QUANTUM COMPUTING SYSTEMS

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 (qubits). Noisy Intermediate-Scale Quantum (NISQ) technology is expected to be available in the near future. This term, coined by John Preskill of CalTech, refers to devices with 50-100 qubits (intermediate-scale), which is too few to have have full error-correction (noisy). Nevertheless, NISQ systems may be able to perform tasks that exceed the capabilities of today's classical digital computers, and may be useful tools for exploring many-body quantum physics. On the theoretical side, significant progress has been made in understanding the fundamental limits of quantum telecommunications systems, giving rise to the subfield of quantum information theory. Moreover, classical information theory has been used to understand the problems in the foundations of physics.

Learning Objective:

The students will learn quantum computing basics and the fundamentals of 1) algorithms for for NISQ technology, 2) quantum telecommunications, and 3) many-body entangled systems, as well as a selected number of more advanced topics of their individual interests.

Instructor: Emina Soljanin (contact info on the web page).

Office hours: By appointment.

Class time and place: Mon & Thu, 8:40 – 10:00 AM, on Webex through 1/28/21, TBD afterwards.

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.

Course notes: given per lecture in separate documents on the class (Sakai) web page.

Recommended reading:

J. D. Hidary, *Quantum Computing: An Applied Approach*, Springer (2019).
M. M. Wilde, *Quantum Information Theory*, Cambridge Univ. Press (2017).
N. D. Mermin, *Quantum Computer Science: An Introduction*, Cambridge U. Press (2007).
L. Susskind and A. Friedman, *Quantum Mechanics: The Theoretical Minimum*.
J. Preskill, *Lecture Notes for Physics 229: Quantum Information and Computation.*
F. W. Byron and R. W. Fuller, *Mathematics of Classical and Quantum Physics*.

Grading: (weekly) quizzes 60%, final take-home exam 20%, project 20%.

Comparison with the Fall course: This course also starts with providing answers to the three essential questions that any newcomer to quantum computing needs to know: How is quantum information represented? How is quantum information processed? How is classical information extracted from quantum states? The Spring course then moves to selected topics in quantum computing, communications, and multi-particle systems. After the initial topics, the Fall course covers selected quantum algorithm.

Quantum Computing Systems [1]

Prof. Emina Soljanin

Lecture #1, January 25

[1] Rutgers, ECE 579, Spring 2021

This lesson introduces qubits and single qubit gates. It addresses the question how quantum information is represented and process.

Classical Information and Bits

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. The assumption here is that the question concerns something we have no prior knowledge about. 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?

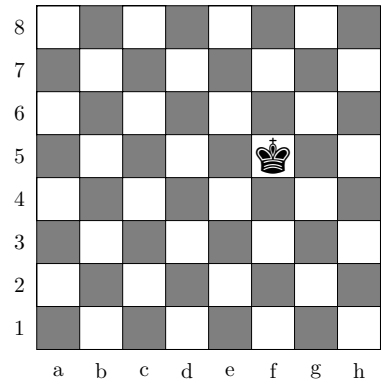


Figure 1: What is the minimum number of yes/no questions that have to be asked to locate the king on a chessboard?

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" for 0 or "down" for 1. Flesh memory cells made from floating-gate transistors act as switches that could be open for 0 or closed for 1. (There are multi-level cell devices that can store more than one bit per cell.)

A physical system with $N = 2^k$ distinguishable physical states can represent k bits of information. Such a system can simply be a collection of k systems with two distinguishable states, i.e., a k-bit register. To specify an object in a set of N, we need $\lceil \log_2 N \rceil$ binary digits.

Operations on Bits and Gates

In this class, we will treat bits as mathematical objects.[2] For us, bits take values in the set $\{0, 1\}$ where we can add and multiply as follows:

[2] Other classes at ECE and Physics study bits as physical systems.

XOR

$\oplus$	0	1
0	0	1
1	1	0

AND

$\cdot$	0	1
0	0	0
1	0	1

Figure 2: Binary arithmetics in $\mathbb{F}_2$.

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 we will formally define below.

Quantum Information and Qubits

A *qubit* is a quantum information/computing counterpart to a bit. We will treat qubits as mathematical objects as well.[3] What we learn in this class is independent of a particular physical realization.

[3] Qubits (as bits) are represented by physical systems.

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

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

then a single qubit $|\psi\rangle$ is mathematically a linear combination of $|0\rangle$ and $|1\rangle$ basis vectors:

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

where α and β are complex numbers such that $|\alpha|^2 + |\beta|^2 = 1$. Observe that the coefficients α and β depend on the choice of the basis. What would these coefficients be if the basis vectors were

$$|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2} \text{ and } |-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$$

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[4]. We say that the quantum state $|\psi\rangle$ above is a superposition of the two basis states. The superposition is instrumental in enabling quantum computing speedup.

[4] In the simplest case, qubit states are *pure* and we mathematically describe them as we described $|\psi\rangle$ here. There are also *mixed* states and a general way to mathematically represent both.

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 ψ and φ, such that [5]

[5] The conjugate of a complex number $c = x + iy$ is $c^* = x - iy$.

1. $\langle\psi|\varphi\rangle = \langle\varphi|\psi\rangle^*$
2. $\langle\alpha\psi + \beta\xi|\varphi\rangle = \alpha\,\langle\psi|\varphi\rangle + \beta\,\langle\xi|\varphi\rangle$, where $\alpha, \beta \in \mathbb{C}$.
3. $\langle\psi|\psi\rangle \geqslant 0$ for any ψ 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.

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* [6] by $\langle\varphi|$. 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[7]

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

Here a column vector times a row vector gives a matrix.

We have used 0 and 1 as labels for the basis in $\mathbb{C}^2$ above:

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

There are other labels in use, e.g., $|+\rangle$ and $|-\rangle$ or $|\downarrow\rangle$ and $|\uparrow\rangle$, and even *dead* and *alive* cats.

[6] $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$
$\implies \langle\varphi| = \alpha^*\langle 0| + \beta^*\langle 1|$

[7] the outer product

Math Interlude - Unitary Matrices

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

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

$U^\dagger$ 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.

Reversible Acting on a Single Qubit

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 2×2 unitary[8] 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:

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

[8] If U is real, we call it is *orthogonal*.

Some Single-Qubit Gates

Mapping the basis states

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

is obtained by matrix multiplication.

- Identity: $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
- Hadamard gate:

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad \begin{aligned} |0\rangle \text{ —}\boxed{H}\text{— } (|0\rangle + |1\rangle)/\sqrt{2} \\ |1\rangle \text{ —}\boxed{H}\text{— } (|0\rangle - |1\rangle)/\sqrt{2} \end{aligned}$$

- Pauli matrices:

$$\sigma_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \begin{aligned} |0\rangle \text{ —}\boxed{X}\text{— } |1\rangle \\ |1\rangle \text{ —}\boxed{X}\text{— } |0\rangle \end{aligned}$$

$$\sigma_Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad \begin{aligned} |0\rangle \text{ —}\boxed{Y}\text{— } i|1\rangle \\ |1\rangle \text{ —}\boxed{Y}\text{— } -i|0\rangle \end{aligned}$$

$$\sigma_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad \begin{aligned} |0\rangle \text{ —}\boxed{Z}\text{— } |0\rangle \\ |1\rangle \text{ —}\boxed{Z}\text{— } -|1\rangle \end{aligned}$$

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.

Any 2×2 complex matrix A (and thus any unitary matrix) can be expressed as a linear combination of the identity I and the Pauli matrices σ_X, σ_Y, and σ_Z:

$$A = \alpha_I I + \alpha_X \sigma_X + \alpha_Y \sigma_Y + \alpha_Z \sigma_Z$$

for some complex numbers α_I, α_X, α_Y, and α_Z.

Problem Set #1:

1. Show that the single qubit gates defined above are indeed unitary.
2. Express the Hadamard matrix H as a linear combination of of the identity I and the Pauli matrices σ_X, σ_Y, and σ_Z.
3. Verify that that the single qubit gates act on the basis vectors $|0\rangle$ and $|1\rangle$ as stated above.

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

Prof. Emina Soljanin

Lecture #2, January 28

This lecture defines the quantum measurement.

Math Interlude

Projection Matrices

A square matrix Π is a projection matrix iff $\Pi^2 = \Pi$. Note that a projection is a linear transformation from a vector space to itself. Recall that $|\varphi\rangle\langle\varphi|$ is a matrix. We say that $|\varphi\rangle\langle\varphi|$ is a rank-1 projection matrix. (Higher rank projection matrices project vectors onto subspaces.)

A projection Π on a Hilbert space $\mathcal{H}$ is an orthogonal projection iff it satisfies $\langle \Pi x, y\rangle = \langle x, \Pi y\rangle \langle \Pi x, y\rangle = \langle x, \Pi y\rangle$ for all $x, y \in \mathcal{H}$. Vector $|\varphi\rangle\langle\varphi| \cdot |\psi\rangle$ is the orthogonal[2] projection of vector $|\psi\rangle$ on vector $|\varphi\rangle$.

[2] To check for orthogonality, consider $\langle\varphi|\left(|\psi\rangle - |\varphi\rangle\langle\varphi| \cdot |\psi\rangle\right)$.

Orthogonal projections on the vectors that form a basis sum to the identity matrix. For example,

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

Eigenvectors and Eigenvalues

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

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

where λ_u is known as the eigenvalue of H corresponding to $|u\rangle$.

Hermitian Matrices

A Hermitian matrix H (or self-adjoint matrix) is a complex square matrix that is equal to its own conjugate transpose $H^\dagger$, i.e., the element in the i-th row and j-th column is equal to the complex conjugate[3] of the element in the j-th row and i-th column, for all indices i and j:

[3] The conjugate of a complex number $c = x + iy$ is $c^* = x - iy$.

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

We call real Hermitian matrices *symmetric*.

Claim:[4] Matrix H is Hermitian if and only if $\langle x|Hx\rangle$ is real for all $|x\rangle$.

[4] Very frequently useful!

If follows that the eigenvalues of a Hermitian operator are real. Why?

Hermitian and unitary matrices are normal[5]. If A is normal, then its eigenvectors corresponding to distinct eigenvalues are orthogonal. For a Hermitian matrix H, there exists a unitary matrix U such that $U^\dagger HU$ is a diagonal matrix:

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

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

$$\left[H|u_1\rangle \ldots H|u_m\rangle\right] = \left[\lambda_1|u_1\rangle \ldots \lambda_m|u_m\rangle\right]$$

and thus $|u_1\rangle, \ldots, |u_m\rangle$ are eigenvectors of H and $\lambda_1, \ldots, \lambda_m$ are the corresponding eigenvalues. Since $|u_1\rangle, \ldots, |u_m\rangle$ are columns of a unitary matrix, they form a basis of $\mathcal{H}^m$. Therefore,

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

How is Classical Information Extracted?

Extraction of classical information from quantum states is connected to a postulate of quantum mechanics which says that 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 normalized (unit-norm) eigenvectors of its Hermitian matrix.

Quantum Observables

The measurement of an observable H always indicates an eigenvalue of H and turns any measured quantum state into the eigenstate of H corresponding to the indicated eigenvalue. The measured state only gives rise to a probability distribution on the set of outcomes, as we explain next.

Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of an $m \times m$ Hermitian matrix H and $|u_1\rangle, \ldots, |u_m\rangle$ be the corresponding eigenvectors. (We assume, for the moment, that all λ_i are different.) Since H is hermitian, we have

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

[5] Matrix A is normal iff $AA^\dagger = A^\dagger A$

Figure 1: What is the color of the shoe? What is the color of the shoelace?

A set of vectors $|u_1\rangle, \ldots, |u_m\rangle$ that satisfies the above two conditions is said to form a *resolution of the identity*. We refer to $|u_1\rangle, \ldots, |u_m\rangle$ as the measurement basis, and say that we *perform a measurement in the basis* or measure in the basis $|u_1\rangle, \ldots, |u_m\rangle$.

Let $|\psi\rangle$ be a state being measured by the observable described by H. Then the measurement result is λ_i and $|\psi\rangle$ collapses to $|u_i\rangle$ with probability (wp) $|\langle\psi|u_i\rangle|^2$, $1 \leqslant i \leqslant N$, as sketched in Fig. 2.

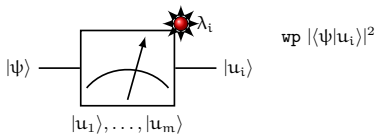


Figure 2: Quantum measurement: The only possible results of "measuring H" are its eigenvalues λ_i, and the only possible states after "measuring H" are its normalized eigenvectors $|u_i\rangle$. When we "see" λ_i (which happens wp $|\langle\psi|u_i\rangle|^2$ when state $|\psi\rangle$ is measured), we know that the state being measured $|\psi\rangle$ has collapsed to $|u_i\rangle$.

Example – Measuements defined by bases: What can we get if we measure qubit $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ in the computational basis $|0\rangle$, $|1\rangle$? What if we use the $|+\rangle$, $|-\rangle$ basis instead?

Example – Measurements defined by Pauli matrices Pauli Matrices are both unitary and Hermitian, and thus can serve to define both quantum gates and quantum measurements. Their eigenvalues with the corresponding eigenvectors are shown in Fig 3.

matrix	action	eigenvalue & eigenvector
$\sigma_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	$\|0\rangle$ — X — $\|1\rangle$	$+1, \quad (\|0\rangle + \|1\rangle)/\sqrt{2}$
	$\|1\rangle$ — X — $\|0\rangle$	$-1, \quad (\|0\rangle - \|1\rangle)\sqrt{2}$
$\sigma_Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$	$\|0\rangle$ — Y — $i\|1\rangle$	$+1, \quad (\|0\rangle + i\|1\rangle)/\sqrt{2}$
	$\|1\rangle$ — Y — $-i\|0\rangle$	$-1, \quad (\|0\rangle - i\|1\rangle)/\sqrt{2}$
$\sigma_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	$\|0\rangle$ — Z — $\|0\rangle$	$+1/\|0\rangle$
	$\|1\rangle$ — Z — $-\|1\rangle$	$-1/\|1\rangle$

Figure 3: Pauli matrices and their eigenvalues with the corresponding normalized eigenvectors.

Mathematical Description of the Quantum Measurement

We have seen above that a measurement on an n–qubit state is defined by a set of $N = 2^n$ basis vectors $|u_i\rangle$, $1 \leqslant i \leqslant N$. When state $|\psi\rangle$ enters the measuring apparatus, it collapses to the state $|u_i\rangle$ wp $|\langle\psi|u_i\rangle|^2$. If we denote by Π_i the rank–1 projection on $|u_i\rangle$, we can equivalently say the the measurement is defined by the set of N orthogonal rank–1 projections $\Pi_i = |u_i\rangle\langle u_i|$ and the measured state $|\psi\rangle$ collapses to $\frac{1}{\sqrt{\langle\psi|\Pi_i|\psi\rangle}}\Pi_i|\psi\rangle$ wp $\langle\psi|\Pi_i|\psi\rangle$. We can now easily generalize our basis defined measurement by removing the requirement that the projections Π_i be rank–1.

Von Neumann Measurement

An observable described by a Hermitian $N \times N$ matrix H may have $m \leqslant N$ different eignvalues. Let Π_i be the projection operator on the eigenspace i of H. The von Neumann projective measurement is defined as follows:

- A set of pairwise orthogonal projection operators $\{\Pi_i\}$ such that $\sum_i \Pi_i = I$.
- For input $|\psi\rangle$, output i happens wp $\langle\psi|\Pi_i|\psi\rangle$, and $|\psi\rangle$ collapses to $\frac{1}{\sqrt{\langle\psi|\Pi_i|\psi\rangle}}\Pi_i|\psi\rangle$.

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 α and β. 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 posses an infinite amount of information? You would if you could read out the values of α and/or β. 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?

Figure 4: 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.

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

Prof. Emina Soljanin

Lecture #3, February 1

This lecture is concerned with multiple qubits and reversible actions on single and multiple qubits.

Math Interlude

Let A be an $m \times n$ matrix[2] and B a $p \times q$ matrix. The Kronecker product $A \otimes B$ is the $mp \times nq$ matrix given by

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

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

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 is distributive over the Kronecker product:

$$(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$$

Compare the expressions for the transpose and inverse of the Kronecker product of matrices with their counterparts for the regular product of matrices?

- $A \otimes B$ has the inverse iff both A and B are invertible, and then

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

How is Quantum Information Represented?

Representation of quantum information is connected to a postulate of quantum mechanics which says that associated to any isolated physical system is a complex vector space with inner product. In this class, and quantum computing in general, we mostly deal with finite dimensional spaces $\mathbb{C}^N$ and often conventionally refer to them as Hilbert spaces $\mathcal{H}_N$.

Single Qubit

Mathematically, independently of a particular physical realization, a qubit is represented by a unit-norm[3] vector in the two-dimensional

[3] The norm is induced by the inner product.

unitary space $\mathbb{C}^2$. If we denote the basis vectors of this space by

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

then a single qubit $|\psi\rangle$ is mathematically a linear combination of $|0\rangle$ and $|1\rangle$, that is

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

where α and β are complex numbers such that $|\alpha|^2 + |\beta|^2 = 1$.

Multiple Qubits

Consider following two qubits: $|\psi_1\rangle = \alpha_1|0\rangle + \beta_1|1\rangle$ and $|\psi_2\rangle = \alpha_2|0\rangle + \beta_2|1\rangle$. The joint state of the pair is the Kronecker product of the individual states:

$$\begin{aligned} |\psi_1\rangle \otimes |\psi_2\rangle &= \big(\alpha_1|0\rangle + \beta_1|1\rangle\big) \otimes \big(\alpha_2|0\rangle + \beta_2|1\rangle\big) \\ &= \alpha_1\alpha_2\,|0\rangle \otimes |0\rangle + \alpha_1\beta_2\,|0\rangle \otimes |1\rangle + \\ &\quad \beta_1\alpha_2\,|1\rangle \otimes |0\rangle + \beta_1\beta_2\,|1\rangle \otimes |1\rangle \end{aligned}$$

where the vectors

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

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

form a basis for $\mathbb{C}^4 = \mathbb{C}^2 \otimes \mathbb{C}^2$. In general[4], 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*[5] states.

[4] separable and entangled stats

[5] Entangled states are responsible for much of "quantum magic".

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

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

This state is known as the Bell state or the EPR pair. [6]

[6] EPR stands for Einstein, Podolsky and Rosen, who were the first to point out the "strange" properties of this state.

An n–qubit state $|\phi\rangle$ is a unit-norm vector in $\mathbb{C}^{2^n}$, which we commonly refer to as the Hilbert space $\mathcal{H}_{2^n} = \mathcal{H}_2^{\otimes n} = \underbrace{\mathcal{H}_2 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_2}_{n}$. We will also often use the common notation $N = 2^n$. For an n–qubit state $|\phi\rangle \in \mathcal{H}_{2^n}$, we have

$$|\phi\rangle = \sum_{i=0}^{2^n-1} \alpha_i |i_0 i_1 \dots i_{n-1}\rangle, \qquad \sum_{i=0}^{2^n-1} |\alpha_i|^2 = 1,$$

where the binary string $i_0 i_1 \dots i_{n-1}$ is the binary representation of i, and $|i_0 i_1 \dots i_{n-1}\rangle$ is a shorthand notation for $|i_0\rangle \otimes |i_1\rangle \otimes \cdots \otimes |i_{n-1}\rangle$ (the i-th basis vector of $\mathcal{H}_{2^n}$). Other commonly used shorthand notation is

$$\begin{aligned} |i_0\rangle \otimes |i_1\rangle \otimes \cdots \otimes |i_{n-1}\rangle &\equiv |i_0\rangle|i_1\rangle \dots |i_{n-1}\rangle \\ &\equiv |i_0, i_1, \dots, i_{n-1}\rangle \\ &\equiv |i_0 i_1 \dots i_{n-1}\rangle. \end{aligned}$$

There is a notion of a *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.

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.

How is Quantum Information Processed?

Processing of quantum information is connected to a postulate of quantum mechanics which says that the evolution of a closed quantum system is described by a unitary[7] transformation. Therefore, in a closed quantum system, a qubit state $|\psi\rangle \in \mathcal{H}$ can be transformed to some other state in $\mathcal{H}$, say $|\varphi\rangle$, only by some unitary operator U, that is,

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

where U is a 2×2 unitary matrix over $\mathbb{C}$. Note that quantum evolution is reversible. 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 $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ since the evolution (matrix multiplication) is a linear operation, and thus

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

Unitary action U maps the computational basis $|0\rangle$, $|1\rangle$ into the basis $U|0\rangle$, $U|1\rangle$.

Actions on an n–qubit state are described by $2^n \times 2^n$ unitary matrices, which may or may not be Kronecker products of matrices of smaller dimensions. When $U = U_0 \otimes U_1 \otimes \cdots \otimes U_{n-1}$, where U_i is a 2×2 unitary matrix, then its action on the basis vector $|i_0\rangle \otimes |i_1\rangle \otimes \cdots \otimes |i_{n-1}\rangle \in \mathcal{H}_{2^n}$ is given by

$$U|i_0 i_1 \dots i_{n-1}\rangle = \boxed{U_0|i_0\rangle} \otimes \boxed{U_1|i_1\rangle} \otimes \cdots \otimes \boxed{U_{n-1}|i_{n-1}\rangle}$$

Single Qubit Gates

In classical computing, `NOT` is the only single bit gate, that is, in addition to the `I` "gate" (identity). In quantum computing, any 2×2

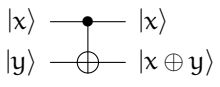
[7] Unitary evolution in a closed quantum system is a consequence of the Schrödinger equation.

unitary matrix specifies a single-qubit gate. The most commonly used single-qubit gates are the Pauli and Hadamard matrices, which we worked with in the previous classes.

Two Qubit Gates

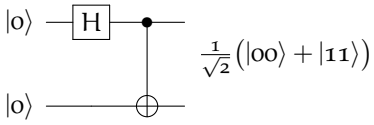
The two-qubit quantum gate known as quantum `XOR` or controlled-not gate `CNOT` is specified as a map and a circuit as follows:

$$\texttt{CNOT} : |x, y\rangle \to |x, x \oplus y\rangle \qquad x, y \in \{0, 1\}$$



An example ...

We can use the Hadamard and the `CNOT` gates to create entanglement:



The No-Cloning Theorem

The requirement that any evolution be unitary gives rise to the famous no-cloning theorem, which asserts that there is no unitary operator U_c on $\mathcal{H} \times \mathcal{H}$ that takes state $|\psi\rangle \otimes |\omega\rangle$ to $|\psi\rangle \otimes |\psi\rangle$ for all states $|\psi\rangle \in \mathcal{H}$ and some fixed state $\omega \in \mathcal{H}$.

To prove the no-cloning theorem, we suppose that there is a unitary matrix U_c such that for two arbitrary sates $|\psi\rangle$ and $|\varphi\rangle$, we have

$$\begin{aligned} U_c(|\psi\rangle \otimes |\omega\rangle) &= |\psi\rangle \otimes |\psi\rangle \\ U_c(|\varphi\rangle \otimes |\omega\rangle) &= |\varphi\rangle \otimes |\varphi\rangle \end{aligned}$$

where ω is some fixed state. Note the following identities:

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

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

2. Since U_c is unitary, that is $U_c^\dagger \cdot U_c = I$, then by the properties of the Kronecker product, we have

$$\begin{aligned} \langle\psi|\varphi\rangle &= \big(\langle\psi| \otimes \langle\omega|\big) \cdot \big(|\varphi\rangle \otimes |\omega\rangle\big) \\ &= \big(\langle\psi| \otimes \langle\omega|\big) U_c^\dagger \cdot U_c \big(|\varphi\rangle \otimes |\omega\rangle\big) \\ &= \big(\langle\psi| \otimes \langle\psi|\big) \cdot \big(|\varphi\rangle \otimes |\varphi\rangle\big) \\ &= \langle\psi|\varphi\rangle \otimes \langle\psi|\varphi\rangle = \langle\psi|\varphi\rangle^2 \end{aligned}$$

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

The no-cloning theorem is often misunderstood to be more restrictive than it is. Note that it does not prohibit the following map:

$$\underbrace{\alpha|0\rangle + \beta|1\rangle}_{\in\mathcal{H}_2} \quad \rightarrow \quad \underbrace{\alpha|000\rangle + \beta|111\rangle}_{\in\mathcal{H}_{23}}$$

Problem Set #2:

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

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

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

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

$$\underbrace{\alpha|0\rangle + \beta|1\rangle}_{\in\mathcal{H}_2} \quad \rightarrow \quad \underbrace{\alpha|000\rangle + \beta|111\rangle}_{\in\mathcal{H}_{23}}$$

   You are allowed to use additional fixed-state quantum systems. The operator can be a circuit consisting of gates you have seen in class.

