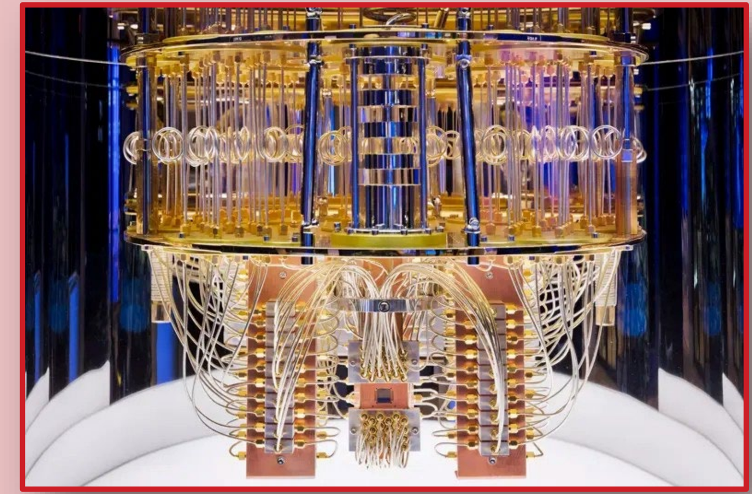
Quantum Correlations Enabling Quantum Advantage in Machine Learning



Presented By:

Sadia Afrin Purba

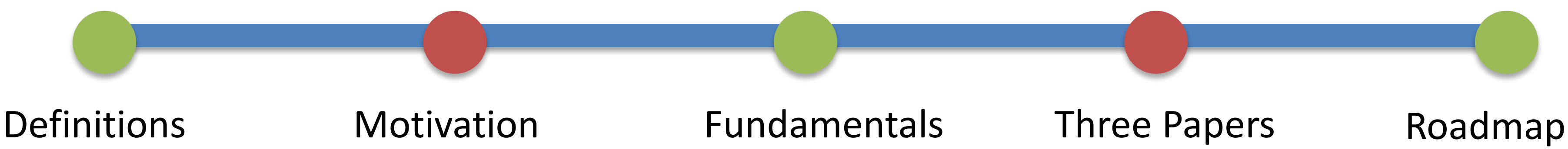
Dept. of Elect. and Comp. Eng.
Temple University

Submitted to:

Advisor:
Dr. Joseph Picone, Dept. of Electrical and Computer Engineering

Committee:
Dr. Iyad Obeid, Dept. of Electrical and Computer Engineering
Dr. Chang-Hee Won, Dept. of Electrical and Computer Engineering
Dr. Andrew Spence, Dept. of Bioengineering
Dr. Yaroslav Koshka, Dept. of Electrical and Computer Engineering, Mississippi State University

Outline


Definitions
Motivation
Fundamentals
Three Papers
Roadmap

Definitions: Quantum Advantages (QA)

- Quantum refers to the principles of **quantum mechanics**, the physics governing subatomic particles like electrons and photons.

- Many physicists and computer scientists describe quantum mechanics as a generalization of classical probability theory.

- Quantum mechanics deals with two fundamental properties of subatomic particles.
  - Superposition
  - Entanglement

- Quantum computing is a type of computation that uses the principles of quantum mechanics to process information in ways that classical computers cannot.

- In quantum computing, quantum advantage is the goal of demonstrating that a programmable quantum computer can solve a problem that no classical computer can solve in any feasible amount of time, irrespective of the usefulness of the problem.

- The term was coined by John Preskill in 2011, but the concept dates to Yuri Manin's 1980 and Richard Feynman's 1981 proposals of quantum computing.

Definitions: Quantum Advantages (QA)

Example	Problem Type	Advantage Over Classical	Current Status
Random Circuit Sampling (Google Sycamore, 2019)	Sampling from random quantum circuits (benchmark task)	Performed in 200 seconds, estimated 10,000 years on a classical supercomputer	Achieved – Demonstration of quantum supremacy
Searching an unsorted database for a specific entry (Grover's Algorithm, 1996)	Unstructured search / optimization	Quadratic speedup: classical search requires $O(N)$ queries; Grover's requires only $O(\sqrt{N})$	Demonstrated on small-scale NISQ devices (up to a few qubits).
Integer Factorization (Shor's Algorithm, 1994)	Factoring large integers (used in encryption)	Exponential speedup over best-known classical methods	Theoretical – Not yet realized at scale

Q-Day!!

Definitions: Quantum Correlations (QCorr)

- Classical correlation is a statistical measure that describes the extent to which two or more random variables are linearly related—meaning they tend to change together.
  - this statistical measure obeys classical probability theory and local-realism theory*,
  - it describes joint statistics of separate entities.
  - e.g., coin toss results that depend on each other

- Quantum correlation is a general term in quantum mechanics that refers to non-classical statistical relationships between two or more quantum particles.
  - can violate local-realism theory, demonstrating nonlocality,
  - correlations physically entangle subsystems so that their total information exists only in the global state — not in any part alone.
  - e.g., two quantum dice always rolling the same number no matter how far apart they are

- The strongest form of quantum correlation is quantum entanglement.

- A more general type of quantum correlation that can exist even in separable (non-entangled) states is quantum discord.

** Things have definite properties (realism), and those properties can only be influenced by nearby events (locality).*

Motivations

- The goal of ML task is to model the distribution $p(y|x)$ using a parameterized classical model $p_\theta(y|x)$ based on a set of training data $\{x^{(i)}, y^{(i)}\}_{i=1}^{N}, x \sim p(x), y \sim p(y \mid x)$.

- This is achieved by tuning the parameter θ to minimize certain loss function that represents how far p_θ is from p.

- Classical models rely on joint and conditional probabilities
  - capture only statistical dependence, not intrinsic physical relationships
  - correlations are local— information is shared only through explicit data features or parameters
  - constrained by the curse of dimensionality — scaling poorly with data size and complexity

Can non-local quantum correlations provide a technological or quantum advantage for machine learning tasks?

Motivations

- In quantum ML (QML), quantum advantage is the ability of quantum systems to perform learning tasks faster, with fewer resources, or greater accuracy.

- Quantum correlations (entanglement, discord) allow nonlocal dependencies — relationships that can't be described by classical probability distributions.
  - can learn information beyond the training distribution,
  - store and process information globally across subsystems, rather than locally per variable, enhancing model expressivity

- Phase correlations allow constructive/destructive interference
  - boosting learning efficiency and speedup

- Entanglement-based correlations provide
  - coherent information transfer,
  - reducing energy and communication costs compared to classical models

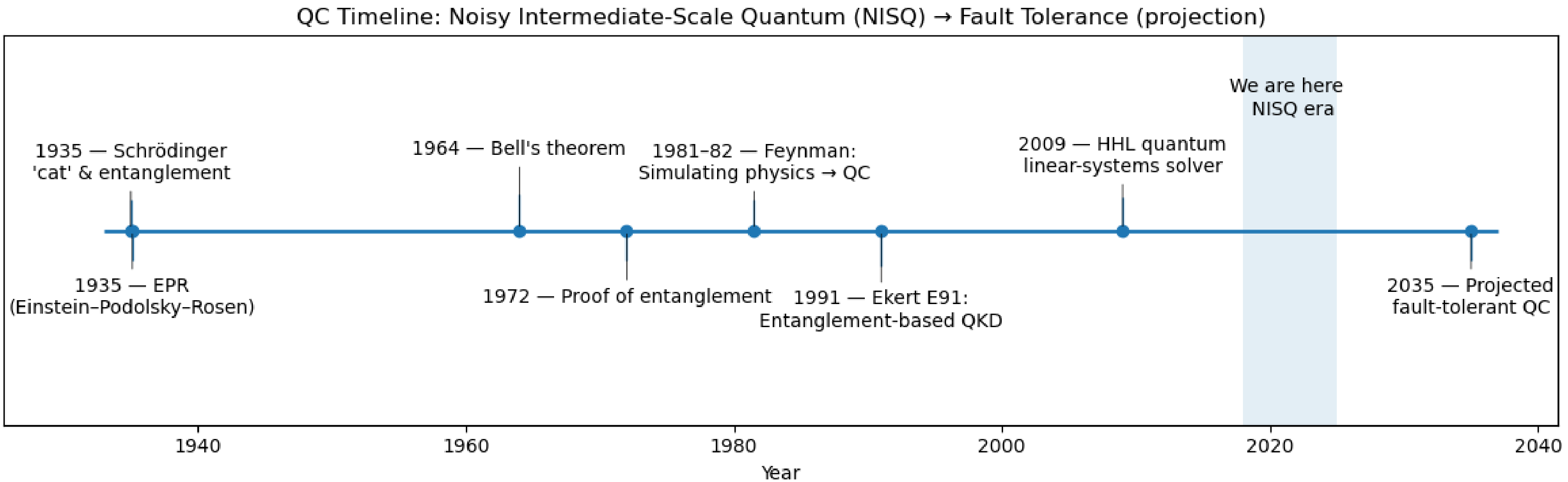
Motivations

- **Entanglement-based Advantage:** H. Zhao and D.-L. Deng, "Entanglement-induced provable and robust quantum learning advantages," *npj Quantum Inf*, vol. 11, no. 1, p. 127, Jul. 2025, doi: *10.1038/s41534-025-01078-x.*

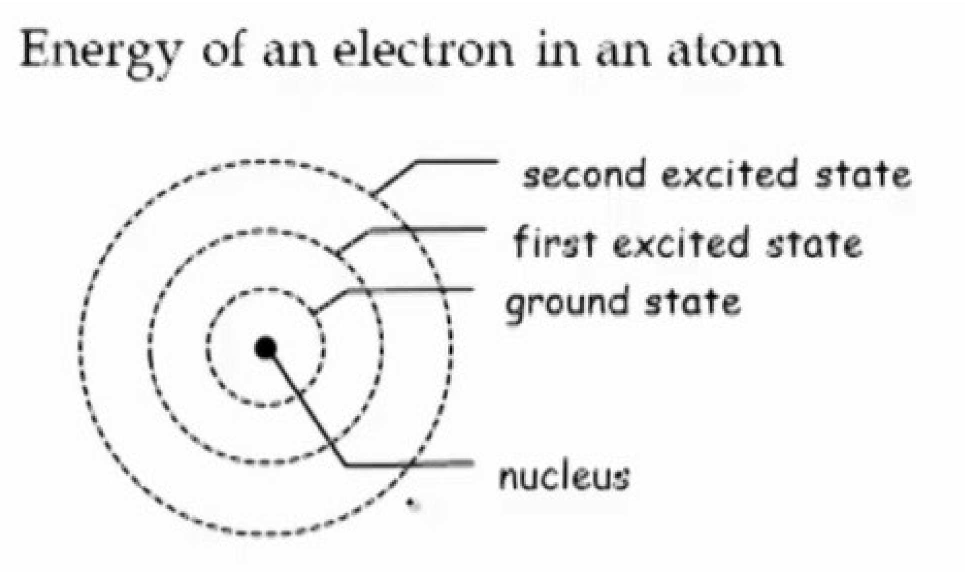
- **Quantum No-free-lunch (NFL):** K. Sharma, M. Cerezo, Z. Holmes, L. Cincio, A. Sornborger, and P. J. Coles, "Reformulation of the No-Free-Lunch Theorem for Entangled Datasets," *Phys. Rev. Lett.*, vol. 128, no. 7, p. 070501, Feb. 2022, doi: *10.1103/PhysRevLett.128.070501*.

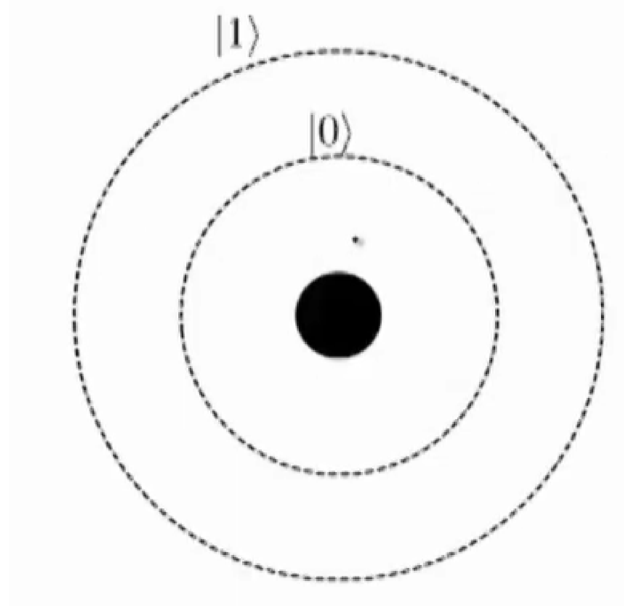
- **Discord-based Advantage:** M. Karimi, A. Javadi-Abhari, C. Simon, and R. Ghobadi, "The power of one clean qubit in supervised machine learning," *Sci Rep*, vol. 13, no. 1, p. 19975, Nov. 2023, doi: *10.1038/s41598-023-46497-y*.

