An Introduction to Quantum Machine Learning

Jeongbin Jo[1, *]

[1]*Department of Physics, Yonsei University, Seoul, 03722, Republic of Korea*

(Dated: May 4, 2026)

Quantum machine learning (QML) sits at the intersection of quantum information science and modern artificial intelligence, offering a promising route toward computational advantages that are inaccessible by classical means alone. This report provides a self-contained introduction to the theoretical foundations necessary to understand and critically evaluate QML proposals. Starting from the postulates of quantum mechanics—state, dynamics, and measurement—we develop the quantum circuit model with single- and multi-qubit gates, establish the universality of quantum gate sets, and examine the no-cloning theorem and elementary quantum communication protocols. We then discuss reversible computation, Landauer's principle, the Hadamard and swap tests with complete circuit derivations, and the basics of quantum computational complexity, culminating in a tutorial treatment of the HHL linear-systems algorithm, the classical SVM and least-squares SVM formulations that reduce training to a linear system, their connection to quantum speedups, and an overview of further landmark QML proposals together with central open questions around quantum advantage. Throughout, formal derivations are presented at a level of detail beyond standard textbook treatments, with the aim of equipping the reader with both intuition and rigorous mathematical foundations.

CONTENTS

* jeongbin033@yonsei.ac.kr

I. INTRODUCTION

A. Motivation: The Age of Quantum Computing

For over half a century, classical computing has followed Moore's Law: the number of transistors on a microprocessor doubles approximately every two years, delivering exponential growth in computational power at constant cost [1]. However, as transistor feature sizes approach atomic dimensions, quantum mechanical effects such as tunneling and thermal fluctuations undermine deterministic switching. This physical barrier signals the end of the classical scaling era and motivates the search for radically different computational paradigms.

Quantum computing exploits the principles of quantum mechanics—superposition, entanglement, and interference—to represent and process information in ways that classical computers cannot efficiently simulate. A classical register of n bits can store exactly one of 2^n binary strings at any given time. By contrast, an n-qubit quantum register can exist in a superposition of all 2^n computational basis states simultaneously (see Postulate I in Section III A for a formal definition). The amplitudes $\{\alpha_x\}$ encode correlations across an exponentially large state space, and unitary operations manipulate all 2^n amplitudes in parallel—a wave-like processing step. Measurement then collapses this superposition, yielding a bit-string x with probability $|\alpha_x|^2$—a particle-like readout step. This *analog-digital duality* underlies the power of quantum computing [2].

B. Classical vs. Quantum Information

Classical information theory studies the representation, compression, and transmission of information encoded in discrete, deterministic symbols [3]. A probabilistic extension replaces each bit state by a probability distribution $\mathbf{p} = (p_0, p_1)^\top$, $p_0 + p_1 = 1$, $p_i \geq 0$. Operations on probabilistic bits are described by stochastic matrices—matrices whose columns are probability vectors.

Quantum information theory is a mathematical generalization of classical probabilistic information. Specifically, whereas classical transitions are modeled by a *Markov chain* where a stochastic matrix M acts on a probability vector $\mathbf{x}$, quantum transitions are governed by *unitary evolution* where a unitary matrix U acts on a ket state $|\psi\rangle$. The key correspondences are summarized below:

	Classical (Markov)	Quantum (Unitary)
State	Prob. vector $\mathbf{x} \in \mathbb{R}_+^d$	Ket $\|\psi\rangle \in \mathbb{C}^d$
Normalization	L_1 norm ($\sum x_i = 1$)	L_2 norm ($\sum \|\alpha_i\|^2 = 1$)
Evolution	Stochastic M ($\sum_i M_{ij} = 1$)	Unitary U ($U^\dagger U = \mathbf{I}$)
Observation	Sampling from $\mathbf{x}$	Born-rule measurement

This structural analogy reveals that quantum mechanics is, in a precise sense, a *complex-amplitude* generalization of probability theory [4]. The crucial difference is that complex amplitudes can interfere constructively and destructively—classical probabilities cannot. Interference is the key resource exploited by quantum algorithms to suppress wrong answers and amplify correct ones [2].

C. Machine Learning as an Optimization Problem

Machine Learning (ML) aims to enable computers to learn underlying properties of data to perform tasks without explicit programming. Mathematically, ML represents an optimization problem where we search for a model $f(z, x)$ parameterized by z that minimizes an expected loss L over a data distribution $\mathcal{D}$ [5]:

$$\min_{z \in \Omega} \mathbb{E}_{x \sim \mathcal{S}} \big\{ L[f(z, x)] \big\}, \tag{1}$$

where $\mathcal{S} = \{x_i\}_{i=1}^m \sim \mathcal{D}$ is the available sample. Typical objective functions can be derived from the principle of *Maximum Likelihood Estimation* (MLE). For a dataset $Y = \{y_1, \ldots, y_m\}$, the log-likelihood $L(\theta)$ is maximized:

$$\max_\theta L(\theta) = \sum_{i=1}^m \log(p(y_i|x_i, \theta)). \tag{2}$$

Under the assumption of i.i.d. Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$ in observations $y_i = f(x_i, \theta) + \epsilon$, MLE is equivalent to mini-

mizing the sum of squared errors (SSE/MSE):

$$L(\theta) \propto -\sum_{i=1}^{m}(y_i - f(x_i,\theta))^2 + \text{const.} \tag{3}$$

Similarly, for classification tasks with Bernoulli labels $y_i \in \{0,1\}$, MLE yields the cross-entropy loss function:

$$L(\theta) = \sum_{i=1}^{m}\big[y_i \log\theta_i + (1-y_i)\log(1-\theta_i)\big]. \tag{4}$$

D. Scope and Organization of This Report

This report follows STA4121 (through Week 10) at a level of mathematical detail intended to complement and, in key derivations, surpass standard textbook treatments. The organization is as follows. Section I introduces the motivation and the machine learning optimization framework. Section II reviews prerequisites including Dirac notation, computational complexity, and classical information theory. Section III develops the three postulates of quantum mechanics. Section IV treats entanglement and density matrices. Section V introduces the quantum circuit model and elementary protocols including Superdense Coding and Teleportation. Section VI discusses reversible computation, Landauer's principle, and the Hadamard/swap tests. Section VII presents the quantum query model and Grover search. Section VIII develops the quantum Fourier transform and phase estimation as subroutines underpinning exponential quantum speedups. Section IX connects these tools to machine learning: the HHL algorithm, classical and least-squares support vector machines, and the quantum LS-SVM pipeline, followed by a concise survey of other QML ideas. Section X concludes.

II. MATHEMATICAL PRELIMINARIES

A. Dirac Notation and Hilbert Space

All quantum states live in a complex Hilbert space $\mathcal{H}$, a complete inner-product space over $\mathbb{C}$. Following Dirac's notation:

- A *ket* $|\psi\rangle \in \mathcal{H}$ is a column vector representing a quantum state.
- A *bra* $\langle\psi| = |\psi\rangle^\dagger$ is the conjugate-transpose row vector.
- The *inner product* $\langle\phi|\psi\rangle \in \mathbb{C}$ satisfies linearity in the second argument and $\langle\psi|\psi\rangle \geq 0$, with equality iff $|\psi\rangle = 0$.
- The *outer product* $|\psi\rangle\langle\phi|$ is a linear operator on $\mathcal{H}$.

For an n-qubit system, $\mathcal{H} = \mathbb{C}^{2^n}$ with the standard inner product $\langle\phi|\psi\rangle = \sum_x \phi_x^* \psi_x$. The *computational basis* $\{|x\rangle\}_{x\in\{0,1\}^n}$ is the set of standard basis vectors, which forms a complete orthonormal set satisfying $\langle x|y\rangle = \delta_{xy}$ and $\sum_x |x\rangle\langle x| = \mathbf{I}$.

B. Asymptotic Complexity Notation

In the analysis of both classical and quantum algorithms, it is essential to characterize how the computational cost (time or space) scales with the input size N as $N \to \infty$ [6].

1. **Big-O** (O): $f(N) = O(g(N))$ if there exist constants $c, N_0 > 0$ such that $f(N) \leq c \cdot g(N)$ for all $N \geq N_0$. This provides an *upper bound* on growth.
2. **Big-Omega** (Ω): $f(N) = \Omega(g(N))$ if $f(N) \geq c \cdot g(N)$ for all $N \geq N_0$. This provides a *lower bound*.
3. **Big-Theta** (Θ): $f(N) = \Theta(g(N))$ if $f(N) = O(g(N))$ and $f(N) = \Omega(g(N))$. This represents a *tight bound*.
4. **Little-o** (o) and **Little-omega** (ω): Used for strict upper and lower bounds, respectively.

An algorithm is generally considered *efficient* if its complexity is polynomial in N, i.e., $O(N^k)$ for some constant k. Key examples include:

- Inner product $\mathbf{x}^\top\mathbf{y}$ costs $O(N)$.
- Matrix-vector product $A\mathbf{x}$ costs $O(MN)$ for $A \in \mathbb{R}^{M\times N}$.
- Classical matrix inversion costs $O(N^3)$, while the HHL algorithm offers an exponential speedup to $O(\text{poly}\log N)$ for sparse, well-conditioned systems [7].

C. Classical Information: Probabilistic States and Stochastic Matrices

Let $\mathcal{S} = \{0,1\}$ denote the classical state set of a bit X. A *probabilistic state* of X is a column vector $\mathbf{p} = (p_0, p_1)^\top \in \mathbb{R}^2$ where $p_a = \Pr(X = a)$, so $p_0, p_1 \geq 0$ and $p_0 + p_1 = 1$.

A *deterministic operation* mapping $a \mapsto f(a)$ is represented by the matrix M satisfying $M|a\rangle = |f(a)\rangle$ for all $a \in \mathcal{S}$. A *probabilistic operation* introducing randomness is represented by a *stochastic matrix*: a matrix M satisfying $M_{ij} \geq 0$ and $\sum_i M_{ij} = 1$ for all j. Stochastic matrices are exactly those that map probability vectors to probability vectors. Composition of operations:

$$\mathbf{p} \xrightarrow{M_1} M_1\mathbf{p} \xrightarrow{M_2} M_2M_1\mathbf{p}, \tag{5}$$

and the product M_2M_1 is again stochastic.

III. POSTULATES OF QUANTUM MECHANICS

A. Postulate I: State

Postulate I (State). *Any isolated physical system is completely described by a unit vector $|\psi\rangle$ in a complex Hilbert space $\mathcal{H}$, called the state space. For an n-qubit system $\mathcal{H} = \mathbb{C}^{2^n}$ and $\langle\psi|\psi\rangle = 1$.*

For a single qubit, $\mathcal{H} = \mathbb{C}^2$ and the most general state is

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad \alpha,\beta \in \mathbb{C}, \quad |\alpha|^2 + |\beta|^2 = 1. \tag{6}$$

a. Bloch sphere parameterization. Writing $\alpha = e^{i\gamma}\cos(\theta/2)$ and $\beta = e^{i(\gamma+\phi)}\sin(\theta/2)$, the global phase $e^{i\gamma}$ is physically unobservable (it cancels in every Born-rule probability $p(m) = |\langle m|\psi\rangle|^2$). Absorbing it, the canonical form is

$$|\psi\rangle = \cos\tfrac{\theta}{2}|0\rangle + e^{i\phi}\sin\tfrac{\theta}{2}|1\rangle, \quad \theta \in [0,\pi],\ \phi \in [0,2\pi). \tag{7}$$

The pair (θ, ϕ) defines a point on the unit sphere—the *Bloch sphere*—giving a bijective correspondence between physically distinct pure single-qubit states and S^2. Antipodal points correspond to orthogonal states: the poles $|0\rangle$ and $|1\rangle$ are mutually orthogonal.