4. Describe each of the following four vectors as linear combinations of either $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$ or $\langle 00|$, $\langle 01|$, $\langle 10|$, and $\langle 11|$:

$$\begin{bmatrix} \alpha \\ \beta \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} \alpha \\ \beta \\ 0 \\ 0 \end{bmatrix}^\dagger, \quad \begin{bmatrix} \alpha & \beta & \alpha & \beta \end{bmatrix}, \quad \begin{bmatrix} \alpha & \beta & \alpha & \beta \end{bmatrix}^\dagger, \text{ where } \alpha, \beta \in \mathbb{C}.$$

Quantum Computing Systems [1]

Prof. Emina Soljanin

Lecture #4, February 4

[1] Rutgers, ECE 579, Spring 2021

This lecture is concerned with quantum measurement and quantum parallelism.

Mathematical Description of the Quantum Measurement

We have seen above that a measurement on an n–qubit state is defined by a set of $N = 2^n$ basis vectors $|u_i\rangle$, $1 \leqslant i \leqslant N$. When state $|\psi\rangle$ enters the measuring apparatus, it collapses to the state $|u_i\rangle$ wp $|\langle\psi|u_i\rangle|^2$. If we denote by Π_i the rank–1 projection on $|u_i\rangle$, we can equivalently say the the measurement is defined by the set of N orthogonal rank–1 projections $\Pi_i = |u_i\rangle\langle u_i|$ and the measured state $|\psi\rangle$ collapses to $\frac{1}{\sqrt{\langle\psi|\Pi_i|\psi\rangle}}\Pi_i|\psi\rangle$ wp $\langle\psi|\Pi_i|\psi\rangle$. We can now easily generalize our basis defined measurement by removing the requirement that the projections Π_i be rank–1.

Von Neumann Measurement

An observable described by a Hermitian $N \times N$ matrix H may have $m \leqslant N$ different eignvalues. Let Π_i be the projection operator on the eigenspace i of H, $1 \leqslant i \leqslant m$. The von Neumann projective measurement is defined as follows:

- A set of pairwise orthogonal projection operators $\{\Pi_i\}$ such that $\sum_i \Pi_i = I$.
- For input $|\psi\rangle$, output i happens wp $\langle\psi|\Pi_i|\psi\rangle$, and $|\psi\rangle$ collapses to $\frac{1}{\sqrt{\langle\psi|\Pi_i|\psi\rangle}}\Pi_i|\psi\rangle$.

Figure 1: You may imagine quantum measurements as loaded dice whose bias is determined by the object being measured.

Von Neumann Measurement - Example

The measurement is defined by the computational basis vectors $|0\rangle$ and $|1\rangle$. The angle between $|\psi_0\rangle$ and $|0\rangle$ is $\pi/12$, and so is the angle between $|\psi_1\rangle$ and $|1\rangle$, as in Fig. 2.

Consider measuring two single-qubit states $|\psi_0\rangle$ and $|\psi_1\rangle$. The angle between these vectors is $2\pi/6$. No matter which state is measured, the resulting state after the measurement is either $|0\rangle$ or $|1\rangle$. If state $|\psi_0\rangle$ is measured, it will collapse either to state $|0\rangle$ with probability $|\langle\psi_0|0\rangle|^2 = \cos^2(\pi/12) = 1/2 + \sqrt{3}/4$ or to state $|1\rangle$ with probability $|\langle\psi_0|1\rangle|^2 = 1 - |\langle\psi_0|0\rangle|^2 = \sin^2(\pi/12) = 1/2 - \sqrt{3}/4$. We can make similar observations when state $|\psi_1\rangle$ is measured.

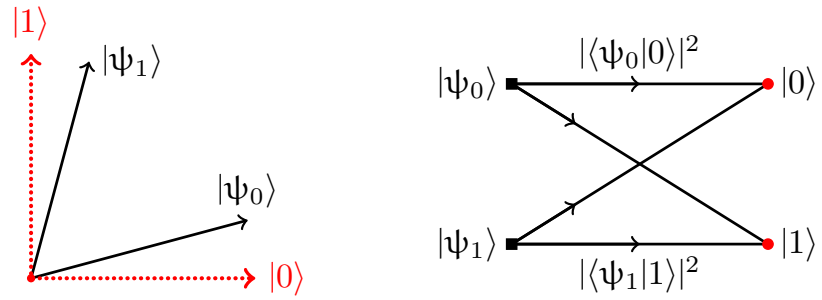


Figure 2: (left) States $|\psi_0\rangle$ and $|\psi_1\rangle$ with the bases $|0\rangle$, $|1\rangle$ used for a von Neumann measurement. (right) The states before and after the measurement with the possible transitions. (Some labels are omitted for clarity of the figure.)

Positive Operator-Valued Measure (POVM)

We can further generalize the von Neumann measurement of an n–qubit state $|\psi\rangle \in \mathcal{H}_{2^n}$ by observing that we can add an ancillary m–qubit state in $\mathcal{H}_{2^m}$ to $|\psi\rangle$ and perform a von Neumann measurement to the joint state in $\mathcal{H}_{2^n} \otimes \mathcal{H}_{2^m}$. If we restrict our attention[2] to $\mathcal{H}_{2^n}$, the measurement is defined as follows:

- Any set of positive-semidefinite operators $\{E_i\}$ such that $\sum_i E_i = I$.
- For input $|\psi\rangle$, output i happens wp $\langle\psi| E_i |\psi\rangle$, and $|\psi\rangle$ collapses to $\frac{1}{\sqrt{\langle\psi|E_i|\psi\rangle}} \Pi_i |\psi\rangle$.

[2] *Restricting our attention* to a part of the system is a formal mathematical notion.

Positive Operator Value Measure (POVM) - Example

The measurement is defined by the projections on vectors $|\varphi_0\rangle$, $|\varphi_1\rangle$, $|\varphi_2\rangle$. The angle between $|\varphi_0\rangle$ and $|\varphi_i\rangle$, $i = 1, 2$, is $2\pi/3$, and it is easy to see that properly normalized projections on these vectors form a resolution of the identity I_2. Vectors $|\psi_0\rangle$ and $|\varphi_1\rangle$ are orthogonal, and so are $|\varphi_0\rangle$ and $|\psi_1\rangle$, as in Fig. 3.

No matter which state is measured, the resulting state after the measurement is one of the states $|\varphi_0\rangle$, $|\varphi_1\rangle$, $|\varphi_2\rangle$. If state $|\psi_0\rangle$ is measured, it will collapse either to state $|\varphi_0\rangle$ with probability $|\langle\psi_0|\varphi_0\rangle|^2 = \cos^2(2\pi/6) = 1/4$ or to state $|\varphi_2\rangle$ with probability $|\langle\psi_0|\varphi_2\rangle|^2 = \cos^2(\pi/6) = 3/4$. Note that the probability of state $|\psi_0\rangle$ collapsing to $|\varphi_1\rangle$ is zero. We can make similar observations when state $|\psi_1\rangle$ is measured.

Measurements Defined by Pauli Matrices

Pauli Matrices are both unitary and Hermitian, and thus can serve to define both quantum gates and quantum measurements. Their eigenvalues with the corresponding eigenvectors are shown in Table 1.

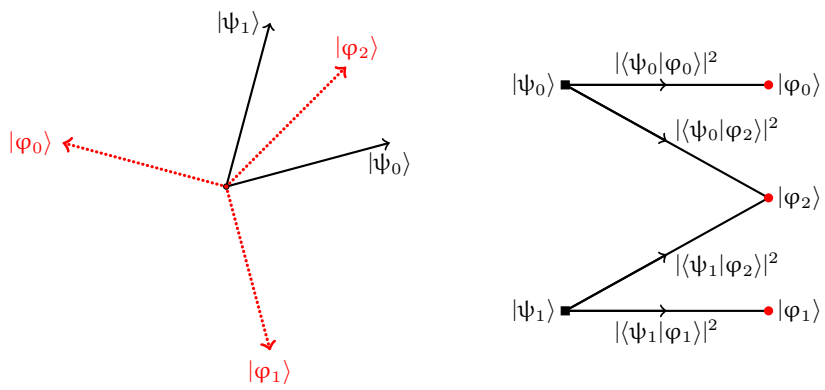


Figure 3: (left) States $|\psi_0\rangle$ and $|\psi_1\rangle$ with the vectors $|\varphi_0\rangle$, $|\varphi_1\rangle$, and $|\varphi_1\rangle$ used for a POVM. (right) The states before and after the measurement with the possible transitions.

matrix	eigenvalus and eigenvectors	
σ_X	$+1,\ (|0\rangle + |1\rangle)/\sqrt{2}$	$-1,\ (|0\rangle - |1\rangle)\sqrt{2}$
σ_Y	$+1,\ (|0\rangle + i|1\rangle)/\sqrt{2}$	$-1,\ (|0\rangle - i|1\rangle)/\sqrt{2}$
σ_Z	$+1,\ |0\rangle$	$-1,\ |1\rangle$

Table 1: Pauli matrices and their eigenvalues with the corresponding normalized eigenvectors.

The Expected Value of a Measurement

Regardless of which state $|\psi\rangle$ is being measured by the observable described by H, the only possible outcomes are the eigenvalues of H. The expected value of the measurement depends on $|\psi\rangle$ as follows:[3]

$$\begin{aligned}\sum_{i=1}^{N} \lambda_i |\langle\psi|u_i\rangle|^2 &= \sum_{i=1}^{N} \lambda_i \langle\psi|u_i\rangle\langle u_i|\psi\rangle \\ &= \langle\psi| \left(\textstyle\sum_{i=1}^{N} \lambda_i |u_i\rangle\langle u_i|\right) |\psi\rangle = \langle\psi| H |\psi\rangle\end{aligned}$$

where we have used the equality $H = \sum_{i=1}^{N} \lambda_i |u_i\rangle\langle u_i|$.

[3] Observe the convenience of the Dirac notation in this simple derivation.

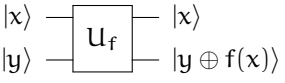
Creating Quantum Parallelism

We can evaluate an m-bit valued function f of an n-bit string by what is known as the *function evaluation gate*. The evaluation gate for a function

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

is described as follows:[4]

$$U_f : |x, y\rangle \to |x, y \oplus f(x)\rangle \qquad x \in \{0,1\}^n,\ y \in \{0,1\}^m$$



Note that U_f is a unitary operator acting on vectors in $\mathcal{H}_2^{\otimes n} \otimes \mathcal{H}_2^{\otimes m}$.

[4] U_{CNOT} is a special case of U_f.

We have seen above that the Hadamard gate action on $|0\rangle$ creates a uniform superposition of the computational basis states. It is easy to show that applying the n–qubit Hadamard product gate $H^{\otimes n}$ to $|0\rangle^{\otimes n}$ creates the uniform superposition of the computational basis of $\mathcal{H}_{2^n}$:

$$|0\rangle^{\otimes n} \; \boxed{H^{\otimes n}} \; \frac{1}{2^{n/2}} \sum_{x \in \{0,1\}^n} |x\rangle$$

Quantum function evaluation parallelism is achieved by first creating the uniform superposition of the computational basis of $\mathcal{H}_{2^n}$ and then applying the U_f unitary transform to simultaneously evaluate f on its entire domain, as follows:

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

If we could also simultaneously read all the evaluations (which we cannot), we would achieve a quantum speedup. We will see what is possible in the next section which describes how we can extract classical information from quantum states.

It is natural to wonder whether these probabilistic measurements can be useful. Recall that quantum parallelism allows us to evaluate $f : \{0,1\}^n \to \{0,1\}^m$ on its entire domain (see map (1)). But we have just seen that we cannot simultaneously extract all values by a single measurement. How is then quantum speedup achieved? Many quantum algorithms prescribe further processing of the state $\frac{1}{2^{n/2}} \sum_{x \in \{0,1\}^n} |x, f(x)\rangle$ so that, when a measurement is eventually performed, the probability of getting the answer to the posed question is (close to) 1. Moreover, the questions usually ask about some global property such as whether a function is balanced or constant or what is its period rather than the explicit function evaluation on its entire domain.

Problem Set # 3

1. Compute $|\varphi_0\rangle\langle\varphi_0| + |\varphi_1\rangle\langle\varphi_1| + |\varphi_2\rangle\langle\varphi_2|$, where
   $$|\varphi_0\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix},\ |\varphi_1\rangle = \begin{bmatrix} \sqrt{3}/2 \\ 1/2 \end{bmatrix},\ |\varphi_2\rangle = \begin{bmatrix} \sqrt{3}/2 \\ -1/2 \end{bmatrix}.$$
2. Show that $\sigma_Z \otimes \sigma_X$ defines an observable.
3. Suppose that state $(|00\rangle + |11\rangle)/\sqrt{2}$ is measured according to the observable $\sigma_Z \otimes \sigma_X$.
   (a) What are the possible measurement outcomes?
   (b) What is the probability of each outcome?
   (c) How would your answers change for the observable $\sigma_Z \otimes \sigma_Z$?

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

Prof. Emina Soljanin

Lecture #5, February 9

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.

Hadamard and `CNOT` Gates – Review

Hadamard gate is a single qubit gate:

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

$$|0\rangle \text{ —H— } (|0\rangle + |1\rangle)/\sqrt{2}$$

$$|1\rangle \text{ —H— } (|0\rangle - |1\rangle)/\sqrt{2}$$

`CNOT` gate is a two qubit gate:

$$\texttt{CNOT} : |x, y\rangle \to |x, x \oplus y\rangle$$

$$x, y \in \{0, 1\}$$

$$|x\rangle \text{ —●— } |x\rangle$$

$$|y\rangle \text{ —⊕— } |x \oplus y\rangle$$

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*[2] states.

[2] Entangled states are responsible for much of "quantum magic".

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

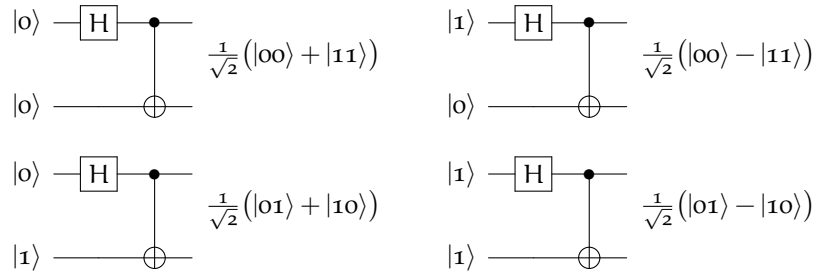


Figure 1: Creating Bell states by a 2-qubit entanglement gate.

The 4 entangled states in Fig. 1 are known as Bell[3] 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.

[3] We will learn more about John Bell and his inequalities later.

Entangled states have some "surprising" properties. To see that, we consider the EPR pair:[4]

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

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:*[5] 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.

[4] EPR stands for Einstein, Podolsky and Rosen, who were the first to point out the "strange" properties of this state.

[5] Einstein's phrase; he was not comfortable with the notion of non-deterministic measurements and entanglement.

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 α and/or β. He can only possibly apply some unitary transformation to $|\psi\rangle$ and then perform a measurement, which would give him at most one bit.

The most Alice can communicate to Bob by sending him a single qubit is a single bit of information, unless they share an EPR pair.

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

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

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:

$$\begin{aligned}
(I \otimes I)|\Psi\rangle &= \frac{1}{\sqrt{2}}\left(|0_A 0_B\rangle + |1_A 1_B\rangle\right)\\
(\sigma_X \otimes I)|\Psi\rangle &= \frac{1}{\sqrt{2}}\left(|1_A 0_B\rangle + |0_A 1_B\rangle\right)\\
(\sigma_Z \otimes I)|\Psi\rangle &= \frac{1}{\sqrt{2}}\left(|0_A 0_B\rangle - |1_A 1_B\rangle\right)\\
(\sigma_Z\sigma_X \otimes I)|\Psi\rangle &= \frac{1}{\sqrt{2}}\left(-|1_A 0_B\rangle + |0_A 1_B\rangle\right)
\end{aligned}$$

Note that Alice is able to create 4 orthogonal states.[6] If after per-

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

forming 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.,

$$\begin{aligned} 00: &\quad (I \otimes I) \\ 01: &\quad (\sigma_X \otimes I) \\ 10: &\quad (\sigma_Z \otimes I) \\ 11: &\quad (\sigma_Z \sigma_X \otimes I) \end{aligned}$$

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

Teleportation

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

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

and then Alice took qubit A and Bob took qubit B. Now, Alice has another qubit in the state[7]

$$|\psi_a\rangle = \alpha\,|0\rangle_a + \beta\,|1\rangle_a$$

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?

[7] We will use a (new) and A (entangled with Bob) subscripts to distinguish the two qubits on Alice's side.

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

$$\begin{aligned} |\psi_a\rangle\,|\Psi_{AB}\rangle &= \big(\alpha\,|0\rangle_a + \beta\,|1\rangle_a\big)\frac{1}{\sqrt{2}}\big(|0_A\rangle\,|0_B\rangle + |1_A\rangle\,|1_B\rangle\big) \\ &= \alpha\,|0\rangle_a\,\frac{1}{\sqrt{2}}\big(|0_A 0_B\rangle + |1_A 1_B\rangle\big) + \beta\,|1\rangle_a\,\frac{1}{\sqrt{2}}\big(|0_A 0_B\rangle + |1_A 1_B\rangle\big) \end{aligned}$$

The following protocol , known as teleportation, results in Bob's qubit (member of the entangled pair) assuming the state $|\psi\rangle$:[8]

[8] Is teleportation cloning?

1. Alice first applies a `CNOT` gate to her two qubits

$$|x\rangle_a\,|x_A\rangle \to |x_a\rangle\,|x_a \oplus x_A\rangle$$

and the 3-qubit state becomes

$$|\Phi_{aAB}\rangle = \alpha\,|0\rangle_a\,\frac{1}{\sqrt{2}}\big(|0_A 0_B\rangle + |1_A 1_B\rangle\big) + \beta\,|1\rangle_a\,\frac{1}{\sqrt{2}}\big(|1_A 0_B\rangle + |0_A 1_B\rangle\big)$$

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

$$\begin{aligned}(H \otimes I \otimes I)\,|\Phi_{aAB}\rangle &= \alpha\frac{1}{\sqrt{2}}(|0\rangle_a + |1\rangle_a)\frac{1}{\sqrt{2}}\big(|0_A 0_B\rangle + |1_A 1_B\rangle\big)\\ &+ \beta\frac{1}{\sqrt{2}}(|0\rangle_a - |1\rangle_a)\frac{1}{\sqrt{2}}\big(|1_A 0_B\rangle + |0_A 1_B\rangle\big)\\ &= \frac{1}{2}|00\rangle_{aA}\left(\alpha|0\rangle_B + \beta|1\rangle_B\right) + \frac{1}{2}|01\rangle_{aA}\left(\alpha|1\rangle_B + \beta|0\rangle_B\right)\\ &+ \frac{1}{2}|10\rangle_{aA}\left(\alpha|0\rangle_B - \beta|1\rangle_B\right) + \frac{1}{2}|11\rangle_{aA}\left(\alpha|1\rangle_B - \beta|0\rangle_B\right)\end{aligned}$$

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:

$$\begin{aligned}\alpha|0\rangle_B + \beta|1\rangle_B &= I|\psi\rangle\\ \alpha|1\rangle_B + \beta|0\rangle_B &= \sigma_X|\psi\rangle\\ \alpha|0\rangle_B - \beta|1\rangle_B &= \sigma_Z|\psi\rangle\\ \alpha|1\rangle_B - \beta|0\rangle_B &= \sigma_Z\sigma_X|\psi\rangle\end{aligned}$$

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[9] 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 σ_X or σ_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.

[9] by the entanglement magic

Observe that there is only one copy[10] 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.

[10] Teleportation is not cloning.

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

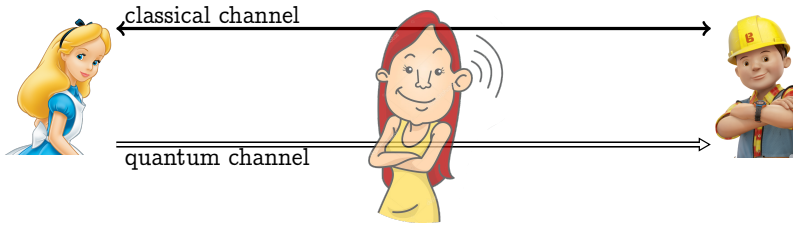
Prof. Emina Soljanin

Lecture #6, February 11

This lecture describes two protocols for quantum key distribution (QKD).

Traditional data encryption methods, based on using public keys, are threatened by the advances in quantum computing algorithms promising to efficiently solve so far intractable problems that make public key encryption currently secure. However, it is precisely quantum information processing advances that are also expected to enable secure communications by allowing efficient and secure private key distribution. The main advantage of private key encryption is that as long as the key strings are truly secret, it is provably secure, that is, insensitive to advances in computing.

A Quantum Key Distribution (QKD) protocol describes how two parties, commonly referred to as Alice and Bob, can establish a secret key by communicating over a quantum and a public classical channel when both channels can be accessed by an eavesdropper Eve.



The basic observation behind QKD protocols is that, since Eve cannot clone qubits, she can only gain information by measuring the original qubit. Therefore, when non-orthogonal qubits are transmitted from Alice to Bob, then Eve cannot gain any information from the qubits without disturbing their states, thus alerting Alice and Bob of her presence. We next describe two important QKD protocols. Substantial progress has been made towards building practical schemes based on these protocols.

BB84 Protocol

The BB84 was developed by Bennett and Brassard in 1984, hence the name. We outline the steps that Alice and Bob make under this protocol in order to generate a secret key of $O(n)$ bits for an arbitrary integer n.

1. Alice creates a sequence of $(4+\delta)n$ random data bits B_A, which she will map into qubits for transmission over the quantum channel between her and Bob.

2. For each data bit, Alice tosses a fair coin. If she gets a tail (T), she maps her data bit into either $|0\rangle$ (if her data bit is 0) or $|1\rangle$ (if her data bit is 1). If she gets a head (H), she maps her data bit into either $|+\rangle$ (if her data bit is 0) or $|-\rangle$ (if her data bit is 1).

   We will refer to the sequence of heads and tails that Alice generated as C_A, and to the sequence of qbits she prepares as Q_A. We will call $\{|0\rangle, |1\rangle\}$ the T basis and $\{|-\rangle, |+\rangle\}$ the H basis, according to the corresponding coin faces.

3. Alice sends the resulting $(4+\delta)n$ qubits to Bob over their public quantum communication channel. Each qubit may be altered by the noise in the channel and/or measured by Eve. Note that, at this point, Eve has no knowledge of C_A and thus what measurement basis she should use for an intercepted qubit in order to learn the corresponding bit. She can only guess the preparation basis for a qubit, and if her guess is wrong, she will alter its state, thus leaving a proof of eavesdropping.

4. Upon receiving a qubit, Bob then tosses a fair coin and then, depending on the toss outcome, he measures the qubit in either the H or the T basis. If he uses the T bases and gets $|0\rangle$, or the H bases and gets $|+\rangle$, he records bit 0; otherwise he records bit 1.

