Ph 220: Quantum Learning Theory
Lecture Note 8: Hardness of Learning and Pseudorandom States

Hsin-Yuan Huang (Robert)

California Institute of Technology

1 Introduction

In previous lectures, we have focused on what is *possible* in quantum learning: randomized measurement allows us to learn approximate classical descriptions of any quantum state, shadow tomography allows us to learn many properties using very few samples, local inversion enables us to learn shallow quantum neural networks, and reshaping techniques provides us the tools for reducing the problem of learning many-body Hamiltonians to learning single-qubit Hamiltonians. Today, we pivot to the converse question: **What is fundamentally hard to learn?**

A fundamental question focuses on how to learn to distinguish simple states from complicated states. Let us imagine an unknown n-qubit state coming from two different worlds:

- **World A (Polynomial Complexity):** The quantum state $|\psi\rangle$ can be generated by a polynomial-size quantum circuit acting on the zero state $|0\rangle^{\otimes n}$. These are the states we can physically prepare in a lab or on a future quantum computer. Due to the polynomial complexity, such a state must be much more structured than a Haar-random state.
- **World B (Exponential Complexity):** The quantum state $|\psi\rangle$ can only be generated by an exponential-size quantum circuit acting on the zero state $|0\rangle^{\otimes n}$. By simple counting arguments (or volume arguments), the vast majority of states in the Hilbert space (all except an exponentially small fraction) require exponential circuit depth to approximate. Hence, we focus on states sampled from the Haar measure as the states in this world[1].

Suppose I hand you a black box that produces copies of a quantum state $|\psi\rangle$. Your task is to act as a discriminator. You must decide:

Did this state come from World A (efficient/structured) or World B (inefficient/random)?

If there exists an ensemble of states in World A that mimics random states in World B so well that no polynomial-time algorithm can distinguish them, we call these **Pseudo-Random States (PRS)**. If PRS exist, it implies a fundamental hardness of learning: there are efficient states that "look" maximally complex, hiding their structure from any efficient learner.

2 Statistical Indistinguishability

Before constructing these states, we must rigorously define what it means for two ensembles to be "indistinguishable." This relies on the concept of **Trace Distance**.

[1] There is an exponentially small chance that a Haar-random state has a polynomial complexity. The probability is so small that a random state from all states with exponential complexity is the same as a random state from all states for all physical purposes.

3 Statistical Indistinguishability

To prove that learning is hard, we must show that no efficient algorithm can distinguish between states from World A and World B.

3.1 The Distinguishing Game

Consider a challenger and a learner (the algorithm $\mathcal{A}$) that consumes t copies of the unknown state $|\psi\rangle$ to distinguish between the two worlds. The game proceeds as follows:

1. The challenger flips a fair coin $b \in \{0, 1\}$.
   - If $b = 0$, they consider the Pseudo-Random ensemble (World A).
   - If $b = 1$, they consider the Haar-Random ensemble (World B).
2. The challenger samples a *specific* pure state $|\psi\rangle$ from the chosen ensemble $\mathcal{E}_b$.
3. The challenger gives the learner t copies of this specific state: $|\psi\rangle^{\otimes t}$.
4. The learner applies a quantum algorithm (measurements) and outputs a guess $b' \in \{0, 1\}$.

If the algorithm $\mathcal{A}$ runs in polynomial time, then t must be polynomial.

3.2 Prediction Accuracy and Linearity

We define the learner's success by their **Prediction Accuracy**:

$$\mathrm{Acc}(\mathcal{A}) = \Pr[b' = b] = \frac{1}{2}\Pr[b' = 0|b = 0] + \frac{1}{2}\Pr[b' = 1|b = 1]. \tag{1}$$

Any quantum learning algorithm with binary output corresponds to a POVM measurement $\{M_0, M_1\}$, where $M_0 + M_1 = I$. If the input state is ρ_{in}, the probability of outputting 1 is $\mathrm{Tr}(M_1\rho_{\text{in}})$. In our game, the input state is $|\psi\rangle^{\otimes t}$. Thus:

$$\Pr[b' = 1|\text{state is } \psi] = \mathrm{Tr}(M_1|\psi\rangle\langle\psi|^{\otimes t}).$$

However, we want the probability averaged over the entire ensemble $\mathcal{E}_b$. By the linearity of the trace and the expectation:

$$\begin{aligned}\Pr[b' = 1|b = 1] &= \mathbb{E}_{\psi\sim\text{Haar}}\left[\mathrm{Tr}(M_1|\psi\rangle\langle\psi|^{\otimes t})\right] \\ &= \mathrm{Tr}\left(M_1\,\mathbb{E}_{\psi\sim\text{Haar}}[|\psi\rangle\langle\psi|^{\otimes t}]\right).\end{aligned} \tag{2}$$