Sketch of proof. We verify Eq. (7) represents all and only unit vectors (up to global phase). Any $(\alpha, \beta) \in \mathbb{C}^2$ with $|\alpha|^2+|\beta|^2 = 1$ can be written as $\alpha = r_0 e^{i\phi_0}$, $\beta = r_1 e^{i\phi_1}$ with $r_0^2 + r_1^2 = 1$, $r_i \geq 0$. Setting $r_0 = \cos(\theta/2)$, $r_1 = \sin(\theta/2)$ (unique for $\theta \in [0, \pi]$), and factoring out $e^{i\phi_0}$ as the global phase gives $\phi = \phi_1 - \phi_0$. The parameterization is surjective onto the sphere and two-to-one only at the poles ($\theta = 0, \pi$), where ϕ is irrelevant—consistent with $|0\rangle, |1\rangle$ being pole states. ◊ ◊

B. Postulate II: Dynamics

Postulate II (Dynamics). *The time evolution of a closed quantum system is governed by the Schrödinger equation $i\hbar\,\partial_t |\psi(t)\rangle = H(t)|\psi(t)\rangle$, where $H(t)$ is the Hermitian Hamiltonian of the system. Equivalently, $|\psi(t)\rangle = U(t)|\psi(0)\rangle$ for some unitary $U(t)$.*

a. General solution via Dyson series. Formally integrating the Schrödinger equation gives

$$|\psi(t)\rangle = \mathcal{T}\exp\left(-\frac{i}{\hbar}\int_0^t H(t')\,dt'\right)|\psi(0)\rangle\,, \tag{8}$$

where $\mathcal{T}$ denotes time-ordering. For a *time-independent* Hamiltonian $H(t) = H$, the series resums exactly:

$$|\psi(t)\rangle = e^{-iHt/\hbar}|\psi(0)\rangle =: U(t)|\psi(0)\rangle\,. \tag{9}$$

Sketch of proof. Split $[0, t]$ into L equal intervals of width $\delta = t/L$. For small δ, $U(\delta) \approx I - iH\delta/\hbar$. The total evolution is $U(t) \approx (I - iH\delta/\hbar)^L \xrightarrow{L\to\infty} e^{-iHt/\hbar}$ by the definition of the matrix exponential. Hermiticity of H gives $U^\dagger = e^{iH^\dagger t/\hbar} = e^{iHt/\hbar}$, hence $U^\dagger U = e^{iHt/\hbar}e^{-iHt/\hbar} = I$, confirming unitarity. ◊ ◊

b. Spectral form and inner-product preservation. Since H is Hermitian, its spectral decomposition reads $H = \sum_k E_k |E_k\rangle\langle E_k|$, giving

$$U(t) = \sum_k e^{-iE_k t/\hbar}|E_k\rangle\langle E_k|\,. \tag{10}$$

Because $|e^{-iE_k t/\hbar}| = 1$, the evolution merely rotates phases; it preserves the norm and inner product: $\langle\psi(t)|\phi(t)\rangle = \langle\psi(0)|U^\dagger U|\phi(0)\rangle = \langle\psi(0)|\phi(0)\rangle$.

c. Circuit model. In quantum computing we set $\hbar = 1$. A quantum circuit decomposes $U = U_L \cdots U_2 U_1$ where each U_k acts on one or two qubits. Because a product of unitaries is unitary ($(VU)^\dagger(VU) = U^\dagger V^\dagger V U = I$), every circuit realizes a valid unitary.

C. Postulate III: Measurement

Postulate III (Measurement). *Quantum measurements are described by a collection $\{M_m\}$ of measurement operators satisfying the completeness relation $\sum_m M_m^\dagger M_m = \mathbf{I}$. If the pre-measurement state is $|\psi\rangle$, the outcome m occurs with probability $p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle$, and the post-measurement state is $M_m|\psi\rangle/\sqrt{p(m)}$.*

a. Projective measurements. A *projective* (von Neumann) measurement is associated with a Hermitian observable $O = \sum_m \mu_m P_m$, where μ_m are real eigenvalues and $P_m = |\mu_m\rangle\langle\mu_m|$ are orthogonal projectors: $P_m P_{m'} = \delta_{mm'} P_m$, $\sum_m P_m = \mathbf{I}$. Setting $M_m = P_m$:

$$p(\mu_m) = \langle\psi|P_m|\psi\rangle\,, \qquad |\psi'\rangle = \frac{P_m|\psi\rangle}{\sqrt{p(\mu_m)}}. \tag{11}$$

b. Expectation value.

$$\langle O\rangle = \sum_m \mu_m p(\mu_m) = \langle\psi|O|\psi\rangle = \mathrm{Tr}(O\,|\psi\rangle\langle\psi|). \tag{12}$$

The second equality follows from $\sum_m \mu_m\langle\psi|P_m|\psi\rangle = \langle\psi|(\sum_m \mu_m P_m)|\psi\rangle$. The third uses the cyclic property of the trace.

c. Computational basis measurement. Setting $M_m = |m\rangle\langle m|$ for $m \in \{0,1\}^n$ and $|\psi\rangle = \sum_x \alpha_x |x\rangle$, outcome m occurs with probability $|\alpha_m|^2$, and the post-measurement state collapses to $|m\rangle$.

IV. MULTI-QUBIT SYSTEMS

A. Composite Systems and Tensor Products

Postulate IV (Composite Systems). *The state space of a composite physical system is the tensor product $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ of the component Hilbert spaces. If subsystem k is in state $|\psi_k\rangle$, the joint state is $|\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle$.*

For two qubits with $|\psi_1\rangle = \alpha_1|0\rangle + \beta_1|1\rangle$ and $|\psi_2\rangle = \alpha_2|0\rangle + \beta_2|1\rangle$, the joint state is

$$\begin{aligned}|\psi_1\rangle \otimes |\psi_2\rangle = {}& \alpha_1\alpha_2|00\rangle + \alpha_1\beta_2|01\rangle \\ & + \beta_1\alpha_2|10\rangle + \beta_1\beta_2|11\rangle\,,\end{aligned} \tag{13}$$

which coincides with the Kronecker product of the column vectors. An n-qubit Hilbert space has dimension 2^n, growing exponentially—the root of both the power and the classical-simulation difficulty of quantum computation.

B. Quantum Entanglement

A bipartite state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is *separable* if $|\Psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$; otherwise it is *entangled*. The four *Bell states*,

$$|\Phi^\pm\rangle = \tfrac{1}{\sqrt{2}}(|00\rangle \pm |11\rangle), \tag{14}$$

$$|\Psi^\pm\rangle = \tfrac{1}{\sqrt{2}}(|01\rangle \pm |10\rangle), \tag{15}$$

are maximally entangled two-qubit states and form an orthonormal basis for $\mathbb{C}^4$ (the Bell basis).

a. Schmidt decomposition and separability criterion. Any $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ admits the *Schmidt decomposition*

$$|\Psi\rangle = \sum_k \sqrt{\lambda_k}\, |a_k\rangle\, |b_k\rangle\,, \quad \lambda_k \geq 0, \quad \sum_k \lambda_k = 1, \tag{16}$$

where $\{|a_k\rangle\}$ and $\{|b_k\rangle\}$ are orthonormal. The Schmidt rank (number of non-zero λ_k) characterizes entanglement: a state is separable iff its Schmidt rank is 1.

Sketch of proof. Represent $|\Psi\rangle$ as a matrix $M_{ij} = \alpha_{ij}$ via $|\Psi\rangle = \sum_{i,j} \alpha_{ij}\, |a_i\rangle\, |b_j\rangle$. Applying the singular value decomposition (SVD) $M = U\Sigma V^\dagger$ gives $|\Psi\rangle = \sum_k \sigma_k\, |\tilde{a}_k\rangle\, |\tilde{b}_k\rangle$ where $|\tilde{a}_k\rangle = \sum_i U_{ik}\, |a_i\rangle$ and $|\tilde{b}_k\rangle = \sum_j V^*_{jk}\, |b_j\rangle$ are new orthonormal sets. Setting $\sqrt{\lambda_k} = \sigma_k$ and normalizing completes the proof. ◊ ◊

C. Partial Measurement

Consider a two-qubit state $|\Psi\rangle = \sum_{i,j} \alpha_{ij}\, |ij\rangle$ with $\sum_{i,j} |\alpha_{ij}|^2 = 1$. Measuring only the first qubit with $M_0 = |0\rangle\langle 0| \otimes \mathbf{I}$ and $M_1 = |1\rangle\langle 1| \otimes \mathbf{I}$:

- Outcome 0 with probability $p(0) = |\alpha_{00}|^2 + |\alpha_{01}|^2$; post-state $|0\rangle \otimes \frac{\alpha_{00}|0\rangle + \alpha_{01}|1\rangle}{\sqrt{p(0)}}$.
- Outcome 1 with probability $p(1) = |\alpha_{10}|^2 + |\alpha_{11}|^2$; post-state $|1\rangle \otimes \frac{\alpha_{10}|0\rangle + \alpha_{11}|1\rangle}{\sqrt{p(1)}}$.

Notably, if $|\Psi\rangle = |\Phi^+\rangle$, measuring the first qubit and obtaining 0 instantly collapses the second qubit to $|0\rangle$, regardless of distance. This is *not* superluminal signaling: the outcome probabilities are always $1/2$.

D. Pauli Matrices

The Pauli matrices

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{17}$$

together with $I \equiv \sigma_0$ satisfy $P^2 = I$, $[X, Y] = 2iZ$ (and cyclic), and $\{P, Q\} = 2\delta_{PQ} I$. The set $\mathcal{P}_n = \{I, X, Y, Z\}^{\otimes n}$ (with phases $\{\pm 1, \pm i\}$) forms the n-qubit Pauli group and provides a basis for $2^n \times 2^n$ Hermitian matrices. This makes Pauli operators the natural language for quantum error correction and VQA cost functions.

E. Density Matrix Formalism

When the state of a system is only known probabilistically—system is in $|\psi_k\rangle$ with probability p_k—the *density operator*

$$\rho = \sum_k p_k\, |\psi_k\rangle\langle\psi_k| \tag{18}$$

provides a complete description. Key properties: (i) $\rho \geq 0$, (ii) $\mathrm{Tr}(\rho) = 1$, (iii) $\mathrm{Tr}(\rho^2) \leq 1$, with equality iff ρ is pure. Under unitary evolution, $\rho \to U\rho U^\dagger$. Under measurement $\{M_m\}$:

$$p(m) = \mathrm{Tr}(M_m^\dagger M_m \rho), \qquad \rho' = \frac{M_m \rho M_m^\dagger}{p(m)}. \tag{19}$$

The expectation value generalizes Eq. (12) to $\langle O\rangle = \mathrm{Tr}(O\rho)$.

F. Reduced Density Matrix and Partial Trace

A central problem in quantum information is describing a subsystem A of a composite system AB when we lack access to system B. If the joint system is in state ρ_{AB}, the *reduced density matrix* ρ_A is defined via the *partial trace* over B:

$$\rho_A \equiv \mathrm{Tr}_B(\rho_{AB}) = \sum_j \langle j|_B\, \rho_{AB}\, |j\rangle_B\,, \tag{20}$$

where $\{|j\rangle_B\}$ is any orthonormal basis for $\mathcal{H}_B$.

a. Proof of Physical Consistency. The reduced density matrix ρ_A is the unique operator that correctly predicts the expectation values of all local observables O_A on system A.

