CSCI3350 Introduction to Quantum Computing (2026 Spring)

Xiao Liang

https://xiao-liang.github.io

Department of Computer Science and Engineering
The Chinese University of Hong Kong

(Updated on February 3, 2026)

Disclaimer: These lecture notes originate from the Spring 2026 offering of the course. They have been lightly edited for accuracy and completeness, and may still contain typographical or other errors.

Contents

Chapter 0

Physics and Computing: from Newton to Shrödinger

Nothing from Chapter 0 will be tested on quizzes or exams!

Our computational power is bounded—and enabled—by our understanding of physics. Different physical laws suggest different ways to represent and process information. As we move from gears and pulleys to transistors and integrated circuits and, finally, to quantum systems, we get new primitives for encoding and manipulating information. In this chapter, we will move conceptually:

- from classical mechanics (Newton) and mechanical computation,
- to electromagnetism (Maxwell) and electronic computation,
- to quantum mechanics (Schrödinger) and quantum computation.

0.1 Era I: Classical Mechanics and Mechanical Computing

Classical mechanics is governed by Newton's laws, which describe how forces influence motion. The foundational equation is Newton's second law:

$$\mathbf{F} = m\,\mathbf{a} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}. \tag{1}$$

Here:

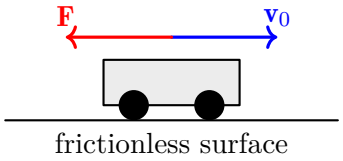
- $\mathbf{F}$ is the net force vector acting on an object (magnitude and direction).
- m is the mass (a scalar), measuring how strongly the object resists changes in motion.
- $\mathbf{a} = \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}$ is the acceleration vector, the time derivative of velocity.
- $\mathbf{p} = m\,\mathbf{v}$ is the momentum vector (for constant mass m and velocity $\mathbf{v}$).
- $\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}$ is the time derivative of momentum, equal to the net force.

Given forces, masses, and initial conditions, we can predict trajectories. The exemplary systems (see Figure 1) that we have solved in secondary school speaks for this idea.

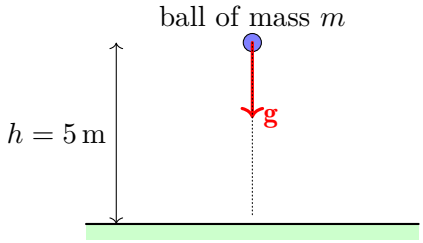
This predictive power underpins early computational devices built on mechanical principles: if we can engineer interacting motions (gears, levers, pulleys), we can compute. The lesson is simple: with precise control over mechanical components, we can implement arithmetic and logic through motion and force.

We now present two examples of mechanical computing devices: abacus and gear calculators.

Mechanical Computing Devices: Abacus



(a) Cart on a frictionless surface



(b) Ball dropped from height

Figure 1: Exemplary Systems Solvable by Classical Mechanics

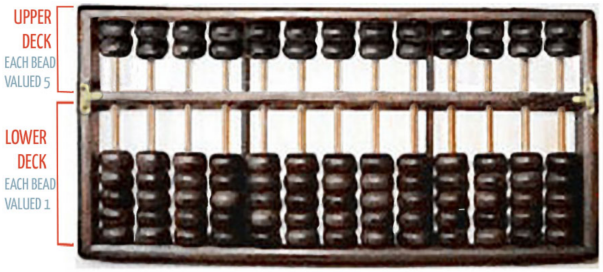
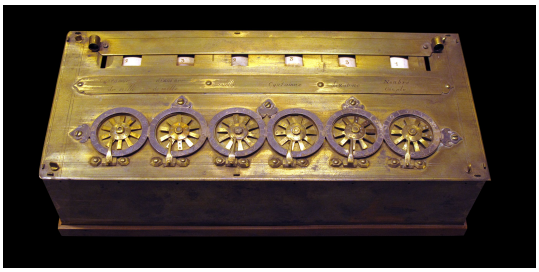


Figure 2: Abacus

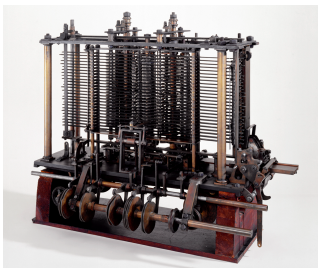
An abacus represents numbers using beads on rods and implements arithmetic by moving beads according to rules. It is simple yet powerful for certain calculations.

Mechanical Computing Devices: Gear Calculators

Gear-based calculators (e.g., Pascal's Pascaline; Babbage's designs) encode numbers as gear positions; addition and subtraction follow from rotations and carries propagating through gear trains.



(a) Pascaline



(b) Babbage's analytical engine (concept)

Mechanical computation excels when problems can be mapped to a small number of coupled motions, but it struggles with tasks requiring deep memory, branching logic, or billions of fast operations.

0.2 Era II: Electromagnetism and Electronic Computing

A second revolution comes with Maxwell's unification of electricity and magnetism. The governing equations are:

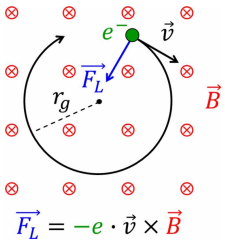
$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0} \qquad \text{(Gauss's law)}$$

$$\nabla \cdot \mathbf{B} = 0 \qquad \text{(Gauss's law for magnetism)}$$

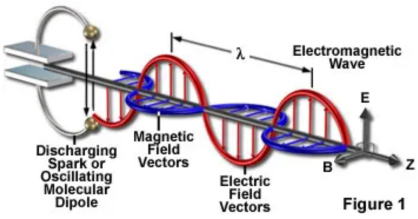
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{(Faraday's law)}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \qquad \text{(Ampère–Maxwell law),}$$

where $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic fields, ρ is charge density, $\mathbf{J}$ is current density, ε_0 vacuum permittivity, and μ_0 vacuum permeability. These laws predict electromagnetic waves and enable circuitry, antennas, and information transmission (see Figure 4).



(a) An electron in a magnetic field



(b) Propagation of an electromagnetic wave

Figure 4: Illustrative Systems in Electromagnetism

If we can harness fields and waves (wires, antennas, circuits), we can compute with electrons and light.

Electronic Computing Devices

Modern devices—from calculators and desktops to smartphones and cloud servers—encode information electronically.

Two common and crucial tasks for all these devices are *encoding* and *computing*:

- **Encoding**
  - Bits as high/low voltage levels.
  - Wires carry bits; memory stores them as charge states or magnetic domains.
- **Computing**
  - Logic gates (AND, OR, NOT, NAND, etc.) built from transistors implement Boolean functions.
  - Synchronous circuits orchestrated by clocks compose gates into processors.
  - Algorithms are sequences of logic operations on registers of bits.

0.2.1 Mechanical vs. Electronic Computing

Mechanical calculators are well-suited for a few coupled motions (gears, cams, analog integrators). They struggle when a task requires:

- exploring enormous combinatorial possibilities,
- processing very large datasets with high precision,
- running millions to billions of steps quickly,
- flexible memory and branching logic.

Electronics handle these via fast arithmetic, hierarchical memory, and parallelism. Examples include:

- high-dimensional PDEs and large-scale simulation,
- large-scale linear algebra and optimization,
- symbolic computation and automated reasoning,
- data-intensive analytics.

0.3 Era III: Quantum Mechanics and Quantum Computing

At microscopic scales, classical laws fail to explain experiments. Quantum mechanics provides new rules for evolution and measurement, altering how we encode and process information.

Concrete examples for microscopic-scale systems include:

- Atoms and molecules ($\sim$ 0.1–1 nm): explains spectral lines, chemical bonds.
- Tunneling through thin barriers (few nm): exploited in flash memory and scanning tunneling microscopes.
- Quantum dots (1–10 nm): nanocrystals with size-tunable colors used in displays.

0.3.1 Schrödinger's Equation

Quantum systems are governed by the Schrödinger equation:[1]

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle \;=\; \hat{H}\, |\psi(t)\rangle \,. \tag{2}$$

Here:

- $i := \sqrt{-1}$ is the imaginary unit. (Complex numbers are essential in quantum theory. We will return to this topic in section 1.5.).
- $\hbar$ is Planck's reduced constant, $\hbar \approx 1.055 \times 10^{-34}\,\mathrm{J \cdot s}$.[2]

[1] What we present in Equation (2) is the simplest form of the (time-dependent) Schrödinger equation. There exist more advanced formulations that capture a broader class of quantum systems.

[2] Named after Max Planck, Nobel Prize in Physics (1918).

- $|\psi(t)\rangle$ is the state vector of the system at time t.
- $\hat{H}$ is the so-called *Hamiltonian operator*. We will discuss it in more details soon.

We do not need a fully rigorous understanding of the Schrödinger equation—this is not a quantum mechanics course. But it helps to have an intuitive feel for it. We will build that intuition by drawing an analogy with classical mechanics.

0.3.2 An Inaccurate-yet-Helpful Analogy

Consider the cart on a frictionless surface with initial velocity $\mathbf{v}_0$ and an applied force $\mathbf{F}$ (shown in Figure 5). In classical mechanics, knowing $\mathbf{F} = m\mathbf{a}$ and initial conditions lets us predict where the cart will be at time t.

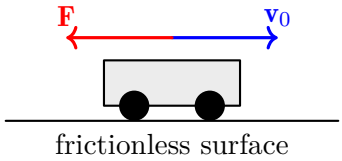


Figure 5: Cart on a frictionless surface

By analogy, the Schrödinger equation plays the role of the "law of motion" in quantum mechanics:

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle \;=\; \hat{H}\, |\psi(t)\rangle\,.$$

Think of:

- $|\psi(t)\rangle$ as a *vector* describing the system's state at time t;
- $\hat{H}$ (the Hamiltonian) as a *rule-book* or *matrix* encoding how the system behaves (energies, forces, boundaries);

With the above understanding, the equation says: *Change in the system's state over time is determined by the rule-book acting on the state!*

In more details:

- $|\psi(t)\rangle$: Think it as "where/what the system is like right now." It's like a vector you update over time.
- $\hat{H}$: Think it as "all the ingredients of the system"—mass, potential energy, constraints, fields, etc. It acts like a matrix that "pushes" the system's state forward in time.
- $i\hbar \dfrac{\partial}{\partial t}$: It reads as "how fast the state changes with time," scaled by a constant i times $\hbar$.

In Summary: The following is all you need to know about Shrödinger's equation:

- Pretend the quantum state is like a vector that tells us where the system could be. The Hamiltonian is a big matrix that encodes the rules of the world. The Schrödinger equation says the matrix pushes the vector forward in time. If we know the rule and where we started, we can predict what we'll see later.

Some caveats

The previous analogy is not accurate. For example,

- The "vector" $|\psi\rangle$ is not a single, deterministic status of the system; it actually encodes *probabilities of many possible status*.
- $\hat{H}$ is *total energy operator* (kinetic + potential).

Indeed, there is an alternative formulation of classical mechanics, equivalent to Newton's, called *Hamiltonian mechanics*. The Schrödinger equation often feels more natural to those familiar with this formalism. If you are interested, you can easily find online resources on the topic, e.g., David Tong's course.

But again, our CSCI3350 is not a quantum mechanics course. We do not need a fully rigorous interpretation of the equation. Instead, we will focus on its major implications for how we perform computation (i.e., *the Four Postulates* that we will discuss later).

We may return to the interpretation of Shrödinger's equation when needed—for example, when discussing Hamiltonian complexity or applications of quantum computing to quantum simulation and quantum chemistry, where the meaning of the Hamiltonian $\hat{H}$ becomes more relevant.

0.4 Our Focus: Quantum Computing

With the quantum background sketched above, our goal is to understand how Schrödinger evolution enables us to encode information and perform computation.

We will first utilize a few lectures to build up the basic knowledge.

Once we have the basics, we will explore the following aspects of quantum computing:

- **Classical vs. quantum power:** Are there problems quantum computers can solve that classical computers cannot? What features are uniquely quantum?
- **Limits:** Can quantum computers solve all problems we care about? Which problems remain hard even for quantum computers?
- **Applications and interfaces:** Quantum machine learning, quantum networks, quantum cryptography, quantum chemistry, and more.

Course Style and Scope

Theory-first, no programming: The course style is similar to CSCI3160 Design and Analysis of Algorithms, focusing on the algorithmic ideas, pseudocode, and theoretical analysis. It does not involve programming, real-world implementation, or software engineering considerations.

Beyond algorithms: We also explore the fundamental principles of quantum information — including concepts like superposition, entanglement, and measurement. The course may also cover several important topics that are not algorithmic in nature but are central to the field of quantum computing, such as quantum error correction, fault-tolerant computation, proofs of quantumness, and non-local games.

What This Course Is Not: This course is *not* intended for students who:

- want to learn how to build a quantum computer,
- want hands-on practice in a specific quantum programming language,
- expect a full quantum mechanics/physics course.

Note that this arrangement is not unique to our course of CSCI3350. It is a common practice for quantum computing courses offered by most CS departments. You can explore similar courses at:

- COMS4281 at Columbia University
- CS498QC at UIUC
- 15-859BB at CMU
- CS358H at UT Austin

Chapter 1

Preliminaries: Linear Algebra over Complex Vector Spaces

Quantum computing is formulated in the language of linear algebra over complex vector spaces. This mathematical framework serves as the foundation for quantum states, measurements, and unitary evolutions. To prepare for these ideas, we first establish the necessary linear-algebraic basics.

We begin by recalling linear algebra on real vector spaces—the material typically covered in an undergraduate course. This review anchors the discussion in concepts you are already familiar with: span, linear independence, bases, dimension, linear transformations, and their matrix representations.

Next, we review the arithmetic of complex numbers. We will cover their algebraic form, geometric interpretation on the complex plane, conjugation, modulus and argument, Euler's formula, and basic operations and identities that recur throughout quantum computing.

Finally, we extend the familiar setting of real vector spaces to complex vector spaces. We will introduce complex vector spaces and inner products, emphasize conjugate-linearity in inner products, and discuss norms, orthogonality, and unitary operators. With these tools in place, we will be ready to develop the quantum formalism.

1.1 Vector Spaces: A Review

Start from what we are familiar with: the real vector space $\mathbb{R}^3$. This vector space consists of:

- An underlying number field $\mathbb{R}$. Elements in $\mathbb{R}$ are called *scalars*.
- A set $\mathbb{R}^3 = \{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \ : \ x_1, x_2, x_3 \in \mathbb{R}\}$. Elements in $\mathbb{R}^3$ are called *vectors*.
- An *addition operation* "$+$" between elements in $\mathbb{R}^3$: $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{bmatrix}$, where $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3$ and $\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \in \mathbb{R}^3$. Note that the result of the addition is again an element in $\mathbb{R}^3$.
- An *multiplication operation* "$\cdot$" between the field $\mathbb{R}$ and the set $\mathbb{R}^3$: $c \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} c \cdot x_1 \\ c \cdot x_2 \\ c \cdot x_3 \end{bmatrix}$, where $c \in \mathbb{R}$ and $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3$. Note that the result of the multiplication is again an element in $\mathbb{R}^3$.

Teaching Suggestions

During the lecture, show an example with $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, and $c = 3$.

1.1.1 Formal Definition of Vector Space

The above example illustrates all the important components of a vector space. But the real definition is more general. We now recall the full definition in Definition 1.1.1.

Definition 1.1.1 (Vector Space)**.** *A vector space over a field* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$[1] *is a non-empty set* V *together with two binary operations*

$$\cdot : \mathbb{F} \times V \to V \quad \text{and} \quad + : V \times V \to V$$

which satisfy the eight axioms listed below. In this context, the elements of V *are commonly called* vectors*, the elements of* $\mathbb{F}$ *are called* scalars*, the operations "*$\cdot$*" is called* scalar multiplication*, and the operations "*$+$*" is called* vector addition *or simply* addition*.*

To have a vector space, the following eight axioms must be satisfied for every $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$*, and* $a, b \in \mathbb{F}$*:*

1. **Associativity of vector addition:** $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
2. **Commutativity of vector addition:** $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
3. **Identity element of vector addition:** *There exists an element* $\mathbf{0} \in V$*, called the* zero vector*, such that* $\mathbf{v} + \mathbf{0} = \mathbf{v}$ *for all* $\mathbf{v} \in V$*.*
4. **Inverse elements of vector addition:** *For every* $\mathbf{v} \in V$*, there exists an element* $-\mathbf{v} \in V$*, called the additive inverse of* $\mathbf{v}$*, such that* $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$*.*
5. **Compatibility of scalar multiplication with field multiplication:**

   $$a \cdot (b \cdot \mathbf{v}) = (ab) \cdot \mathbf{v},$$

   where the "ab" in the right-hand side of the equation means the field multiplication of $\mathbb{F}$ *between its elements* a *and* b*.*
6. **Identity element of scalar multiplication:** $1 \cdot \mathbf{v} = \mathbf{v}$*, where* 1 *denotes the multiplicative identity in* $\mathbb{F}$*.*
7. **Distributivity of scalar multiplication with respect to vector addition:** $a \cdot (\mathbf{u} + \mathbf{v}) = a \cdot \mathbf{u} + a \cdot \mathbf{v}$.
8. **Distributivity of scalar multiplication with respect to field addition:** $(a + b) \cdot \mathbf{v} = a \cdot \mathbf{v} + b \cdot \mathbf{v}$.

$\Diamond$

[1] Actually, the concept of "field" can be more general. However, a formal treatment is out of the scope of this course. For this course, we will only use two fields: the real number field $\mathbb{R}$ and the complex number field $\mathbb{C}$.

More Examples of Vector Spaces. Anything satisfies the requirements in Definition 1.1.1 would be a vector space. Besides the familiar $\mathbb{R}^n$, let us see some more examples:

- **Space $C[a,b]$:** Each point of this space is a continuous real-valued function on (the real interval) $[a,b]$. The set of all these functions forms a real vector space with the algebraic operations defined in the usual way:

$$(x+y)(t) = x(t) + y(t),$$

$$(ax)(t) = ax(t) \quad (a \in \mathbb{R}).$$

  Namely, this an instantiation of Definition 1.1.1 with $V = C[a,b]$ and $\mathbb{F} = \mathbb{R}$.

- **Space $P(\mathbb{R})$:** All polynomials with coefficients from $\mathbb{R}$ form a vector space, under the standard polynomial arithmetic. Namely, this an instantiation of Definition 1.1.1 with $V = P(\mathbb{R})$ and $\mathbb{F} = \mathbb{R}$.

It is not hard to verify that these two spaces satisfy all the requirements stated in Definition 1.1.1.

Note also that as long as we have a vector space—regardless of what it looks like—it already possesses all the concepts we will review in Section 1.1.2.

1.1.2 Reviewing Basic Concepts in Vector Spaces

In a vector space, we study fundamental concepts such as span, linear independence, basis, dimension, and linear operators/transformations (represented by matrices). Let's begin by reviewing these ideas.

There are additional important concepts—such as eigenvalues and eigenvectors—which we will review later as needed.

Span

Definition 1.1.2 (Linear combination)**.** *Given vectors $v_1, \ldots, v_k \in V$ and scalars $a_1, \ldots, a_k \in \mathbb{F}$, a* linear combination *of the v_i is $a_1 v_1 + \cdots + a_k v_k$.* ◊

Definition 1.1.3 (Subspace)**.** *For a non-empty subset $S \subseteq V$, given vector $v, w \in S$ and scalars $a, b \in \mathbb{F}$, if $av + bw \in S$, then S is a subspace of V.* ◊

Definition 1.1.4 (Span)**.** *For a subset $S \subseteq V$, the* span *of S is*

$$\mathsf{Span}(S) = \left\{ \sum_{i=1}^{k} a_i v_i \ : \ k \in \mathbb{N},\ v_i \in S,\ a_i \in \mathbb{F} \right\}.$$

If $S = \{v_1, \ldots, v_n\}$ is finite we write $\mathsf{Span}\{v_1, \ldots, v_n\}$. ◊

Proposition 1.1.1. *$\mathsf{Span}(S)$ is a subspace of V and is the smallest subspace of V containing S (i.e., contained in every subspace that contains S).*

Example 1.1.1. *In $\mathbb{R}^3$, $\mathsf{Span}\{(1,0,0),(0,1,0)\} = \{(x,y,0) : x, y \in \mathbb{R}\}$, the xy-plane.*

Linear Independence

Definition 1.1.5 (Linear independence)**.** *A list (or set) of vectors $v_1, \ldots, v_n \in V$ is* linearly independent *if*

$$a_1 v_1 + \cdots + a_n v_n = 0 \quad \Rightarrow \quad a_1 = \cdots = a_n = 0.$$

Otherwise they are linearly dependent*.* ◊

Example 1.1.2. *In $\mathbb{R}^3$, the vectors $(1,0,0), (0,1,0), (1,1,0)$ are dependent because $(1,1,0) = (1,0,0) + (0,1,0)$.*

Basis

Definition 1.1.6 (Basis)**.** *A* basis *of V is a set of vectors $\mathcal{B} = \{b_1, \ldots, b_n\}$ such that*

1. *$\mathcal{B}$ spans V, i.e., $\mathsf{Span}(\mathcal{B}) = V$, and*
2. *$\mathcal{B}$ is linearly independent.*

◊

Theorem 1.1.1 (Uniqueness of coordinates)**.** *If $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is a basis of V, every $\mathbf{v} \in V$ can be written uniquely as $\mathbf{v} = \sum_{i=1}^{n} a_i \mathbf{b}_i$ with scalars $a_i \in \mathbb{F}$. The vector $(a_1, \ldots, a_n)$ is the coordinate vector of $\mathbf{v}$ relative to $\mathcal{B}$.* ◊

Example 1.1.3. *The standard basis of $\mathbb{R}^n$ is $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, where $\mathbf{e}_i$ has a 1 in the i-th position and 0 elsewhere.*

Dimension

Definition 1.1.7 (Dimension)**.** *If V has a finite basis, its* dimension*, denoted as $\dim V$, is the number of vectors in any basis. If no finite basis exists, V is infinite-dimensional.* ◊

Theorem 1.1.2 (Well-definedness)**.** *If V is finite-dimensional, all bases of V have the same cardinality. Hence $\dim V$ is well-defined.* ◊

Proposition 1.1.2 (Dimension inequalities)**.** *Let V be finite-dimensional.*

1. *Any (linearly) independent set has size at most $\dim V$; any spanning set has size at least $\dim V$.*
2. *If U is a subspace of V, then $\dim U \leq \dim V$, with equality iff $U = V$.*

Linear maps and Matrices

Definition 1.1.8 (Linear map (or operator))**.** *Let V and W be vector spaces over the same field $\mathbb{F}$. A **linear map** (or operator or transformation) from V to W is a function $T : V \to W$ with the following properties:*

- ***Additivity:***

$$T(u+v) = T(u) + T(v) \quad \text{for all } u, v \in V.$$

- ***Homogeneity:***

$$T(av) = aT(v) \quad \textit{for all } a \in \mathbb{F} \textit{ and all } v \in V.$$

We use $L(V, W)$ to denote the set of all linear maps from V to W. $\Diamond$

Definition 1.1.9 (Matrix of a linear map)**.** *Fix ordered bases $\mathcal{B} = (v_1, \ldots, v_n)$ of V and $\mathcal{C} = (w_1, \ldots, w_m)$ of W. For any linear map $T : V \to W$, there exists a unique matrix $M_T \in \mathbb{F}^{m\times n}$ such that for all $x \in V$,*

$$[T(x)]_{\mathcal{C}} = M_T \cdot [x]_{\mathcal{B}},$$

where $[x]_{\mathcal{B}}$ denotes the coordinates of vector $x \in V$ with respect to $\mathcal{B}$, and $[T(x)]_{\mathcal{C}}$ denotes the coordinates of vector $T(x) \in W$ with respect to $\mathcal{C}$.

Moreover, the j-th column of M_T is $[T(v_j)]_{\mathcal{C}}$, i.e., the coordinates of $T(v_j)$ with respect to $\mathcal{C}$. $\Diamond$

Example 1.1.4 (A 4×3 matrix as the matrix of a linear map)**.** *Let $V = \mathbb{R}^3$ and $W = \mathbb{R}^4$ over $\mathbb{R}$. Fix ordered bases*

$$\mathcal{B} = (v_1, v_2, v_3), \quad v_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \ v_2 = \begin{bmatrix}1\\1\\0\end{bmatrix}, \ v_3 = \begin{bmatrix}0\\1\\1\end{bmatrix},$$

and

$$\mathcal{C} = (w_1, w_2, w_3, w_4), \quad w_1 = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \ w_2 = \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \ w_3 = \begin{bmatrix}0\\0\\1\\0\end{bmatrix}, \ w_4 = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}.$$

Define a linear map $T : V \to W$ by prescribing its values on the basis $\mathcal{B}$:

$$T(v_1) = \begin{bmatrix}2\\1\\0\\-1\end{bmatrix}, \quad T(v_2) = \begin{bmatrix}0\\3\\1\\2\end{bmatrix}, \quad T(v_3) = \begin{bmatrix}-1\\0\\4\\1\end{bmatrix}.$$

Because $\mathcal{C}$ is the standard basis of $\mathbb{R}^4$, the coordinate vectors $[T(v_j)]_{\mathcal{C}}$ are exactly the columns written above. Therefore, by definition,

$$M_T = \big[T\big]_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{bmatrix}2 & 0 & -1\\ 1 & 3 & 0\\ 0 & 1 & 4\\ -1 & 2 & 1\end{bmatrix} \in \mathbb{R}^{4\times 3},$$

whose j-th column is $[T(v_j)]_{\mathcal{C}}$.