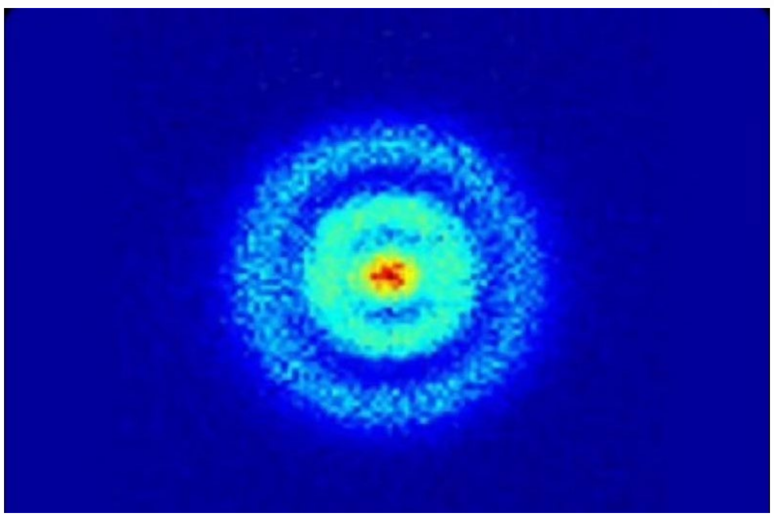
Fundamentals: Quantum Computing



Fundamentals: Quantum Computing







first-ever direct visualization of an electron orbital

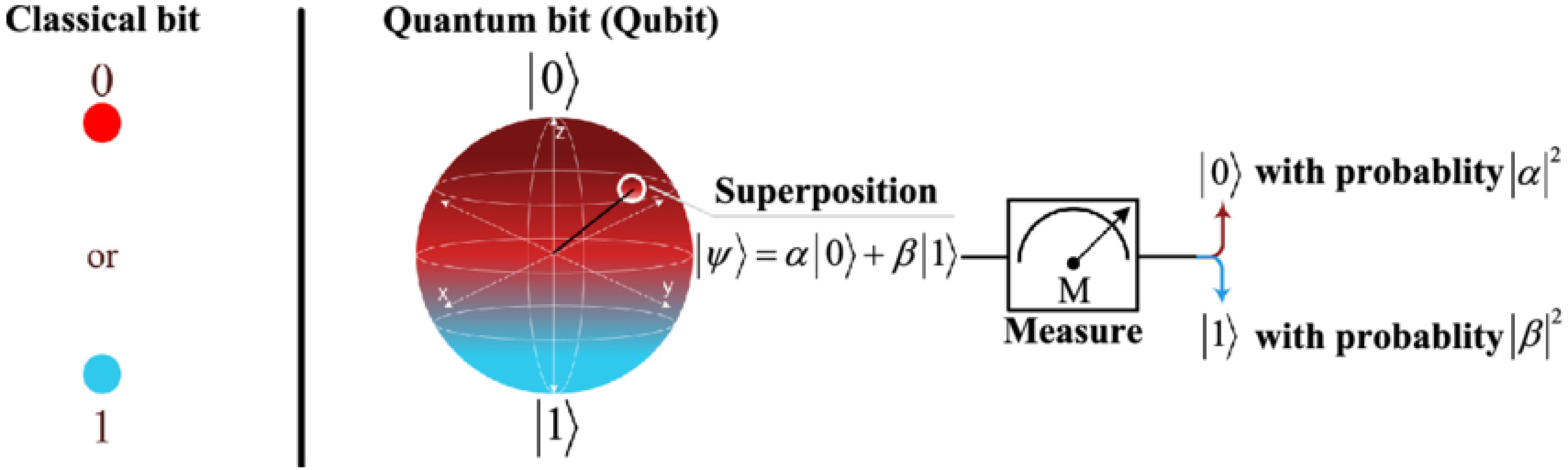
$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

Probability amplitude

Quantum bit (qubit)

$|\psi\rangle$ $(sigh)$ is the quantum state encoding the information about a quantum particle.

Fundamentals: Quantum Computing



classical bit vs qubit [1]

In the computational basis,

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

$|0\rangle$: The probability of finding the qubit in ground state or up-spin is 100%.
$|1\rangle$: The probability of finding the qubit in excited state or down-spin is 100%.

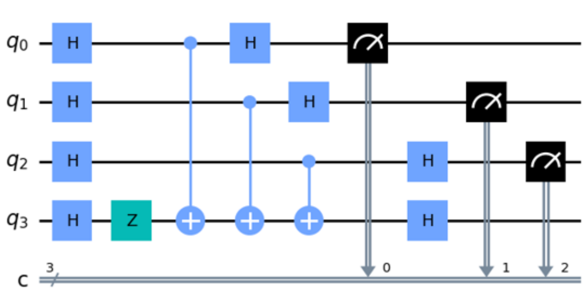
More generally, a qubit $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is

$$|\psi\rangle = \begin{bmatrix}\alpha\\\beta\end{bmatrix}, \quad where\ probabilty, \quad p_0 = |\alpha|^2, p_1 = |\beta|^2, and\ p_0 + p_1 = 1$$

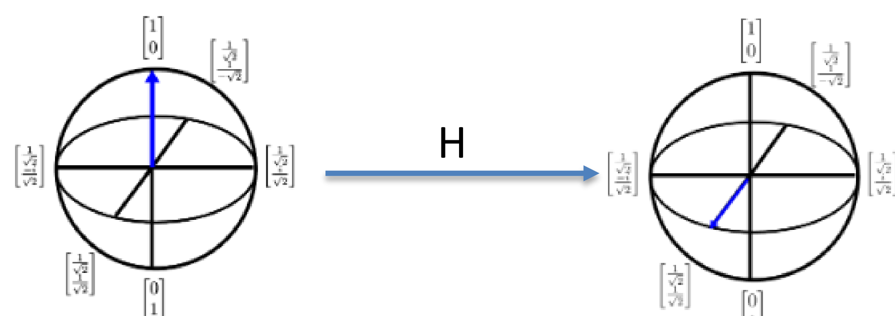
Quantum state is represented as $|.\rangle$ dirac notation.

Fundamentals: Quantum Computing

- The Bloch sphere gives a way of describing a single-qubit quantum state (which is a two-dimensional complex vector) as a three-dimensional real-valued vector.
- Quantum circuit/model is a sequence of linear transformations using quantum gates on the qubits, followed by measurements to extract a classical outcome.
- Common quantum gates (MUST be unitary):
  - Hadamard (H): create superposition (e.g., $|0\rangle \rightarrow (|0\rangle + |1\rangle)/\sqrt{2}$).
  - Pauli X, Y, Z: single-qubit rotations around x, y, z axes.
  - CNOT (Controlled-Not), CZ: two-qubit gate; flip the target if the control qubit is $|1\rangle$; creates entanglement.
- Why must quantum gates be unitary?
  - Reversibility: Quantum evolution is information-preserving; unitary U ensures a unique inverse $U^\dagger$.
  - Probability conservation: Unitarity enforces $U^\dagger U = I$, keeping total probability $= 1$.
- How is measurement performed (computational basis):
  - Measure each qubit in the Z basis $\{|0\rangle, |1\rangle\}$.
  - The state collapses to $|0\rangle$ or $|1\rangle$ with probabilities given by the squared amplitudes.
  - Repeated shots (many runs) build histograms to estimate outcome probabilities.

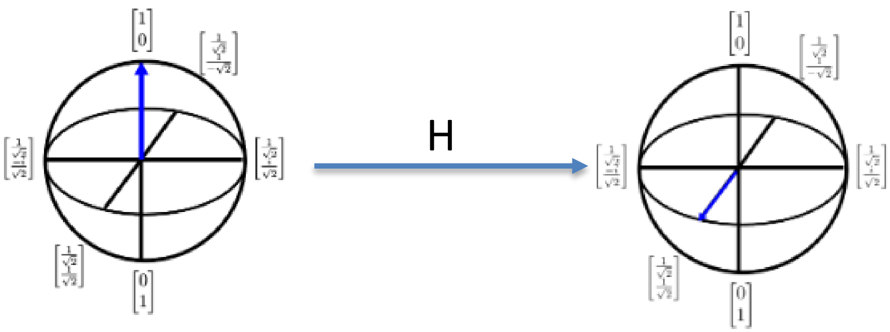


A quantum circuit



A Bloch sphere

Fundamentals: Quantum Computing

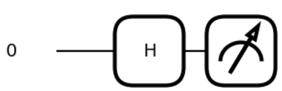


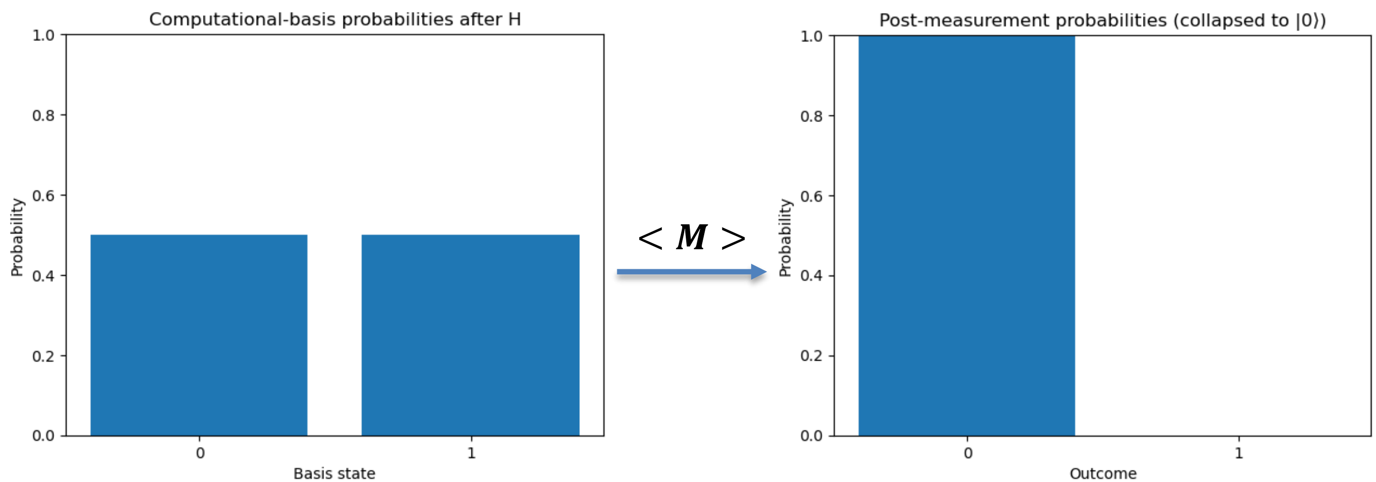
$$H|0\rangle = [1 \quad 0] \cdot \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} = \begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$$

Probability amplitude, $\alpha = \beta = \frac{1}{\sqrt{2}}$

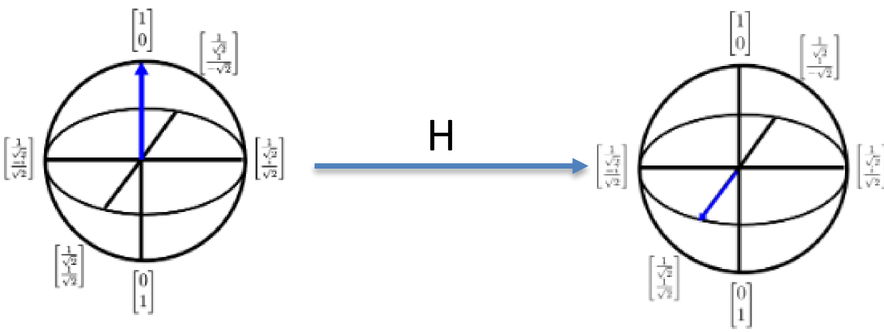
Thus, probability, $p_0 = p_1 = 0.5 = 50\%$

A qubit can be in a state of both 0 and 1 at the same time, a phenomenon called superposition.





