

# Pseudorandomness from Subset States: Representation-Theoretic and Spectral Views

Lenny Liu   Tzu-Hsiang Huang   Yueying Wu

Representation Methods in QIP

December 11, 2025

# Recall: Haar measure unitary

(i) Normalization:  $\int_{\mathcal{U}_2} dU = 1$ .

(ii) Left- and right-invariance: For any function  $f$  and an arbitrary unitary  $V$ , we have

$$\int_{\mathcal{U}_2} dU f(VU) = \int_{\mathcal{U}_2} dU f(UV) = \int_{\mathcal{U}_2} dU f(U).$$

Haar random state  $|\psi\rangle \leftarrow \mu$ : A uniform distribution over pure quantum states

$\mu$  can be defined as  $U|0\rangle$ , where  $U$  is a Haar measure unitary.

# Pseudorandom states (PRS)

A function  $G : k \mapsto |\phi_k\rangle$  is a pseudorandom states if

1.  $G$  is efficiently computable
2. Pseudorandomness



$$b \leftarrow \{0,1\}, \quad k \leftarrow \{0,1\}^\lambda$$

if  $b = 0$  :

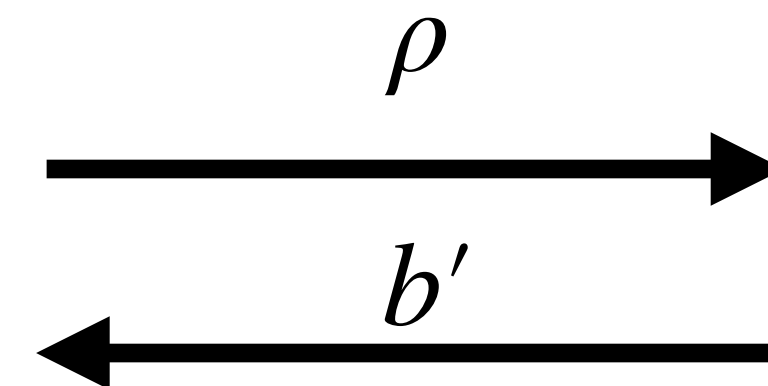
$$\rho := |\phi_k\rangle^{\otimes poly(\lambda)}$$

if  $b = 1$  :

$$|\psi\rangle \leftarrow \mu$$

$$\rho := |\psi\rangle^{\otimes poly(\lambda)}$$

$$Pr[b = b'] \approx \frac{1}{2}$$



# Pseudorandom states (PRS)

A function  $G : k \mapsto |\phi_k\rangle$  is a pseudorandom states if

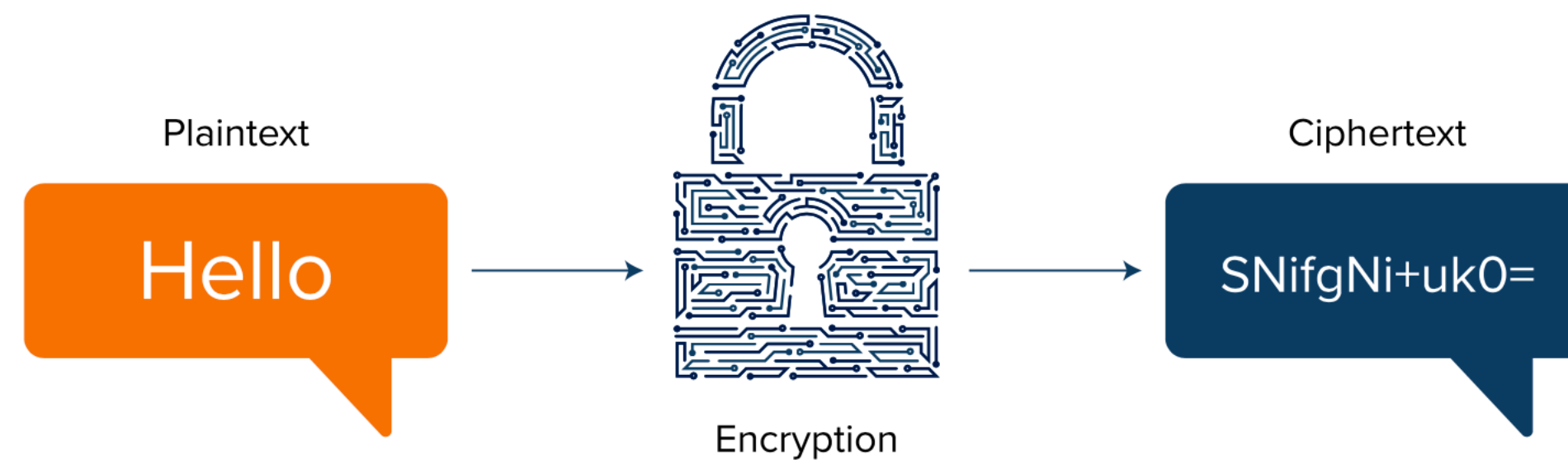
1.  $G$  is efficiently computable
2. Pseudorandomness