Now take any $x \in V$. Write x in $\mathcal{B}$-coordinates:

$$[x]_{\mathcal{B}} = \begin{bmatrix}a\\b\\c\end{bmatrix} \quad \textit{meaning} \quad x = a\,v_1 + b\,v_2 + c\,v_3.$$

By linearity,

$$T(x) = a\,T(v_1) + b\,T(v_2) + c\,T(v_3),$$

so in $\mathcal{C}$-coordinates,

$$[T(x)]_{\mathcal{C}} = a\,[T(v_1)]_{\mathcal{C}} + b\,[T(v_2)]_{\mathcal{C}} + c\,[T(v_3)]_{\mathcal{C}} = M_T \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

This verifies the defining relation $[T(x)]_{\mathcal{C}} = M_T\,[x]_{\mathcal{B}}$.

Concretely, pick $[x]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, i.e., $x = 1 \cdot v_1 + 2 \cdot v_2 - 1 \cdot v_3$. Then

$$[T(x)]_{\mathcal{C}} = M_T \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \cdot 1 + 0 \cdot 2 + (-1) \cdot (-1) \\ 1 \cdot 1 + 3 \cdot 2 + 0 \cdot (-1) \\ 0 \cdot 1 + 1 \cdot 2 + 4 \cdot (-1) \\ -1 \cdot 1 + 2 \cdot 2 + 1 \cdot (-1) \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ -2 \\ 2 \end{bmatrix}.$$

Independently, compute using $T(v_j)$:

$$T(x) = 1 \cdot T(v_1) + 2 \cdot T(v_2) - 1 \cdot T(v_3) = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -1 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 3 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} -1 \\ 0 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ -2 \\ 2 \end{bmatrix},$$

which matches $[T(x)]_{\mathcal{C}}$ above, as expected.

Kernel, Image, Support, Rank-Nullity

Definition 1.1.10 (Image (aka Range or Column Space) and Rank)**.** *The* image *of a linear operator $T \in L(V, W)$, denoted by $\mathsf{Im}(T)$, is the set defined as*

$$\mathsf{Im}(T) = \{T(v) : v \in V\}.$$

It is also known as the range *of T.*

Note that $\mathsf{Im}(T)$ is exactly that vector space spanned by the column vectors of (the matrix representation of) T. For this reason, $\mathsf{Im}(T)$ is also known as the column space *of T.*

The dimension of $\mathsf{Im}(T)$ is known as the rank *of T, denoted as $\mathsf{rank}(T)$.* ◊

Definition 1.1.11 (Kernel (aka Null Space) and Nullity)**.** *The* kernel *of a linear operator $T \in L(V, W)$, denoted by $\mathsf{Ker}(T)$, is the set defined as*

$$\mathsf{Ker}(T) = \{v \in V : T(v) = \mathbf{0}\}.$$

It is also known as the null space *of T.*

The dimension of $\mathsf{Ker}(T)$ is known as the nullity *of T, denoted as $\mathsf{nullity}(T)$.* ◊

Theorem 1.1.3 (Rank–Nullity)**.** *For a linear map $T \in L(V, W)$ between finite-dimensional spaces V and W, it holds that*

$$\dim V = \mathsf{rank}(T) + \mathsf{nullity}(T).$$

◊

Proposition 1.1.3. *For $T : V \to W$ with V finite-dimensional:*

(a) T is injective $\iff$ $\mathsf{Ker}(T) = \{0\}$ $\iff$ $\mathsf{nullity}(T) = 0$.

(b) T is surjective $\iff$ $\operatorname{rank}(T) = \dim W$.

(c) If $\dim V = \dim W$, then injective $\iff$ surjective $\iff$ bijective.

1.2 Complex Vector Spaces

We emphasize again that the properties listed in Section 1.1.2 are general. They hold for all vector spaces.

We will be particularly interested in complex vector spaces, i.e., the case of Definition 1.1.1 with $\mathbb{F} = \mathbb{C}$. Even after fixing $\mathbb{F} = \mathbb{C}$, there remain many ways to construct vector spaces satisfying Definition 1.1.1—by choosing different sets V and defining addition and scalar multiplication accordingly. All such constructions are called complex vector spaces.

Among these possibilities, the specific case with $\mathbb{F} = \mathbb{C}$ and $V = \mathbb{C}^n$ for some $n \in \mathbb{N}$ (with the natural addition and multiplication) is sufficient for our purposes. This will be our main focus, and we will rarely discuss other vector spaces in this course.

1.2.1 Reviewing Complex Number Arithmetic

A *complex number* is a number of the form

$$z = a + bi,$$

where $a, b \in \mathbb{R}$, and i is the imaginary unit defined by

$$i^2 = -1.$$

The set of all complex numbers is denoted by $\mathbb{C}$.

For $z = a + bi$,

- a is called the *real part* of $z = a + bi$, denoted $\operatorname{Re}(z) = a$.
- b is called the *imaginary part* of z, denoted $\operatorname{Im}(z) = b$.

Euler's Formula and Polar Form

Probably the most important formula involving complex numbers is Euler's formula:

$$e^{i\theta} = \cos(\theta) + i\sin(\theta). \tag{1.1}$$

Equation (1.1) implies that every complex number z can be written in the so-called *polar form*

$$z = re^{i\theta},$$

with $r \geq 0$ and $\theta \in \mathbb{R}$. Think why.

The following two identities are direct consequences of Euler's formula and are highly useful.

$$\cos\theta = \frac{1}{2}\left(e^{i\theta} + e^{-i\theta}\right) \quad \text{and} \quad \sin\theta = \frac{1}{2i}\left(e^{i\theta} - e^{-i\theta}\right).$$

One way to appreciate Euler's formula is by recognizing how it can be used to derive various trigonometric identities—identities that many students struggle to learn and memorize when first encountering them in secondary school.

As an example, let me show how to derive the identity $\cos(\alpha+\beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$ using Euler's formula:

$$\begin{aligned}
\cos(\alpha+\beta) &= \mathrm{Re}\left(e^{i(\alpha+\beta)}\right) \\
&= \mathrm{Re}\left(e^{i\alpha}e^{i\beta}\right) \\
&= \mathrm{Re}\left((\cos(\alpha) + i\sin(\alpha)) \cdot (\cos(\beta) + i\sin(\beta))\right) \\
&= \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta),
\end{aligned}$$

where $\mathrm{Re}(\cdot)$ denotes the real part of a complex number.

Basic Arithmetic Operations

Let $z_1 = a + bi$ and $z_2 = c + di$ be two complex numbers. Then:

$$z_1 + z_2 = (a+c) + (b+d)i,$$

$$z_1 - z_2 = (a-c) + (b-d)i.$$

The product of $z_1 = a + bi$ and $z_2 = c + di$ is given by:

$$z_1 \cdot z_2 = (a+bi)(c+di) = (ac - bd) + (ad + bc)i.$$

I assume you are quite familiar with the above basic arithmetic operations on complex numbers. The only operations that might be somewhat tricky to memorize are division and roots. Let us quickly review them.

- **Division of Complex Numbers:** For two complex numbers $z_1 = a + bi$ and $z_2 = c + di$, divide by multiplying the numerator and denominator by the conjugate of z_2:

$$\frac{z_1}{z_2} = \frac{(a+bi)(c-di)}{(c+di)(c-di)} = \frac{(a+bi)(c-di)}{c^2+d^2} = \frac{(ac+bd)+(bc-ad)i}{c^2+d^2} = \frac{(ac+bd)}{c^2+d^2} + \frac{(bc-ad)}{c^2+d^2}i.$$

- **Roots of Complex Numbers:** Let n be a positive integer. For a complex number $z = r(\cos\theta + i\sin\theta)$, there are exactly n distinct the n-th roots. They are given by:

$$z_k = r^{\frac{1}{n}}\left(\cos\left(\frac{\theta+2k\pi}{n}\right) + i\sin\left(\frac{\theta+2k\pi}{n}\right)\right),$$

  where $k = 0, 1, 2, \ldots, n-1$. These roots are the n equally spaced solutions on the complex plane.

De Moivre's Formula

De Moivre's formula (also known as de Moivre's theorem or de Moivre's identity) states that for any real number θ and integer n, the following holds:

$$\big(\cos(\theta) + i\sin(\theta)\big)^n = \cos(n\theta) + i\sin(n\theta).$$

Modulus of a Complex Number

The *modulus* (or absolute value) of z, denoted $|z|$, is the distance from z to the origin in the complex plane:

$$|z| \;=\; \sqrt{a^2 + b^2}.$$

Equivalent characterizations. Using complex conjugation: if $\overline{z} = a - bi$, then

$$|z|^2 = z\,\overline{z} = (a+bi)(a-bi) = a^2 + b^2, \quad \text{so} \quad |z| = \sqrt{z\,\overline{z}}.$$

In polar form $z = re^{i\theta}$ with $r \geq 0$ and $\theta \in \mathbb{R}$,

$$|z| = r.$$

Basic Properties. For all $z, w \in \mathbb{C}$ and $\lambda \in \mathbb{C}$:

$$\begin{aligned}
&|z| \geq 0, \quad |z| = 0 \iff z = 0,\\
&|\overline{z}| = |z|, \quad |-z| = |z|,\\
&|zw| = |z|\,|w|, \quad \Big|\frac{z}{w}\Big| = \frac{|z|}{|w|} \quad (w \neq 0),\\
&|z+w| \leq |z| + |w| \quad \text{(triangle inequality)},\\
&\big||z| - |w|\big| \leq |z-w| \quad \text{(reverse triangle inequality)},\\
&\text{If } z = re^{i\theta}, \text{ then } |z^n| = |z|^n = r^n \text{ for } n \in \mathbb{Z}.
\end{aligned}$$

1.3 Inner Products: Adding Geometry to Vector Spaces

We have defined vector spaces and reviewed several foundational notions in Section 1.1.2. These concepts capture the algebraic structure of vector spaces and allow us to manipulate and represent vectors symbolically and computationally.

Despite this, the bare notion of a vector space is not yet rich enough for many of our goals. In particular, a general vector space does not come equipped with any "geometric structure:" there is no canonical way to talk about lengths or volumes, nor about angles or distances between vectors.

For example, it is not a priori clear how to define length and angle in the vector space $C[a,b]$ and $P(\mathbb{R})$ we discussed on Page 10.

Without such notions, we cannot formalize geometric ideas like orthogonality, projections, or "nearest" vectors, all of which are indispensable in analysis, optimization, numerical linear algebra, signal processing, and our focus of quantum computing!

To address this, we now endow vector spaces with geometry via the concept of an *inner product*.

1.3.1 Starting Point: the Euclidean Space $\mathbb{R}^3$

It is widely known that the $\mathbb{R}^3$ is a *Euclidean space*, which means that there are well-defined geometry concepts. Let's first recall this fact.

In Euclidean 3-space $\mathbb{R}^3$, vectors are triples $\mathbf{v} = (v_1, v_2, v_3)$ with real components.

We have a notion of length of vectors by the so-called *Euclidean norm*:

$$\|v\|_2 \coloneqq \sqrt{v_1^2 + v_2^2 + v_3^2},$$

where the "2" in the subscript of $\|\cdot\|_2$ is because that Euclidean norm is also called the quadratic norm, L^2 norm, 2 norm, or square norm.

We also have a notion of distance between the "end points" of vectors by the so-called *Euclidean distance*:

$$\mathsf{dist}(u, v) \coloneqq \|u - v\|_2 = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2}.$$

Actually, all these geometric notions can be unified by the concept of *inner product*. In the special case of $\mathbb{R}^3$, the inner product of $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ is

$$\langle \mathbf{v}, \mathbf{w} \rangle \;=\; v_1 w_1 + v_2 w_2 + v_3 w_3.$$

This single operation encodes several geometric notions: length (norm), distance, angles, and orthogonality. Let's see how.

Norm induced by inner product

The *norm* (length) induced by the inner product is

$$\|\mathbf{v}\| \;=\; \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \;=\; \sqrt{v_1^2 + v_2^2 + v_3^2}.$$

Example 1.3.1. *If* $\mathbf{v} = (3, -4, 12)$*, then* $\|\mathbf{v}\| = \sqrt{3^2 + (-4)^2 + 12^2} = \sqrt{9 + 16 + 144} = \sqrt{169} = 13$.

Distance induced by norm (induced by inner product)

The distance between $\mathbf{v}$ and $\mathbf{w}$ is the norm of their difference:

$$d(\mathbf{v}, \mathbf{w}) \;=\; \|\mathbf{v} - \mathbf{w}\| \;=\; \sqrt{(\mathbf{v} - \mathbf{w}) \cdot (\mathbf{v} - \mathbf{w})}.$$

Example 1.3.2. *For* $\mathbf{v} = (1, 2, 3)$ *and* $\mathbf{w} = (4, 0, -1)$*,*

$$\mathbf{v} - \mathbf{w} = (-3, 2, 4), \qquad d(\mathbf{v}, \mathbf{w}) = \sqrt{(-3)^2 + 2^2 + 4^2} = \sqrt{9 + 4 + 16} = \sqrt{29}.$$

Angle induced by inner product

The angle $\theta \in [0, \pi]$ between nonzero vectors is defined by the *cosine formula*

$$\cos\theta \coloneqq \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\|\|\mathbf{w}\|} \quad \Longleftrightarrow \quad \langle \mathbf{v}, \mathbf{w} \rangle \;=\; \|\mathbf{v}\|\|\mathbf{w}\| \cos\theta.$$

Rationale behind this definition: Place the tails of vectors $\mathbf{a}$ and $\mathbf{b}$ at the same point (say, the origin). The three sides of the triangle are:

- side 1: the segment representing $\mathbf{a}$, length $\|\mathbf{a}\|$,
- side 2: the segment representing $\mathbf{b}$, length $\|\mathbf{b}\|$,
- side 3: the segment from the head of $\mathbf{b}$ to the head of $\mathbf{a}$, which is $\mathbf{a}-\mathbf{b}$, length $\|\mathbf{a}-\mathbf{b}\|$.

Let θ be the angle between $\mathbf{a}$ and $\mathbf{b}$, i.e., the smaller angle in $[0,\pi]$ formed when both are drawn from the origin.

Next, apply the law of cosines to that triangle. Recall that

Lemma 1.3.1 (law of cosines)**.** *For any triangle with side lengths x, y, z and included angle θ between x and y, it holds that*

$$z^2 = x^2 + y^2 - 2xy\cos\theta.$$

◊

Identify $x = \|\mathbf{a}\|$, $y = \|\mathbf{b}\|$, and $z = \|\mathbf{a}-\mathbf{b}\|$. Therefore,

$$\|\mathbf{a}-\mathbf{b}\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2\,\|\mathbf{a}\|\,\|\mathbf{b}\|\cos\theta. \tag{1.2}$$

Expand $\|\mathbf{a}-\mathbf{b}\|^2$ using the inner product. By the inner-product definition of the norm,

$$\|\mathbf{a}-\mathbf{b}\|^2 = \langle \mathbf{a}-\mathbf{b},\, \mathbf{a}-\mathbf{b}\rangle = \langle \mathbf{a},\mathbf{a}\rangle - 2\langle \mathbf{a},\mathbf{b}\rangle + \langle \mathbf{b},\mathbf{b}\rangle = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 - 2\langle \mathbf{a},\mathbf{b}\rangle. \tag{1.3}$$

Compare the two expressions. Equation (1.2) and Equation (1.3) and canceling common terms yields

$$\langle \mathbf{a},\mathbf{b}\rangle = \|\mathbf{a}\|\,\|\mathbf{b}\|\cos\theta.$$

Example 1.3.3. *For $\mathbf{v} = (1,1,0)$ and $\mathbf{w} = (1,-1,0)$,*

$$\langle \mathbf{v},\mathbf{w}\rangle = 1\cdot 1 + 1\cdot(-1) + 0\cdot 0 = 0,$$

so $\cos\theta = 0$ and $\theta = \pi/2$ (they are perpendicular).

Orthogonality and projections

Vectors are *orthogonal* if their inner product is zero.

Given a nonzero vector $\mathbf{u}$, the projection of $\mathbf{v}$ onto $\mathbf{u}$ is

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\langle \mathbf{u},\mathbf{v}\rangle}{\langle \mathbf{u},\mathbf{u}\rangle}\mathbf{u} = \|\mathbf{v}\|\cos(\theta)\frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

The component orthogonal to $\mathbf{u}$ is $\mathbf{v} - \mathrm{proj}_{\mathbf{u}}(\mathbf{v})$. See Figure 1.1.

Example 1.3.4. *Let $\mathbf{u} = (1,2,2)$ and $\mathbf{v} = (3,0,1)$. Then $\langle \mathbf{v},\mathbf{u}\rangle = 3\cdot 1 + 0\cdot 2 + 1\cdot 2 = 5$ and $\langle \mathbf{u},\mathbf{u}\rangle = 1+4+4 = 9$, so*

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{5}{9}(1,2,2) = \left(\tfrac{5}{9}, \tfrac{10}{9}, \tfrac{10}{9}\right),$$

and $\mathbf{v} - \mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{22}{9}, -\frac{10}{9}, -\frac{1}{9}\right)$ is orthogonal to $\mathbf{u}$.

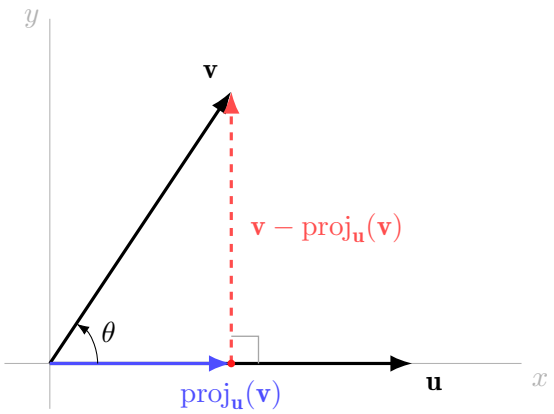


Figure 1.1: Illustrating Projection

Key inequalities in $\mathbb{R}^3$

For all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$:

$$\begin{aligned} \textbf{Cauchy–Schwarz:} \quad & |\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \|\mathbf{w}\|, \\ \textbf{Triangle inequality:} \quad & \|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|. \end{aligned}$$

These follow from the inner product structure.

1.3.2 Complex Inner-Product Space $\mathbb{C}^n$

Note that the geometric concepts we reviewed in Section 1.3.1 treat the inner product as a "black box"—they do not depend on how the inner product is defined or computed. Consequently, as long as we equip a vector space with a suitable inner product, it will induce all the geometric notions we discussed in Section 1.3.1.

Next, we introduce in Definition 1.3.1 an additional definition concerning linear operators on vector spaces (equipped with an inner product). This definition is important and should be understood in contrast to Definition 1.1.10. However, since it involves the notion of an inner product, we have to postpone its introduction until this point.

Definition 1.3.1 (Support (aka Co-image or Row Space))**.** *Let V be an inner-product space and W be a vector space. The* support *of a linear operator $T \in L(V, W)$, denoted by $\mathsf{Supp}(T)$, is defined to be the orthogonal complement of its kernel:*

$$\mathsf{Supp}(T) \coloneqq \mathsf{Ker}(T)^{\perp} \coloneqq \{v \in V : \langle v, u \rangle = 0 \ \forall u \in \mathsf{Ker}(T)\}.$$

It is also known as the coimage *of T.*

Note that $\mathsf{Supp}(T)$ is exactly that vector space spanned by the row vectors of (the matrix representation of) T. For this reason, $\mathsf{Supp}(T)$ is also known as the row space *of T.*

$\Diamond$

Inner Product on $\mathbb{C}^n$

As we said, quantum computing uses the particular complex vector spaces $\mathbb{C}^n$. We now proceed

to define a inner product for it:

Definition 1.3.2 (Standard inner product on $\mathbb{C}^n$)**.** *For $v = (v_1, \ldots, v_n)$, $w = (w_1, \ldots, w_n) \in \mathbb{C}^n$, define*

$$\langle v, w \rangle \coloneqq \sum_{i=1}^{n} \overline{v_i} \cdot w_i = \mathbf{v}^\dagger \mathbf{w},$$

where $\overline{v_i}$ is the complex conjugate of v_i. ◊

Why $\overline{v_i}$ in Definition 1.3.2? The most natural option—reusing the same definition as for real vector spaces (which replaces $\overline{v_i}$ with v_i)—does not make much sense for $\mathbb{C}^n$. For example, if we reuse the real inner product definition to define inner product for $\mathbb{C}^n$, then the inner product induced norm may no longer be a real number (which cannot be a reasonable measure for "length"). On the other hand, it is easy to see that Definition 1.3.2 always ensure that $\|v\| \coloneqq \sqrt{\langle v, v \rangle}$ is a real number.

Moreover, it is easy to see that Definition 1.3.2, when restricted to real vector spaces, always becomes the same definition as

$$\langle v, w \rangle \coloneqq \sum_{i=1}^{n} v_i \cdot w_i,$$

This is simply because that $\overline{r} = r$ for all $r \in \mathbb{R}$.

Thus, Definition 1.3.2 should be considered as the more general, more "correct" definition for inner product.

Geometry Structure over $\mathbb{C}^n$. With Definition 1.3.2, we can define (almost) all the geometry concepts as in Section 1.3.1 for $\mathbb{C}^n$, including norm, distance, orthogonality, projection, Cauchy–Schwarz inequality and Triangle inequality. Note that the only exception is the concept of angle; its definition will be more tricky as naively copying that same definition will lead to a complex value for $\cos(\theta)$. Fortunately, we will not need the notion anyway in this course. Thus, we do not present further discussion here.

Example 1.3.5 (Worked examples in $\mathbb{C}^3$)**.** *Let $v = (1+i,\ 2,\ i)$ and $w = (2,\ 1-i,\ -i)$ in $\mathbb{C}^3$. Using the convention $\langle v, w \rangle = \sum_{k=1}^{3} \overline{v_k}\, w_k$, we have*

$$\begin{aligned}
\langle v, w \rangle &= \overline{1+i} \cdot 2 \ + \ \overline{2} \cdot (1-i) \ + \ \overline{i} \cdot (-i) \\
&= (1-i) \cdot 2 \ + \ 2(1-i) \ + \ (-i) \cdot (-i) \\
&= 2 - 2i \ + \ 2 - 2i \ - \ 1 \\
&= 3 - 4i,
\end{aligned}$$

$$\|v\|^2 = |1+i|^2 + |2|^2 + |i|^2 = 2 + 4 + 1 = 7, \quad \|v\| = \sqrt{7},$$
$$\|w\|^2 = |2|^2 + |1-i|^2 + |-i|^2 = 4 + 2 + 1 = 7, \quad \|w\| = \sqrt{7}.$$

Orthogonality holds iff $\langle v, w \rangle = 0$.

1.4 Hilbert Space—the Playground for Quantum Computing

Our course on quantum computing will take place entirely in the complex vector space $\mathbb{C}^n$, equipped with the (standard) inner product:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} \overline{u_i} v_i = \mathbf{u}^\dagger \mathbf{v}.$$

This vector space is an example of the so-called *Hilbert space*, which is named after the famous German mathematician David Hilbert.

You will often hear people in quantum computing talk about Hilbert space. In most cases, this simply refers to this inner-product space $\mathbb{C}^n$ we defined so far (or its infinite-dimensional analogues).

There are two caveats worth noting:

- The term "Hilbert space" refers to a class of inner-product spaces that satisfy certain axioms. The space $\mathbb{C}^n$ with the inner product above is one example, but there are many others (including infinite-dimensional spaces) that also qualify as Hilbert spaces.
- The formal definition of a Hilbert space involves some math-analytic conditions, notably *completeness*. We will not develop the full theory in this undergraduate course. For those who are interested, please refer to Supplementary Box 1.4.1.

> Nothing from Supplementary Box 1.4.1 will be tested on quizzes or exams!

Supplementary Box 1.4.1: What is a Hilbert Space?

A Hilbert space is a *complete* (real or complex) inner-product space. That is, it is a *complete* vector space equipped with an inner product.

We have already explained what a vector space and inner product are. The only thing left is the notion of *completeness*. We now explain this term:

- This term is used to describe a *metric space*. We do not formally define a metric space. Roughly, it is a space that allows us to talk about the concept of "distance". An inner-product space is always a metric space, because inner product induces a norm, and a norm always induces a metric $d(\cdot, \cdot)$ as follows:

$$d(\mathbf{u}, \mathbf{v}) \coloneqq \|\mathbf{u} - \mathbf{v}\| \coloneqq \sqrt{\langle \mathbf{u}, \mathbf{v} \rangle}.$$

- In mathematical analysis, a metric space M is called *complete* (or a Cauchy space) if every Cauchy sequence of points in M has a limit that is also in M.