Fundamentals: Quantum Computing

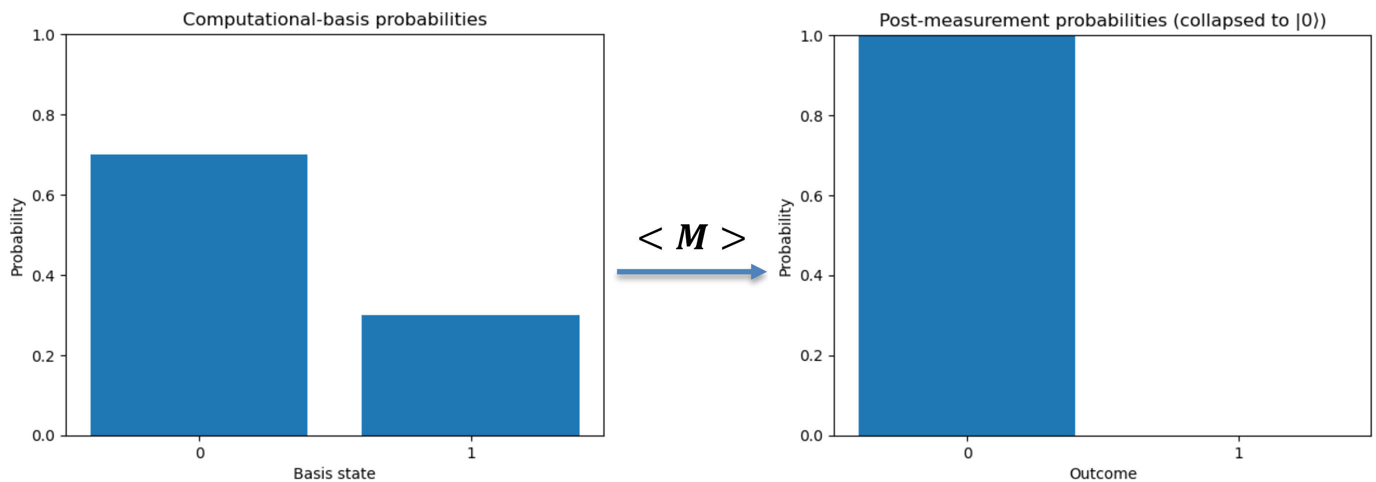


$$H|0\rangle = \begin{bmatrix}1 & 0\end{bmatrix} \cdot \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} = \begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$$

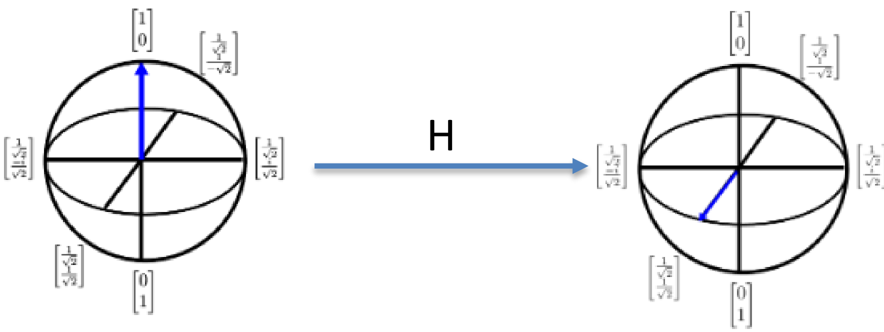
Probability amplitude, $\alpha = \beta = \frac{1}{\sqrt{2}}$

Thus, probability, $p_0 = p_1 = 0.5 = 50\%$

A qubit can be in a state of both 0 and 1 at the same time, a phenomenon called superposition.



Fundamentals: Quantum Computing

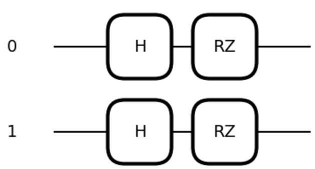


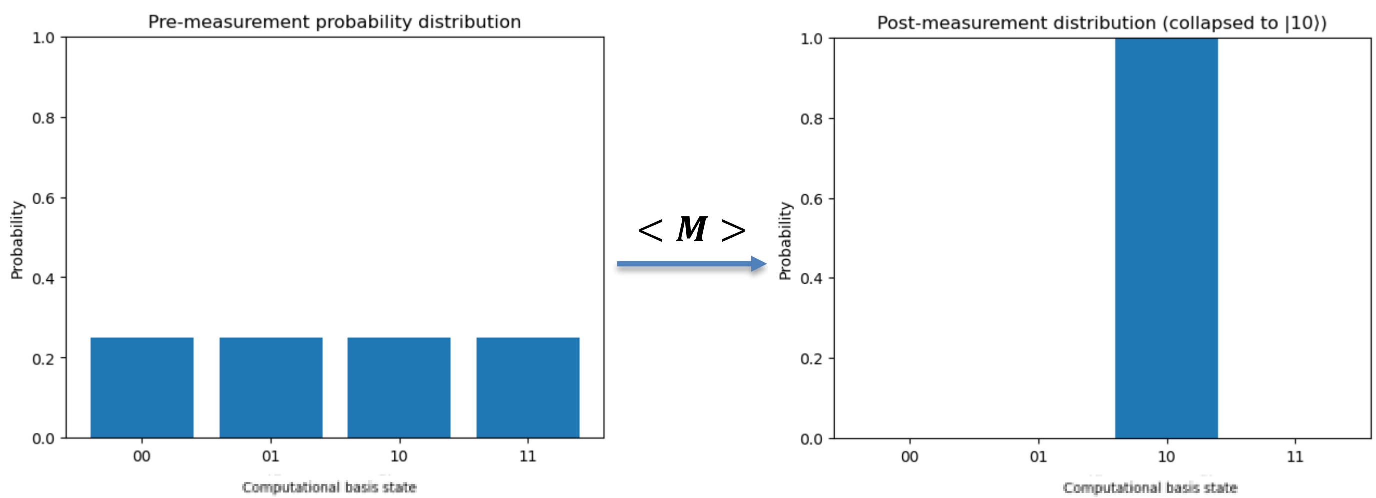
$$H|0\rangle = [1 \quad 0] \cdot \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} = \begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$$

Probability amplitude, $\alpha = \beta = \frac{1}{\sqrt{2}}$

Thus, probability, $p_0 = p_1 = 0.5 = 50\%$

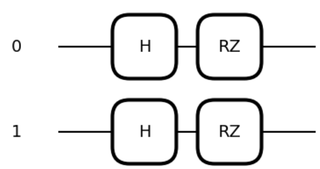
A qubit can be in a state of both 0 and 1 at the same time, a phenomenon called superposition.

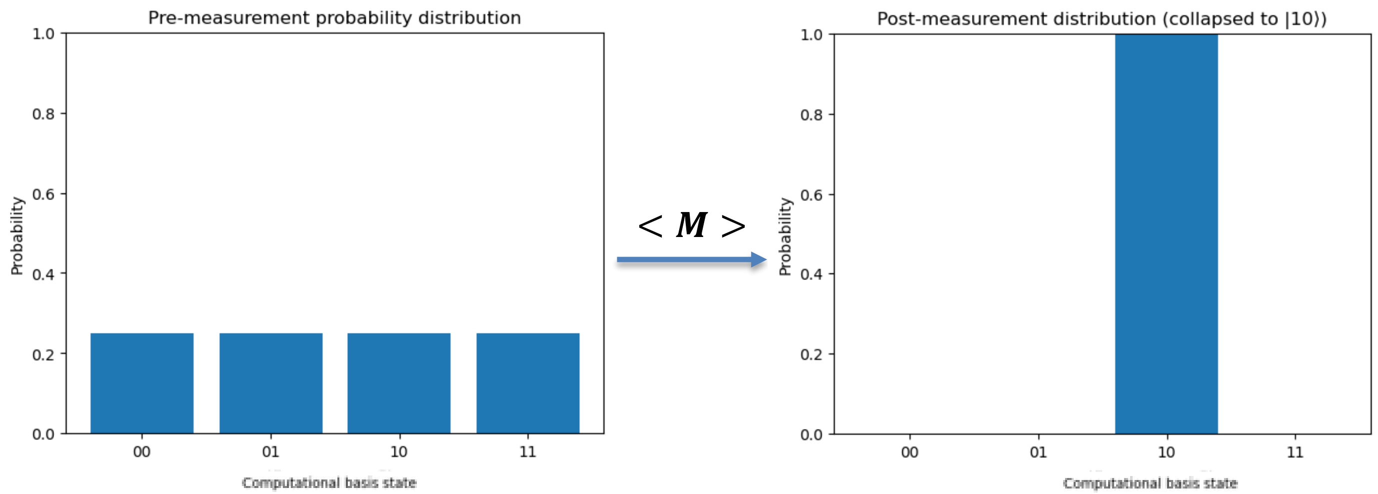




Fundamentals: Quantum Computing

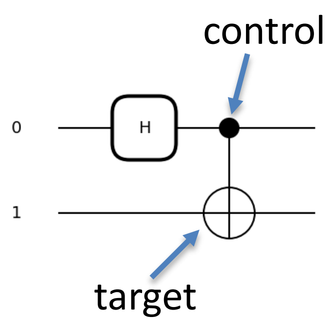
- An n-qubit register has 2^n computational-basis states $\{|x\rangle : x \in \{0,1\}^n\}$.
- Applying H to each qubit $(H^{\otimes n})$ yields a uniform superposition with equal outcome probability $1/2^n$ when measured in the computational basis.
- Uniform superposition is cool, but not a speedup—until we add correlations that boost the ‘right’ states and cancel the rest.
- For that need quantum correlation: *entanglement or discord*.

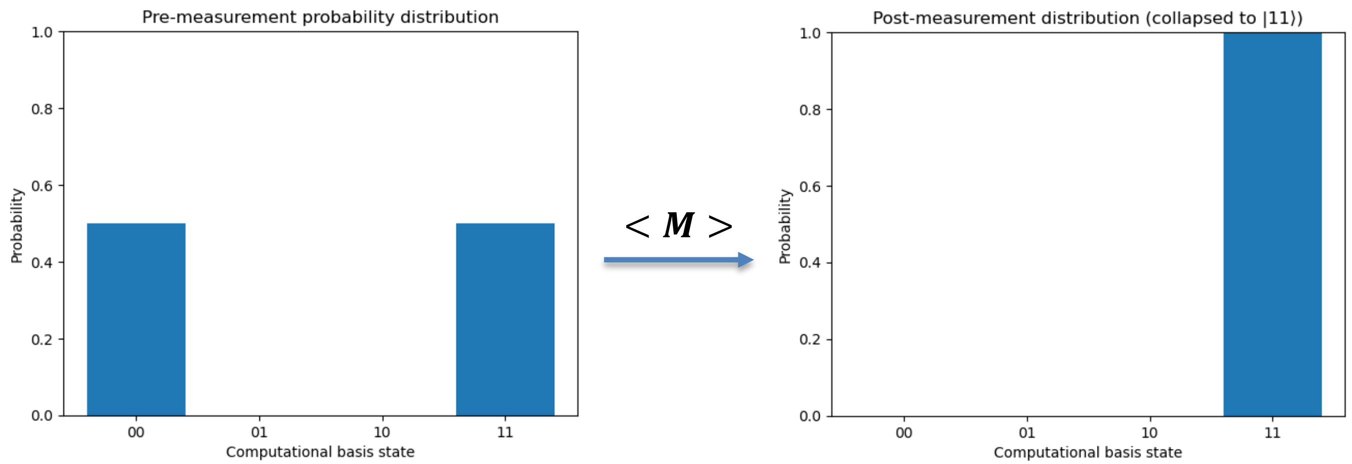




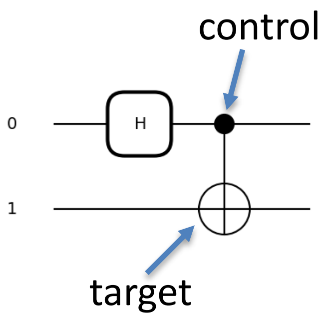
Fundamentals: Quantum Computing

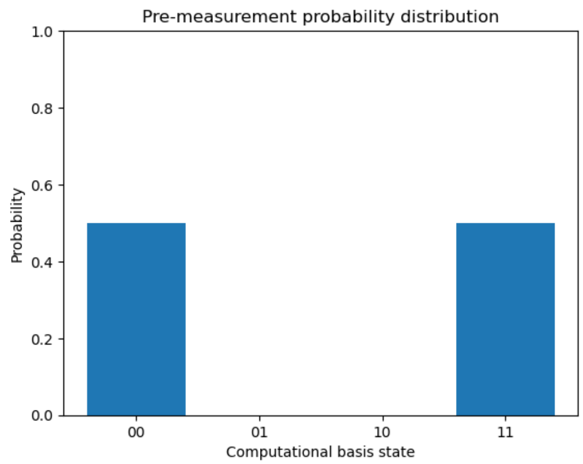
- An n-qubit register has 2^n computational-basis states $\{|x\rangle : x \in \{0,1\}^n\}$.
- Applying H to each qubit $\left(H^{\otimes n}\right)$ yields a uniform superposition with equal outcome probability $1/2^n$ when measured in the computational basis.
- Uniform superposition is cool, but not a speedup—until we add correlations that boost the ‘right’ states and cancel the rest.
- For that need quantum correlation: *entanglement or discord*.
- Use CNOT gate to create entanglement.
- CNOT logic: Flip the target qubit iff the control qubit is 1. (Control is unchanged.)