   We will refer to the sequence of heads and tails generated by Bob as C_B, and to the sequence of bits generated by his measurements as B_B. Here is a possible outcome of this protocol:

B_A	0	1	0	1	0	0	0	1	1	0	1	1
C_A	H	H	T	H	H	T	H	T	T	H	H	H
Q_A	$\|+\rangle$	$\|-\rangle$	$\|0\rangle$	$\|-\rangle$	$\|+\rangle$	$\|0\rangle$	$\|+\rangle$	$\|1\rangle$	$\|1\rangle$	$\|+\rangle$	$\|-\rangle$	$\|-\rangle$
C_B:	H	T	H	H	T	T	H	H	T	T	T	H
B_B:												

5. Once Bob receives $(4+\delta)n$ qubits, Alice publicly announces C_A and Bob publicly announces C_B.

6. Alice and Bob discard the bits from B_A and B_B where sequences C_A and C_B differ (that is, when Bob measured a qubit a in the different basis than Alice used for its preparation). With high probability, there are at least $2n$ bits left (if not, repeat the protocol). They keep $2n$ bits.

7. Alice selects a subset of n bits from the remaining $2n$ that will serve to check Eve's interference, and tells Bob which bits she selected.

8. Alice and Bob announce and compare the values of the n check bits. If more than an acceptable number disagree, they abort the protocol. (The acceptable number is determined by e.g., the noise in the channels.)

9. Alice and Bob perform classical information reconciliation and privacy amplification on the remaining n bits to obtain $O(n)$ shared key bits.

Quantum Computing Systems [1]

Prof. Emina Soljanin

Lecture #7, February 16

[1] Rutgers, ECE 579, Spring 2021

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

$$|\varphi\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B\big) \tag{1}$$

Qubit A is given to Alice and qubit B 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.

Local Measurements in the Computational Basis

What happens if Alice measures her qubit in the computational basis[2] and Bob does nothing? There are two possibilities: [3]

1. $|\varphi\rangle$ collapses to state

$$\frac{(|0\rangle\langle 0| \otimes I)\cdot|\varphi\rangle}{\|(|0\rangle\langle 0| \otimes I)\cdot|\varphi\rangle\|} = |0\rangle_A \otimes |0\rangle_B,$$

   which happens with probability $\langle\varphi|\cdot|0\rangle\langle 0|\otimes I\cdot|\varphi\rangle = 1/2$

2. $|\varphi\rangle$ collapses to state

$$\frac{(|1\rangle\langle 1| \otimes I)\cdot|\varphi\rangle}{\|(|1\rangle\langle 1| \otimes I)\cdot|\varphi\rangle\|} = |1\rangle_A \otimes |1\rangle_B,$$

   which happens with probability $\langle\varphi|\cdot|1\rangle\langle 1|\otimes I\cdot|\varphi\rangle = 1/2$

Observe that if Bob now measures his qubit in the computational basis, he will get a state that is identical to Alice's.[4]

[2] σ_Z measurement

[3] Recall that the projective measurement is defined as follows:

- A set of pairwise orthogonal projection operators $\{\Pi_i\}$ such that $\sum_i \Pi_i = I$.
- For input $|\psi\rangle$, output i happens wp $\langle\psi|\Pi_i|\psi\rangle$, and $|\psi\rangle$ collapses to $\frac{1}{\sqrt{\langle\psi|\Pi_i|\psi\rangle}}\Pi_i|\psi\rangle$.

[4] Einstain called to this phenomenon "spooky action at the distance" or *quantum non-locality*.

Local Measurements in the Hadamard Basis

What happens if Alice measures her qubit in the Hadamard basis[5] and Bob does nothing? There are two possibilities:

1. $|\varphi\rangle$ collapses to state

$$\frac{(|+\rangle\langle +| \otimes I)\cdot|\varphi\rangle}{\|(|+\rangle\langle +| \otimes I)\cdot|\varphi\rangle\|} = |+\rangle_A \otimes |+\rangle_B,$$

   which happens with probability $\langle\varphi|\cdot|+\rangle\langle +|\otimes I\cdot|\varphi\rangle = 1/2$

[5] σ_X measurement

2. $|\varphi\rangle$ collapses to state

$$\frac{(|1\rangle\langle 1| \otimes I) \cdot |\varphi\rangle}{\|(|-\rangle\langle -| \otimes I) \cdot |\varphi\rangle\|} = |-\rangle_A \otimes |-\rangle_B,$$

which happens with probability $\langle\varphi| \cdot |-\rangle\langle -| \otimes I \cdot |\varphi\rangle = 1/2$

Observe that if Bob now measures his qubit in the Hadamard basis, he will get a state that is identical to Alice's.

Local Measurements in the Computational Basis

What happens if Alice measures her qubit in the computational basis and Bob measures his qubit in the Hadamard basis?

Simultaneous Local Measurements

Suppose that Alice measures her qubit in the basis $\{|A_0\rangle, |A_1\rangle\}$ and Bob measurews his qubit in the basis $\{|B_0\rangle, |B_1\rangle\}$, where $|A_0\rangle$ and $|B_0\rangle$ can be expressed in the computational basis as follows:

$$|A_0\rangle = \cos\alpha|0\rangle + \sin\alpha|1\rangle \;\text{ and }\; |B_0\rangle = \cos\beta|0\rangle + \sin\beta|1\rangle$$

Note that Bob's basis can be obtained from Alice's by rotation, where the rotation angle is $\theta = \alpha - \beta$.

We denote the Alice's measurement result by a where $a \in \{0, 1\}$, and when Bob's measurement by b where $b \in \{0, 1\}$. We next show that Alice and Bob will have identical outputs wp $\cos^2\theta$. Since Alice and Bob perform their measurements locally, the measurement on the shared EPR pair $|\varphi_{AB}\rangle$ is effectively performed in the Kronceker product basis $|A_i\rangle\langle A_i| \otimes |B_j\rangle\langle B_j|$, $i, j \in \{0, 1\}$. Furthermore, we have[6]

$$\begin{aligned}
P(a = b = 0) &= \langle\varphi_{AB}| \left(|A_0\rangle\langle A_0| \otimes |B_0\rangle\langle B_0|\right) |\varphi_{AB}\rangle \\
&= \frac{1}{\sqrt{2}}\left(\langle 0|A_0\rangle \langle A_0| \otimes \langle 0|B_0\rangle \langle B_0| + \langle 1|A_0\rangle \langle A_0| \otimes \langle 1|B_0\rangle \langle B_0|\right) |\varphi_{AB}\rangle \\
&= \cos^2\alpha\cos^2\beta + 2\cos\alpha\cos\beta\sin\alpha\sin\beta + \sin^2\alpha\sin^2\beta \\
&= \frac{1}{2}(\cos\alpha\cos\beta + \sin\alpha\sin\beta)^2 = \frac{1}{2}\cos^2(\alpha - \beta)
\end{aligned}$$

[6] We use the following identities:

$$\langle 0|A_0\rangle = \langle A_0|0\rangle = \cos\alpha$$
$$\langle 0|B_0\rangle = \langle B_0|0\rangle = \cos\beta$$

It follows from a simple geometric argument that

$$P(a = b = 0) = P(a = b = 1)$$

Therefore, when Bob's basis can be obtained from Alice's by the angle $\alpha - \beta$ rotation, we have

$$P(a = b) = \cos^2(\alpha - \beta) \;\text{ and }\; P(a \neq b) = \sin^2(\alpha - \beta).$$

We conclude that when Alice and Bob measure in the same basis (i.e., $\alpha - \beta = 0$), they get identical results.

Hidden Variables and Bell's Inequalities

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.

The CHSH Game

The CHSH game demonstrates how two players Alice and Bob, who cannot communicate with each other once the game starts, can benefit from shared entanglement much more than from shared classical randomness in winning the game [7]. It is rooted in a paper by Clauser, Horne, Shimony, and Holt, hence the name [8].

[7]

[8]

Game Rules

In the CHSH game, Alice is given a binary input $x \in \{0, 1\}$ and Bob is given a binary input $y \in \{0, 1\}$ by a referee who guarantees that each combination of the inputs is equally likely. Upon receiving the input, Alice generates her output a and Bob his output b. They send the outputs to the referee who declares them the winners if $x \cdot y = a \texttt{ xor } b$. In other words, if $x = y = 1$, Alice and Bob win if their outputs are different. In all other cases, they win if their outputs are identical. Alice and Bob are allowed to agree on a strategy in advance, and to share random bits or entangled qubits, but once the game starts, they cannot communicate. Is there any advantage to be had from sharing entangled qubits?

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

Prof. Emina Soljanin

Lecture #8, February 18

This lecture explains the famous CHSH game and discusses quantum vs. classical correlations.

The CHSH Game

The CHSH game demonstrates how two players Alice and Bob, who cannot communicate with each other once the game starts, can benefit from shared entanglement much more than from shared classical randomness in winning the game. It is rooted in a paper by Clauser, Horne, Shimony, and Holt, hence the name.

In the CHSH game, Alice is given a binary input $x \in \{0, 1\}$ and Bob is given a binary input $y \in \{0, 1\}$ by a referee who guarantees that each combination of the inputs is equally likely. Upon receiving the input, Alice generates her output a and Bob his output b. They send the outputs to the referee who declares them the winners if $x \cdot y = a \texttt{ xor } b$. In other words, if $x = y = 1$, Alice and Bob win if their outputs are different. In all other cases, they win if their outputs are identical. Alice and Bob are allowed to agree on a strategy in advance, and to share random bits or entangled qubits, but once the game starts, they cannot communicate. Is there any advantage to be had from sharing entangled qubits?

The Best Classical Strategy

A classical strategy to maximize the winning probability is that Alice sends to the referee $a = 0$ regardless of the value of her input, and Bob does the same. With this strategy, we always have $a = b$, and Alice and Bob loose only when x and y are both 1. Therefore, they win the game with probability 0.75. It is straightforward to check that this is an optimal strategy among the 16 different deterministic strategies (ways to map four possible inputs to four possible outputs). Any shared classical randomness would essentially randomize among the 16 possible deterministic strategies, and thus cannot beat the best. The question then becomes if Alice and Bob can benefit from sharing EPR pairs. The answer to this question is yes, and we next describe a strategy with the winning probability of about 0.85.

The Best Quantum Strategy

Consider a strategy where Alice and Bob share an entangled pair of qubits in the state $|\varphi_{AB}\rangle$ given as

$$|\varphi\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B\big). \tag{1}$$

Upon receiving the input, each player measures his/her qubit in one of the two possible bases depending on whether the input is 0 or 1. They then generate their outputs according to the result of the measurement. Alice's chooses the computational basis for her input $x = 0$, and the Hadamard basis for her input $x = 1$. Bob's chooses the computational basis rotated by $\pi/8$ for his input $y = 0$ and the computational basis rotated by $-\pi/8$ for his input $y = 1$. Thus, there are four possible combinations of Alice/Bob measurement bases corresponding to the four different input pairs x and y, as shown in Fig. **??**. In order to find the winning probability of this strategy, we next prove a general result about local measurements of entangled qubits.

Figure 1: Choices of bases that Alice and Bob make based on their respective inputs x and y to measure their respective qubits. Each player selects the basis to measure based solely on the local input. The angle θ is chosen to be $\pi/8$. The strategy of the game is that, regardless of which basis is used, when Alice's measurement result is $i \in \{0, 1\}$, she outputs $a = i$, and when Bob's measurement result is $j \in \{0, 1\}$, he outputs $b = j$.

Recall that Alice and Bob share an EPR pair in the state $|\varphi_{AB}\rangle$ given by (**??**). Suppose that Alice measures her qubit in the basis $\{|A_0\rangle, |A_1\rangle\}$ and Bob measurews his qubit in the basis $\{|B_0\rangle, |B_1\rangle\}$, where $|A_0\rangle$ and $|B_0\rangle$ can be expressed in the computational basis as follows:

$$|A_0\rangle = \cos\alpha|0\rangle + \sin\alpha|1\rangle \quad \text{and} \quad |B_0\rangle = \cos\beta|0\rangle + \sin\beta|1\rangle$$

Note that Bob's basis can be obtained from Alice's by rotation, where the rotation angle is $\alpha - \beta$.

The strategy of the game is that, regardless of which basis is used, when Alice's measurement result is $i \in \{0, 1\}$, she outputs $a = i$, and when Bob's measurement result is $j \in \{0, 1\}$, he outputs is $b = j$. We next show that Alice and Bob will have identical outputs wp $\cos^2\theta$. Since Alice and Bob perform their measurements locally, the measurement on the shared EPR pair $|\varphi_{AB}\rangle$ is effectively performed in the Kronceker product basis $|A_i\rangle\langle A_i| \otimes |B_j\rangle\langle B_j|$, $i, j \in \{0, 1\}$. It follows from the definition of quantum measurement and simple geometry (see Fig. **??**) that

$$P(a = b = 0) = P(a = b = 1) \text{ and}$$

$$P(a = b = 0) = \langle\varphi_{AB}| \left(|A_0\rangle\langle A_0| \otimes |B_0\rangle\langle B_0|\right) |\varphi_{AB}\rangle .$$

Furthermore, we have

$$\begin{aligned}
P(a=b) &= 2\cdot\langle\varphi_{AB}|\left(|A_0\rangle\langle A_0|\otimes|B_0\rangle\langle B_0|\right)|\varphi_{AB}\rangle \\
&= \frac{2}{\sqrt{2}}\big(\langle 0|A_0\rangle\,\langle A_0|\otimes\langle 0|B_0\rangle\,\langle B_0| + \\
&\qquad \langle 0|A_0\rangle\,\langle A_0|\otimes\langle 0|B_0\rangle\,\langle B_0|\big)\,|\varphi_{AB}\rangle \\
&= \cos^2\alpha\cos^2\beta + 2\cos\alpha\cos\beta\sin\alpha\sin\beta + \sin^2\alpha\sin^2\beta \\
&= (\cos\alpha\cos\beta+\sin\alpha\sin\beta)^2 = \cos^2(\alpha-\beta)
\end{aligned}$$

where we have used the following identities:

$$\langle 0|A_0\rangle = \langle A_0|0\rangle = \cos\alpha \text{ and } \langle 0|B_0\rangle = \langle B_0|0\rangle = \cos\beta.$$

Therefore, when Bob's basis can be obtained from Alice's by the angle $\alpha-\beta$ rotation, we have

$$P(a=b)=\cos^2(\alpha-\beta) \;\text{ and }\; P(a\neq b)=\sin^2(\alpha-\beta). \tag{2}$$

We are now ready to derive the probability that Alice and Bob win the CHSH game. Observe that 1) the angle between Alice's and Bob's measurement bases is $3\pi/8$ when $x=y=1$, and $\pi/8$ for all other input combinations (see Fig. **??**), and 2) Alice and Bob win if they generate different outputs $a\neq b$ for inputs $x=y=1$, and identical outputs for all other input combinations. Therefore, by (**??**), the winning probability P_{win} can be computed as follows:

$$\begin{aligned}
P_{win} &= P(xy=0)P(a+b=0|xy=0)| + \\
&\qquad P(xy=1)P(a+b=1|xy=1) \\
&= \frac{3}{4}P(a=b|xy=0)| + \frac{1}{4}P(a\neq b|xy=1)] \\
&= \frac{3}{4}\cdot\cos^2(\pi/8) + \frac{1}{4}\cdot\sin^2(3\pi/8) \\
&= \cos^2(\pi/8) = (2+\sqrt{2})/4 \gtrsim 0.853.
\end{aligned}$$

A natural question to ask is whether Alice and Bob can achieve even better winning probability by sharing some other entangled state or by using some other sets of bases or both. The answer to these questions is no.

The significance of the CHSH game and similar tools is that they show that there is a limit to what can be done with classical (possibly hidden) randomness. If experiments involving shared entanglement show that this limit can be beaten (as they have), then there must be some "spooky action at a distance" like the one reflected in (**??**). And that is where the weirdness and the power of quantum computing reside.

E91-like QKD Protocols

The E91 protocol for quantum key distribution was proposed by Ekert in 1991, hence the name. The scheme distributes entangled pairs of photons so that Alice and Bob each end up with one photon from each entangled pair. The creation and distribution of photons can be done by Alice, by Bob, or by some third party.

Suppose Alice and Bob share a set of n entangled pairs of qubits in the state $(|00\rangle + |11\rangle)/\sqrt{2}$ and Eve is not present. If they measure their respective states in the computational basis, they will get identical sequences of completely random bits. Thus, the scheme benefits from two properties of shared entanglement: randomness and correlation. To check if Eve was present, Alice and Bob can, for example, select a random subset of the shared entangled pairs, and test to see if they are entangled (instead of using them to generate the key bits). They can do that, e.g., by playing the CHSH game.

Quantum Computing Systems [1]

Prof. Emina Soljanin

Lecture #9, February 23

[1] Rutgers, ECE 579, Spring 2021

This lecture introduces the notion of mixed state, and is only concerned with representing, processing, and measuring mixed states.

Mixed States

There are many scenarios when we do not know the state of a quantum system, but do know that it is in the state $|\psi_j\rangle$ with probability p_j, $j = 1, \dots, k$. We say then that the quantum system is in a *mixed state*, and refer to the collection of pairs $\{|\psi_j\rangle, p_j\}_{j=1}^{k}$ as an *ensemble of states*. The states we worked with so far, which can be described by a vector, are known as *pure states*. A mixed state arises e.g., when we know that a measurement of a pure state has been performed but do not know the outcome.

The Density Matrix Formalism

So far, we used unit-norm vectors in Hilbert spaces to mathematically specify quantum states. We can instead describe a quantum state, say $|\psi\rangle$, by the projection matrix $\rho_\psi = |\psi\rangle\langle\psi|$. We refer to ρ_ψ as the *density matrix* of $|\psi\rangle$. To verify that this is a valid model, we need 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.

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

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii}$$

The trace is invariant under cyclic permutations, e.g.,

$$\mathrm{Tr}(ABCD) = \mathrm{Tr}(BCDA)$$

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, $|\psi\rangle \xrightarrow{U} U|\psi\rangle$ is replaced by $\rho_\psi \xrightarrow{U} U\rho_\psi U^\dagger$.

2. Suppose that a measurement defined by the basis $|u_1\rangle, \dots, |u_N\rangle$ is performed on the state $|\psi\rangle$. We know that the resulting state will be $|u_i\rangle$ with probability (wp) $|\langle\psi|u_i\rangle|^2$, $1 \leqslant i \leqslant N$. We observe that $|\langle\psi|u_i\rangle|^2 = \langle\psi|u_i\rangle\langle u_i|\psi\rangle = \mathrm{Tr}\big(|u_i\rangle\langle u_i| \cdot |\psi\rangle\langle\psi|\big) = \mathrm{Tr}\big(|u_i\rangle\langle u_i| \cdot \rho_\psi\big)$.

   Therefore, in terms of density matrices, we have
   $\rho_\psi \to |u_i\rangle\langle u_i|$ wp $\mathrm{Tr}\big(|u_i\rangle\langle u_i|\rho_\psi\big)$.

Note that to describe the state, evolution, and measurement, we used density matrices rather than state vectors. The advantage of the density matrix formalism is that it allows us to compactly describe mixed states. A mixed state, that is, a quantum system about which we only know that it is in the state $|\psi_j\rangle$ with probability p_j has a

density matrix defined as follows:

$$\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|. \tag{1}$$

The states that have rank–1 density matrices $\rho_\psi = |\psi\rangle\langle\psi|$ are known as *pure states*. In general, a density matrix ρ is a Hermitian, positive semi-definite, trace–1 matrix. These properties easily follow from (1).

Observe that two different ensembles of states can have identical density matrices, and therefore quantum mechanically represent identical states. Fig. 1 shows two different ensembles with the density matrix equal to $\frac{1}{2}\mathrm{I}$.

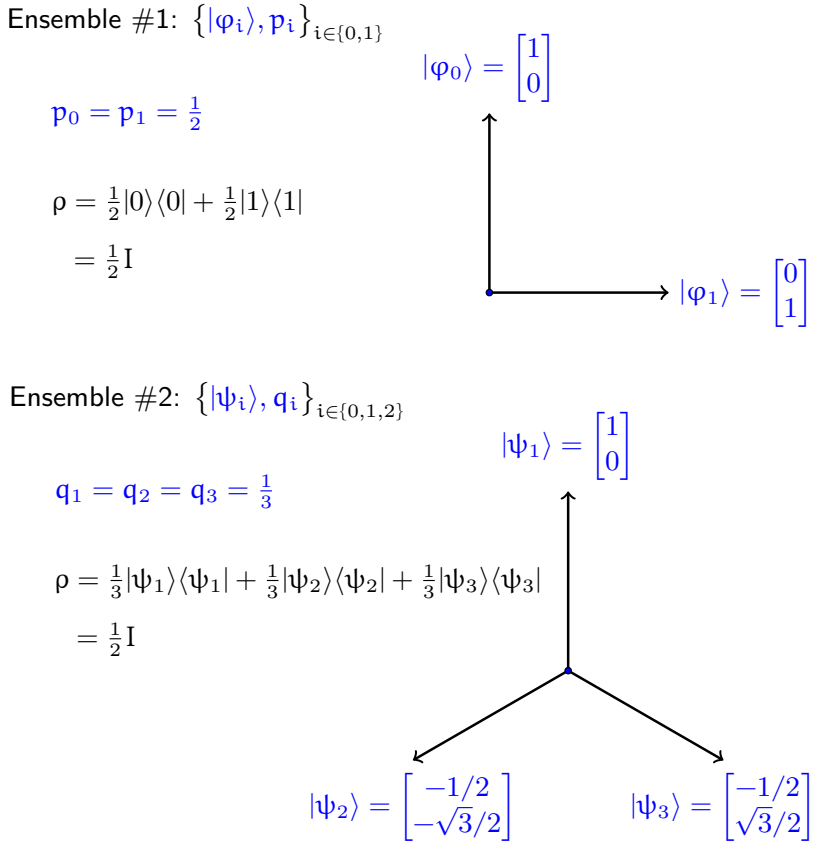


Figure 1: Two "different" ensembles of pure states with identical density matrices equal to $\frac{1}{2}\mathrm{I}$.

When a $d \times d$ density matrix is equal to $\frac{1}{2}\mathrm{I}$, we say that the system is in the maximally mixed state. These density matrices are quantum counterparts to classical uniform distributions.

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 $U\rho_j U^\dagger$, as we showed above. But we only know that the system is in the state ρ_j with probability p_j. Therefore, the mixed state will evolve to the state $U\rho_j U^\dagger$ with probability p_j. Therefore, the system with density matrix (1) will evolve into another mixed state, whose density matrix is given by

$$\sum_j p_j U|\psi_j\rangle\langle\psi_j|U^\dagger = U\Big(\sum_j p_j|\psi_j\rangle\langle\psi_j|\Big)U^\dagger = U\rho U^\dagger$$

Therefore, $\rho \xrightarrow{U} U\rho U^\dagger$.

Measuring Mixed States

We next look into what happens when we perform a quantum measurement defined by operators Π_i on a mixed state whose density matrix is $\rho = \sum_j p_j|\psi_j\rangle\langle\psi_j|$. Again, let $\rho_j = |\psi_j\rangle\langle\psi_j|$. If the state being measured is $|\psi_j\rangle$ (which happens with probability p_j), then the probability of getting measurement result i is $\mathrm{Tr}(\Pi_i\rho_j)$. Therefore, by the total probability formula, when measuring ρ, we get outcome i with probability

$$\sum_j p_j \underbrace{\mathrm{Tr}(\Pi_i|\psi_j\rangle\langle\psi_j|)}_{\Pr(i|j)} = \mathrm{Tr}(\Pi_i \sum_j p_j|\psi_j\rangle\langle\psi_j|) = \mathrm{Tr}(\Pi_i\rho)$$

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

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 $\frac{\Pi_i\rho_j\Pi_i}{\mathrm{Tr}(\Pi_i\rho_j)}$. Therefore, if we observe outcome i, the system is in the mixed state

$$\sum_j p_j \frac{\Pi_i\rho_j\Pi_i}{\mathrm{Tr}(\Pi_i\rho_j)} = \frac{\Pi_i\rho\Pi_i}{\mathrm{Tr}(\Pi_i\rho)}.$$

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? Note that, mathematically, the state of the system is for us described based on our *ignorance/knowledge*. We saw that we get state $\frac{\Pi_i\rho\Pi_i}{\mathrm{Tr}(\Pi_i\rho)}$ wp $\mathrm{Tr}(\Pi_i\rho)$. If we lost the measurement record, we would have a state described by the density matrix

$$\sum_{i=1} \mathrm{Tr}(\Pi_i\rho) \cdot \frac{\Pi_i\rho\Pi_i}{\mathrm{Tr}(\Pi_i\rho)} = \sum_{i=1} \Pi_i\rho\Pi_i. \tag{2}$$

Elementary probability (e.g., the total probability expression) is used for derivations.