  Intuitively, a space is complete if there are "no points missing from it" (inside or at the boundary). For instance, the set of rational numbers is not complete—for example, the number e is "missing" from it, even though one can construct a Cauchy sequence of rational numbers (i.e., the following Maclaurin series) that converges to e.

  - Note that $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.
  - Let $s_i = \sum_{k=0}^{i} \frac{1}{k!}$. Then, $\{s_i\}$ is a Cauchy sequence as it satisfies the Cauchy property that: $\forall \varepsilon > 0$, $\exists N$ such that $\forall m, n > N$, it holds that $|s_m - s_n| < \varepsilon$.

– Also, $\lim_{n\to\infty} s_n = e$.

- Essentially, a complete metric space is the minimal place where you can talk about concepts like *limit* and *convergence*. It has nice math-analytical properties.

1.5 *Bonus Section: Why Complex Numbers for Quantum Mechanics?

Nothing from Section 1.5 will be tested on quizzes or exams!

We keep saying that quantum mechanics and quantum computing need to use the mathematical language of complex vector spaces—more precisely, Hilbert spaces. A natural question is why we must use complex numbers, and why we cannot simply treat everything using real vector spaces, which we are more familiar and comfortable with.

In this bonus section, we provide some partial explanations for the necessity of complex numbers in quantum mechanics.

Please note that what we provide here should be viewed as rationale and clues pointing to the indispensable role of complex numbers in quantum mechanics. As for a definitive reason why quantum mechanics must require complex numbers, that remains beyond the current reach of human knowledge in physics.

1.5.1 Why Complex Numbers?

Shore Answer

Real-valued quantum mechanics fails to capture all phenomena. Attempts to describe quantum mechanics with only real numbers result in a restricted theory that cannot reproduce all experimental results, such as entanglement and certain interference effects.

Longer answer

We provide more details. But what provided below are not really an "answer" yet. They are just some justification/rationale behind the essential role of complex numbers in quantum mechanics.

1. Complex numbers are not unique to quantum mechanics. They are already there in classical physics, specially in the topis involving oscillatory phenomena and **wave functions**. See Section 1.5.2 and Section 1.5.3 as an example—it is more mathematically convenient and straightforward to express these concepts in terms of complex exponentials rather than real trigonometric functions.

2. Quantum mechanics have a **wave-like nature**:

   - As demonstrated by phenomena such as interference and diffraction (check this short YouTube video on double-split experiment).
   - Schrödinger equation inherently uses complex numbers. For a non-relativistic particle, the time-dependent Schrödinger equation is:

$$i\hbar \frac{\partial}{\partial t} |\psi(x,t)\rangle = \hat{H} |\psi(x,t)\rangle \,, \tag{1.4}$$

where:

- $\psi(x,t)$ is the wavefunction (a complex-valued function),
- $\hbar$ is the reduced Planck constant,
- $\hat{H}$ is the Hamiltonian operator (energy operator).

Moreover, the exemplary systems described in Sections 1.5.2 and 1.5.3 are second-order differential equations. Consine functions are already sufficient to serve as their solutions; it is fine if you think exponential functions as being overkilling. In contrast, the Schrödinger's equation is 1st order differential equation; this makes exponential function necessary. This serves as one of the clue about the necessity of the e^{it} wave function form in QM.

3. The explanations above merely state that "complex numbers are a matter of mathematical convenience," but they do not prove that complex numbers are truly necessary. One can still ask:

   Question: *Are complex numbers really essential for quantum mechanics?*

   Indeed, the answer is NO for classical physics like oscillatory phenomenon and wave functions. So, it might be natural to hope that complex numbers are not essential for quantum mechanics as well. However, for quantum mechanics, the answer to this question is far more nuanced. Even in the past five years, both experimental and theoretical research has continued to explore this topic. Eventually, the answer is YES! We will not dive into further details. For more insights, I recommend reading Quantum Mechanics Must Be Complex, by Alessio Avella.

1.5.2 Complex Numbers in Simple Harmonic Oscillation

The restoring force in a simple harmonic oscillator is given by Hooke's law:

$$\mathbf{F} = -k\mathbf{x},$$

where:

- $\mathbf{F}$ is the restoring force,
- k is the spring constant (a measure of stiffness),
- $\mathbf{x}$ is the displacement from equilibrium.

The negative sign indicates that the force acts in the opposite direction to the displacement.

We apply Newton's second law:

$$\mathbf{F} = m\mathbf{a},$$

where

$$\mathbf{a} = \ddot{\mathbf{x}}$$

(the second derivative of x with respect to time, i.e., acceleration). Substituting $F = -kx$, we get:

$$m\ddot{\mathbf{x}} = -k\mathbf{x}.$$

Rearranging, we obtain the simple harmonic oscillator equation:

$$\ddot{\mathbf{x}} + \frac{k}{m}\mathbf{x} = 0.$$

Define

$$\omega^2 = \frac{k}{m},$$

where ω is the *angular frequency* of the oscillator. The equation becomes:

$$\ddot{\mathbf{x}} + \omega^2\mathbf{x} = 0 \tag{1.5}$$

Solving the Equation. Equation (1.5) is a second-order differential equation. Most of you have learned the following solution

$$x(t) = C\cos(\omega t + \phi),$$

where:

- C is the amplitude (maximum displacement),
- ϕ is the phase angle (determined by the initial conditions).

This form is often more intuitive, as it directly describes the motion in terms of amplitude, frequency, and phase.

However, it is not hard to see that $x(t) = C \cdot e^{i(\omega t+\phi)}$ is also a solution to the simply harmonic oscillation equation. Indeed, $C\cos(\omega t + \phi)$ is nothing but the real part of $C \cdot e^{i(\omega t+\phi)}$, by Euler's formula. The form $x(t) = C \cdot e^{i(\omega t+\phi)}$ enjoys the following advantages:

1. On the complex plane, the real part of $C \cdot e^{i(\omega t+\phi)}$ can be perceived as the "shadow" (formally, the projection) of $C \cdot e^{i(\omega t+\phi)}$ on the x axis while $x(t) = C \cdot e^{i(\omega t+\phi)}$ moves as a circle. That is, while $x(t)$ is rotating around the original, the real part $\mathrm{Re}(x(t)) = |C|\cos(\omega t + \phi)$ is a point bouncing between $-|C|$ and $|C|$ on the real line, as illustrated in Figure 1.2.
2. Exponential functions are much more convenient for many calculations, such as differentiation and integration.

1.5.3 Complex Numbers in Electromagnetic Wave Equations

The electromagnetic wave equations in free space is given as:

$$\nabla^2\mathbf{E} - \mu_0\varepsilon_0\frac{\partial^2\mathbf{E}}{\partial t^2} = 0,$$
$$\nabla^2\mathbf{B} - \mu_0\varepsilon_0\frac{\partial^2\mathbf{B}}{\partial t^2} = 0.$$

These equations describe how the electric field ($\mathbf{E}$) and magnetic field ($\mathbf{B}$) propagate in free space. Let us break down and explain the terms:

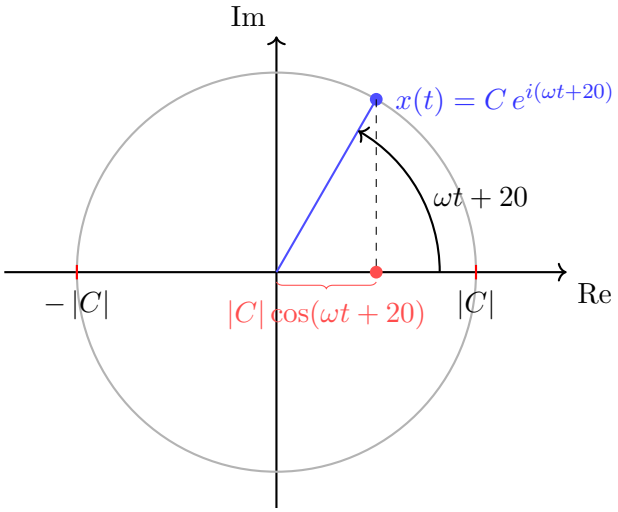


Figure 1.2: Example for $x(t) = C \cdot e^{i(\omega t+\phi)}$ with $\phi = 20$

1. **E**: The **electric field**, a vector field that represents the force experienced by a charged particle in the presence of an electric field. It is a function of position ($\mathbf{r}$) and time (t):

$$\mathbf{E} = \mathbf{E}(\mathbf{r}, t).$$

   In an electromagnetic wave, **E** oscillates perpendicularly to the direction of propagation and to the magnetic field **B**.

2. **B**: The **magnetic field**, a vector field that represents the force on a moving charged particle in the presence of a magnetic field. It is also a function of position and time:

$$\mathbf{B} = \mathbf{B}(\mathbf{r}, t).$$

   In an electromagnetic wave, **B** oscillates perpendicularly to the direction of propagation and to the electric field **E**.

3. ∇^2: The **Laplacian operator**, which represents the spatial variation of the field. In Cartesian coordinates, it is defined as:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

   For example, $\nabla^2\mathbf{E}$ describes how the electric field varies in space.

4. $\frac{\partial^2}{\partial t^2}$: The **second time derivative**, which describes how the field (either **E** or **B**) varies with time. For example, $\frac{\partial^2\mathbf{E}}{\partial t^2}$ describes the temporal change of the electric field.

5. μ_0: The **permeability of free space**, a physical constant that describes how magnetic fields interact with the vacuum. Its value is:

$$\mu_0 = 4\pi \times 10^{-7}\,\mathrm{N/A}^2.$$

6. ε_0: The **permittivity of free space**, a physical constant that describes how electric fields

interact with the vacuum. Its value is:

$$\varepsilon_0 \approx 8.854 \times 10^{-12}\,\mathrm{F/m}.$$

7. $\mu_0\varepsilon_0$: This combination determines the **speed of light** in free space, given by:

$$c = \frac{1}{\sqrt{\mu_0\varepsilon_0}}.$$

Numerically:

$$c \approx 3 \times 10^8\,\mathrm{m/s}.$$

Wave Solutions

The solutions to the electromagnetic wave equation are sinusoidal traveling waves. For the electric field **E**, the solution can be written as:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi),$$

where:

- $\mathbf{E}_0$: The amplitude of the electric field,
- $\mathbf{k}$: The wavevector, which gives the direction of propagation and the wavenumber ($k = 2\pi/\lambda$, where λ is the wavelength),
- ω: The angular frequency ($\omega = 2\pi f$, where f is the frequency),
- ϕ: The phase constant, determined by initial conditions,
- $\mathbf{r}$: The position vector, (x, y, z),
- t: Time.

Similarly, for the magnetic field **B**, the solution is:

$$\mathbf{B}(\mathbf{r}, t) = \mathbf{B}_0 \cos(\mathbf{k} \cdot \mathbf{r} - \omega t + \phi).$$

Supplementary Box 1.5.1: Relation between Electric and Magnetic Fields

The electric field (**E**) and magnetic field (**B**) in an electromagnetic wave are related as follows:

1. **E**, **B**, and the wavevector **k** (direction of wave propagation) are mutually perpendicular:

$$\mathbf{E} \perp \mathbf{B} \perp \mathbf{k}.$$

2. The magnitudes of **E** and **B** are related by:

$$|\mathbf{B}| = \frac{|\mathbf{E}|}{c},$$

where $c = \frac{1}{\sqrt{\mu_0\varepsilon_0}}$ is the speed of light.

Chapter 2
Postulate (1/4): State Space

Let us start by stating the first postulate:

Postulate 1: An (isolated) physical system is completely described by its state vector, which is a *unit* vector in some Hilbert space.

Recall that we defined a Hilbert space in Section 1.4. In particular, we stated that we will focus exclusively on a specific Hilbert space, namely $\mathbb{C}^n$ for some dimension n. Under this convention, **Postulate 1** states that for any given physical system, there exists a dimension parameter n such that the state of the system is completely described by a vector $\mathbf{v} \in \mathbb{C}^n$ with $\|\mathbf{v}\| = 1$.

Some important highlights:

- Recall that the norm $\|\mathbf{v}\|$ on $\mathbb{C}^n$ is defined to be the inner product induced norm
$$\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\mathbf{v}^\dagger \mathbf{v}}.$$
- The dimension n depends on the physical system under consideration. Different physical systems may require different values of n in $\mathbb{C}^n$ to describe their states.
- We emphasize the *unit-norm* requirement: although there are infinitely many vectors in the vector space $\mathbb{C}^n$, only those with length 1 (i.e., unit vectors) can be used to represent quantum states.

2.1 Two-Dimensional Quantum System—the Qubit

The simplest quantum system is when $n = 2$, i.e., the space $\mathbb{C}^2$.

The state of such a 2D quantum system can always be fully described by a unit vector $v \in \mathbb{C}^2$, and every unit vector $v \in \mathbb{C}^2$ could possible be corresponding to a state of the system.

Such a 2D quantum system is referred to as a *qubit*.

Before we dive into more mathematical discussions of such 2D quantum systems, let us first see two examples to help us imagine the them.

Remark 2.1.1 (How does the microscopic world look like?)**.** *Before we begin, a quick clarification: we will sometimes use simplified, intuitive diagrams that depict atoms or other microscopic particles as little dots or balls moving in fixed orbits. Please keep in mind that this is not accurate. In the microscopic realm, our everyday intuitions about position and shape no longer apply. It is a quantum world, where familiar quantities such as position, speed, and kinetic energy are represented by* **wave functions**.

For example, you should not imagine an electron as a tiny ball located at a precise point in space. Rather, an electron is better described by a probability distribution: before measurement, it is spread out and "exists" everywhere that its wave function has support, and upon observation it

appears at a particular location with a probability determined by that wave function. For a more accurate visualization of these ideas, see this YouTube video.

2.1.1 Example 1: Spin-$\frac{1}{2}$ in a Magnetic Field

Nothing from Section 2.1.1 will be tested on quizzes or exams!

The first example involves the physical concept of *spin*.

What Spin Is. Many particles (electrons, protons, some nuclei) behave as if each carries a built-in, permanent tiny bar magnet. This built-in magnetism is called *spin*. Spin is as basic to a particle as its mass or electric charge. It's not something added later; the particle simply has it. In a magnetic field, this tiny magnet cannot take on any arbitrary orientation with any value; only specific outcomes appear when you measure[1] along a chosen direction. For the simplest case (spin-$\frac{1}{2}$, like the electron), measurement along a given axis yields one of two results: "up" or "down" relative to that axis.

What Spin Is Not. The particle is not a tiny ball physically spinning in space. Modeling the electron as a spinning sphere leads to contradictions (like surface speeds faster than light). "Spin" is a quantum property that *acts like* angular momentum and magnetism.

Spin is neither a literal rotation nor a hidden arrow you can simply peek at. We draw an arrow for intuition, but measuring along different directions yields outcomes that no single pre-existing classical arrow can fully explain. That is the quantum part.

How We Know Spin Is Real:

- **Stern-Gerlach experiment.** A beam of atoms sent through a non-uniform magnetic field splits into two distinct spots, not a smear. This matches the idea of a tiny magnet with only two allowed outcomes along the field direction.
- **Magnetic resonance.** In a magnetic field, applying an oscillating, sideways field at just the right frequency flips spins in a controlled way. This is the basis of NMR (nuclear magnetic resonance) and MRI.

How Spin Interacts with Magnetic Fields

- **Energy preference.** In a magnetic field, "aligned with the field" and "against the field" have slightly different energies. The difference grows with stronger fields.
- **A natural resonance frequency.** That energy difference corresponds to a natural rhythm (Larmor frequency). Driving the system with a sideways oscillating magnetic field at this frequency moves the spin between the two orientations.

Measuring Spin. A measurement along a chosen axis yields one of the allowed discrete results (for spin-$\frac{1}{2}$: "up" or "down"). Importantly, note that *measure along a different axis and the outcomes and their probabilities change.*

[1] Technically, we have not yet introduced the concept of "measurement" in a formal way; we will do so in a later chapter. For now, you can think of "measuring" intuitively as taking some action to detect the current value of a particle's spin.

2.1.2 Example 2: Two-Level Atom (or "Artificial Atom")

Nothing from Section 2.1.2 will be tested on quizzes or exams!

What a Two-Level Atom Is. Many physical systems (real atoms, trapped ions, and "artificial atoms" such as superconducting circuits) have many possible energy levels. In practice, we often choose *two* of these—a lower-energy *ground* level and a higher-energy *excited* level—and arrange our experiment so that everything we do mainly affects just this pair.

In that regime, the atom effectively behaves like a system with two options, much like a switch that can be "ground" or "excited." The energy gap between these two levels sets a specific color of light (for optical transitions) or a specific microwave frequency (for artificial atoms) that most strongly interacts with the pair.

What a Two-Level Atom Is Not. It is not that the atom truly has only two levels; rather, we *select* two and isolate them from the rest by careful tuning. We avoid accidentally using light at other colors (or microwaves at other frequencies) that would drive unwanted transitions. So, "two-level" is a well-engineered approximation: extremely accurate when we respect the conditions that keep other levels quiet.

How We Know This Picture Works:

- **Resonant absorption and emission.** When we shine light at the right color (or apply the right microwave frequency), the system efficiently moves *population*[2] between ground and excited and can emit light back at that same color. Off-color light barely does anything by comparison.
- **Rabi flopping (controlled swapping).** If we turn on the right-frequency drive for a controlled time, the population sloshes back and forth between ground and excited in a regular, predictable way. Turning the drive off at the right moment lets us stop in ground, excited, or any desired mix of the two.

How the Two-Level Atom Interacts with Light (or Microwaves).

- **Energy gap sets the "right" frequency.** The energy difference between ground and excited corresponds to a specific frequency. Driving at this frequency most effectively couples the two states; driving far away from it has little effect.
- **Pulse duration and phase control the outcome.** Short bursts nudge population partway; longer bursts can complete a swap. Adjusting the timing and the *phase* of the drive changes *how* the two states are mixed, giving flexible control.

Measuring a Two-Level Atom. We often arrange things so the two states respond differently to our probe:

[2] In this context, the term "population" means how many atoms (or what fraction/probability for a single atom) are in each of the two states: ground vs. excited. For a single atom, "population" means the chance you would find it in ground or in excited if you measured it right now. For many identical atoms, it is literally how many are in each state.

- **Atoms/ions (optical readout).** A detection laser makes only one state glow ("bright"), while the other stays dark. Counting the glow tells us which state the atom was in.
- **Artificial atoms (circuit readout).** A nearby microwave resonator acts like a sensitive listener whose pitch shifts slightly depending on whether the system is in ground or excited; we infer the state from that shift.

Everyday Intuition and Classroom Analogy. Think of a playground swing. It naturally moves back and forth at its own rhythm. A gentle push at exactly that rhythm transfers energy efficiently, carrying the swing from low to high and back. In the same way, light (or microwaves) at the right frequency efficiently moves population between ground and excited. Pushing off-rhythm does much less.

The above intuitive analogy is useful for beginners. But we do feel obligated to highlight the following caveats:

- Two levels vs a continuum: A swing moves through a continuum of heights; a two-level atom has only two energy levels.
- Probability, not position: "Population" means probability (or fraction) in ground/excited, not a physical angle like a swing.
- Superposition and phase: The atom can be in a superposition with a phase; the swing has no true analog of quantum phase.
- "Measurement" matters: Measuring the atom changes its state; watching a swing does not affect anything.
- Off-resonance still works: Light slightly off the exact frequency can still drive the atom (less efficiently); the swing analogy suggests almost no effect.

2.2 The Math for a Single Qubit

From now on, we will set aside the physics and take a purely mathematical view of a qubit: it is simply a unit vector in $\mathbb{C}^2$, corresponding to some quantum physical system (like those discussed in Section 2.1) that we no longer need to consider.

Analogy to a classical bit. Classically, we think of a bit $b \in \{0, 1\}$ as something stored in a register within the CUP. Note that this is also a mathematical abstraction: b is logical data, and the register is some piece of computer hardware whose implementation details we can ignore.

A *qubit* is just a *quantum bit*. By the same analogy, it is logical data encoded as a unit vector $\mathbf{v} \in \mathbb{C}^2$, stored in a "quantum register." The "quantum register" is some quantum computer hardware whose implementation we do not need to specify (which is exactly the underlying 2D quantum physical system).

Ultimately, the only thing we care about is how the math for a bit (and now, a qubit) works: how to describe it and how to use it to perform computation.

How to specify a single qubit. It is a vector. So, do what we will do for vectors:

- Find a basis and write it as a linear combination of the basis vectors.

Recall that the linear algebra theory is invariant to the choice of basis. Let us start with our favorite basis for $\mathbb{C}^2$: the standard basis

$$\mathbf{e}_0 \coloneqq \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_1 \coloneqq \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

(Notice that computer scientists prefer that the subscript starts from 0.)

Using the standard basis, a qubit $\mathbf{v}$ can always be written as

$$\mathbf{v} = \alpha_0 \cdot \mathbf{e}_0 + \alpha_1 \cdot \mathbf{e}_1,$$

where $\alpha_0, \alpha_1 \in \mathbb{C}^2$ *and* $|\alpha_0|^2 + |\alpha_1|^2 = 1$. The coefficients α_b ($b \in \{0, 1\}$) are referred to as *amplitudes*.

Teaching Suggestions

Derive the $|\alpha_0|^2 + |\alpha_1|^2 = 1$ condition from the "unit" requirement

$$1 = \|\mathbf{v}\| \coloneqq \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{\mathbf{v}^\dagger \mathbf{v}}.$$

(Help students recall that the definition of inner product in $\mathbb{C}^n$ requires the *conjugate* transpose $\mathbf{v}^\dagger$.)
This condition is often known as the *normalization condition* for state vectors.

Note that if you choose a different basis, the value of the amplitudes will change. You have complete freedom to choose the basis. We typically prefer the standard basis for its mathematical convenience, but for certain calculations a different basis may be more convenient. **Always be aware of the basis you are working in!**

Data encoding by a single qubit. Note that we can use a classical bit to encode a binary bit of either 0 or 1, and do some computation over it.

We do similar things on a quantum computer. We use a qubit to encode data and will operate the qubit to finish some computation to our interest.

Note that there are infinitely many unit vectors in $\mathbb{C}^2$. This immediately leads to a strange thing:

- In contrast to a classical bit, it seems a qubit can encode an infinite amount of information?!

Yes, it does! However, this does not mean we have infinite encoding or computational power. Quantum mechanics imposes quite stringent restrictions on how we can operate on a qubit and how we can read the data encoded in it. We will learn these rules as the following postulates. For now, the main takeaway is:

- *A single qubit can, in principle, encode an infinite amount of information, but quantum mechanics restricts how we can access and use this unbounded resource in precise ways that we will study later!*

2.3 Quantum Systems of Higher Dimensions

We have already learned a great deal about two-dimensional (2D) quantum systems. Before moving on, it is helpful to recognize that higher-dimensional quantum systems also exist and are widely used.

While providing concrete physical examples would require additional physics background beyond the scope of this course, we can still develop the mathematics cleanly. In fact, such systems are naturally understood as straightforward generalizations of the 2D case to an arbitrary dimension parameter n:

- **State space.** An n-dimensional quantum system corresponds to the Hilbert space $\mathbb{C}^n$. That is, its (pure) state can be completely specified by a unit vector $\mathbf{v} \in \mathbb{C}^n$, and each unit vector in this space can in principle describe a possible state of the system under appropriate conditions (for instance, at different times or after different operations).
- **Remark on physical realizations.** Such n-dimensional systems do exist in the real world; however, they are not our focus here, so we will not provide specific examples.
- **Coordinates via a basis.** To describe the state of an n-dimensional system concretely, we choose a basis of $\mathbb{C}^n$ and express the state vector $\mathbf{v}$ as a linear combination of the basis vectors, exactly as we did in the 2D case.
- **A convenient choice: the standard basis.** A particularly convenient choice is the standard basis for $\mathbb{C}^n$:
$$\mathbf{e}_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_1 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad \mathbf{e}_{n-1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$
Here, $\mathbf{e}_j$ (for $j \in \{0, 1, \dots, n-1\}$) is the vector whose j-th entry is 1 and all other entries are 0. (Note that we follow the computer-science convention: the first entry of a vector is indexed as the 0-th entry.)
- **Expression of a unit vector:** In the standard basis, a unit vector $\mathbf{v} \in \mathbb{C}^n$ can always be written as
$$\mathbf{v} = \sum_{j=0}^{n-1} \alpha_j \cdot \mathbf{e}_j,$$
where $\alpha_0, \dots, \alpha_{n-1} \in \mathbb{C}^n$ satisfy the normalization condition $\sum_{j=0}^{n-1} |\alpha_j|^2 = 1$.

2.4 Dirac Notation

Dirac[3] introduced a concise and powerful notation that is extremely convenient for quantum calculations. This notation has become the standard in quantum mechanics and quantum computing, although it is not widely adopted by mathematicians. One reason is that the notation is tailored to Hilbert spaces, whereas mathematicians often work in more general linear-algebraic settings. In spaces that are not Hilbert (for example, where no inner product is defined), Dirac's notation is less natural or not directly applicable.

