variables $Y_1, \ldots, Y_n$ where $Y_i = -\log P_X(X_i)$. Note that the following holds:

1. $Y_i$ are i.i.d. and $E(Y_i) = H(X)$.
2. $-\log P_X(X_1, \ldots, X_n) = \log \prod_{i=1}^{n} P_X(X_i) = \sum_{i=1}^{n} Y_i$

The AEP follows from applying the WLLN to the process $Y_1, \ldots, Y_n$.

Weak Typicality

We say that a source sequence $x_1, \ldots, x_n$ is weakly $\varepsilon$-typical (aka entropy $\varepsilon$-typical) if

$$2^{-n(H(X)+\varepsilon)} \leq P_X(x_1, x_2, \ldots, x_n) \leq 2^{-n(H(X)-\varepsilon)}$$

The set of all such sequences is known as the typical set $A^{(n)}_\varepsilon$.

The AEP implies that, asymptotically, the probability of the typical set is large, while (unless $P_X$ is uniform), its size is small. More precisely, given any $\varepsilon > 0$, one can choose $n$ such that:

1. $\Pr[A^{(n)}_\varepsilon] \geq 1 - \varepsilon$
2. $(1 - \varepsilon)2^{n[H(X)-\varepsilon]} \leq |A^{(n)}_\varepsilon| \leq 2^{n[H(X)+\varepsilon]}$

Therefore, the fraction of sequences that are typical is

$$\frac{|A^{(n)}_\varepsilon|}{|X^n|} \leq \frac{2^{nH(X)+\varepsilon}}{2^{n\log_2 |X|}} = 2^{-n[\log_2 |X|-H(X)-\varepsilon]} \to 0$$

as $n \to \infty$ since $H(X) < \log_2 |X|$. 

Asymptotic Equipartition

Consider a sequence of RVs $X_1, \ldots, X_n$ produced by a DMS characterized by the RV $X$ whose probability distribution is $P_X$ and entropy is $H(X)$. The asymptotic equipartition property (AEP) is a theorem that states that the following is true:

$$\lim_{n \to \infty} \Pr \left[ \left| -\frac{1}{n} \log P_X(X_1, X_2, \ldots, X_n) - H(X) \right| > \varepsilon \right] = 0 \quad \forall \varepsilon > 0.$$  

The AEP is a direct consequence of the weak law of large numbers. To see that, consider the sequence of random variables $Y_1, \ldots, Y_n$ where $Y_i = -\log P_X(X_i)$. Note that the following holds:

1. $Y_i$ are i.i.d. and $E(Y_i) = H(X)$.
2. $-\log P_X(X_1, \ldots, X_n) = \log \prod_{i=1}^{n} P_X(X_i) = \sum_{i=1}^{n} Y_i$

The AEP follows from applying the WLLN to the process $Y_1, \ldots, Y_n$.

Weak Typicality

We say that a source sequence $x_1, \ldots, x_n$ is weakly $\varepsilon$-typical (aka entropy $\varepsilon$-typical) if

$$2^{-n(H(X)+\varepsilon)} \leq P_X(x_1, x_2, \ldots, x_n) \leq 2^{-n(H(X)-\varepsilon)}$$

The set of all such sequences is known as the typical set $A^{(n)}_\varepsilon$.

The AEP implies that, asymptotically, the probability of the typical set is large, while (unless $P_X$ is uniform), its size is small. More precisely, given any $\varepsilon > 0$, one can choose $n$ such that:

1. $\Pr[A^{(n)}_\varepsilon] \geq 1 - \varepsilon$
2. $(1 - \varepsilon)2^{n[H(X)-\varepsilon]} \leq |A^{(n)}_\varepsilon| \leq 2^{n[H(X)+\varepsilon]}$

Therefore, the fraction of sequences that are typical is

$$\frac{|A^{(n)}_\varepsilon|}{|X^n|} \leq \frac{2^{nH(X)+\varepsilon}}{2^{n\log_2 |X|}} = 2^{-n[\log_2 |X|-H(X)-\varepsilon]} \to 0$$

as $n \to \infty$ since $H(X) < \log_2 |X|$.
(Asymptotically) Lossless Source Coding

How many bits do we need to represent the \(2^{n \log_2 |X|}\) source sequences of length \(n\)? On average, we need only about \(nH(X)\) bits if we use e.g., one of the following two procedures of assigning strings of bits to the sequences of letters.\(^4\)

Enumeration

Here is an example of a lossless source encoding scheme:

1. **Partitioning**: Divide all sequences in \(X^n\) into two sets: the typical set \(A^{(n)}\) and its complement.

2. **Ordering**: Order the elements in each set according to some (e.g., lexicographic) order. Give each sequence an index corresponding to the order it has within its set.

3. **Labeling**: For each sequence in \(A^{(n)}\) use a \(0\) followed by \(\lceil n(H(X) + \epsilon) \rceil\) bits that correspond to its index. For each sequence in the complement of \(A^{(n)}\), use \(1\) followed by \(\lceil n \log |X| \rceil\) bits\(^5\) that correspond to its index. Here, the initial bit acts as a flag bit to indicate the length of the codeword that follows. The mapping from source sequences to bit strings is thus one-to-one and look-up table decodable.

Random Binning

Here is an example of an asymptotically lossless source coding scheme:

1. **Encoding**:\(^6\) For each sequence in \(X^n\), we draw an index at random with replacement from the set \(\{1, 2, \ldots, 2^{nR}\}\). This procedure is identical to randomly throwing sequences from \(X^n\) into \(2^{nR}\) bins labeled by \(\{1, 2, \ldots, 2^{nR}\}\). Note that the encoder does not need to know the typical set.

2. **Decoding**:\(^7\) Given an index (bin), we look for a typical source sequence in the bin. If there is one and only one typical sequence in the bin, we declare it to be the estimate of the source sequence; otherwise, we declare an error. Note that the decoder does need to know the typical set.

Observe that there are two ways that the decoder may declare an error for an index \(i\): 1) there is more than one typical sequence with the index \(i\), and 2) there is no typical sequence with the index \(i\). But if the number of bins is much larger than the number of typical sequences, the probability that there is more than one typical sequence in a bin is very small, and hence the probability that a typical sequence will result in an error is very small.

\(^4\) More efficient algorithms are used in practice, and real sources are seldom DMS.

\(^5\) \(\lceil n \log |X| \rceil\) bits are sufficient, but \(\lceil n \log |X| \rceil\) is good enough for efficient compression.

\(^6\) SOURCE SEQUENCE \(\rightarrow\) INDEX

\(^7\) INDEX \(\rightarrow\) SOURCE SEQUENCE
 Problems – Homework due Oct. 29

1. Let $\mathcal{X}$ be a finite set (alphabet) and $\mathcal{P}(\mathcal{X})$ be the set of all probability distributions on $\mathcal{X}$. Show that

$$\max_{P \in \mathcal{P}(\mathcal{X})} H(P) \leq \log|\mathcal{X}|$$

Hint: Use Jensen’s inequality.

2. Find the expected codeword length when the enumeration algorithm is used to encode source sequences.

3. Find the probability of decoding error when the binning (hashing) algorithm is used to encode source sequences.
Introduction to Quantum Information Science

Prof. Emina Soljanin
Lecture #10, October 22

This lecture introduces 1) the density matrix formalism of quantum mechanics and 2) the notion of quantum mixed states.

Math Interlude – The Trace of a Matrix

Let \( A \) be an \( n \times n \) complex matrix. The trace of \( A \) is defined to be the sum of the elements on the diagonal of \( A \):

\[
\text{tr}(A) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}
\]

Some properties of the trace:

1. The trace is invariant under cyclic permutations: \(^2\)

\[
\text{tr}(ABCD) = \text{tr}(BCDA) = \text{tr}(CDAB) = \text{tr}(DABC).
\]

We will often use this property of the trace.

2. The trace of the Kronecker product of two square matrices is the product of their traces:

\[
\text{tr}(X \otimes Y) = \text{tr}(X) \text{tr}(Y).
\]

3. The trace is a linear operator:

\[
\text{tr}(\alpha X + \beta Y) = \alpha \text{tr}(X) + \beta \text{tr}(Y).
\]

Here \( X \) and \( Y \) are square matrices and \( \alpha \) and \( \beta \) are scalars.