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

Prof. Emina Soljanin

Lecture #10, February 25

This lecture talks about the Bloch sphere representation of qubits.

How is Quantum Information Represented?

Representation of quantum information is connected to a postulate of quantum mechanics which says that associated to any isolated physical system is a complex vector space with inner product. In this class, and quantum computing in general, we mostly deal with finite dimensional spaces $\mathbb{C}^N$ and often conventionally refer to them as Hilbert spaces $\mathcal{H}_N$.

The quantum information and computing counterpart to the bit is the *qubit*. Qubits (as bits) are represented by physical systems. Mathematically, independently of a particular physical realization, a qubit is represented by a unit-normvector in the two-dimensional unitary space $\mathbb{C}^2$. If we denote the basis vectors of this space by

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

then a single qubit $|\psi\rangle$ is mathematically a linear combination of $|0\rangle$ and $|1\rangle$, that is

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

where α and β are complex numbers such that $|\alpha|^2 + |\beta|^2 = 1$. Because we can easily normalize any vector to have a unit norm, a quantum state can be thought of as a ray in a Hilbert space. It is an equivalence class of vectors that differ by multiplication by a nonzero scalar.

We can represent a unit norm state $|\psi\rangle$ uniquely as follows:

$$|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\phi}\sin(\theta/2)\,|1\rangle \tag{2}$$

$$\begin{aligned} \implies \quad |\psi\rangle\langle\psi| &= \cos^2(\theta/2)|0\rangle\langle 0| + e^{-i\phi}\cos(\theta/2)\sin(\theta/2)|0\rangle\langle 1| + \\ &\quad e^{i\phi}\cos(\theta/2)\sin(\theta/2)|1\rangle\langle 0| + \sin^2(\theta/2)|1\rangle\langle 1| \\ &= \begin{bmatrix} \cos^2(\theta/2) & e^{-i\phi}\cos(\theta/2)\sin(\theta/2) \\ e^{i\phi}\cos(\theta/2)\sin(\theta/2) & \cos^2(\theta/2) \end{bmatrix} \end{aligned}$$

$$\begin{aligned} |\psi\rangle\langle\psi| - \frac{1}{2}I &= \frac{1}{2}\begin{bmatrix} \cos\theta & e^{-i\phi}\sin\theta \\ e^{i\phi}\sin\theta & \cos\theta \end{bmatrix} \\ &\quad \frac{1}{2}\left(\sin\theta\cos\phi\cdot\sigma_X + \sin\theta\sin\phi\cdot\sigma_Y + \cos\theta\cdot\sigma_Z\right) \end{aligned}$$

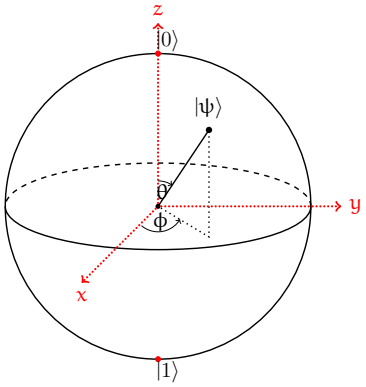


Figure 1: Bloch Sphere

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \sigma_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\sigma_Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad \sigma_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Bloch Sphere

The Bloch sphere provides a useful way to represent and visualize both pure and mixed states, and is traditionally used in quantum mechanics. It is also used in quantum computing platforms, such as IBM-Q, since actions of single-qubit gates on pure states are easy to see within the Bloch sphere framework.

Any 2×2 complex matrix, and thus any density matrix ρ, can be expressed as a linear combination of the identity I and the Pauli matrices σ_X, σ_Y, and σ_Z:

$$\rho = \alpha_I I + \alpha_X \sigma_X + \alpha_Y \sigma_Y + \alpha_Z \sigma_Z$$

for some complex numbers α_I, α_X, α_Y, and α_Z. Since a density matrix is Hermitian and has trace one, these numbers will satisfy certain constraints.

Note that σ_X, σ_Y, and σ_Z have trace equal to 0. Therefore

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

where β_X, β_Y, and β_Z are real numbers. The latter holds because ρ is a Hermitian matrix and

$$\rho = \frac{1}{2}\begin{bmatrix} 1+\beta_Z & \beta_X - i\beta_y \\ \beta_X + i\beta_y & 1-\beta_Z \end{bmatrix}. \tag{3}$$

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

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

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

For pure states, we have $\mathrm{Tr}(\rho^2) = 1$, and thus

$$1 = \mathrm{Tr}(\rho^2) = \frac{1}{2}\left(1 + |\vec{\beta}|^2\right) \quad \Leftrightarrow \quad |\vec{\beta}| = 1$$

Therefore, 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.

We can also see that pure states are points on the Bloch sphere by considering the representation of $|\psi\rangle$ we introduced earlier in the class:

$$|\psi\rangle = \cos(\theta/2)\,|0\rangle + e^{i\phi}\sin(\theta/2)\,|1\rangle.$$

Comparing $|\psi\rangle\langle\psi|$ with the matrix (3), we find that the Bloch vector of $|\psi\rangle$ makes an angle of θ with the z axis, and its projection in the $x-y$ plane makes an angle of ϕ with the x axis, as shown in Fig. 2. With

Pauli matrices:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \sigma_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\sigma_Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad \sigma_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

form a basis for $\mathbb{C}^{2\times 2}$.

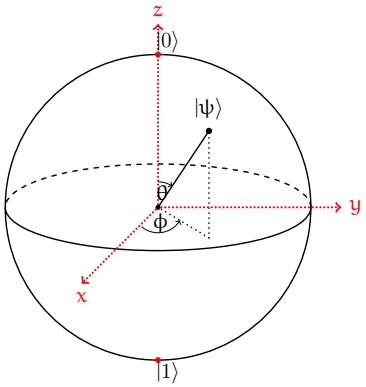


Figure 2: Bloch Sphere

this representation, it is easy to see that any two diametrically opposite (antipodal) points correspond to a pair of mutually orthogonal pure state vectors. In particular, $\beta_X = \beta_Y = 0$ and $\beta_Z = 1$ gives $\rho = |0\rangle\langle 0|$, while $\beta_X = \beta_Y = 0$ and $\beta_Z = -1$ gives $\rho = |1\rangle\langle 1|$.

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

Prof. Emina Soljanin

Lecture #11, March 2

This lecture 1) discusses bipartite quantum systems and 2) defines the van Neumann entropy.

Bipartite Quantum States

Quantum systems are often shared between two, usually spatially separated, parties we refer to Alice and Bob. Let Alice's subsystem be in the Hilbert space $\mathcal{H}_A$ and Bob's in $\mathcal{H}_B$ Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be finite-dimensional with basis states $\{|a_i\rangle\}_{i=1}^n$ and $\{|b_j\rangle\}_{j=1}^m$, respectively. Then the state space of the composite (compound bipartite) system is the tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$ with the basis $\{|a_i\rangle \otimes |b_j\rangle\}$, or in more compact notation $\{|a_i b_j\rangle\}$.

Bipartite Pure States

Let $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$. We say that $|\psi\rangle_{AB}$ is a bipartite pure state of a composite system with subsystems A and B. Any pure state of the composite system can be written as

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

where $c_{i,j}$ are complex numbers. If a pure state $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ can be written in the form $|\psi\rangle_{AB} = |\psi_A\rangle \otimes |\psi_B\rangle$, 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 pure states to its subsystems.

Schmidt Decomposition – Theorem

For any vector $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$, there exist orthonormal vectors $|u_i\rangle \in \mathcal{H}_A$, $|v_i\rangle \in \mathcal{H}_B$, and real, non-negative numbers α_i, $i = 1, 2, \ldots, \ell$ such that

$$|\psi\rangle_{AB} = \sum_{i=1}^{\ell} \alpha_i \, |u_i\rangle \otimes |v_i\rangle \tag{1}$$

where $\ell = \min\{m, n\}$ and α_i are unique up to re-ordering. Expression (1) is the Schmidt decomposition of $|\psi\rangle_{AB}$ and α_i are its Schmidt coefficients.

The Schmidt Decomposition is useful for the separability characterization of pure states:

1. The state $|\psi\rangle_{AB}$ is separable iff it has only one non-zero Schmidt coefficient. Otherwise it is entangled.
2. If all the Schmidt coefficients are non-zero and equal, then the state is said to be *maximally* entangled.[2]

[2] We will discuss entanglement measures next time.

Bipartite Mixed States

Let ρ_{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 ρ_{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
$$\rho = \sum_k p_k \rho_A^k \otimes \rho_B^k,$$
3. an entangled state, otherwise.[3]

[3] 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,

$$\operatorname{Tr} A = \sum_{i=1}^{n} \operatorname{Tr}\big(|e_i\rangle\langle e_i|A\big) = \sum_{i=1}^{n} \langle e_i| A |e_i\rangle$$

Let ρ_{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:[4]

$$\rho_A = \operatorname{Tr}_B \rho_{AB} = \sum_i \big(I \otimes \langle b_i|\big)\rho_{AB}\big(I \otimes |b_i\rangle\big)$$

[4] You will often see a shorthand expression $\operatorname{Tr}_B \rho_{AB} = \sum_b \langle b| \rho_{AB} |b\rangle$

We say that ρ_A is a reduced density operator[5] obtained from ρ_{AB} by *tracing out* the subsystem B.

[5] Reduced density operators are quantum counterparts to marginal distributions in the classical world.

Example #1 – Product State:
Suppose a quantum system is in the product state $\rho_{AB} = \rho_A \otimes \rho_B$ where ρ_A is a density operator for system A, and ρ_B is a density operator for system B. Then

$$\rho_A = \operatorname{Tr}_B(\rho_A \otimes \rho_B) = \rho_A \operatorname{Tr} \rho_B = \rho_A$$

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

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

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

$$\begin{aligned}\rho_A = \mathrm{Tr}_B \rho_{AB} &= (I \otimes \langle 0|)\rho_{AB}(I \otimes |0\rangle) + (I \otimes \langle 1|)\rho_{AB}(I \otimes |1\rangle) \\ &= \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \boxed{\frac{1}{2}I}\end{aligned}$$

Shannon Entropy

The Shannon entropy of a probability distribution P on the sample space $\mathcal{X}$ is defined as

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

When $\mathcal{X}$ has two elements, one with probability p and the other with $1-p$, the Shannon entropy is known as the binary entropy, shown in Fig. 1. 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 a probability distribution. It is maximized by the uniform distribution among all distributions with equal sample space sizes.

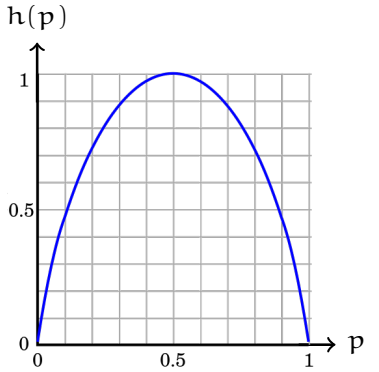


Figure 1: Shannon's Binary Entropy $h(p) = -p\log(p) - (1-p)\log(1-p)$

Math Interlude – Matrix Functions

Matrix Exponential

Let X be an $n \times n$ complex matrix. The exponential of X, denoted by e^X or $\exp(X)$, is the $n \times n$ matrix given by the power series

$$e^X = \sum_{k=0}^{\infty} \frac{X^k}{k!}$$

where X^0 is defined to be the $n \times n$ identity matrix.[6]
e^X has the following properties:

1. $e^0 = I$
2. $\exp(X^T) = (\exp X)^T$
3. $\exp(X^*) = (\exp X)^*$
4. If Y is invertible, then $e^{YXY^{-1}} = Ye^XY^{-1}$.
5. If $XY = YX$ then $e^Xe^Y = e^{X+Y}$.

[6] What is e^X when X is diagonal?

Matrix Logarithm

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

$$e^X = \sum_{k=0}^{\infty} \frac{X^k}{k!}.$$

Since $e^{VXV^{-1}} = Ve^XV^{-1}$, we have $\log Y = V(\log V^{-1}XV)V^{-1}$.

Let D_ρ be a matrix obtained by diagonalizing ρ. Then

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

Von Neumann Entropy

Recall that the Shannon entropy measures the expected uncertainty associated with a classical probability distribution.[7] The quantum counterpart of a probability distribution is a density matrix ρ. The von Neumann entropy is an older concept that generalizes the Shannon entropy. It is given by

$$S(\rho) = -\operatorname{Tr} \rho \log \rho$$

If λ_i are eigenvalues of ρ, 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 ρ 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.

Example #3:
Consider pure bipartite state $|\psi_{AB}\rangle$ with the density operator $\rho_{AB} = |\phi_{AB}\rangle\langle\phi_{AB}|$. Show that $S(\rho_A) = S(\rho_B)$.[8]

[7] Shannon entropy has multiple *operational* meanings, including the compression rate of classical DMS information sources.

[8] Hint: Use Schmidt decomposition.

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

Prof. Emina Soljanin

Lecture #12, March 4

This lecture discusses entanglement entropy, entanglement measures, and monogamy of entanglement.

Reduced States of a Pure state

Consider a pure bipartite state $|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ with the density matrix ρ_{AB}. Its reduced states are

$$\rho_A = \mathrm{Tr}_B(\rho_{AB}) \text{ and } \rho_B = \mathrm{Tr}_A(\rho_{AB})$$

Let the Schmidt decomposition of $|\psi\rangle_{AB}$ be

$$|\psi\rangle_{AB} = \sum_{i=1}^{\ell} \alpha_i |u_i\rangle \otimes |v_i\rangle$$

From the decomposition, we see that each reduced state has the von Neumann entropy equal to

$$-\sum_{i}^{\ell} |\alpha_i|^2 \log |\alpha_i|^2$$

with orthonormal vectors $|u_i\rangle \in \mathcal{H}_A$, $|v_i\rangle \in \mathcal{H}_B$.[2] The bipartite von Neumann entanglement entropy of ρ_{AB} is defined as the von Neumann entropy of either of its reduced states.

[2] Show that

$$\rho_A = \sum_{i=1}^{\ell} \alpha_i^2 |u_i\rangle\langle u_i|$$

$$\rho_B = \sum_{i=1}^{\ell} \alpha_i^2 |v_i\rangle\langle v_i|$$

Entanglement Measures

Entanglement Measures for Pure Bipartitie States

The entanglement between subsystems A and B of a pure bipartite state is measured by *the bipartite Von Neumann entanglement entropy*, which is the Von Neumann entropy of either subsystem A or subsystem B. It follows from the Schmidt decomposition.[3]

$\Longrightarrow$

For a pure state ρ_{AB}, the entanglement is measured by

$$S(\rho_A) = -\mathrm{Tr}[\rho_A \log \rho_A] = -\mathrm{Tr}[\rho_B \log \rho_B] = S(\rho_B)$$

where

$$\rho_A = \mathrm{Tr}_B(\rho_{AB}) \text{ and } \rho_B = \mathrm{Tr}_A(\rho_{AB})$$

are the reduced density matrices of the subsystems.

[3] The Schmidt decomposition refers to a particular way of expressing a vector in the Kronecker product of two Hilbert spaces.

Example #1 – Product State:
Consider the bipartite state $|\psi_{AB}\rangle = (|00\rangle + |01\rangle)/\sqrt{2}$. Note that this is a product state $|\psi_{AB}\rangle = |0\rangle \otimes (|0\rangle + |1\rangle)/\sqrt{2} = |0\rangle \otimes |+\rangle$ with the density operator

$$\rho_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}| = \frac{1}{2}|00\rangle\langle 00| \otimes |+\rangle\langle +|$$

with 0 entanglement by the measure $S(\rho_A) = S(\rho_B) = 0$.

Example #1 – Bell State:
Consider the bipartite state $|\phi_{AB}\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. This is a pure state with the density operator

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

The reduced density matrices are

$$\rho_A = \mathrm{Tr}_B \rho_{AB} = \rho_B = \mathrm{Tr}_A \rho_{AB} = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \frac{1}{2}I$$

Since $S(\rho_A) = S(\rho_B) = 1$, state $|\phi_{AB}\rangle$ is maximally entangled.

Entanglement Measures for Mixed Bipartitie States

There is no unique measure of entanglement for mixed bipartite states, but there are certain conditions that any measure of entanglement should satisfy:

1. $E(\rho) = 0$ iff ρ is separable.
2. $E(\rho)$ is invariant under local unitary operations:
   $E\big((U_A \otimes U_B)\rho(U_A \otimes U_B)\big) = E(\rho)$.
3. E does not increase on average under LOCC.[4]

There are several entanglement measures proposed in the literature for mixed states, but no single one is standard. Many entanglement measures reduce to the entropy of entanglement when evaluated on pure states.

[4] LOCC refers to *local operations and classical communication*. Here a local (product) operation is performed on part of the system, and the result of that operation is communicated classically to another part where usually another local operation is performed based on the information received.

Monogamy of Entanglement

Informally speaking, a quantum state cannot be maximally entangled with multiple states simultaneously. More precisely, given a tripartite state ρ_{ABC}, the entanglement entropy of subsystem A cannot be maximal for both ρ_{AB} and ρ_{AC}.

Example #1 – The GHZ *State*

The GHZ state refers to the folowing entangled quantum state of three qubits A, B, C:

$$|GHZ\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle).$$

Any single qubit is maximally entangle with the rest of the system. We can see that by computing a single qubit entanglement entropy:

$$\rho_A^{GHZ} = \rho_B^{GHZ} = \rho_C^{GHZ} = \frac{1}{2}I \implies$$
$$S(\rho_A^{GHZ}) = S(\rho_B^{GHZ}) = S(\rho_C^{GHZ}) = 1.$$

However, any two qubit subsystem is not an entangled state. To see that consider

$$\rho_{AB}^{GHZ} = \frac{1}{2}(|00\rangle\langle 00| + |11\rangle\langle 11|) = \frac{1}{2}(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|)$$

Therefore ρ_{AB}^{GHZ} is a separable mixed state, and subsystem A is not entangled with B alone.

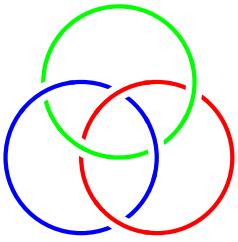


Figure 1: The GHZ state could be envisioned as the Borromean Rings, which are three rings linked so that no single ring can be removed from this arrangement without cutting, while cutting out any single ring leaves the other two separated.

Example #2 – The W State

The W state refers to the following entangled quantum state of three qubits A, B, C:

$$|W\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle)$$

We denote $\rho_{ABC}^W = |W\rangle\langle W|$. What can we say about the entanglement of qubit A with each of the subsystems BC, B, C?

We need to find ρ_{BC}^W, ρ_{AB}^W, and ρ_{AC}^W. Observe that the Schmidt decomposition of W is

$$|W\rangle = \sqrt{\frac{2}{3}}\,|0\rangle\,|\phi_{BC}\rangle + \frac{1}{\sqrt{3}}\,|1\rangle\,|00\rangle$$

where $|\phi_{BC}\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$. We have

$$\rho_{BC}^W = \mathrm{Tr}_A|W\rangle\langle W| = \frac{2}{3}|\phi_{BC}\rangle\langle\phi_{BC}| + \frac{1}{3}|00\rangle\langle 00|$$

$\implies$

1) A is not maximally entangled with BC since $S(\rho_{BC}^W) = h(1/3 < 1$

2) ρ_{BC}^W is an entangled state.

Because of the symmetry in the W state, we have

$$\rho_{AB}^W = \mathrm{Tr}_C|W\rangle\langle W| = \frac{2}{3}|\phi_{AB}\rangle\langle\phi_{AB}| + \frac{1}{3}|00\rangle\langle 00|$$

$$\implies \quad \rho_A^W = \mathrm{Tr}_B\,\rho_{AB}^W = \frac{2}{3}|0\rangle\langle 0| + \frac{1}{3}|1\rangle\langle 1|$$

A is entangled with both B and C alone, but not maximally with either.

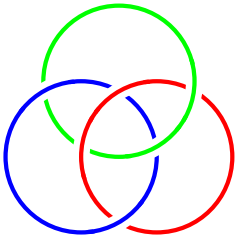


Figure 2: The W state could be envisioned three rings linked so that no single ring can be removed from this arrangement without a cut and no two rings can be removed without 2 cuts.

Sharable States

A bipartite state ρ_{AB} is said to be n-sharable if it is possible to find a quantum state $\rho_{AB_1B_2\ldots B_n}$ such that $\rho_{AB_1} = \rho_{AB_2} = \cdots = \rho_{AB_n}$ Such $\rho_{AB_1B_2\ldots B_n}$ if it exists, is called as an n-extension of ρ_{AB}. It can can be shown that a bipartite state is n-sharable for any n if and only if it is separable.

Entanglement Entropy in Many-Qubit Systems

The entanglement entropy is also used to measure the level of entanglement of a subsystem in a many-qubit pure quantum state. Given the density matrix of a many-qubit system, we can perform the partial trace operation to obtain a reduced density matrix of the a subsystem and compute its entropy. Such computations are often very difficult.

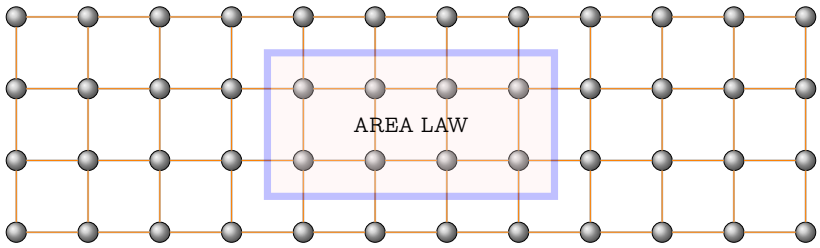


Figure 3: A square lattice many body system.

A key problem in quantum many-body physics is understanding the entanglement structure of a ground state. Consider, for example, a system with local interactions defined by a two-dimensional square lattice as shown in Fig. 3. We are interested in a gapped system in the ground state, when the number of qubits goes to infinity, and a subsystem as in the figure. The entanglement entropy of the subsystem is proportional to the length of the boundary of the region rather than to the volume of the region. This surprising property is known as the *entanglement area law*.

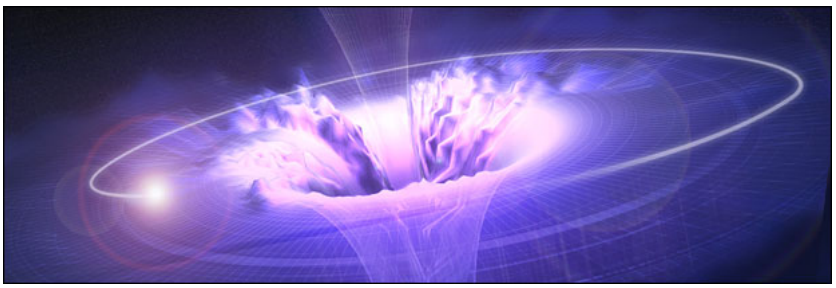


Figure 4: Studying area laws is of interest in e.g., quantum complexity, quantum information theory, and the holographic principle of black holes.