This reveals that the relevant mathematical object is not the tensor power of the average, but the average of the tensor powers. Let us define these moment operators:

$$\rho_{\text{A}}^{(t)} \coloneqq \mathbb{E}_{\psi\sim\text{PRS}}[|\psi\rangle\langle\psi|^{\otimes t}], \qquad \rho_{\text{B}}^{(t)} \coloneqq \mathbb{E}_{\psi\sim\text{Haar}}[|\psi\rangle\langle\psi|^{\otimes t}]. \tag{3}$$

We can now rewrite the prediction accuracy. Let the algorithm guess "1" if it measures outcome M_1. The accuracy is:

$$\mathrm{Acc}(\mathcal{A}) = \frac{1}{2}\left(1 - \mathrm{Tr}(M_1\rho_{\text{A}}^{(t)})\right) + \frac{1}{2}\mathrm{Tr}(M_1\rho_{\text{B}}^{(t)}) \tag{4}$$

$$= \frac{1}{2} + \frac{1}{2}\mathrm{Tr}(M_1(\rho_{\text{B}}^{(t)} - \rho_{\text{A}}^{(t)})). \tag{5}$$

3.3 Relating Accuracy to Trace Distance

We define the **Advantage** of the algorithm as how much better it performs than random guessing:

$$\mathrm{Acc}(\mathcal{A}) = \frac{1}{2} + \frac{1}{2}\mathrm{Adv}(\mathcal{A}).$$

From the equation above, maximizing accuracy is equivalent to maximizing the quantity $\mathrm{Tr}(M_1(\rho_{\mathrm{B}}^{(t)} - \rho_{\mathrm{A}}^{(t)}))$. This brings us to the Helstrom bound.

Definition 1 (Trace Distance)**.** *The trace distance between two density matrices ρ and σ is $\|\rho - \sigma\|_1 = \mathrm{Tr}\sqrt{(\rho-\sigma)^\dagger(\rho-\sigma)}$.*

Theorem 1 (Helstrom Bound)**.** *The maximum distinguishing advantage is exactly half the trace distance between the ensemble moments:*

$$\max_{\mathcal{A}} Adv(\mathcal{A}) = \frac{1}{2}\left\|\rho_A^{(t)} - \rho_B^{(t)}\right\|_1. \tag{6}$$

Proof. Let $\Delta = \rho_{\mathrm{B}}^{(t)} - \rho_{\mathrm{A}}^{(t)}$. This matrix is Hermitian and traceless. We can diagonalize it and separate positive eigenvalues (P) and negative eigenvalues (N): $\Delta = P - N$. The trace norm is $\|\Delta\|_1 = \mathrm{Tr}(P) + \mathrm{Tr}(N)$. Since $\mathrm{Tr}(\Delta) = 0$, we have $\mathrm{Tr}(P) = \mathrm{Tr}(N) = \frac{1}{2}\|\Delta\|_1$. The advantage is $\mathrm{Tr}(M_1\Delta)$. To maximize this, we choose M_1 to be the projector onto the subspace spanned by the eigenvectors of Δ with positive eigenvalues (the support of P). Then:

$$\mathrm{Tr}(M_1\Delta) = \mathrm{Tr}(P) = \frac{1}{2}\|\Delta\|_1.$$

This completes the proof. □

Together, we can see that if we can prove that the trace distance between the t-th moments of the Pseudo-Random ensemble and the Haar ensemble is negligible for any polynomial t (e.g., $\mathcal{O}(t^2/2^n)$), then **no** polynomial-time quantum learning algorithm (no matter how clever) can distinguish the two worlds with probability visibly better than 1/2.

4 Cryptographic Ingredients: PRFs

We have established that if we can construct an ensemble from World A that is statistically indistinguishable from the Haar measure (World B), learning becomes hard. To construct such an ensemble efficiently, we rely on one of the most important cryptographic primitives known as a **Pseudo-Random Function (PRF)**.

4.1 The Concept of a PRF

Intuitively, a PRF is a function that can be computed efficiently using a secret key, but whose output looks completely random to anyone who does not know the key. Let $\mathcal{K}$ be the key space (e.g., $\{0,1\}^\lambda$). A PRF family is a set of functions:

$$\mathcal{F} = \{f_k : \{0,1\}^n \to \{0,1\} \mid k \in \mathcal{K}\}.$$

Security is defined via a game. A distinguisher D is given oracle access to a function $\mathcal{O}$ and must decide if $\mathcal{O}$ is:

1. A specific function f_k chosen randomly from the family $\mathcal{F}$ (Pseudo-Random).
2. A truly random function h chosen uniformly from the set of all functions $\{0,1\}^n \to \{0,1\}$.

4.2 Quantum Security and Superposition Access

This is where the definition becomes subtle in the quantum setting. In classical cryptography, the adversary D can only query the oracle on specific inputs x and receive $f(x)$. However, a quantum learner can interact with the black box in significantly more powerful ways. A **Quantum-Secure PRF** must remain indistinguishable even if the adversary can query the function *in superposition*. This is modeled by a quantum oracle $\mathcal{U}_{\mathcal{O}}$ that acts as:

$$\mathcal{U}_{\mathcal{O}} \sum_{x,y} \alpha_{x,y} |x\rangle |y\rangle = \sum_{x,y} \alpha_{x,y} |x\rangle |y \oplus \mathcal{O}(x)\rangle . \tag{7}$$

Equivalently, for boolean functions, we often consider the phase oracle:

$$\mathcal{O}_{\pm} \sum_{x} \alpha_x |x\rangle = \sum_{x} \alpha_x (-1)^{\mathcal{O}(x)} |x\rangle . \tag{8}$$

The two oracle definitions are equivalent as each one can easily implement the other one.