Sketch of proof. Let $O = O_A \otimes I_B$ be an observable acting only on system A. Its expectation value in state ρ_{AB} is:

$$\begin{aligned}
\langle O_A \otimes I_B\rangle &= \mathrm{Tr}\big((O_A \otimes I_B)\rho_{AB}\big) \\
&= \sum_i \sum_j \langle i|_A\, \langle j|_B\, (O_A \otimes I_B)\rho_{AB}\, |i\rangle_A\, |j\rangle_B \\
&= \sum_i \langle i|_A\, O_A \underbrace{\left(\sum_j \langle j|_B\, \rho_{AB}\, |j\rangle_B\right)}_{\mathrm{Tr}_B(\rho_{AB})} |i\rangle_A \\
&= \sum_i \langle i|_A\, O_A \rho_A\, |i\rangle_A = \mathrm{Tr}(O_A \rho_A).
\end{aligned}$$

This identity proves that ρ_A contains all information necessary to describe local experiments on A [3]. ◊ ◊

b. Entanglement and Mixedness. Consider the maximally entangled Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Its joint density matrix is $\rho_{AB} = |\Phi^+\rangle\langle\Phi^+|$. Computing the partial trace over B:

$$\begin{aligned}
\rho_A &= \langle 0|_B\, \rho_{AB}\, |0\rangle_B + \langle 1|_B\, \rho_{AB}\, |1\rangle_B \\
&= \tfrac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \tfrac{1}{2}\mathbf{I}.
\end{aligned} \tag{21}$$

While the joint state ρ_{AB} is pure ($\mathrm{Tr}(\rho_{AB}^2) = 1$), the reduced state ρ_A is *maximally mixed* ($\mathrm{Tr}(\rho_A^2) = 1/2$). This conversion of global entanglement into local classical uncertainty (entropy) is a hallmark of quantum correlations [2].

V. QUANTUM CIRCUIT MODEL

A. Motivation: Classical Circuits

Classical digital computation is described by Boolean circuits: directed acyclic graphs of *gates* with *wires* carrying bits. The set $\{\mathrm{NAND}, \mathrm{FANOUT}\}$ is universal. However, NAND is *irreversible* (many-to-one), which by Landauer's principle implies thermodynamic energy dissipation (Sec. VI A).

B. Single-Qubit Gates

Single-qubit gates are 2×2 unitary matrices. The most important examples:

$$H = \tfrac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{22}$$

$$S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}, \quad T^2 = S,\ S^2 = Z. \tag{23}$$

Rotation gates about Bloch-sphere axes:

$$R_{\hat{n}}(\theta) := e^{-i\theta\hat{n}\cdot\vec{\sigma}/2} = \cos\tfrac{\theta}{2}\, I - i\sin\tfrac{\theta}{2}\,\hat{n}\cdot\vec{\sigma}, \tag{24}$$

where $\vec{\sigma} = (X, Y, Z)$. Geometrically, $R_{\hat{n}}(\theta)$ rotates the Bloch-sphere vector by angle θ about the $\hat{n}$-axis.

a. Euler decomposition. Every single-qubit unitary $U \in SU(2)$ can be written as

$$U = e^{i\alpha} R_z(\beta)\, R_y(\gamma)\, R_z(\delta) \tag{25}$$

for some real $\alpha, \beta, \gamma, \delta$.

Sketch of proof. Since U is 2×2 unitary and $\det U = 1$ (up to global phase), write $U = \left(\begin{smallmatrix} a & -b^* \\ b & a^* \end{smallmatrix}\right)$ with $|a|^2 + |b|^2 = 1$. Set $a = e^{i(\alpha-\beta/2-\delta/2)}\cos(\gamma/2)$ and $b = e^{i(\alpha+\beta/2-\delta/2)}\sin(\gamma/2)$. Matching this to $R_z(\beta)R_y(\gamma)R_z(\delta)$ computed explicitly via Eq. (24) confirms the decomposition for all $U \in SU(2)$. ◊ ◊

A hardware-convenient parameterization for superconducting qubits is

$$U(\theta,\phi,\lambda) = R_z(\phi)\, R_x\!\left(-\tfrac{\pi}{2}\right) R_z(\theta)\, R_x\!\left(\tfrac{\pi}{2}\right) R_z(\lambda). \tag{26}$$

C. Two-Qubit Gates and Entanglement Generation

Product operators $U_A \otimes U_B$ cannot generate entanglement. The canonical entangling gates are:

$$CX = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X, \tag{27}$$
$$CZ = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes Z. \tag{28}$$

CX flips the target conditioned on the control being $|1\rangle$. Applying CX to $(H|0\rangle)\otimes|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\otimes|0\rangle$:

$$CX \cdot \tfrac{1}{\sqrt{2}}(|00\rangle + |10\rangle) = \tfrac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = |\Phi^+\rangle, \tag{29}$$

demonstrating Bell-state generation from a product state.

D. Universality of Quantum Gates

Theorem (Universality). *A gate set $\mathcal{G}$ is* universal for quantum computation *if any n-qubit unitary can be approximated to precision ε by a circuit of size $O\left(\mathrm{poly}\left(\log\frac{1}{\varepsilon}\right)\right)$ from $\mathcal{G}$. The set $\{H, T, CX\}$ is universal [2].*

Sketch of proof. The argument proceeds in two steps. (1) *Single-qubit universality*: H and T together generate a dense subgroup of $SU(2)$—this can be shown by computing HTH and checking that the resulting rotations are irrational multiples of π, ensuring density. (2) *Multi-qubit reduction*: any n-qubit unitary can be decomposed into two-qubit unitaries via a sequence of controlled-U gates, each decomposable into CX and single-qubit gates (Gray-code argument). The Solovay-Kitaev theorem then guarantees that approximation within ε requires only $O(\mathrm{poly}\log(1/\varepsilon))$ gates. ◊ ◊

E. No-Cloning Theorem

Theorem (No-Cloning, Wootters & Zurek 1982). *There is no unitary operation U such that $U|\psi\rangle|0\rangle = |\psi\rangle|\psi\rangle$ for all $|\psi\rangle$ [8].*

Proof. Suppose such U exists and let $|\psi\rangle, |\phi\rangle$ be arbitrary. Then $U|\psi\rangle|0\rangle = |\psi\rangle|\psi\rangle$ and $U|\phi\rangle|0\rangle = |\phi\rangle|\phi\rangle$. Taking the inner product of both equations and using unitarity ($U^\dagger U = I$):

$$\langle\psi|\phi\rangle \underbrace{\langle 0|0\rangle}_{=1} = \langle\psi|\phi\rangle\langle\psi|\phi\rangle = \langle\psi|\phi\rangle^2. \tag{30}$$

Hence $\langle\psi|\phi\rangle(\langle\psi|\phi\rangle - 1) = 0$, forcing $\langle\psi|\phi\rangle \in \{0, 1\}$. Any two states that U clones must be either identical or orthogonal, contradicting the assumption that U clones *all* states. ◊

a. Implication for QML. No-cloning forbids direct copying of unknown quantum data, restricting classical-style data augmentation. It simultaneously motivates quantum-native protocols (superdense coding, teleportation) that exploit entanglement instead of copying.

F. Basic Quantum Protocols

a. Superdense coding. By sharing the Bell pair $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ in advance, Alice can transmit *2 classical bits* to Bob by sending only a *single qubit* [9]. By applying a local unitary $X^a Z^b$ ($a, b \in \{0, 1\}$) to her qubit, Alice can transform the globally shared entangled state into any of the four orthogonal Bell states. Bob then performs a Bell-basis measurement to distinguish these states with 100% certainty.

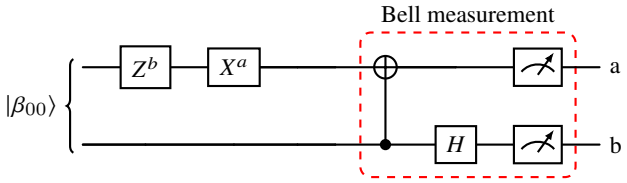

FIG. 1. Superdense coding circuit. Alice applies local operators Z^b and X^a to her half of a shared Bell pair (q_A). She then sends q_A to Bob, who performs a Bell-basis measurement (CNOT and H) to recover both bits a, b.

The protocol succeeds because the set $\{(E_i \otimes I)|\Phi^+\rangle\}$ forms an orthonormal basis for the four-dimensional Hilbert space $(\mathbb{C}^2)^{\otimes 2}$. Alice effectively picks which basis vector the system resides in, and Bob performs the change-of-basis back to the computational basis for sampling.

b. Quantum teleportation. Alice transmits an unknown qubit $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ to Bob using only a shared Bell pair and two classical bits [10]. The circuit (Fig. 2) evolves as follows. The three-qubit initial state is

$$|\psi\rangle \otimes |\Phi^+\rangle = \tfrac{1}{\sqrt{2}}(\alpha|0\rangle + \beta|1\rangle)(|00\rangle + |11\rangle). \tag{31}$$

After Alice applies CX then H on her qubits, the state becomes

$$\tfrac{1}{2}\big[\,|00\rangle(\alpha|0\rangle + \beta|1\rangle) + |01\rangle(\alpha|1\rangle + \beta|0\rangle) + |10\rangle(\alpha|0\rangle - \beta|1\rangle) + |11\rangle(\alpha|1\rangle - \beta|0\rangle)\big]. \tag{32}$$

TABLE I. Alice's encoding and resulting Bell states for Superdense Coding. The bit b (determined by the wire with the H gate) controls the Z gate.

Requested Bits (ba)	Alice's Gate Logic ($X^a Z^b$)	Resulting Bell State (on wire A)	Bob's Measurement Result (ba)
00	I	$\|\Phi^+\rangle = \frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle)$	00
01	X	$\|\Psi^+\rangle = \frac{1}{\sqrt{2}}(\|10\rangle + \|01\rangle)$	01
10	Z	$\|\Phi^-\rangle = \frac{1}{\sqrt{2}}(\|00\rangle - \|11\rangle)$	10
11	XZ	$\|\Psi^-\rangle = \frac{1}{\sqrt{2}}(\|10\rangle - \|01\rangle)$	11

TABLE II. Alice's measurement outcomes and Bob's corresponding recovery gates. The bit a (from the wire with H) controls the Z gate.

Outcome (ba)	Alice's Projector	Bob's Received State	Correction Logic ($Z^a X^b$)
00	$\|\Phi^+\rangle\langle\Phi^+\|$	$\alpha\|0\rangle + \beta\|1\rangle$	I
10	$\|\Psi^+\rangle\langle\Psi^+\|$	$\alpha\|1\rangle + \beta\|0\rangle$	X
01	$\|\Phi^-\rangle\langle\Phi^-\|$	$\alpha\|0\rangle - \beta\|1\rangle$	Z
11	$\|\Psi^-\rangle\langle\Psi^-\|$	$\alpha\|1\rangle - \beta\|0\rangle$	ZX

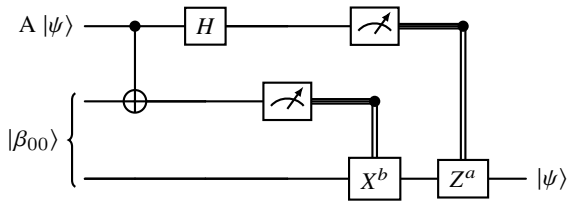

FIG. 2. Quantum teleportation circuit. Alice performs a Bell measurement on her state A and her half of a shared Bell pair A. The outcomes a and b are used to apply gates X^b and Z^a to Bob's qubit B, recovering $|\psi\rangle$.