Quantum Computing Systems [1]

Prof. Emina Soljanin

Lecture #13, March 11

[1] Rutgers, ECE 579, Spring 2021

This lecture is about the measurements in the classical penny weighing problem.

The Problem and the Algorithm

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?

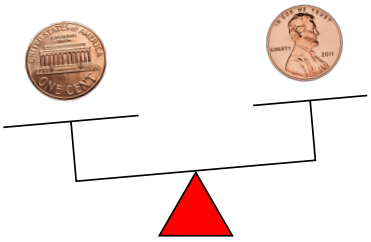


Figure 1: How would you use a balance scale to determine which of the 8 pennies has a different weight?

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[2] is given in Table 1. 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 raw 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.

[2] an algorithm

		PENNY							
		⓪	①	②	③	④	⑤	⑥	⑦
ON SCALE	M1	0	0	0	0	L	L	R	R
	M2	0	0	L	L	0	0	R	R
	M3	0	L	0	L	0	R	0	R

Table 1: Pennies placement on the scale in three measurements. A penny can be placed left (L), right (R) or not at all (0).

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. 2.

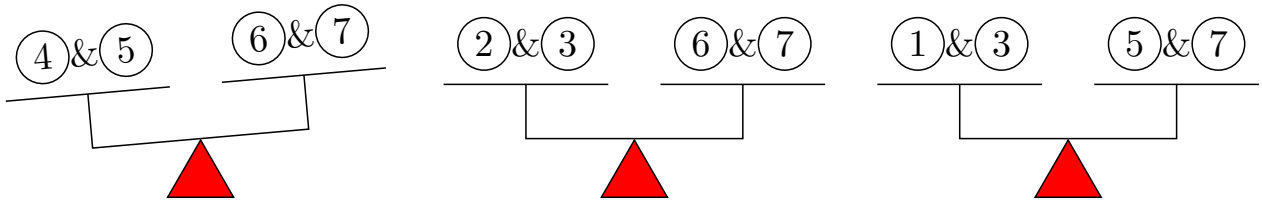


Figure 2: An example of measurement outcomes. Which penny has different weight?

Observe that since there is only one penny of different weight, a

measurement will result in an unbalanced state of the scale iff the penny of different weight is placed on the scale in that measurement. Therefore, the possible measurement outcomes are as given in Table 2. 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.

		DIFFERENT PENNY							
		(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
SCALE STATE	M1	0	0	0	0	1	1	1	1
	M2	0	0	1	1	0	0	1	1
	M3	0	1	0	1	0	1	0	1

Table 2: Scale states corresponding to measurements for each of the 8 "different panny" possibilities. The scale can be either balanced (0) or unbalanced (1).

Suppose you have a balance scale as in Fig. 1. Find a set of 3 measurements that you can use to identify the different penny if you know that it is heavier (or lighter) than the other seven.

Some Observations

1. We have committed to the way we perform the three measurements before the measuring process started. That is, we do not *adapt*[3] 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. It can also be helpful in practice.

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. 3, and you can only use the left tray to place pennies?

[3] Non-adaptive measuring can be as powerful as adaptive.

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. 1. 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. 3. How many measurements would you need *on average* to find the single penny that does not have the unit weight?

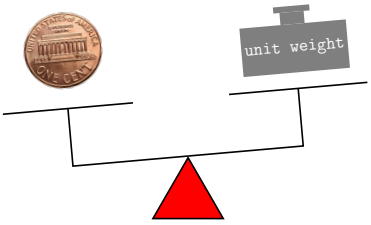


Figure 3: In this scale, there is some unit weight fixed to the right tray.

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

Prof. Emina Soljanin

Lecture #14, March 23

This lecture introduces classical information sources and (asymptotically) lossless classical source coding (data compression).

Block Diagram of a Communication System

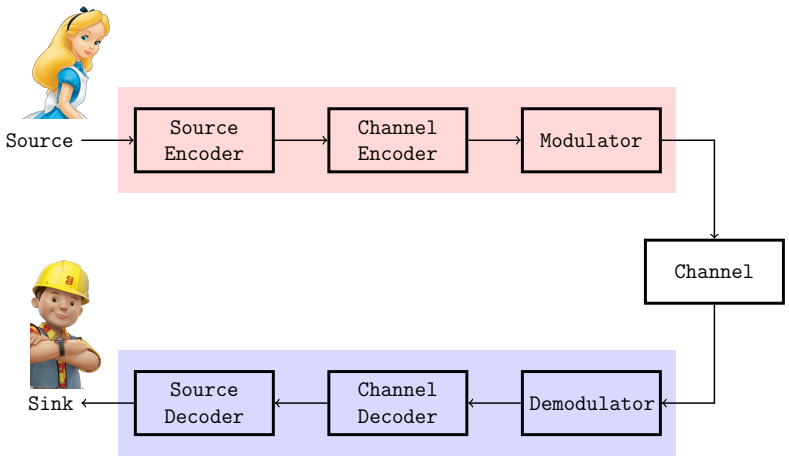


Figure 1: Transmission or storage system.

Math Interlude

A sequence of random variables (RVs) $X_1, X_2, \ldots, X_n$ that are mutually independent and have the same distribution[2] is called a sequence of independent trials or an independent trials process.

[2] We say that they are i.i.d. (independent identically distributed).

Example:
A *Bernoulli trials process* is a sequence of chance experiments s.t.[3]

[3] A coin toss is a Bernoulli trial.

1. Each experiment has two possible outcomes, which we may call H & T, 1 & 0, 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[4] with $E(X_j) = \mu < \infty$ and $V(X_j) = \sigma^2 < \infty$ (finite mean and variance). Let $S_n = X_1 + X_2 + \cdots + X_n$.

[4] The mean and variance are only defined for numerically valued RVs.

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

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

Jensen's Inequality:[5] For an RV X with expectation $E(X)$ and a convex function f, we have

$$f(E[X]) \leqslant E[f(X)].$$

[5] The form often used in probability.

A Function of a Random variable

If X is an RV with the range Ω_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

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

- The mean of Y is given by

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

Example:[6] 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

[6] An unusual example but we need to understand it.

X	⚀	⚁	⚂	⚃	⚄	⚅
P^1_X	1/6	1/6	1/6	1/6	1/6	1/6
P^2_X	1/3	1/6	1/12	1/6	1/6	1/12

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.

A Classical Source of Information

A discrete memoryless source[7] (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

[7] What information theorists call a binary source is mathematically a Bernoulli trials process.

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.

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[8] $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ occurs with probability

$$P_X(\mathbf{x}) = P_X(x_1) \cdot \ldots \cdot P_X(x_n).$$

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

[8] Compression algorithms deal with source sequences rather than individual letters.

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 δ for some very small δ. 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.

Figure 2: How much information is in an outcome of rolling a fair 8-faced die? What if only even numbers are possible.

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. 3. The binary entropy function attains its maximum value at $p = \frac{1}{2}$ (cf. unbiased coin flip).

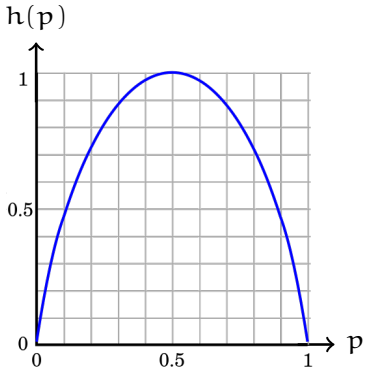


Figure 3: Shannon's Binary Entropy $h(p) = -p \log(p) - (1-p) \log(1-p)$

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.[9]

[9] 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

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

Hint: Use Jensen's inequality.

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:

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

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

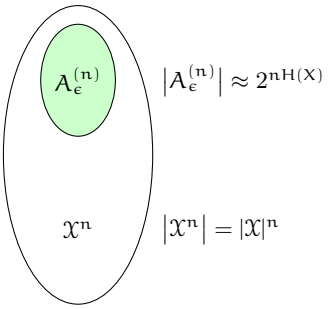
We say that a source sequence $x_1, \ldots, x_n$ is weakly ε-typical (aka entropy ε-typical) if

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

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

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. $\Pr\left[A_\epsilon^{(n)}\right] \geqslant 1 - \varepsilon$
2. $(1-\varepsilon)2^{n(H(X)-\varepsilon)} \leqslant \left|A_\varepsilon{}^{(n)}\right| \leqslant 2^{n(H(X)+\varepsilon)}$



Therefore, the fraction of sequences that are typical is

$$\frac{|A_\epsilon^{(n)}|}{|\mathcal{X}^{(n)}|} \leqslant \frac{2^{nH(X)+\varepsilon}}{2^{n\log_2|\mathcal{X}|}} = 2^{-n(\log_2|\mathcal{X}| - H(X) - \varepsilon)} \to 0$$

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

(Asymptotically) Lossless Source Coding

How many bits do we need to represent the $2^{n \log_2 |\mathcal{X}|}$ 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.[10]

[10] More efficient algorithms are used in practice, and real sources are seldom DMS.

Enumeration

Here is an example of a lossless source encoding scheme:

1. *Partitioning:* Divide all sequences in $\mathcal{X}^n$ into two sets: the typical set $A_\varepsilon^{(n)}$ 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 $A_\varepsilon^{(n)}$, use a 0 followed by $\lceil n(H(x) + \varepsilon) \rceil$ bits corresponding to its index. For each sequence in the complement of $A_\varepsilon^{(n)}$, use 1 followed by $\lceil n \log |\mathcal{X}| \rceil$ bits[11] 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.

[11] $\left\lceil n \log |\mathcal{X} \setminus A_\varepsilon^{(n)}| \right\rceil$ bits are sufficient, but $\lceil n \log |\mathcal{X}| \rceil$ is good enough for efficient compression.

Random Binning

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

1. *Encoding:*[12] For each sequence in $\mathcal{X}^n$, 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 $\mathcal{X}^n$ 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:*[13] 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.

[12] SOURCE SEQUENCE $\longrightarrow$ INDEX

[13] INDEX $\longrightarrow$ SOURCE SEQUENCE

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.

Quantum Computing Systems [1]

Prof. Emina Soljanin

Lecture #15, March 25

[1] Rutgers, ECE 579, Spring 2021

This lecture introduces classical and quantum error correcting codes. Codes play indispensable roles in numerous scientific disciplines and virtually all telecommunications and computing systems. Today, we even ask if the space and timecould be a quantum error-correcting code.

Error correcting codes add redundancy to data in order to make it less sensitive to errors. The most basic form of redundancy is simple replication (cloning), 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, after first finding out (measuring) what the value of the majority is. But could there be a counterpart to this process in the quantum world where the no-cloning theorem holds and the measurements disturb the states?[2] We will first 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.

[2] *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.

A Classical Error Correcting Code

- Encoding is a map that introduces redundancy. In our 1-to-3 bits repetition code example, each bit x is mapped to a 3-bit string (codeward) $x\,x\,x$, that is, the encoding is the following map:

$$0 \to 000 \quad \text{and} \quad 1 \to 111$$

- Decoding is the inverse map of the encoding. It removes the redundancy by the encoder. In this example, decoding reduces to keeping the first bit of the 3-bit codeward.

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

additive error	y_0	y_1	y_2
000	x	x	x
100	$x\oplus 1$	x	x
010	x	$x\oplus 1$	x
001	x	x	$x\oplus 1$

- Measurements: We perform the following matrix vector multiplication (cf. two measurements):

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} y_0 \oplus y_1 \\ y_0 \oplus y_2 \end{bmatrix} \tag{1}$$

  We refer to the vector on the right-hand side in the above equation as the *error syndrome*. Observe that the first bit of the syndrome tells us if whether bits y_0 and y_1 have identical values and the second bit of the syndrome tells us if whether bits y_0 and y_2 have identical values.

- Error Correction: The 2-bit measurement result (syndrome bits $y_0 \oplus y_1$, $y_0 \oplus y_2$) tells us which bit is flipped, and thus instructs us how to correct errors as follows:

y_0	y_1	y_2	$y_0 \oplus y_1$	$y_0 \oplus y_2$	add
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

  The error is corrected by adding the string in the last column to the received word.

A Quantum Error Correcting Code

Quantum error correction has to follow the laws of quantum mechanics. Therefore all actions on qubits (encoding, errors, decoding) have to be either unitary or measurements. We describe the simplest code only to show that quantum error correction under these constraints is feasible, and possibly make the reader interested in this fascinating subject. [3]

Building scalable quantum computers will require not only further research in quantum information processing, but also further research in many relevant classical fields. Error correcting codes will be an indispensable part of any quantum system regardless of its physical qubit realization. However, errors are realization-specific, and thus will require tailored as well as multi-purpose error correction, which will have to be done in both classical and quantum domain.

- 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:

$$\alpha |0\rangle + \beta |1\rangle \rightarrow \alpha |000\rangle + \beta |111\rangle$$

[3] *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.

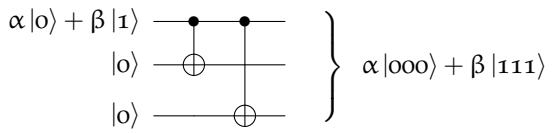


Figure 1: Quantum circuit that maps $(\alpha|0\rangle + \beta|1\rangle) \otimes |0\rangle \otimes |0\rangle$ to $\alpha|000\rangle + \beta|111\rangle$.

The unitary map circuit shown in Fig. 1 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.

- Decoding is the inverse map of the encoding, as in the classical case. How does a unitary map circuit that reverses the action shown in Fig. 1 look like?
- Error Model: We assume that at most one qubit experiences the basis flip (i.e., is acted on by σ_X). The possible 3-qubit error operators and the resulting states they give when acting on $\alpha|000\rangle + \beta|111\rangle$ are as follows:

error operators	resulting state
$I \otimes I \otimes I$	$\alpha\|000\rangle + \beta\|111\rangle$
$\sigma_X \otimes I \otimes I$	$\alpha\|100\rangle + \beta\|011\rangle$
$I \otimes \sigma_X \otimes I$	$\alpha\|010\rangle + \beta\|101\rangle$
$I \otimes I \otimes \sigma_X$	$\alpha\|001\rangle + \beta\|110\rangle$

- Measurements: As in the classical case, the idea is to have two measurements such that one compares qubits 1 and 2, and the other compares qubits 1 and 3. The additional constraint here is that the measuring process leave the measured states unchanged. We perform the following two measurements:

  $\boxed{M_1}$: This measurement is defined by the Hermitian operator $\sigma_Z \otimes \sigma_Z \otimes I$, i.e., the following two orthogonal projection operators:

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

  Π_1 projects on the eigenspace of $\sigma_Z \otimes \sigma_Z \otimes I$ with eigenvalue 1, and Π_2 projects on the eigenspace of $\sigma_Z \otimes \sigma_Z \otimes I$ with eigenvalue -1.

  $\boxed{M_2}$: This measurement is defined by the Hermitian operator $\sigma_Z \otimes I \otimes \sigma_Z$, i.e., the following two orthogonal projection operators:

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

  Π_1 projects on the eigenspace of $\sigma_Z \otimes I \otimes \sigma_Z$ with eigenvalue 1, and Π_2 projects on the eigenspace of $\sigma_Z \otimes I \otimes \sigma_Z$ with eigenvalue -1.

- Error Correction: The results of the two measurements are two eigenvalues (M_1 and M_2 in the table below). As in the classical case, we refer to this result as the error syndrome, which instructs us how to correct errors, as follows:

corrupted state	M_1	M_2	apply
$\alpha\|000\rangle + \beta\|111\rangle$	$+1$	$+1$	$I \otimes I \otimes I$
$\alpha\|100\rangle + \beta\|011\rangle$	-1	-1	$\sigma_X \otimes I \otimes I$
$\alpha\|010\rangle + \beta\|101\rangle$	-1	$+1$	$I \otimes \sigma_X \otimes I$
$\alpha\|001\rangle + \beta\|110\rangle$	$+1$	-1	$I \otimes I \otimes \sigma_X$

  The error is corrected by applying the unitary operator in the last column to the three received qubits.

Remark: The error detecting and correcting 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 von Neumann measurement defined by the following set of projectors:

$$\Pi_0 = |000\rangle\langle 000| + |111\rangle\langle 111| \quad \text{no error}$$
$$\Pi_1 = |100\rangle\langle 100| + |011\rangle\langle 011| \quad \text{bit flip on qubit one}$$
$$\Pi_2 = |010\rangle\langle 010| + |101\rangle\langle 101| \quad \text{bit flip on qubit two}$$
$$\Pi_3 = |001\rangle\langle 001| + |110\rangle\langle 110| \quad \text{bit flip on qubit three}$$

The result of the measurement M (error syndrome) takes values in the set $\{0, 1, 2, 3\}$ corresponding to the four projectors. A measurement result different than 0 means that an error has been detected. Errors are corrected based on the measurement result as follows:

corrupted state	M	apply
$\alpha\|000\rangle + \beta\|111\rangle$	0	$I \otimes I \otimes I$
$\alpha\|100\rangle + \beta\|011\rangle$	1	$\sigma_X \otimes I \otimes I$
$\alpha\|010\rangle + \beta\|101\rangle$	2	$I \otimes \sigma_X \otimes I$
$\alpha\|001\rangle + \beta\|110\rangle$	3	$I \otimes I \otimes \sigma_X$

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.

As their classical counterparts, decoders of quantum error correcting codes can miss-correct or not-detect certain errors. For example, the decoder above will miss-correct the two-qubit error introduced by the operator $\sigma_X \otimes \sigma_X \otimes I$, and it will not detect three-qubit error $\sigma_X \otimes \sigma_X \otimes \sigma_X$ and even a single-qubit error $\sigma_Z \otimes I \otimes I$.

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

Prof. Emina Soljanin

Lecture #16, March 30

This lecture presents two similarity measures for density matrices and discusses blind quantum source coding.

Claude Shannon has written that the *fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point*. To say what *approximately* means in quantum communications, we need to have a notion of distance and/or similarity between quantum states.

How Close are Two Quantum States?

How Close are Two Probability Vectors?

Let $P = \{p_1, \dots, p_k\}$ and $Q = \{q_1, \dots, q_k\}$ 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.[2]

[2] Log of BC is *Bhattacharyya ditance.*

Fidelity and Trace Distance

To measure how faithfully mixed state σ approximates mixed state ω and vice versa, we use the so called *mixed state fidelity* F defined as

$$F(\sigma, \omega) = \left\{ \mathrm{Tr}\left[(\sqrt{\sigma}\omega\sqrt{\sigma})^{1/2} \right] \right\}^2,$$

Besides computing the mixed state fidelity, we can measure how close state σ is to state ω by computing the *trace distance*

$$D(\sigma, \omega) = \frac{1}{2} \mathrm{Tr} |\sigma - \omega|.$$

where $|A|$ is the positive square root of $A^\dagger A$, *i.e.*, $|A| = \sqrt{A^\dagger A}$. The trace distance is a metric on the space of density operators, and is closely related to the fidelity as follows:

$$1 - F(\sigma, \omega) \leqslant D(\sigma, \omega) \leqslant \sqrt{1 - F(\sigma, \omega)^2}. \tag{1}$$

The trace distance and the fidelity generalize the classical measures of distance/similarity between probability distributions. When matrices σ and ω are simultaneously diagonalizable (commute), the trace distance is equal to the total variation between their eigenvalues, and the fidelity is equal to the squared Bhattacharyya coefficient of the their eigenvalues.

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[3] where source letter $a \in \mathcal{X}$ is mapped into qubit $|\psi_a\rangle$.

[3] Mixed states compression has not been fully understood.

The Density Matrix and the Entropy of the Source

The Source density matrix is defined as

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

and its von Neumann entropy is

$$S(\rho) = -\operatorname{Tr} \rho \log \rho = -\sum_i \lambda_i \log \lambda_i,$$

where λ_i are the eigenvalues of ρ.

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 $|\Psi_{\mathbf{x}}\rangle \in \mathcal{H}_d^{\otimes n}$ that corresponds to source sequence $\mathbf{x} = x_1, x_2, \ldots, x_n \in \mathcal{X}^n$ is given by

$$|\Psi_{\mathbf{x}}\rangle = |\psi_{x_1}\rangle \otimes |\psi_{x_2}\rangle \otimes \cdots \otimes |\psi_{x_n}\rangle, \qquad x_i \in \mathcal{X},$$

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

The Typical Subspace Λ_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

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

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 Λ. We say that sequence $z = z_1, \ldots, z_n \in \{0,1\}^n$ is weakly ϵ_n-typical if

$$2^{-n(H(\Lambda)+\epsilon_n)} \leqslant \Lambda(z_1, \ldots, z_n) \leqslant 2^{-n(H(\Lambda)-\epsilon_n)}$$

The set of all such sequences $A^{\Lambda}_{\epsilon_n}$ is the typical set according to distribution Λ. There are approximately $2^{nH(\Lambda)} = 2^{nS(\rho)}$ such sequences.

We define the typical subspace $\boldsymbol{\Lambda_n}$ to be the subspace spanned by the typical states $|\Phi_z\rangle$, $z \in A^{\Lambda}_{\epsilon_n}$. We define the projector to $\boldsymbol{\Lambda_n}$ and its complement:

$\Pi = \sum_{z\in A^{\Lambda}_{\epsilon_n}} |\Phi_z\rangle\langle\Phi_z|$ is the projector to $\boldsymbol{\Lambda_n}$.

$\Pi^{\perp} = \sum_{z\in\{0,1\}^n \setminus A^{\Lambda}_{\epsilon_n}} |\Phi_z\rangle\langle\Phi_z|$ is the projector to $\boldsymbol{\Lambda_n}^{\perp}$.

$\Pi + \Pi^{\perp} = I_{2^n}$. The dimension of $\boldsymbol{\Lambda_n}$ is approximately $2^{nS(\rho)}$.

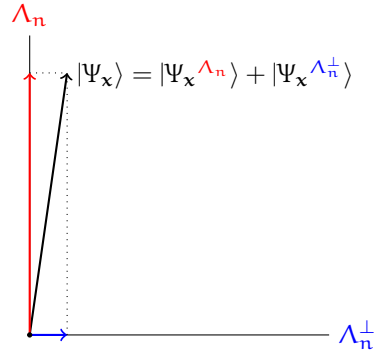


Figure 1: Each source vector state $|\Psi_x\rangle$ is a sum of its projections to $\boldsymbol{\Lambda_n}$ and $\boldsymbol{\Lambda_n}^{\perp}$.