Definition 2 (Quantum-Secure PRF)**.** *A family $\mathcal{F}$ is a Quantum-Secure PRF if for any polynomial-time quantum algorithm D with superposition query access to the oracle:*

$$\left| \Pr_{k\sim\mathcal{K}}[D^{f_k} = 1] - \Pr_{h\sim All}[D^{h} = 1] \right| \leq \epsilon(n), \tag{9}$$

where $\epsilon(n)$ is a negligible function (decays faster than any inverse polynomial).

Some functions might look random if you query inputs one by one (classically secure) but reveal their structure immediately if you query a superposition of all inputs (e.g., via the Bernstein-Vazirani algorithm or Simon's algorithm). For our construction, we assume the existence of PRFs that are robust against such quantum attacks. Post-Quantum Cryptography candidates, such as those based on the Learning With Errors (LWE) problem (or, equivalently, lattice problems), are believed to satisfy this. Furthermore, the existence of quantum-secure PRFs can be proven assuming the existence of quantum-secure one-way functions (a function that can be computed efficiently but its inverse cannot be computed efficiently).

5 The Construction: Binary Phase States

We can now define our candidate for Pseudo-Random States. The idea is to encode the output of a PRF into the phases of a quantum superposition.

Definition 3 (Binary Phase State)**.** *Given a boolean function $f : \{0,1\}^n \to \{0,1\}$, the Binary Phase State $|\psi_f\rangle$ is defined as:*

$$|\psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} (-1)^{f(x)} |x\rangle . \tag{10}$$

5.1 Why is this in World A (i.e., efficient to create)?

If f is a PRF, $|\psi_f\rangle$ can be generated by a polynomial-size circuit:

1. Start with $|0\rangle^{\otimes n}$.
2. Apply Hadamard gates $H^{\otimes n}$ to create the uniform superposition $\frac{1}{\sqrt{2^n}} \sum |x\rangle$.
3. Query the quantum PRF oracle once. This applies the phase $(-1)^{f(x)}$ to each term.

Thus, the state is efficiently preparable.

5.2 The Indistinguishability Argument

Our goal is to prove these states are indistinguishable from Haar-random states (World B). The logic proceeds in two steps:

1. **PRF ≈ RF:** By the definition of a Quantum-Secure PRF, states generated by f_k are computationally indistinguishable from states generated by a truly Random Function (RF).
2. **RF ≈ Haar:** We must prove that states generated by a truly Random Function are *statistically* indistinguishable from Haar-random states.

Step 1 is guaranteed by our cryptographic assumption. Step 2 is an information-theoretic fact that we will prove now.

Theorem 2 (Main Result)**.** *The ensemble of Binary Phase States generated by a truly random function (RF) is statistically indistinguishable from the Haar measure. Specifically, the trace distance between their t-th moments is bounded by:*

$$\left\| \mathbb{E}_{f\sim RF}\left[(|\psi_f\rangle\langle\psi_f|)^{\otimes t}\right] - \mathbb{E}_{\psi\sim Haar}\left[(|\psi\rangle\langle\psi|)^{\otimes t}\right] \right\|_1 \leq \frac{6t^2}{2^n}. \tag{11}$$

Combining with the result that quantum-secure PRF exists if quantum-secure one-way function (OWF) exists immediately implies the following theorem.

Theorem 3 (PRS)**.** *PRS exist if quantum-secure OWF exists.*

6 An Elementary Proof (The Purification Method)

Remark: The original proof of the existence of PRS relied on heavy machinery from Representation Theory (Schur-Weyl duality, Weingarten calculus). Below, we present an elementary proof relying only on basic linear algebra and probability.

6.1 Step 1: Purification and the Environment

We wish to analyze the t-th moment of the ensemble of states $|\psi_f\rangle$. It is convenient to analyze the **purified** state. We introduce a hypothetical Environment register E that stores the function f. Let us define the orthonormal basis for E as the truth table of the function:

$$|f\rangle_E \equiv \bigotimes_{z\in\{0,1\}^n} |f(z)\rangle_E .$$

The joint state of the system A (containing t copies of the state) and the environment E is:

$$|\phi\rangle_{AE} = \sum_f \frac{1}{\sqrt{|F|}} |\psi_f\rangle_A^{\otimes t} \otimes |f\rangle_E \tag{12}$$

$$= \sum_f \frac{1}{\sqrt{|F|}} \left(\sum_{x\in\{0,1\}^n} \frac{(-1)^{f(x)}}{\sqrt{2^n}} |x\rangle_A \right)^{\otimes t} \otimes |f\rangle_E \tag{13}$$

$$= \sum_{x_1,\ldots,x_t} \frac{1}{\sqrt{2^{nt}}} |x_1 \ldots x_t\rangle_A \otimes \underbrace{\left(\sum_f \frac{1}{\sqrt{|F|}} (-1)^{\sum_{i=1}^t f(x_i)} |f\rangle_E \right)}_{|\widetilde{\Phi}_{\vec{x}}\rangle_E} . \tag{14}$$

We have expanded $|\psi_f\rangle$ to reveal the environment states $|\widetilde{\Phi}_{\vec{x}}\rangle_E$ associated with inputs $\vec{x} = (x_1, \ldots, x_t)$.