Fix a positive integer n and consider the space $\mathbb{C}^n$.

The "ket" for column vectors

[3] Paul A. M. Dirac (1902–1984) was a pioneering theoretical physicist and one of the founders of quantum mechanics and quantum field theory. He formulated the Dirac equation for relativistic electrons, predicted the existence of antimatter, and introduced the bra–ket (Dirac) notation that is standard in quantum mechanics. He shared the 1933 Nobel Prize in Physics.

We use the "ket" notation $|\psi\rangle$ to denote a **column vector** in $\mathbb{C}^n$. The symbol $|\psi\rangle$ is read as "ket ψ."

For example, in $\mathbb{C}^4$ we may assign

$$|\psi\rangle = \begin{bmatrix} 1 \\ 2 \\ 5 \\ 5 \end{bmatrix}.$$

This is merely a change of symbols: instead of the familiar $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 5 \\ 5 \end{bmatrix}$, we write $|\psi\rangle$. The label ψ in the notation $|\psi\rangle$ is a variable, just as v is in $\mathbf{v}$; you are free to choose any letter inside the ket. However, the "$|\cdot\rangle$" part indicates that the object is a vector, analogous to using boldface $\mathbf{v}$ to signal a vector.

Notation for an orthonormal basis

In $\mathbb{C}^n$, we denote an **orthonormal** basis by

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

This notation specifies only that these symbols represent some orthonormal basis; it does not, by itself, fix which orthonormal basis is intended. When you use this notation, you should state explicitly which concrete orthonormal basis the symbols refer to, to avoid ambiguity.

For example, in $\mathbb{C}^4$ you might take the standard basis

$$|0\rangle = \mathbf{e}_0 \coloneqq \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad |1\rangle = \mathbf{e}_1 \coloneqq \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad |2\rangle = \mathbf{e}_2 \coloneqq \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad |3\rangle = \mathbf{e}_3 \coloneqq \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Alternatively, you could also use the following orthonormal basis:

$$|0\rangle = \tfrac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad |1\rangle = \tfrac{1}{2}\begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \quad |2\rangle = \tfrac{1}{2}\begin{bmatrix} 1 \\ -1 \\ i \\ -i \end{bmatrix}, \quad |3\rangle = \tfrac{1}{2}\begin{bmatrix} 1 \\ -1 \\ -i \\ i \end{bmatrix}.$$

We emphasize again that $\{|0\rangle, \ldots, |n-1\rangle\}$ is just a naming convention for whichever orthonormal basis you choose. You have total freedom to associate it to a specific orthonormal basis.

In many textbooks and papers on quantum computing, if no basis is explicitly specified, the default convention is the standard basis:

$$|0\rangle = \mathbf{e}_0, \quad |1\rangle = \mathbf{e}_1, \ \ldots, \ |n-1\rangle = \mathbf{e}_{n-1}.$$

Once a basis is fixed, any vector $|\psi\rangle$ can be expressed as a linear combination

$$|\psi\rangle = \alpha_0 |0\rangle + \cdots + \alpha_{n-1} |n-1\rangle,$$

where the coefficients (amplitudes) $\alpha_0, \ldots, \alpha_{n-1} \in \mathbb{C}$.

The "bra" for conjugate transpose

The "bra" $\langle\psi|$ denotes the conjugate transpose of $|\psi\rangle$; it is a row vector. For example, if

$$|\psi\rangle = \begin{bmatrix} 1 \\ 2+3i \\ 5i \\ 7 \end{bmatrix},$$

then

$$\langle\psi| = \begin{bmatrix} 1 & 2-3i & -5i & 7 \end{bmatrix}.$$

Standard linear algebra in Dirac notation

Dirac notation is simply an alternative symbolic system for the same linear-algebraic objects. All computations follow the usual rules of linear algebra.

A few operations to emphasize:

- **Inner product:** $\langle\phi| \cdot |\psi\rangle$ is the inner product of $|\phi\rangle$ and $|\psi\rangle$. It is often abbreviated as $\langle\phi|\psi\rangle$.
- **Ket–bra products:** The operator $|u\rangle \cdot \langle v|$ (often abbreviated as $|u\rangle\langle v|$) is defined as the outer product of $|u\rangle$ and $\langle v|$, i.e., the matrix whose (i,j)-entry is $u_i\overline{v_j}$.
- **Matrix multiplication:** Matrix multiplication with kets/bras behaves as expected. For example, for any $|a\rangle, |b\rangle, |c\rangle \in \mathbb{C}^n$ and $\mathbf{M} \in \mathbb{C}^{n\times n}$, the following equations hold:

$$\begin{aligned} |a\rangle\langle b|\,|c\rangle &= \langle b|c\rangle\ |a\rangle\,, \\ \langle c|\,|a\rangle\langle b| &= \langle c|a\rangle\ \langle b|\,, \\ |a\rangle\langle b|\ \mathbf{M} &= |a\rangle\ (\langle b|\,\mathbf{M}), \\ \mathbf{M}\ |a\rangle\langle b| &= (\mathbf{M}\,|a\rangle)\ \langle b|\,. \end{aligned}$$

These identities are direct consequences of associativity of matrix multiplication and the definitions of inner and outer products.

2.4.1 Worked Examples: Calculations in Dirac Notation

Example 1: Computing a bra from a ket

Let

$$|\psi\rangle = \begin{bmatrix} 1 \\ 2+3i \\ -i \\ 0 \end{bmatrix}.$$

Find $\langle\psi|$.

Solution. Take the conjugate transpose (transpose and conjugate entries):

$$\langle\psi| = \begin{bmatrix} 1 & 2-3i & i & 0 \end{bmatrix}.$$

Example 2: Inner products and norms

Let

$$|\phi\rangle = \tfrac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \qquad |\psi\rangle = \tfrac{1}{2}\begin{bmatrix}1\\i\\-1\\-i\end{bmatrix}.$$

Compute $\langle\phi|\psi\rangle$ and $\| \, |\psi\rangle \, \|$.

Solution.

$$\langle\phi| = \tfrac{1}{2}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}, \quad \langle\phi|\psi\rangle = \tfrac{1}{4}\big(1\cdot 1 + 1\cdot i + 1\cdot(-1) + 1\cdot(-i)\big) = \tfrac{1}{4}(1+i-1-i) = 0.$$

Thus $|\phi\rangle$ and $|\psi\rangle$ are orthogonal.

For the norm,

$$\| \, |\psi\rangle \, \|^2 = \langle\psi|\psi\rangle = \tfrac{1}{4}\big(1 + |i|^2 + |-1|^2 + |-i|^2\big) = \tfrac{1}{4}(1+1+1+1) = 1,$$

so $\| \, |\psi\rangle \, \| = 1$.

Example 3: Expanding a vector in the computational basis

Let the (standard) orthonormal basis of $\mathbb{C}^4$ be $\{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}$ with $|0\rangle = \mathbf{e_0}, \ldots, |3\rangle = \mathbf{e_3}$. Given

$$|\psi\rangle = \begin{bmatrix}2\\-i\\0\\1\end{bmatrix},$$

write $|\psi\rangle$ as a linear combination of $|0\rangle, |1\rangle, |2\rangle, |3\rangle$.

Solution. Read off coordinates:

$$|\psi\rangle = 2\,|0\rangle - i\;|1\rangle + 0\cdot|2\rangle + 1\cdot|3\rangle = 2\,|0\rangle - i\;|1\rangle + |3\rangle\,.$$

Example 4: Outer products (rank-1 operators)

Let

$$|u\rangle = \tfrac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \qquad |v\rangle = \tfrac{1}{\sqrt{2}}\begin{bmatrix}1\\-i\end{bmatrix}.$$

Compute the matrix $|u\rangle\langle v|$ and its action on $|v\rangle$.

Solution.

$$\langle v| = \tfrac{1}{\sqrt{2}}\begin{bmatrix}1 & i\end{bmatrix}, \qquad |u\rangle\langle v| = \Big(\tfrac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}\Big)\Big(\tfrac{1}{\sqrt{2}}\begin{bmatrix}1 & i\end{bmatrix}\Big) = \tfrac{1}{2}\begin{bmatrix}1 & i\\1 & i\end{bmatrix}.$$

Action on $|v\rangle$:

$$|u\rangle\langle v|\,|v\rangle = \langle v|v\rangle\;|u\rangle = 1\cdot|u\rangle = |u\rangle\,,$$

since $|v\rangle$ is normalized.

Example 5: Projectors and their properties

Fix a $|a\rangle \in \mathbb{C}^n$. Define a matrix $P \coloneqq |a\rangle\langle a|$. Such a P corresponds to the projection operation on the "direction"[4] specified by $|a\rangle$.

Show that $P^2 = P$ and $P^\dagger = P$.

Solution. First,

$$P^2 = |a\rangle\langle a|\,|a\rangle\langle a| = |a\rangle \underbrace{\langle a|a\rangle}_{=1} \langle a| = |a\rangle\langle a| = P.$$

Second,

$$P^\dagger = (|a\rangle\langle a|)^\dagger = |a\rangle\langle a|\,,$$

since $(|\cdot\rangle)^\dagger = \langle\cdot|$ and $(\langle\cdot|)^\dagger = |\cdot\rangle$.

Example 6: Orthogonal decomposition with a projector

Let $|a\rangle$ be a unit vector and $P = |a\rangle\langle a|$. For any $|\psi\rangle$, decompose $|\psi\rangle$ into components parallel and orthogonal to $|a\rangle$. That is, write $|\psi\rangle$ as a linear combination of two vectors $|\psi\rangle_{\mathsf{para}}$ and $|\psi\rangle_{\mathsf{orth}}$ such that $|\psi\rangle_{\mathsf{para}}$ is in parallel with $|a\rangle$ and $|\psi\rangle_{\mathsf{orth}}$ is orthogonal to $|a\rangle$.

Solution. The parallel component is

$$|\psi\rangle_{\mathsf{para}} = P\,|\psi\rangle = |a\rangle\langle a|\,|\psi\rangle = \langle a|\psi\rangle\,|a\rangle\,.$$

The orthogonal component is

$$|\psi\rangle_{\mathsf{orth}} = (I - P)\,|\psi\rangle\,.$$

Example 7: Change of basis

In $\mathbb{C}^n$, let $\{|f_0\rangle, \ldots, |f_{n-1}\rangle\}$ be a set of orthonormal basis. Let U be a matrix whose columns are $|f_0\rangle, \ldots, |f_{n-1}\rangle$ in order. Then, for any vector $|\psi\rangle \in \mathbb{C}^n$, its coordinates in the $\{|f_k\rangle\}$ basis are exactly the elements of the vector

$$|\widetilde{\psi}\rangle = U^\dagger\,|\psi\rangle\,.$$

That is, $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_{k=0}^{n-1} \widetilde{\psi}_k\,|f_k\rangle\,,$$

where $\widetilde{\psi}_k$'s are elements of the vector $|\widetilde{\psi}\rangle = \begin{bmatrix} \widetilde{\psi}_0 \\ \widetilde{\psi}_1 \\ \vdots \\ \widetilde{\psi}_{n-1} \end{bmatrix}$.

[4] This is exactly the subspace $\mathsf{span}\{|a\rangle\} \subset \mathbb{C}^n$.

Worked example. Consider the following orthonormal basis of $\mathbb{C}^4$

$$|f_1\rangle = \tfrac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \quad |f_2\rangle = \tfrac{1}{2}\begin{bmatrix}1\\i\\-1\\-i\end{bmatrix}, \quad |f_3\rangle = \tfrac{1}{2}\begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}, \quad |f_4\rangle = \tfrac{1}{2}\begin{bmatrix}1\\-i\\-1\\i\end{bmatrix},$$

and let U be a matrix whose columns are $|f_1\rangle, \ldots, |f_4\rangle$.

Let $|\psi\rangle = \begin{bmatrix}1\\0\\0\\0\end{bmatrix} = |0\rangle$. Then

$$\widetilde{\psi} = U^\dagger |0\rangle = \begin{bmatrix}\langle f_0|\\\langle f_1|\\\langle f_2|\\\langle f_3|\end{bmatrix} |0\rangle = \begin{bmatrix}\langle f_0|0\rangle\\\langle f_1|0\rangle\\\langle f_2|0\rangle\\\langle f_3|0\rangle\end{bmatrix} = \tfrac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}.$$

Hence $|0\rangle = \frac{1}{2}\sum_{k=0}^{3} |f_k\rangle$.

Example 8: Expectation value of an "observable"

Let

$$|\psi\rangle = \tfrac{1}{\sqrt{2}}\begin{bmatrix}1\\i\end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}.$$

We define a value $\mathbf{A}_\psi$ as follows:

$$\mathbf{A}_\psi = \langle\psi| \mathbf{A} |\psi\rangle .$$

This value is called the *expectation* of the *observable* $\mathbf{A}$ in the state $|\psi\rangle$. Do not worry if the terms "expectation" and "observable" are not yet familiar: they arise from the theory of measurements of quantum states. We will introduce these concepts in later chapters.

For now, let us just compute the value $\mathbf{A}_\psi$.

Solution.

$$\langle\psi| = \tfrac{1}{\sqrt{2}}\begin{bmatrix}1 & -i\end{bmatrix}, \quad \mathbf{A}|\psi\rangle = \tfrac{1}{\sqrt{2}}\begin{bmatrix}i\\1\end{bmatrix}.$$

Thus

$$\mathbf{A}_\psi = \tfrac{1}{2}\big(1 \cdot i + (-i) \cdot 1\big) = \tfrac{1}{2}(i - i) = 0.$$

Example 9: Matrix elements in a basis

Let $\{|0\rangle, |1\rangle\}$ be the standard basis of $\mathbb{C}^2$. Given the operator

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

compute its matrix elements $H_{ij} = \langle i| H |j\rangle$, and verify $H = \sum_{i,j\in\{0,1\}} H_{ij}\ |i\rangle\langle j|$.

Solution.

$$H_{00} = \langle 0| H |0\rangle = 1, \quad H_{01} = \langle 0| H |1\rangle = 0, \quad H_{10} = \langle 1| H |0\rangle = 0, \quad H_{11} = \langle 1| H |1\rangle = -1.$$

Therefore

$$H = |0\rangle\langle 0| - |1\rangle\langle 1| .$$

Example 10: Verifying orthonormality

Let

$$|f_1\rangle = \tfrac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \quad |f_2\rangle = \tfrac{1}{2}\begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}, \quad |f_3\rangle = \tfrac{1}{2}\begin{bmatrix}1\\-1\\i\\-i\end{bmatrix}, \quad |f_4\rangle = \tfrac{1}{2}\begin{bmatrix}1\\-1\\-i\\i\end{bmatrix}.$$

Show that $\{|f_k\rangle\}_{k=1}^4$ is an orthonormal basis.

Solution (sketch). Each vector has norm 1 since each has four entries of magnitude $1/2$. Pairwise inner products vanish by direct calculation; e.g.,

$$\langle f_3|f_4\rangle = \tfrac{1}{4}\big(1\cdot 1 + (-1)\cdot(-1) + i\cdot(-i) + (-i)\cdot i\big) = \tfrac{1}{4}(1+1+1-1) = 0.$$

Thus the set is orthonormal and spans $\mathbb{C}^4$.

Chapter 3
Postulate (2/4): State Evolution

Postulate 2: The evolution of a closed quantum system is described by a unitary operator (aka matrix). Notation-wise, $|\psi'\rangle = U|\psi\rangle$, where U is a unitary operator (or transformation or matrix).

Definition of Unitaries. Unitary matrices are square matrices U for which $U^\dagger U = I$. Equivalently, you can define them be requiring that $U^\dagger = U^{-1}$.

Consistency between Postulates 1 and 2. Note that unitary operators preserve length of vectors. If you have a unit vector $|\psi\rangle$, then it must be the case that $U|\psi\rangle$ is also a unit vector. This is consistent with Postulate 1 in the sense that transforming a valid quantum state by a unitary at least will still give you a valid quantum state (i.e., unit vector).

Teaching Suggestions

Prove that for any positive $n \in \mathbb{N}$, any unitary $U \in \mathbb{C}^{n\times n}$, and any $|\psi\rangle \in \mathbb{C}^n$, it holds that $\||\psi\rangle\| = \|U|\psi\rangle\|$.

Why unitaries? The use of unitary operators follows from certain physics reasons: it follows directly from Schrödinger's equation. We will not derive this here; you can simply treat is as an axiom put by the Nature. Interested readers can consult the discussion in [NC10, Section 2.2.2].

3.1 Commonly Used Unitaries

The Pauli operators: bit flip X, phase flip Z, and Y

The Pauli matrices[1] $\{X, Y, Z\}$ act on a single qubit (i.e., on $\mathbb{C}^2$):

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

Key properties:

- All three are unitary: $X^\dagger = X$, $Y^\dagger = Y$, $Z^\dagger = Z$, and $X^\dagger X = Y^\dagger Y = Z^\dagger Z = I$.
- X is the bit-flip: $X|0\rangle = |1\rangle$, $X|1\rangle = |0\rangle$.

[1] Named after the physicist Wolfgang Pauli (1900–1958), a pioneer of quantum mechanics best known for the Pauli exclusion principle. Pauli introduced these 2×2 matrices to describe spin-$\frac{1}{2}$ systems; they generate the Lie algebra $\mathsf{SU}(2)$ and are central to qubit dynamics. He was awarded the 1945 Nobel Prize in Physics for the discovery of the exclusion principle.

- Z is the phase-flip: $Z|0\rangle = |0\rangle$, $Z|1\rangle = -|1\rangle$.
- Y combines a bit and phase flip: $Y|0\rangle = i|1\rangle$, $Y|1\rangle = -i|0\rangle$.
- Eigenstructure:

$$Z|0\rangle = |0\rangle\,, \quad Z|1\rangle = -|1\rangle\,; \qquad X|\pm\rangle = \pm|\pm\rangle\,, \text{ where } |\pm\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle).$$

- Commutation and multiplication:

$$XZ = -ZX = iY, \quad YZ = -ZY = iX, \quad XY = -YX = iZ, \quad X^2 = Y^2 = Z^2 = I.$$

The Hadamard operator

The 2×2 Hadamard operator[2]

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

is unitary. It maps the *computational* basis (another name for standard basis) to the following *Hadamard* basis

$$|+\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \qquad |-\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

Operationally, H is an operator that helps you switch between computational basis and Hadamard basis:

$$\begin{cases} H|0\rangle = |+\rangle \\ H|1\rangle = |-\rangle \end{cases}, \qquad \begin{cases} H|+\rangle = |0\rangle \\ H|-\rangle = |1\rangle \end{cases}$$

Also, it is important to note that the Pauli Z is just the "flip operator" under Hadamard basis:

$$Z|+\rangle = |1\rangle\,, \qquad Z|-\rangle = |+\rangle\,.$$

The above could serve as an "operational" proof of the following equations, which you can of course verify using regular math.

$$HZH = X, \qquad HXH = Z.$$

Also note that $H^2 = I$.

An interesting unitary on $\mathbb{C}^4$: the 4-point Quantum Fourier Transform

Unitary matrices are not only of size 2×2. As a higher-dimension example, consider the so-called 4-point Quantum Fourier Transform (QFT) matrix F_4 defined by:

$$F_4 \coloneqq \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}.$$

[2] Named after the French mathematician Jacques Hadamard (1865–1963). Hadamard studied matrices with entries ±1 whose rows are mutually orthogonal—now called Hadamard matrices—which maximize determinant among ±1 matrices and appear in coding theory, design theory, and signal processing. In quantum computing, the 2×2 Hadamard matrix H creates and interconverts computational and $|\pm\rangle$ bases, serving as a basic "mixing" operation.

This unitary maps the computational basis

$$\{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}$$

to the so-called "Fourier basis"

$$\{|\widetilde{0}\rangle, |\widetilde{1}\rangle, |\widetilde{2}\rangle, |\widetilde{3}\rangle\}$$

that are defined as follows:

$$\forall k \in \{0,1,2,3\}, \quad |\widetilde{k}\rangle := F_4 |k\rangle = \frac{1}{2} \sum_{j=0}^{3} \omega^{jk} |j\rangle,$$

where $\omega = e^{\frac{2\pi i}{4}}$ is the principle 4-th root of unity.

It is unitary since $F_4^\dagger F_4 = I$, and it plays a central role in quantum algorithms (e.g., phase estimation, period finding). We will return to this unitary later in the course. For now, we use it simply to illustrate a unitary operator acting on vectors in $\mathbb{C}^n$ with $n > 2$.

There are many more unitaries

Beyond these, there are numerous commonly used unitaries: phase gates (S, T), rotations (R_x, R_y, R_z), two-qubit gates such as CNOT, CZ, the SWAP gate, controlled-phase gates, and more. We will introduce them as needed throughout the course rather than all at once.

Chapter 4
Postulate (3/4): Measurement (Born Rule)

4.1 What a Quantum Measurement Means (Physical Intuition)

In addition to applying unitaries, we can also perform *measurements* on quantum states. Postulate 3, which we are about to introduce, fully characterizes how quantum measurements operate.

But before giving the mathematical formulation of Postulate 3, it helps to build physical intuition for what "measurement" means in quantum mechanics.

A measurement is an interaction with a device

A quantum measurement is not just "looking" at a system; it is a controlled physical interaction between the system (e.g., the qubit) and a macroscopic apparatus (detector, photodiode, superconducting resonator, etc.). During this interaction:

- information about some property of the system (referred to an *observable*) is transferred to a classical readout (a voltage pulse, a bright/dark spot, a pointer position),
- the system is generally disturbed by the interaction,
- the outcome is intrinsically probabilistic, even when the apparatus is ideal and free of engineering errors.

From everyday measurements to the microscopic world

We already know how measurements work in everyday life:

- **Length of a pencil:** a ruler gives a number (e.g., 12.3 cm).
- **Mass of a ball:** a scale reads off a weight.
- **Speed of a car:** a radar gun displays the speed.

In each case, you use a device that interacts with the object and produces a classical readout (a number on a display or a mark on a ruler).

Key idea: We can do the same in the microscopic world. We build devices that interact with tiny systems (like single electrons or atoms) and produce classical outcomes we can record. The overall logic:

$$system + device \Rightarrow readout$$

is the same.

Examples for microscopic measurements

Let us revisit the single-qubit systems we have introduced as concrete microscopic examples for measurements:

- **Electron spin (Stern–Gerlach–type idea).** An electrons spin can be "up" or "down" along a chosen axis (our qubit basis). To measure it, we let the electron pass through a region where the spin affects the path (conceptually like a Stern–Gerlach magnet) and place detectors at the two possible exit spots. One detector click = "spin up," the other = "spin down." This is directly analogous to the ruler/scale/gun giving a definite readout. If we repeat the experiment on many identically prepared electrons, the fraction of "up" vs. "down" clicks stabilizes to well-defined probabilities.
- **Atomic energy level (bright/dark detection).** Consider an atom that can be in a ground state $|g\rangle$ or an excited state $|e\rangle$ (another qubit). Shine a laser that makes $|e\rangle$ fluoresce. If the atom is in $|e\rangle$, we see photons (bright); if it is in $|g\rangle$, we see none (dark). Bright vs. dark is the classical outcome shown by the detector, just like a scale showing a number. Repeating on many atoms prepared the same way yields stable bright/dark frequencies.

What's similar?

- There is always a *device* that interacts with the system and produces a *classical readout* (click/no click, bright/dark, 0/1).

What's different (specifically quantum)?

There are three major differences between classical measurement and quantum measurement:

1. **Intrinsic randomness:** Even with a perfect device, a single microscopic outcome can be random. The probabilities come from the quantum state, not from hidden imperfections of the apparatus.

   Imagine we prepare identical copies of the same system. A classical measurement would yield the same stable, deterministic outcome every time. In contrast, a quantum measurement can still produce probabilistic outcomes on individual trials (all drawn from the same distribution).

2. **State update ("collapse"):** After you read "spin up," the electron's state is now aligned with "up." After a bright detection, the atom is effectively in $|e\rangle$ (or quickly pumped into $|g\rangle$ depending on the scheme). The measurement changes what the system will do next.

3. **Choice of question matters:** Measuring spin along Z (up/down) answers a different question than measuring along X (left/right). Answering one precisely disturbs the other. This basis-dependence has no direct everyday analog for length, mass, or speed measured at one moment.

Summary

The above examples prepare us for the formal Postulate 3, also known as the Born rule[1]. It is a precise recipe that connects the quantum state, the chosen measurement (the "question"), and the probabilities of each classical outcome.

[1] Named after the German physicist Max Born (1882–1970), a founder of quantum mechanics. Born formulated the probabilistic interpretation of the wavefunction—now called the Born rule—identifying $|\psi(x)|^2$ (or, in finite dimensions, $|\langle\phi|\psi\rangle|^2$) as a probability density (or probability) for measurement outcomes. For this fundamental contribution to quantum theory, Born was awarded the 1954 Nobel Prize in Physics.