$\frac{1}{2} \left\| |\phi\rangle^{\otimes poly} - |\psi\rangle^{\otimes poly} \right\|_1$  is negligible, where  $|\psi\rangle \leftarrow \mu$  is Haar random

# Motivation: Step back to the classical setting

Many cryptography protocol requires assumptions:

Encryption



Commitment



Multiparty Computation

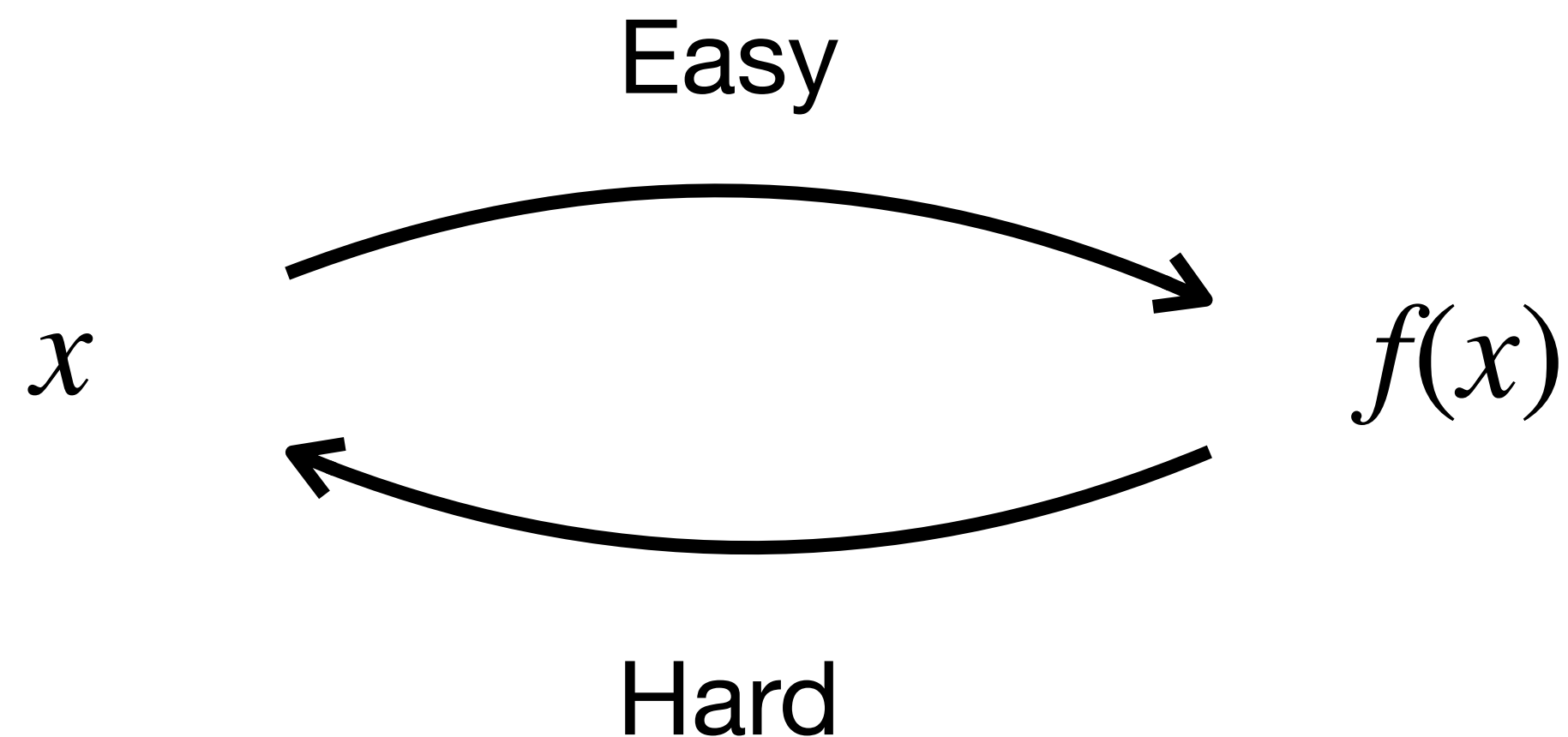


# Motivation: Step back to the classical setting

Many cryptography protocol requires assumptions:

What kind of assumption is needed?