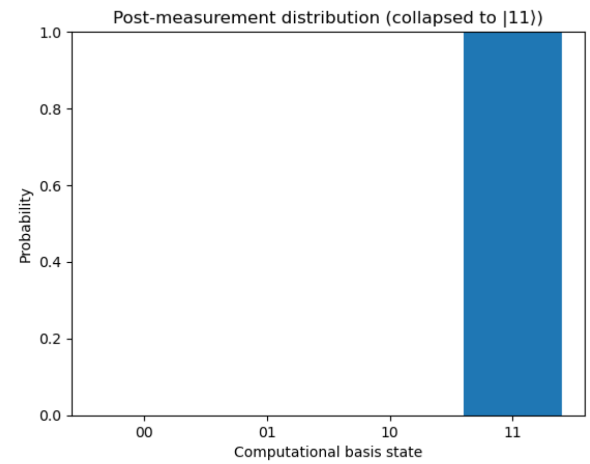


Fundamentals: Quantum Computing



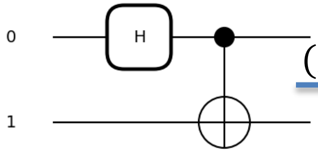


$< \boldsymbol{M} >$



- Quantum entanglement is the phenomenon where the quantum state of each particle in a group cannot be described independently of the state of the others, even when the particles are separated by a large distance.
- $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is entangled iff it is not a product:

$$|\psi\rangle_{AB} \neq |\psi\rangle_A \otimes |\psi\rangle_B.$$

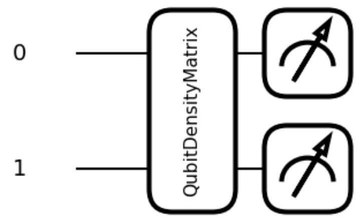


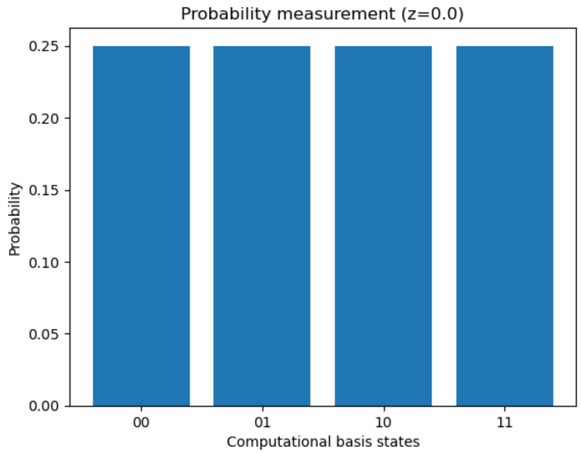
$(\boldsymbol{H}|\boldsymbol{0}\rangle) \otimes |\boldsymbol{1}\rangle$ → $\frac{1}{\sqrt{2}}(|01\rangle + |11\rangle)$ $\xrightarrow{\boldsymbol{CNOT}}$ $\frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$

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

*$|01\rangle$ is shorthand for $|0\rangle \otimes |1\rangle$.

Fundamentals: Quantum Computing





- Quantum correlation (discord) can be created without being in the entangled states.

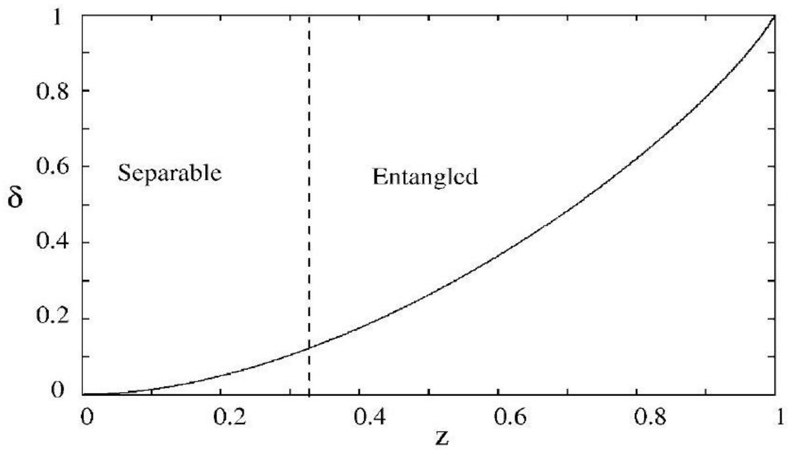
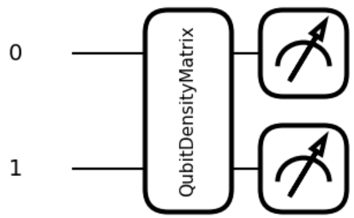
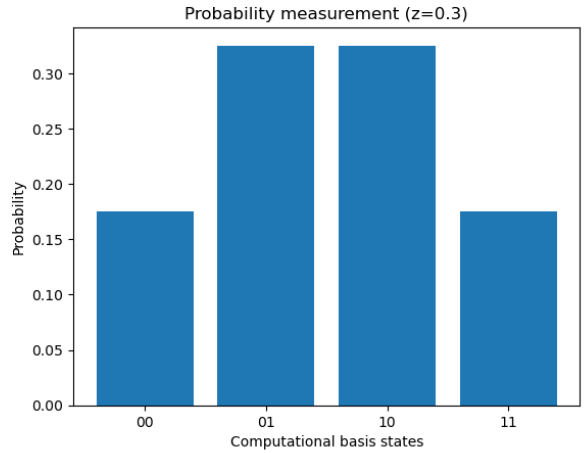


Figure 1: Quantum discord (δ) versus the mixing parameter z for a 2-qubit quantum state [2].

Fundamentals: Quantum Computing





- Quantum correlation (discord) can be created without being in the entangled states.

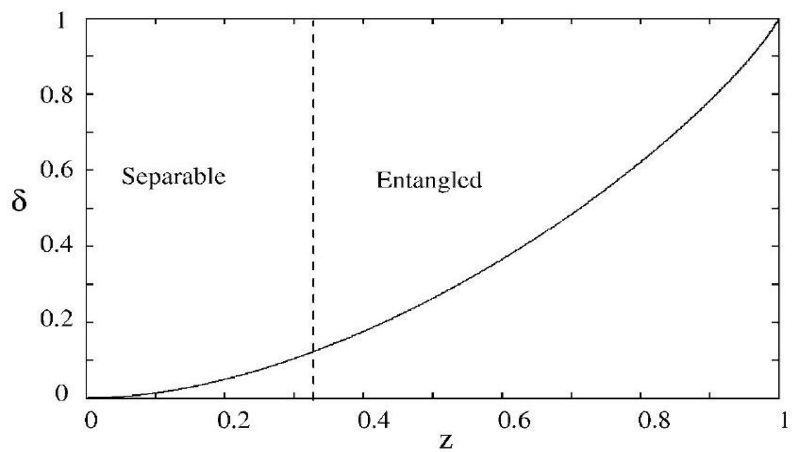
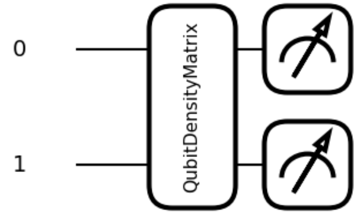
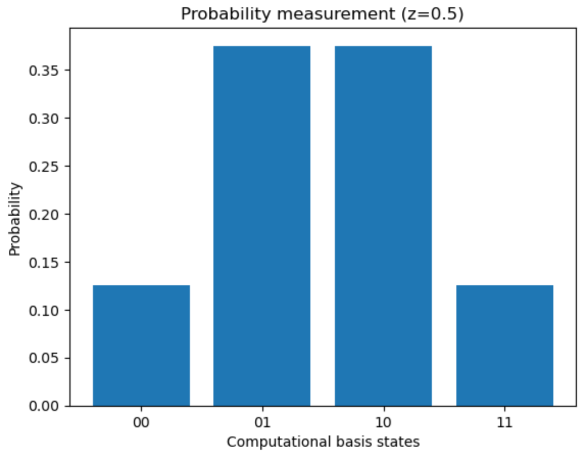


Figure 1: Quantum discord (δ) versus the mixing parameter z for a 2-qubit quantum state [2].

Fundamentals: Quantum Computing





- Quantum correlation (discord) can be created without being in the entangled states.

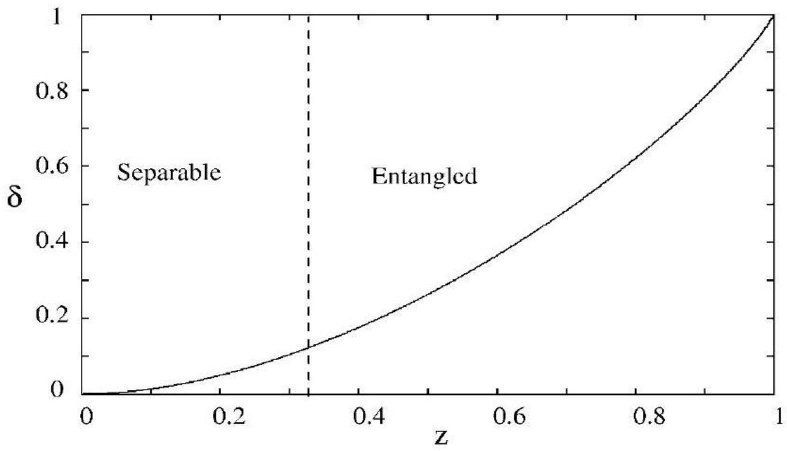
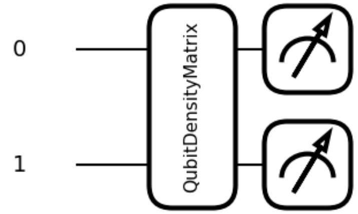
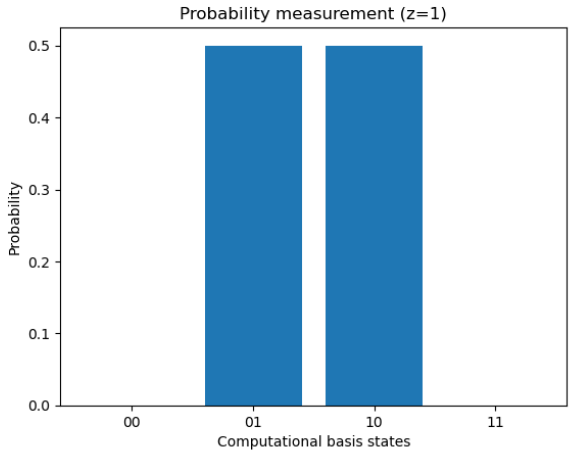


Figure 1: Quantum discord (δ) versus the mixing parameter z for a 2-qubit quantum state [2].

Fundamentals: Quantum Computing





- Quantum correlation (discord) can be created without being in the entangled states.

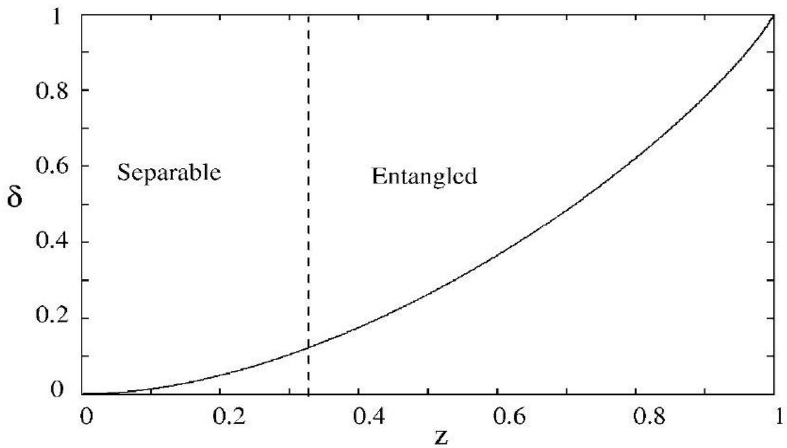


Figure 1: Quantum discord (δ) versus the mixing parameter z for a 2-qubit quantum state [2].