4.2 Postulate 3: the Formal Statement

Postulate 3: Measurement (Born Rule). Besides applying unitaries, we can perform *measurements* on quantum states.
Quantum measurements obey the following rules:

- **Measurement specification:** a quantum measurement is specified by a collection $\{M_m\}$ of *measurement operators* satisfying the *completeness* condition:

$$\sum_m M_m^\dagger M_m = I. \tag{4.1}$$

  These operators acting on the state space of the system being measured. The index m refers to the measurement outcomes that may occur in the experiment.

- **Measurement outcome:** if the state of the quantum system is $|\psi\rangle$ immediately before the measurement, then the probability that result m occurs is given by

$$\Pr[\mathsf{outcome} = m] = \|M_m|\psi\rangle\|^2 \quad \left(= \langle\psi|M_m^\dagger M_m|\psi\rangle\right) \tag{4.2}$$

- **Post-measurement state:** the state of the system after the measurement is

$$\frac{M_m|\psi\rangle}{\|M_m|\psi\rangle\|} \quad \left(= \frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}\right). \tag{4.3}$$

Remarks regarding this postulate:

- We emphasize that $\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}$ is just $\|M_m|\psi\rangle\|$, where $\|\cdot\|$ is the inner-product induced norm.

  Thus, the probability is the inner product between the vector $M_m|\psi\rangle$ and itself. Note that this probability must be positive. Think why.

- The completeness equation Equation (4.1) expresses the fact that probabilities sum to one. Prove it during the lecture.

- Also note that the post-measurement state is really a valid quantum state. That is, Expression (4.3) is consistent with Postulate 1, the condition for being a valid quantum state. Prove it during the lecture.

4.3 Example: Measuring a Single Qubit

We now illustrate the Born rule with concrete single-qubit measurements, written in the general $\{M_m\}$ (measurement-operator) framework we just learned.

4.3.1 Case 1: measurement in the computational (Z) basis

Teaching Suggestions

To make the lecture more interactive and engaging, the instructor can invite students to carry out the following analysis on the whiteboard, stepping in to help only when necessary.

Let the qubit be $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ with $|\alpha|^2 + |\beta|^2 = 1$. Specify the measurement by two operators

$$M_0 = |0\rangle\langle 0|, \qquad M_1 = |1\rangle\langle 1|.$$

They satisfy completeness:

$$M_0^\dagger M_0 + M_1^\dagger M_1 = |0\rangle\langle 0| + |1\rangle\langle 1| = I.$$

Born-rule probabilities:

$$\Pr[0] = \langle\psi| M_0^\dagger M_0 |\psi\rangle = |\alpha|^2, \qquad \Pr[1] = \langle\psi| M_1^\dagger M_1 |\psi\rangle = |\beta|^2.$$

Post-measurement states (conditioned on the outcome), where we use $|\psi_b\rangle$ to denote the post-measurement states corresponding to outcome b ($b \in \{0, 1\}$):

$$|\psi_0\rangle = \frac{M_0|\psi\rangle}{\|M_0|\psi\rangle\|} = |0\rangle, \qquad |\psi_1\rangle = \frac{M_1|\psi\rangle}{\|M_1|\psi\rangle\|} = |1\rangle.$$

In summary, the measurement results are stated as the following:

- with probability $|\alpha|^2$, we observe the measurement outcome 0, and the state collapses to $|0\rangle$;
- with probability $|\beta|^2$, we observe the measurement outcome 1, and the state collapses to $|1\rangle$;

Geometric picture in the 2D computational plane

For this simple measurement of a single qubit, we have a elegant geometric picture for it. See Figure 4.1.

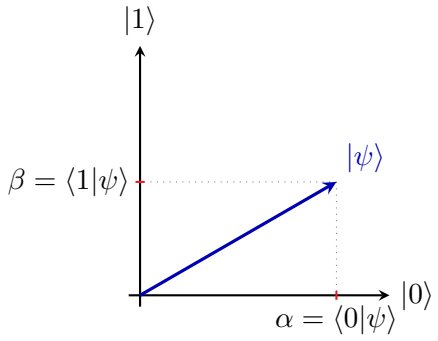


Figure 4.1: Geometric picture in the 2D space

Consider the 2D vector space spanned by $\{|0\rangle, |1\rangle\}$. Write the state as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \qquad |\alpha|^2 + |\beta|^2 = 1.$$

Think of $|0\rangle$ and $|1\rangle$ as two orthonormal axes in this plane. The number $\alpha = \langle 0|\psi\rangle$ is the component of $|\psi\rangle$ along the $|0\rangle$ axis, and $\beta = \langle 1|\psi\rangle$ is the component along the $|1\rangle$ axis. The measurement "collapses" onto one axis, with probabilities given by the squared lengths of these components:

$$\Pr[0] = |\alpha|^2 = |\langle 0|\psi\rangle|^2, \qquad \Pr[1] = |\beta|^2 = |\langle 1|\psi\rangle|^2.$$

If α and β are real and nonnegative, you can visualize the state vector forming an angle θ with the $|0\rangle$ axis so that $\alpha = \cos\theta$ and $\beta = \sin\theta$; then the probabilities are the squared projections onto each axis, $\cos^2\theta$ and $\sin^2\theta$.

In the general complex case, the same idea holds: you still project onto the two orthonormal axes and take the squared magnitudes of those projections to get the outcome probabilities. Note that in this case, it is tricky to define the "angle" between $|\psi\rangle$ and other vectors, say, $|0\rangle$. But it is always possible to define the notion of "projection" in the following way: Projecting a vector $\mathbf{v}$ onto another vector $\mathbf{u}$ leads to the following vector $\mathsf{Proj}_{\mathbf{u}}(\mathbf{v})$:

$$\mathsf{Porj}_{\mathbf{u}}(\mathbf{v}) \coloneqq \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\langle \mathbf{u}, \mathbf{u}\rangle}\mathbf{u}.$$

Therefore, Figure 4.1 is still applicable even if $|\psi\rangle$ contains complex amplitudes. The projection of $|\psi\rangle$ onto $|0\rangle$ is exactly $\langle 0|\psi\rangle|0\rangle = \alpha|0\rangle$.

On the naming

In the 2D picture above, this measurement is a projection of the state onto the orthonormal axes $\{|0\rangle, |1\rangle\}$. For this reason, it is called a measurement in the computational basis (or simply a computational measurement).

Furthermore, the Pauli Z operator has $|0\rangle$ and $|1\rangle$ as eigenstates (i.e., eigenvectors) with eigenvalues $+1$ and -1. Thus, measuring in $\{|0\rangle, |1\rangle\}$ is equivalently a projection to the eigenstates of Z. This is why it is also referred to as measurement in the Z basis.

Later in the course, we will explore in detail how measurements relate to the eigenstates of a special class of matrices called *observables*. For now, you need not worry about these details. Interested readers can consult Section 4.4.

4.3.2 Case 2: measurement in the Hadamard (X) basis

Let the qubit be $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ with $|\alpha|^2 + |\beta|^2 = 1$. Define the Hadamard-basis (X-basis) states

$$|+\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \qquad |-\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

Specify the measurement by two operators

$$M_+ = |+\rangle\langle +|, \qquad M_- = |-\rangle\langle -|.$$

They satisfy completeness:

$$M_+^\dagger M_+ + M_-^\dagger M_- = |+\rangle\langle +| + |-\rangle\langle -| = I.$$

Born-rule probabilities (expressing $|\psi\rangle$ in the $\{|+\rangle, |-\rangle\}$ basis):

$$\langle +|\psi\rangle = \tfrac{\alpha+\beta}{\sqrt{2}}, \qquad \langle -|\psi\rangle = \tfrac{\alpha-\beta}{\sqrt{2}},$$

$$\Pr[+] = \langle\psi| M_+ |\psi\rangle = \left|\tfrac{\alpha+\beta}{\sqrt{2}}\right|^2, \qquad \Pr[-] = \langle\psi| M_- |\psi\rangle = \left|\tfrac{\alpha-\beta}{\sqrt{2}}\right|^2.$$

Post-measurement states (conditioned on the outcome), using $|\psi_b\rangle$ for $b \in \{+,-\}$:

$$|\psi_+\rangle = \frac{M_+ |\psi\rangle}{\|M_+ |\psi\rangle\|} = |+\rangle, \qquad |\psi_-\rangle = \frac{M_- |\psi\rangle}{\|M_- |\psi\rangle\|} = |-\rangle.$$

In summary, the measurement results are:

- with probability $\left|\tfrac{\alpha+\beta}{\sqrt{2}}\right|^2$, we observe the outcome "+" and the state collapses to $|+\rangle$;
- with probability $\left|\tfrac{\alpha-\beta}{\sqrt{2}}\right|^2$, we observe the outcome "−" and the state collapses to $|-\rangle$.

Geometric picture in the same 2D plane

Remain in the 2D vector space spanned by $\{|0\rangle, |1\rangle\}$, but now view $\{|+\rangle, |-\rangle\}$ as an orthonormal pair of axes obtained by a 45° rotation of the computational axes. The components of $|\psi\rangle$ along these axes are $\langle +|\psi\rangle = (\alpha+\beta)/\sqrt{2}$ and $\langle -|\psi\rangle = (\alpha-\beta)/\sqrt{2}$. The measurement collapses onto one of these rotated axes with probabilities given by the squared magnitudes of the corresponding projections:

$$\Pr[+] = |\langle +|\psi\rangle|^2, \qquad \Pr[-] = |\langle -|\psi\rangle|^2.$$

On the naming

This measurement is in the Hadamard or X basis because the Pauli X operator has $|+\rangle$ and $|-\rangle$ as eigenstates with eigenvalues $+1$ and -1: $X|+\rangle = +|+\rangle$, $X|-\rangle = -|-\rangle$. Equivalently, applying a Hadamard gate H maps the computational basis to the X basis ($H|0\rangle = |+\rangle$, $H|1\rangle = |-\rangle$), so "measure in X" is the same procedure as "apply H, then measure in Z".

4.4 *From Matrices to Measurable Properties: why Z represents an observable

Nothing from Section 4.4 will be tested on quizzes or exams!

This is our first concrete example in which a matrix—the Pauli Z—stands for a physical property that one can actually measure. Here is the conceptual bridge:

- **States as vectors.** A qubit's state is a unit vector $|\psi\rangle$ in a two-dimensional complex vector space spanned by $\{|0\rangle, |1\rangle\}$.
- **Physical questions as operators.** A *yes/no* or *which-of-two* question about the system (e.g., "is the spin up or down along the Z direction?" or "is the qubit in $|0\rangle$ or $|1\rangle$?") is

encoded by a Hermitian operator A acting on that same space. Hermitian matrices have real eigenvalues and an orthonormal eigenbasis—exactly what we need to define possible outcomes and the states we collapse to.

- **Eigenvalues = possible outcomes; eigenvectors = post-measurement states.** For a Hermitian operator A, the spectral theorem says

$$A = \sum_m a_m \Pi_m, \qquad \Pi_m = |a_m\rangle\langle a_m| ,$$

where a_m are real eigenvalues (the *numbers* you can read on a device) and Π_m are projectors onto the corresponding eigenvectors $|a_m\rangle$ (the *states* you end up in after getting outcome a_m).

- **Born rule in observable language.** If the system is $|\psi\rangle$, then

$$\Pr[\text{outcome } a_m] \;=\; \langle\psi|\,\Pi_m\,|\psi\rangle \quad \text{and} \quad |\psi_{a_m}\rangle \;=\; \frac{\Pi_m\,|\psi\rangle}{\sqrt{\langle\psi|\,\Pi_m\,|\psi\rangle}}.$$

The operator A thus determines both the *set of outcomes* and the *state update rule*.

Specializing to the Pauli Z

The Pauli Z matrix is

$$Z \;=\; \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

with eigenpairs

$$Z\,|0\rangle = (+1)\,|0\rangle , \qquad Z\,|1\rangle = (-1)\,|1\rangle .$$

Thus:

- The *possible outcomes* are the real numbers $+1$ and -1.
- The *associated projectors* are $\Pi_+ = |0\rangle\langle 0|$ and $\Pi_- = |1\rangle\langle 1|$.
- Measuring the observable Z is the same operational procedure as measuring in the $\{|0\rangle , |1\rangle\}$ basis: outcome $+1$ corresponds to reporting "$|0\rangle$", and outcome -1 corresponds to "$|1\rangle$".

Why this is physically meaningful

A matrix like Z is not merely a computational device; it *labels a concrete experimental question*:

- It specifies which directions in state space are the "definite answers" (its eigenvectors).
- It determines what numbers are read out on the meter (its eigenvalues).
- Through the projectors $\{\Pi_m\}$, it fixes the Born-rule probabilities and the post-measurement states, connecting linear algebra to experimental statistics and dynamics.

Computational vs. physical labels

In quantum computing, we often report the outcome as a classical bit $0/1$. In the observable viewpoint, the same experiment can be labeled by $+1/-1$ (the eigenvalues of Z). These are two notational skins over the same process:

$$\text{“bit} = 0\text{”} \;\equiv\; \text{“eigenvalue } +1 \text{ of } Z\text{”} \;\equiv\; \text{project onto } |0\rangle\,,$$

$$\text{“bit} = 1\text{”} \;\equiv\; \text{“eigenvalue } -1 \text{ of } Z\text{”} \;\equiv\; \text{project onto } |1\rangle\,.$$

Seeing Z as an observable crystallizes the idea that *matrices encode measurable questions*, with their eigenstructures defining what can be observed and how the state changes when we observe it.

Chapter 5

Basic Quantum-Exclusive Effects (1/2): Elitzur-Vaidman Bomb

With the first three postulates, we can already begin discussing some effects that are unique to quantum computing (i.e., they do not appear in classical computing).

In particular, we will discuss the Elitzur-Vaidman bomb tester.

You will explore the Elitzur-Vaidman Bomb problem, a fascinating protocol that demonstrates how quantum mechanics enables the detection of a bomb without triggering it. Your goal is to derive and understand the steps in this protocol using the concepts of quantum rotation, measurement, and probability.

The Setup

Imagine you are in an airport where there is an unattended suitcase that might contain a bomb. More precisely, the suitcase behaves as follows. It has an input wire that allows you to send a bit $b \in \{0, 1\}$ into the suitcase. Then there are two possibilities:

1. **An innocent suitcase:** If there is no bomb inside, the suitcase simply forwards your input bit b to the output wire and returns the value b to you.
2. **A malicious suitcase:** Inside the suitcase, the input wire is connected to a meter that detects the value of your input b:
   - If it detects $b = 0$, the suitcase again forwards your input bit b to the output wire and returns b to you.
   - If it detects $b = 1$, the suitcase triggers the bomb and explodes. Game over.

In classical physics, there is no safe way to test whether this suitcase contains a bomb, because any attempt to interact with it risks triggering an explosion.

However, quantum mechanics provides a surprising solution: by leveraging the principles of superposition, interference, and measurement, it is possible to detect the presence of the bomb with a very high probability, while minimizing the risk of triggering it.

The Quantum Game

Now, suppose that instead of a classical bit, we are allowed to use a qubit, $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$, to represent our decision to query; and in the case of a malicious suitcase, the meter inside is a "quantum" meter that can measure your input qubit.

The behavior of the suitcase in this quantum version is as follows: you can send any qubit $|\psi\rangle$ of your choice into the suitcase, and then:

1. **An innocent suitcase:** If there is no bomb inside, the suitcase simply forwards your input qubit $|\psi\rangle$ to the output wire, returning $|\psi\rangle$ to you.

2. **An malicious suitcase:** the quantum meter will measure your qubit $|\psi\rangle$ in the computational basis $\{|0\rangle, |1\rangle\}$:
   - If the measurement outcome is $|0\rangle$, nothing happens, and the post-measurement state (i.e., $|0\rangle$) is returned to you.
   - If the measurement outcome is $|1\rangle$, the bomb will explode immediately. Game over.

Teaching Suggestions

Let us first make sure that we understand the rules. Assume that you send the query qubit

$$|\psi\rangle = \frac{1}{\sqrt{3}}|0\rangle + \frac{\sqrt{2}}{\sqrt{3}}|1\rangle$$

into the suitcase. Based on the above rules of the quantum Elitzur-Vaidman game, what will happen in each scenario? What is the probability of each possible outcome? What will the query qubit $|\psi\rangle$ look like at the end of each case?

The Quantum Solution

Now, we present in Algo. 5.0.1 a clever way to "quantumly interrogate" the suitcase, so that we have control of the probability of explosion if there is a bomb inside. The algorithm makes use of the rotation matrix

$$R_\varepsilon = \begin{bmatrix} \cos\varepsilon & -\sin\varepsilon \\ \sin\varepsilon & \cos\varepsilon \end{bmatrix}.$$

Convince yourself that:

- this matrix represents a counterclockwise rotation by angle ε in the x–y plane;
- it is unitary. (We need this condition because we will apply it to quantum states in Algo. 5.0.1.)

Algorithm 5.0.1: The Elitzur-Vaidman Algorithm

Parameters. Fix a small real number $\varepsilon \in (0, 1]$. (Concretely, you can think that $\varepsilon = 0.000001$.) Let $T := \lceil \frac{\pi}{2\varepsilon} \rceil$.

Algorithm. Initialize $|\psi_0\rangle = |0\rangle$, then repeat the following procedure for T times, where for each $i \in \{1, 2, \ldots, T\}$

- Update our state as $|\psi_i\rangle = R_\varepsilon |\psi_{i-1}\rangle$;
- Query the suitcase using $|\psi_i\rangle$.

At the end (if the suitcase has not exploded so far), measure $|\psi_T\rangle$ in the computational basis. Claim "no bomb" if the outcome is $|1\rangle$.

Analyze Algo. 5.0.1. Recall that we set $T = \frac{\pi}{2\varepsilon}$.

- If there is no bomb, final state would be $|\psi_T\rangle = \cos(T\varepsilon) \cdot |0\rangle + \sin(T\varepsilon) \cdot |1\rangle$. It is then easy to see that, by $T = \frac{\pi}{2\varepsilon}$, we detect the case of "no bomb" with probability 1.

- If there is a bomb, each iteration explodes with probability $\sin^2(\varepsilon)$. We claims that using a probabilistic technique called the *union bound*, we can prove that we are in bad luck (i.e., the bomb explodes while we are running the algorithm) with probability at most

$$\Pr[\text{the bomb explodes}] \leq T \cdot \sin^2(\varepsilon) \approx T \cdot \varepsilon^2 = \frac{\pi}{2}\varepsilon. \tag{5.1}$$

  This step utilizes the union bound as well as the fact that $\sin^2(\varepsilon) \approx \varepsilon^2$. We defer a formal treatment of these mathematical tools to Sections 5.2 and 5.3. Since Section 5.2 is a basic calculus fact, it is out of the scope of this course and will not be tested on quizzes or exams. However, note that Section 5.3 may appear in quizzes and/or exams.

  We have control over ε. By decreasing its value, we can satisfy any desired risk tolerance (at the cost of increasing the number of iterations in Algo. 5.0.1).

To give you a concrete numeric sense: note that $\frac{\pi}{2} \approx 1.57$. Thus, if we set $\varepsilon = 0.0001$, then we need to run Algo. 5.0.1. for $T = 15,700$ iterations, which yields a "bad luck" probability

$$\Pr[\text{the bomb explodes}] \approx 0.000157.$$

5.1 A Geometric Perspective

We present a minimal, rotation-based picture that captures the essence of the Elitzur–Vaidman protocol using repeated small rotations and intermediate measurements.

Case 1: no bomb. If there is no bomb, no intermediate measurement occurs. Applying T identical small rotations accumulates to a finite rotation:

$$|\psi_T\rangle = R_\varepsilon^T |0\rangle = R_{(T\varepsilon)} |0\rangle = \cos(T\varepsilon) |0\rangle + \sin(T\varepsilon) |1\rangle .$$

By our parameter setting in Algo. 5.0.1

$$|\psi_T\rangle = |1\rangle .$$

Therefore, with no bomb, the state is smoothly rotated from $|0\rangle$ to $|1\rangle$ by coherent accumulation of many small rotations. See Figure 5.1 for an illustration.

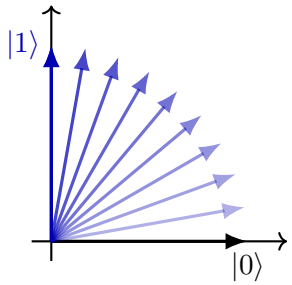


Figure 5.1: No bomb: coherent accumulation of small rotations

Case 2: a bomb is present. If a live bomb is present, each iteration consists of:

1. a small rotation R_ε that slightly moves amplitude toward $|1\rangle$,

2. an immediate measurement in the $\{|0\rangle, |1\rangle\}$ basis performed by the "bomb check."
    - Explosion event: outcome $|1\rangle$, with probability approximately $\sin^2 \varepsilon \approx \varepsilon^2$ at that step.
    - Survival (no explosion): outcome $|0\rangle$, which collapses the state back to $|0\rangle$, erasing the small tilt created by R_ε.

This procedure is illustrated by Figure 5.2.

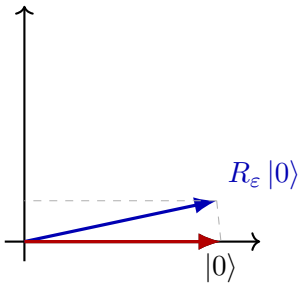


Figure 5.2: Live-bomb case: small tilt + measurement keeps resetting to $|0\rangle$

5.2 Establishing Inequality (5.1): $\sin^2(\varepsilon) \approx \varepsilon^2$

Nothing from Section 5.2 will be tested on quizzes or exams!

We claim that for small ε, the following approximation holds:

$$\sin^2(\varepsilon) \approx \varepsilon^2$$

This is a standard small-angle approximation used in physics and applied mathematics. To understand why it is valid, we begin by recalling the Taylor expansion of the sine function around $\varepsilon = 0$ (aka the Maclaurin series):

$$\sin(\varepsilon) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} \varepsilon^{2n+1} = \varepsilon - \frac{\varepsilon^3}{6} + \frac{\varepsilon^5}{120} - \cdots$$

We are interested in the square of this series:

$$\sin^2(\varepsilon) = \left(\varepsilon - \frac{\varepsilon^3}{6} + \frac{\varepsilon^5}{120} - \cdots\right)^2$$

Squaring a power series: the Cauchy product. To correctly square an infinite power series, we cannot simply square each term. Instead, we must use the *Cauchy product* of power series. Given two power series:

$$f(x) = \sum_{n=0}^{\infty} a_n x^n, \quad g(x) = \sum_{n=0}^{\infty} b_n x^n,$$

their product is defined as:

$$f(x)g(x) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_k b_{n-k} \right) x^n$$

This applies in our case where $f(x) = g(x) = \sin(\varepsilon)$, and we are computing $\sin^2(\varepsilon)$ as a power series.

In our case, $f(\varepsilon) = g(\varepsilon) = \sin(\varepsilon)$, so:

$$\sin^2(\varepsilon) = \sum_{n=0}^{\infty} c_n \varepsilon^n, \quad \text{where } c_n = \sum_{k=0}^{n} a_k a_{n-k}$$

Compute first few terms of $\sin^2(\varepsilon)$. We now compute the first few c_n explicitly using the coefficients of $\sin(\varepsilon)$:

$$\begin{aligned} a_1 &= 1, \\ a_3 &= -\frac{1}{6}, \\ a_5 &= \frac{1}{120}, \\ \text{All even } a_n &= 0. \end{aligned}$$

We compute the coefficients of ε^2, ε^4, and ε^6 in $\sin^2(\varepsilon)$:

- Coefficient of ε^2 $(n = 2)$:

$$c_2 = \sum_{k=0}^{2} a_k a_{2-k} = a_1 a_1 = 1$$

- Coefficient of ε^4 $(n = 4)$:

$$c_4 = \sum_{k=0}^{4} a_k a_{4-k} = a_1 a_3 + a_3 a_1 = 2 \cdot (1) \cdot \left(-\frac{1}{6} \right) = -\frac{1}{3}$$

- Coefficient of ε^6 $(n = 6)$:

$$c_6 = a_1 a_5 + a_3 a_3 + a_5 a_1 = 1 \cdot \frac{1}{120} + \left(-\frac{1}{6} \right)^2 + \frac{1}{120} \cdot 1 = \frac{1}{120} + \frac{1}{36} + \frac{1}{120} = \frac{1}{60} + \frac{1}{36} = \frac{3}{180} + \frac{5}{180} = \frac{2}{45}$$

Thus, we obtain the expansion:

$$\sin^2(\varepsilon) = \varepsilon^2 - \frac{1}{3}\varepsilon^4 + \frac{2}{45}\varepsilon^6 + O(\varepsilon^6)$$

Conclusion. For small values of ε, the higher-order terms (ε^4, ε^6, etc.) become negligible compared to ε^2. Therefore, we can approximate:

$$\sin^2(\varepsilon) \approx \varepsilon^2 \quad \text{for } \varepsilon \ll 1$$

This is the justification for the common small-angle approximation used in physics and engineering.

5.3 Establishing Inequality (5.1): Union Bound

The *union bound* (also known as Boole's inequality) is a fundamental concept in probability theory that provides an upper bound on the probability of a union of events. It states that the probability that at least one of several events occurs is at most the sum of the probabilities of the individual events.