E.g. Prime factoring is hard, One way function (OWF) exists



It turns out that OWF is a minimum assumption that makes crypto exists.

# Motivation: Back to quantum

If we allow quantum computation / communication:

One-way function (OWF) is no longer a minimal assumption.

Instead PRS is **believed** to be a weaker assumption than OWF.

Also, PRS are sufficient to construct some crypto protocol.

E.g. Symmetric key encryption, commitment, multiparty computation...

# A candidate construction

$$\frac{1}{\sqrt{|S|}} \sum_{x \in S} |x\rangle,$$

where  $S$  is a random subset over  $\{0,1\}^n$  with size  $\omega(\text{poly}(n)) \sim 2^n / \omega(\text{poly}(n))$

1. Why the size cannot be too small or too large?
2. This construction also gives us a pseudo entanglement.

# Pseudoentanglement

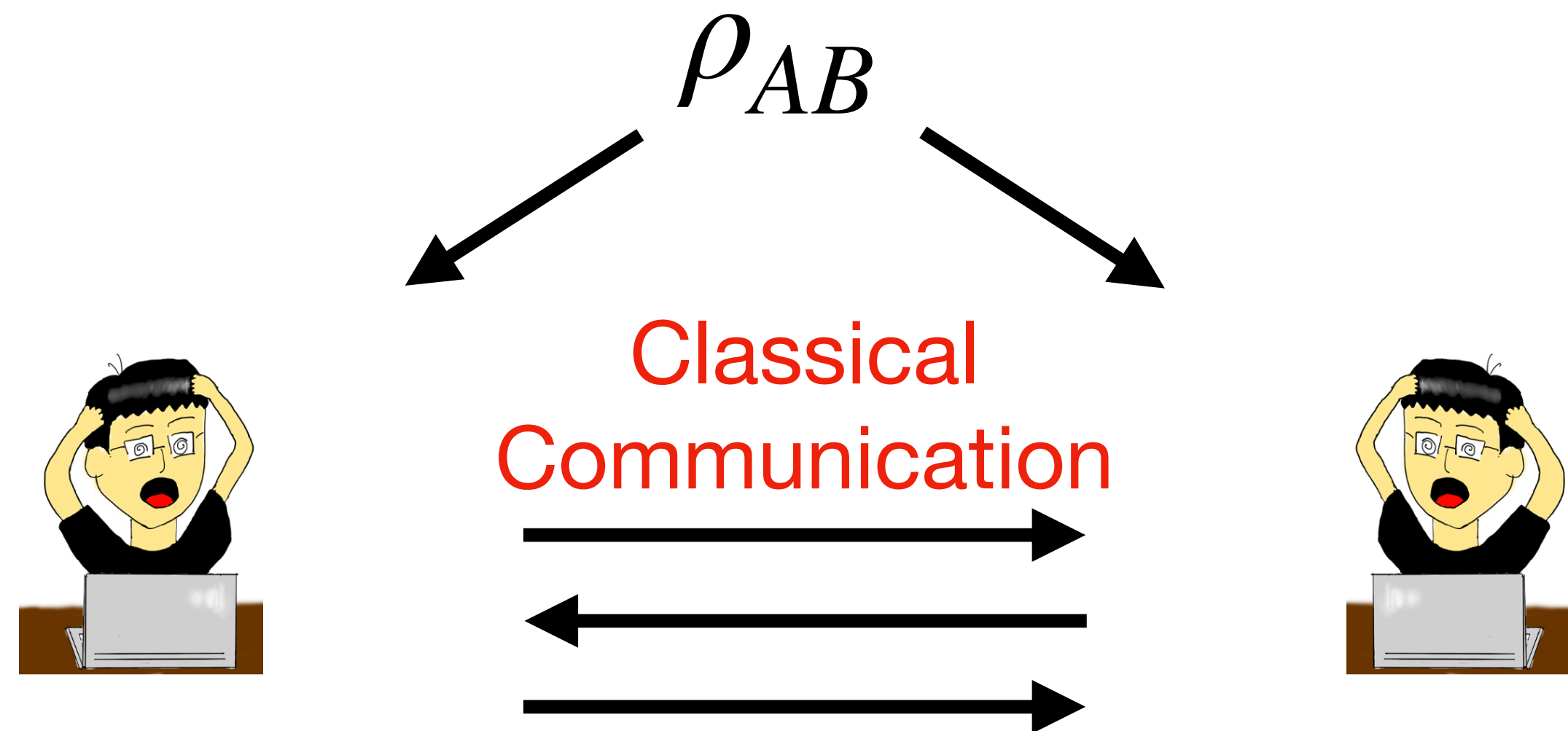
A function  $G : k \mapsto |\phi_k\rangle$  is a pseudoentanglement states if