Alice measures qubits 1 and 2 and sends outcomes (a, b) classically. Bob applies $Z^a X^b$ to recover $\alpha|0\rangle + \beta|1\rangle = |\psi\rangle$ exactly.

Importantly, the "teleportation" is not instantaneous. Although the system collapses into one of the four states in Table II immediately upon Alice's measurement, Bob cannot know which gate to apply without the two classical bits. This preserves causality and satisfies the *no-communication theorem*: entanglement alone cannot transmit information faster than light.

VI. REVERSIBLE COMPUTATION AND COMPLEXITY

A. Landauer's Principle

Landauer's Principle (1961). *Erasing one bit of information irreversibly dissipates at least $k_B T \ln 2$ of energy as heat, where k_B is Boltzmann's constant and T is the temperature of the environment [11].*

Since NAND (and most classical gates) are many-to-one, they are thermodynamically irreversible and incur this minimum energy cost. In contrast, unitary quantum evolution is always reversible ($U^{-1} = U^\dagger$), satisfying Landauer's bound with equality in principle.

B. Toffoli Gate: Universal Reversible Classical Gate

The Toffoli (CCX) gate maps $(a, b, c) \mapsto (a, b, c \oplus ab)$. It is universal for reversible classical computation: both NAND and FANOUT can be simulated using Toffoli gates with ancilla bits. As a quantum gate on three qubits, it is a 8×8 permutation matrix.

a. Quantum oracle for arbitrary functions. Any $f : \{0,1\}^n \to \{0,1\}^m$ can be embedded into the reversible unitary

$$U_f : |x\rangle|y\rangle \mapsto |x\rangle|y \oplus f(x)\rangle, \tag{33}$$

where $\oplus$ denotes bitwise XOR. Applying U_f to a uniform superposition:

$$U_f\left(\frac{1}{\sqrt{2^n}}\sum_x |x\rangle\right)|0\rangle = \frac{1}{\sqrt{2^n}}\sum_x |x\rangle|f(x)\rangle, \tag{34}$$

evaluating f on all 2^n inputs simultaneously—the key mechanism underlying quantum speedups via interference.

C. Gate Decompositions and Computational Cost

In practice, we do not have direct access to arbitrary multi-qubit gates. Complex gates must be decomposed into a set of "native" gates supported by the specific quantum hardware. Typically, this universal set consists of single-qubit rotations and a two-qubit entangling gate such as CNOT (CX) or CZ. The overall computational cost is heavily influenced by the number of required two-qubit gates.

a. SWAP Gate Decomposition. The SWAP gate, which exchanges the quantum states of two qubits, can be elegantly decomposed into three consecutive CNOT gates alternating their control and target qubits: This identity is easily verified by

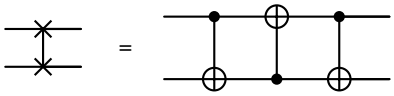

FIG. 3. Decomposition of the SWAP gate into three CNOT gates.

evaluating the action of the three CNOTs on the computational basis states $|00\rangle, |01\rangle, |10\rangle, |11\rangle$.

b. Toffoli (CCX) Gate Decomposition. The Toffoli gate is a three-qubit controlled-controlled-NOT gate. While universal for reversible classical computation, it is not native to most quantum hardware and requires decomposition. A standard decomposition into the Clifford+T gate set requires 6 CNOT gates and several single-qubit T, $T^\dagger$, and H gates: Because

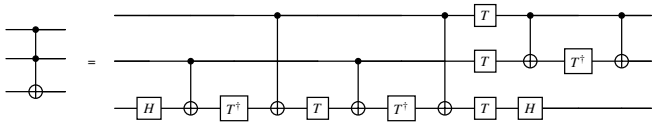

FIG. 4. Decomposition of the Toffoli (CCX) gate using 6 CNOT gates and single-qubit operations.

the Toffoli gate is foundational, this 6-CNOT cost serves as a critical benchmark for the overhead in fault-tolerant quantum algorithms.

c. Fredkin (CSWAP) Gate Decomposition. The Fredkin gate is a controlled-SWAP gate. If the control qubit is $|1\rangle$, it swaps the two target qubits. It is heavily utilized in algorithms evaluating state overlaps, such as the swap test. Using the fact that a SWAP gate is three CNOTs, a CSWAP can be decomposed into two CNOTs and one Toffoli gate: Given that

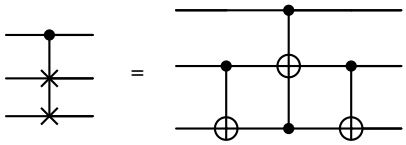

FIG. 5. Decomposition of the Fredkin (CSWAP) gate into two CNOT gates and one Toffoli gate.

a Toffoli gate costs 6 CNOTs, the CSWAP gate requires an overall cost of 8 CNOT gates.

D. Hadamard Test and Swap Test

a. Hadamard test. Given a unitary U and a state $|\psi\rangle$, the Hadamard test (Fig. 6) estimates $\langle\psi|U|\psi\rangle \in \mathbb{C}$. A full step-by-step derivation follows.

Complete derivation of Hadamard test. Step 1. Initialize ancilla and system: $|0\rangle|\psi\rangle$.

Step 2. Apply H to ancilla: $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|\psi\rangle$.

Step 3. Apply controlled-U (if ancilla is $|1\rangle$, apply U to system):

$$\tfrac{1}{\sqrt{2}}\big(|0\rangle|\psi\rangle + |1\rangle U|\psi\rangle\big). \tag{35}$$

Step 4. Apply final H to ancilla:

$$\begin{aligned}&\tfrac{1}{2}\big[(|0\rangle + |1\rangle)|\psi\rangle + (|0\rangle - |1\rangle)U|\psi\rangle\big]\\ &= \tfrac{1}{2}|0\rangle(|\psi\rangle + U|\psi\rangle) + \tfrac{1}{2}|1\rangle(|\psi\rangle - U|\psi\rangle).\end{aligned} \tag{36}$$

Step 5. Measure ancilla. The outcome probabilities are:

$$P(0) = \tfrac{1}{4}\||\psi\rangle + U|\psi\rangle\|^2 = \tfrac{1}{4}(2 + 2\,\mathrm{Re}\,\langle\psi|U|\psi\rangle), \tag{37}$$

$$P(1) = \tfrac{1}{4}(2 - 2\,\mathrm{Re}\,\langle\psi|U|\psi\rangle). \tag{38}$$

Hence $P(0) - P(1) = \mathrm{Re}\,\langle\psi|U|\psi\rangle$.

Imaginary part. Insert an $S^\dagger$ phase gate before the final H; a parallel argument gives $P(0) - P(1) = \mathrm{Im}\,\langle\psi|U|\psi\rangle$. Together, both runs fully specify $\langle\psi|U|\psi\rangle \in \mathbb{C}$. ◊ ◊

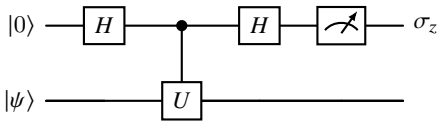

FIG. 6. Hadamard test circuit. The ancilla qubit $|0\rangle$ controls the unitary U on the system register $|\psi\rangle$. A final Hadamard on the ancilla followed by σ_z measurement yields $P(0) - P(1) = \mathrm{Re}\,\langle\psi|U|\psi\rangle$. Inserting $S^\dagger$ before the final H yields the imaginary part.

b. Swap test. Setting $U = \mathrm{SWAP}$ and $|\psi\rangle = |\psi_1\rangle|\psi_2\rangle$, the Hadamard test gives

$$P(0) - P(1) = \langle\psi_1|\langle\psi_2|\,\mathrm{SWAP}\,|\psi_1\rangle|\psi_2\rangle = |\langle\psi_1|\psi_2\rangle|^2. \tag{39}$$

The identity $\langle\psi_1|\langle\psi_2|\,\mathrm{SWAP}(|\psi_1\rangle, |\psi_2\rangle) = (\langle\psi_1| \otimes \langle\psi_2|)(|\psi_2\rangle \otimes |\psi_1\rangle) = |\langle\psi_1|\psi_2\rangle|^2$ follows directly from the action of SWAP.

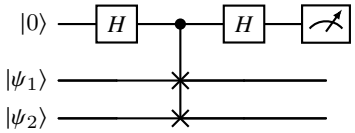

FIG. 7. Swap test circuit. A controlled-SWAP between $|\psi_1\rangle$ and $|\psi_2\rangle$ is mediated by the ancilla qubit. The outcome probability difference satisfies $P(0) - P(1) = |\langle\psi_1|\psi_2\rangle|^2$, providing an efficient quantum estimate of state overlap used in kernel-based QML [12].

E. Classical Complexity Classes

To understand the potential advantages of quantum computing, it is essential to review the foundations of classical complexity theory. Complexity theory categorizes problems based on the resources (time or space) required to solve them.

a. Decision Problems. A decision problem asks a yes-or-no question. Many complex optimization problems (e.g., "minimize $f(x)$") can be recast as decision problems (e.g., "is there an x such that $f(x) \leq L$?").

b. The Classes P and NP.

- **P** (Polynomial time): The class of decision problems that can be solved efficiently by a deterministic classical computer in polynomial time $O(n^k)$, where n is the input size. Examples include sorting or finding a substring.

- **NP** (Nondeterministic Polynomial time): The class of decision problems for which a "yes" instance has a proof (or witness) that can be verified efficiently in polynomial time. For example, finding a subset of numbers that sums to zero is difficult, but verifying a proposed subset takes only linear time.

While clearly $\mathrm{P} \subseteq \mathrm{NP}$, determining whether $\mathrm{P} = \mathrm{NP}$ is widely considered the most important open question in theoretical computer science.

c. NP-Hard and NP-Complete. A problem is **NP-hard** if every problem in NP can be reduced to it in polynomial time. It is **NP-complete** if it is both NP-hard and in NP. Classic NP-complete problems include Boolean satisfiability (SAT) and graph coloring.

F. Quantum Complexity Classes

- **BQP** (Bounded-error Quantum Polynomial time): decision problems solvable by a polynomial-time quantum circuit with error $\leq 1/3$. $\mathrm{BPP} \subseteq \mathrm{BQP} \subseteq \mathrm{PSPACE}$.
- **QMA** (Quantum Merlin-Arthur): the quantum analogue of NP; YES instances have a polynomial-size quantum witness verifiable by a BQP verifier.

Whether $\mathrm{BQP} \not\subseteq \mathrm{NP}$ or $\mathrm{BQP} \neq \mathrm{BPP}$ remains open [6]. The believed picture is $\mathrm{BPP} \subsetneq \mathrm{BQP}$, with factoring as a candidate separation (Shor's algorithm).

VII. QUERY MODEL OF COMPUTATION AND ALGORITHMS

A. Query Model of Computation

One of the potential advantages of quantum computers is to provide faster solutions to computational problems. To rigorously analyze computational complexity and demonstrate quantum advantages, we introduce the *query model of computation*. In this model, the input is provided via an oracle or black box that evaluates a Boolean function $f : \{0,1\}^n \to \{0,1\}$. The complexity of an algorithm is measured by the number of queries made to the oracle.