6.2 Step 2: Projection to Distinct Subspace

The geometry of this ensemble is determined by the inner product of the environment states. For any two sequences of inputs $\vec{x}$ and $\vec{x}'$:

$$\langle\widetilde{\Phi}_{\vec{x}'}|\widetilde{\Phi}_{\vec{x}}\rangle = \frac{1}{|F|}\sum_f (-1)^{\sum_i f(x_i)+\sum_i f(x_i')}\langle f|f\rangle \tag{15}$$

$$= \mathbb{E}_f\left[(-1)^{\sum_{z\in\{0,1\}^n} f(z)\cdot c_z}\right], \tag{16}$$

where c_z counts how many times z appears in the multiset $\{\vec{x},\vec{x}'\}$. This expectation is 1 if the exponent is 0 (mod 2) for all f, and 0 otherwise. Suppose we only consider the case where there are **no collisions** in $x_1,\ldots,x_t$, i.e., all x_i in the sequence are distinct. Then we have the following

- If the sets $\{x_i\}_i$ and $\{x_i'\}_i$ are identical, the inner product is 1.
- If the sets are distinct, the inner product is 0.

This means the (unnormalized) state $|\widetilde{\Phi}_{\vec{x}}\rangle$ for distinct $x_1,\ldots,x_t$ is isomorphic to the normalized state $|\{x_1,\ldots,x_t\}\rangle = \frac{1}{\sqrt{t!}}\sum_{\sigma\in S_t}\sigma|x_1,\ldots,x_t\rangle$, where σ is a permutation over the t copies of the n-qubit system. Note that $\frac{1}{\sqrt{t!}}\sum_{\sigma\in S_t}\sigma|x_1,\ldots,x_t\rangle$ is not normalized if $x_1,\ldots,x_t$ are not distinct (we will calculate the norm later). Furthermore, we also have the following fact:

- For any distinct $x_1,\ldots,x_t$ and any non-distinct $x_1',\ldots,x_t'$, the inner product is 0.

Let us define Π_E as the projector onto the subspace of the environment spanned by $|\Phi_{\vec{x}}\rangle$ where $\vec{x}$ contains distinct elements:

$$\Pi_E \equiv \sum_{\vec{x}:\text{distinct}} |\Phi_{\vec{x}}\rangle\langle\Phi_{\vec{x}}|_E.$$

Using the above facts, we have:

$$\Pi_E|\phi\rangle_{AE} = \sum_{x_1,\ldots,x_t:\text{distinct}} \frac{1}{\sqrt{2^{nt}}}|x_1\ldots x_t\rangle_A \otimes |\widetilde{\Phi}_{\vec{x}}\rangle_E. \tag{17}$$

This projection shrinks the state $|\phi\rangle_{AE}$ minimally. To see this, let us calculate the norm of the projected state $\Pi_E|\phi\rangle_{AE}$, which is equal to the probability that t random n-bit strings are distinct.

$$\mathrm{Tr}(\Pi_E|\phi\rangle\langle\phi|_{AE}) = \sum_{\vec{x}:\text{distinct}} \frac{1}{2^{nt}}\langle\Phi_{\vec{x}}|\Phi_{\vec{x}}\rangle \tag{18}$$

$$= \frac{1}{2^{nt}} \times (\text{Number of distinct sequences}) \tag{19}$$

$$= \frac{1}{2^{nt}}\cdot 2^n(2^n-1)\ldots(2^n-t+1) \tag{20}$$

$$= 1\cdot\left(1-\frac{1}{2^n}\right)\ldots\left(1-\frac{t-1}{2^n}\right). \tag{21}$$

Using the bound $(1-a_1)\ldots(1-a_t)\geq 1-\sum a_i$, we have:

$$\|\Pi_E|\phi\rangle_{AE}\|^2 \geq 1-\sum_{k=0}^{t-1}\frac{k}{2^n} \geq 1-\frac{t^2}{2^n}. \tag{22}$$

We can see that for any t polynomial in n, the length is exponentially close to 1.