1.  $G$  is efficiently computable
2. Low entanglement entropy across all cut
3. Indistinguishable from a high entanglement state, e.g., Haar random state

$$\frac{1}{\sqrt{|S|}} \sum_{x \in S} |x\rangle \quad \text{has at most } \log(|S|) \text{ entanglement entropy}$$

# Application of Pseudoentanglement

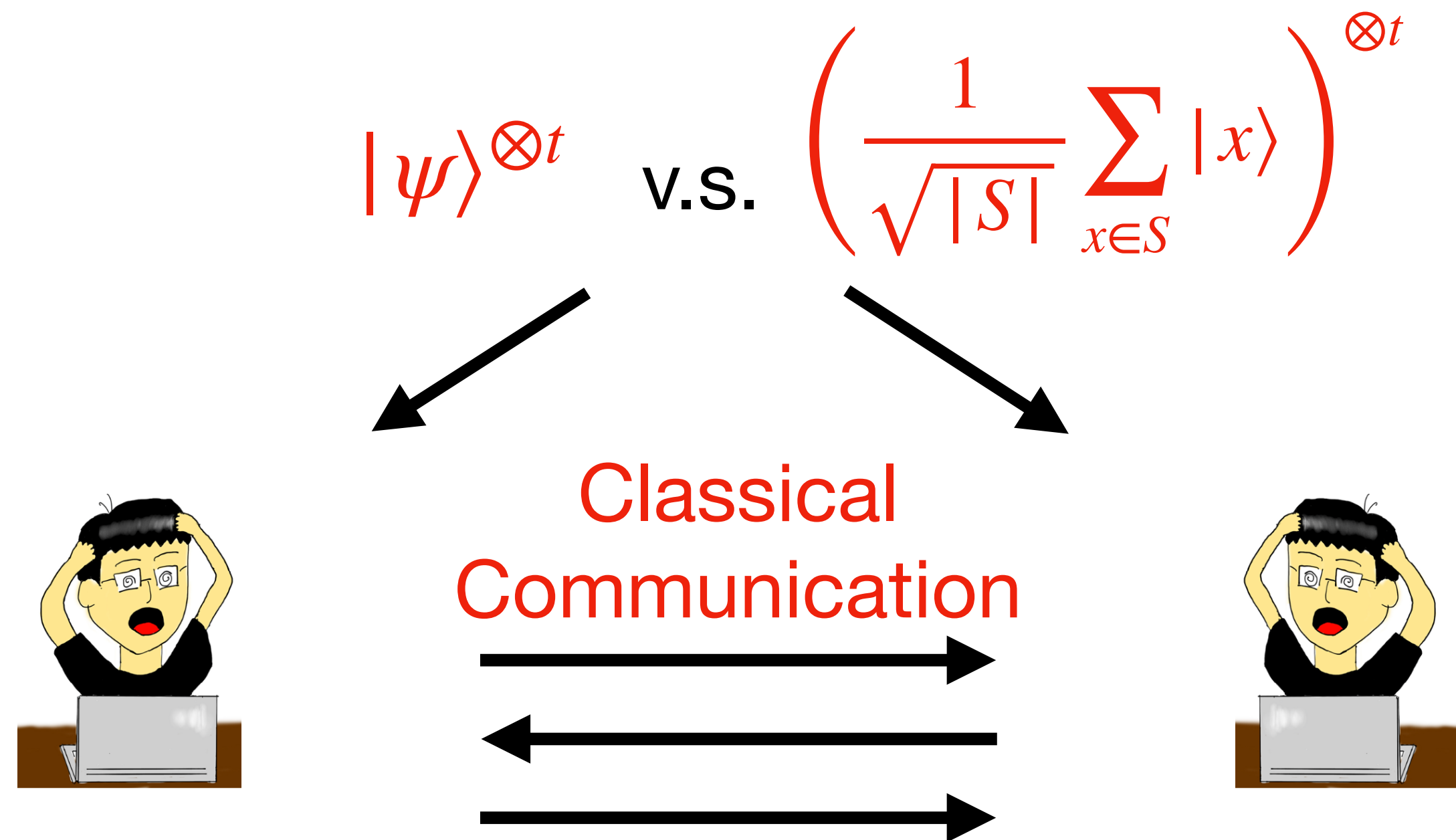
**Proposition 3.2:** There is non efficient distillation protocol that produced  $\log S(\rho)$  EPR pairs