Classically, the oracle is evaluated as $x \mapsto f(x)$. In the quantum query model, the oracle is represented as a unitary operator U_f that acts on a superposition of states:

$$U_f \sum_{x\in\{0,1\}^n} \alpha_x |x\rangle |y\rangle = \sum_{x\in\{0,1\}^n} \alpha_x |x\rangle |y \oplus f(x)\rangle . \quad (40)$$

This allows the quantum computer to query all possible inputs simultaneously, leveraging quantum parallelism.

B. The Phase Kick-back Trick

A recurring and foundational principle in many quantum algorithms (such as Deutsch-Jozsa, Simon's, Grover's, and Phase Estimation) is the *phase kick-back* trick. In classical reversible computation, an oracle U_f evaluates a Boolean function $f(x)$ by flipping a target bit y based on the result: $U_f |x\rangle |y\rangle = |x\rangle |y \oplus f(x)\rangle$.

However, in quantum computation, if we prepare the target register y in the specific superposition state $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, applying the oracle yields a remarkable effect. Consider the action of U_f on a single basis state $|x\rangle$ and the target $|-\rangle$:

$$\begin{aligned} U_f |x\rangle |-\rangle &= U_f \left(|x\rangle \otimes \frac{|0\rangle - |1\rangle}{\sqrt{2}} \right) \\ &= \frac{1}{\sqrt{2}} \left(U_f |x\rangle |0\rangle - U_f |x\rangle |1\rangle \right) \\ &= \frac{1}{\sqrt{2}} \left(|x\rangle |f(x)\rangle - |x\rangle |1 \oplus f(x)\rangle \right) . \end{aligned} \quad (41)$$

If $f(x) = 0$, the state becomes $|x\rangle \frac{|0\rangle - |1\rangle}{\sqrt{2}} = |x\rangle |-\rangle$. If $f(x) = 1$, the state becomes $|x\rangle \frac{|1\rangle - |0\rangle}{\sqrt{2}} = -|x\rangle |-\rangle$. We can succinctly write this as:

$$U_f |x\rangle |-\rangle = (-1)^{f(x)} |x\rangle |-\rangle . \quad (42)$$

Notice that the target state $|-\rangle$ remains entirely unchanged. Instead, the function evaluation $f(x)$ is "kicked back" as a global phase factor $(-1)^{f(x)}$ onto the control register $|x\rangle$. When $|x\rangle$ is in a superposition, these relative phases create destructive and constructive interference patterns, which quantum algorithms exploit to efficiently extract global properties of f.

More generally, if a unitary operator U has an eigenvector $|\psi\rangle$ with eigenvalue $e^{i\theta}$, a controlled-U gate targeting $|\psi\rangle$ will kick back the phase $e^{i\theta}$ to the control qubit. This generalized phase kick-back is the core mechanism behind the Quantum Phase Estimation algorithm.

C. Quantum Query Algorithms

Several foundational algorithms demonstrate the power of the quantum query model, exhibiting separations in complexity between classical and quantum computation.

a. Deutsch-Jozsa Algorithm. The Deutsch-Jozsa algorithm solves a specialized problem: given a boolean function $f : \{0,1\}^n \to \{0,1\}$ that is promised to be either constant (same output for all inputs) or balanced (outputs 0 for exactly half of the inputs and 1 for the other half), determine which it is. Classically, this requires $O(2^{n-1} + 1)$ queries in the worst case. Quantumly, it requires only a single query. The quantum circuit is as follows:

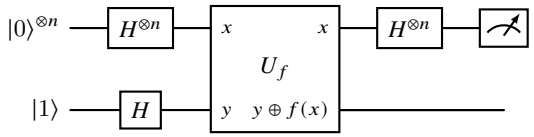

FIG. 8. Quantum circuit for the Deutsch-Jozsa and Bernstein-Vazirani algorithms.

The algorithm proceeds by initializing the system to $|0\rangle^{\otimes n} |1\rangle$ and applying Hadamard gates to all qubits, preparing the state:

$$|\psi_1\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} |x\rangle \otimes |-\rangle . \quad (43)$$

Querying the oracle U_f applies a phase kick-back effect, encoding $f(x)$ into the amplitude:

$$|\psi_2\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} (-1)^{f(x)} |x\rangle \otimes |-\rangle . \quad (44)$$

Applying the final n-qubit Hadamard gate $H^{\otimes n}$ transforms the state using the identity $H^{\otimes n} |x\rangle = \frac{1}{\sqrt{2^n}} \sum_z (-1)^{x\cdot z} |z\rangle$:

$$|\psi_3\rangle = \frac{1}{2^n} \sum_{x\in\{0,1\}^n} \sum_{z\in\{0,1\}^n} (-1)^{x\cdot z + f(x)} |z\rangle \otimes |-\rangle . \quad (45)$$

If f is constant, the amplitude of $|z = 0\rangle$ is ± 1, so measuring the first register yields 0^n with probability 1. If f is balanced, the amplitude of $|z = 0\rangle$ evaluates to $\sum_x (-1)^{f(x)}/2^n = 0$. Hence, a single measurement perfectly distinguishes the two cases.

b. Bernstein-Vazirani Algorithm. This algorithm finds a hidden n-bit string a for a linear function $f(x) = a \cdot x$. Classically, identifying a requires $O(n)$ queries by probing each

basis vector. Quantumly, the identical circuit (Fig. 8) finds a in 1 query. Substituting $f(x) = a \cdot x$ into $|\psi_3\rangle$, we get:

$$|\psi_3\rangle = \frac{1}{2^n} \sum_{x\in\{0,1\}^n} \sum_{z\in\{0,1\}^n} (-1)^{(a+z)\cdot x} |z\rangle \otimes |-\rangle . \quad (46)$$

The amplitude is non-zero if and only if $z = a$, yielding the state $|a\rangle |-\rangle$. Thus, a single measurement reveals the hidden string a exactly.

c. Simon's Algorithm. Simon's algorithm solves a problem with an exponential separation between quantum and randomized classical algorithms. We are given $f : \{0,1\}^n \to \{0,1\}^m$, promised that there is a hidden string s such that $f(x) = f(y)$ if and only if $x = y \oplus s$ or $x = y$. Classically, finding s requires $\Omega(2^{n/2})$ queries. The quantum algorithm finds s in $O(n)$ queries.

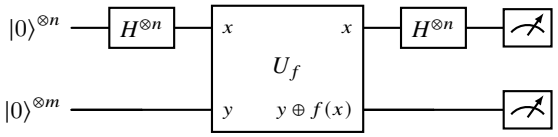

FIG. 9. Quantum circuit for Simon's algorithm.

The state before the oracle is $|\psi_1\rangle = \frac{1}{\sqrt{2^n}} \sum_x |x\rangle |0\rangle^{\otimes m}$. The oracle yields $|\psi_2\rangle = \frac{1}{\sqrt{2^n}} \sum_x |x\rangle |f(x)\rangle$. Measuring the second register gives some $f(x)$ and collapses the first register into an equal superposition of the two pre-images:

$$|\psi_2'\rangle = \frac{1}{\sqrt{2}}(|x\rangle + |x \oplus s\rangle) \otimes |f(x)\rangle . \quad (47)$$

Applying the final $H^{\otimes n}$ to the first register yields:

$$\begin{aligned} |\psi_3\rangle &= \frac{1}{\sqrt{2^{n+1}}} \sum_{z\in\{0,1\}^n} \left[(-1)^{x\cdot z} + (-1)^{(x\oplus s)\cdot z}\right] |z\rangle \\ &= \frac{1}{\sqrt{2^{n+1}}} \sum_{z\in\{0,1\}^n} (-1)^{x\cdot z} \left[1 + (-1)^{s\cdot z}\right] |z\rangle . \end{aligned} \quad (48)$$

The amplitude is non-zero only when $s \cdot z = 0 \pmod 2$. By repeating the algorithm $O(n)$ times, we obtain a system of linear equations $z_i \cdot s = 0$, which can be solved efficiently on a classical computer to find s.

VIII. QUANTUM FOURIER TRANSFORM AND PHASE ESTIMATION

A. Quantum Fourier Transform (QFT)

The Fourier Transform is a fundamental mathematical tool for transforming signals between domains. The Discrete Fourier Transform (DFT) maps a vector $(x_0, \ldots, x_{N-1})$ to $(y_0, \ldots, y_{N-1})$ where $y_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_j \exp(2\pi i jk/N)$. This requires $O(N \log N)$ operations classically using the Fast Fourier Transform (FFT).

The *Quantum Fourier Transform* (QFT) is the quantum analogue, performing the transformation on the amplitudes of a quantum state. For an n-qubit system with $N = 2^n$ basis states, the QFT is defined as:

$$\mathrm{QFT}\,|j\rangle = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{2\pi i jk/N} |k\rangle . \quad (49)$$

Since it is unitary, the inverse QFT ($\mathrm{QFT}^\dagger$) is straightforwardly defined by taking the complex conjugate of the phase.

a. Product State Representation. To build the quantum circuit, it is highly instructive to rewrite the output state in a factored tensor product form. Let the integer j be represented in binary as $j = j_1 j_2 \ldots j_n$, and define the binary fraction $0.j_l \ldots j_n = \sum_{m=1}^{n-l+1} j_{l+m-1} 2^{-m}$. The QFT transformation can be elegantly factored as:

$$\begin{aligned} \mathrm{QFT}\,|j_1 j_2 \ldots j_n\rangle = \frac{1}{\sqrt{2^n}} &\left(|0\rangle + e^{2\pi i 0.j_n} |1\rangle\right) \\ &\otimes \left(|0\rangle + e^{2\pi i 0.j_{n-1} j_n} |1\rangle\right) \\ &\otimes \cdots \otimes \left(|0\rangle + e^{2\pi i 0.j_1 j_2 \ldots j_n} |1\rangle\right) . \end{aligned} \quad (50)$$

b. Quantum Circuit for QFT. This product form implies that the QFT can be implemented using single-qubit Hadamard gates (H) and two-qubit controlled-phase gates (R_k), where $R_k = \mathrm{diag}(1, e^{2\pi i/2^k})$. The Hadamard gate creates the equal superposition and the leading phase bit, while the controlled-R_k gates incrementally add the fractional phase contributions from the other qubits. A standard circuit for a 3-qubit QFT is shown below. Note that the output qubits are in reverse order (k_3, k_2, k_1), requiring a final set of SWAP gates if the original ordering is strictly needed.

The QFT requires $O(n^2) = O(\log^2 N)$ gates, offering an exponential speedup over the classical FFT's $O(N \log N)$ operations.

B. Quantum Phase Estimation (QPE)

Quantum Phase Estimation is one of the most important subroutines in quantum computing, forming the basis for Shor's algorithm (integer factoring) and the HHL algorithm (solving linear systems). Given a unitary operator U and its eigenvector $|\psi\rangle$ with an unknown eigenvalue $e^{2\pi i\theta}$, the goal of QPE is to estimate the phase $\theta \in [0, 1)$.

a. Algorithm Steps. The QPE algorithm uses two registers. The first (estimation) register contains t qubits initialized to $|0\rangle^{\otimes t}$, where t determines the precision of the phase estimate. The second (target) register is initialized to the eigenvector $|\psi\rangle$. The algorithm proceeds in three main steps:

1. **Superposition**: Apply Hadamard gates to all t qubits in the first register to create an equal superposition of all computational basis states.