4. The trace is similarity-invariant. This property follows from the property 1. above:

\[
\text{tr} \left( P^{-1}AP \right) = \text{tr} \left( (P^{-1}AP) \right) = \text{tr} \left( (AP)P^{-1} \right) = \text{tr} \left( (PP^{-1})A \right) = \text{tr}(A).
\]

Therefore, the trace is invariant to the change of basis, and thus

\[
\text{tr}(A) = \sum_{i=1}^{n} \lambda_i, \quad \text{where } \lambda_i \text{ are the eigenvalues of } A.
\]

5. Trace in Dirac’s notation: \(^3\) Let \( |e_i\rangle, i = 1, \ldots, n \) be an orthonormal basis of \( \mathbb{C}^n \). Then \( |e_i\rangle\langle e_i|A \) is a matrix whose \( i \)-th diagonal element is \( a_{ii} \) and all other elements are 0. Therefore,

\[
\text{tr} A = \sum_{i=1}^{n} \text{tr}(|e_i\rangle\langle e_i|A) = \sum_{i=1}^{n} \langle e_i | A | e_i \rangle
\]

For the last equality, we have used the property 1. above.
**The Density Matrix Formalism**

To describe quantum states, so far we used vectors in Hilbert spaces. Can we instead describe a quantum state, say $|\psi\rangle$, by the matrix $|\psi\rangle\langle\psi|$? We will use the notation $\rho_\psi = |\psi\rangle\langle\psi|$ and refer to $\rho_\psi$ as the **density matrix** of state $|\psi\rangle$.

Any description of a state should allow us to describe 1) how a state evolves when a unitary transformation is applied to it and 2) what happens to a state and with what probability when a measurement is performed on it.

1. Suppose that unitary operator $U$ acts on state $|\psi\rangle$ giving the state $|\varphi\rangle = U|\psi\rangle$. We have $|\varphi\rangle\langle\varphi| = U|\psi\rangle\langle\psi|U^\dagger$. Therefore,

$$|\psi\rangle \xrightarrow{U} U|\psi\rangle \iff \rho_\psi \xrightarrow{U} U\rho_\psi U^\dagger$$

2. Suppose a measurement defined by the matrices $\{\Pi_i\}_{i=1}^k$ is performed on the state $|\varphi\rangle$. We know that the resulting state will be $|\varphi\rangle = \frac{\Pi_i|\psi\rangle}{||\Pi_i|\psi||}$ with probability $\langle\psi|\Pi_i|\psi\rangle$. Therefore,

$$\rho_\psi = \frac{\Pi_i|\psi\rangle\langle\psi|\Pi_i}{\langle\psi|\Pi_i|\psi\rangle} \text{ wp } \langle\psi|\Pi_i|\psi\rangle$$

or, in terms of density matrices,

$$\rho_\psi \rightarrow \frac{\Pi_i\rho_\psi\Pi_i}{\text{tr}(\Pi_i\rho_\psi)} \text{ wp } \text{tr}(\Pi_i\rho_\psi)$$

by observing that $\text{tr}(\Pi_i\rho_\psi) = \langle\Pi_i|\psi|\psi\rangle = \langle\psi|\Pi_i|\psi\rangle$.

**Mixed States**

**Density matrix**

The density matrix formalism allows us to compactly describe a quantum system about which we only know that it is in the state $|\psi_j\rangle$ with probability $p_j$ as follows:

$$\rho = \sum_j p_j|\psi_j\rangle\langle\psi_j|$$

We refer to such quantum systems as **mixed states**.

The states we worked with so far that can be described by a vector, say $|\psi\rangle$, or, equivalently, the corresponding rank-1 density matrix $\rho_\psi = |\psi\rangle\langle\psi|$ are known as **pure states**. In general, a density matrix $\rho$ is a Hermitian, positive semi-definite, trace one matrix. It follows that $\rho$ can be diagonalized by a unitary matrix, and has eigenvalues that are all real, nonnegative, and sum to one.

---

*Note that $\rho^2 = \rho$ for pure states. We will use this property when we discuss the Bloch sphere.*
Unitary Evolution of Mixed States

What happens to a mixed state when a unitary transform $U$ is applied to it? If the system described by the mixed state is actually in pure state $|\psi_j\rangle$ with the density matrix $\rho_j = |\psi_j\rangle\langle\psi_j|$, then it will evolve to the state $U\rho_j U^\dagger$. But we only know that the system is in the state $\rho_j$ with probability $p_j$. Therefore, the mixed state will evolve to the state $U\rho_j U^\dagger$ with probability $p_j$, that is, another mixed state, whose density matrix is given by

$$
\sum_j p_j U|\psi_j\rangle\langle\psi_j|U^\dagger = U\left(\sum_j p_j|\psi_j\rangle\langle\psi_j|\right)U^\dagger
$$

Therefore, $\rho \rightarrow U\rho U^\dagger$.

Ensembles of States

We call the set of pure states together with the associated probabilities an ensemble of states. Observe that two different ensembles can have identical density matrices, and therefore quantum mechanically represent identical states. Figure 1 shows two different ensembles with the density matrix equal to $I$.

Mixed State #1:

$p_0 = p_1 = \frac{1}{2} \implies$

$$
\rho = \frac{1}{2}|0\rangle\langle0| + \frac{1}{2}|1\rangle\langle1| = \frac{1}{2}I
$$

Mixed State #2:

$p_1 = p_2 = p_3 = \frac{1}{3} \implies$

$$
\rho = \frac{1}{3}|\psi_1\rangle\langle\psi_1| + \frac{1}{3}|\psi_2\rangle\langle\psi_2| + \frac{1}{3}|\psi_3\rangle\langle\psi_3| = \frac{1}{2}I
$$

Figure 1: Two “different” mixtures of pure states with identical density matrices.
This lecture discusses measuring quantum mixed states and introduces the von Neumann entropy.

**Measuring Mixed States**

Measurements are also easily described in the density matrix formalism. Recall that the projective measurement is defined by a set of $m \times m$ matrices $\{\Pi_i\}_{i=1}^\ell$ such that

1. $\{\Pi_i\}$ are pairwise orthogonal projection operators
2. $\{\Pi_i\}$ form a complete resolution of the identity, that is, $\Pi_1 + \Pi_2 + \cdots + \Pi_\ell = I_m$.
3. The measured state $|\psi\rangle$ gets projected (collapses) to state $\Pi_i |\psi\rangle \langle \psi| \Pi_i$ with probability $\frac{\Pi_i |\psi\rangle \langle \psi| \Pi_i}{\|\Pi_i |\psi\rangle\|}$.

Therefore, pure state $\rho_\psi = |\psi\rangle \langle \psi|$ collapses to $\rho_\psi \Pi_i = \sum_j \frac{\Pi_i |\psi_j\rangle \langle \psi_j|}{\|\Pi_i |\psi_j\rangle\|}$ with probability $\frac{\Pi_i |\psi\rangle \langle \psi| \Pi_i}{tr(\Pi_i \rho_\psi)}$.

We next look into what happens when we perform such a measurement on a mixed state whose density matrix is $\rho = \sum_j p_j |\psi_j\rangle \langle \psi_j|$.

We denote $\rho_j = |\psi_j\rangle \langle \psi_j|$. If the state being measured is $|\psi_j\rangle$ (which happens with probability $p_j$), then the probability of getting measurement result $i$ is $\text{tr}(\Pi_i |\psi_j\rangle \langle \psi_j|)$. Therefore, by the total probability formula, when measuring $\rho$, we get outcome $i$ with probability $\sum_j p_j \frac{\Pi_i |\psi_j\rangle \langle \psi_j|}{\|\Pi_i |\psi_j\rangle\|} = \text{tr}(\Pi_i \rho)$.

Is the state corresponding to outcome $i$ pure or mixed? If the state being measured is $|\psi_j\rangle$ and the measurement result is $i$, then the system is in the state $\frac{\Pi_i (\rho_j \Pi_i)}{\text{tr}(\Pi_i \rho_j)}$. Therefore, if we observe outcome $i$, the system is in the mixed state $\sum_j p_j \frac{\Pi_i \rho_j \Pi_i}{\text{tr}(\Pi_i \rho_j)} = \frac{\Pi_i \rho_j \Pi_i}{\text{tr}(\Pi_i \rho)}$.

Note that different ensembles $\{\psi_j, p_j\}$ with the same $\rho$ will give outcome $i$ with the same probability $\text{tr}(\Pi_i \rho)$, which depends only on $\rho$. 

\[ \text{Introduction to Quantum Information Science}^1 \]

\[ Prof. Emina Soljanin \]

\[ Lecture #11, October 24 \]

1. Rutgers, ECE 579, Fall 2018
Note that we ended up having a mixed state after the measurement resulted in outcome $i$, because we started with a mixed state.

Which state would we have if we lost the measurement record? We saw that we get state $\frac{\Pi_i \rho \Pi_i}{\text{tr}(\Pi_i \rho)}$. We would therefore have

$$
\sum_{i=1}^{\ell} \text{tr}(\Pi_i \rho) \cdot \frac{\Pi_i \rho \Pi_i}{\text{tr}(\Pi_i \rho)} = \sum_{i=1}^{\ell} \Pi_i \rho \Pi_i
$$

Is this a valid density matrix?

**Math Interlude – Matrix Functions**

A logarithm of a square complex matrix $A$ is any matrix $B$ such that $e^B = A$, where the exponential of a matrix $B$ is defined by

$$
e^B = \sum_{n=0}^{\infty} \frac{B^n}{n!}.
$$

Note that $e^{VBV^{-1}} = Ve^B V^{-1}$, and thus $\log A = V \log V^{-1} AV V^{-1}$.