Three Papers

- **Entanglement-based Advantage:** H. Zhao and D.-L. Deng, "Entanglement-induced provable and robust quantum learning advantages," *npj Quantum Inf*, vol. 11, no. 1, p. 127, Jul. 2025, doi: *10.1038/s41534-025-01078-x.*

- **Quantum No-free-lunch (NFL):** K. Sharma, M. Cerezo, Z. Holmes, L. Cincio, A. Sornborger, and P. J. Coles, "Reformulation of the No-Free-Lunch Theorem for Entangled Datasets," *Phys. Rev. Lett.*, vol. 128, no. 7, p. 070501, Feb. 2022, doi: *10.1103/PhysRevLett.128.070501*.

- **Discord-based Advanatge:** M. Karimi, A. Javadi-Abhari, C. Simon, and R. Ghobadi, "The power of one clean qubit in supervised machine learning," *Sci Rep*, vol. 13, no. 1, p. 19975, Nov. 2023, doi: *10.1038/s41598-023-46497-y*.

Paper I: Quantum Entanglement-based Advantage

Title: Entanglement-Induced Provable and Robust Quantum Learning Advantages

Problem Statement

Classical sequence learning models face a communication bottleneck where internal latent dimension d must scale linearly, $d=\Omega(n)$, with the input sequence length n, limiting speed and efficiency.

Solution Method

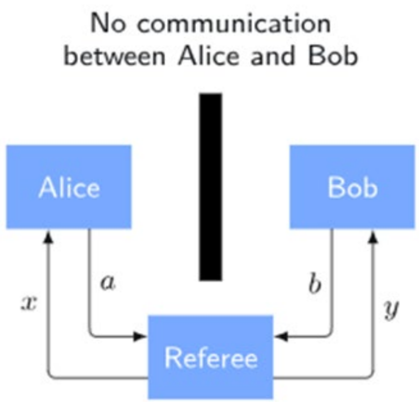
Replace classical communication with pre-computed quantum entanglement. The model leverages the Mermin–Peres magic square game to construct a simple sequence-to-sequence task. The quantum model uses a constant-depth, $O(1)$-parameter quantum circuit with $2n$ Bell pairs (n = sequence length).

How Entanglement Solves it

Non-local correlations provided by entanglement substitute the need for explicit communication required to coordinate remote parts of the input sequence, reducing classical complexity from $\Omega(n)$ to $O(1)$.

Paper I: Quantum Entanglement-based Advantage

The Mermin–Peres Magic Square Game



$x, y \in \{1,2,3\}, \mathrm{a}, \mathrm{b} \in \{0, 1\}^3$
$x = row\ num, y = col\ num$

Win conditions (all must hold):

- Each **row** must have **odd parity** (the XOR of Alice's three bits = 1).
- Each **column** must have **even parity** (the XOR of Bob's three bits = 0).
- At the intersection cell (row x_A ,column y_B), their bits agree.

1		
0, 0	1	0
1		

$x = 2, y = 1$

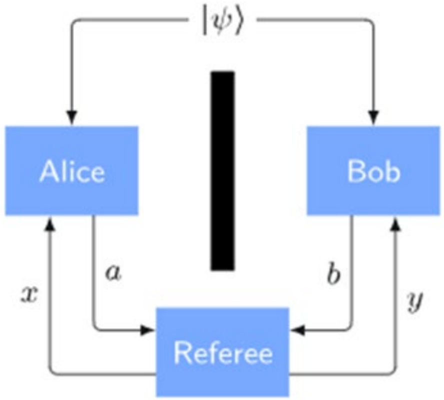
	1	
0	1,0	0
	0	

$x = 2, y = 2$

0,1	0,0	1,1
0, 0	1,1	0,0
1, 1	0,1	0,1

In the classical world, the best they could do was win 8/9 of the time. But if Alice and Bob fix a quantum entangled state before the game, the winning probability is 100% [3].

The Mermin–Peres Magic Square Game



Quantum Strategy:

- Alice and Bob fix a quantum state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$:
$$|\psi\rangle = \frac{1}{2}|0011\rangle + \frac{1}{2}|0110\rangle + \frac{1}{2}|1001\rangle + \frac{1}{2}|1100\rangle$$
- Alice can only measure the part of the quantum state in $\mathcal{H}_A$, and similarly for Bob.
- Alice and Bob come up with measurements that they will implement depending on what row/column the referee assigns.

Example: Suppose the referee gives Alice row 2 and Bob column 3. Alice and Bob "rotate" $|\psi\rangle$ by $X_2 \otimes Z_3$, and then the resulting quantum state "snaps" into one of these basis vectors with equal probability (since the coefficients all have the same modulus).

Suppose $X_2 \otimes Z_3|\psi\rangle \mapsto |0110\rangle$. They will win. This works for every $A_{row} \otimes Z_{col}$, and every possible measurement outcome.

		1
0	1	0, 0
		1

$x = 2, y = 3$

Paper I: Quantum Entanglement-based Advantage

Translation of the Magic Square Game to QML

- Sequence-to-Sequence Task Setup: The task R is a sequence translation problem, $R: \{0,1\}^{4n}\, x\, \{0,1\}^{4n} \rightarrow \{0,1\}$, constructed as an n −fold parallel repetition of a sub-task, R_0. The input sequence x is split into two remote halves (Player A and Player B).
- Sequence Validity Condition: $R(x, y) = 1$: The sequence translation if and only if:
  - All n parallel sub-tasks R_0 are satisfied simultaneously.
  - The i-th sub-task (R_i) is satisfied if at least one of these holds:
    - Trivial Case: Either input query x^Aor x^B (the two-bit queries for the sub-task) is 00
    - Magic Square Rule: The coordinated outputs satisfy the index-mapped equality: $y^B_{\{I(x^A)\}} = y^A_{\{I(x^B)\}}$, where, $I(x) = 2x_1 + x_2$ maps the 2-bit query to a specific bit index $\{1, 2, 3\}$ of the opponent's 3-bit parity-constrained answer, confirming non-local coordination.

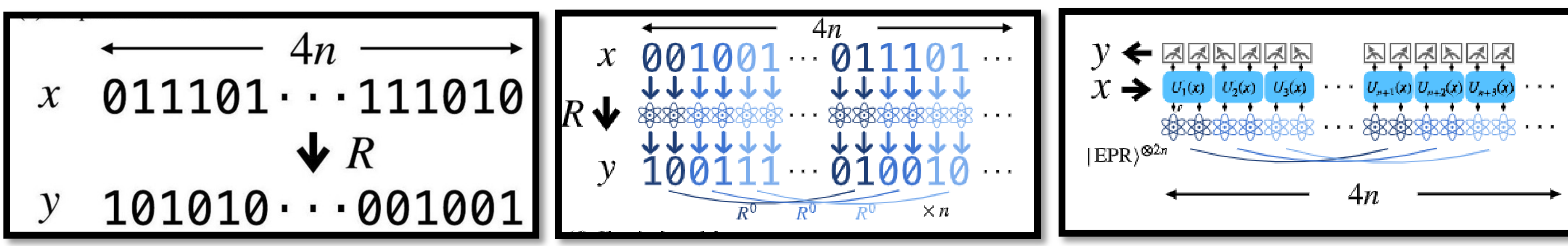


Figure 2: Schematic overview of the machine learning tasks and models.

Translation of the Magic Square Game to QML

- Sequence Validity Condition: $R(x, y) = 1$: The sequence translation if and only if:
  - All n parallel sub-tasks R_0 are satisfied simultaneously.
  - The i-th sub-task (R_i) is satisfied if at least one of these holds:
    - Trivial Case: Either input query x^Aor x^B (the two-bit queries for the sub-task) is 00
    - Magic Square Rule: The coordinated outputs satisfy the index-mapped equality: $y^B_{\{I(x^A)\}} = y^A_{\{I(x^B)\}}$, where, $I(x) = 2x_1 + x_2$ maps the 2-bit query to a specific bit index $\{01, 10, 11\} \Rightarrow \{1, 2, 3\}$ of the opponent's 3-bit parity-constrained answer, confirming non-local coordination.
    - For example: if $x = 01$, then $x_1 = 0, x_2 = 1 \Rightarrow I = 1$.

Example A - trivial (auto-valid)
Queries: $x_A = 00, x_B = 10$.
Because one query is 00 , the sub-task is valid no matter what y_A, y_B are.

Translation of the Magic Square Game to QML

- Sequence Validity Condition: $R(x, y) = 1$: The sequence translation if and only if:
  - All n parallel sub-tasks R_0 are satisfied simultaneously.
  - The i-th sub-task (R_i) is satisfied if at least one of these holds:
    - Trivial Case: Either input query x^Aor x^B (the two-bit queries for the sub-task) is 00
    - Magic Square Rule: The coordinated outputs satisfy the index-mapped equality: $y^B_{\{I(x^A)\}} = y^A_{\{I(x^B)\}}$, where, $I(x) = 2x_1 + x_2$ maps the 2-bit query to a specific bit index $\{01, 10, 11\}$ => $\{1, 2, 3\}$ of the opponent's 3-bit parity-constrained answer, confirming non-local coordination.
    - For example: if $x = 01$, then $x_1 = 0, x_2 = 1 \Rightarrow I = 1$.

Example B - non-trivial and valid
Queries: $x_A = 01$ (so $I(x_A) = 1$), $x_B = 11$ (so $I(x_B) = 3$).
Pick answers (2 bits each): $y_A = (1,0), y_B = (0,1)$.
Extend by parity:

$$y_{A3} = 1 \oplus 0 \oplus 1 = 0 \Rightarrow y_A = (1,0,0)$$
$$y_{B3} = 0 \oplus 1 = 1 \Rightarrow y_B = (0,1,1).$$

Magic-square check:
compare $y_B[I(x_A) = 1]$ with $y_A[I(x_B) = 3]$: $y_B[1] = 0$ and $y_A[3] = 0 \Rightarrow$ equal $\Rightarrow$ valid.

Paper I: Quantum Entanglement-based Advantage

Translation of the Magic Square Game to QML

- Sequence Validity Condition: $R(x, y) = 1$: The sequence translation if and only if:
  - All n parallel sub-tasks R_0 are satisfied simultaneously.
  - The i-th sub-task (R_i) is satisfied if at least one of these holds:
    - Trivial Case: Either input query x^Aor x^B (the two-bit queries for the sub-task) is 00
    - Magic Square Rule: The coordinated outputs satisfy the index-mapped equality: $y^B_{\{I(x^A)\}} = y^A_{\{I(x^B)\}}$, where, $I(x) = 2x_1 + x_2$ maps the 2-bit query to a specific bit index $\{01, 10, 11\} => \{1, 2, 3\}$ of the opponent's 3-bit parity-constrained answer, confirming non-local coordination.
    - For example: if $x = 01$, then $x_1 = 0, x_2 = 1 \Rightarrow I = 1$.

Example C — non-trivial and invalid
Queries: $x_A = 10$ (so $I = 2$), $x_B = 11$ (so $I = 3$).
Pick answers: $y_A = (0,0), y_B = (0,0)$.
Extend:

$$y_{A3} = 0 \oplus 0 \oplus 1 = 1 \Rightarrow y_A = (0,0,1)$$
$$y_{B3} = 0 \oplus 0 = 0 \Rightarrow y_B = (0,0,0)$$

Check: $y_B[I(x_A) = 2] = y_B[2] = 0$ vs $y_A[I(x_B) = 3] = y_A[3] = 1 \Rightarrow$ not equal $\Rightarrow$ invalid.

Paper I: Quantum Entanglement-based Advantage

Translation of the Magic Square Game to QML

- **Classical Bottleneck: Communication:**
  - Classical sequence models must use internal hidden states or latent dimensions d that scale linearly ($\Omega(n)$) to carry long-range context and coordinate the two halves (i.e., communication capacity $c = O(d) \geq \Omega(n)$.
  - Without this linearly increasing resource, the classical success probability decays exponentially.