2. **Controlled Unitaries**: Apply a sequence of controlled-U^{2^k} operations, controlled by the k-th qubit of the first register and targeting the second register. Because $U|\psi\rangle = e^{2\pi i\theta}|\psi\rangle$, applying U^{2^k} kicks back a phase of $e^{2\pi i\theta 2^k}$ to the control qubit. After all controlled unitaries, the state of the first register is:

$$\frac{1}{\sqrt{2^t}} \sum_{k=0}^{2^t-1} e^{2\pi i\theta k} |k\rangle . \quad (51)$$

3. **Inverse QFT**: The state of the first register is precisely the Quantum Fourier Transform of the state $|2^t\theta\rangle$. By applying the Inverse QFT ($\mathrm{QFT}^\dagger$) to the first register, the state transforms into the basis state $|2^t\theta\rangle$. Measuring the first register yields a binary string representing the best t-bit approximation of the phase θ.

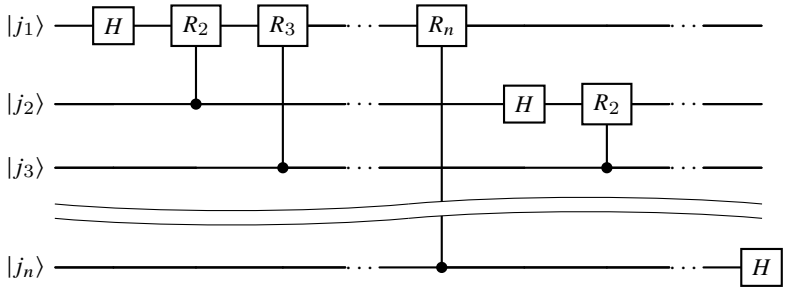

FIG. 10. Quantum circuit for an n-qubit Quantum Fourier Transform. SWAP gates at the end are omitted for brevity.

b. Quantum Circuit for QPE. The schematic circuit for Quantum Phase Estimation is illustrated below:

If the true phase θ cannot be exactly represented with t bits, the measured result will be the closest t-bit string with high probability. The required number of qubits t is chosen based on the desired accuracy and the allowed failure probability.

IX. TOWARDS QUANTUM MACHINE LEARNING

A. Quantum Advantage in Machine Learning

Quantum machine learning (QML) investigates whether quantum computers can accelerate or qualitatively improve machine learning tasks [13]. Three criteria define meaningful quantum advantage:

1. *Physical feasibility* (with or without QEC),
2. *Computational advantage* over the best known classical algorithm,
3. *Industry relevance* for real-world ML tasks.

Meeting all three simultaneously is an open problem. Quantum speedups in supervised learning often rely on quantum RAM (QRAM) for efficient data loading—a hardware resource whose practical feasibility remains uncertain.

B. Quantum linear systems: the HHL algorithm

Many QML constructions reduce training or inference to solving a linear system $A\mathbf{x} = \mathbf{b}$. The quantum algorithm of Harrow, Hassidim, and Lloyd (HHL) [7] targets this problem when $A \in \mathbb{C}^{N\times N}$ is Hermitian and *s-sparse* (at most s nonzero entries per row). The idealized output is a normalized quantum state

$$|x\rangle \;\propto\; A^{-1}\,|b\rangle\,, \tag{52}$$

where $|b\rangle$ encodes $\mathbf{b}$ in an amplitude vector of dimension $N = 2^n$. Typical classical solvers for sparse, well-conditioned systems (such as conjugate gradient) cost

$$T_{\mathrm{cl}} \;=\; \tilde{O}\big(Ns\,\kappa\log(1/\varepsilon)\big) \tag{53}$$

up to logarithmic factors, where $\kappa = \lambda_{\max}/\lambda_{\min}$ is the spectral condition number and ε the accuracy. Under an ideal matrix-entry query model and QRAM-like state preparation, HHL achieves

$$T_{\mathrm{q}} \;=\; \mathrm{poly}\big(\log N,\, s,\, \kappa,\, 1/\varepsilon\big), \tag{54}$$

i.e., polynomial time in $\log N$ rather than in the dimension N [5, 7].

a. Spectral reduction. For Hermitian invertible A,

$$A = \sum_{j=0}^{N-1} \lambda_j\,|u_j\rangle\,\langle u_j|\,, \tag{55}$$

$$A^{-1} = \sum_{j=0}^{N-1} \lambda_j^{-1}\,|u_j\rangle\,\langle u_j|\,. \tag{56}$$

Writing $|b\rangle = \sum_j \beta_j\,|u_j\rangle$,

$$A^{-1}\,|b\rangle = \sum_{j=0}^{N-1} \frac{\beta_j}{\lambda_j}\,|u_j\rangle\,. \tag{57}$$

HHL prepares (57) coherently. Quantum phase estimation (Sec. VIII B) with $U = e^{iAt}$ entangles each $|u_j\rangle$ with a binary encoding of λ_j. Let $C > 0$ satisfy $|C/\lambda_j| \le 1$ on the relevant eigenvalue support. A single-qubit rotation on an *ancilla*, controlled on that encoding, can load amplitudes proportional to C/λ_j. After inverting the QPE unitary, unmeasured data registers hold a coherent superposition whose renormalization is the target (52).

b. Non-Hermitian systems. If A is square but not Hermitian, embed into the Hermitian dilation

$$\tilde{A} = \begin{pmatrix} 0 & A \\ A^{\dagger} & 0 \end{pmatrix}. \tag{58}$$

Then, with block vectors $\mathbf{0}, \mathbf{x}, \mathbf{b} \in \mathbb{C}^N$,

$$\tilde{A}\begin{pmatrix}\mathbf{0}\\ \mathbf{x}\end{pmatrix} = \begin{pmatrix} A\mathbf{x}\\ \mathbf{0}\end{pmatrix} = \begin{pmatrix}\mathbf{b}\\ \mathbf{0}\end{pmatrix} \quad\Longleftrightarrow\quad A\mathbf{x} = \mathbf{b}. \tag{59}$$

c. Circuit outline (conceptual). Figure 12 summarizes the core flow with wires q_1 (data), q_2 (clock), q_3 (ancilla), matching the lecture slides:

1. Prepare the amplitude-encoded state $|b\rangle = \sum_j \beta_j\,|u_j\rangle$.
2. Apply QPE with $U = e^{iAt}$ so that

$$|b\rangle \;\mapsto\; \sum_j \beta_j\,|u_j\rangle\,|\tilde{\lambda}_j\rangle\,, \tag{60}$$

   where $|\tilde{\lambda}_j\rangle$ is a t-bit approximation of λ_j.
3. **Conditioned rotation on q_3.** Work in the computational basis of the ancilla and adopt the usual y-rotation

$$R_y(\theta) \;=\; e^{-i\theta Y/2} \;=\; \begin{pmatrix}\cos(\theta/2) & -\sin(\theta/2)\\ \sin(\theta/2) & \cos(\theta/2)\end{pmatrix}, \qquad Y = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}. \tag{61}$$

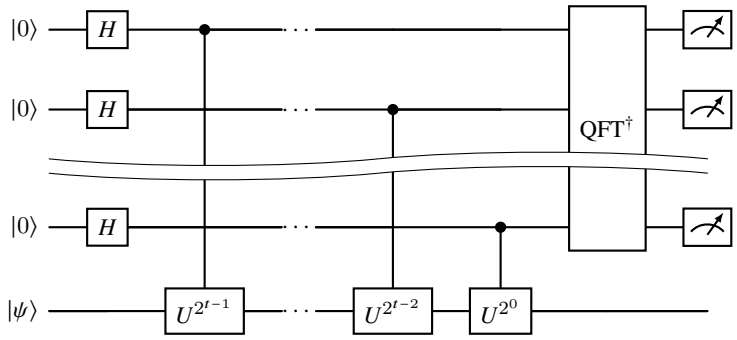

FIG. 11. Quantum circuit for the Quantum Phase Estimation algorithm using an arbitrary t-qubit precision register.

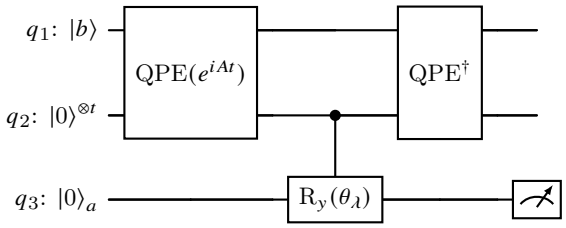

FIG. 12. Schematic HHL subroutine in the same qubit stacking as the course slides (top: amplitude loading of $|b\rangle$ on q_1, middle: t-qubit phase register q_2, bottom: workspace ancilla q_3 for eigenvalue inversion). The QPE(e^{iAt}) block entangles q_1 (eigenstate register) with the encoded eigenvalues on q_2 (cf. Fig. 11; bit order within q_2 follows that figure's QFT convention). The gate $R_y(\theta_\lambda)$ on q_3 is controlled by the phase register q_2 (filled control dot). After QPE† clears q_2, post-selecting q_3 on $|1\rangle$ yields $q_1 \propto A^{-1}|b\rangle$ [7].

Hence $R_y(\theta)|0\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle$. Fix a scale $C > 0$ such that $|C/\lambda_j| \le 1$ on every eigenvalue that receives amplitude from $|b\rangle$ (for signed λ_j, one first infers $|\lambda_j|$ coherently and absorbs any phase into an additional z-type rotation—the idealized positive case suffices for the amplitude logic). Define

$$\theta_j \;:=\; 2\arcsin\!\left(\frac{C}{\lambda_j}\right), \tag{62}$$

so that the *uncontrolled* rotation would map

$$R_y(\theta_j)|0\rangle = \sqrt{1-\frac{C^2}{\lambda_j^2}}\;|0\rangle + \frac{C}{\lambda_j}\;|1\rangle\,. \tag{63}$$

The controlled gate applies $R_y(\theta_j)$ (or the QPE-limited analogue with $\tilde{\lambda}_j$) on q_3 whenever q_2 encodes branch j.

4. **Inverse QPE.** Applying QPE† disentangles the clock, leaving

$$|\Psi_{\rm pre}\rangle = \sum_{j=0}^{N-1}\beta_j\,|u_j\rangle_{q_1}\otimes|0\rangle_{q_2}^{\otimes t}\otimes\left(\sqrt{1-\frac{C^2}{\lambda_j^2}}\;|0\rangle_{q_3}+\frac{C}{\lambda_j}\;|1\rangle_{q_3}\right). \tag{64}$$

(Again, replace $\lambda_j \to \tilde{\lambda}_j$ to track finite-precision QPE.)

5. **Post-select $|1\rangle$ on q_3.** Measuring q_3 in the computational basis gives outcome probabilities

$$\Pr(\text{outcome } 1) = \sum_{j=0}^{N-1}|\beta_j|^2\,\frac{C^2}{\lambda_j^2} \;=:\; p_{\rm succ}, \tag{65}$$

$$\Pr(\text{outcome } 0) = \sum_{j=0}^{N-1}|\beta_j|^2\left(1-\frac{C^2}{\lambda_j^2}\right). \tag{66}$$

The joint Born weight on "1" for branch j is the product of $|\beta_j|^2$ from q_1 and $\sin^2(\theta_j/2) = C^2/\lambda_j^2$ from (63). Conditioning on outcome 1 collapses q_1 to

$$|x_{\rm post}\rangle = \frac{1}{\sqrt{p_{\rm succ}}}\sum_{j=0}^{N-1}\beta_j\,\frac{C}{\lambda_j}\;|u_j\rangle = \frac{C}{\sqrt{p_{\rm succ}}}\,A^{-1}|b\rangle\,, \tag{67}$$

which is exactly the target direction (57), up to the known global factor $C/\sqrt{p_{\rm succ}}$. Large κ makes some $|\lambda_j|^{-2}$ terms dominate (65), so naive post-selection can cost $\Omega(\kappa^{-2})$ attempts; amplitude amplification reduces the query overhead to $\tilde{O}(\kappa)$ in the analysis of Ref. [7].