Vector-Sequences and Fidelity

Recall that source vector-sequences $|\Psi_x\rangle$ are in $\mathcal{H}^{2^n}$, $(x \in \mathcal{X}^n)$. Vector $|\Psi_x\rangle$ is compressed and then reproduced as $\left|\widehat{\Psi_x}\right\rangle$. The fidelity between $|\Psi_x\rangle$ and $\left|\widehat{\Psi_x}\right\rangle$ is

$$F(|\Psi_x\rangle, \left|\widehat{\Psi_x}\right\rangle) = |\langle\Psi_x|\widehat{\Psi_x}\rangle|^2$$

For asymptotically lossless compression, the expected fidelity

$$\bar{F} = \sum_{x\in\mathcal{X}^n} P(x)F(|\Psi_x\rangle, \left|\widehat{\Psi_x}\right\rangle)$$

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 Π, $\Pi^{\perp}$. The state after the measurement is

1. $\Pi \cdot |\Psi_x\rangle / \sqrt{\langle\Psi_x|\Pi|\Psi_x\rangle} = \left|\Psi_x{}^{\Lambda_n}\right\rangle$ 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 $\left|\Psi_x{}^{\Lambda_n}\right\rangle$. Note that this is still an n-qubit state. However, Alice can apply a unitary change of basis U that takes each state $\left|\Psi_x{}^{\Lambda_n}\right\rangle$ to a state of the form $|\Psi_x{}^c\rangle \otimes |o_r\rangle$, where state $|\Psi_x{}^c\rangle$ consists of $|A_{\epsilon_n}^{\Lambda}|$ (approximately $2^{nS(\rho)}$) qubits, and $|o_r\rangle$ is the $(2^n - |A_{\epsilon_n}^{\Lambda}|)$ - fold Kronecker product of $|o\rangle$ states. Alice then sends $|\Psi_x{}^c\rangle$ to Bob. Bob recovers $\left|\Psi_x{}^{\Lambda_n}\right\rangle$ 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 Alice:

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 $\left|\Psi_x{}^{\Lambda_n}\right\rangle$ (result 1) or $\left|\Psi_x{}^{\Lambda_n^{\perp}}\right\rangle$ (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.
$$\left|\Psi_x{}^{\Lambda_n}\right\rangle \xrightarrow{U} |\Psi_x{}^c\rangle \otimes |o_r\rangle$$
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. $|\Psi_x{}^c\rangle \otimes |o_r\rangle \xrightarrow{U} \left|\Psi_x{}^{\Lambda_n}\right\rangle$

Bob's reconstructed state is the normalized state $\left|\widehat{\Psi_x}\right\rangle = \Pi \cdot |\Psi_x\rangle \,/ \sqrt{\langle\Psi_x|\,\Pi\,|\Psi_x\rangle}$

Blind Compression Fidelity

How close are $|\Psi_x\rangle$ and $\left|\widehat{\Psi_x}\right\rangle$ on average?

$$\begin{aligned}
\bar{F} &= \sum_{x\in\mathcal{X}^n} P(x) F(|\Psi_x\rangle, \left|\widehat{\Psi_x}\right\rangle) \\
&= \sum_{x\in\mathcal{X}^n} P(x) \left[\Pr(\text{result } 1|x) F(|\Psi_x\rangle, \left|\widehat{\Psi_x}\right\rangle) + \Pr(\text{result } 2|x) F(|\Psi_x\rangle, \left|\widehat{\Psi_x}\right\rangle)\right] \\
&\geqslant \sum_{x\in\mathcal{X}^n} P(x) \Pr(\text{result } 1|x) F(|\Psi_x\rangle, \left|\widehat{\Psi_x}\right\rangle) \\
&= \sum_{x\in\mathcal{X}^n} P(x) \,\langle\Psi_x|\,\Pi\,|\Psi_x\rangle \left[\langle\Psi_x|\,\Pi\,|\Psi_x\rangle \,/ \sqrt{\langle\Psi_x|\,\Pi\,|\Psi_x\rangle}\right]^2 \\
&= \sum_{x\in\mathcal{X}^n} P(x) |\,\langle\Psi_x|\,\Pi\,|\Psi_x\rangle\,|^2 \geqslant \sum_{x\in\mathcal{X}^n} P(x)(1 - 2\,\langle\Psi_x|\,\Pi\,|\Psi_x\rangle) = -1 + 2\,\mathrm{Tr}(\Pi\rho^{\otimes n}) \\
&= -1 + 2\,\mathrm{Tr}\Big\{\Big[\sum_{z\in A_{\epsilon_n}^{\Lambda}} |\Phi_z\rangle\langle\Phi_z|\Big] \cdot \Big[\sum_{z\in\{0,1\}^n} \lambda(z) |\Phi_z\rangle\langle\Phi_z|\Big]\Big\} = 1 - 2\epsilon_n
\end{aligned}$$

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

Prof. Emina Soljanin

Lecture #17, April 1

This lecture uncovers a little more about quantum error correction.

Consider again the code that maps a single-qubit state into a 3-qubit state as follows:

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

Recall that this code corrects single σ_X errors[2] that can be identified by e.g., the von Neumann measurement defined by the following set of projectors:

error operators	resulting state
$I \otimes I \otimes I$	$\alpha\|000\rangle + \beta\|111\rangle$
$\sigma_X \otimes I \otimes I$	$\alpha\|100\rangle + \beta\|011\rangle$
$I \otimes \sigma_X \otimes I$	$\alpha\|010\rangle + \beta\|101\rangle$
$I \otimes I \otimes \sigma_X$	$\alpha\|001\rangle + \beta\|110\rangle$

[2]

$$\Pi_0 = |000\rangle\langle 000| + |111\rangle\langle 111| \quad \text{no error}$$
$$\Pi_1 = |100\rangle\langle 100| + |011\rangle\langle 011| \quad \text{bit flip on qubit one}$$
$$\Pi_2 = |010\rangle\langle 010| + |101\rangle\langle 101| \quad \text{bit flip on qubit two}$$
$$\Pi_3 = |001\rangle\langle 001| + |110\rangle\langle 110| \quad \text{bit flip on qubit three}$$

We need to know which errors are possible before we design a code.

Correcting Phase Flips

When the error operator $E = \sigma_Z \otimes I \otimes I$ acts on the encoded state $|\psi\rangle$, the result is the corrupted state $|\varphi\rangle$:

$$E(\underbrace{\alpha|000\rangle + \beta|111\rangle}_{|\psi\rangle}) = \underbrace{\alpha|000\rangle - \beta|111\rangle}_{|\varphi\rangle}.$$

Observe that both $|\psi\rangle$ and $|\varphi\rangle = E|\psi\rangle$ belong to the subspace of $\mathbb{C}^8$ spanned by $|000\rangle$ and $|111\rangle$, and are, in general, not orthogonal.[3]

[3] Is there a quantum measurement that can distinguish between $|\psi\rangle$ and $|\varphi\rangle$, that is, error and no-error?

Consider the code that maps one information qubit (and some ancillary qubits) into three encoded qubits by the encoder shown in Fig. 1.

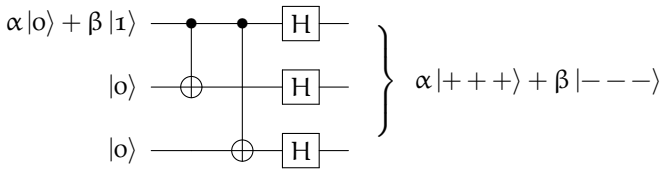


Figure 1: Quantum circuit that maps $(\alpha\,|0\rangle + \beta\,|1\rangle) \otimes |0\rangle \otimes |0\rangle$ to $\alpha|+++\rangle + \beta|---\rangle$.

Note that the basis vectors are mapped as follows:

$$|0\rangle \quad \to \quad \big((|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle)\big)/2\sqrt{2}$$
$$|1\rangle \quad \to \quad \big((|0\rangle - |1\rangle) \otimes (|0\rangle - |1\rangle) \otimes (|0\rangle - |1\rangle)\big)/2\sqrt{2}$$

Show that this code correct single σ_Z (phase) errors but not single σ_X errors (bit flips).

Correcting Both Bit and Phase Flips

We can correct both bit and phase flips, if we first encode the qubit using the phase flip code into three qubits, and then encode each of these three qubits using the bit flip code.[4] This code maps one information qubit (and some ancillary qubits) into nine encoded qubits as follows:

[4] This code is known as Shor's Code.

$$\begin{aligned} |0\rangle &\rightarrow (|000\rangle + |111\rangle) \otimes (|000\rangle + |111\rangle) \otimes (|000\rangle + |111\rangle) \\ |1\rangle &\rightarrow (|000\rangle - |111\rangle) \otimes (|000\rangle - |111\rangle) \otimes (|000\rangle - |111\rangle) \end{aligned}$$

We can use the measurements defined by the following observables to learn about the error syndrome, and make error corrections accordingly.

$$\begin{aligned} &(\sigma_Z \otimes \sigma_Z \otimes I) \otimes (I \otimes I \otimes I) \otimes (I \otimes I \otimes I) \\ &(\sigma_Z \otimes I \otimes \sigma_Z) \otimes (I \otimes I \otimes I) \otimes (I \otimes I \otimes I) \\ &(I \otimes I \otimes I) \otimes (\sigma_Z \otimes \sigma_Z \otimes I) \otimes (I \otimes I \otimes I) \\ &(I \otimes I \otimes I) \otimes (\sigma_Z \otimes I \otimes \sigma_Z) \otimes (I \otimes I \otimes I) \\ &(I \otimes I \otimes I) \otimes (I \otimes I \otimes I) \otimes (\sigma_Z \otimes \sigma_Z \otimes I) \\ &(I \otimes I \otimes I) \otimes (I \otimes I \otimes I) \otimes (\sigma_Z \otimes I \otimes \sigma_Z) \\ &(\sigma_X \otimes \sigma_X \otimes \sigma_X) \otimes (\sigma_X \otimes \sigma_X \otimes \sigma_X) \otimes (I \otimes I \otimes I) \\ &(I \otimes I \otimes I) \otimes (\sigma_X \otimes \sigma_X \otimes \sigma_X) \otimes (\sigma_X \otimes \sigma_X \otimes \sigma_X) \end{aligned}$$

Quantum and Classical Error Correcting Codes

An $[n, k]_q$ classical linear error correcting code (ECC) is a k dimensional subspace $\mathcal{C}$ of $\mathbb{F}_q^n$. The encoder for an $[n, k]_q$ linear code is a linear map $\mathbb{F}_q^k \rightarrow \mathbb{F}_q^n$, which maps k bit-strings (data words, information words) into n-bit strings (codewords).

A quantum error-correcting encoder maps k-qubit strings and $n-k$ ancillary qubits into n-bit qubits.[5] A quantum error-correcting code is a 2^k dimensional subspace $\mathcal{C}$ of the 2^n-dimensional Hilbert space.

[5] The encoder is a unitary map that acts on an n-qubit register.

Requirements for Error Correction

For a classical ECC, errors $\boldsymbol{e}_p$ and $\boldsymbol{e}_q$ are both correctable if and only if they send different codewords into different bit-strings, that is,

$$(\boldsymbol{c}_i \oplus \boldsymbol{e}_p) \oplus (\boldsymbol{c}_j \oplus \boldsymbol{e}_q) \neq 0$$

for all codewords $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$.

For a quantum error correcting code (QECC), errors E and F are correctable if and only if the following holds:

1. We can distinguish error events from no-error events, that is,

$$\langle\psi| \cdot E |\psi\rangle = 0 \quad \text{for all} \quad |\psi\rangle \in \mathcal{C}.$$

2. We can distinguish error E from any other error F,[6] that is,

$$\langle i| E \cdot F |j\rangle = 0$$

for all basis states $|i\rangle$ and $|j\rangle$ (including $i = j$). Note that if we include the identity (no error) as a possibility for E or F, we obtain the first condition above.

[6] Otherwise, we may attempt to correct error E when error F occurred.

A General Error Model

Why is it important to be able to correct errors described by Pauli matrices? If a QECC corrects errors E and F, it also corrects any linear combination of E and F. Recall that σ_X, σ_Y, σ_Z, and I span the space of 2×2 matrices, and matrices $E_0 \otimes \cdots \otimes E_{n-1}$, $E_i \in \{I, \sigma_X, \sigma_Y, \sigma_Z\}$, span the space of $2^n \times 2^n$ matrices.

The Stabilizer Codes – A General Class of QECCs

One way to ensure that the conditions we listed above for quantum error correction hold is to have code $\mathcal{C}$ lie in the +1-eigenspace of some operator M:

$$\mathcal{C} = \{|\psi\rangle : M|\psi\rangle = |\psi\rangle\}$$

Note that if EF anticommutes[7] with M, then EF will take states from the +1-eigenspace of M to the -1-eigenspace of M, which is orthogonal to the +1-eigenspace, that is,

[7] Matrices A and B anticommute if $AB = -BA$.

$$\begin{aligned} &\langle i| EF |j\rangle = \langle i| EF \cdot M |j\rangle = -\langle i| M \cdot EF |j\rangle = -\langle i| EF |j\rangle \\ &\implies \langle i| EF |j\rangle = 0. \end{aligned}$$

It is convenient to pick M to be a Krinecker product of the Pauli matrices, because then other products of Pauli matrices[8] will always either commute or anticommute with M. By choosing our QECC $\mathcal{C}$ to be in the +1-eigenspace of enough such operators,[9] we can make sure that EF anticommutes with one of the M's for any pair of E and F. The set of such operators is called the *stabilizer* of the code.

[8] the errors we want to correct

[9] $\implies$ the operators must commute with each other.

Example: A one-to-five qubit QECC that corrects one general error[10] has the following stabilizer:

[10] The highest rate such code.

$$\begin{aligned} M_1 &= X \otimes Z \otimes Z \otimes X \otimes I \\ M_2 &= I \otimes X \otimes Z \otimes Z \otimes X \\ M_3 &= X \otimes I \otimes X \otimes Z \otimes Z \\ M_4 &= Z \otimes X \otimes I \otimes X \otimes Z \end{aligned}$$

Not all codes can be described by a stabilizer.

Outline of a Formal Definition – Advanced

- $\mathcal{G}_n$: the group of $2^n \times 2^n$ matrices of the form

$$U = U_0 \otimes U_1 \otimes \cdots \otimes U_{n-1},$$

  $U_i \in \{\pm I, \pm X, \pm Y, \pm Z\}$, $X = \sigma_X$, $Z = \sigma_Z$, $Y = -i\sigma_Y$.

- $\mathcal{S}$: an Abelian subgroup of $\mathcal{G}_n$.

- $\mathcal{C}$: the simultaneous eigenspace of the elements of $\mathcal{S}$ corresponding to the trivial character:

$$\mathcal{S} = \{M \in \mathcal{G}_n : M|\psi\rangle = |\psi\rangle \text{ if } |\psi\rangle \in \mathcal{C}\}$$

- $\mathcal{C}$ is a QECC and $\mathcal{S}$ its *stabilizer*.

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

Prof. Emina Soljanin

Lecture #18, April 6

This lecture describes the Variational Quantum Eigensolver (VQE), which is quantum/classical hybrid algorithms for NISQ computing.

Noisy Intermediate-Scale Quantum (NISQ) technology is expected to be available in the near future. This term, coined by John Preskill, refers to devices with 50-100 qubits (intermediate-scale), too few to have have full error-correction, hence the attribute noisy. Nevertheless, NISQ systems may be able to perform tasks that exceed the capabilities of today's classical digital computers, and may be useful tools for exploring many-body quantum physics. One such task is finding the minimum eigenvalue a large matrix. This is an optimization problems that is crucial in many domains, ranging from Google's Page Rank and aircraft design to quantum simulation and quantum chemistry.

We will use a quantum/classical hybrid algorithm known as the Variational Quantum Eigensolver (VQE) to find the minimum eigenvalue of a large matrix that represents the Hamiltonian of an n qubit system. A Hamiltonian is a Hermitian matrix, but the algorithm can be extended to include arbitrary matrices. Hybrid quantum/classical algorithms iterate between quantum and classical processing, as shown in Fig. 1.

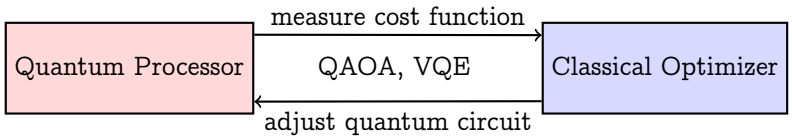


Figure 1: NISQ Quantum/Classical hybrid architecture

The quantum processor performs computation that would be hard to perform classically. In the VQE algorithm where the goal is to compute the smallest eiginvelue of a large Herimitian matrix H, the quantum processor evaluates expressions of the form $\langle\psi| H |\psi\rangle$, where $|\psi\rangle$ is a parameterized quantum state. This computation is relevant because of what is known as *the variational principle,* which we explain next.

Physics Interlude

Schrödinger's Equation

The time evolution of the state of a closed quantum system $|\psi(t)\rangle$ is governed by the Schrödinger equation:

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = H \cdot |\psi(t)\rangle \tag{1}$$

where $\hbar$ is the reduced Plank's constant and H is a Hermitian matrix known as the *Hamiltonian* of the system. If the quantum system consists of n qubits, it can be represented by a unit norm vector in an $N = 2^n$ dimensional Hilbert space. When H does not depend on time, and we know the state at time 0, then

$$|\psi(t)\rangle = U(t) |\psi(0)\rangle \tag{2}$$

where $U(t) = \exp\left(-\frac{i}{\hbar} Ht\right)$ is a unitary matrix. [2]

[2] U is unitary because H is Hermitian.

The Hamiltonian describes the physical model. The Schrödinger equation tells us how a state-vector evolves in time given the physical model described by the Hamiltonian.[3] The ground state is defined to be the eigenvector of H with the smallest eigenvalue.

[3] How do the sates that happen to be the eigenvectors of the Hamiltonian evolve in time?

Solving a Schrödinger Equation

Let state $|\varphi\rangle$ be an eigenvector of H, i.e., $H|\varphi\rangle = E|\varphi\rangle$. Then state $|\phi\rangle = |\varphi\rangle \exp\left(-\frac{i}{\hbar} Et\right)$ is a solution of the Schrödinger equation (1). Check! (When we know all eigenvectors of H, we know all solutions of (1), but this is not our concern at the moment.)

Let $|\varphi_i\rangle$ and the E_i for $i = 1, \ldots, N$ be the eigenstates and eigenvalues of H, i.e., [4]

$$H|\varphi_i\rangle = E_i |\varphi_i\rangle, \quad i = 1, \ldots, N.$$

[4] Recall that $|\varphi_i\rangle$ are orthonormal: $\langle\varphi_i|\varphi_j\rangle = \delta_{ij}$ and $\sum |\varphi_i\rangle\langle\varphi_i| = I_N$.

Since $|\varphi_i\rangle$, $i = 1, \ldots, N$, form a basis, any other vector ψ can be expressed as their linear combination:

$$|\psi\rangle = \sum_{i=1}^{N} c_i |\varphi_i\rangle, \quad c_i = \langle\varphi_i|\psi\rangle, \tag{3}$$

where $\sum_{i=1}^{N} |c_i|^2 = 1$ when $|\psi\rangle$ is a quantum (thus unit-norm) state. Note that H can be expressed as follows:

$$H = \sum_{i=1}^{N} E_i |\varphi_i\rangle\langle\varphi_i|. \tag{4}$$

We can think of H as an observable associated with the energy of the quantum system. If we measure state $|\psi\rangle$ in the basis $|\varphi_i\rangle$, we

will get the result E_i (energy level) with probability $|c_i|^2 = |\langle\psi|\varphi_i\rangle|^2$. Therefore, the expected value of the measurement is equal to

$$= \sum_i E_i |c_i|^2 = \sum_{i=1}^{N} |\langle\psi|\varphi_i\rangle|^2 E_i$$

By using the identities (3) and (4), we can show that

$$\langle\psi| H |\psi\rangle = \sum_{i=1}^{N} |\langle\psi|\varphi_i\rangle|^2 E_i.$$

as follows:

$$\begin{aligned}\langle\psi| H |\psi\rangle &= \sum_i c_i^* \langle\varphi_i| H \sum_j c_j \left|\varphi_j\right\rangle \\ &= \sum_{ij} c_i^* c_j \langle\varphi_i| H \left|\varphi_j\right\rangle = \sum_i E_i |c_i|^2\end{aligned}$$

Note that for eigenstate $|\varphi_i\rangle$, we have $\langle\varphi_i| H |\varphi_i\rangle = E_i$.

Let $E_0 < E_1 < E_2 < \cdots < E_{N-1}$. We say that $|\varphi_0\rangle$ is the ground-state, and $|\varphi_i\rangle$ the i-th excited state. Thus, $|\psi_0\rangle$ is the ground-state, $|\psi_1\rangle$ the first excited state, and so on. We see that

$$\langle\varphi_0| H |\varphi_0\rangle = E_0 < E_i = \langle\varphi_i| H |\varphi_i\rangle \quad \text{for} \quad i > 0.$$

Consider now $\langle\psi| H |\psi\rangle$ for an arbitrary state $|\psi\rangle = \sum_{i=1}^{N} c_i |\varphi_i\rangle$:

$$\langle\psi| H |\psi\rangle = \sum_i E_i |c_i|^2 \geqslant \sum_i E_0 |c_i|^2 = E_0$$

Therefore, the ground state energy is always smaller than the expectation of the energy calculated for any other state $|\psi\rangle$. Thus, by varying ψ until $\langle\psi| H |\psi\rangle$ is minimized,[5] we can obtain an approximation to the state and energy of the ground state. This is known as the variational principle. The variational principle is more general than its applications in quantum computing, physics, and chemistry.

[5] This is often refered as minimizing of the expectation of H.

VQE

In variational methods, we start with a best guess, called *ansatz*, for the ground state. We parameterize the ansatz by a set of parameters θ. Let $|\psi(\theta)\rangle$ denote the parameterized quantum state. Note that our ansatz may not be equal to the ground state that we want to find for any value of the parameter.

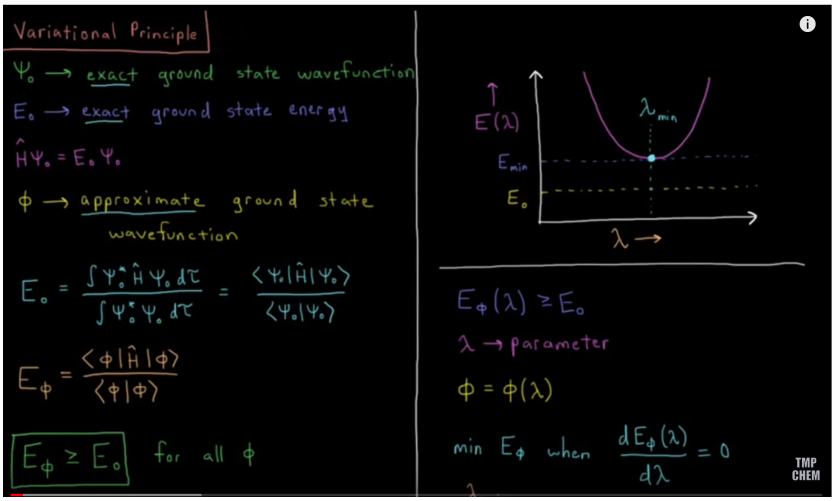


Figure 2: A video explanation of the variational principle.

The Hamiltonian for VQE

Any 2×2 complex matrix, and thus any Hamiltonian H, can be expressed as a linear combination of the identity I and the Pauli matrices σ_X, σ_Y, and σ_Z:

$$H = \alpha_I I + \alpha_X \sigma_X + \alpha_Y \sigma_Y + \alpha_Z \sigma_Z$$

We represent the n qubit Hamiltonian H as a weighted sum of tensor products of Pauli matrices:

$$H = \sum_{i=1}^{m} c_i H_i \tag{5}$$

A NISQ VQE only considers the Hamiltonians where m grows at most polynomially in the system size, that is, $m = \mathcal{O}(n^k)$ which is a reasonable assumption for many physical systems of interest.

Pauli matrices form a basis for $\mathbb{C}^{2\times 2}$

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \sigma_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\sigma_Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad \sigma_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

The n-fold tensor products of Pauli matrices (I, σ_x, σ_y, σ_z) form an orthogonal basis for the vector space of $2^n \times 2^n$ complex matrices.

A Hybrid Quantum/Classical VQE Algorithm

A variational algorithm iterates between two modules. The first module computes the expected energy for the ansatz. The second module finds the parameters for which the expected energy is minimized, and possibly comes up with another parameterized ansatz to give back to the first module for the expected energy evaluation.

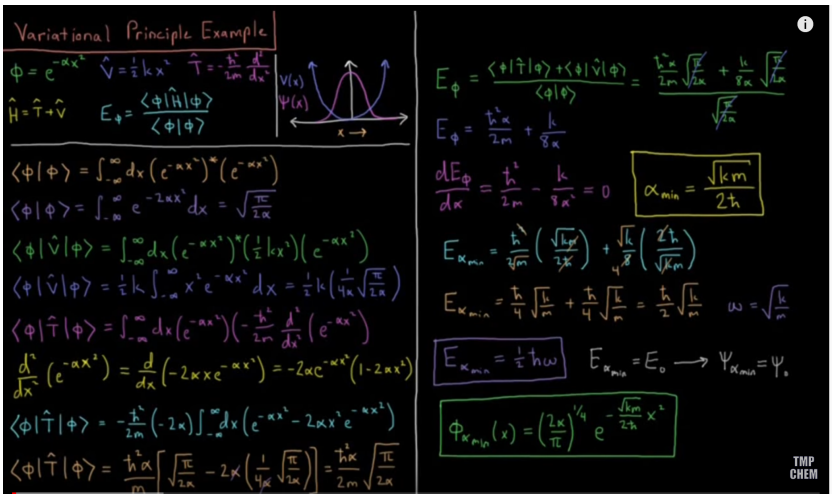


Figure 3: A video example of using the variational principle.