6.3 Step 3: The Invariant State $|\tilde{\eta}\rangle$

We now introduce an (unnormalized) reference state $|\tilde{\eta}\rangle_{AE}$ constructed to be perfectly symmetric under permutations of the computational basis. We define it as:

$$|\tilde{\eta}\rangle_{AE} = \sum_{\vec{x}} \frac{1}{\sqrt{2^{nt}}} |\vec{x}\rangle_A \otimes \left(\frac{1}{\sqrt{t!}} \sum_{\sigma \in S_t} \sigma |x_1, \dots, x_t\rangle_E \right). \tag{23}$$

This state serves as a "Bridge" between the structured World A and the random World B.

Lemma 1 (The Bridge)**.** *In the collision-free (distinct) subspace, the random function state and the invariant state are identical:*

$$\Pi_E |\phi\rangle_{AE} = \Pi_E |\tilde{\eta}\rangle_{AE}.$$

6.4 Step 4: The Norm of the Invariant State

A crucial detail is that the state $|\tilde{\eta}\rangle_{AE}$ is not normalized to 1. To use it in our distance bounds, we must calculate its norm explicitly.

$$\langle\tilde{\eta}|\tilde{\eta}\rangle = \sum_{\vec{x}\in\{0,1\}^{nt}} \frac{1}{2^{nt}} \left\| \frac{1}{\sqrt{t!}} \sum_{\sigma \in S_t} \sigma |x_1 \dots x_t\rangle_E \right\|^2. \tag{24}$$

We calculate the norm of the term inside the summation for a fixed string $\vec{x}$. Let $|\psi_{\vec{x}}\rangle = \frac{1}{\sqrt{t!}} \sum_{\sigma \in S_t} |x_{\sigma(1)} \dots x_{\sigma(t)}\rangle_E$. The squared norm is the inner product of the state with itself:

$$\langle\psi_{\vec{x}}|\psi_{\vec{x}}\rangle = \frac{1}{t!} \sum_{\sigma \in S_t} \sum_{\tau \in S_t} \langle x_{\sigma(1)} \dots x_{\sigma(t)} | x_{\tau(1)} \dots x_{\tau(t)}\rangle. \tag{25}$$

The inner product $\langle x_{\sigma(\cdot)} | x_{\tau(\cdot)}\rangle$ is 1 if and only if the permuted strings are identical (i.e., $x_{\sigma(i)} = x_{\tau(i)}$ for all i), and 0 otherwise.

Step 4a: Counting the Permutations for a Single String Let us fix a specific string of inputs $\vec{x} = (x_1, \dots, x_t)$. We need to calculate the norm of the unnormalized environment state associated with this string:

$$|\eta_{\vec{x}}\rangle = \frac{1}{\sqrt{t!}} \sum_{\sigma \in S_t} |x_{\sigma(1)} \dots x_{\sigma(t)}\rangle_E.$$

The squared norm is the inner product of this state with itself:

$$\langle\eta_{\vec{x}}|\eta_{\vec{x}}\rangle = \frac{1}{t!} \sum_{\sigma \in S_t} \sum_{\tau \in S_t} \langle x_{\sigma(1)} \dots x_{\sigma(t)} | x_{\tau(1)} \dots x_{\tau(t)}\rangle. \tag{26}$$

The inner product inside the sum is equal to 1 if the two permuted sequences are identical, and 0 otherwise. To count how many times this happens, let us group the elements of $\vec{x}$ by value. Let $N = 2^n$. Let $z_1, \dots, z_N$ be the set of all possible n-bit strings. Let k_j be the number of times the string z_j appears in the sequence $\vec{x}$. Note that $\sum_{j=1}^{N} k_j = t$.

Fix a specific permutation σ. We ask: for how many permutations τ is the sequence $x_{\tau(\cdot)}$ identical to $x_{\sigma(\cdot)}$? Since the positions of identical values are interchangeable, any permutation τ that only swaps positions containing the same value will result in the same sequence. The number of ways to permute the k_1 positions containing z_1 is $k_1!$. Similarly, we have $k_2!$ ways for z_2, and so

on. Thus, for every fixed σ, there are exactly $k_1!k_2!\dots k_N!$ matching τ's. Since there are $t!$ choices for σ, the total sum is:

$$\sum_{\sigma\in S_t}\sum_{\tau\in S_t}\langle\dots\rangle = t!\times(k_1!k_2!\dots k_N!). \tag{27}$$

Substituting this back into the norm equation (and cancelling the $1/t!$ prefactor):

$$\||\eta_{\vec{x}}\rangle\|^2 = \prod_{j=1}^{N} k_j!. \tag{28}$$

Step 4b: Summing over all Strings We now calculate the total norm of $|\tilde{\eta}\rangle_{AE}$ by summing over all possible strings $\vec{x}$.

$$\langle\tilde{\eta}|\tilde{\eta}\rangle = \frac{1}{2^{nt}}\sum_{\vec{x}\in\{0,1\}^{nt}}\||\eta_{\vec{x}}\rangle\|^2. \tag{29}$$