The procedure assumes efficient oracles for preparing $|b\rangle$ and simulating e^{-iAt} (Hamiltonian simulation for sparse Hermitian A). The speedup is stated relative to this *quantum input model*; quantitative comparisons to classical numerical linear algebra require care [5].

C. Linear classifiers and support vector machines

Binary classifiers often seek a hyperplane in feature space $\mathbb{R}^d$ that splits two classes. Any affine hyperplane can be written as $H = \{\mathbf{x} : \langle \mathbf{w}, \mathbf{x}\rangle + b = 0\}$ with normal $\mathbf{w} \in \mathbb{R}^d$ and offset $b \in \mathbb{R}$. The *linear score* or *logit*

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x}\rangle + b \tag{68}$$

is positive on one side of H (predict class +1) and negative on the other (predict −1). For labeled training pairs $\{(\mathbf{x}_k, y_k)\}$ with $y_k \in \{+1, -1\}$, correct orientation means $y_k f(\mathbf{x}_k) > 0$ on every example used for training.

a. Geometric margin. The Euclidean distance from a point $\mathbf{x}$ to H is $|f(\mathbf{x})|/\|\mathbf{w}\|$. The quantity $\gamma_k := y_k f(\mathbf{x}_k)$ is the

functional margin (up to sign it is the "score confidence"). Dividing by $\|\mathbf{w}\|$ gives the *geometric margin* $\gamma_k/\|\mathbf{w}\|$—the signed distance to the decision boundary in direction $\mathbf{w}$. Rescaling $(\mathbf{w}, b) \mapsto (c\mathbf{w}, cb)$ leaves the hyperplane fixed but changes $\|\mathbf{w}\|$; SVM fixes the scale by imposing $y_k f(\mathbf{x}_k) \geq 1$ on support vectors so that the slab between the two hyperplanes $f = \pm 1$ has half-width $1/\|\mathbf{w}\|$.

b. Maximum-margin (hard) SVM. For linearly separable data, the support vector machine (SVM) finds the unique hyperplane that maximizes the smallest geometric margin—equivalently, it shrinks $\|\mathbf{w}\|$ while keeping all training points outside the $f = \pm 1$ slab. The primal problem is

$$\min_{\mathbf{w},b} \quad \tfrac{1}{2}\|\mathbf{w}\|^2 \tag{69}$$

$$\text{s.t.} \quad y_k\big(\langle \mathbf{w}, \mathbf{x}_k\rangle + b\big) \geq 1, \quad k = 1, \ldots, M. \tag{70}$$

Any constraint satisfied with equality defines a *support vector*; non-support points can be removed without changing the optimal $(\mathbf{w}, b)$. Introducing Lagrange multipliers $\alpha_k \geq 0$ yields the *dual*:

> Maximize $\sum_k \alpha_k - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j\rangle$ subject to $\alpha_k \geq 0$ and $\sum_k \alpha_k y_k = 0$. Sparsity in α (many α_k=0) follows from the KKT complementarity $\alpha_k\big[y_k f(\mathbf{x}_k) - 1\big] = 0$.

Only inner products $\langle \mathbf{x}_i, \mathbf{x}_j\rangle$ ever appear, which motivates the kernel trick $K(\mathbf{x}_i, \mathbf{x}_j)$ and the quantum feature-map viewpoint [12].

c. Soft-margin SVM. Real data often overlap. Nonnegative slack variables ξ_k tolerate points inside the margin or even on the wrong side of H, while a penalty $C\sum_k \xi_k$ limits violations:

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \quad \tfrac{1}{2}\|\mathbf{w}\|^2 + C\sum_{k=1}^{M} \xi_k \tag{71}$$

$$\text{s.t.} \quad y_k\big(\langle \mathbf{w}, \mathbf{x}_k\rangle + b\big) \geq 1 - \xi_k, \quad \xi_k \geq 0. \tag{72}$$

Large C enforces a stricter margin (risk of over-fitting); small C allows more misclassification slack but fattens the effective margin.

D. Least-squares SVM as a linear system

Least-squares SVM (LS-SVM) replaces margin inequalities by *equality* constraints with squared loss on slack [14]. One seeks

$$\min_{\mathbf{w},b,\mathbf{e}} \quad \tfrac{1}{2}\|\mathbf{w}\|^2 + \tfrac{\gamma}{2}\sum_{k=1}^{M} e_k^2 \tag{73}$$

$$\text{s.t.} \quad y_k\big(\langle \mathbf{w}, \mathbf{x}_k\rangle + b\big) = 1 - e_k, \quad k = 1, \ldots, M, \tag{74}$$

with regularization parameter $\gamma > 0$. Introducing Lagrange multipliers α_k, stationarity gives $\mathbf{w} = \sum_k \alpha_k y_k \mathbf{x}_k$, the bias constraint $\sum_k \alpha_k = 0$, and residuals $e_k = \alpha_k/\gamma$. Eliminating $\mathbf{w}$ and $\mathbf{e}_k$ produces a symmetric $(M+1)\times(M+1)$ linear system in the unknowns $(b, \alpha_1, \ldots, \alpha_M)^\top$. With the *kernel matrix* $K_{ij} = \langle \mathbf{x}_i, \mathbf{x}_j\rangle$ (linear kernel) and label-weighted block $\Omega_{ij} = y_i y_j K_{ij}$, the block form is

$$\begin{pmatrix} 0 & \mathbf{y}^\top \\ \mathbf{y} & \Omega + \gamma^{-1} I_M \end{pmatrix} \begin{pmatrix} b \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{1}_M \end{pmatrix}, \tag{75}$$

up to the standard sign conventions for multipliers—the key point for QML is that (75) is *a linear system with structured Hermitian data*, so its solution vector can in principle be prepared via HHL-style methods given suitable oracle access to the matrix.

E. Quantum least-squares SVM and classification

Rebentrost, Mohseni, and Lloyd [15] promote the least-squares SVM linear system to a *quantum* subroutine while keeping the same classical optimality conditions (Sec. IX D). The construction separates into (i) preparing the optimal KKT vector in a quantum register and (ii) turning the SVM decision score into overlaps measurable on a quantum device. Figure 15 summarizes the workflow.

a. Quantum formulation of the KKT system. Collect the LS-SVM unknowns in $\mathbf{z} = (b, \alpha_1, \ldots, \alpha_M)^\top$. Writing (75) as a single matrix equation,

$$F\,\mathbf{z} = \mathbf{r}, \qquad F := \begin{pmatrix} 0 & \mathbf{y}^\top \\ \mathbf{y} & \Omega + \gamma^{-1} I_M \end{pmatrix}, \qquad \mathbf{r} := \begin{pmatrix} 0 \\ \mathbf{1}_M \end{pmatrix}, \tag{76}$$

where $\Omega_{ij} = y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j\rangle$ for the linear kernel. The block F is real symmetric, hence Hermitian, so it falls inside the scope of the HHL input model once suitable sparsity or block-encoded access is assumed [5, 7].

Let $|r\rangle$ denote a normalized amplitude encoding of $\mathbf{r}$ (after possibly appending dummy components if one prefers a power-of-two dimension). HHL with $A = F$ produces a state proportional to

$$|z\rangle \;\propto\; F^{-1}\,|r\rangle\,, \tag{77}$$

matching the solution $\mathbf{z}^\star = (b^\star, \boldsymbol{\alpha}^\star)^\top$ up to global normalization and finite-error phase estimation of e^{iFt}. In practice one does not always *read* every $\alpha_k^\star$ classically; instead $|z\rangle$ is kept coherent for subsequent controlled rotations or as an input to overlap-based inference, trading full tomography for the specific linear functional needed at test time.

b. Decision score and swap-test overlaps. The classical LS-SVM classifier uses the affine score

$$s(\tilde{\mathbf{x}}) \;:=\; b^\star + \sum_{k=1}^{M} \alpha_k^\star\, y_k\, \langle \mathbf{x}_k, \tilde{\mathbf{x}}\rangle, \tag{78}$$

and predicts $\tilde{y} = \operatorname{sign} s(\tilde{\mathbf{x}})$. Using $\mathbf{w}^\star = \sum_k \alpha_k^\star y_k \mathbf{x}_k$ recovers the equivalent form $s(\tilde{\mathbf{x}}) = \langle \mathbf{w}^\star, \tilde{\mathbf{x}}\rangle + b^\star$.

c. Lecture-style state construction and Hadamard test. The lecture notes implement the score evaluation by embedding the bias and training coefficients into a single "index" superposition state produced by HHL. Writing the HHL output for the KKT solution schematically as an amplitude encoding of $(b^\star, \alpha_1^\star, \ldots, \alpha_M^\star)$,

$$|b, \boldsymbol{\alpha}\rangle \;:=\; \frac{1}{\sqrt{N_b}}\Big(b^\star\,|0\rangle + \sum_{k=1}^{M} \alpha_k^\star\,|k\rangle\Big), \qquad N_b := |b^\star|^2 + \sum_{k=1}^{M} |\alpha_k^\star|^2, \tag{79}$$

one then entangles each training index k with its feature vector $\mathbf{x}_k \in \mathbb{R}^d$ by a data-loading unitary (QRAM model):

$$|k\rangle\,|0\rangle \;\mapsto\; |k\rangle\,|\mathbf{x}_k\rangle \;\approx\; |k\rangle \sum_{j=1}^{d} (x_k)_j\,|j\rangle\,, \tag{80}$$

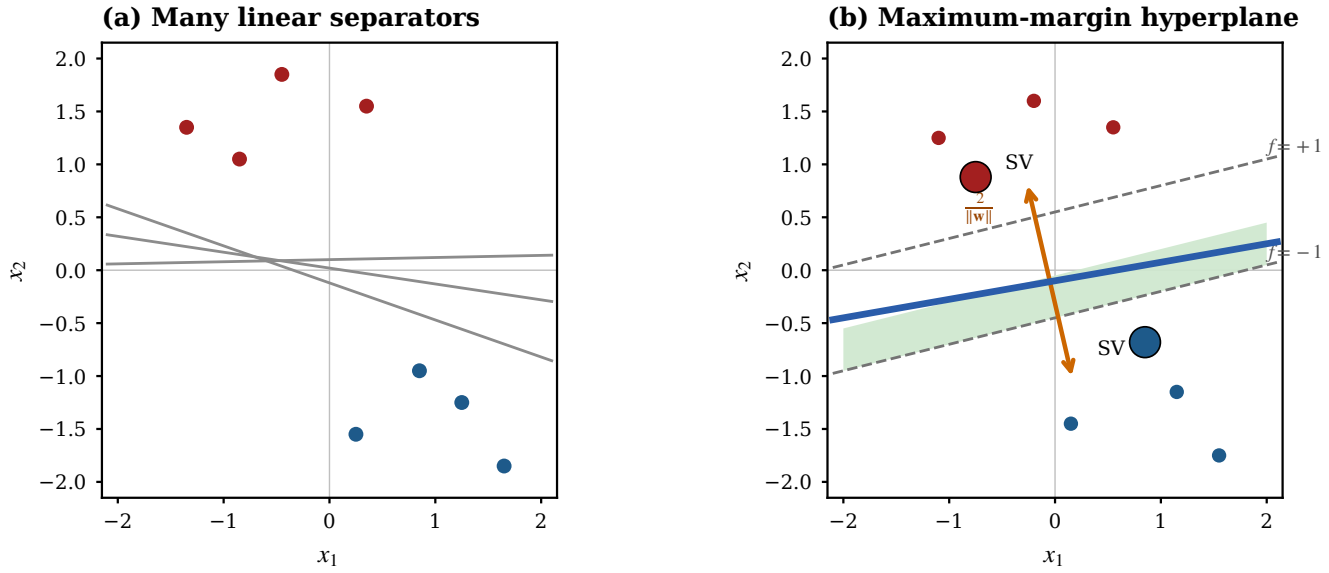

FIG. 13. Two-dimensional schematic of linear classification (vector PDF from `scripts/plot_svm_figures.py`). **(a)** A perceptron only needs *some* hyperplane that separates the classes, so many slopes are admissible (gray). **(b)** A hard-margin SVM picks the separator that maximizes the distance between class clouds; the shaded band is between $f = \pm 1$. Support vectors (marked "SV") lie on the dashed margin lines and have active constraints in (70).