- **Quantum Solution: Entanglement Substitution:**
  - Role of Pre-computed Bell Pairs (EPR States): Entanglement provides the perfect non-local correlation necessary to coordinate the outputs of the two halves instantly.
  - Why $2n$ Bell Pairs? The total task R is an n-fold parallel repetition of the basic magic square sub-task, R_0. Since the quantum winning strategy for the base sub-task requires two Bell pairs to achieve certainty, the total resource scales linearly with the sequence length n, requiring $2n$ Bell pairs.
  - Outcome: This resource substitution enables the quantum model to achieve a perfect score ($\omega^* = 1$) using a constant, $O(1)$-parameter circuit, replacing the necessary $\Omega(n)$ classical communication bits.

Paper I: Quantum Entanglement-based Advantage

Title: Entanglement-Induced Provable and Robust Quantum Learning Advantages

Theorem 1: Noiseless Inference

For the magic square translation task $R: \{0,1\}^{4n} \times \{0,1\}^{4n} \rightarrow \{0,1\}$, there exists an $O(1)$-parametersize quantum model $\mathcal{M}_Q$ that can achieve a score $S(\mathcal{M}_Q) = 1$ using $2n$ Bell pairs. Meanwhile, any communication-bounded classical model $\mathcal{M}_C$ that can achieve a score $S(\mathcal{M}_C) \geq 2^{-o(n)}$ must have $\Omega(n)$ parameter size.

Main Idea

Establishes exponential separation in complexity (constant vs. linear scaling).

Paper I: Quantum Entanglement-based Advantage

Title: Entanglement-Induced Provable and Robust Quantum Learning Advantages

Theorem 2: Inference Advantage with Constant Noise

For any noise strength $p \leq p^\star \approx 0.0064$, there exists a noisy magic square translation task $R_p: \{0,1\}^{4n} \times \{0,1\}^{4n} \to \{0,1\}$, such that it can be solved by an $O(1)$-parameter-size quantum model $\mathcal{M}_Q$ under depolarization noise of strength p with score $S(\mathcal{M}_Q) \geq 1 - 2^{-\Omega(n)}$ using $2n$ Bell pairs. Meanwhile, any communication bounded classical model $\mathcal{M}_C$ that can achieve a score $S(\mathcal{M}_C) \geq 2^{-o(n)}$ must have $\Omega(n)$ parameter size.

Main Idea

Demonstrates robustness; advantage persists under moderate, constant noise levels.

Paper I: Quantum Entanglement-based Advantage

Title: Entanglement-Induced Provable and Robust Quantum Learning Advantages

Theorem 3: Training Advantage

There exists a training algorithm that, with probability at least $2/3$, takes $\{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$ with size $M = \Theta(nN) = O(1)$ as input and outputs the optimal quantum model $\mathcal{M}_Q$ for the task R. Moreover, the running time of this training algorithm is $T = O(1)$.

Main Idea

Shows constant resource requirements for training the quantum model.

Paper I: Quantum Entanglement-based Advantage

Quantum Model:

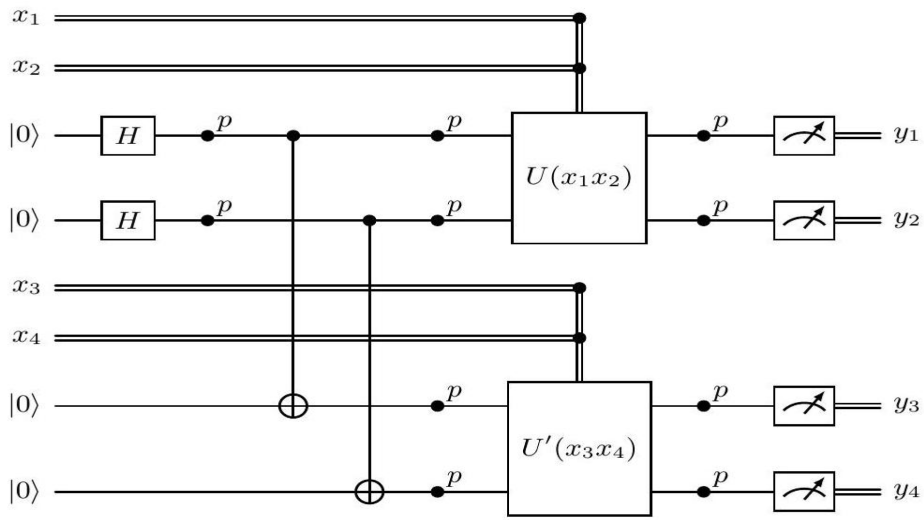
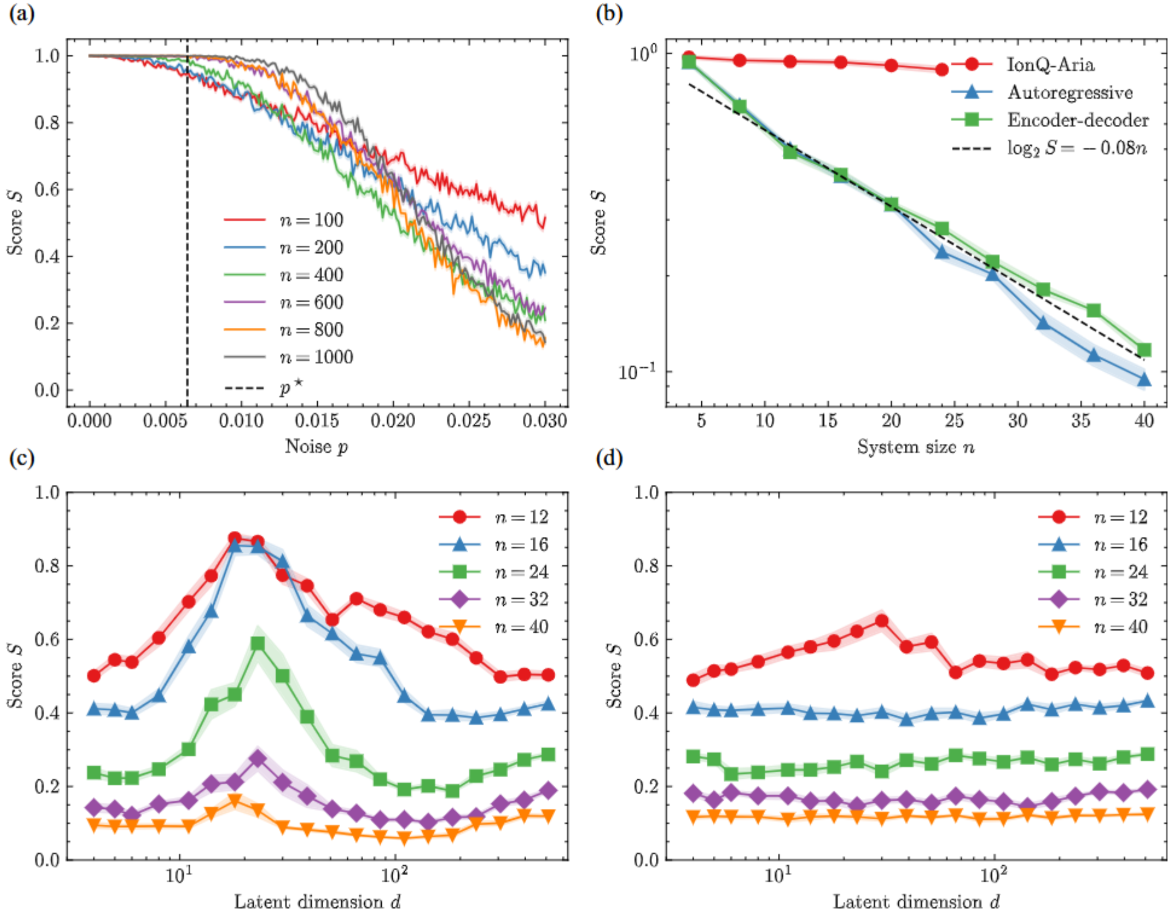


Figure 3: Illustration of the quantum model for R_0under depolarization noise of strength p.

Alg	Scaling	Formula	Train Params	Depth
QC	$O(1)$	$Q \times \dim U(4) \times \|\{0,1\}^2\|$ = $2 \times 4 \times 4 \times 4$ = 128	128	$O(1)$
GRU (1-layer)	$\Omega(n^2)$	$\sim 3(h^2 + mh + h), h \propto n$	$\sim 3 \times 10^4$ (with $h = n, m \leq 4$)	$\Omega(n)$
Enc-Dec	$\Omega(n^2)$	$\sim 2 \times 3(h^2 + mh + h)$	$\sim 6 \times 10^4$	$\Omega(n)(enc) + \Omega(n)(dec)$

Table 1: Comparative parameter count and depth for sequence length, $n = 100$.
h = hidden size; m= input size per step; Q=number of qubit

Paper I: Quantum Entanglement-based Advantage



The quantum model achieves a **perfect score** on *this task* while using **about two to three orders of magnitude fewer** trainable params than GRU-based autoregressive and encoder–decoder models—and its parameter count stays **constant** as inputs get longer.

Figure 4: Results on numerical simulations and trapped-ion experiments on IonQ Aria (a,b). Performance of classical autoregressive (c) and encoder-decoder (d) models on the magic sqaure translation task R with different problem size n and model size d.

Paper I: Quantum Entanglement-based Advantage

Critical Analysis:

Why the advantage arises:

- The proof reframes sequence learning as **communication capacity**: classical models must *carry* information across the sequence; entanglement sidesteps this by **reducing required classical communication**, so the quantum model's parameters stay $O(1)$while classical size must be $\Omega(n)$.

Open questions:

- **Task generality:** Does the advantage hold beyond this very specific rule-based sequence task?
- **Continuous inputs:** How do these results—proved for bitstring inputs/outputs—extend to continuous inputs: which encoding (amplitude, angle, or basis) preserves the advantage, and how does that encoding affect the communication lower bounds and overall quantum advantage?
- **Expressivity ceiling (gate set):** Given the model uses only Clifford gates (Pauli→Pauli under conjugation; generated by H, S, CNOT), can we add a minimal set of non-Clifford operations while keeping constant parameter count and depth, and how would that change expressivity and provable advantage?
- **Noise tolerance:** Can the advantage survive higher, device-realistic noise levels?

Paper II: Entangled Data and Quantum NFL

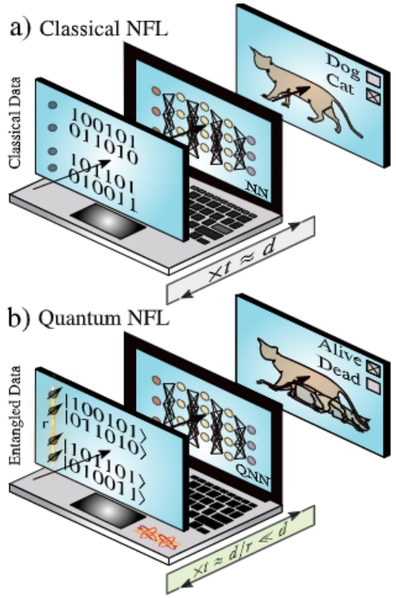
Title: Reformulation of the No-Free-Lunch Theorem for Entangled Datasets

Classical No-Free-Lunch (NFL) Theorem:

- Learn an unknown function $f: X \to Y$ from a training set S of t labeled examples; evaluate generalization via the risk.
- Core claim (averaged over all S and all f): the average risk is lower-bounded only by dataset and alphabet sizes-not by the learning method:

$$\mathbb{E}_f\left[\mathbb{E}_S\left[R_f(h_S)\right]\right] \geq \left(1 - \frac{1}{d_Y}\right)\left(1 - \frac{t}{d_X}\right)$$

- More data (t) ⇒ lower average error; the bound hits 0 only when the training set covers the whole domain $(t = d_X)$.



Paper II: Entangled Data and Quantum NFL

Title: Reformulation of the No-Free-Lunch Theorem for Entangled Datasets

Problem Statement

The original Quantum No-Free-Lunch (Q-NFL) theorem suggested QML required an exponential number of training samples ($t = d = 2^n$) to learn an unknown unitary process completely.

Solution Method

The reformulation addresses the task of learning an unknown unitary using entangled training set. The Schmidt rank (r) quantifies the strength of entanglement used as a resource.

How Entanglement Solves it

A single maximally entangled training pair ($r = d$) is theoretically sufficient to uniquely specify the unknown unitary, drastically reducing sample complexity.