Formally, if $A_1, A_2, \ldots, A_n$ are events, then:

$$\Pr\left[\bigcup_{k=1}^{n} A_k\right] \leq \sum_{i=1}^{n} \Pr[A_k].$$

Note that this inequality holds even if the events $\{A_1, \ldots, A_n\}$ are *not* mutually independent (i.e., there exist j and k such that $\Pr[A_j \cap A_k] \neq \Pr[A_j] \cdot \Pr[A_k]$). It is very useful in situations where calculating the exact probability of the union of events is difficult, but the individual probabilities of the events are easier to compute.

The union bound is often used in fields like computer science, statistics, and machine learning to provide worst-case guarantees. For example, it is used in analyzing algorithms, bounding error probabilities, or studying rare events.

In Elitzur-Vaidman Analysis

Recall that in the Elitzur-Vaidman algorithm discussed in our lecture, if there is a bomb inside the suitcase, it will be triggered with probability $\sin^2(\varepsilon)$ during each iteration.

We can use the union bound to derive a lower bound for the probability that none of the T iterations actually triggers the bomb. To do that, simply let A_k ($k \in \{1, \ldots, T\}$) denote the even that the explode bomb in the i-th iteration of Algo. 5.0.1. As we learned during the lecture, it holds that

$$\Pr[A_k] \leq \sin^2(\varepsilon),$$

where note that the "$<$" part in "$\leq$" is due to that possibility that there was no bomb in the suitcase.

Then,

$$\Pr[\text{the bomb explodes}] = \Pr\left[\bigcup_{k=1}^{T} A_k\right] \leq \sum_{k=1}^{T} \Pr[A_k] \leq T \cdot \sin^2(\varepsilon).$$

On More Example: Random Graphs

Next, let us see one more example for the application of the union bound.

Random graphs are widely used when analyzing social networks, wireless networks, and the internet. A simple model for random graphs is the Erdös-Rényi model $G(n, p)$. In this model, we have n nodes in the graph, and every pair of nodes are connected by an edge with probability p. The occurrence of each edge in the graph is independent from other edges in the graph. An *isolated node* is a node that is not connected to any other nodes in the graph. Let E_n be the event that a graph randomly generated according to the $G(n, p)$ model has at least one isolated node.

Question. Prove that if we set $p = \frac{(1+\varepsilon)\ln n}{n}$ with $0 < \varepsilon \leq 1$ being a constant, then it holds that

$$\lim_{n\to\infty} \Pr[E_n] = 0.$$

It is worth noting that this result reveals an interesting fact: a random graph is unlikely to contain isolated vertices, when p is only moderately large, in the asymptotic sense.

Solution. We first claim that $\Pr[E_n] \leq n(1-p)^{n-1}$. In the following, we show this inequality via the union bound.

Let A_i be the event that node i is *isolated*, i.e., it has no edges connecting it to any other node in the graph. Then:

$$E_n = \bigcup_{i=1}^{n} A_i$$

Using the union bound:

$$\Pr[E_n] = \Pr\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \Pr[A_i]$$

Now, we compute $\Pr[A_i]$ for a fixed node i.

In the $G(n,p)$ model, each of the $n-1$ possible edges from node i to the other nodes exists independently with probability p. Therefore, node i is isolated if *none* of these $n-1$ edges exist. So:

$$\Pr[A_i] = (1-p)^{n-1}$$

Hence:

$$\Pr[E_n] \leq \sum_{i=1}^{n} (1-p)^{n-1} = n(1-p)^{n-1}$$

Next, we show that $\lim_{n\to\infty} \Pr(E_n) = 0$ when $p = \frac{(1+\varepsilon)\ln n}{n}$ with $0 < \varepsilon \leq 1$:

$$\lim_{n\to\infty} \Pr(E_n) \leq \lim_{n\to\infty} n(1-p)^n = \lim_{n\to\infty} n\left[1 - \frac{(1+\varepsilon)\ln n}{n}\right]^n = \lim_{n\to\infty} n\left[1 - \frac{1+\varepsilon}{\frac{n}{\ln n}}\right]^n$$

$$= \lim_{n\to\infty} n\left(\left[1 - \frac{1+\varepsilon}{\frac{n}{\ln n}}\right]^{\frac{n}{\ln n}}\right)^{\ln n} = \lim_{n\to\infty} ne^{-(1+\varepsilon)\ln n} = \lim_{n\to\infty} \frac{1}{n^{\varepsilon}} = 0.$$

Note that the above derivation of the limit is not totally rigorous. In particular, I used the limit $\lim_{n\to\infty}(1-1/n)^n = e^{-1}$ in the middle of calculation, which should not be allowed. Note that only if $\lim f(x)$ and $\lim g(x)$ exist simultaneously can one do "limit by part" when computing $\lim(f(x)g(x))$. The rigorous way to solve this question would be using Taylor expansion and to track the order of infinitesimal. But this is too technical and not our focus. So, we do not present the full details.

Chapter 6

Postulate (4/4): Composition of Quantum Systems

6.1 Statement of Postulate 4

We now arrive at the final postulate of quantum mechanics, which explains how to combine separate quantum systems into a single, larger system. This postulate is indispensable: without it, we would lack a mathematical description of multi-part systems (e.g., two qubits, an atom plus a field mode, or a register of many qubits), and we could not even formalize the simple idea of "looking at two systems at once."

In more detail, suppose we have two quantum systems, A and B, prepared and manipulated independently. Physically, nothing prevents us from considering them *together* as a single composite entity AB—for instance, by placing both on the same lab bench or by routing both optical modes into the same apparatus. To reason about preparations, measurements, correlations, and dynamics of the joint setup, we need a precise mathematical framework that:

1. represents the state space of the combined system in terms of the state spaces of the parts;
2. represents product (independently prepared) states and allows for more general states that capture correlations between the parts;
3. lifts local (unitary) transformations and measurements on A and B to well-defined transformations and measurements on the composite system AB;
4. ensures consistency with the previous postulates we have already introduced.

The statement. Postulate 4 is precisely formulated to meet these requirements, utilizing the mathematical notion of *tensor product*. Its formal statement is:

> **Postulate 4:** The state space of a composite physical system is the *tensor product* of the state spaces of the component physical systems.

On tensor product. To fully understand **Postulate 4**, we need to conduct a formal treatment of the notion "tensor product."

My approach to defining the tensor product differs from most resources, such as [NC10] the many other lecture notes available on-line. Many of them did not provide a mathematically rigorous definition of the tensor product. However, offering a rigorous definition requires more effort, such as viewing vectors as mappings over indices and discussing isomorphisms in detail. One resource that I have found that does this rigorously is the CS 766/QIC 820: Theory of Quantum Information (Fall 2011) lecture notes by John Watrous.

Here is the approach we will take. We will define tensor products only in the special case of finite-dimensional complex Hilbert spaces $\mathbb{C}^n$ (with possibly different dimension parameters n),

and we will only realize the "tensor" operation via the Kronecker product. Of course, this is just one concrete instantiation among many possible ways to construct tensor products (even among spaces of the form $\mathbb{C}^n$). By restricting attention to this setting, we sacrifice the generality of tensor products. However, this finite-dimensional, Kronecker-based construction is the one that overwhelmingly covers the use cases in quantum computing, and it is the only tensor product we will need throughout this course.

6.2 Tensor Products among $\mathbb{C}^n$'s

We now provide the formal treatment of tensor products.

Organization: we will

1. in Section 6.2.1, introduce a matrix operation called Kronecker product;
2. in Section 6.2.2, show how Kronecker product help us to "tensor" Hilbert spaces of the form $\mathbb{C}^n$, to form larger Hilbert space, which is called the tensor product space;

6.2.1 Kronecker Product

Definition 6.2.1 (Kronecker[1] Product)**.** *If* $\mathbf{A}$ *is an* $m \times n$ *matrix and* $\mathbf{B}$ *is a* $p \times q$ *matrix over* $\mathbb{C}$*, then the* Kronecker product $\mathbf{A} \otimes \mathbf{B}$ *is the* $pm \times qn$ *block matrix:*

$$\mathbf{A} \otimes \mathbf{B} \coloneqq \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}.$$

◊

Best illustrated by examples:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \otimes \begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} = \begin{bmatrix} 1\begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} & 2\begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} \\ 3\begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} & 4\begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ 0 & 15 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{bmatrix}.$$

The properties of Kronecker product: Following are most useful properties of Kronecker product. They are by means exhaustive though.

1. **Mixed-Product Property.** Given matrices A, B, C and D such that the dimension of them allow us to form the matrix products AC and BD, then it holds that
$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD).$$
2. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ and $(A \otimes B)^{\dagger} = A^{\dagger} \otimes B^{\dagger}$.
3. **Non-commutativity.** In general, $A \otimes B \neq B \otimes A$.

[1] Named after Leopold Kronecker (1823-1891), a German mathematician known for foundational work in algebra and arithmetic.

4. $\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\,\mathrm{tr}(B)$. (Note that $\mathrm{tr}(AB) \neq \mathrm{tr}(A)\,\mathrm{tr}(B)$)

5. **"As-you-expected" Properties:**

$$\begin{aligned}
\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) &= \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C}, \\
(\mathbf{B} + \mathbf{C}) \otimes \mathbf{A} &= \mathbf{B} \otimes \mathbf{A} + \mathbf{C} \otimes \mathbf{A}, \\
(k\mathbf{A}) \otimes \mathbf{B} &= \mathbf{A} \otimes (k\mathbf{B}) = k(\mathbf{A} \otimes \mathbf{B}), \\
(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} &= \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}), \\
\mathbf{A} \otimes \mathbf{0} &= \mathbf{0} \otimes \mathbf{A} = \mathbf{0},
\end{aligned}$$

You are required to memorize all of these properties!

6.2.2 Hilbert Space formed via Kronecker Product

Recall that both $\mathbb{C}^n$ and $\mathbb{C}^m$ form vector spaces over the field $\mathbb{C}$. Let $\mathbb{C}^n \otimes \mathbb{C}^m$ be the set where elements are

$$\mathbb{C}^n \otimes \mathbb{C}^m := \{\mathbf{v} \otimes \mathbf{w} \mid \mathbf{v} \in \mathbb{C}^n, \mathbf{w} \in \mathbb{C}^m\},$$

where we emphasize that:

1. the "$\otimes$" between $\mathbb{C}^n$ and $\mathbb{C}^m$ is just a symbol;
2. however, the "$\otimes$" between $\mathbf{v}$ and $\mathbf{w}$ is the Kronecker product defined above.

We next switch to Dirac notation from now on, so that

$$\mathbb{C}^n \otimes \mathbb{C}^m := \{|v\rangle \otimes |w\rangle \mid |v\rangle \in \mathbb{C}^n, |w\rangle \in \mathbb{C}^m\}.$$

The following theorem is easy to verify:

Theorem 6.2.1 (Vector Space via Kronecker Product)**.** *The set $\mathbb{C}^n \otimes \mathbb{C}^m$ form a Hilbert space over $\mathbb{C}$, under the following vector addition, scalar multiplication, and inner product:*

1. *$|v_1\rangle \otimes |w_1\rangle + |v_2\rangle \otimes |w_2\rangle$ is the natural vector addition.*
2. *$a \cdot (|v\rangle \otimes |w\rangle) = (a \cdot |v\rangle) \otimes |w\rangle = |v\rangle \otimes (a \cdot |w\rangle)$, where $a \in \mathbb{C}$.*
3. *The inner product between $|v_1\rangle \otimes |w_1\rangle$ and $|v_2\rangle \otimes |w_2\rangle$ is defined as:*

$$\begin{aligned}
(|v_1\rangle \otimes |w_1\rangle)^\dagger (|v_2\rangle \otimes |w_2\rangle) &= (\langle v_1| \otimes \langle w_1|)(|v_2\rangle \otimes |w_2\rangle) \\
&= \langle v_1|v_2\rangle \otimes \langle w_1|w_2\rangle \\
&= \langle v_1|v_2\rangle \cdot \langle w_1|w_2\rangle
\end{aligned}$$

we usually write $|v\rangle \otimes |w\rangle$ as $|v, w\rangle$ or simply $|vw\rangle$.

This Hilber space $\mathbb{C}^n \otimes \mathbb{C}^m$ is referred to as the tensor product space of $\mathbb{C}^n$ and $\mathbb{C}^m$. ◊

Note that Theorem 6.2.1 ensures that **Postulate 4** is consistent with **Postulate 1**, as the "tensor composition" leads to a Hilbert space again.

Technically, when the symbol $\otimes$ is used as a matrix operation, as in $\mathbf{A} \otimes \mathbf{B}$ or $|\psi\rangle \otimes |\phi\rangle$, it is more precise to call it the *Kronecker product* (of matrices or vectors). When the symbol is used between vector spaces, as in $\mathbb{C}^n \otimes \mathbb{C}^m$ in Theorem 6.2.1, it denotes the *tensor product* (of spaces). In other words, the term "tensor product" should properly be reserved for the construction on spaces, not for matrix operations; this distinction reflects standard mathematical usage.
However, in quantum physics and quantum computing, it is conventional to use the term "tensor product" for both contexts, even where "Kronecker product" would be the strictly accurate name for arrays. This is a community convention that we will follow throughout these notes.

6.3 Structures of Tensor Product Space

Basis and Dimension

The following lemma is important for understanding the structure of tensor product spaces.

Lemma 6.3.1. *Let $\{|d_j\rangle\}_{j=0}^{n-1}$ be a basis of $\mathbb{C}^n$. Let $\{|e_k\rangle\}_{j=0}^{m-1}$ be a basis of $\mathbb{C}^m$. Then it holds that $\{|d_j\rangle \otimes |e_k\rangle\}_{j,k}$ form a basis of the tensor product space $\mathbb{C}^n \otimes \mathbb{C}^m$.* $\Diamond$

The proof of Lemma 6.3.1 is left for you as an exercise.

Lemma 6.3.1 implies that the vector space $\mathbb{C}^n \otimes \mathbb{C}^m$ has dimension nm. We claim (without a formal proof) that we can naturally identify $\mathbb{C}^n \otimes \mathbb{C}^m$ as $\mathbb{C}^{nm}$. This can be seen by consider the simple case where $\{|d_j\rangle\}_{j=0}^{n-1}$ and $\{|e_k\rangle\}_{j=0}^{m-1}$ are simply standard basis for $\mathbb{C}^n$ and $\mathbb{C}^m$—in this case, $\{|d_j\rangle \otimes |e_k\rangle\}_{j,k}$ is exactly the standard basis for $\mathbb{C}^{nm}$.

Therefore, one can say that $\mathbb{C}^n \otimes \mathbb{C}^m$ and $\mathbb{C}^{nm}$ are just the same vector space up to a notation change—"renaming" the basis. This effect has a formal mathematical definition called "isomorphism," denoted as $\mathbb{C}^n \otimes \mathbb{C}^m \cong \mathbb{C}^{nm}$. We provide a formal treatment of this concept in Section 6.3.1 as optional reading for interested students.

A low-dimensional example: $\mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathbb{C}^4$. Let $\{|0\rangle, |1\rangle\}$ be the standard basis of $\mathbb{C}^2$. Then

$$\left\{ |0\rangle \otimes |0\rangle,\ |0\rangle \otimes |1\rangle,\ |1\rangle \otimes |0\rangle,\ |1\rangle \otimes |1\rangle \right\}$$

is a basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$. Note that these are exactly the standard basis $\{|e_0\rangle, |e_1\rangle, |e_2\rangle, |e_3\rangle\}$ of $\mathbb{C}^4$ via

$$|0\rangle|0\rangle = \begin{bmatrix}1\\0\\0\\0\end{bmatrix} = |e_0\rangle, \quad |0\rangle|1\rangle = \begin{bmatrix}0\\1\\0\\0\end{bmatrix} = |e_1\rangle, \quad |1\rangle|0\rangle = \begin{bmatrix}0\\0\\1\\0\end{bmatrix} = |e_2\rangle, \quad |1\rangle|1\rangle = \begin{bmatrix}0\\0\\0\\1\end{bmatrix} = |e_3\rangle,$$

Hence any state $a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle$ corresponds to the vector $(a, b, c, d)^\top \in \mathbb{C}^4$.

6.3.1 The Isomorphism $\mathbb{C}^n \otimes \mathbb{C}^m \cong \mathbb{C}^{nm}$

Nothing from Section 6.3.1 will be tested on quizzes or exams!

In this section we explain what it means for two vector spaces to be *isomorphic* and then prove that the tensor product space $\mathbb{C}^n \otimes \mathbb{C}^m$ is (canonically) isomorphic to the ordinary vector space $\mathbb{C}^{nm}$. We also give several equivalent descriptions of the same isomorphism that are useful in quantum computing.

Definition 6.3.1 (Linear isomorphism)**.** *Let V and W be complex vector spaces. A map $T : V \to W$ is a* linear isomorphism *if:*

1. *T is linear: $T(\alpha v + \beta w) = \alpha T(v) + \beta T(w)$ for all $v, w \in V$ and $\alpha, \beta \in \mathbb{C}$;*
2. *T is bijective (one-to-one and onto).*

If such a T exists, we say V and W are isomorphic *and write $V \cong W$.* ◊

Intuitively, $V \cong W$ means V and W are "the same" as vector spaces: after a (linear) change of coordinates, their algebraic structure agrees. In finite dimensions, linear isomorphism is equivalent to equality of dimensions:

Lemma 6.3.2. *Two finite-dimensional vector spaces are isomorphic iff they have the same dimension.* ◊

This lemma, together with Lemma 6.3.1, implies that $\mathbb{C}^n \otimes \mathbb{C}^m \cong \mathbb{C}^{nm}$.

Concrete Example: $\mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathbb{C}^4$

Fix the computational basis $\{|0\rangle, |1\rangle\}$ for each copy of $\mathbb{C}^2$. The product basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$ is

$$\mathcal{B}_\otimes = \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}.$$

Let $\{|f_1\rangle, |f_2\rangle, |f_3\rangle, |f_4\rangle\}$ be an arbitrary orthonormal basis of $\mathbb{C}^4$. Define the linear isomorphism $F : \mathbb{C}^2 \otimes \mathbb{C}^2 \to \mathbb{C}^4$ by

$$F(|00\rangle) = |f_1\rangle, \quad F(|01\rangle) = |f_2\rangle, \quad F(|10\rangle) = |f_3\rangle, \quad F(|11\rangle) = |f_4\rangle,$$

and extend F linearly.

Take concrete vectors in $\mathbb{C}^2 \otimes \mathbb{C}^2$:

$$|\psi\rangle = (1+i)\,|00\rangle + 2\,|01\rangle - i\,|10\rangle + 3i\,|11\rangle,$$

$$|\varphi\rangle = (-2+i)\,|00\rangle + (1-i)\,|01\rangle + 4\,|10\rangle + (-3+i)\,|11\rangle.$$

Under F, their images in $\mathbb{C}^4$ (in the $\{|f_i\rangle\}$ coordinates) are

$$F(|\psi\rangle) = \begin{bmatrix} 1+i \\ 2 \\ -i \\ 3i \end{bmatrix}_{\{|f_i\rangle\}}, \qquad F(|\varphi\rangle) = \begin{bmatrix} -2+i \\ 1-i \\ 4 \\ -3+i \end{bmatrix}_{\{|f_i\rangle\}}. \tag{6.1}$$

(1) Addition and scalar multiplication. Let $\alpha = 2 - i$ and $\beta = 1 + i$. Compute in $\mathbb{C}^2 \otimes \mathbb{C}^2$:

$$\alpha \left|\psi\right\rangle + \beta \left|\varphi\right\rangle = \sum_{b \in \{00,01,10,11\}} c_b \left|b\right\rangle,$$

with coefficients

$$\begin{aligned}
c_{00} &= (2-i)(1+i) + (1+i)(-2+i) = (3+i) + (-3-i) = 0. \\
c_{01} &= (2-i)\cdot 2 + (1+i)\cdot(1-i) = 4 - 2i + \left(1 - i^2\right) = 4 - 2i + 2 = 6 - 2i. \\
c_{10} &= (2-i)\cdot(-i) + (1+i)\cdot 4 = (-2i + i^2) + (4 + 4i) = (-2i - 1) + 4 + 4i = 3 + 2i. \\
c_{11} &= (2-i)\cdot 3i + (1+i)\cdot(-3+i) = (6i - 3i^2) + \left(-3 + i - 3i + i^2\right) \\
&= (6i + 3) + \left(-3 - 2i - 1\right) = (3 - 3 - 1) + (6i - 2i) = -1 + 4i.
\end{aligned}$$

Therefore

$$\alpha \left|\psi\right\rangle + \beta \left|\varphi\right\rangle = 0 \cdot \left|00\right\rangle + (6 - 2i)\left|01\right\rangle + (3 + 2i)\left|10\right\rangle + (-1 + 4i)\left|11\right\rangle.$$

Apply F:

$$F(\alpha \left|\psi\right\rangle + \beta \left|\varphi\right\rangle) = \begin{bmatrix} 0 \\ 6 - 2i \\ 3 + 2i \\ -1 + 4i \end{bmatrix}_{\{|f_i\rangle\}}.$$

Compute the same entirely in $\mathbb{C}^4$:

$$\alpha F(|\psi\rangle) + \beta F(|\varphi\rangle) = (2-i)\begin{bmatrix} 1+i \\ 2 \\ -i \\ 3i \end{bmatrix} + (1+i)\begin{bmatrix} -2+i \\ 1-i \\ 4 \\ -3+i \end{bmatrix} = \begin{bmatrix} 0 \\ 6-2i \\ 3+2i \\ -1+4i \end{bmatrix}_{\{|f_i\rangle\}},$$

which matches exactly.

(2) Inner product. Compute the inner product in $\mathbb{C}^2 \otimes \mathbb{C}^2$ (conjugate-linear in the first slot):

$$\langle \psi, \varphi \rangle = \overline{(1+i)}(-2+i) + \overline{2}(1-i) + \overline{(-i)}\,4 + \overline{3i}\,(-3+i).$$

Term-by-term:

$$\begin{aligned}
\overline{1+i}(-2+i) &= (1-i)(-2+i) = -2 + i + 2i - i^2 = -1 + 3i, \\
\overline{2}(1-i) &= 2(1-i) = 2 - 2i, \\
\overline{-i}\cdot 4 &= (i)\cdot 4 = 4i, \\
\overline{3i}\,(-3+i) &= (-3i)(-3+i) = 9i - 3i^2 = 9i + 3.
\end{aligned}$$

Sum:

$$\langle \psi, \varphi \rangle = (-1 + 3i) + (2 - 2i) + 4i + (3 + 9i) = (-1 + 2 + 3) + (3i - 2i + 4i + 9i) = 4 + 14i.$$

Now compute in $\mathbb{C}^4$ using coordinates (with the same convention):

$$\langle F(\psi), F(\varphi)\rangle = \overline{1+i}\,(-2+i) + \overline{2}\,(1-i) + \overline{-i}\,4 + \overline{3i}\,(-3+i) = 4+14i,$$

agreement as expected. (Note that in this calculation, we are using the expressions of the transformed vectors in Equation (6.1) and the fact that the $\{|f_j\rangle\}$ basis is orthonormal.)

6.3.2 More Examples to Build Intuition

We present more concrete, low-dimensional examples. Throughout, we use the computational (standard) bases unless otherwise stated.

Example 0: Tensor product of orthonormal vectors

Note that the tensor product of two unit vectors is again a unit vector. This is important to ensure that **Postulate 4** is consistent with **Postulate 1**.

This can be easily shown using the Kronecker product properties to calculate the inner product of $\||\psi\rangle \otimes |\phi\rangle\|$.

It helps to see a concrete example of $\mathbb{C}^2 \otimes \mathbb{C}^2$. Let

$$(\alpha_0 |0\rangle + \alpha_1 |1\rangle) \otimes (\beta_0 |0\rangle + \beta_1 |1\rangle) = \alpha_0\beta_0 |00\rangle + \alpha_0\beta_1 |01\rangle + \alpha_1\beta_0 |10\rangle + \alpha_1\beta_1 |11\rangle .$$

Given $|\alpha_0|^2 + |\alpha_1|^2 = 1$ and $|\beta_0|^2 + |\beta_1|^2 = 1$, it is easy to see that

$$|\alpha_0\beta_0|^2 + |\alpha_0\beta_1|^2 + |\alpha_1\beta_0|^2 + |\alpha_1\beta_1|^2 = 1.$$

Next, let $\{|d_j\rangle\}_{j=1}^n$ and $\{|e_k\rangle\}_{k=1}^m$ be orthonormal. Then the tensor product family $\{|d_j\rangle \otimes |e_k\rangle\}_{j,k}$ is also orthonormal:

$$\langle d_j \otimes e_k | d_{j'} \otimes e_{k'}\rangle = \langle d_j | d_{j'}\rangle \, \langle e_k | e_{k'}\rangle = \delta_{jj'}\delta_{kk'}, \quad \text{where} \quad \delta_{xy} \coloneqq \begin{cases} 1 & x = y \\ 0 & x \neq y \end{cases}.$$

Hence the nm product vectors are mutually orthonormal and therefore form an orthonormal basis of $\mathbb{C}^n \otimes \mathbb{C}^m$. This is a concrete example of lemma 6.3.1.