Let $D_\rho$ be a matrix obtained by diagonalizing $\rho$. Then

$$
D_\rho = \rho \rho^{-1} \implies \log D_\rho = \log (\rho^{-1} D_\rho \rho^{-1}) = \rho (\log \rho) \rho^{-1}
$$

**Von Neumann Entropy**

Recall that the Shannon entropy measures the expected uncertainty associated with a classical probability distribution. The quantum counterpart of a probability distribution is a density matrix $\rho$. The von Neumann entropy is an older concept that generalizes the Shannon entropy. It is given by

$$
S(\rho) = -\text{tr} \rho \log \rho
$$

If $\lambda_i$ are eigenvalues of $\rho$, we have

$$
S(\rho) = -\sum_i \lambda_i \log \lambda_i
$$

The von Neumann entropy of a density matrix is, therefore, the Shannon entropy of the set of its eigenvalues. Properties of the Shannon entropy imply that 1) the von Neumann entropy is nonnegative, and zero if and only if the state is pure and 2) if $\rho$ is in a $d$-dimensional Hilbert space, then the entropy is at most $\log d$. The entropy is equal to $\log d$ if and only if the system is in the mixed state $I/d$. 

\(^1\)Here “we” is used on purpose to stress that, mathematically, the state of the system is described based on our ignorance.

\(^3\)Shannon entropy has multiple operational meanings, including the compression rate of classical DMS information sources.
This lecture discusses the Von Neumann entropy and some relations between two probability distributions and two density matrices.

**Math Interlude**

**Majorization**

Majorization is a preorder\(^2\) relation on vectors of real numbers. We will use it to relate two probability distributions. Let \( P = \{ p_1, \ldots, p_k \} \) and \( Q = \{ q_1, \ldots, q_k \} \) be vectors of probabilities, and \( P^\uparrow \) and \( Q^\uparrow \) the vector with identical components as \( P \) and \( Q \) respectively, but sorted in descending order. We write \( Q \succ P \) say that \( Q \) majorizes \( P \), or dominates \( P \), or \( P \) is majorized by \( Q \) when

\[
\sum_{j=1}^{i} p_j^\uparrow \leq \sum_{j=1}^{i} q_j^\uparrow \quad \text{for } i = 1, \ldots, k-1.
\]

Note that \( \sum_{j=1}^{i} p_j^\uparrow = \sum_{j=1}^{i} q_j^\uparrow = 1 \). Vectors \( P \) and \( Q \) do not have to have the same support, since we can pad the smaller support vector with zeros.

**Schur Concavity**

We say that \( f : \mathbb{R}^k \to \mathbb{R} \) is Schur concave when we have that \( a \succ b \) implies \( f(a) \leq f(b) \).

Similarly, \( f \) is Schur convex when \( a \succ b \) implies \( f(a) \geq f(b) \).

Every concave symmetric function is Schur-concave.\(^3\)

**Examples**

1. For quasi-uniform vectors with \( d \) components, we have

\[
\left\{ \frac{1}{d}, \ldots, \frac{1}{d} \right\} \prec \left\{ \frac{1}{d-1}, \ldots, \frac{1}{d-1}, 0 \right\} \prec \cdots \prec \{ 1, 0, \ldots, 0 \}
\]

2. For any probability distribution \( P \) with support size \( d \), we have

\[
\left\{ \frac{1}{d}, \ldots, \frac{1}{d} \right\} \prec P
\]
How to Mix a Density Matrix

We have seen that different quantum ensembles \( \{ p_j, |\psi_j\rangle \} \) can have identical density matrices. Can we characterize ensembles that have a given density matrix? Can we characterize probability distributions that are consistent with a given density matrix.

Let \( \{ p_j \} \) be a PD, and \( p_1 \geq p_2 \geq \cdots \geq p_k \). Let \( \rho \) be a density matrix, and \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \) its eigenvalues. There exist vectors \( |\psi_i\rangle \) such that \( \rho = \sum_{i=1}^{k} p_i |\psi_i\rangle \langle \psi_i| \) iff

\[
\sum_{i=1}^{n} p_i \leq \sum_{i=1}^{n} \lambda_i \quad \text{for all } n < d.
\]

For \( \rho = \frac{1}{d} I_d \), this condition becomes \( p_1 \leq \frac{1}{d} \).

von Neumann Entropy

Recall that the Shannon entropy measures the expected uncertainty associated with a classical probability distribution.\(^4\) The quantum counterpart of a probability distribution is a density matrix \( \rho \). The von Neumann entropy is an older concept that generalizes the Shannon entropy. It is given by

\[
S(\rho) = - \text{tr} \rho \log \rho
\]

If \( \lambda_i \) are eigenvalues of \( \rho \), we have

\[
S(\rho) = - \sum_i \lambda_i \log \lambda_i
\]

The von Neumann entropy of a density matrix is, therefore, the Shannon entropy of the set of its eigenvalues. Properties of the Shannon entropy imply that

1. the von Neumann entropy is nonnegative, and zero if and only if the state is pure
2. if \( \rho \) is in a \( d \)-dimensional Hilbert space, then the entropy is at most \( \log d \). The entropy is equal to \( \log d \) if and only if the system is in the mixed state \( I/d \).

Let \( \rho \) be a \( d \)-dimensional density matrix with eigenvalues \( \{ \lambda_j \}_{j=1}^{d} \), and \( \{ p_j, |\psi_j\rangle \}_{j=1}^{k} \), \( k \geq d \), an ensemble of distinct pure states, then

\[
\{ p_j \}_{j=1}^{k} \preceq \{ \lambda_j \}_{j=1}^{k}, \quad \text{where } \lambda_i = 0 \text{ for } i = d+1, \ldots, k
\]

Therefore, \( H(\{ p_j \}) \geq S(\rho) \).
How Close are Two Quantum States?

How Close are Two Probability Vectors?

Let \( P = \{p_1, \ldots, p_k\} \) and \( Q = \{q_1, \ldots, q_k\} \) be vectors of probabilities. We can tell how close these vectors are by

1. **total variation**

\[
D(P, Q) = \frac{1}{2} \sum_i |p_i - q_i|
\]

which measures the distance, and

2. **Bhattacharyya coefficient**

\[
BC(P, Q) = \sum_i \sqrt{p_i q_i}
\]

which measures the amount of overlap.\(^5\)

\(^5\) Log of BC is Bhattacharyya distance.

**Fidelity and Trace Distance**

To measure how faithfully mixed state \( \sigma \) approximates mixed state \( \omega \) and vice versa, we use the so called **mixed state fidelity** \( F \) defined as

\[
F(\sigma, \omega) = \left\{ \text{tr} \left[ (\sqrt{\sigma} \omega \sqrt{\sigma})^{1/2} \right] \right\}^2
\] (1)

Besides computing the mixed state fidelity (1), one can measure how close state \( \sigma \) is to state \( \omega \) by computing the **trace distance**

\[
D(\sigma, \omega) = \frac{1}{2} \text{tr} |\sigma - \omega|.
\]

Here \(|A|\) denotes the positive square root of \( A \dagger A \), i.e., \(|A| = \sqrt{A \dagger A}\).

The trace distance and the fidelity are closely related and the following holds:

\[
1 - F(\sigma, \omega) \leq D(\sigma, \omega) \leq \sqrt{1 - F(\sigma, \omega)^2}.
\] (2)

The trace distance is a metric on the space of density operators, and therefore the triangle inequality is true:

\[
D(\sigma, \omega) \leq D(\sigma, \tau) + D(\tau, \omega).
\] (3)

It has some other useful properties, as well. When we need one of those properties, we shall switch from the fidelity to the trace distance and back by making use of the inequalities in (2).
Mixed State #1:

\[ p_0 = p_1 = \frac{1}{2} \implies \rho = \frac{1}{2} |0\rangle \langle 0| + \frac{1}{2} |1\rangle \langle 1| = \frac{1}{2} I \]

\[ |0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \]

\[ |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \]

Mixed State #2:

\[ p_1 = p_2 = p_3 = \frac{1}{3} \implies \rho = \frac{1}{3} |\psi_1\rangle \langle \psi_1| + \frac{1}{3} |\psi_2\rangle \langle \psi_2| + \frac{1}{3} |\psi_3\rangle \langle \psi_3| = \frac{1}{3} I \]

\[ |\psi_1\rangle = \begin{bmatrix} -1/2 \\ -\sqrt{3}/2 \end{bmatrix} \]

\[ |\psi_2\rangle = \begin{bmatrix} -1/2 \\ \sqrt{3}/2 \end{bmatrix} \]

Problems - Homework Due November 7

1. Compute the Shannon entropy of the probability distributions and the von Neumann entropy of the two mixed states in the figure above.