The hybrid quantum/classical algorithm algorithm computes expectation values of each term H_i in (5) using a NISQ circuit, and then adds the total energy classically. A classical optimizer finds the parameters for which expected energy is minimized, and possibly comes up with another parameterized ansatz to give back to the quantum processor for the expected energy evaluation.

Quantum Computing Systems [1]

Prof. Emina Soljanin

Lecture #19, April 8

[1] Rutgers, ECE 579, Spring 2021

This lecture is concerned with local Hamiltonians of many qubit systems and some associated problems of quantum computational complexity.

Many-Qubit Hamiltonians

An n-qubit state is a vector in $\mathcal{H}_{2^n} = \underbrace{\mathcal{H}_2 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_2}_{n}$. The Hamiltonian H of an n qubit system is a $2^n \times 2^n$ matrix. Many-body Hamiltonians associated with naturally occurring many-body systems exhibit locality. Some examples are shown in Fig. 1. But, not all Hamiltonins that we need for quantum computing can be build based on the naturally occurring.

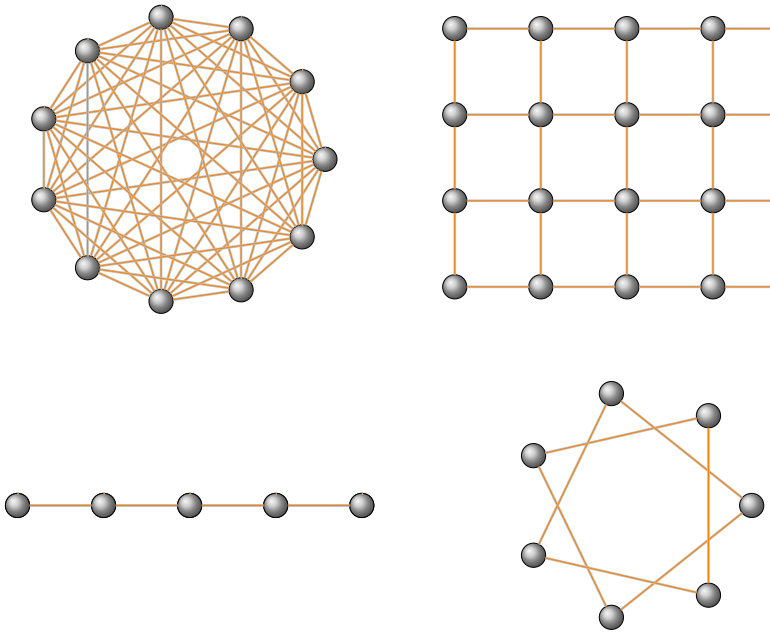


Figure 1: Interactions in many-qubit systems.

We are often interested in the minimum energy E_o and the ground state $|\varphi_o\rangle$ of the system with the Hamiltonian H. Recall that $|\varphi_o\rangle$ is the eigenvector of H corresponding to the minimum eigenvalue E_o, i.e., $H|\varphi_o\rangle = E_o|\varphi_o\rangle$. What can we say about E_o in some special cases?[2]

[2] By the variational principle, we know that $E_o = \min_{|\psi\rangle} \langle\psi| H |\psi\rangle$. Last time, we explained how a NISQ system can use this principle to compute E_o.

Local Hamiltonians

Recall that any 2×2 complex matrix, and thus any Hamiltonian H acting on a qubit, can be expressed as a linear combination of the identity I and the Pauli matrices σ_X, σ_Y, and σ_Z:

$$H = \alpha_I I + \alpha_X \sigma_X + \alpha_Y \sigma_Y + \alpha_Z \sigma_Z$$

The n-fold tensor products of Pauli matrices (I, σ_x, σ_y, σ_z) form an orthogonal basis for the vector space of $2^n \times 2^n$ complex matrices. Thus, we can represent the n qubit Hamiltonian H as a weighted sum of Kronecker products of Pauli matrices.

The Hamiltonian H of an n qubit system is usually given as a sum of terms H_i, each describing an interaction among a (small) subset of qubits, that is, H is a Hermitian matrix with a succinct representation of the form

$$H = \sum_{i=1}^{m} H_i \tag{1}$$

We say that H is a k-local Hamiltonian if each H_i acts "non-trivially" only on some small subset of k qubits where k is a constant that is independent of the system size n. Here, each H_i should be thought of as a "quantum constraint" or a "clause", analogous to the notion of a k-local clause in classical constraint satisfaction problems.

There are $\binom{n}{k}$ ways to select k interacting qubits. However, the qubits interact only if they are "near each other". As a concrete example, consider the *Quantum Ising model*[3] shown as the chain in Fig. 1. The system's Hamiltonian H is given by

[3] named after the physicist Ernst Ising

$$H = -J \sum_{(i,j)} \sigma_Z^{(i)} \sigma_Z^{(j)} - \mu \sum_i \sigma_X^{(j)}, \tag{2}$$

where J is the coupling constant, and μ is the magnitude of an external field. The notation[4] (i, j) indicates that sites i and j are nearest neighbors (see Fig. 1.) $\sigma_Z^{(j)}$ refers the Pauli Z matrix acting on the j-th qubit only. The Hamiltonian H is a $2^n \times 2^n$ matrix, written here in a succinct form where $\sigma_Z^{(i)} \sigma_Z^{(j)}$ stands for the n-fold Kronceker product where each factor is $I_{2\times 2}$ except for the i-th and the j-th which are σ_Z. Similarly, $\sigma_X^{(j)}$ refers the Pauli X matrix acting on the j-th qubit, and $\sigma_X^{(j)}$ stands for the n-fold Kronceker product where each factor is $I_{2\times 2}$ except for the j-th which is σ_X.

[4] ♩ Be aware of notation!

Are local Hamiltonins better understood than the general Hamiltonians describing global interactions?

k-*Local Hamiltonian Problem (k-LH)*

Given a local Hamiltonian of an n-qubit system, can we determine its ground state energy E_o, namely, the minimum eigenvalue of H. We have seen that this problem is hard and described how it can be addressed by a VQE on NISQ.

> Given a k-local Hamiltonian $H = \sum_{i=1}^{m} H_i$ acting on n qubits and threshold parameters $a, b \in \mathbb{R}$ such that $b - a \geqslant 1/p(n)$ for some polynomial p, determine which of the following statements is true:
>
> $$E_o \leqslant a \quad \text{or} \quad E_o \geqslant b$$

It is believed that this problem is not likely to be efficiently solvable even by the quantum computer.

The Quantum Marginals Problem

Suppose we want to find the ground-state energy E_o of the Hamiltonian H. From the variational principle, we know that

$$E_o = \min_{|\psi\rangle} \langle\psi| H |\psi\rangle = \min_{\rho} \mathrm{Tr}(H\rho) \quad \text{where} \quad \rho = |\psi\rangle\langle\psi| \tag{3}$$

We described how we can attempt to solve this problem by an VQE on a NISQ device (Lecture # 18). We said above that this problem is hard to solve even when H is a local Hamiltonian. Are there are other implications of this hardness?

Because H is local, we have $H = \sum_{i=1}^{m} H_i$, and thus

$$E_o = \min_{\rho} \mathrm{Tr}(H\rho) = \min_{\{\rho_i\}} \sum_{i=1}^{m} \mathrm{Tr}(H_i \rho_i) \tag{4}$$

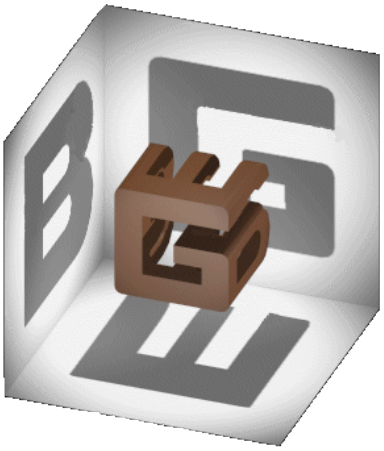
where ρ_i is the reduced density matrix[5] for at most k qubits that H_i is acting on. Therefore, in order to find the ground-state energy E_o, instead of having to find an n-qubit state ρ that minimizes (3), we can do minimization over the set $\{\rho_i\}$ instead. Is this problem simpler than the original?

The original problem in (3) has to find 2^n unknowns to specify the ground state $|\phi_o\rangle$, while the latter problem in (4) has to find unknowns whose number is polynomial in n. This is because the set $\{\rho_i\}$ consists of $2^k \times 2^k$ matrices where k is small and independent of n, and the set size m is polynomial in n.

However, the minimization problem in (4) is not unconstrained since there has to exist an n-qubit state $\rho = |\psi\rangle\langle\psi|$ such that the matrices $\{\rho_i\}$ can be obtained from it by the appropriate density matrix reduction determined by the non-trivial actions of H_i. Therefore, one has to first determine the condition such that the density matrices $\{\rho_i\}$ are indeed

[5] Reduced density matrix for the subsystem A of the bipartite system AB is $\rho_A = \mathrm{Tr}_B\, \rho_{AB}$.

reduced density matrices of ρ. This leads to the quantum marginals problem:



> Given a set of local density matrices $\{\rho_i\}$, determine whether there exists an n-qubit density matrix ρ such that $\{\rho_i\}$ are reduced density matrices of ρ.

The quantum marginal problem is as hard as the local Hamilonian problem.

Suppose we have a three qubit system ABC. Given density matrices ρ_{AB} and ρ_{AC}, we ask whether there is a three-qubit state ρ_{ABC} such that $\rho_{AB} = \mathrm{Tr}_C \rho_{ABC}$ and $\rho_{AC} = \mathrm{Tr}_B \rho_{ABC}$. We can answer this question only when $\rho_{AB} = \rho_{AC}$:
A two-qubit state ρ_{AB} has a symmetric extension if and only if

$$\mathrm{Tr}(\rho_B^2) \geqslant \mathrm{Tr}(\rho_{AB}^2) - 4\sqrt{\det(\rho_{AB})}$$

where $\rho_B = \mathrm{Tr}_A \rho_{AB}$.

Frustration-Free Hamiltonians

Consider again the following minimization problem:

$$E_o = \min_{|\psi\rangle} \langle\psi| H |\psi\rangle \quad \text{and} \quad |\varphi_o\rangle = \arg\min_{|\psi\rangle} \langle\psi| H |\psi\rangle$$

where $H = \sum_{i=1}^m H_j$. In general $|\varphi_o\rangle$ is the ground state of each H_i.[6]

A surprisingly diverse and important class of local Hamiltonians is defined by the additional property of being *frustration-free*, which is that the ground state $|\varphi_o\rangle$ of H is also a ground state of each H_i. In general, frustration-free conditions provide analytic control of ground state properties in otherwise mostly hard quantum problems.

When a local Hamiltonian $H = \sum_{i=1}^m H_i$ is frustration-free, then its ground-state energy $E_o = \sum_j E_{oj}$, where E_{oi} is the ground-state energy of H_i. Here E_{oi} are easy to find since each H_i acts non-trivially on only k particles. However, one first needs to determine whether $H = \sum_{i=1}^m H_i$ is frustration-free. In the theory of quantum computational complexity, this problem is formulated as follows:

> Given a k-local, n-qubit Hamiltonian $H = \sum_{i=1}^m H_i$, where each H_i has ground state energy $E_{oi} = 0$, and threshold $b \geqslant c/p(n)$ for some polynomial p and constant $c > 0$, determine which of the following statements is true:
>
> $$E_o \leqslant 0 \quad \text{or} \quad E_o \geqslant b$$

This problem is as hard as the local Hamilonian problem for $k \geqslant 3$.

[6] An H_i with different ground states is frustrated.

Consider the n-qubit Ising chain with the following Hamiltonian:

$$H = -\sum_{i=1}^{n} J_i \sigma_Z^{(i)} \sigma_Z^{(i+1)} \qquad J_i > 0.$$

1. Find the ground states of H.
2. Is H a frustration free Hamiltonian? Why?

Gapped Systems

When the system size $n \to \infty$, one important property of its Hamiltonian H is the gap Δ. Let H_n be the Hamiltonian of the size-n system. The system is called gapped if one of the following is true:

- As $n \to \infty$, the ground state degeneracy m_n of H_n is upper bounded by a finite integer m, and the gap Δ_n between the ground states and the first excited states of H_n is lower bounded by a finite positive number Δ.
- As $n \to \infty$, there are a finite number m of lowest energy states with energy separations among themselves that is exponentially small in n, and the energy separation of these lowest energy states to all the other states is lower bounded by a finite number Δ for arbitrary n.

Quantum Computing Systems [1]

[1] Rutgers, ECE 579, Spring 2021

Prof. Emina Soljanin

Lecture #20, April 20

This lecture describes the Quantum Approximate Optimization Algorithm (QAOA).

Quantum Approximate Optimization Algorithm (QAOA)

Algorithms that are appropriate for Noisy Intermediate-Scale Quantum (NISQ) systems

- should not need extensive error correction (can tolerate noise), and
- should not need very large numbers of qubits (intermediate scale),
- but should exhibit quantum speedup and solve useful problems.

The QAOA is a low-depth, and thus would not need too much coherence. Some results indicate that it is fairly robust to errors. When it was proposed (by Farhi et al. in 2014), it worked better than the best known classical algorithm, but just for a few months.

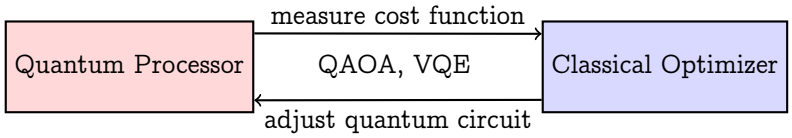


Figure 1: NISQ Quantum/Classical hybrid architecture

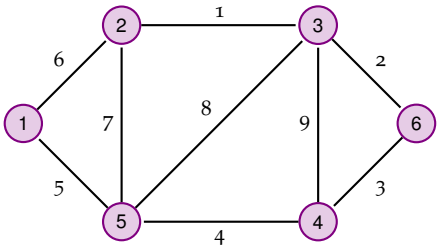
The goal of QAOA is to optimize a real-valued function f acting on n-bit strings, $f : \{0,1\}^n \to \mathbb{R}$, of the following form:

$$f(\mathbf{b}) = \sum_{j=1}^{m} f_j(\mathbf{b}),$$

where $f_j(\mathbf{b})$ depends on only a few of the n bits. QAOA seems a promising technique for optimization problems that have only local constraints. More general optimization problems may need to deal with global constraints. (For example, the traveling salesman problem.) In principle any optimization problem with global constraints can be turned into one with purely local constraints but that may add a big overhead. We will study QAOA on the max-cut problem in graph theory.

CS Interlude – The Max-Cut Problem

A *cut* of a graph is a partition of its vertices into two subsets. A cut is defined by a cut-set, that is, the set of edges that have one endpoint in each subset of the partition. The *max-cut* problem seeks to find a cut with the largest cut-set. (For a weighted graph, the max cut problem seeks to maximize the sum of weights of the edges in the cut.)



Consider a graph with n vertices labeled $1, \ldots, n$. If nodes j and k are connected, we label the edge between them as (j, k). Edge (j, k) is in a cut if the nodes j and k are in two different subsets of the partition. Suppose we associate with node j, $j = 1, \ldots, n$, a variable $z_j \in \{-1, 1\}$. Note that the set of edges for which $z_j z_k = -1$ is a cut-set between the vertices that are assigned 1 and the vertices that are assigned -1. Therefore, the following maximization gives the maximum cut.

$$\max_{z_i \in \{-1,1\}} f(z_1, \ldots, z_n) = \max_{z_i \in \{-1,1\}} \sum_{(j,k)} (1 - z_j z_k)$$

Here, the sum is over all edges in the graph, and each clause $(1 - z_j z_k)$ contributes a non-zero term to the cost iff $z_j z_k = -1$, that is, the corresponding edge is in the cut.

We will later need the notion of distance and diameter in graphs. The *distance* between two vertices in a graph is the number of edges in a shortest or minimal path between them in the graph. The *diameter* of a graph is the maximum distance between any pair of vertices.

A QAOA for the Max-Cut Problem

Consider now an n qubit state, where each qubit is associated with a vertex of the graph, and the following Hamiltonian:

$$H_C = \sum_{(j,k)} \left(I - \sigma_Z^{(j)} \sigma_Z^{(k)}\right)$$

where $I_{2^n \times 2^n}$ is the identity matrix and $\sigma_Z^{(j)}$ refers the Pauli Z matrix acting on the j-th qubit. The Hamiltonian H_C is a $2^n \times 2^n$ matrix, written here in a succinct form where $\sigma_Z^{(j)} \sigma_Z^{(k)}$ denotes the n-fold Kronceker product where each factor is $I_{2\times 2}$ except for the j-th and the k-th which are σ_Z.

We refer to H_C as the cost Hamiltonian. Note that H_C is a diagonal the computational basis. If qubit j is measured in the σ_Z basis, we can use the result of the measurement (an eigenvalue of σ_Z) to decide to which subset of the partition to put vertex j into. Thus measuring all qubits defines a cut.

For the cost Hamiltonian H_C, we define the unitary operator[2]

[2] depends on the objective function

$$U_{C,\gamma} = \exp(-i\gamma H_C) = \prod_{(j,k)} \exp\left[-i\gamma\left(I - \sigma_Z^{(j)}\sigma_Z^{(k)}\right)\right]$$

Here the product is over all edges in the graph. We next define operator H_B, known as a mixer[3] Hamiltonian, and its associated unitary matrix as follows:

[3] depends on the domain

$$H_B = \sum_{j=1}^{n} \sigma_X^{(j)} \quad \text{and} \quad U_{B,\beta} = \exp(-i\beta H_B) = \prod_{j=1}^{n} \exp\left(-i\beta\sigma_Z^{(j)}\right)$$

where $\sigma_X^{(j)}$ denotes the Pauli X matrix acting on the j-th qubit.

The idea for Max-Cut QAOA goes as follows: We assign a qubit for each vertex of our graph, and prepare the n qubit state as a uniform superposition of all basis states. Note that each basis state describes a possible partition of the graph. We then apply a sequence of unitary operators to the set of qubits, and measure the resulting state by the σ_Z observable. Each qubit yields a value of 1 or -1. We use these values to define a partition to the graph. We try to find a sequence of unitary operators that maximizes the expected value of our cut.

Algorithm:

1. Prepare the n qubit state $|s\rangle$ as a uniform superposition of of all basis states:
$$|s\rangle = \frac{1}{\sqrt{2^n}} \sum_{\mathbf{b}\in\{0,1\}^n} |\mathbf{b}\rangle$$

2. For a graph with diameter p, repeat the following step p times with different parameters γ and β:

   Transform (evolve) the current state by $U_C(\gamma)$ followed by $U_B(\beta)$.

   The end state is
$$|\psi(\boldsymbol{\gamma},\boldsymbol{\beta})\rangle = \prod_{j=1}^{p} U_B(\beta_i)U_C(\gamma_i)\,|s\rangle$$
   where $(\boldsymbol{\gamma},\boldsymbol{\beta}) = (\gamma_1,\beta_1,\gamma_2,\beta_2,\ldots,\gamma_p,\beta_p)$.

3. Compute the expectation
$$F_p(\boldsymbol{\gamma},\boldsymbol{\beta}) = \langle\psi(\boldsymbol{\gamma},\boldsymbol{\beta})|\,H_C\,|\psi(\boldsymbol{\gamma},\boldsymbol{\beta})\rangle$$
   by repeating the above state preparation, evolution, and measuring. (Each measurement produces a sample value.)

4. Use a (classical) optimization algorithm to (approximately) find the vector $(\boldsymbol{\gamma},\boldsymbol{\beta})$ that optimizes $F_p(\boldsymbol{\gamma},\boldsymbol{\beta})$.[4]

[4] A key to success for the algorithm is having good values for these parameters.

Possibly run the entire procedure again, and return the best problem solution.

Quantum Computing Systems [1]

Prof. Emina Soljanin

Lecture #21, April 22

[1] Rutgers, ECE 579, Spring 2021

This lecture discusses classical and quantum transmission of classical information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently, the messages have meaning; ...[2]

[2] A quote by Claude E. Shannon, who is considered to be the father of *Information Theory.*

Binary-Input Communications Channel

The transmitter Alice has a sequence of bits to send to the receiver Bob. She represents her 0s and 1s into two distinguishable physical quantities that can be physically carried by the communications channel.

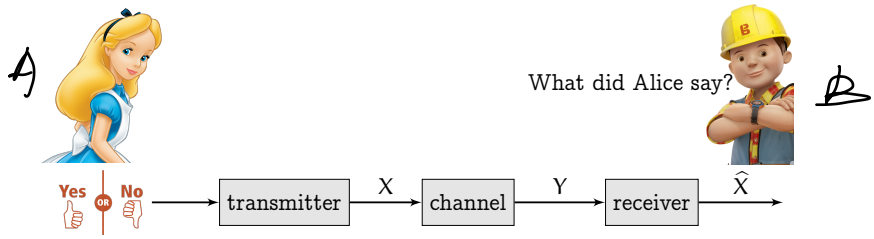


Figure 1: A classical communications system with binary input.

Mathematically, a communications channel is a probabilistic device. It involves at least 2 random variables:

- X – the channel input; its range $\mathcal{X}$ is called the input alphabet[3]
- Y – the channel output; its range $\mathcal{Y}$ is called the output alphabet

[3] Recall that in our probability class, we used Ω_X to denote the range of random variable X.

Channel output RV Y is a noisy version of its input RV X. The relation between the input and the output is described by, e.g.,

- the conditional probability of the output given the input $W(Y \mid X)$. We call W the channel transition probability.
- a noise random variable Z added to the input s.t. $Y = X + Z$.

A discrete memoryless channel (DMC)[4] 1) has discrete RVs at its inputs and outputs with finite alphabets, and 2) the output symbol at time i depends only on the input symbol at time i and not on any of the previous input symbols *(memoryless)*, that is,

[4] We will only consider DMCs.

$$P(Y_1 = y_1, \ldots, Y_n = y_n \mid X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(Y_i = y_i \mid X_i = x_i)$$

Based on the received y, Bob makes an estimate $\hat{x}$ of the true x.

Examples of Classical Channels

The Binary Symmetric Channel *BSC(p)*

Binary input and output: $\mathcal{X} = \mathcal{Y} = \{0, 1\}$

$$W(0 \mid 0) = W(1 \mid 1) = 1 - p$$
$$W(1 \mid 0) = W(0 \mid 1) = p$$

We can instead say

$$Y = X + Z \mod 2 \quad \text{where } Z \sim \text{Bernoulli}(p).$$

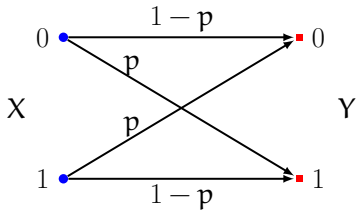


Figure 2: Binary symmetric channel BSC(p).

The Z *Channel*

Binary input and output: $\mathcal{X} = \mathcal{Y} = \{0, 1\}$

$$W(0 \mid 0) = 1 \qquad W(1 \mid 0) = 0$$
$$W(0 \mid 1) = p \qquad W(1 \mid 1) = 1 - p$$

This channel models the optical ON/OFF channel, since even when light is ON, we can have no photons reaching the detector.

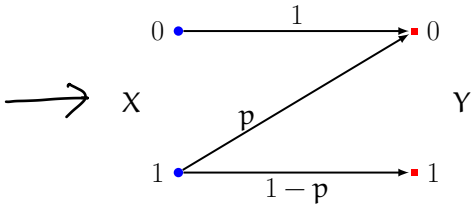


Figure 3: Z channel.