Example 1: Two qubits ($\mathbb{C}^2 \otimes \mathbb{C}^2$)

Let $\{|0\rangle, |1\rangle\}$ be a basis of $\mathbb{C}^2$ for both subsystems. Then the set

$$\{|0\rangle \otimes |0\rangle,\ |0\rangle \otimes |1\rangle,\ |1\rangle \otimes |0\rangle,\ |1\rangle \otimes |1\rangle\}$$

forms a basis of $\mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathbb{C}^4$. Any two-qubit state can be written as

$$|\psi\rangle = a_{00} |00\rangle + a_{01} |01\rangle + a_{10} |10\rangle + a_{11} |11\rangle ,$$

with complex coefficients a_{jk}. For instance,

$$|\psi\rangle = \tfrac{1}{2} |00\rangle + \tfrac{1}{2} |01\rangle + \tfrac{1}{2} |10\rangle + \tfrac{1}{2} |11\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \ \otimes \ \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle).$$

This shows a product state.

By contrast, the Bell state

$$\left|\Phi^+\right\rangle = \tfrac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

cannot be written as a single product $|\phi\rangle \otimes |\zeta\rangle$ (Think why); it is *entangled*. Nevertheless, it still expands in the basis $\{|j\rangle \otimes |k\rangle\}_{j,k\in\{0,1\}}$.

Example 2: Qubit–qutrit ($\mathbb{C}^2 \otimes \mathbb{C}^3$)

Let $\{|0\rangle, |1\rangle\}$ be a basis of $\mathbb{C}^2$ and $\{|0\rangle, |1\rangle, |2\rangle\}$ a basis of $\mathbb{C}^3$. Then the $2 \times 3 = 6$ vectors

$$\{|0\rangle \otimes |0\rangle,\ |0\rangle \otimes |1\rangle,\ |0\rangle \otimes |2\rangle,\ |1\rangle \otimes |0\rangle,\ |1\rangle \otimes |1\rangle,\ |1\rangle \otimes |2\rangle\}$$

form a basis of $\mathbb{C}^2 \otimes \mathbb{C}^3 \cong \mathbb{C}^6$. Any state admits the expansion

$$|\psi\rangle = \sum_{j\in\{0,1\}} \sum_{k\in\{0,1,2\}} a_{jk}\ |j\rangle \otimes |k\rangle .$$

Example 3: Changing the basis on one side

Keep the qubit standard basis $\{|0\rangle, |1\rangle\}$ for the left space, but use the Hadamard basis $\{|+\rangle, |-\rangle\}$ on the right, where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$. Then

$$\{|0\rangle \otimes |+\rangle,\ |0\rangle \otimes |-\rangle,\ |1\rangle \otimes |+\rangle,\ |1\rangle \otimes |-\rangle\}$$

is also a basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$. For any $|\psi\rangle = \sum_{j,k\in\{0,1\}} a_{jk}\,|j\rangle \otimes |k\rangle$ we can regroup terms as

$$|\psi\rangle = |\phi_+\rangle \otimes |+\rangle + |\phi_-\rangle \otimes |-\rangle ,$$

where

$$|\phi_\pm\rangle = \frac{1}{\sqrt{2}}\big((a_{00} \pm a_{01})\,|0\rangle + (a_{10} \pm a_{11})\,|1\rangle\big).$$

6.4 Bipartite Quantum System and Partial Measurement

In quantum mechanics, a *bipartite* quantum system is a quantum system that consists of two subsystems. The space $\mathbb{C}^n \otimes \mathbb{C}^m$ is a typical example. Physically, you can think it as a mathematical description of two "quantum" registers.

- Fix a quantum state (i.e., a unit vector) $|\psi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$. You can think $\mathbb{C}^n$ is the mathematical description of a physical "quantum" register, and $\mathbb{C}^m$ of another "quantum" registers. Then, $|\psi\rangle$ is a state that is "stored" (jointly) within the two "quantum" registers.

We call such a state $|\psi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$ an *bipartite* state because it is from the tensor product of two Hilbert spaces.

Bipartite states are fundamental in quantum information and many-body physics. They capture correlations and entanglement between subsystems. Postulate 4 provides a formal mathematical framework to characterize bipartite quantum systems.

6.4.1 Separability on Tensor Product Space

Since $\mathbb{C}^n \otimes \mathbb{C}^m$ is also a Hilbert space, its elements are vectors and linear operators on it is again matrices. All we know about linear algebra applies to this space.

Here is one thing about vectors and linear operators on this $\mathbb{C}^n \otimes \mathbb{C}^m$ that worth emphasizing, something we refer to as "separability."

Entanglement

For a bipartite state $|\psi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^m$, there are two possibilities:

- There exist two quantum states $|\psi_L\rangle \in \mathbb{C}^n$ and $|\psi_R\rangle \in \mathbb{C}^n$ such that $|\psi\rangle = |\psi_L\rangle \otimes |\psi_R\rangle$. Think why. In this case, we say that $|\psi\rangle$ is a separable state, or a (tensor) product state.
- There is no way to write $|\psi\rangle$ as the tensor product of two quantum states $|\psi_L\rangle \in \mathbb{C}^n$ and $|\psi_R\rangle \in \mathbb{C}^n$. In this case, we say that $|\psi\rangle$ is an entangled state. (We have already seen an example of entangled state: $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.)

Entanglement is a very important concept in quantum computing. It is a property unique to quantum information. We will discuss more about it in future lectures.

Separability of Linear Operators

Similar phenomenon appears for linear operators on $\mathbb{C}^n \otimes \mathbb{C}^m$ as well.

Let us focus only on unitary operators for now. Fix a unitary operator $U \in L(\mathbb{C}^n \otimes \mathbb{C}^m)$. Then, it might be possible (but not always true) that such an U can be decomposed as $U = U_1 \otimes U_2$ for some $U_1 \in L(\mathbb{C}^n)$ and $U_2 \in L(\mathbb{C}^m)$.

Teaching Suggestions

Show $X \otimes H$ as an example of a separable unitary operator.

First, let us see how this operator applies to a separable states:

$$(U_1 \otimes U_2)\,|v\rangle \otimes |w\rangle = U_1\,|v\rangle \otimes U_2\,|w\rangle\,.$$

Note that this simply follows from the property of Kronecker product. Intuitively, this represents the physical operation that we apply U_1 to the state stored in the first quantum register and U_2 to the state stored in the second registers, separately.

Some remarks:

1. Easy to see that $U_1 \otimes U_2$ is again a linear operator: Kronecker product results in a matrix; linear operators are nothing but matrices.
2. To be consistent with **Postulate 2** (state evolution), need to prove that $U_1 \otimes U_2$ is a unitary if both U_1 and U_2 are unitaries.
3. Note that the converse of the above is not true. Many unitary matrices do not have this

separable structure. Counter example:

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

This matrix is unitary. But it cannot be written as $U_1 \otimes U_2$. Think why.

In more detail, this matrix transforms the computational basis

$$\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$$

into the Bell basis

$$\{\left|\Phi^+\right\rangle, \left|\Psi^+\right\rangle, \left|\Psi^-\right\rangle, \left|\Phi^-\right\rangle\},$$

which consists of maximally entangled states:

$$\left|\Phi^+\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle),$$
$$\left|\Phi^-\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle),$$
$$\left|\Psi^+\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle),$$
$$\left|\Psi^-\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle).$$

6.4.2 Partial Unitaries

Consider the setting that we have a (separable) bipartite state $|\psi\rangle_{AB} = |\phi\rangle_A \otimes |\zeta\rangle_B \in \mathcal{H}_A \otimes \mathcal{H}_B$, and we want to apply a unitary U_A on the A system, without touching the B system.

Here are two point of views for the (same) operation:

- **Perspective 1:** Ignore system B temporarily. We are only applying U_A to $|\phi\rangle_A$. This will convert the A system to $U_A |\phi\rangle_A$. Since we did not touch B at all, it remains as $|\zeta\rangle_B$., Therefore, the overall state after applying U_A to A becomes $(U_A |\phi\rangle_A) |\zeta\rangle_B$.
- **Perspective 2:** We do not isolate the two systems. We always view them as a whole. Then, we need a proper mathematical framework that allows us to talk about "applying a unitary to (only) a part of the overall system". Fortunately, the $(A \otimes B) \cdot (C \otimes D) = (AB) \otimes (CD)$ property serves for this purpose—we can simply write the operator as $U_A \otimes I_B$ and apply it to the overall state $|\phi\rangle_A |\zeta\rangle_B$ which leads to the same output:

$$(U_A \otimes I_B) |\phi\rangle_A |\zeta\rangle_B = (U_A |\phi\rangle_A) |\zeta\rangle_B.$$

Indeed, using the notation in Perspective 2, we can do more general operations:

$$(U_A \otimes U_B) |\phi\rangle_A |\zeta\rangle_B = (U_A |\phi\rangle_A) \otimes (U_B |\zeta\rangle_B),$$

which means that we apply U_A to system A and U_B to system B.

It is worth mentioning that when $|\psi\rangle_{AB}$ is not separable (i.e., it is a entangled state), it is still valid to write the expression $(U_A \otimes U_B)|\psi\rangle_{AB}$ to express the meaning that we apply U_A to system A and U_B to system B. And mathematically everything is correct. But we may not be able to claim that the expression is equal to $(U_A|\phi\rangle_A) \otimes (U_B|\zeta\rangle_B)$ anymore—in this case, $|\phi\rangle_A$ and $|\zeta\rangle_B$ do not exist as $|\psi\rangle_{AB}$ is not separable.

6.4.3 Partial Measurements

Setup and notation

Let $\mathcal{H}_A \coloneqq \mathbb{C}^n$ and $\mathcal{H}_B \coloneqq \mathbb{C}^m$ be Hilbert spaces for two subsystems (registers) A and B. A bipartite pure state is a unit vector

$$|\psi\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B.$$

We consider measurements that act only on one register, say A, while leaving B untouched.

Born rule recap (global form)

Quantum measurements obey the following rules:

- **Measurement specification:** A measurement is given by a collection $\{M_m\}$ of measurement operators on the system being measured that satisfy the completeness relation

$$\sum_m M_m^\dagger M_m = I. \tag{6.2}$$

  The index m labels the possible measurement outcomes.

- **Measurement outcome:** If the pre-measurement state is $|\psi\rangle$, then

$$\Pr[\mathsf{outcome} = m] \;=\; \|M_m|\psi\rangle\|^2 = \langle\psi| M_m^\dagger M_m |\psi\rangle\,. \tag{6.3}$$

- **Post-measurement state:** Conditional on obtaining outcome m, the post-measurement state is

$$\frac{M_m|\psi\rangle}{\|M_m|\psi\rangle\|} \;=\; \frac{M_m|\psi\rangle}{\sqrt{\langle\psi| M_m^\dagger M_m|\psi\rangle}}\,. \tag{6.4}$$

Partial measurement on subsystem A

To measure only subsystem A while leaving B intact, we lift the measurement operators on A to the joint system by tensoring with the identity on B:

$$\{M_m \text{ on } \mathcal{H}_A\} \quad \rightsquigarrow \quad \{M_m \otimes I_B \text{ on } \mathcal{H}_A \otimes \mathcal{H}_B\}.$$

Because $\sum_m M_m^\dagger M_m = I_A$, we have completeness on AB:

$$\sum_m (M_m \otimes I_B)^\dagger (M_m \otimes I_B) = \sum_m (M_m^\dagger M_m) \otimes I_B = I_A \otimes I_B = I_{AB}.$$

Applying the Born rule to the joint state $|\psi\rangle_{AB}$:

$$\Pr[m] = \langle\psi|_{AB} \left((M_m^\dagger M_m) \otimes I_B\right) |\psi\rangle_{AB},$$
$$\left|\psi_{AB}^{(m)}\right\rangle = \frac{(M_m \otimes I_B)\,|\psi\rangle_{AB}}{\sqrt{\langle\psi|_{AB} \left((M_m^\dagger M_m) \otimes I_B\right) |\psi\rangle_{AB}}}.$$

This is precisely the Born rule instantiated for a partial (local) measurement.

Example EPR pair measured on one share

Consider the EPR pair on two qubits (A and B):

$$\left|\Phi^+\right\rangle_{AB} = \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big).$$

We examine measurements on A only, first in the computational basis $\{|0\rangle, |1\rangle\}$, then in the Hadamard (X) basis $\{|+\rangle, |-\rangle\}$, where

$$|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}}.$$

You will observe a surprising effect: we perform an operation on quantum register A without directly interacting with quantum register B, yet the state of register B appears to change in response. This is the power of entanglement—a uniquely quantum phenomenon with no classical counterpart.

Teaching Suggestions

To make the lecture more interactive and engaging, the instructor can invite students to carry out the following analysis on the whiteboard, stepping in to help only when necessary.

Measurement of A in the computational (Z) basis. Take $M_0 = |0\rangle\langle 0|$ and $M_1 = |1\rangle\langle 1|$ on A. The joint measurement operators are $M_0 \otimes I$ and $M_1 \otimes I$.

Outcome probabilities:

$$\Pr[0] = \left\langle\Phi^+\right| \left((|0\rangle\langle 0|) \otimes I\right) \left|\Phi^+\right\rangle = \frac{1}{2},$$
$$\Pr[1] = \left\langle\Phi^+\right| \left((|1\rangle\langle 1|) \otimes I\right) \left|\Phi^+\right\rangle = \frac{1}{2}.$$

Post-measurement joint states:

$$\left|\Phi^+\right\rangle_{AB}^{(0)} = \frac{\left((|0\rangle\langle 0|) \otimes I\right) |\Phi^+\rangle}{\sqrt{\Pr[0]}} = \frac{\frac{1}{\sqrt{2}}\,|00\rangle}{\sqrt{1/2}} = |00\rangle,$$
$$\left|\Phi^+\right\rangle_{AB}^{(1)} = \frac{\left((|1\rangle\langle 1|) \otimes I\right) |\Phi^+\rangle}{\sqrt{\Pr[1]}} = \frac{\frac{1}{\sqrt{2}}\,|11\rangle}{\sqrt{1/2}} = |11\rangle.$$

Measurement of A in the Hadamard (X) basis

Now take $M_+ = |+\rangle\langle+|$ and $M_- = |-\rangle\langle-|$ on A. (That is, we take the measurement $\{M_+ \otimes I_B, M_- \otimes I_B\}$ in the overall state consisting of both the A and B systems.)

To ease our calculation, we could first wirte $|\Phi^+\rangle$ in the X basis on A:

$$|0\rangle = \tfrac{1}{\sqrt{2}}(|+\rangle + |-\rangle), \qquad |1\rangle = \tfrac{1}{\sqrt{2}}(|+\rangle - |-\rangle),$$

so

$$\begin{aligned}\left|\Phi^+\right\rangle_{AB} &= \frac{1}{\sqrt{2}}\left(|0\rangle|0\rangle + |1\rangle|1\rangle\right) = \frac{1}{2}\Big[|+\rangle\left(|0\rangle + |1\rangle\right) + |-\rangle\left(|0\rangle - |1\rangle\right)\Big] \\ &= \frac{1}{\sqrt{2}}\Big(|+\rangle|+\rangle + |-\rangle|-\rangle\Big).\end{aligned}$$

Outcome probabilities:

$$\Pr[+] = \Pr[-] = \frac{1}{2}.$$

Post-measurement joint states:

$$\begin{aligned}\left|\Phi^+\right\rangle_{AB}^{(+)} &= \frac{\left((|+\rangle\langle+|) \otimes I\right)|\Phi^+\rangle}{\sqrt{1/2}} = |+\rangle|+\rangle, \\ \left|\Phi^+\right\rangle_{AB}^{(-)} &= \frac{\left((|-\rangle\langle-|) \otimes I\right)|\Phi^+\rangle}{\sqrt{1/2}} = |-\rangle|-\rangle.\end{aligned}$$

Consistency with the Born rule. In both bases, we used the measurement operators on A tensored with I_B and applied the global Born rule:

$$\Pr[m] \;=\; \langle\Phi^+|_{AB}\Big((M_m^\dagger M_m) \otimes I_B\Big)\left|\Phi^+\right\rangle_{AB}, \qquad \left|\Phi^+\right\rangle_{AB}^{(m)} \;=\; \frac{(M_m \otimes I_B)\,|\Phi^+\rangle_{AB}}{\sqrt{\Pr[m]}}.$$

The outcomes and post-measurement states computed above are exactly those predicted by these formulas. In particular, measuring one share of an EPR pair prepares the other share in the same eigenstate of the measured basis, with outcome probabilities equal to the squared amplitudes (here, uniform 1/2-1/2).

Chapter 7
Basic Quantum-Exclusive Effects (2/2)

We talk about the following effects in order:

1. no-cloning (Section 7.1),
2. quantum teleportation (Section 7.2),
3. superdense coding (Section 7.3)
4. EPR paradox and CHSH game (Section 7.4)

7.1 No-Cloning Theorem

In classical information, copying is both natural and ubiquitous. We routinely duplicate files, fan-out signals on a circuit, and back up data with bit-perfect fidelity.

- **Classical copying is cheap and exact.** A classical bit $b \in \{0,1\}$ can be copied deterministically. The operation $b \mapsto (b,b)$ is well-defined for all $b \in \{0,1\}$.
- **Known quantum states can be "re-prepared."** If the identity of a state is known, e.g., we know "the input is $\alpha|0\rangle + \beta|1\rangle$" (namely, we know the exact values α and β), then we can trivially create as many copies as desired by preparing $\alpha|0\rangle + \beta|1\rangle$ repeatedly. No-cloning does not forbid *re-preparation*; it forbids a *universal, state-independent copying operation* that works for arbitrary unknown inputs.

It is then natural to ask for a general quantum states:

- *Can we build a (universal) quantum algorithm which takes as input any quantum state $|\psi\rangle$ and produces two perfect copies $|\psi,\psi\rangle$?*

The *no-cloning theorem* answers this with a definitive "no" for arbitrary and unknown quantum states.

Teaching Suggestions

Explain the role of $|e\rangle_B$ in Theorem 7.1.1.

Theorem 7.1.1 (No-Cloning Theorem)**.** *Let $\mathcal{H} = \mathbb{C}^2$. There is no unitary operator U on $\mathcal{H} \otimes \mathcal{H}$ together with a state $|e\rangle_B \in \mathcal{H}$ that satisfies*

$$\forall |\psi\rangle_A \in \mathcal{H}, \quad U(|\psi\rangle_A |e\rangle_B) = |\psi\rangle_A |\psi\rangle_B .$$

◊

Proof. Assume for the sake of contradiction that we have a universal cloning machine U and a $|e\rangle_B$ such that for an arbitrary state $|\psi\rangle_A$,

$$U(|\psi\rangle_A|e\rangle_B) = |\psi\rangle_A|\psi\rangle_B. \tag{7.1}$$

Now, choose $|\psi\rangle_A = \frac{1}{\sqrt{2}}(|0\rangle_A + |1\rangle_A)$.

According to Equation (7.1), we have

$$U\left(\frac{1}{\sqrt{2}}(|0\rangle_A + |1\rangle_A)|e\rangle_B\right) = \frac{1}{\sqrt{2}}(|0\rangle_A + |1\rangle_A)\frac{1}{\sqrt{2}}(|0\rangle_B + |1\rangle_B). \tag{7.2}$$

On the other hand, by the linearity of U, we obtain

$$U\left(\frac{1}{\sqrt{2}}(|0\rangle_A + |1\rangle_A)|e\rangle_B\right) = \frac{1}{\sqrt{2}}U(|0\rangle_A|e\rangle_B) + \frac{1}{\sqrt{2}}U(|1\rangle_A|e\rangle_B) = \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B) + \frac{1}{\sqrt{2}}(|1\rangle_A|1\rangle_B). \tag{7.3}$$

However, Equation (7.2) is not equal to Equation (7.3), which leads to a contradiction. ■

7.2 Quantum Teleportation

Quantum teleportation is a protocol that transfers an *unknown* qubit state from Alice to Bob without moving the physical carrier (or the "quantum registers") of that state. It uses two ingredients:

- pre-shared entanglement between Alice and Bob, and
- two bits of classical communication from Alice to Bob.

Crucially, teleportation does not violate the no-cloning theorem: by the time Bob recovers the state, Alice's original qubit has been irreversibly measured and no longer contains the unknown state.

7.2.1 Preparation: CNOT gate and EPR pair

The CNOT operator. We will utilize a 4-by-4 unitary operator called CNOT. It is a 2-qubit operator. It has the following matrix representation

$$\mathsf{CNOT} \coloneqq \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

As usual, let us understand it by investigating its effect on computational basis:

$$\forall a, b \in \{0,1\}, \quad \mathsf{CNOT}\,|a,b\rangle = |a, a \oplus b\rangle\,.$$

- $|a\rangle$ is called the "control qubit" and $|b\rangle$ is called the "target qubit."
- Essentially, CNOT flips b iff a is 1.

The EPR pair. We will also use a special two-qubit entangled state known as the EPR pair, named after Albert Einstein, Boris Podolsky, and Nathan Rosen. In their 1935 paper "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" [EPR35],[1] EPR introduced a thought experiment intended to show that, under seemingly reasonable assumptions of locality and realism, quantum mechanics yields paradoxical correlations—suggesting to them that the theory might be incomplete and that "elements of reality" not captured by the wavefunction might exist. Their argument sparked a decades-long debate about the foundations of quantum theory.

Historically, EPR's challenge motivated John Bell to formulate Bell's theorem in 1964, which showed that no local hidden-variable theory can reproduce all quantum predictions. Subsequent experiments—beginning with Aspect et al. in the 1980s and culminating in modern loophole-free tests—have confirmed quantum non-local correlations. The EPR pair (also called a Bell state) is the canonical example of such correlations and underpins many quantum information protocols, including teleportation, superdense coding, and device-independent cryptography. We will return to these foundational issues in Section 7.4.

For now, we record the definition of the state:

$$\left|\Phi^+\right\rangle := \frac{|00\rangle + |11\rangle}{\sqrt{2}}.$$

It can be prepared from a product state by applying a Hadamard to the first qubit followed by a CNOT:

$$\mathsf{CNOT}(H \otimes I)\,|00\rangle = \mathsf{CNOT} \cdot \frac{|0\rangle + |1\rangle}{\sqrt{2}}\,|0\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}.$$

7.2.2 Protocol

- Alice has an unknown qubit $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$ that she wants Bob to possess.
- Alice and Bob already share an entangled EPR pair:

$$\left|\Phi^+\right\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}.$$

  Alice holds one half in her quantum register; Bob holds the other in his quantum register.

The teleportation protocol then works as follows:

1. Alice puts her $|\psi\rangle$ on wire 1, and puts her share of $|\Phi^+\rangle$ on wires 2.
2. Bot puts his share of $|\Phi^+\rangle$ on wires 3.
3. Alice performs a "Bell-basis measurement" on wires 1 and 2. Operationally, this can be implemented by CNOT (with control being wire 1 and target being wire 2), then H on wire 1, followed by computational-basis measurements on both wires 1 and 2. Let the outcomes be z (from wire 1) and x (from wire 2).
4. Alice sends the two classical bits (z, x) to Bob over a classical channel.
5. Bob applies the correction $X^x Z^z$ to wire 3. We claim that Bob will obtain $|\psi\rangle$. (We will prove it in Section 7.2.3.)

This procedure is illustrated as a "quantum circuit diagram" in Figure 7.1. See Supplementary Box 7.2.1 to learn how to read the diagram.

[1] Interestingly, according to Google Scholar, this is now Einstein's most-cited paper (as of January 21, 2026).

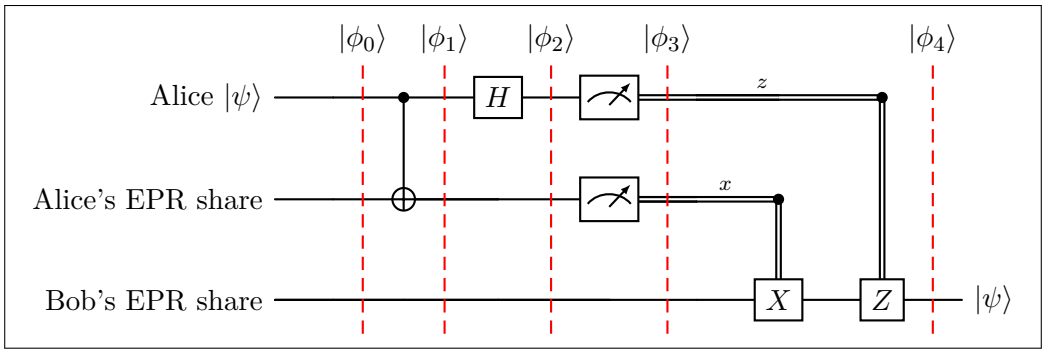


Figure 7.1: Quantum Circuit Diagram for Quantum Teleportation

Supplementary Box 7.2.1: How to Read Quantum Circuit Diagrams

We will give a formal treatment of the quantum circuit model later in **??**. For now, here is some basic background to help you read the circuit diagram in Figure 7.1.
A quantum circuit is a left-to-right picture of a computation. It usually consists of the following components:

Wires. Each single line represents a quantum system (a qubit in our course). Each double line represents a classical wire. Time flows left to right.