2. Consider two density matrices \( \sigma \) and \( \omega \) such that \( \sigma \omega = \omega \sigma \). Express the fidelity and the trace distance between \( \sigma \) and \( \omega \) in terms of their eigenvalues.

3. Consider two pure states \( |\psi\rangle \) and \( |\varphi\rangle \). Find the fidelity and the trace distance between \( \rho_{\psi} \) and \( \rho_{\varphi} \).

4. Let \( \rho \) be a density operator. Show that \( \text{tr}(\rho^2) \leq 1 \) with equality if and only if \( \rho \) is a pure state.
**Introduction to Quantum Information Science**

Prof. Emina Soljanin

Lecture #13, October 31

This lecture 1) introduces the Bloch sphere, 2) discusses bipartite states, and 3) describes quantum information processing.

**Bloch Sphere**

Any $2 \times 2$ complex matrix, and thus any density matrix $\rho$, can be expressed as a linear combination of the identity $I$ and the Pauli matrices $\sigma_X, \sigma_Y, \text{and} \sigma_Z$. Note that $\sigma_X, \sigma_Y, \text{and} \sigma_Z$ have trace equal to 0. Since a density matrix is Hermitian and has trace one, we have

$$
\rho = \frac{1}{2} (I + \beta_X \sigma_X + \beta_Y \sigma_Y + \beta_Z \sigma_Z)
$$

where $\beta_X, \beta_Y, \text{and} \beta_Z$ are real numbers. To see that we write the above expression for $\rho$ as follows:

$$
\rho = \frac{1}{2} \begin{bmatrix}
1 + \beta_Z & \beta_X - i \beta_Y \\
\beta_X + i \beta_Y & 1 - \beta_Z
\end{bmatrix}
$$

We call $\vec{\beta} = (\beta_X, \beta_Y, \beta_Z)$ the Bloch vector of $\rho$. Since $\rho$ is positive semi-definite, we have $\det(\rho) > 0$:

$$
0 \leq \det(\rho) = 1 - (\beta_X^2 + \beta_Y^2 + \beta_Z^2) = 1 - |\vec{\beta}|^2
$$

which implies $|\vec{\beta}|^2 \leq 1$. The set of all vectors that satisfy this condition is a ball in $\mathbb{R}^3$, known as the Bloch sphere.

Furthermore, for pure states, we have $\text{tr}(\rho^2) = 1$, and thus

$$
1 = \text{tr}(\rho^2) = \frac{1}{2} (1 + |\vec{\beta}|^2) \quad \Leftrightarrow \quad |\vec{\beta}| = 1
$$

As a consequence, the surface of the Bloch sphere represents all the pure states of a two-dimensional quantum system, whereas the interior corresponds to all the mixed states. In particular,

- $\beta_X = \beta_Y = 0$ and $\beta_Z = 1$ gives $\rho = |0\rangle \langle 0|$
- $\beta_X = \beta_Y = 0$ and $\beta_Z = -1$ gives $\rho = |1\rangle \langle 1|$

Sometimes $|\uparrow\rangle$ and $|\downarrow\rangle$ (spin up and down) are used to denote $|0\rangle$ and $|1\rangle$ respectively.

Any two diametrically opposite (antipodal) points correspond to a pair of mutually orthogonal pure state vectors. Why?
Bipartite Quantum States

Let $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. We say that $|\psi\rangle$ is a bipartite quantum state of a composite system with subsystems $A$ and $B$. Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be finite-dimensional Hilbert spaces with basis states $\{|a_i\rangle\}_{i=1}^n$ and $\{|b_j\rangle\}_{j=1}^m$, respectively. Then the state space of the composite system is the tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$ with the basis $\{|a_i\rangle \otimes |b_j\rangle\}$. Any pure state of the composite system can be written as

$$|\psi\rangle = \sum_{i=1}^n \sum_{j=1}^m c_{ij} |a_i\rangle \otimes |b_j\rangle,$$

where $c_{ij}$ are complex numbers. If a pure state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ can be written in the form $|\psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$, it is said to be separable. Otherwise it is called entangled. When a system is in an entangled pure state, it is not possible to assign states to its subsystems.

Let $\rho_{AB}$ be a density matrix in the product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. A mixed state of the bipartite system described by $\rho_{AB}$ can be

1. a product state if $\rho_{AB} = \rho_A \otimes \rho_B$ or
2. a separable state if there exist a probability distribution $\{p_k\}$, and $\{\rho_{A}^k\}$ and $\{\rho_{B}^k\}$ which are mixed states of the respective subsystems such that

$$\rho = \sum_k p_k \rho_{A}^k \otimes \rho_{B}^k.$$

Otherwise $\rho_{AB}$ is an entangled state. Note that, for mixed states, separable and product are different notions.

Reduced Density Operator

Recall the trace expression in Dirac’s notation: Let $|e_i\rangle$, $i = 1, \ldots, n$ be an orthonormal basis of $\mathbb{C}^n$. Then $|e_i\rangle\langle e_i|A$ is a matrix whose $i$-th diagonal element is $a_{ii}$ and all other elements are 0. Therefore,

$$\text{tr} A = \sum_{i=1}^n \text{tr} (|e_i\rangle \langle e_i|A) = \sum_{i=1}^n \langle e_i| A |e_i\rangle.$$

Let $\rho_{AB}$ be a density matrix in the product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, and let $|b_i\rangle$ be an orthonormal basis for $\mathcal{H}_B$. Then the partial trace over the Hilbert space $\mathcal{H}_B$ is defined as follows:\footnote{You will often see a shorthand expression $\text{tr}_B \rho_{AB} = \sum_b \langle b| \rho_{AB} |b\rangle$}

$$\rho_A = \text{tr}_B \rho_{AB} = \sum_i (I \otimes |b_i\rangle) \rho_{AB} (I \otimes |b_i\rangle).$$
**Example #1 – Product State:**

Suppose a quantum system is in the product state $\rho_{AB} = \rho_A \otimes \rho_B$ where $\rho_A$ is a density operator for system A, and $\rho_B$ is a density operator for system B. Then

$$\rho_A = \text{tr}_B(\rho_A \otimes \rho_B) = \rho_A \text{ tr } \rho_B = \rho_A$$

**Example #2 – Bell State:**

Consider the bipartite state $|\phi_{AB}\rangle = (|00\rangle + |11\rangle) / \sqrt{2}$. This is a pure state with the density operator

$$\rho_{AB} = |\phi_{AB}\rangle \langle \phi_{AB}| = \frac{1}{2}(|00\rangle \langle 00| + |01\rangle \langle 01| + |10\rangle \langle 11| + |11\rangle \langle 11|)$$

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

$$\rho_A = \text{tr}_B \rho_{AB} = (I \otimes |0\rangle \langle 0|) \rho_{AB} (I \otimes |0\rangle \langle 0|) + (I \otimes |1\rangle \langle 1|) \rho_{AB} (I \otimes |1\rangle \langle 1|)$$

$$= \frac{1}{2}(|0\rangle \langle 0| + |1\rangle \langle 1|) = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

**Quantum Information Systems – Recap & Generalization**

**Quantum States**

Quantum states are, in the simplest case, mathematically represented as unit length column vectors in a $d$-dimensional Hilbert space $\mathcal{H}$. Such quantum states are called pure. When $d = 2$, quantum states are called qubits. A pure state is mathematically described by its density matrix equal to the outer product $|\varphi\rangle \langle \varphi|$

In a more complex case, all we know about a quantum state is that it is one of a finite number of possible pure states $|\varphi_i\rangle$ with probability $p_i$. Such quantum states are called mixed. A mixed state is also described by its density matrix which is equal to

$$\rho = \sum_i p_i |\varphi_i\rangle \langle \varphi_i|.$$ 

Note that a density matrix is a $d \times d$ Hermitian trace-one positive semidefinite matrix.

When we consider sources of quantum states, we often deal with sequences rather than with individual states. The quantum state corresponding to a sequence of length $n$ has a $d^n \times d^n$ density matrix, equal to the tensor product of density matrices corresponding to the states in the sequence.
**Quantum Evolutions**

A quantum state \( \rho \) can be transformed to another state \( \mathcal{E}(\rho) \) only by a physical process consistent with the laws of quantum theory. Such a process is, in the simplest case, mathematically described as a *unitary* evolution:

\[
\mathcal{E}(\rho) = U \rho U^\dagger \quad \text{where} \quad UU^\dagger = I,
\]

and, in the general case, as an evolution by a *completely positive, trace-preserving* map:

\[
\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger \quad \text{where} \quad \sum_k E_k^\dagger E_k = I.
\]

It is envisioned that a quantum computer (like a classical one) would implement such evolutions by using universal quantum gates. An example of a two-qubit quantum gate is the \( \text{XOR} \).

\[
\text{XOR} : |x, y\rangle \rightarrow |x, x \oplus y\rangle \quad U_{\text{XOR}} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0
\end{bmatrix}
\]

**Quantum Measurements**

A quantum measurement is a physical process applied to determine the state of the quantum system being measured. Only when the possible states, say \( \{ |\psi_j\rangle, \ j = 1, \ldots, k \} \), are orthogonal can a quantum measurement be designed to give an unambiguous answer.