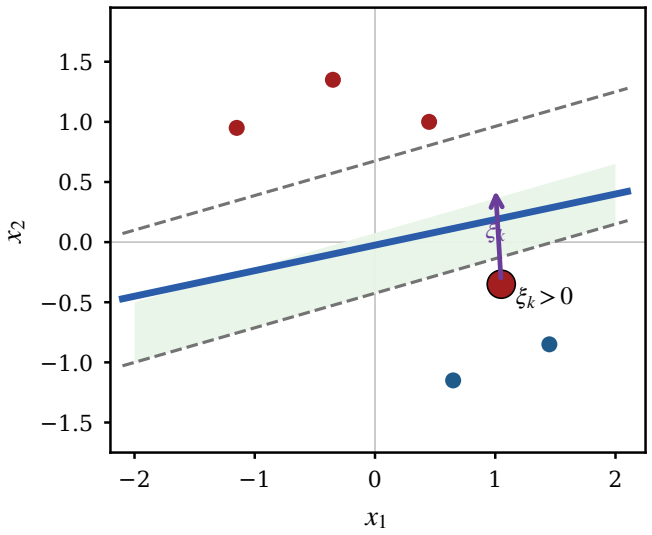

FIG. 14. Soft-margin notion: most positive-class points satisfy $y_k f(\mathbf{x}_k) \geq 1$, but an outlier (red) lies inside the margin tube or on the wrong side. Its slack ξ_k measures how much the margin constraint (72) is relaxed.

and similarly for the test point $\tilde{\mathbf{x}}$. With this convention, the lecture constructs the joint "model" state

$$|u\rangle \;:=\; \frac{1}{\sqrt{N_u}}\Big(b^\star\,|0\rangle\,|0\rangle + \sum_{k=1}^{M} \alpha_k^\star\,|k\rangle\,|\mathbf{x}_k\rangle\Big), \qquad N_u := |b^\star|^2 + \sum_{k=1}^{M} |\alpha_k^\star|^2\,\|\mathbf{x}_k\|^2 \tag{81}$$

and the "query" state for the new data

$$|\tilde{x}\rangle \;:=\; \frac{1}{\sqrt{N_{\tilde{x}}}}\Big(|0\rangle\,|0\rangle + \sum_{k=1}^{M} |k\rangle\,|\tilde{\mathbf{x}}\rangle\Big), \qquad N_{\tilde{x}} := 1 + M\|\tilde{\mathbf{x}}\|^2. \tag{82}$$

Taking the inner product gives

$$\langle \tilde{x}|u\rangle \;=\; \frac{1}{\sqrt{N_u N_{\tilde{x}}}}\left(b^\star + \sum_{k=1}^{M} \alpha_k^\star\,\langle \tilde{\mathbf{x}}|\mathbf{x}_k\rangle\right) \;\propto\; b^\star + \sum_{k=1}^{M} \alpha_k^\star\,\langle \mathbf{x}_k, \tilde{\mathbf{x}}\rangle, \tag{83}$$

where the proportionality hides known normalization factors and the last step uses the real-vector convention $\langle \tilde{\mathbf{x}}|\mathbf{x}_k\rangle = \langle \mathbf{x}_k, \tilde{\mathbf{x}}\rangle$. Thus, up to the label-folding convention ($\alpha_k^\star \to \alpha_k^\star y_k$), the classification rule can be written as

$$\tilde{y} = \operatorname{sign}\big(\langle \tilde{x}|u\rangle\big). \tag{84}$$

The overlap $\langle \tilde{x}|u\rangle$ is estimated by the Hadamard test (Sec. VI D), which returns the real (and, with a phase shift, imaginary) part of $\langle \tilde{x}|u\rangle$ rather than only its magnitude.

Quantumly, each term $\langle \mathbf{x}_k, \tilde{\mathbf{x}}\rangle$ may be estimated by preparing feature states $|\psi_k\rangle \approx |\mathbf{x}_k\rangle / \|\mathbf{x}_k\|$, $|\tilde{\psi}\rangle \approx |\tilde{\mathbf{x}}\rangle / \|\tilde{\mathbf{x}}\|$ and invoking the swap test (Sec. VI D). Denoting the control outcome probability of seeing $|0\rangle_a$ on the ancilla by p_0, one has the standard relation

$$p_0 = \frac{1 + |\langle \psi_k|\tilde{\psi}\rangle|^2}{2}, \qquad \text{hence} \qquad |\langle \psi_k|\tilde{\psi}\rangle|^2 = 2p_0 - 1 \tag{85}$$

(in the ideal, unmitigated protocol). Recovering the *signed* inner product needed in (78) may require additional phase information or data preprocessing so that all overlaps share a known global phase convention; the lecture-level takeaway is that the bottleneck becomes accurate quantum evaluation of similarities $\langle \mathbf{x}_k, \tilde{\mathbf{x}}\rangle$, possibly in superposition over k, before combining them with the trained weights $\alpha_k^\star y_k$ and bias $b^\star$.

d. Kernelized QSVM. Nothing in (76) is specific to dot products: replacing $\Omega_{ij} = y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$ with a positive semidefinite kernel embeds nonlinear margins while preserving the Hermitian KKT matrix for HHL-class solvers. Quantum feature maps $K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x})|\phi(\mathbf{x}')\rangle$ connect this picture to modern kernel-based QML [12].

e. Assumptions and caveats. The advertised polylogarithmic scaling presupposes efficient preparation of $|r\rangle$

(QRAM-like or structured data access), sparse or block-encoded Hamiltonian simulation of e^{-iFt}, and a moderate condition number κ of F so that HHL post-selection (65) remains tractable after amplitude amplification. Classical *dequantization* shows that comparable sampling models can erase the exponential separation on some linear-algebra problems [5], and constant-depth overlap readouts on NISQ devices inherit hardware noise not reflected in idealized complexity displays. These issues are discussed in depth in Refs. [13, 15].

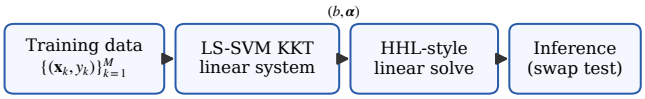

FIG. 15. Block view of Rebentrost–Mohseni–Lloyd quantum LS-SVM [15]: encode the training problem as a structured Hermitian system, prepare the solution amplitudes with HHL-class primitives, then classify new features using estimated overlaps (Sec. VI D).

F. Other landmark QML algorithms

Beyond the LS-SVM construction, we briefly recall two widely cited directions.

a. Quantum principal component analysis. Lloyd, Mohseni, and Rebentrost [16] estimate principal components of a density matrix ρ in time $O(\log N)$ under a strong QRAM input model—exponentially faster than naive classical diagonalization in matrix dimension.

b. Variational quantum algorithms. VQAs use short parameterized circuits $U(\boldsymbol{\theta})$ and optimize $\boldsymbol{\theta}$ classically to minimize $C(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|O|\psi(\boldsymbol{\theta})\rangle$ [17]. VQAs are the primary paradigm for NISQ-era QML [1].

G. Open Questions and Future Outlook

Despite significant theoretical progress, fundamental questions remain:

- *Dequantization*: Can classical sampling-based algorithms match quantum speedups [5]?
- *Barren plateaus*: Do VQA gradients vanish exponentially with system size, making training intractable?
- *End-to-end advantage*: Is a provable, QRAM-free quantum speedup achievable on relevant ML tasks?

Progress on these questions will determine whether QML transcends academic interest to become a practical technology.

X. CONCLUSION

This report has developed the theoretical foundations of quantum machine learning from first principles. Beginning with the postulates of quantum mechanics and the mathematical framework of Hilbert spaces and Dirac notation, we derived the quantum circuit model, established universality, proved the no-cloning theorem, and analyzed reversible computation and Landauer's principle. Detailed circuit derivations for the Hadamard and swap tests illuminated how quantum expectation values and state overlaps are estimated. Building toward algorithms, we reviewed the query model, Grover search, QFT, and quantum phase estimation—ingredients that feed the Harrow–Hassidim–Lloyd linear-systems solver.

Week 10 material is synthesized in Section IX: we outlined the HHL pipeline (spectral inversion, Hermitian embedding for non-Hermitian A, and a conceptual circuit checklist), summarized hard- and soft-margin SVMs, rewrote least-squares SVM as the Hermitian linear system (75), and connected training plus classification to HHL and the swap test following Rebentrost *et al.* [15]. We briefly recalled quantum PCA and variational QML and highlighted open questions around dequantization, barren plateaus, and hardware assumptions.

Natural extensions for later lectures include kernelized dual forms, deeper Hamiltonian-simulation oracles, and fault-tolerant resource estimates for end-to-end QML pipelines.

[1] J. Preskill, Quantum computing in the nisq era and beyond, Quantum **2**, 79 (2018).
[2] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, Cambridge, UK, 2010).
[3] M. M. Wilde, *Quantum Information Theory*, 2nd ed. (Cambridge University Press, Cambridge, UK, 2017).
[4] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, UK, 2018).
[5] M. Schuld and F. Petruccione, *Machine Learning with Quantum Computers*, Quantum Science and Technology (Springer, Cham, Switzerland, 2021).
[6] E. Bernstein and U. Vazirani, Quantum complexity theory, SIAM Journal on Computing **26**, 1411 (1997).
[7] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum algorithm for linear systems of equations, Physical Review Letters **103**, 150502 (2009).
[8] W. K. Wootters and W. H. Zurek, A single quantum cannot be cloned, Nature **299**, 802 (1982).
[9] C. H. Bennett and S. J. Wiesner, Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states, Physical Review Letters **69**, 2881 (1992).
[10] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels, Physical Review Letters **70**, 1895 (1993).
[11] R. Landauer, Irreversibility and heat generation in the computing process, IBM Journal of Research and Development **5**, 183 (1961).
[12] M. Schuld and N. Killoran, Quantum machine learning in feature hilbert spaces, Physical Review Letters **122**, 040504 (2019).
[13] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature **549**, 195 (2017).
[14] J. A. K. Suykens and J. Vandewalle, Least squares support vector machine classifiers, Neural Processing Letters **9**, 293 (1999).
[15] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum support vector machine for big data classification, Physical Review Letters **113**, 130503 (2014).

[16] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum principal component analysis, Nature Physics **10**, 631 (2014).
[17] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nature Reviews Physics **3**, 625 (2021).