Goal: Output EPR pairs as many as they could

# Application of Pseudoentanglement

**Proposition 3.2:** There is non efficient distillation protocol that produced  $\log S(\rho)$  EPR pairs



Goal: Output EPR pairs as many as they could

## Core information-theoretic statement

Let  $\Phi$  be the  $k$ -copy mixture of random subset states of size  $s$ , and  $\Psi$  the  $k$ -copy Haar moment. Then

$$\|\Psi - \Phi\|_1 \leq O\left(\frac{k^2}{d} + \frac{k}{\sqrt{s}} + \frac{sk}{d}\right).$$

**Method:** collision-free reduction + spectra of generalized Johnson graphs.

# Background: subset states and Haar moments

## Subset states.

- Work in  $\mathbb{C}^d$  with computational basis  $\{|i\rangle\}_{i \in [d]}$ .
- For  $S \subseteq [d]$  with  $|S| = s$ , define

$$|S\rangle := \frac{1}{\sqrt{s}} \sum_{i \in S} |i\rangle, \quad \phi_S := |S\rangle\langle S|.$$

## $k$ -copy ensembles.

$$\Phi := \mathbb{E}_{|S|=s} \phi_S^{\otimes k}, \quad \Psi := \int |\psi\rangle\langle\psi|^{\otimes k} d\mu(\psi)$$

where  $\mu$  is the Haar measure on pure states.

**Haar  $k$ -moment.** Representation theory gives

$$\Psi = \frac{1}{\binom{d+k-1}{k}} \Pi_{\text{sym}},$$

where  $\Pi_{\text{sym}}$  projects onto the symmetric subspace of  $(\mathbb{C}^d)^{\otimes k}$ .

## Background: trace distance and distinguishability

For density matrices  $\rho, \sigma$ ,

$$\|\rho - \sigma\|_1 := \text{Tr} \sqrt{(\rho - \sigma)^\dagger (\rho - \sigma)}.$$

### Operational meaning (Holevo-Helstrom)

The optimal success probability when discriminating between  $\rho$  and  $\sigma$  is

$$\frac{1}{2} + \frac{1}{2} \|\rho - \sigma\|_1.$$

So bounding  $\|\Psi - \Phi\|_1$  says: even an all-powerful adversary with  $k$  copies cannot distinguish subset states from Haar random by more than that amount.

# Sketch of the proof strategy

- 1 Project both  $\Phi$  and  $\Psi$  onto the *no-collision* subspace in the computational basis.
- 2 Show that outside this subspace the trace norm contribution is small (birthday-type bounds).
- 3 Inside the no-collision subspace, exploit symmetry: entries depend only on intersection size  $\Rightarrow$  we can express the difference using Johnson graphs.
- 4 Use known spectra of Johnson graphs to bound the trace norm.

The rest of the slides unpack each of these steps.

## Step A: project to the collision-free subspace

Define

$$A([d], k) = \{(i_1, \dots, i_k) \in [d]^k : i_a \neq i_b \forall a \neq b\},$$

and let  $\Pi$  project onto

$$\mathcal{H}_{\text{nocoll}} = \text{span}\{|i_1, \dots, i_k\rangle : (i_1, \dots, i_k) \in A([d], k)\}.$$

**Define**

$$\tilde{\Psi} := \Pi\Psi\Pi, \quad \tilde{\Phi} := \Pi\Phi\Pi \text{ (with a convenient renormalization).}$$

- Haar:  $\|\Psi - \tilde{\Psi}\|_1 = O(k^2/d)$  (birthday bound in  $[d]$ ).
- Subset:  $\|\Phi - \tilde{\Phi}\|_1 = O(k/\sqrt{s})$  (collisions inside a size- $s$  subset).