The simplest model of quantum measurement is known as the von Neumann’s measurement. Mathematically, this type of measurement is defined by a set of pairwise orthogonal projection operators \( \{ \Pi_i \} \) which form a complete resolution of the identity, that is, \( \sum_i \Pi_i = I \). For input \( |\psi_j\rangle \), the *classical* output \( \Pi_i |\psi_j\rangle / \sqrt{\text{tr}(\Pi_i |\psi_j\rangle \langle \psi_j|)} \) happens with probability \( \text{tr}(\Pi_i |\psi_j\rangle \langle \psi_j|) \).

In a more general case, the pairwise orthogonal projection operators \( \{ \Pi_i \} \) are replaced by any positive-semidefinite operators \( \{ E_i \} \) which form a complete resolution of the identity. This type of measurement is known as *positive operator-valued measure* (POVM).
Cloning, Broadcasting, and Deleting

Quantum information cannot be cloned, broadcast or deleted in the sense made precise below, unless we are dealing with states with commuting density matrices.

The No-Cloning Principle: There is no physical process that for all $|\psi\rangle$ leads to an evolution

$$|\phi\rangle \otimes |s\rangle \rightarrow |\phi\rangle \otimes |\phi\rangle$$

where $|\phi\rangle$ is an arbitrary state and $|s\rangle$ is a fixed state. (Approximate cloning, on the other hand, is possible.)

The No-Deleting Principle: There is no physical process that for all $|\psi\rangle$ leads to an evolution

$$|\phi\rangle \otimes |\phi\rangle \rightarrow |\phi\rangle \otimes |s\rangle$$

where $|\phi\rangle$ is an arbitrary state and $|s\rangle$ is a fixed state.

The No-Broadcasting Principle: Suppose that quantum system A is in an unknown state $\rho$ and quantum system B is in some known or standard state $\omega$. Then there is no physical process that for all $\rho$ leads to an evolution

$$\rho \otimes \omega \rightarrow \sigma \text{ s.t. } tr_A(\sigma) = \rho \text{ and } tr_B(\sigma) = \rho,$$

that is, so that both subsystem A and subsystem B evolve into state $\rho$. 

This lecture introduces quantum information sources and quantum visible compression.

Quantum DMS

A discrete memoryless source (DMS) of information produces a sequence of independent, identically distributed random variables taking values in a finite set called the source alphabet. In quantum systems, source letters are mapped into quantum states for quantum transmission or storage. We will concentrate on pure states where source letters are transmitted as $d$-dimensional unit-length vectors in $\mathcal{H}_d$. When $d = 2$ we have qubits. We will work with qubits without loss of generality.

Example: The 4 letters from alphabet $\mathcal{X} = \{0, 1, 2, 3\}$ are mapped onto 4 qubits

$$
\begin{align*}
|\psi_0\rangle &= \alpha_0|0\rangle + \beta_0|1\rangle & |\psi_1\rangle &= \alpha_1|0\rangle + \beta_1|1\rangle \\
|\psi_2\rangle &= \alpha_2|0\rangle + \beta_2|1\rangle & |\psi_3\rangle &= \alpha_3|0\rangle + \beta_3|1\rangle
\end{align*}
$$

where $|0\rangle$ and $|1\rangle$ are the basis vectors of $\mathcal{H}_2$: 

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

Sequence of two letters of this source are vectors in the $\mathcal{H}_4$ with the basis vectors

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

The Density Matrix and the Entropy of the Source

The Source density matrix is defined as

$$
\rho = \sum_{a \in \mathcal{X}} P_a \frac{|\psi_a\rangle \langle \psi_a|}{\rho_a}
$$

and its von Neumann entropy is

$$
S(\rho) = -\text{tr} \rho \log \rho = -\sum_i \lambda_i \log \lambda_i,
$$
where $\lambda_i$ are the eigenvalues of $\rho$.

Observe that two different sources can have identical density matrices, and therefore identical entropies. Figure 1 shows two different sources with the density matrix equal to $\frac{1}{2}I$ and von Neumann entropy equal to 1. Which source has larger Shannon entropy?

**Source #1:**

$$X = \{0, 1\}$$

$$p_0 = p_1 = \frac{1}{2} \implies$$

$$\rho = \frac{1}{2}\langle 0|0\rangle + \frac{1}{2}\langle 1|1\rangle = \frac{1}{2}I$$

$$S(\rho) = 1$$

**Source #2:**

$$X = \{1, 2, 3\}$$

$$p_1 = p_2 = p_3 = \frac{1}{3} \implies$$

$$\rho = \frac{1}{3}\langle \psi_1|\psi_1\rangle + \frac{1}{3}\langle \psi_2|\psi_2\rangle + \frac{1}{3}\langle \psi_3|\psi_3\rangle = \frac{1}{2}I$$

$$S(\rho) = 1$$

Figure 1: Two classically different quantum sources with identical density matrices.

**Source Sequences and a Probabilistically Large Set**

We say that a source sequence $x = x_1, x_2, \ldots, x_n \in \mathcal{X}^n$ is weakly $\epsilon_n$-typical (aka entropy $\epsilon_n$-typical) if

$$2^{-n(H(P) + \epsilon_n)} \leq P(x_1, x_2, \ldots, x_n) \leq 2^{-n(H(P) - \epsilon_n)}$$

The set of all such sequences is known as the typical set $A^P_{\epsilon_n}$ according to distribution $P$.

The AEP implies that, asymptotically, the probability of the typical set is large, while (unless $P$ is uniform) its size is small. More precisely, given any $\epsilon_n > 0$, one can choose $n$ such that$^4$

1. $\Pr[A^P_{\epsilon_n}] \geq 1 - \epsilon_n$

2. $(1 - \epsilon_n)2^n(H(P) - \epsilon_n) \leq |A^P_{\epsilon_n}| \leq 2^n(H(P) + \epsilon_n)$

$^4$We have shown this result by using the weak low of large numbers.
Therefore, the fraction of sequences that are typical is
\[
\frac{|\mathcal{A}_P^n|}{|X^n|} \leq \frac{2^{nH(P) + \epsilon_n}}{2^n \log_2 |X|} = 2^{-n \left( \log_2 |X| - H(P) - \epsilon_n \right)} \to 0
\]
as \(n \to \infty\) since \(H(P) < \log_2 |X|\).
Set \(\mathcal{A}_{\epsilon_n}^P\) is probabilistically large but exponentially smaller than \(|X^n|\).

(Asymptotically) Lossless Classical Data Compression

We talked about two algorithms for classical data compression: 1) enumeration and 2) random binning. The enumeration algorithm implements lossless compression and the binning algorithm implements asymptotically lossless compression.

In general, Alice, a source encoder, compresses source sequence \(x\) to a \(c(x)\). Then Bob, a source decoder, decompresses \(c(x)\) into a reproduction \(\hat{x} \in X^n\).\(^5\) The fidelity of the reproduction in the classical case is defined as
\[
F(x, \hat{x}) = \begin{cases} 
1 & \text{if } x = \hat{x} \\
0 & \text{otherwise.}
\end{cases}
\]
Therefore the average compression fidelity is
\[
\bar{F} = \sum_{x \in X^n} P(x) F(x, \hat{x}).
\]
The average fidelity should approach 1 as \(n \to \infty\). In the enumeration algorithm, we have \(x = \hat{x}\), and thus \(\bar{F} = 1\) for all \(x\). In the binning algorithm, the fidelity is 0 for 1) all non-typical sequences, and 2) for any typical sequence \(x\) if there is another typical sequence \(x'\) such that \(c(x) = c(x')\). The probability of the former event goes to 0 as \(n \to \infty\). For a sufficiently large number of bins, the probability of the latter event also goes to 0 as \(n \to \infty\). See the homework.\(^6\)

Vector Sequences

Suppose that the classical source letter \(a \in X\) is mapped into the quantum source state \(|\psi_a\rangle \in \mathcal{H}_d\). Then quantum state \(|\Psi_x\rangle \in \mathcal{H}_d^\otimes n\) that corresponds to source sequence \(x = x_1, x_2, \ldots, x_n \in X^n\) is given by
\[
|\Psi_x\rangle = |\psi_{x_1}\rangle \otimes |\psi_{x_2}\rangle \otimes \cdots \otimes |\psi_{x_n}\rangle, \quad x_i \in X,
\]
which we will refer to as the quantum source vector-sequence. State \(|\Psi_x\rangle\) is produced by the source with probability \(P_x = P_{x_1} \cdot P_{x_2} \cdots P_{x_n}\).
The states that correspond to typical sequences are called typical states. Thus there are approximately \(2^{nH(P)}\) typical states.