Instead of summing over the N^t individual strings, we sum over the possible counts $\{k_1,\dots,k_N\}$. How many distinct strings $\vec{x}$ result in the same set of counts $\{k_j\}$? This is a standard combinatorial result given by the **Multinomial Coefficient**:

$$\text{Number of strings with counts } \{k_j\} = \binom{t}{k_1,\dots,k_N} = \frac{t!}{k_1!k_2!\dots k_N!}.$$

We can now rewrite the summation by grouping terms with the same counts:

$$\langle\tilde{\eta}|\tilde{\eta}\rangle = \frac{1}{N^t}\sum_{k_1+\dots+k_N=t}(\#\text{ of strings})\times(\text{Norm of each string}) \tag{30}$$

$$= \frac{1}{N^t}\sum_{k_1+\dots+k_N=t}\left(\frac{t!}{\prod_{j=1}^{N}k_j!}\right)\times\left(\prod_{j=1}^{N}k_j!\right). \tag{31}$$

Notice that the combinatorial factors involving $k_j!$ cancel perfectly:

$$\langle\tilde{\eta}|\tilde{\eta}\rangle = \frac{t!}{N^t}\sum_{k_1+\dots+k_N=t}1. \tag{32}$$

Step 4c: Stars and Bars Argument The term $\sum_{k_1+\dots+k_N=t}1$ counts the number of non-negative integer solutions to $k_1+\dots+k_N=t$. Imagine we have t indistinguishable items ("stars") that we want to distribute into N distinct bins ("bars"). The number of ways to do this is given by the binomial coefficient:

$$\sum_{k_1+\dots+k_N=t}1 = \binom{N+t-1}{t}.$$

Substituting this back:

$$\langle\tilde{\eta}|\tilde{\eta}\rangle = \frac{t!}{N^t}\binom{N+t-1}{t} \tag{33}$$

$$= \frac{t!}{N^t}\frac{(N+t-1)!}{t!(N-1)!} \tag{34}$$

$$= \frac{1}{N^t}\frac{(N+t-1)(N+t-2)\dots(N)\cdot(N-1)!}{(N-1)!} \tag{35}$$

$$= \frac{1}{N^t}\prod_{j=0}^{t-1}(N+j) \tag{36}$$

$$= \left(\frac{N}{N}\right)\cdot\left(\frac{N+1}{N}\right)\cdot\left(\frac{N+2}{N}\right)\cdots\left(\frac{N+t-1}{N}\right) \tag{37}$$

$$= 1\cdot\left(1+\frac{1}{N}\right)\left(1+\frac{2}{N}\right)\dots\left(1+\frac{t-1}{N}\right). \tag{38}$$

To bound this value, we use the standard inequality $1+x \le e^x$ for all real x.

$$\langle\tilde{\eta}|\tilde{\eta}\rangle = \prod_{j=0}^{t-1}\left(1+\frac{j}{N}\right) \tag{39}$$

$$\le \prod_{j=0}^{t-1}\exp\left(\frac{j}{N}\right) \tag{40}$$

$$= \exp\left(\frac{1}{N}\sum_{j=0}^{t-1} j\right). \tag{41}$$

The sum of integers from 0 to $t-1$ is $\frac{t(t-1)}{2}$. Thus:

$$1 \le \langle\tilde{\eta}|\tilde{\eta}\rangle \le \exp\left(\frac{t(t-1)}{2N}\right) < \exp\left(\frac{t^2}{2\cdot 2^n}\right) \le 1+\frac{t^2}{2^n}, \tag{42}$$

whenever $t^2 \le 2^{n+1}$. This provides a strict upper bound on the norm of the invariant state.

6.5 Step 5: Unitary Invariance

Why did we construct $|\tilde{\eta}\rangle$? Because it possesses a powerful property: **Unitary Invariance**. This is what connects it to the Haar measure.

Lemma 2 (Unitary Invariance)**.** $\mathrm{Tr}_E(U_A^{\otimes t}|\tilde{\eta}\rangle\langle\tilde{\eta}|_{AE}U_A^{\dagger,\otimes t}) = \mathrm{Tr}_E(|\tilde{\eta}\rangle\langle\tilde{\eta}|_{AE})$.

Proof. We use the following linear algebraic identity. For any unitary U and the unnormalized maximally entangled state $|\Omega\rangle = \sum_x |x\rangle|x\rangle$:

$$\sum_x U|x\rangle\otimes|x\rangle = \sum_{x,y} U_{yx}|y\rangle\otimes|x\rangle = \sum_{x,y}|y\rangle\otimes U_{yx}|x\rangle = \sum_y |y\rangle\otimes U^T|y\rangle. \tag{43}$$