So it remains to bound  $\|\tilde{\Phi} - \tilde{\Psi}\|_1$  on  $\mathcal{H}_{\text{nocoll}}$ .

## Step B: dependence only on union size

For no-collision tuples  $\mathbf{i}, \mathbf{j} \in A([d], k)$ , let

$$\ell(\mathbf{i}, \mathbf{j}) := |\{i_1, \dots, i_k, j_1, \dots, j_k\}|$$

be the number of distinct indices used.

Then the projected subset moment has entries

$$\tilde{\Phi}(\mathbf{i}, \mathbf{j}) = \frac{1}{s^k} \Pr_{|S|=s} [\mathbf{i}, \mathbf{j} \subseteq S] = \frac{s_\ell}{s^k d_\ell},$$

where

$$s_\ell = s(s-1)\cdots(s-\ell+1), \quad d_\ell = d(d-1)\cdots(d-\ell+1).$$

### Consequence

$\tilde{\Phi}$  depends only on  $\ell(\mathbf{i}, \mathbf{j})$  (equivalently, only on  $|A \cap B|$  after grouping). So on the no-collision subspace,  $\tilde{\Phi}$  lies in the Johnson association scheme.

## Step C: from tuples to $k$ -subsets and Johnson graphs

Each no-collision tuple  $\mathbf{i}$  corresponds to:

- a  $k$ -subset  $A \subseteq [d]$  (its support), and
- a permutation in  $S_k$  specifying the ordering.

So

$$A([d], k) \simeq \binom{[d]}{k} \times S_k.$$

After factoring out the  $S_k$  part:

- The difference operator on  $\mathcal{H}_{\text{nocoll}}$  has the form

$$D = \tilde{D} \otimes J_{k!},$$

where  $J_{k!}$  is the all-ones matrix on permutations.

- The nontrivial part  $\tilde{D}$  acts on  $k$ -subsets  $A, B \in \binom{[d]}{k}$ .

we can write Johnson adjacency matrices

$$\tilde{D} = \sum_{t=0}^{k-1} \alpha_t D_t,$$

## Step D: spectral bound via Johnson graphs

- The matrices  $D_t$  generate the Bose–Mesner algebra of the Johnson scheme.
- Their eigenvalues and multiplicities are explicitly known.
- Schatten norm inequalities give

$$\|\tilde{D}\|_1 \leq \sum_{t=0}^{k-1} |\alpha_t| \|D_t\|_1 = O\left(\frac{sk}{d}\right) \quad \text{for } k \ll \sqrt{d}.$$

Combining with the small rescaling difference between  $\tilde{\Psi}$  and the exact Haar moment yields

$$\|\tilde{\Psi} - \tilde{\Phi}\|_1 = O\left(\frac{sk}{d} + \frac{k^2}{d}\right).$$

# Assembling the bounds & tightness

Triangle inequality:

$$\|\Psi - \Phi\|_1 \leq \|\Psi - \tilde{\Psi}\|_1 + \|\tilde{\Psi} - \tilde{\Phi}\|_1 + \|\tilde{\Phi} - \Phi\|_1.$$

- $\|\Psi - \tilde{\Psi}\|_1 = O(k^2/d)$  (Haar collisions).
- $\|\tilde{\Phi} - \Phi\|_1 = O(k/\sqrt{s})$  (subset collisions / renormalization).
- $\|\tilde{\Psi} - \tilde{\Phi}\|_1 = O(sk/d)$  (Johnson-graph spectral bound).

$$\Rightarrow \|\Psi - \Phi\|_1 \leq O\left(\frac{k^2}{d} + \frac{k}{\sqrt{s}} + \frac{sk}{d}\right).$$

**Tightness intuition:**

- If  $s$  is too small, collisions on  $\approx s + 1$  copies distinguish.
- If  $s$  is too large, swap tests against  $|+^n\rangle$  distinguish with gap  $\approx s/d$ .

# Why a “Goldilocks” subset size is unavoidable (tightness intuition)

Two simple distinguishers force constraints on  $m$ :

- 1 **Birthday/collision test:** measure each copy in computational basis.

$$\Pr[\text{collision}] \approx \Theta(t^2/m) \Rightarrow m \gg t^2 \text{ needed.}$$

- 2 **Overlap with uniform superposition:**  $|+\rangle^n = \frac{1}{\sqrt{N}} \sum_x |x\rangle$ .

$$|\langle +^n | S \rangle|^2 = \frac{m}{N} \Rightarrow m/N \text{ must be negligible.}$$

So we need:  $\omega(\text{poly}(n)) < m < o(2^n)$  (roughly).