\(^5\)cf. Alice & Bob of channel coding
\(^6\)Is there a price to pay for perfect fidelity of the enumeration algorithm?
Typical States and Visible Compression

Suppose that source sequences $x$ are visible by the encoder Alice, that is, she knows the sequence. She can compress with perfect fidelity the typical states. Instead of $n$ bits of $x$, she transmits a different string $c(x)$ of approximately $nH(P)$ bits to Bob for each typical state and some fixed string $c_0$ of approximately $nH(P)$ bits for each non-typical state. The decoder Bob prepares $|\psi_x\rangle$ as $|\hat{\psi}_x\rangle$ for typical $x$ (when he receives $c(x)$), and some fixed state $|\Phi\rangle$ for each non-typical state (when he receives $c_0$). How well does this strategy perform in terms of the average fidelity and the compression rate?

Since perfect reconstruction is guaranteed only for the typical sequences, we have

$$F(x, \hat{x}) = \begin{cases} 1 & \text{if } x = \hat{x} \\ 0 & \text{otherwise.} \end{cases}$$

$$\bar{F} = \sum_{x \in \mathcal{X}^n} P(x) F(x, \hat{x}) = \sum_{x \in A_p^n} P(x) F(x, \hat{x}) + \sum_{x \in \mathcal{X}^n \setminus A_p^n} P(x) F(x, \hat{x})$$
$$\geq P^n(A_p^n) \geq 1 - \epsilon_n.$$ 

Set $A_p^n$ contains approximately $2^{nH(P)}$ sequences. Therefore, Alice has to send approximately $nH(P)$ bits as opposed to $n \log |\mathcal{X}|$ bits or $n$ qubits.
This lecture discusses blind quantum source coding.

**Quantum DMS**

A discrete memoryless source (DMS) of information produces a sequence of independent, identically distributed random variables taking values in a finite set called the source alphabet $\mathcal{X}$. The source produces letter $a \in \mathcal{X}$ with probability $P_a$. In quantum systems, source letters are mapped into quantum states for quantum transmission or storage. We will concentrate on pure states where source letter $a \in \mathcal{X}$ is mapped into qubit $|\psi_a\rangle$.

**The Density Matrix and the Entropy of the Source**

The Source density matrix is defined as

$$
\rho = \sum_{a \in \mathcal{X}} P_a |\psi_a\rangle \langle \psi_a|,
$$

and its von Neumann entropy is

$$
S(\rho) = -\text{tr} \rho \log \rho = -\sum_i \lambda_i \log \lambda_i,
$$

where $\lambda_i$ are the eigenvalues of $\rho$.

**Vector Sequences**

Suppose that the classical source letter $a \in \mathcal{X}$ is mapped into the quantum source state $|\psi_a\rangle \in \mathcal{H}_d$. Then quantum state $|\Psi_x\rangle \in \mathcal{H}_d^\otimes n$ that corresponds to source sequence $x = x_1, x_2, \ldots, x_n \in \mathcal{X}^n$ is given by

$$
|\Psi_x\rangle = |\psi_{x_1}\rangle \otimes |\psi_{x_2}\rangle \otimes \cdots \otimes |\psi_{x_n}\rangle, \quad x_i \in \mathcal{X},
$$

which we will refer to as the quantum source vector-sequence. State $|\Psi_x\rangle$ is produced by the source with probability $P_x = P_{x_1} \cdot P_{x_2} \cdots P_{x_n}$. The states that correspond to typical sequences are called typical states. Thus there are approximately $2^{nH(P)}$ typical states.

**The Typical Subspace $\Lambda_n$**

We represent the source density matrix $\rho = \sum_{a \in \mathcal{X}} P(a) |\psi_a\rangle \langle \psi_a|$ in terms of its eigenvectors and eigenvalues as

$$
\rho = \lambda_0 |\phi_0\rangle \langle \phi_0| + \lambda_1 |\phi_1\rangle \langle \phi_1|,
$$

Mixed states compression has not been fully understood.
Recall that $\Lambda = \{\lambda_0, \lambda_1\}$ is a PD on $\{0, 1\}$ and $\langle \phi_0 | \phi_1 \rangle = 0$, and thus we can define typical sequences according to distribution $\Lambda$. We say that sequence $z = z_1, \ldots, z_n \in \{0, 1\}^n$ is weakly $\epsilon_n$-typical if
\[ 2^{-n(H(\Lambda) + \epsilon_n)} \leq \Lambda(z_1, \ldots, z_n) \leq 2^{-n(H(\Lambda) - \epsilon_n)} \]
The set of all such sequences $\mathcal{A}_{\epsilon_n}$ is the typical set according to distribution $\Lambda$. There are approximately $2^{nH(\Lambda)} = 2^{nS(\rho)}$ such sequences.
We define the typical subspace $\Lambda_n$ to be the subspace spanned by the typical states $|\Phi_z\rangle$, $z \in \mathcal{A}_{\epsilon_n}$. We define the projector to $\Lambda_n$ and its complement:
\[ \Pi = \sum_{z \in \mathcal{A}_{\epsilon_n}} |\Phi_z\rangle\langle\Phi_z| \text{ is the projector to } \Lambda_n. \]
\[ \Pi^\perp = \sum_{z \in \{0, 1\}^n \setminus \mathcal{A}_{\epsilon_n}} |\Phi_z\rangle\langle\Phi_z| \text{ is the projector to } \Lambda_n^\perp. \]
Note that $\Pi + \Pi^\perp = I_{2^n}$.

The dimension of $\Lambda_n$ is approximately $2^{nS(\rho)}$.

Vector-Sequences and Fidelity

Recall that source vector-sequences $|\Psi_x\rangle$ are in $\mathcal{H}^{2^n}$, $(x \in \{0, 1\}^n)$. Vector $|\Psi_x\rangle$ is compressed and then reproduced as $|\tilde{\Psi}_x\rangle$. The fidelity between $|\Psi_x\rangle$ and $|\tilde{\Psi}_x\rangle$ is
\[ \text{F}(|\Psi_x\rangle, |\tilde{\Psi}_x\rangle) = |\langle \Psi_x | \tilde{\Psi}_x \rangle|^2 \]
For asymptotically lossless compression, the expected fidelity
\[ \bar{\text{F}} = \sum_{x \in \{0, 1\}^n} P(x) \text{F}(|\Psi_x\rangle, |\tilde{\Psi}_x\rangle) \]
should approach 1 as $n \to \infty$.

Typical Subspace and Blind Compression

In blind quantum compression, Alice cannot see source sequences, and has to compress quantum vector states by using operations allowed by quantum mechanics. To compress an $n$-qubit source vector-sequence $|\Psi_x\rangle \in \mathcal{H}^{2^n}$, Alice performs measurement defined by $\Pi$, $\Pi^\perp$. The state after the measurement is
1. $\Pi \cdot |\Psi_x\rangle / \sqrt{\langle \Psi_x | \Pi | \Psi_x \rangle} = |\Psi_x^{\Lambda_n}\rangle$ with probability $\langle \Psi_x | \Pi | \Psi_x \rangle$
2. $\Pi^\perp \cdot |\Psi_x\rangle / \sqrt{\langle \Psi_x | \Pi^\perp | \Psi_x \rangle}$ with probability $\langle \Psi_x | \Pi^\perp | \Psi_x \rangle$

If Alice gets result 1, her resulting state is $|\Psi_x^{\Lambda_n}\rangle$. Note that this is still an $n$-qubit state. However, Alice can apply a unitary change of basis $U$ that takes each state $|\Psi_x^{\Lambda_n}\rangle$ to a state of the form $|\Psi_x^{\cal E}\rangle \otimes |0_r\rangle$, where state $|\Psi_x^{\cal E}\rangle$ consists of $|\Lambda_{\epsilon_n}^{\perp}\rangle$ (approximately $2^{nS(\rho)}$) qubits, and $|0_r\rangle$ is the $(2^n - |\Lambda_{\epsilon_n}^{\perp}|)$-fold Kronecker product of $|0\rangle$ states. Alice
then sends $|\Psi_x^c\rangle$ to Bob. Bob recovers $|\Psi_x^{\Lambda_n}\rangle$ by first appending $|o_r\rangle$ to $|\Psi_x^c\rangle$ and applying $U$ to the resulting bipartite state. In summary, the compression algorithm operates as follows:

**Encoder:**

For each source vector state $|\Psi_x\rangle$ of $n$ qubits,

1. Alice performs the measurement defined by the projection to the typical subspace and its complement to obtain $|\Psi_x^{\Lambda_n}\rangle$ (result 1) or $|\Psi_x^{\Lambda_{n'}}\rangle$ (result 2).

2. If Alice gets result 2, she sends some fixed state of $2^n S(\rho)$ qubits to Bob. Otherwise, she applies a unitary transform $U$ s.t.

   $$|\Psi_x^{\Lambda_n}\rangle \xrightarrow{U} |\Psi_x^c\rangle \otimes |o_r\rangle$$

   She then sends $|\Psi_x^c\rangle$ to Bob.