Hence, we have

$$U_A^{\otimes t} |\tilde{\eta}\rangle_{AE} = \sum_{\vec{x}} \frac{1}{\sqrt{2^{nt}}} U_A^{\otimes t} |x_1, \ldots, x_t\rangle_A \otimes \left(\frac{1}{\sqrt{t!}} \sum_{\sigma \in S_t} \sigma |x_1, \ldots, x_t\rangle_E \right) \tag{44}$$

$$= \sum_{\vec{x}} \frac{1}{\sqrt{2^{nt}}} |x_1, \ldots, x_t\rangle_A \otimes \left(\frac{1}{\sqrt{t!}} \sum_{\sigma \in S_t} \sigma U_E^{T,\otimes t} |x_1, \ldots, x_t\rangle_E \right) \tag{45}$$

$$= \sum_{\vec{x}} \frac{1}{\sqrt{2^{nt}}} |x_1, \ldots, x_t\rangle_A \otimes \left(\frac{1}{\sqrt{t!}} \sum_{\sigma \in S_t} U_E^{T,\otimes t} \sigma |x_1, \ldots, x_t\rangle_E \right) = U_E^{T,\otimes t} |\tilde{\eta}\rangle_{AE} . \tag{46}$$

This completes the proof. □

6.6 Step 6: The Gentle Measurement Lemma (Purified Version)

We need to bound the distance between the unprojected states and the projected states. We use the following lemma, which allows us to convert the trace distance norm into a simple trace calculation.

Lemma 3 (Gentle Measurement for Purifications)**.** *Let ρ_{AE} be a (possibly unnormalized) density matrix, i.e., Hermitian PSD matrix. Let Π_E be a projector on the environment which is diagonal in the basis used for the partial trace. Then:*

$$\|\mathrm{Tr}_E(\rho_{AE}) - \mathrm{Tr}_E(\Pi_E \rho_{AE} \Pi_E)\|_1 = \mathrm{Tr}(\rho_{AE}) - \mathrm{Tr}(\Pi_E \rho_{AE}).$$

Proof. By the linearity of the partial trace, we can expand the term inside the norm:

$$\mathrm{Tr}_E(\rho_{AE}) = \mathrm{Tr}_E(\Pi_E \rho_{AE}) + \mathrm{Tr}_E((I - \Pi_E)\rho_{AE}) \tag{47}$$

$$= \mathrm{Tr}_E(\Pi_E \rho_{AE} \Pi_E) + \mathrm{Tr}_E((I - \Pi_E)\rho_{AE}(I - \Pi_E)). \tag{48}$$

Substituting this into the LHS of the lemma statement:

$$\text{LHS} = \|(\mathrm{Tr}_E(\Pi_E \rho_{AE} \Pi_E) + \mathrm{Tr}_E((I - \Pi_E)\rho_{AE}(I - \Pi_E))) - \mathrm{Tr}_E(\Pi_E \rho_{AE} \Pi_E)\|_1 \tag{49}$$

$$= \|\mathrm{Tr}_E((I - \Pi_E)\rho_{AE}(I - \Pi_E))\|_1 . \tag{50}$$

Let $M = (I - \Pi_E)\rho_{AE}(I - \Pi_E)$. Since ρ_{AE} is a positive operator (it is a state), M is also a positive operator. The partial trace of a positive operator is also positive. For any positive operator A, the trace norm is simply the trace: $\|A\|_1 = \mathrm{Tr}(A)$.

$$\text{LHS} = \mathrm{Tr}\left(\mathrm{Tr}_E((I - \Pi_E)\rho_{AE}(I - \Pi_E))\right) \tag{51}$$

$$= \mathrm{Tr}\left((I - \Pi_E)\rho_{AE}(I - \Pi_E)\right) \quad \text{(Trace of partial trace is total trace)} \tag{52}$$

$$= \mathrm{Tr}\left((I - \Pi_E)^2 \rho_{AE}\right) \quad \text{(Cyclicity of trace)} \tag{53}$$

$$= \mathrm{Tr}\left((I - \Pi_E)\rho_{AE}\right) \quad \text{(Projector property } P^2 = P\text{)} \tag{54}$$

$$= \mathrm{Tr}(\rho_{AE}) - \mathrm{Tr}(\Pi_E \rho_{AE}). \tag{55}$$

This completes the proof. □

6.7 Step 7: The Triangle Inequality Roadmap

We now have all the ingredients to rigorously bound the trace distance between the Random Function ensemble and the Haar ensemble. Let $\rho_{\text{RF}} = \text{Tr}_E(|\phi\rangle\langle\phi|_{AE})$ be the t-th moment of the Random Function ensemble. Let ρ_{Haar} be the t-th moment of the Haar ensemble. Let $\tilde{\sigma} = \text{Tr}_E(|\tilde{\eta}\rangle\langle\tilde{\eta}|_{AE})$ be the reduced density matrix of our invariant state. We use the triangle inequality to bound the distance via the intermediate state $\tilde{\sigma}$.

$$\|\rho_{\text{RF}} - \rho_{\text{Haar}}\|_1 \leq \|\rho_{\text{RF}} - \tilde{\sigma}\|_1 + \|\tilde{\sigma} - \rho_{\text{Haar}}\|_1 .$$