Paper II: Entangled Data and Quantum NFL

Title: Reformulation of the No-Free-Lunch Theorem for Entangled Datasets

Theorem 1: Quantum NFL

For d-dimensional input/output spaces and a training set $\mathcal{S}_Q$ of t input-output pairs whose states all have Schmidt rank, r across $\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{R}$, the average quantum risk of any perfectly trained hypothesis unitary $V_{\mathcal{S}_Q}$ obeys

$$\mathbb{E}_U\left[\mathbb{E}_{\mathcal{S}_Q}\left[R_U\left(V_{\mathcal{S}_Q}\right)\right]\right] \geq 1 - \frac{r^2t^2 + d + 1}{d(d+1)}$$

The inequality is tight (saturated when the training inputs are linearly independent).

Main Idea

Entanglement in the training data reduces the fundamental error floor: the product $r\times t$ is the key resource. With no entanglement (r=1), vanishing risk needs t=d (exponentially large). With maximal entanglement (r=d), a single training pair can drive the lower bound to zero.

Paper II: Entangled Data and Quantum NFL

Title: Reformulation of the No-Free-Lunch Theorem for Entangled Datasets

- Quantum risk $R_U\left(V_{S_Q}\right)$ is nonnegative by definition because it is an average of squared trace distances between the target unitary, U and hypothesis V. So, we can rewrite the Quantum NFL theorem as:

$$\mathbb{E}[R] \geq \max\left\{0, 1 - \frac{r^2t^2 + d + 1}{d(d+1)}\right\}.$$

- Numerically, if we have only one training example, $t = 1$, and maximal entanglement $(d = r = 8)$, then,

$$1 - \frac{8^2 1^2 + 8 + 1}{8(8+1)} = -\frac{1}{72}.$$

So, $\mathbb{E}[R] = 0$.

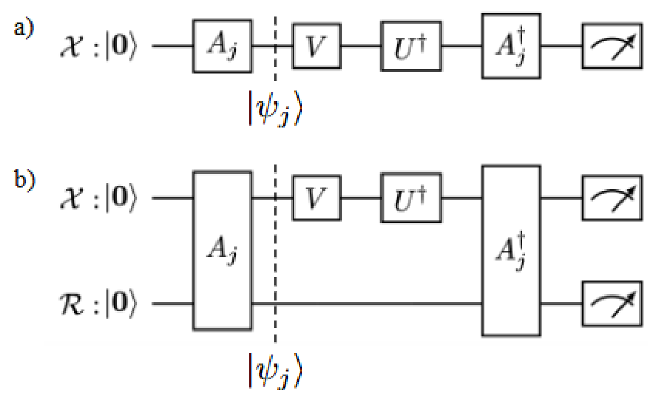
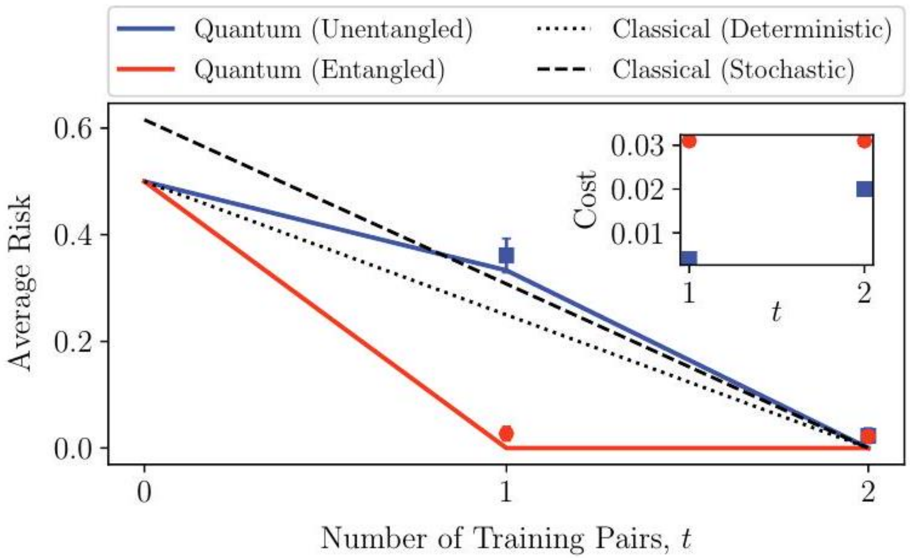


Figure 5: Diagram of quantum circuits w/o ent (a), and with ent (b).

Paper II: Entangled Data and Quantum NFL

- The average risk was calculated over 10 random unitaries and 10 random training sets for the 2-dimensional implementation on the Rigetti quantum computer.
- Over 10 random unitaries and 100 random training sets in the case of the 64-dimensional implementation on the simulator.
- Baselines: Theoretical classical NFL bounds — deterministic/permutation maps and stochastic maps (no specific classical ML models used).



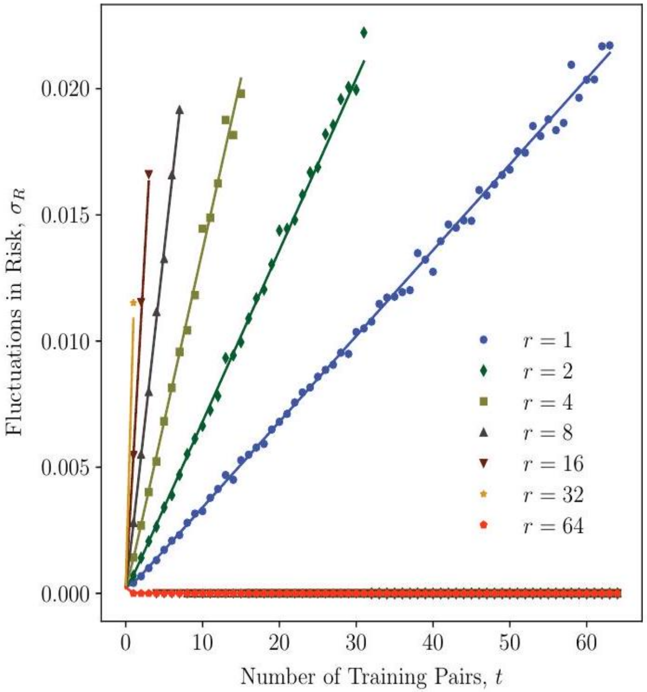


Figure 6: Average risk vs. training pairs t; entangled data outperforms unentangled and classical baselines (left). Risk fluctuations drops as entanglement r and t increase (right).

Paper II: Entangled Data and Quantum NFL

Critical Analysis:

Why the advantage arises:

- Entanglement turns *process learning* into *state learning* meaning an entangled input probes all input directions at once. The bound depends only on $r \times t$ (effective information), not on the learning algorithm.

Open questions:

- **Imperfect training:** If training isn't exact, how does the bound deform?
- **Noise & robustness:** How do gate noise alter the $r \times t$ scaling? Can we derive **noise-aware** lower bounds?
- **Entangled dataset:** Where do entangled datasets come from in practice? How can we detect and use entangled signals in a real-world classical dataset?

Paper III: Quantum Discord-based Advantage

Title: The power of one clean qubit in supervised machine learning.

Problem Statement

Entanglement is highly susceptible to real-world noise (a major constraint in the NISQ era). The challenge is developing an efficient, robust quantum advantage for supervised learning that avoids fragile multi-qubit measurements

Solution Method

Use the minimal-resource Deterministic Quantum Computing with One Qubit (DQC1) framework to estimate complex kernel functions for classification. The architecture uses only a single, high-quality ("clean") qubit, coupling it to a register of target qubits that can be in a mixed (noisy) state.

How Discord Solves it

Quantum discord ensures robustness against noise, and the estimation cost (shot complexity) is independent of the system size n (qubit count).

Paper III: Quantum Discord-based Advantage

Title: The power of one clean qubit in supervised machine learning.

Quantum Coherence

It describes a system's ability to maintain a stable, synchronized phase relationship between its quantum states, allowing it to exist in multiple states simultaneously through superposition.

Theorem 1: Coherence Consumption Metric

The change in coherence ΔC of the control qubit is directly linked to the kernel's discriminative power:

$$\Delta C(\mathbf{x}, \mathbf{x}') = H_2\left(\frac{1 - |K(\mathbf{x}, \mathbf{x}')|}{2}\right).$$

where $H_2(x)$ is the binary Shannon entropy, and $|K(\mathbf{x}, \mathbf{x}')|$ is the kernel method which can also be expressed as $K(\mathbf{x}, \mathbf{x}') = \frac{\mathrm{tr}\left(U_n(\mathbf{x}, \mathbf{x}')\right)}{2^n}$.

Main Idea

If zero coherence is consumed, the kernel cannot learn.

Paper III: Quantum Discord-based Advantage

Title: The power of one clean qubit in supervised machine learning

Theorem 2: Discord-Coherence Trade-off

Discord is bounded by consumed coherence:

$$D_G(\rho_f) \leq \Delta C$$

where, D_G is geometric discord, a distance-based measure of "quantum-but-not-entangled" correlations, and ρ_f is the final quantum state.

Main Idea

The nonclassical correlation we create (quantum discord) is upper-bounded by how much coherence we spend from the single clean qubit. This ties the kernel signal we read out to a concrete resource—coherence—giving a simple, hardware-friendly knob for performance vs. noise.

Paper III: Quantum Discord-based Advantage

Title: The power of one clean qubit in supervised machine learning.

- Suppose we build a data-dependent unitary:

$$U_n(x) = \exp\left(i \sum_{j=1}^{n} \phi_j(x) Z_j + i \sum_{j<k} \phi_{jk}(x) Z_j Z_k \right).$$

where x is the classical data point (feature vector), and the functions $\phi_j(x)$ and $\phi_{jk}(x)$ are the feature maps that convert features of x (and pairwise interactions of features) into rotation angles.

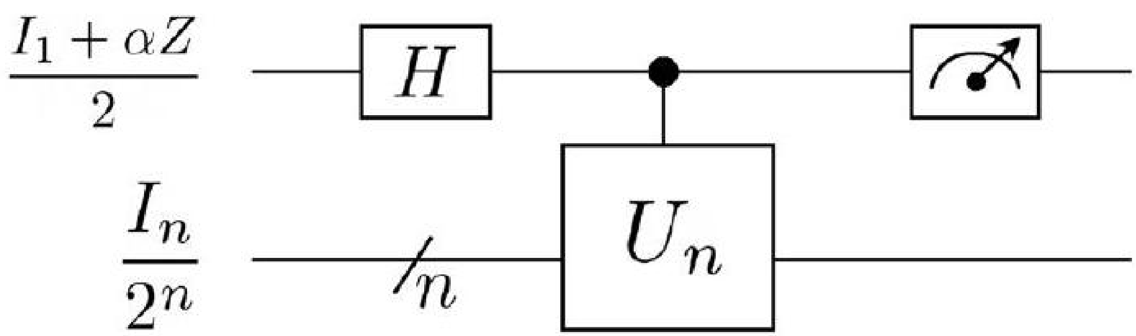


Figure 6: The circuit representation of the DQC1 algorithm. The input states for control and target qubits are $\frac{I_1+\alpha Z}{2}$, with $\alpha \in [0,1]$ and $\frac{I_n}{2^n}$, respectively. H and Z denote the Hadamard and Pauli Z gates.

Paper III: Quantum Discord-based Advantage

Title: The power of one clean qubit in supervised machine learning.

- In DQC1, the control register is the single clean qubit (not a multi-qubit block) that controls the data-dependent unitary acting on the mixed n-qubit register.
- The control is initialized in a partially pure state $\rho_c = \frac{1}{2}(I + \alpha Z)$ (purity $\alpha \in [0,1]$), apply a Hadamard, and couple it to the mixed register $I/2^n$ via a controlled- U_n.
- The control qubit accumulates the global interference from U_n; its off-diagonals encode $\mathrm{tr}(U_n)/2^n$. Measuring only the control (e.g., X and Y) gives

$\langle X \rangle = \alpha \mathrm{Re}\left(\frac{\mathrm{tr}(U_n)}{2^n}\right), \langle Y \rangle = -\alpha \mathrm{Im}\left(\frac{\mathrm{tr}(U_n)}{2^n}\right)$, which is exactly the kernel entry we need.