States at the left. Labels like $|\psi\rangle$ or $|0\rangle$ appearing at the left-most column indicate the initial state on a wire.

Single-qubit gates. Boxes labeled as H, X, and Z denote unitary operations on that wire.

Controlled gates. A control dot on one wire connected vertically to a gate symbol (often $\oplus$ for X) on another wire means "apply the target gate if the control qubit is $|1\rangle$." This is the controlled-X (CNOT). Controls can also target Z or other gates.

Measurement. A measurement is shown by a meter-like symbol with a double line (i.e., classical wire) carrying a classical bit to later boxes. Classical wires can control classically-conditioned operations, shown as gates with small arrows or labels like "X^x."

Barriers (optional). Vertical dashed lines sometimes mark logical stages; they have no operational effect.

7.2.3 Analysis

Teaching Suggestions

To make the lecture more interactive and engaging, the instructor can invite students to carry out the following analysis on the whiteboard, stepping in to help only when necessary.

We analyze the quantum teleportation protocol by calculating each of the intermediate states in Figure 7.1.

$$|\phi_0\rangle = |\psi\rangle \otimes \left|\Phi^+\right\rangle$$

$$
\begin{aligned}
&= (\alpha\,|0\rangle + \beta\,|1\rangle) \otimes \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big) \\
&= \frac{1}{\sqrt{2}}\big(\alpha\,|000\rangle + \alpha\,|011\rangle + \beta\,|100\rangle + \beta\,|111\rangle\big) \\
|\phi_1\rangle &= \mathsf{CNOT} \otimes I\,|\phi_0\rangle = \frac{1}{\sqrt{2}}\big(\alpha\,|000\rangle + \alpha\,|011\rangle + \beta\,|110\rangle + \beta\,|101\rangle\big) \\
|\phi_2\rangle &= H \otimes I \otimes I\,|\phi_1\rangle \\
&= \frac{1}{\sqrt{2}}\Big(\alpha(H\,|0\rangle)\,|00\rangle + \alpha(H\,|0\rangle)\,|11\rangle + \beta(H\,|1\rangle)\,|10\rangle + \beta(H\,|1\rangle)\,|01\rangle\Big) \\
&= \frac{1}{\sqrt{2}}\Big(\alpha\,|+\rangle\,|00\rangle + \alpha\,|+\rangle\,|11\rangle + \beta\,|-\rangle\,|10\rangle + \beta\,|-\rangle\,|01\rangle\Big) \\
&= \frac{1}{2}\Big(\alpha\,|000\rangle + \alpha\,|100\rangle + \alpha\,|011\rangle + \alpha\,|111\rangle + \beta\,|010\rangle - \beta\,|110\rangle + \beta\,|001\rangle - \beta\,|101\rangle\Big) \qquad (7.4)
\end{aligned}
$$

Measurements occur immediately after the state reaches $|\phi_3\rangle$. Thus, we must first determine the measurement statistics before analyzing the state $|\phi_4\rangle$. There are four possible outcomes for (z, x). We will work through one case, $(z = 0, x = 1)$, to show that in this instance $|\phi_4\rangle = |\psi\rangle$. It is left as an exercise to verify that in all four cases, the appropriate correction yields $|\phi_4\rangle = |\psi\rangle$.

Case $(z = 0, x = 1)$

Alice measures her first two qubits in the computational basis and obtains $(z, x) \in \{0, 1\}^2$. For $(z = 0, x = 1)$ we project onto $|01\rangle$ on Alice's two qubits and keep only terms whose first two bits are 01.

From Equation (7.4), the $|01\rangle$ terms are $\alpha\,|011\rangle$ and $\beta\,|010\rangle$. Hence the (unnormalized) post-measurement state is

$$
\left|\widetilde{\phi}_4^{(0,1)}\right\rangle = \frac{1}{2}\big(\alpha\,|011\rangle + \beta\,|010\rangle\big) = \frac{1}{2}\,|01\rangle \otimes \big(\beta\,|0\rangle + \alpha\,|1\rangle\big). \qquad (7.5)
$$

The probability of this outcome is

$$
p_{01} = \left\|\left|\widetilde{\phi}_4^{(0,1)}\right\rangle\right\|^2 = \frac{1}{4}\left(|\alpha|^2 + |\beta|^2\right) = \frac{1}{4}. \qquad (7.6)
$$

Conditioned on $(z, x) = (0, 1)$, the normalized post-measurement state on Bob's qubit is

$$
\left|\phi_4^{(0,1)}\right\rangle = \frac{\left|\widetilde{\phi}_4^{(0,1)}\right\rangle}{\sqrt{p_{01}}} = |01\rangle \otimes \big(\beta\,|0\rangle + \alpha\,|1\rangle\big) = |01\rangle \otimes X\,|\psi\rangle\,, \qquad (7.7)
$$

since $X\,|\psi\rangle = X(\alpha\,|0\rangle + \beta\,|1\rangle) = \alpha\,|1\rangle + \beta\,|0\rangle = \beta\,|0\rangle + \alpha\,|1\rangle$. Therefore, Bob applies the Pauli correction X and recovers $|\psi\rangle$.

7.3 Superdense Coding

Motivation: *Can we encode more informaiton using a single qubit?* A qubit lives in a 2D Hilbert space, which in principle, have infinitely many values encoded by its amplitudes. So it is natural to hope that we could encode more than a single bit.

The answer is no, which follows immediately from the Holevo theorem. This theorem is out of the scope of this introductory course. Here, we simply state this fact without proof:

Corollary 7.3.1 (Corollary of Holevo theorem)**.** *An n-qubit channel can reliably transmit at most n classical bits.* ◊

However, by exploiting quantum entanglement, we can achieve the following:

$$(1 \text{ e-bit}) + (1 \text{ q-bit}) = (2 \text{ c-bits}),$$

where the (1 e-bit) denotes that each party holds one qubit of a shared EPR pair (i.e., their qubits are maximally entangled).

Teaching Suggestions

Note that teleportation can be thought of as:

$$(1 \text{ e-bit}) + (2 \text{ c-bits}) = (1 \text{ q-bit}).$$

7.3.1 Protocol

the setup is as follows: Alice has two classical bits (a, b) that she wants to send to Bob. Given that they share an EPR pair, this task can be done by Alice sending a single qubit to Bob. The protocol is illustrated in Figure 7.2. (Recall that we have learned how to read such a quantum circuit diagram in Supplementary Box 7.2.1.)

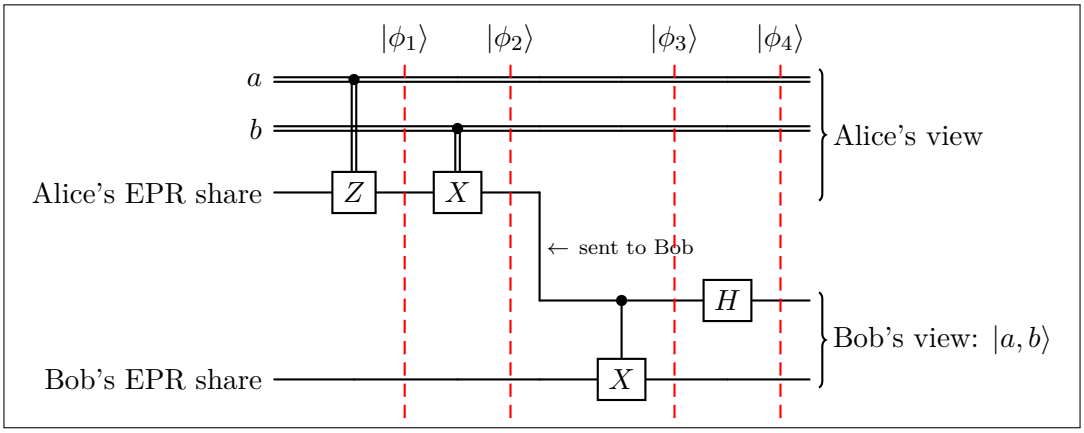


Figure 7.2: Quantum Circuit Diagram for Superdense Coding

7.3.2 Analysis

To verify the protocol, we track how the EPR pair evolves through the circuit; the results are summarized in Table 7.1.

Teaching Suggestions

To make the lecture more interactive and engaging, the instructor can invite students to derive (some of) the rows of Table 7.1 on the whiteboard, stepping in to help only when necessary.

(a, b)	$\|\phi_1\rangle$	$\|\phi_2\rangle$	$\|\phi_3\rangle$	$\|\phi_4\rangle$
$(0, 0)$	$\frac{\|00\rangle + \|11\rangle}{\sqrt{2}}$	$\frac{\|00\rangle + \|11\rangle}{\sqrt{2}}$	$\frac{\|0\rangle + \|1\rangle}{\sqrt{2}} \|0\rangle$	$\|00\rangle$
$(0, 1)$	$\frac{\|00\rangle + \|11\rangle}{\sqrt{2}}$	$\frac{\|10\rangle + \|01\rangle}{\sqrt{2}}$	$\frac{\|1\rangle + \|0\rangle}{\sqrt{2}} \|1\rangle$	$\|01\rangle$
$(1, 0)$	$\frac{\|00\rangle - \|11\rangle}{\sqrt{2}}$	$\frac{\|00\rangle - \|11\rangle}{\sqrt{2}}$	$\frac{\|0\rangle - \|1\rangle}{\sqrt{2}} \|0\rangle$	$\|10\rangle$
$(1, 1)$	$\frac{\|00\rangle - \|11\rangle}{\sqrt{2}}$	$\frac{\|10\rangle - \|01\rangle}{\sqrt{2}}$	$\frac{\|1\rangle - \|0\rangle}{\sqrt{2}} \|1\rangle$	$-\|11\rangle$

Table 7.1: Intermediate States of Superdense Coding

Remark. The last row of Table 7.1 contains the state $-|11\rangle$. For now, please treat this as equivalent to $|11\rangle$ and do not worry too much about this detail.

In short, the minus sign in front of $-|11\rangle$ represents the so-called "*global phase*," which has no physical effect on the quantum state. Therefore, $|11\rangle$ and $-|11\rangle$ are considered the same quantum state. A rigorous discussion of this concept requires the formalism of *density matrices*, which is beyond the scope of this undergraduate-level course in quantum computing.

7.4 EPR Paradox and CHSH Game

The EPR pair $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ is named after three physicists Albert Einstein, Boris Podolsky, and Nathan Rosen. We have already introduced a brief history of EPR pairs earlier when we discuss quantum teleportation.

The Weirdness of EPR pair. Alice and Bob always get the same outcomes if they measure an EPR pair $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ in the same basis (say computational basis).

EPR (Einstein, Podolsky, and Rosen) found this game weird—they believed that a complete physical theory should satisfy:

- **Realism:** Physical properties of a system have definite values whether or not they are measured. This means that a measurement simply reveals a pre-existing value, rather than creating the value upon measurement.
- **Locality:** Physical influences cannot propagate faster than the speed of light. This means that actions performed on one system cannot instantaneously affect the state of another spatially separated system.

Why do EPR believe that? Here's a breakdown of their reasoning:

Realism:

- **Correspondence with Classical Physics:** Classical physics embodies realism. An object has a definite position, momentum, energy, etc., regardless of whether we measure them. The act of measurement simply reveals these pre-existing properties. EPR saw no reason why this shouldn't hold true at the quantum level as well. They viewed the probabilistic nature of quantum mechanics as a sign of its incompleteness, not a fundamental feature of reality.
- **Intuitive Understanding of Reality:** Realism aligns with our everyday experience. We perceive objects as having definite properties independent of observation. The idea that a property only comes into existence when measured seemed counter-intuitive and philosophically unsettling to EPR.

Locality:

- **Special Relativity:** Einstein's theory of special relativity postulates that nothing can travel faster than light. This principle of locality is a cornerstone of modern physics. Allowing instantaneous action at a distance would open the door to potential paradoxes and violations of causality.
- **Principle of Local Causality:** This principle states that an event can only be influenced by events in its immediate spacetime neighborhood. It is a fundamental assumption in how we understand cause and effect. Non-locality would imply that an event here could instantaneously cause an effect somewhere else in the universe, which seemed highly problematic to EPR.
- **Philosophical Concerns:** The idea of instantaneous action at a distance seemed "spooky" and unphysical to EPR. It contradicted their understanding of how physical influences should propagate.

Local Hidden Variable theory

To resolve the obvious conflicts between the physical facts of EPR pair and the belief hold by them, they (Einstein, Podolsky, and Rosen) proposed the so-called *Local Hidden Variable theory*, which says

- There must be some underlying "hidden variables" that predetermine the measurement outcome (e.g., the spin values of both particles) from the moment they become entangled.
- The measurement would then simply reveal these pre-existing values.

According this this theory, there should be no "spooky" action at a distance, but rather that QM is not complete in that it does not touch/reveal the real-truth—the hidden variables.

7.4.1 Bell Test and CHSH Game

John Bell (1964) proved a landmark result: no local hidden variable theory can reproduce all the predictions of quantum mechanics. In other words, if one assumes locality (that spacelike separated systems cannot influence each other instantaneously) and pre-existing hidden variables that determine outcomes, then certain experimentally testable inequalities must hold. However, quantum mechanics predicts violations of these inequalities. Consequently, Bell devised an experiment—now

called a *Bell test*—whose outcomes distinguish quantum predictions from any local hidden variable (LHV) model.

To make Bell's test accessible and operational, Clauser, Horne, Shimony, and Holt proposed a simplified scenario in the 1970s, known as the CHSH game. This game captures the essence of non-local correlations in a clean, two-player setting.

CHSH Game

There are three parties:

- Two cooperating but spatially separated players, Alice and Bob.
- A referee who runs the game with Alice and Bob.

Crucially, Alice and Bob may agree on a strategy beforehand, but after the game begins, they cannot communicate.

The game proceeds as follows:

1. The referee flips two fair coins to generate random bits $x, y \in \{0, 1\}$.
2. The referee sends x to Alice and y to Bob.
3. Alice replies with a bit a, and Bob replies with a bit b.
4. Alice and Bob win if and only if $a \oplus b = x \wedge y$.

We present the win condition in Table 7.2:

x	y	Win condition
0	0	$a = b$
0	1	$a = b$
1	0	$a = b$
1	1	$a \neq b$

Table 7.2: Table of CHSH Win Condition

We focus on the following question:

- *What is the maximum achievable winning probability?*

In what follows, we will show that quantum strategies yield answers that differ from those of all classical strategies, including any local hidden-variable theory. The elegance of this result lies in the fact that:

1. It demonstrates that quantum information is fundamentally different from classical information; there is genuinely something new in quantum mechanics that cannot be captured by classical theories.
2. It rules out every classical attempt to subsume quantum behavior, without requiring us to specify the details of any particular classical model!

7.4.2 Classical Strategies for CHSH Game

Deterministic Strategies. First, let us consider deterministic strategies. In this case, Alice's answer is a fixed function of her question, $a = a(x)$, and Bob's answer is a fixed function $b = b(y)$. Because there are only four choices for each player's function, one can enumerate them or reason directly. It turns out that the best possible deterministic strategy wins with probability 3/4.

For instance, a simple strategy achieving 3/4 is to always output 0: that is, $a(x) = 0$ and $b(y) = 0$. Then Alice and Bob succeed whenever $(x, y) \neq (1, 1)$, which occurs with probability 3/4.

Local Hidden Variable Strategies. Next, we consider local hidden variable (LHV) strategies. Here, there is an underlying random variable λ drawn from some distribution L before the questions are chosen. The model assumes:

1. The hidden variable λ is sampled from L prior to the game.
2. The questions (x, y) are sampled independently of λ.
3. Alice's answer is $a = a(x, \lambda)$.
4. Bob's answer is $b = b(y, \lambda)$.

Intuitively, λ can correlate Alice and Bob's responses, but it **cannot** depend on the questions (which are independent). By conditioning on λ, the strategy becomes deterministic for fixed λ. Hence,

$$\Pr[\text{win}] = \sum_{\lambda} \Pr[\lambda] \cdot \Pr[\text{win} \mid \lambda] \leq \sum_{\lambda} \Pr[\lambda] \cdot \frac{3}{4} = \frac{3}{4}.$$

Therefore, even with shared randomness via λ, the maximum winning probability remains 3/4. Historically, this aligns with Einstein's expectation that local realism should limit correlations in precisely this way.

7.4.3 Quantum Strategy for CHSH Game

Having seen the classical limit, we now turn to quantum mechanics. Remarkably, quantum entanglement enables a strategy that surpasses the 3/4 bound.

Entangled State and Measurement Plan

Before the game starts, Alice and Bob meet to prepare a maximally entangled two-qubit state (an EPR pair):

$$|\Phi\rangle = \frac{1}{\sqrt{2}}\big(\,|00\rangle + |11\rangle\,\big).$$

Then they separate, each carrying one qubit. Upon receiving their respective questions x and y, they perform local measurements on their qubits according to the following plan, where we describe their strategies one by one.

Alice's Strategy:

- If $x = 0$, Alice measures in the computational basis $\{|0\rangle, |1\rangle\}$ and outputs $a = 0$ for outcome $|0\rangle$, $a = 1$ for outcome $|1\rangle$.

- If $x = 1$, Alice measures in the Hadamard basis $\{|+\rangle, |-\rangle\}$, and outputs $a = 0$ for $|+\rangle$, $a = 1$ for $|-\rangle$.

We illustrate the basis for Alice's measurements in Figures 7.3 and 7.4.

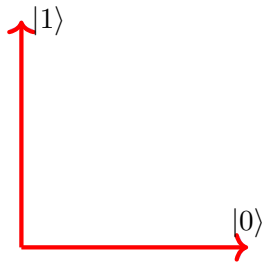


Figure 7.3: Alice's computational basis ($x = 0$)

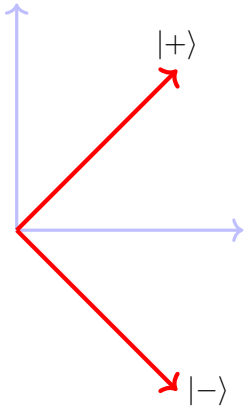


Figure 7.4: Alice's Hadamard basis ($x = 1$)

Bob's Strategy:

- If $y = 0$, Bob measures in the basis $\{|s_0\rangle, |s_1\rangle\}$ defined by

$$|s_0\rangle = \cos\left(\tfrac{\pi}{8}\right)|0\rangle + \sin\left(\tfrac{\pi}{8}\right)|1\rangle, \quad |s_1\rangle = -\sin\left(\tfrac{\pi}{8}\right)|0\rangle + \cos\left(\tfrac{\pi}{8}\right)|1\rangle,$$

  and outputs $b = 0$ for $|s_0\rangle$, $b = 1$ for $|s_1\rangle$.

- If $y = 1$, Bob measures in the basis $\{|t_0\rangle, |t_1\rangle\}$ defined by

$$|t_0\rangle = \cos\left(\tfrac{\pi}{8}\right)|0\rangle - \sin\left(\tfrac{\pi}{8}\right)|1\rangle, \quad |t_1\rangle = \sin\left(\tfrac{\pi}{8}\right)|0\rangle + \cos\left(\tfrac{\pi}{8}\right)|1\rangle,$$

  and outputs $b = 0$ for $|t_0\rangle$, $b = 1$ for $|t_1\rangle$.

We illustrate the basis for Bob's measurements in Figures 7.5 and 7.6.

7.4.4 Performance Analysis

To analyze joint measurements on the two separate qubits, we can, without loss of generality, imagine that Alice measures first and then Bob; the joint probabilities are the same as if Bob measured first. Consider, for example, the case $x = 0, y = 0$.

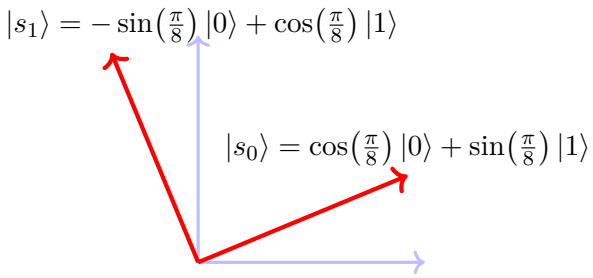


Figure 7.5: Bob's s-basis (y = 0)

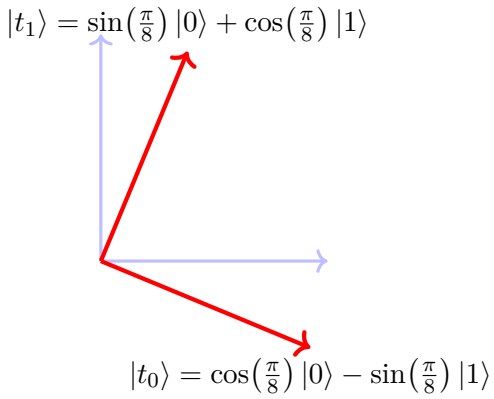


Figure 7.6: Bob's t-basis ($y = 1$)

Case $x = 0, y = 0$**.** Alice measures in the computational basis. With probability 1/2, she obtains $|0\rangle$, in which case the post-measurement state collapses to $|0\rangle \otimes |0\rangle$. To win, Bob must output $b = 0$, i.e., obtain the outcome $|s_0\rangle$. Since his reduced state is $|0\rangle$, the success probability in this branch is

$$|\langle 0|s_0\rangle|^2 = \cos^2\left(\frac{\pi}{8}\right) \approx 0.8535.$$

Similarly, with probability 1/2, Alice obtains $|1\rangle$, collapsing the state to $|1\rangle \otimes |1\rangle$. To win, Bob must output $b = 1$, i.e., obtain $|s_1\rangle$. Given his state is $|1\rangle$, the success probability is

$$|\langle 1|s_1\rangle|^2 = \cos^2\left(\frac{\pi}{8}\right) \approx 0.8535.$$

Therefore, for $(x, y) = (0, 0)$, the overall winning probability is $\cos^2(\pi/8)$.

Case $x = 1, y = 1$**.** Now Alice measures in the diagonal basis. If she obtains $|+\rangle$, the joint state collapses to $|+\rangle \otimes |+\rangle$. To win in this input setting, Alice and Bob must produce different bits. Under Bob's $y = 1$ measurement, he succeeds by obtaining $|t_1\rangle$ (mapped to $b = 1$) when Alice outputs $a = 0$. The probability of Bob obtaining $|t_1\rangle$ from $|+\rangle$ equals

$$|\langle +|t_1\rangle|^2 = \cos^2\left(\frac{\pi}{8}\right).$$

On the other hand, if Alice obtains $|-\rangle$, the state collapses to $|-\rangle \otimes |-\rangle$. Then Bob should obtain

$|t_0\rangle$ (mapped to $b = 0$) to differ from $a = 1$. This again occurs with probability

$$|\langle -|t_0\rangle|^2 = \cos^2\left(\frac{\pi}{8}\right).$$

Other cases. The remaining two input pairs $(x, y) = (1, 0)$ and $(0, 1)$ are handled analogously, yielding the same success probability. We leave them as exercises for you.

Overall success probability. Aggregating all four input pairs, the quantum strategy wins with probability

$$\cos^2\left(\frac{\pi}{8}\right) \approx 0.8535,$$

which strictly exceeds the classical/LHV maximum of $3/4 = 0.75$. This quantum advantage arises from entanglement, which enables correlations stronger than any achievable by local hidden variables. Nonetheless, it does not allow signaling, preserving relativistic causality.

Optimality (Tsirelson's bound). It turns out that $\cos^2(\pi/8)$ is the optimal quantum winning probability for the CHSH game. More precisely, Tsirelson's theorem shows that the maximal quantum value of the CHSH expression equals $2\sqrt{2}$, which translates into a winning probability of $\cos^2(\pi/8)$. Therefore, while quantum mechanics outperforms classical physics in this task, it still respects a precise upper limit on nonlocal correlations.

7.4.5 Conclusion

To summarize, the CHSH game provides a crisp, operational way to test the predictions of quantum mechanics against local hidden variable theories. Deterministic and LHV strategies cannot exceed a 3/4 winning probability, whereas an entanglement-based quantum strategy achieves approximately 85.35%. Finally, Tsirelson's theorem proves that this quantum advantage is optimal, and a wealth of experiments confirms that nature indeed exhibits such nonlocal correlations.

Experimental Confirmation Since 1972, numerous Bell tests have been conducted, including many based on the CHSH scenario and others employing different setups. Repeatedly, experiments have observed violations of classical (LHV) bounds consistent with the quantum prediction near $\cos^2(\pi/8)$. Consequently, two high-level conclusions have emerged:

1. Quantum mechanics is fundamentally non-classical; moreover, nature appears to be quantum mechanical.

2. Physical outcomes are intrinsically probabilistic, not merely the result of hidden deterministic variables obeying locality.

Furthermore, increasingly sophisticated "loophole-free" Bell tests have reinforced these conclusions, closing detection and locality loopholes and thereby strengthening the empirical case against local hidden variable theories.

Bibliography

[EPR35] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935. 72

[NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, USA, 10th edition, 2010. 39, 57