Binary Erasure Channel *BEC(ϵ)*

Binary input and ternary output: $\mathcal{X} = \{0, 1\}$, $\mathcal{Y} = \{0, 1, -\}$

$$W(1 \mid 0) = W(0 \mid 1) = 0$$
$$W(0 \mid 0) = W(1 \mid 1) = 1 - \epsilon$$
$$W(- \mid 0) = W(- \mid 1) = \epsilon$$

Is this channel additive?

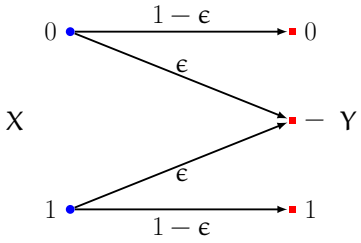


Figure 4: Binary erasure channel BEC(ϵ).

Binary Input Additive Gaussian Noise Channel

-1 or 1 input and real-valued output: $\mathcal{X} = \{-1, 1\}$, $\mathcal{Y} = \mathbb{R}$

$$W(y \mid -1) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x+1)^2}{2\sigma^2}}$$
$$W(y \mid +1) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-1)^2}{2\sigma^2}}$$

We can instead say $Y = X + Z$ where $Z \sim N(0, \sigma^2)$.

If you know the PMF of X and the channel, can you find the PMF (PDF) the output?

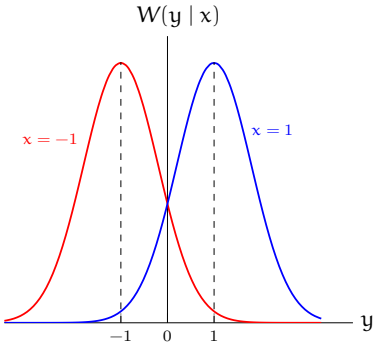


Figure 5: Binary input additive white Gaussian noise channel AWGN(σ).

Quantum Measurement as a Communications Channel

Assume an ensemble $\{|\psi_i\rangle, p_i\}$ of pure states is measured by a quantum measurement. We can consider this process as a communications channel which has as its possible inputs vectors $|\psi_i\rangle$, $i \in \mathcal{X}$. The channel outputs (vectors $|\varphi_j\rangle$) and the transition probabilities are determined by the chosen measurement.

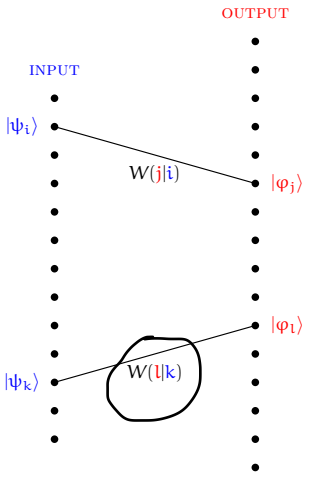
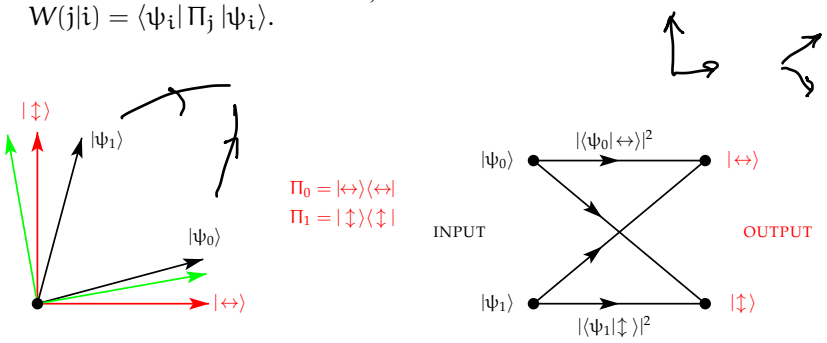


Figure 6: A two-qubit ensamble with a von Neumann measurement.

van Neumann Measurement

- A set of pairwise orthogonal projection operators $\{\Pi_j\}$ that form a complete resolution of the identity: $\sum_j \Pi_j = I$.
- For input $|\psi_i\rangle$, the output $\frac{1}{\|\Pi_j|\psi_i\rangle\|}\Pi_j|\psi_j\rangle$ happens with probability $W(j|i) = \langle\psi_i|\Pi_j|\psi_i\rangle$.



Positive Operator-Valued Measure (POVM)

- Any set of positive-semidefinite operators $\{E_j\}$ that form a complete resolution of the identity: $\sum_j E_j = I$.
- For input $|\psi_i\rangle$, the output $\frac{1}{\|E_j|\psi_i\rangle\|}E_j|\psi_i\rangle$ happens with probability $W(j|i) = \langle\psi_i|E_j|\psi_i\rangle$.

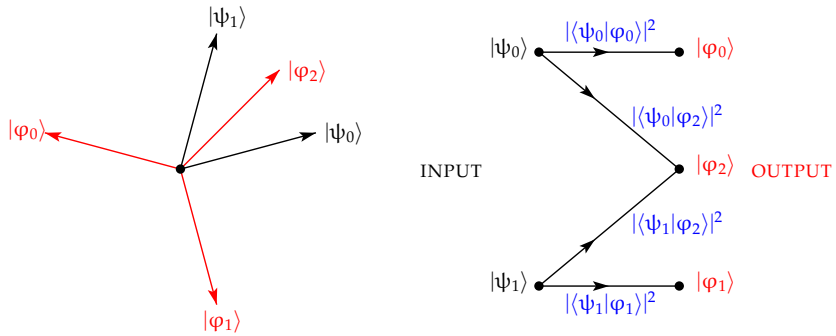


Figure 7: A 3 element POVM that gives a binary erasure channel because $\langle\varphi_0|\psi_1\rangle = 0$ and $\langle\varphi_1|\psi_0\rangle = 0$.

We would like to select a measurement to e.g., minimize error rate.

Channel Coding

We protect messages from errors/erasures by transmitting redundant information, namely, error/erasure correcting coding.

BSC(p) with Repetition Coding

Encoder:

message	transmit
0	000
1	111

Random Variables in a Repetition Code on the BSC

M - RV associated with the message, $\Omega_M = \{0, 1\}$.
We assume that $P(0) = P(1) = 1/2$, i.e., equal priors.

$\bar{X}$ - at the channel input, $\bar{\mathcal{X}} = \{000, 111\}$.

$\bar{Y}$ - at the channel output, $\bar{\mathcal{Y}} = \{000, 001, 010, 100, 111, 110, 101, 011\}$.

Decoder:

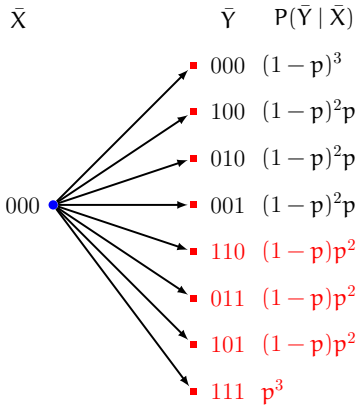
Q: What should the optimal receiver do when we have equal priors?

A: On a DMC, the receiver should implement majority logic.

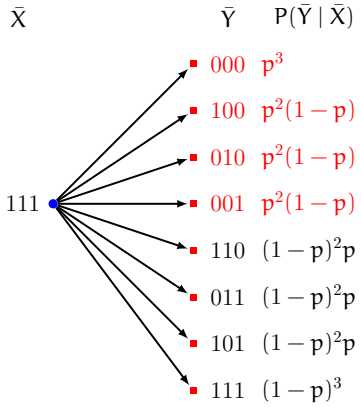
Q: What if we did not have a memoryless channel and equal priors.

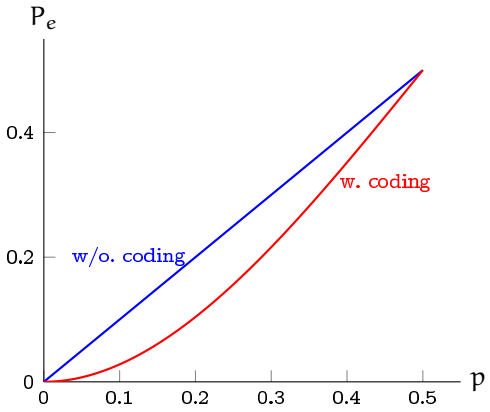
What is the error rate P_e?

Transitions from 000 input.



Transitions from the 111 input.





- Without coding, the probability of error is $P_e = p$.
- With coding, the probability of error is $P_e = p^3 + 3p^2(1-p)$ (corresponding to 3 or 2 flips).

BEC(ϵ) with Parity-Check Coding

Encoder:

message	transmit
00	000
01	011
10	101
11	110

Decoder:
The decoder is able to reconstruct the message as long as none or one codeword symbols are erased. What is the erasure rate P_x?

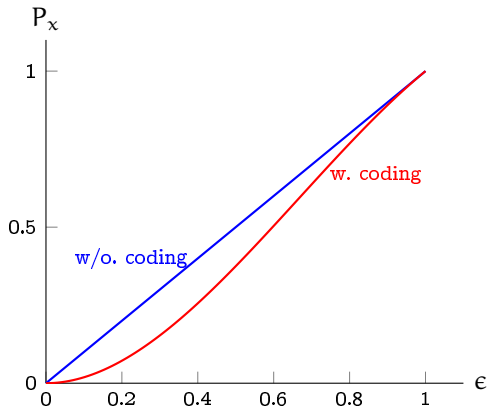


Figure 8: Probability of erasure with a rate 2/3 parity chcek code.

- Without coding, the bit-erasure probability ϵ.
- With coding, the bit erasure probability is $P_x = \epsilon\left[\epsilon^2 + 2\epsilon(1-\epsilon)\right]$.

What is the message-erasure probability?

Quantum Computing Systems [1]

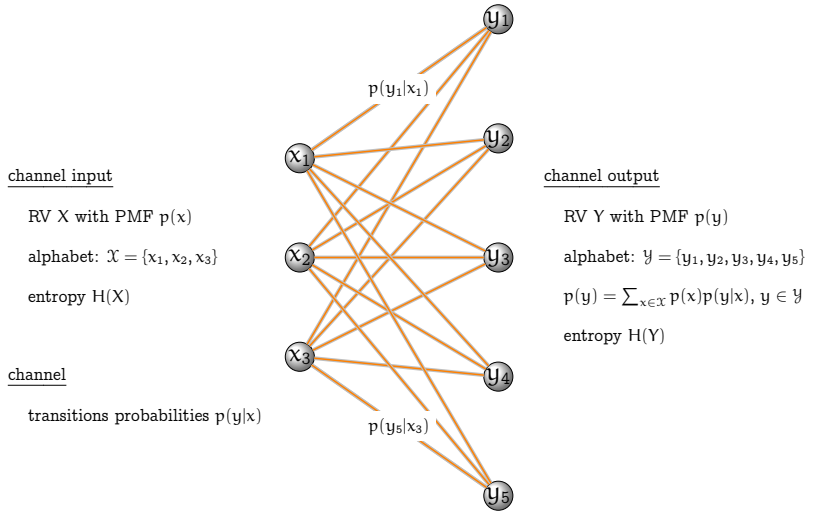
Prof. Emina Soljanin

Lecture #22, April 27

[1] Rutgers, ECE 579, Fall 2020

This lecture discusses 1) achievable information rates over classical communications channels, 2) the accessible information of a quantum ensemble and 3) the Holevo bound.

Communications Channel - Example



Classical Information Measures

Shannon Entropy of RV X

Let X be a discrete random variable with alphabet (range) $\mathcal{X}$ and PMF P_X. The Shannon entropy $H(X)$ (in bits) is defined as follows:

$$H(X) = -\sum_{x \in \mathcal{X}} P_X(x) \log_2 P_X(x)$$

$H(X)$ tells us how many typical sequences the source can generate.

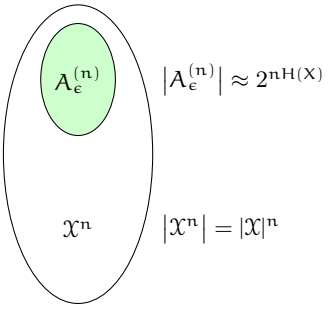


Figure 1: The high probability subset in $\mathcal{X}^n$ in contains $2^{nH(X)}$ equally likely sequences.

Conditional Entropy

Consider two discrete random variables: X with alphabet $\mathcal{X}$ and Y with alphabet $\mathcal{Y}$ (corresponding to the channel inputs and outputs). Consider RV Y conditioned on RV X taking a certain value. Its PMF is

$p(y|x)$, $y \in \mathcal{Y}$, and thus its entropy $H(Y|X = x)$ is given by

$$H(Y|X = x) = -\sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x).$$

The conditional entropy $H(Y|X)$ of Y given X is defined as the expected value of $H(Y|X = x)$ over X:

$$H(Y|X) = \sum_{x \in \mathcal{X}} p(x)\, H(Y|X = x) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)}.$$

$H(Y|X)$ tells us how many typical sequences at the channel output correspond to a typical channel input sequence.

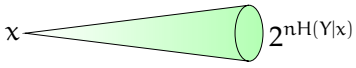


Figure 2: $x \in \mathcal{X}^n$ at the channel input gives raise to $2^{nH(Y|x)}$ typical sequences at the channel output.

Observe the following:

- $H(Y|X) = H(Y)$ iff Y and X are independent.
- $H(Y|X) \leqslant H(Y)$ – conditioning reduces entropy for dependent RVs.
- $H(Y|X) = 0$ iff the value of Y is completely determined by the value of X, as in the noiseless channel.

Mutual Information of RVs X and Y

We can now address the following questions: How many messages can be selected at the channel input that is described by RV X? How large is a subset of these messages s.t. a nosy version of any message in this subset is likely to be distinguishable from a noisy version of any other message in the subset?

Note from the illustration and explanation in Fig. 3, that we cannot have more than $2^{nH(Y)}/2^{nH(Y|X)}$ input sequences and hope that their noisy versions are different. The classical channel coding theorem shows that approximately $2^{nH(Y)}/2^{nH(Y|X)} = 2^{n[H(Y)-H(Y|X)]}$ indeed can be found s.t. the sets of their likely noisy versions have diminishing intersections as n increases.[2]

The quantity $H(Y) - H(Y|X)$ is known as the *mutual information* between random variables X and Y, and is defined as follows:

$$I(X;Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p_{(X,Y)}(x, y) \log \frac{p_{(X,Y)}(x, y)}{p_X(x)\, p_Y(y)} \tag{1}$$

We have

$$I(X;Y) = H(Y) - H(Y|X) = H(X) - H(X|Y) \tag{2}$$

We here provided only some general reasoning. For proofs, please see e.g. *Elements of Information Theory* by Cover and Thomas.

[2] Therefore, the achievable information rate, that is, the average number of bits per channel use, is equal to $I(X;Y)$.

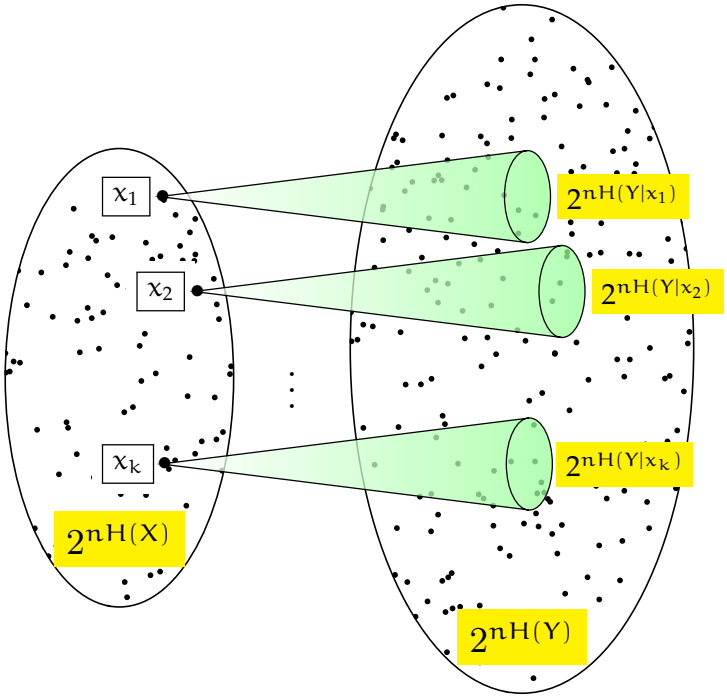


Figure 3: For each of the $2^{nH(X)}$ typical input sequences, there are approximately $2^{nH(Y|X)}$ conditionally typical Y sequences, all of them equally likely. In order to be able to decide what was the input, we have to ensure that no two input sequences produce the same output sequence.

Accessible Information

Suppose Alice has a classical random variable X with alphabet $\mathcal{X}$ with $|\mathcal{X}|$ letters and letter probabilities $\{p_1, p_2, \ldots, p_{|\mathcal{X}|}\}$. When X assumes letter $a \in \mathcal{X}$, Alice prepares quantum state with density matrix ρ_a and gives this state to Bob, whose goal is to find the value a of X that Alice has. In order to achieve his goal Bob performs a measurement on the received state obtaining a classical outcome, namely, a random variable which we denote with Y.

As we discussed last time, the process of measurement is mathematically equivalent to classical information transmission through a classical channel determined by the selected measurements.[3] Therefore, the amount of information that Bob can get about the variable X through the associated quantum ensemble $\mathcal{E} = \{\rho_a, p_a\}_{a\in\mathcal{X}}$ is the maximum value of the mutual information $I(X;Y)$ between the random variables X and Y over all the possible measurements that Bob can make on $\mathcal{E}$. We refer to this information as the accessible information Acc of $\mathcal{E}$:

$$\mathrm{Acc}(\mathcal{E}) = \max_{\text{measurements}} I(X;Y)$$

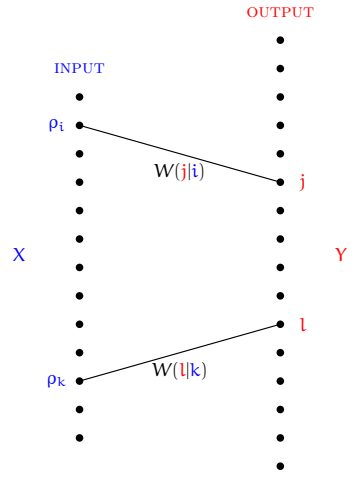


Figure 4: For input $|\psi_i\rangle$, the output $\frac{1}{\|E_j|\psi_i\rangle\|}E_j|\psi_i\rangle$ happens with probability $W(j|i) = \langle\psi_i|E_j|\psi_i\rangle$

[3] The channel output and the transition probabilities are determined by the selected measurement.

Example

Consider an ensemble of pure states $|\psi_0\rangle$ and $|\psi_1\rangle$, as shown in Figs. 5 and 6. A von Neumann measurement will give us a binary-output channel, as in Fig. 5. A POVM measurement will give us a channel

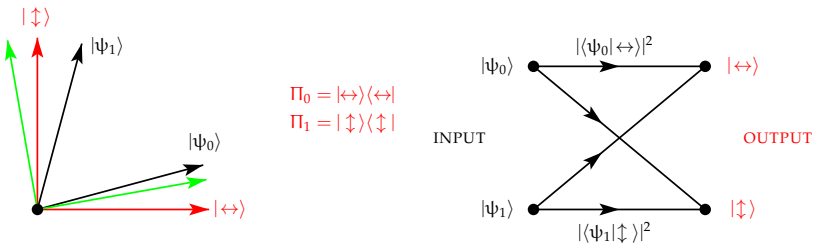


Figure 5: A two-qubit ensamble with a von Neumann measurement.

with more outputs. Note that the transition probabilities also depend on the applied measurement. How do we choose a measurement to

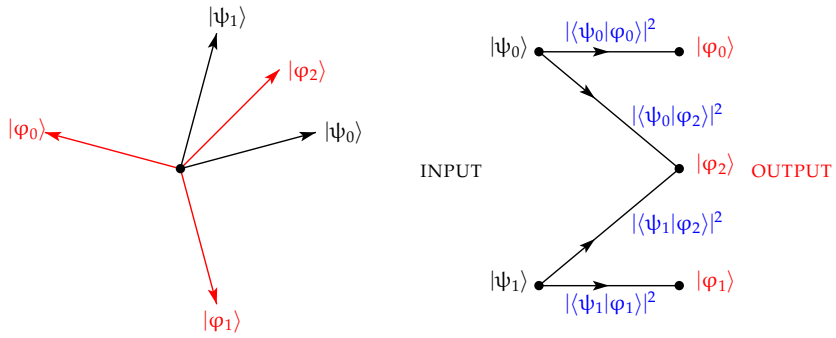


Figure 6: A two-qubit ensemble with a POVM measurement.

maximize the accessible information?

The general formula formula to compute the accessible information for an ensemble $\mathcal{E} = \{\rho_a, p_a\}_{a\in\mathcal{X}}$ is not known. The best known upper bound on accessible information is the Holevo bound:

$$\mathrm{Acc}(\mathcal{E}) \leqslant S(\rho) - \sum_{a\in\mathcal{X}} p_a S(\rho_a) \tag{3}$$

where $\rho = \sum_{a\in\mathcal{X}} p_a \rho_a$ is the ensemble density matrix. The quantity

$$\chi = S(\rho) - \sum_{a\in\mathcal{X}} p_a S(\rho_a)$$

is called the Holevo information or the Holevo χ quantity.

In order prove the Holevo bound (3), we devise a tripartite quantum system we refer to ABM. The subsystem A[4] corresponds to the Alice's RV X and is described by the ensemble $\{|a\rangle, p_a\}_{a\in\mathcal{X}}$ where $|a\rangle, a \in \mathcal{X}$ are orthogonal states. The subsystem B corresponds to the quantum

[4] channel input RV X

state prepared by Alice and given to Bob and is described by the ensemble $\mathcal{E} = \{\rho_a, p_a\}_{a\in\mathcal{X}}$. The subsystem M[5] is where Bob imprints his measurement result.

[5] channel output RV Y

The joint system ABM is prior to the measurement in the state

$$\rho_{ABM} = \sum_{a\in\mathcal{X}} p_a |a\rangle\langle a| \otimes \rho_a \otimes |o\rangle\langle o|$$

The state of the subsystem M before the measurement is some known state of the register, here $|o\rangle\langle o|$. If the B subsystem is in state ρ_a (for some $a \in \mathcal{X}$, and Bob performers the von Neumann measurement $\{\Pi_m\}$, then Bob is left with the state[6]

$$\rho_a \otimes |o\rangle\langle o| \longrightarrow \sum_m \Pi_m \rho_a \Pi_m \otimes |m\rangle\langle m|$$

[6] Recall that measuring a system in state σ leaves the system in state $\sum_m \Pi_m \rho_a \Pi_m$.

Therefore, the tripartite system state ρ_{ABM} is mapped into ρ'_{ABM} as follows:

$$\rho_{ABM} \longrightarrow \rho'_{ABM} = \sum_{a\in\mathcal{X}} p_a |a\rangle\langle a| \otimes \sum_m \Pi_m \rho_a \Pi_m \otimes |m\rangle\langle m|$$

By the strong subadditibvity of the quantum entropy,[7] we have

$$S(\rho'_{ABM}) + S(\rho'_M) \leqslant S(\rho'_{AM}) + S(\rho'_{BM}), \tag{4}$$

[7] Both the Shannon and van Neumann entropy satisfy the subadditivity property.

which translates into the Holevo bound (see the problems below).

Problems

1. Show that the following identities hold:
   (a) $S(\rho'_{ABM}) = H(X) + \sum_{a\in\mathcal{X}} p_a S(\rho_a)$
   (b) $S(\rho'_M) = H(Y)$
   (c) $S(\rho'_{AM}) = H(X, Y)$
   (d) $S(\rho'_{BM}) = S(\rho)$
2. Show that the Holevo bound follows from the inequality (4) and the claims in Problem 1.