**Decoder Bob:**

1. Receives $|\Psi_x^c\rangle$ and appends ancillary qubits to get $|\Psi_x^c\rangle \otimes |o_r\rangle$

2. Applies a unitary transform $U$ s.t.

   $$|\Psi_x^c\rangle \otimes |o_r\rangle \xrightarrow{U} |\Psi_x^{\Lambda_n}\rangle$$

Bob’s reconstructed state is the normalized state

$$|\tilde{\Psi}_x\rangle = \Pi \cdot |\Psi_x\rangle / \sqrt{\langle \Psi_x||\Pi||\Psi_x\rangle}$$

**Blind Compression Fidelity**

$$\tilde{F} = \sum_{x \in X^n} P(x) F(\langle \Psi_x||\tilde{\Psi}_x\rangle)$$

$$= \sum_{x \in X^n} P(x) \left[ \Pr(\text{result 1}|x) F(\langle \Psi_x||\tilde{\Psi}_x\rangle) + \Pr(\text{result 2}|x) F(\langle \Psi_x||\tilde{\Psi}_x\rangle) \right]$$

$$\geq \sum_{x \in X^n} P(x) \Pr(\text{result 1}|x) F(\langle \Psi_x||\tilde{\Psi}_x\rangle)$$

$$= \sum_{x \in X^n} P(x) |\langle \Psi_x||\Pi||\Psi_x\rangle| / \sqrt{\langle \Psi_x||\Pi||\Psi_x\rangle^2}^2$$

$$= \sum_{x \in X^n} P(x)|\langle \Psi_x||\Pi||\Psi_x\rangle|^2 \geq \sum_{x \in X^n} P(x) (1 - 2\langle \Psi_x||\Pi||\Psi_x\rangle)$$

$$= -1 + 2 \text{tr}(\Pi \rho^{\otimes n})$$

$$= -1 + 2 \text{tr} \left\{ \left[ \sum_{z \in A_n^\Lambda} |\Phi_z\rangle\langle\Phi_z| \right] \left[ \sum_{z \in \{0,1\}^n} \lambda(z)|\Phi_z\rangle\langle\Phi_z| \right] \right\}$$

$$= 1 - 2\epsilon_n$$
Simon’s Problem

Problem Description

In the Simon’s problem, we are given a function

\[ f : \{0,1\}^n \rightarrow \{0,1\}^m, \]

and know that, for some \( a \in \{0,1\}^n \), we have, for all \( x, y \in \{0,1\}^n \),

\[ f(x) = f(y) \quad \text{if and only if} \quad x \oplus y = a. \]

In other words, \( f(x) = f(x \oplus a) \), for all \( x \in \{0,1\}^n \), and the problem is to find \( a \).

Observe that the above condition requires that \( f \) be a one-to-one function when \( a = 0 \), and two-to-one function, when \( a \neq 0 \). Note that \( x \oplus y = 0^n \) if and only if \( x = y \).

\[ |\psi_0 \rangle = \frac{1}{\sqrt{m}} \sum_{k=0}^{m-1} |x_0 + kr\rangle \]

**Parallel Function Evaluation**

We can evaluate an \( m \)-bit valued function \( f \) of an \( n \)-bit string \( x \):

\[ f : \{0,1\}^n \rightarrow \{0,1\}^m \]

We create a superposition of all basis states by applying the Hadamard \( H^{\otimes n} \) gate to state \( |0\rangle^{\otimes n} \), and then apply the \( U_f \) gate, we can simultaneously compute the value of \( f \) on its entire domain:

\[ U_f \left( H^{\otimes n} \otimes I_m \right) (|0\rangle^{\otimes n} \otimes |0\rangle^{\otimes m}) = \frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} |x, f(x)\rangle \]

But can we see the result?
Measurement followed by a Hadamard Transform

What happens when we measure the right part of the register

\[ \frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} |x, f(x)\rangle \]

and get the result \( f(s_0) \)? Then the state of the left register becomes

\[ \ket{\psi_0} = \frac{1}{\sqrt{2}} (\ket{s_0} + \ket{s_0 + a}) \].

We next apply the Hadamard transform\(^4\) on \( \ket{\psi_0} \) and get \( \ket{\psi_1} \):

\[
\ket{\psi_1} = H^\otimes n \ket{\psi_0} = H^\otimes n \frac{1}{\sqrt{2}} (\ket{s_0} + \ket{s_0 + a}) \\
= \frac{1}{\sqrt{2^{n+1}}} \sum_{y=0}^{2^n-1} ((-1)^{s_0 \cdot y} + (-1)^{(s_0 + a) \cdot y}) \ket{y} \\
= \frac{1}{\sqrt{2^{n+1}}} \sum_{y=0}^{2^n-1} (-1)^{s_0 \cdot y} (1 + (-1)^a \cdot y) \ket{y} \\
= \frac{2}{\sqrt{2^{n+1}}} \sum_{y=0, a \cdot y=0}^{2^n-1} (-1)^{s_0 \cdot y} \ket{y}
\]

If we measure state \( \ket{\psi_1} \) in the computational basis, the state will collapse to some \( \ket{y} \) for which \( a \cdot y = 0 \) (to any such \( y \) with equal probability)\(^5\).

Therefore, we get one equation with \( n \) unknowns of the form \( a \cdot y_1 = 0 \). We need \( n \) linearly independent equations to solve for \( a \). We can get \( n-1 \) of such equations with high probability\(^6\) by repeating the above procedure a finite number of times, and getting linearly independent\(^7\) vectors \( y_1 \) s.t. \( a \cdot y_1 = 0 \). We can get the \( n \)-th equation by picking any vector \( y_n \) in \( \mathbb{F}_2^n \) which is not in the span of the \( y_1 \)'s (and therefore not orthogonal to \( a \)). The resulting system of equations is

\[
a \cdot y_i = 0, \quad i = 1, \ldots, n - 1 \\
a \cdot y_n = 1
\]

Observe that the Simon’s algorithm is different from the D-J algorithm in that we need to run the quantum part many times to discover the value of \( a \).

Math Interlude – Factors and Periods

The order of an integer \( b \) modulo \( M \) is the smallest integer \( r > 0 \) such that \( b^r = 1 \mod M \); if no such integer exists, the order is said to be

\(^1\) Recall that

\[ H^\otimes n \ket{x} = \frac{1}{\sqrt{2^n}} \sum_{y=0}^{2^n-1} (-1)^{x \cdot y} \ket{y} \]

\[ x \cdot y = x_0 y_0 + x_1 y_1 + \cdots + x_{n-1} y_{n-1} \]

\(^5\) How many such \( y \)-s are there? How many of them can be linearly independent in \( \mathbb{F}_2^n \)?

\(^6\) requires proof

\(^7\) Check for linear independence classically!
We know that \( r \) is finite when \( b \) and \( M \) are relatively prime.\(^8\)

Consider the function

\[
f(x) = b^x \mod M.
\]

What is \( f(x + r) \)? Because \( b^x = b^{x+r} \mod M \) if and only if \( b^r = 1 \mod M \), for \( b \) relatively prime to \( M \), the order \( r \) of \( b \) modulo \( M \) is the period of \( f(x) = b^x \mod M \).

Finding the period of \( f(x) = b^x \mod M \) allows us to find a factor of \( M \). To see that, note that if \( b^r = 1 \mod M \) and \( r \) is even, we can write

\[
(b^{r/2} + 1)(b^{r/2} - 1) = 0 \mod M.
\]

As long as neither \( b^{r/2} + 1 \) nor \( b^{r/2} - 1 \) is a multiple of \( M \), both \( b^{r/2} + 1 \) and \( b^{r/2} - 1 \) have nontrivial common factors with \( M \). These factors can be found efficiently by e.g., the Euclidean algorithm.

### Shor’s Algorithm Outline

We want to factor \( M = pq \) where \( p \) and \( q \) are odd primes.\(^9\)

1. Pick a positive integer \( b \) smaller than \( M \). Find the greatest common divisor (GCD)\(^10\) \( y \) of \( b \) smaller than \( M \). If \( y > 1 \), then a non-trivial factor of \( M \) has been found. Otherwise, \( y = 1 \) meaning that \( b \) and \( M \) are relatively prime.

2. Create an \( n \) qubit superposition of all basis states in \( \mathcal{H}^{2^n} \) for some \( n \) s.t. \( M^2 \leq 2^n \leq 2M^2 \), and use quantum parallelism to compute \( f(x) = b^x \mod M \) on the superposition of inputs.

3. Measure the target (right) register and apply a quantum Fourier transform to the data register of the resulting state.

4. Measure the data register in the computational basis. With high probability, a value \( v \) close to a multiple of \( 2^n/r \) will be obtained.

5. Use classical methods to obtain a conjectured period \( r \) from the value \( v \).

6. If \( r \) is even, use the Euclidean algorithm to check efficiently whether \( b^{r/2} + 1 \) (or \( b^{r/2} - 1 \)) has a nontrivial common factor with \( M \).