- Work in  $\text{Sym}^t([N])$  (support of both Haar and subset moments).
- Find a **typical subspace** capturing almost all of  $\text{Sym}^t([N])$ .
- Exploit **basis-label permutation symmetry** ( $S_N$  acting on  $[N]$ ).
- Identify the relevant homogeneous space as the **Johnson scheme**

$$X = S_N / (S_t \times S_{N-t}),$$

which is a **Gelfand pair**  $\Rightarrow$  multiplicity-free decomposition.

- Invariant matrices diagonalize on  $S_N$ -irrep blocks; analyze the *largest block*.

## Step 0: a “type basis” for the symmetric subspace

A basis of  $\text{Sym}^t([N])$  can be indexed by *types* (multisets)  $\theta$  of size  $t$  from  $[N]$ .

$$\dim \text{Sym}^t([N]) = \binom{N+t-1}{t}.$$

### Unique types

A type is *unique* if no element repeats (i.e., it is a  $t$ -subset). Let  $\Sigma_{\text{unique}} \subseteq \text{Sym}^t([N])$  be the span of unique types. Then

$$\dim \Sigma_{\text{unique}} = \binom{N}{t} \quad \text{and} \quad \frac{\binom{N}{t}}{\binom{N+t-1}{t}} = 1 - O\left(\frac{t^2}{N}\right).$$

Interpretation: when  $t \ll \sqrt{N}$ , almost all of  $\text{Sym}^t$  consists of “no-collision” configurations.

## Step 1: restrict the moment operator to $\Sigma_{\text{unique}}$

Let

$$\rho := \mathbb{E}_{S:|S|=m} |S\rangle\langle S|^{\otimes t}.$$

Index an orthonormal basis of  $\Sigma_{\text{unique}}$  by  $t$ -subsets  $\alpha, \beta \in \binom{[M]}{t}$ .

### Concrete matrix entries

One computes

$$\langle \alpha | \rho | \beta \rangle = \frac{t!}{m^t} \Pr_{S:|S|=m} [\alpha \cup \beta \subseteq S] = \frac{t!}{m^t} \cdot \frac{\binom{N-|\alpha \cup \beta|}{m-|\alpha \cup \beta|}}{\binom{N}{m}}.$$

Key point: entries depend only on  $|\alpha \cup \beta|$  (equivalently intersection size), i.e. only on a *distance* in the Johnson geometry.

## Step 2: identify $\Sigma_{\text{unique}}$ with a homogeneous space

Think of a  $t$ -subset as the image of  $[t]$  under a basis permutation:

$$\alpha = g([t]) \quad \text{for } g \in S_N.$$

Then  $t$ -subsets are cosets:

$$X = S_N / (S_t \times S_{N-t}) \cong \binom{[N]}{t}.$$

So  $\Sigma_{\text{unique}} \cong L(X)$  (the permutation representation on  $X$ ).

### Why this helps

The subset-averaged operator is invariant under relabeling basis strings (the  $S_N$  action), so it becomes an  $S_N$ -invariant (“group-circulant”) matrix on  $L(X)$ .

# Gelfand pairs: multiplicity-free decomposition

## Definition (finite Gelfand pair)

$(G, K)$  is a Gelfand pair if  $L(G/K)$  decomposes into irreps *without multiplicity*.

For

$$(G, K) = (S_N, S_t \times S_{N-t}),$$

we have the classic **Johnson scheme** Gelfand pair.

## Irrep content (two-row Young diagrams)

$$L(X) \cong \bigoplus_{q=0}^t V_{[N-q, q]}.$$

Exactly  $t + 1$  blocks, each appearing once.

This is the representation-theoretic backbone behind “diagonalizing by symmetry.”

# Invariant matrices diagonalize on irrep blocks (non-Abelian Fourier)

- An  $S_N$ -invariant matrix  $M$  on  $L(X)$  depends only on the double coset  $K \backslash G / K$ , which here is indexed by Johnson distance  $p \in \{0, \dots, t\}$ .
- By Schur's lemma + multiplicity-free property:

$M$  is block-scalar on each  $V_{[N-q, q]}$ .