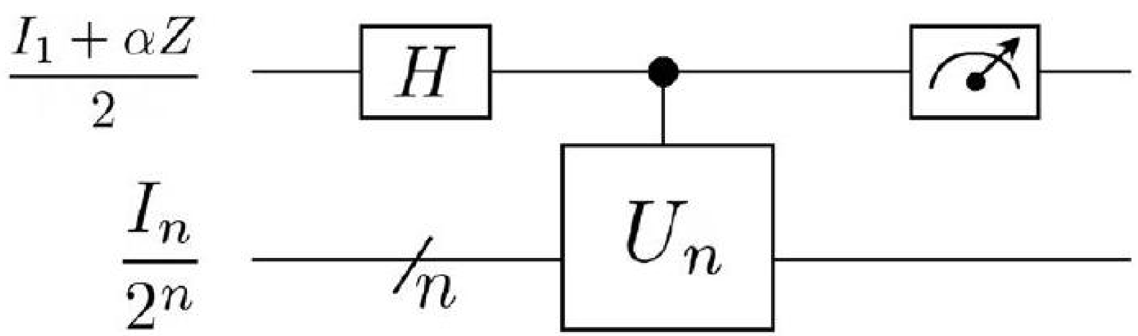


Figure 7: The circuit representation of the DQC1 algorithm. The input states for control and target qubits are $\frac{I_1+\alpha Z}{2}$, with $\alpha \in [0,1]$ and $\frac{I_n}{2^n}$, respectively. H and Z denote the Hadamard and Pauli Z gates.

Paper III: Quantum Discord-based Advantage

Title: The power of one clean qubit in supervised machine learning.

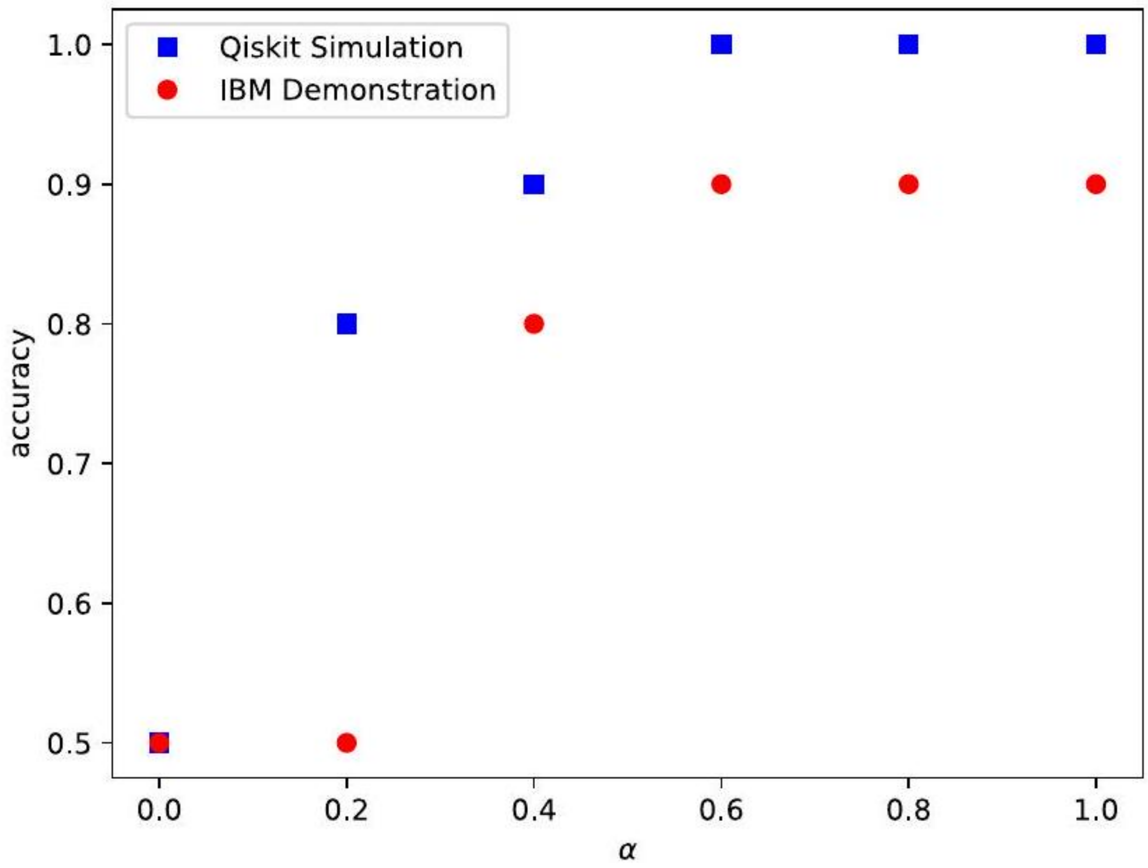


Figure 8: The accuracy as a function of the control qubit's purity for the “adhoc_dataset” is shown. Note that when $\alpha = 0$, the state is in a completely mixed state, and when $\alpha = 1$, the state is pure. The blue curve (square) indicates simulation results, and the red curve (circle) shows the results obtained from IBM hardware.

Paper III: Quantum Discord-based Advantage

Critical Analysis:

Why the advantage arises:

- Measuring one qubit per kernel entry is a big advatange given readout is one of the noisiest ops on superconducting devices. The shot complexity ($O(\epsilon^{-2}\alpha^{-2}\log(1/\delta))$) is independent of the qubit size. This lowers readout noise and decouples sampling cost from n, yielding a hardware-friendly estimator.

Open questions:

- **Expressivity vs. robustness:** When, precisely, do discord-based DQC1 kernels underperform entanglement-dependent quantum kernels?
- **Benchmarking vs. classical baselines:** How do DQC1 kernels compare to strong classical kernels regarding accuracy and computational speed?
- **Benchmarking vs. other quantum methods:** Where do DQC1 kernels sit relative to alternative quantum kernel estimators?
- **Role of the feature map on accuracy:** How exactly does the choice of feature map $Un(x)$ impact the accuracy and generalization?

Summary

QCorr	Advantage Domain	Mechanism of Advantage	NISQ Relevance
Entanglement	Computational Speed in sequence translation task	Substitutes Classical Communication	High robustness to constant noise
Entanglement	Data Efficiency (Learning Theory)	Amplifies information per sample ($r \times t$).	Theoretical, sensitive to noise
Discord	Noise Resilience (Kernel Estimation)	Converts local Coherence into Correlation (DQC1)	High, due to resilience in mixed states

- **Gap:** Most proofs assume discrete data, perfect training conditions and a constant device noise, but in practice, machine learning deals with continuous data, training error and limited-noisy quantum hardware.

- We need a QML research roadmap that is **theoretically complete, measurable, and deployable**.

Roadmap

generalized learning theory — data/encoding — metric — benchmarks — system — application

1) Generalized Q-NFL: bounds on expected generalization risk as a function of n (samples), usable correlation $\tilde{r}$ (post-noise Schmidt rank), depth L, and noise params $\vec{\eta}$.

2) Generalized ML task-based advantage theorem for training and inference.

1) Scalable encoding for continuous, high D-data.

2) Theory bound on diff. encoding method (angle/amplitude) as a function of accuracy and resource.

1) Eval metric beyond p-value and classical statistical significance.

2) Quantum-native metrics such as fidelity, coherence used, wall clock, quantifiable Qcorr, energy consumption, etc.

1) Generate benchmark datasets using tunable entanglement measures.

2) Beyond MNIST classical dataset.

2) Comprehensive classical and quantum algorithms' performance on the benchmark datasets.

1) Hybrid pipeline.

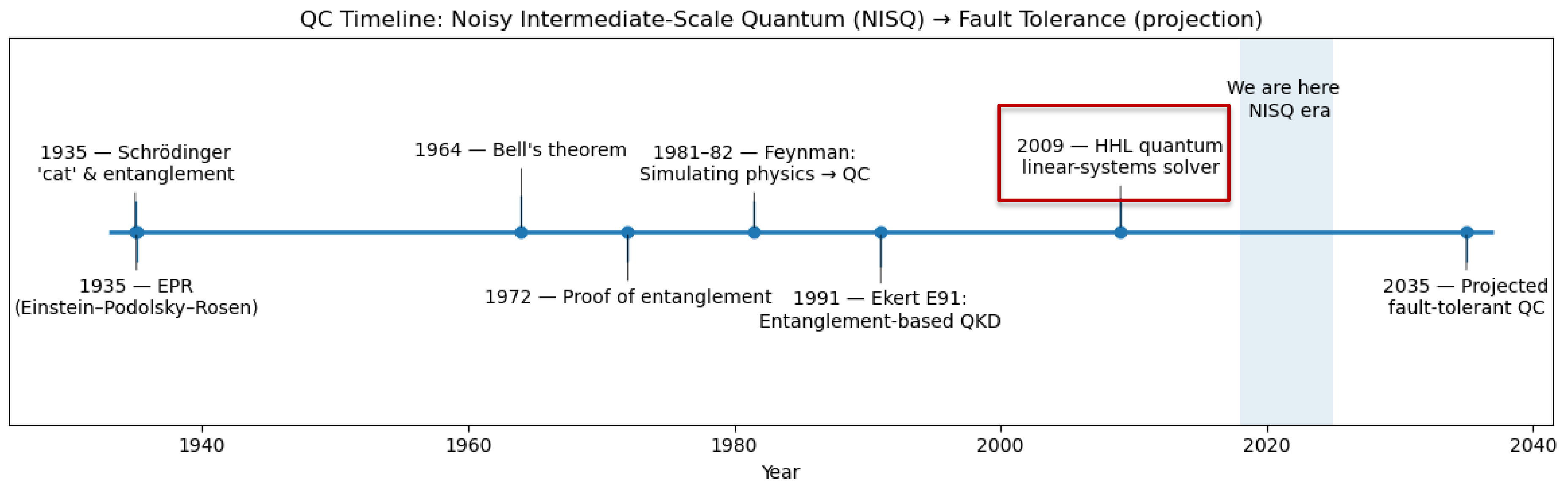
2) Entanglement for expressivity.

3) Discord for robustness with classical head.

1) Pilot domain: EEG signal segmentation, multimodal classification.

Success: ↑accuracy at equal shots/wall-clock or ↓shots at equal accuracy.

Closing Thoughts



Closing Thoughts

- It's too early to either claim or dismiss quantum advantages for most ML tasks—uncertainty here is what makes the field exciting.
- Why are we interested in machine learning?
  - "solve intelligence", "wants to understand how the human brain works",..
- By focusing on the human brain, we risk ignoring other intelligences—animal cognition, plant signaling/coordination, and non-biological forms [4].
- Quantum ideas (e.g., Penrose-Hameroff Orch OR hypothesis [5]) keep open the possibility that aspects of cognition may exploit quantum effects.
- Also, experimentally, scientists have discovered quantum properties [6] in neurons—miraculously, these properties persist despite the noisy environment.
- None of this proves the brain is a quantum computer; it simply argues against premature conclusions.
- Because quantum mechanics underlies all physical processes, it's rational to explore quantum approaches alongside classical ones in the quest for building artificial general/super intelligence (AGI/ASI).

References

[1] S. Golestan, M. R. Habibi, S. Y. Mousazadeh Mousavi, J. M. Guerrero, and J. C. Vasquez, "Quantum computation in power systems: An overview of recent advances," *Energy Reports*, vol. 9, pp. 584–596, Dec. 2023, doi: *10.1016/j.egyr.2022.11.185*.

[2] H. Ollivier and W. H. Zurek, "Quantum Discord: A Measure of the Quantumness of Correlations," *Phys. Rev. Lett.*, vol. 88, no. 1, p. 017901, Dec. 2001, doi: *10.1103/PhysRevLett.88.017901*.

[3] P. Ball, "Researchers Use Quantum 'Telepathy' to Win an 'Impossible' Game," Scientific American. Accessed: Oct. 24, 2025. [Online]. Available: https://www.scientificamerican.com/article/researchers-use-quantum-telepathy-to-win-an-impossible-game/

[4] M. Bennett, *A Brief History of Intelligence: Evolution, AI, and the Five Breakthroughs That Made Our Brains*. Mariner Books, 2023. [Online]. Available: https://www.amazon.com/Brief-History-Intelligence-Humans-Breakthroughs/dp/0063286343

[5] "Orchestrated objective reduction," *Wikipedia*. Oct. 17, 2025. Accessed: Oct. 24, 2025. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Orchestrated_objective_reduction&oldid=1317310693

[6] N. S. Babcock, G. Montes-Cabrera, K. E. Oberhofer, M. Chergui, G. L. Celardo, and P. Kurian, "Ultraviolet Superradiance from Mega-Networks of Tryptophan in Biological Architectures," *J. Phys. Chem. B*, vol. 128, no. 17, pp. 4035–4046, May 2024, doi: *10.1021/acs.jpcb.3c07936*.

Thank You!