7. Repeat all steps 2–5 if necessary.

\(^8\) Two integers are relatively prime if they have no common factors.

\(^9\) Such numbers are used in RSA.

\(^10\) e.g., by the Euclidean algorithm
Quantum Fourier Transform

The quantum Fourier transform on \( n \) qubits\(^{11} \) is the map that can be described by its action on the basis states \( |x\rangle \) of \( \mathcal{H}^n \) as follows:

\[
U_{\text{FT}} : |y\rangle \mapsto \frac{1}{\sqrt{N}} \sum_{y=0}^{N-1} \omega_N^y |y\rangle.
\]

where \( N = 2^n \) and \( \omega_N = e^{2\pi i / N} \) is a primitive \( N \)th root of unity. The \( N \times N \) unitary matrix \( F_N \) of the quantum Fourier transform is given by

\[
U_{\text{FT}} = \frac{1}{\sqrt{N}} \begin{bmatrix}
1 & 1 & 1 & 1 & \cdots & 1 \\
1 & \omega_N & \omega_N^2 & \omega_N^3 & \cdots & \omega_N^{N-1} \\
1 & \omega_N^2 & \omega_N^4 & \omega_N^6 & \cdots & \omega_N^{2(N-1)} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & \omega_N^{N-1} & \omega_N^{2(N-1)} & \omega_N^{3(N-1)} & \cdots & \omega_N^{(N-1)(N-1)}
\end{bmatrix}.
\]

The Quantum Fourier transform is related but not identical\(^{12} \) to the Hadamard transform. Quantum Fourier transform can be carried out by a quantum circuit built entirely out of 1-qubit and 2-qubit gates.\(^{13} \)

Measurement followed by a Hadamard Transform

Measuring the right part of the register \( \frac{1}{2^{n/2}} \sum_{x=0}^{2^n-1} |x, f(x)\rangle \) projects the state of the left register to

\[
|\psi_0\rangle = \frac{1}{\sqrt{m}} \sum_{k=0}^{m-1} |x_0 + kr\rangle
\]

where \( f(x_0) \) is the measured value. Here \( x_0 \) is the smallest value of \( x \) \((0 \leq x_0 < r)\) for which \( f(x_0) = f_0 \), and \( m \) is the smallest integer for which \( mr + x_0 \geq 2^n \).

\[
U_{\text{FT}} |\psi_0\rangle = \frac{1}{\sqrt{N}} \sum_{y=0}^{N-1} \frac{1}{\sqrt{m}} \sum_{k=0}^{m-1} \omega_N^{y(x_0 + kr)} |y\rangle
\]

\[
= \sum_{y=0}^{N-1} \omega_N^{y x_0} \frac{1}{\sqrt{Nm}} \sum_{k=0}^{m-1} \omega_N^{y kr} |y\rangle
\]
This lecture discusses entanglement (quantum vs. classical correlations), hidden variables theories, quantum non-locality, and Bell’s inequalities.

**Alice and Bob Share an EPR Pair**

Consider a bipartite system consisting of two entangled qubits whose joint state is

\[ |\varphi\rangle = \frac{1}{\sqrt{2}} (|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B) \]

Qubit A (Pixie) is given to Alice and qubit B (Dixie) to Bob. Note that \(|\varphi\rangle\) is a Bell state (aka EPR pair) we discussed before. EPR stands for Einstein, Podolsky and Rosen, who were the first to point out the “strange” properties of this state in 1935. Recall that each individual qubit is in the mixed state \((|0\rangle + |1\rangle)/\sqrt{2}\). We have shown earlier that if Alice measures her qubit in the computational basis, she will get either state \(|0\rangle\) or \(|1\rangle\), each with probability \(1/2\). But if Bob then measures his qubit in the computational basis, he will get a state that is identical to Alice’s. Einstain referred to this phenomenon as “spooky action at the distance” or quantum non-locality.

Einstein was not comfortable with the notion of non-deterministic measurements and entanglement. He believed that there exist some “hidden variables” that determine measurement outcomes, and in general govern the reality. He did not question the predictions of quantum mechanics, but declared it *incomplete* since it does not take into account existence of hidden variables that could explain the spooky actions at the distance.

Until John Bell’s work in 1964, no circumstances were known where predictions provided by any theory with hidden variables disagreed with those provided by quantum mechanics. John Bell came up with scenarios where these predictions were not identical, and thus which one is true could be determined by experiments. In the past half a century, many such experiments were conducted, but it was only in 2015 that experiments showed non existence of hidden variables in a most complete manner possible.

We will next go over a common example (involving the state \(|\varphi\rangle\) above) that shows a disagreement between the predictions provided by quantum mechanics and those provided by a hidden variable theory.
Two Measurement Scenarios

1. Suppose Alice measures her qubit in some orthonormal basis \{b_0, b_1\}, and gets \(|b_0\rangle\). This measurement result implies that Alice’s post-measurement state is her original state multiplied by \(|b_0\rangle\langle b_0|\) (and properly normalized), while the action on Bob state is described by the identity operator \(I\). What happens to state \(|\varphi\rangle\)? To see that \(|\varphi\rangle\) collapses to \(|b_0\rangle\), consider the following:

\[
(|b_0\rangle\langle b_0| \otimes I) |\varphi\rangle = \langle b_0|\langle b_0| \otimes |0\rangle + \langle b_0|\langle 1| \otimes |1\rangle
\]

\[
= |b_0\rangle \otimes (\langle b_0|\langle 0| + \langle b_0|\langle 1| \otimes |0\rangle)
\]

\[
= |b_0\rangle \otimes |b_0\rangle
\]

We have used here the fact that \(|b_0\rangle = \langle b_0|\langle 0| + \langle b_0|\langle 1| \otimes |0\rangle\).

2. Suppose Alice measures her qubit in an orthonormal basis \{b_0, b_1\} obtained by rotating the computational basis by 120 degrees, that is, \(|b_0\rangle = \frac{1}{2} |0\rangle + \frac{1}{2} |1\rangle\). Then the probability that Alice’s qubit collapses to \(|b_0\rangle\) is \(|\langle b_0|\langle 0| \otimes |0\rangle|^2 = 1/4\).

A Measurement Protocol

We consider a protocol in which Alice and Bob can perform measurements in three possible orthonormal bases: \{a_0, a_1\}, \{b_0, b_1\}, or \{c_0, c_1\} that are 120 degrees apart, see Fig. 1. The measurement equipment is connected to a light indicator which shows a red light when the result of the measurement is either \(|a_0\rangle\) or \(|b_0\rangle\) or \(|c_0\rangle\), and a green light when the result of the measurement is either \(|a_1\rangle\) or \(|b_1\rangle\) or \(|c_1\rangle\).

Note that 1 and 2 imply that if Alice and Bob measure their entangled qubits in the identical bases they get identical results with probability \(1\), and if they measure their entangled qubits in the bases that are 120 degrees apart, they get identical results with probability \(1/4\).

Figure 1: Alice and Bob randomly pick a basis by rolling a 3-sided die.
roll their respective dice. These rolls are independent of each other and of the previous rolls. Based on their rolls outcomes, Alice and Bob choose one of the three possible basis and measure their respective qubits, and observe the light indicator. How often will Alice and Bob see the lights of the same color?

Note that Alice and Bob will chose the same basis with probability 1/2, and a pair of bases that are 120 degrees apart with probability 2/3. Whenever they choose the same basis, they will see the light of the same color with probability 1. Whenever they choose different bases, they will see the the same color with probability 1/4. Therefore, Alice and Bob will see the same color with probability

$$\frac{1}{3} \cdot 1 + \frac{2}{3} \cdot \frac{1}{4} = \frac{1}{2}$$

Is such outcome a result of some “spooky action at a distance” or some “hidden variables” e.g., a pre-determined light color response to each of the three basis which is somehow hidden within the particles?

**Talking Mice**

Can the result of the above measurement protocol be explained by assuming that the particles had agreed in advance which light will be turned on for each basis? For example, Pixie and Dixie of Fig. 2 could agree that for a measurement in bases \(\{|a_0\rangle, |a_1\rangle\}\) and \(\{|b_0\rangle, |b_1\rangle\}\), they will always turn the red light on, and for a measurement in basis \(\{|c_0\rangle, |c_1\rangle\}\), they will always turn the green light on, as in Fig. 2.

Note that there are 8 possible agreements, and 9 possible Alice and Bob basis pairs, as shown in Table 1. For each basis pair, we can check if a particular agreement will result in Alice and Bob observing the same (S) or different (D) light colors. Table 1 shows that no agreement (row in the table) results in Alice and Bob observing the same light color 50% of the time.

**Table 1:** Alice and Bob can observe the same (S) or different (D) light colors depending on their choice of the basis pair and the Pixie-Dixie agreement.

---

**Figure 2:** Pixie and Dixie agree in advance on which light will be turned on for each of the three possible bases.
That’s All Folks