Bound 1: Distance between RF and Invariant State We apply the Gentle Measurement Lemma (Lemma 3) to both ρ_{RF} and $\tilde{\sigma}$, using the fact that they share the same projection onto the collision-free subspace (Lemma 1). First, for the Random Function state $|\phi\rangle_{AE}$:

$$\begin{aligned}
\|\rho_{\text{RF}} - \text{Tr}_E(\Pi_E|\phi\rangle\langle\phi|\Pi_E)\|_1 &= \text{Tr}(|\phi\rangle\langle\phi|) - \text{Tr}(\Pi_E|\phi\rangle\langle\phi|) && (56)\\
&= 1 - \|\Pi_E|\phi\rangle\|^2 && (57)\\
&\leq \frac{t^2}{2^n}. \quad \text{(From Step 2)} && (58)
\end{aligned}$$

Second, for the Invariant state $|\tilde{\eta}\rangle_{AE}$:

$$\begin{aligned}
\|\tilde{\sigma} - \text{Tr}_E(\Pi_E|\tilde{\eta}\rangle\langle\tilde{\eta}|\Pi_E)\|_1 &= \text{Tr}(|\tilde{\eta}\rangle\langle\tilde{\eta}|) - \text{Tr}(\Pi_E|\tilde{\eta}\rangle\langle\tilde{\eta}|) && (59)\\
&= \langle\tilde{\eta}|\tilde{\eta}\rangle - \|\Pi_E|\tilde{\eta}\rangle\|^2 . && (60)
\end{aligned}$$

Recall from Lemma 1 that $\Pi_E|\tilde{\eta}\rangle = \Pi_E|\phi\rangle$. Thus $\|\Pi_E|\tilde{\eta}\rangle\|^2 = \|\Pi_E|\phi\rangle\|^2 \geq 1 - t^2/2^n$. Also, from Step 4, we know $\langle\tilde{\eta}|\tilde{\eta}\rangle \leq 1 + t^2/2^n$. Thus:

$$\|\tilde{\sigma} - \text{Tr}_E(\Pi_E|\tilde{\eta}\rangle\langle\tilde{\eta}|\Pi_E)\|_1 \leq \left(1 + \frac{t^2}{2^n}\right) - \left(1 - \frac{t^2}{2^n}\right) = \frac{2t^2}{2^n}. \quad (61)$$

Combining these using the triangle inequality (and noting the projected states are identical):

$$\begin{aligned}
\|\rho_{\text{RF}} - \tilde{\sigma}\|_1 &\leq \|\rho_{\text{RF}} - \text{Tr}_E(\Pi|\phi\rangle\langle\phi|\Pi)\|_1 + \|\text{Tr}_E(\Pi|\tilde{\eta}\rangle\langle\tilde{\eta}|\Pi) - \tilde{\sigma}\|_1 && (62)\\
&\leq \frac{t^2}{2^n} + \frac{2t^2}{2^n} = \frac{3t^2}{2^n}. && (63)
\end{aligned}$$

Bound 2: Distance between Invariant State and Haar We know that ρ_{Haar} is the average of the Random Function ensemble over all unitary rotations:

$$\rho_{\text{Haar}} = \mathbb{E}_{U\sim\text{Haar}}\left[U^{\otimes t}\rho_{\text{RF}}(U^\dagger)^{\otimes t}\right].$$

We also established in Step 5 (Lemma 2) that $\tilde{\sigma}$ is unitary invariant:

$$\tilde{\sigma} = \mathbb{E}_{U\sim\text{Haar}}\left[U^{\otimes t}\tilde{\sigma}(U^\dagger)^{\otimes t}\right].$$

Using the convexity of the trace norm (or simply averaging the distance):

$$\begin{aligned}
\|\rho_{\text{Haar}} - \tilde{\sigma}\|_1 &= \left\|\mathbb{E}_U[U^{\otimes t}\rho_{\text{RF}}(U^\dagger)^{\otimes t}] - \mathbb{E}_U[U^{\otimes t}\tilde{\sigma}(U^\dagger)^{\otimes t}]\right\|_1 && (64)\\
&= \left\|\mathbb{E}_U\left[U^{\otimes t}(\rho_{\text{RF}} - \tilde{\sigma})(U^\dagger)^{\otimes t}\right]\right\|_1 && (65)\\
&\leq \mathbb{E}_U\left\|U^{\otimes t}(\rho_{\text{RF}} - \tilde{\sigma})(U^\dagger)^{\otimes t}\right\|_1 && (66)\\
&= \|\rho_{\text{RF}} - \tilde{\sigma}\|_1 . && (67)
\end{aligned}$$

Since the trace norm is unitarily invariant, the distance remains unchanged inside the expectation. Thus, this distance is also bounded by $\frac{3t^2}{2^n}$.

Final Result Adding the two bounds together:

$$\|\rho_{\mathrm{RF}} - \rho_{\mathrm{Haar}}\|_1 \leq \frac{3t^2}{2^n} + \frac{3t^2}{2^n} = \frac{6t^2}{2^n}. \tag{68}$$

This completes the proof. Since $6t^2/2^n$ is negligible for any polynomial t, the Random Function states are statistically indistinguishable from Haar-random states. Consequently, replacing the random function with a quantum-secure PRF yields a Pseudo-Random State ensemble.