## Spherical functions give eigenvalues

For each  $\lambda = [N - q, q]$ , the eigenvalue is a distance-sum

$$\mu_\lambda = \sum_{p=0}^t \nu(p) \#\{\text{pairs at distance } p\} \Phi_\lambda(p),$$

where  $\nu(p)$  is the circulant kernel of  $M$  and  $\Phi_\lambda$  is the spherical function of the Gelfand pair.

## Typicality inside $L(X)$ : the largest block dominates

The irrep dimensions for two-row partitions satisfy (hook-length special case):

$$\dim V_{[N-q,q]} = \begin{cases} 1 & q = 0, \\ \binom{N}{q} - \binom{N}{q-1} & 1 \leq q \leq t. \end{cases}$$

In particular, the **largest block** is  $\lambda^* = [N - t, t]$  and

$$\frac{\dim V_{[N-t,t]}}{\dim \text{Sym}^t([N])} = 1 - O\left(\frac{t^2}{N}\right).$$

So controlling the eigenvalue on  $V_{[N-t,t]}$  almost controls the whole operator.

## Key estimate: the dominant eigenvalue is near Haar

After rescaling so Haar corresponds to identity on  $\Sigma_{\text{unique}}$ , the dominant eigenvalue satisfies

$$\mu_{[N-t,t]} = \left(1 - \frac{m}{N}\right)^t \left(1 + O\left(\frac{t^2}{m}\right)\right) = 1 + O\left(\frac{tm}{N}\right) + O\left(\frac{t^2}{m}\right).$$

### Interpretation of the two error terms

- $O(t^2/m)$ : collisions when sampling  $t$  elements from a set of size  $m$ .
- $O(tm/N)$ : excess overlap with  $|+^n\rangle$  when  $m$  is too large vs.  $N$ .

# From dominant block control to full trace distance

Use a general lemma: if a density matrix matches the maximally mixed state on a subspace that occupies  $1 - \varepsilon$  of the dimension, then overall trace distance is  $O(\delta + \varepsilon)$ .

Here:

$$\varepsilon = O\left(\frac{t^2}{N}\right), \quad \delta = O\left(\frac{tm}{N}\right) + O\left(\frac{t^2}{m}\right).$$

## Main bound (Paper 1)

For  $t = \text{poly}(n)$  and  $\omega(\text{poly}(n)) < m < o(2^n)$ ,

$$\text{TD}(\mathbb{E}_S |S\rangle\langle S|^{\otimes t}, \mathbb{E}_{\psi \sim \text{Haar}} |\psi\rangle\langle \psi|^{\otimes t}) \leq O\left(\frac{tm}{N}\right) + O\left(\frac{t^2}{m}\right).$$

# Cryptographic corollaries (high-level)

## PRS from a quantum-secure PRP

Given  $\text{PRP}_k : [M] \rightarrow [M]$ , define a keyed family

$$|\phi_k\rangle = \frac{1}{\sqrt{m}} \sum_{x \in [m]} |\text{PRP}_k(x)\rangle.$$

Hybrid argument: replace PRP by truly random permutation  $\Rightarrow$  truly random subset. Then use the information-theoretic bound to conclude pseudorandomness.

## Pseudoentanglement

Across any bipartition, Schmidt rank  $\leq m \Rightarrow$  entanglement entropy  $\leq \log m$ , yet the ensemble is computationally indistinguishable from Haar.

# Unifying viewpoint: Johnson scheme = common core

## Same algebra, two languages

- Paper 1:  $S_N$ -harmonic analysis on  $X = S_N / (S_t \times S_{N-t})$  (a Gelfand pair)

$$L(X) \cong \bigoplus_{q=0}^t V_{[N-q,q]}, \quad \text{invariant matrices are block-scalar.}$$

- Paper 2: adjacency matrices  $D_t$  span the Bose–Mesner algebra of the Johnson association scheme; diagonalization is “spectra of generalized Johnson graphs”.

Morally: **representation theory explains why only  $t + 1$  eigenvalues matter**, and why the “dominant block” governs the trace distance.

## References (primary)

- T. Giurgică-Tiron and A. Bouland, *Pseudorandomness from Subset States*, arXiv:2312.09206.
- F. Granha Jeronimo, N. Magrafta, and P. Wu, *Pseudorandom and Pseudoentangled States from Subset States*, arXiv:2312.15285.