Local Indistinguishability of Probability Distributions

PAN Weiheng, LYU Xin

Aug 22, 2019

Given two distributions p and q with disjoint supports on bit strings of length n, the goal of a **distinguisher** D is to tell whether a sample is from p or q by looking at $k(k \leq n)$ bits of the sample.

The process is as follows:

1. Take a sample from p or q with equal probability (both $1/2$)
2. The distinguisher queries a bit from the sample k times
3. The distinguisher decides which distribution the sample is from by outputting 1 if it is from p and 0 if it is from q

Recall that for some $s \in \{0,1\}^k$,

$$D(s) = \begin{cases} 1 & \text{if it decides that } s \in \mathrm{supp}(p) \\ 0 & \text{if it decides that } s \in \mathrm{supp}(q) \end{cases}$$

To provide a metric for the distinguishing ability of distinguishers, we define the notion of (distinguishing) **advantage** of a distinguisher D:

$$\mathrm{adv}(D) = \mathbb{E}[D(X)] - \mathbb{E}[D(Y)]$$

where

$$X \sim p, Y \sim q$$

And w.l.o.g. we can assume that

$$0 \leq \mathrm{adv}(D) \leq 1$$

Since there are different ways to decide where to look at in the sampled bit string, there are also different types of distinguishers:

1. **Non-adaptive distinguishers**: The bit positions that they look at are fixed.

2. **Adaptive distinguishers**: They have strategies and can decide on where to look at next based on the observed bits.

3. **Quantum distinguishers**: They are quantum circuits in which the querying process is implemented using quantum oracles.

For the three types of distinguishers, we define the largest advantages they can achieve when p, q are fixed, respectively:

adv_{NA} - The supremum of the advantages of **non-adaptive distinguishers**

adv_{AD} - The supremum of the advantages of **adaptive distinguishers**

adv_{Q} - The supremum of the advantages of **quantum distinguishers**

And we have a basic relationship between them:

$$\mathrm{adv}_{NA} \leq \mathrm{adv}_{AD} \leq \mathrm{adv}_{Q}$$

We are interested in the following quantities:

- Gaps: $\mathrm{adv}_{AD} - \mathrm{adv}_{NA}$ and $\mathrm{adv}_{Q} - \mathrm{adv}_{AD}$
- Ratios: $\dfrac{\mathrm{adv}_{AD}}{\mathrm{adv}_{NA}}$ and $\dfrac{\mathrm{adv}_{Q}}{\mathrm{adv}_{AD}}$

About:

- How large can they be?
- What are their asymptotic behaviors?

Secret sharing

- Distributes a secret among n parties
- Any group with t or more parties can reconstruct the secret
- But any group with strictly less than t parties cannot reconstruct
- Called an (n, t)-threshold scheme
- Examples
  - Bit-string XOR: $(2,2)$-threshold scheme
  - Intersection of n hyperplanes: (n, n)-threshold scheme
  - Chinese remainder theorem

p	0 0 1	0 1 0	1 0 0	1 1 1
	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
q	0 0 0	0 1 1	1 0 1	1 1 0
	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Let $0 \leq \varepsilon \leq \frac{1}{4}$, then $\varepsilon \leq \frac{1}{2} - \varepsilon$.

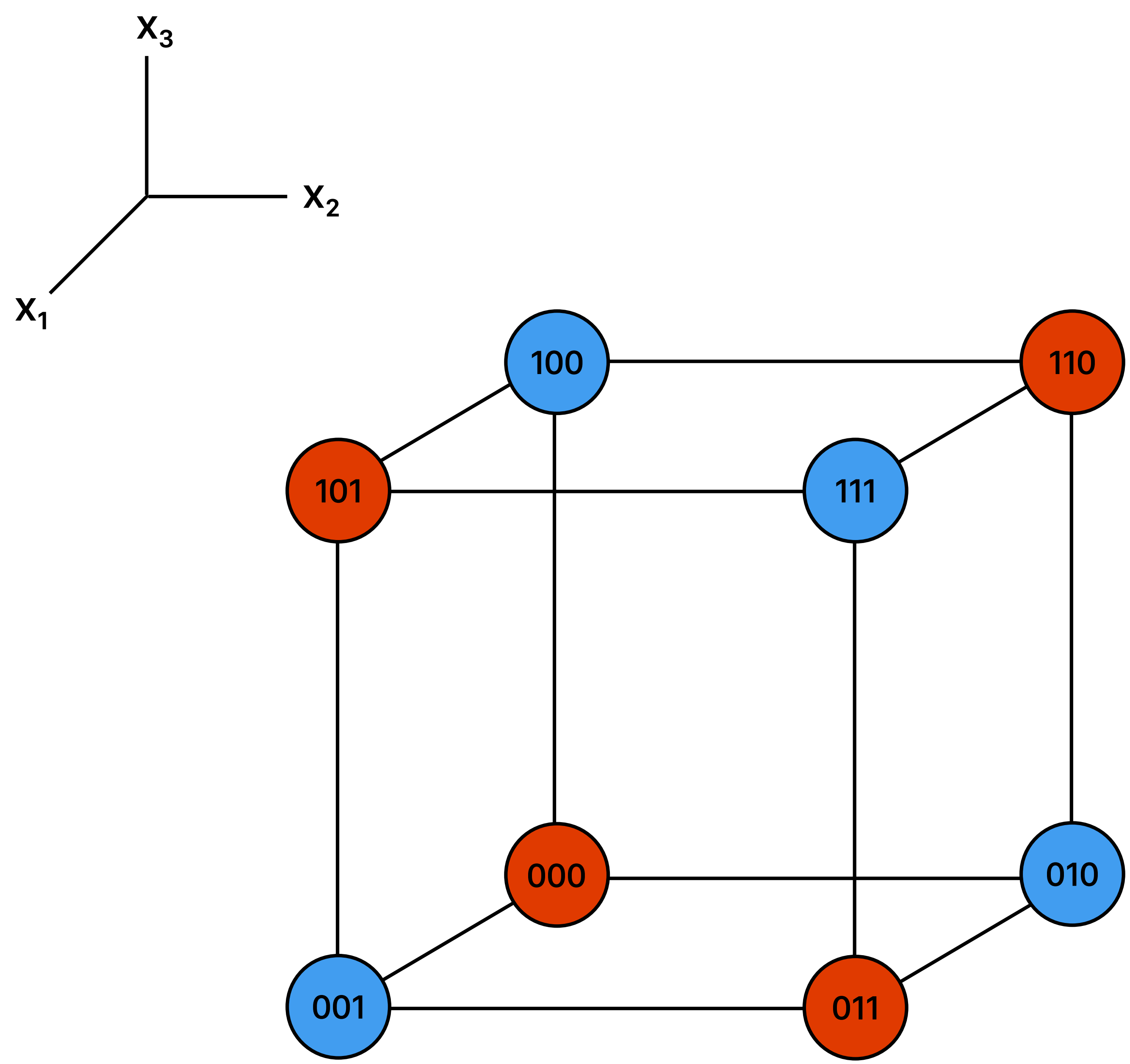
p	0 0 1	0 1 0	1 0 0	1 1 1
	ε	$\frac{1}{2} - \varepsilon$	ε	$\frac{1}{2} - \varepsilon$
q	0 0 0	0 1 1	1 0 1	1 1 0
	$\frac{1}{2} - \varepsilon$	ε	ε	$\frac{1}{2} - \varepsilon$

Let $0 \le \varepsilon \le \frac{1}{4}$. Then $\varepsilon \le \frac{1}{2} - \varepsilon$.

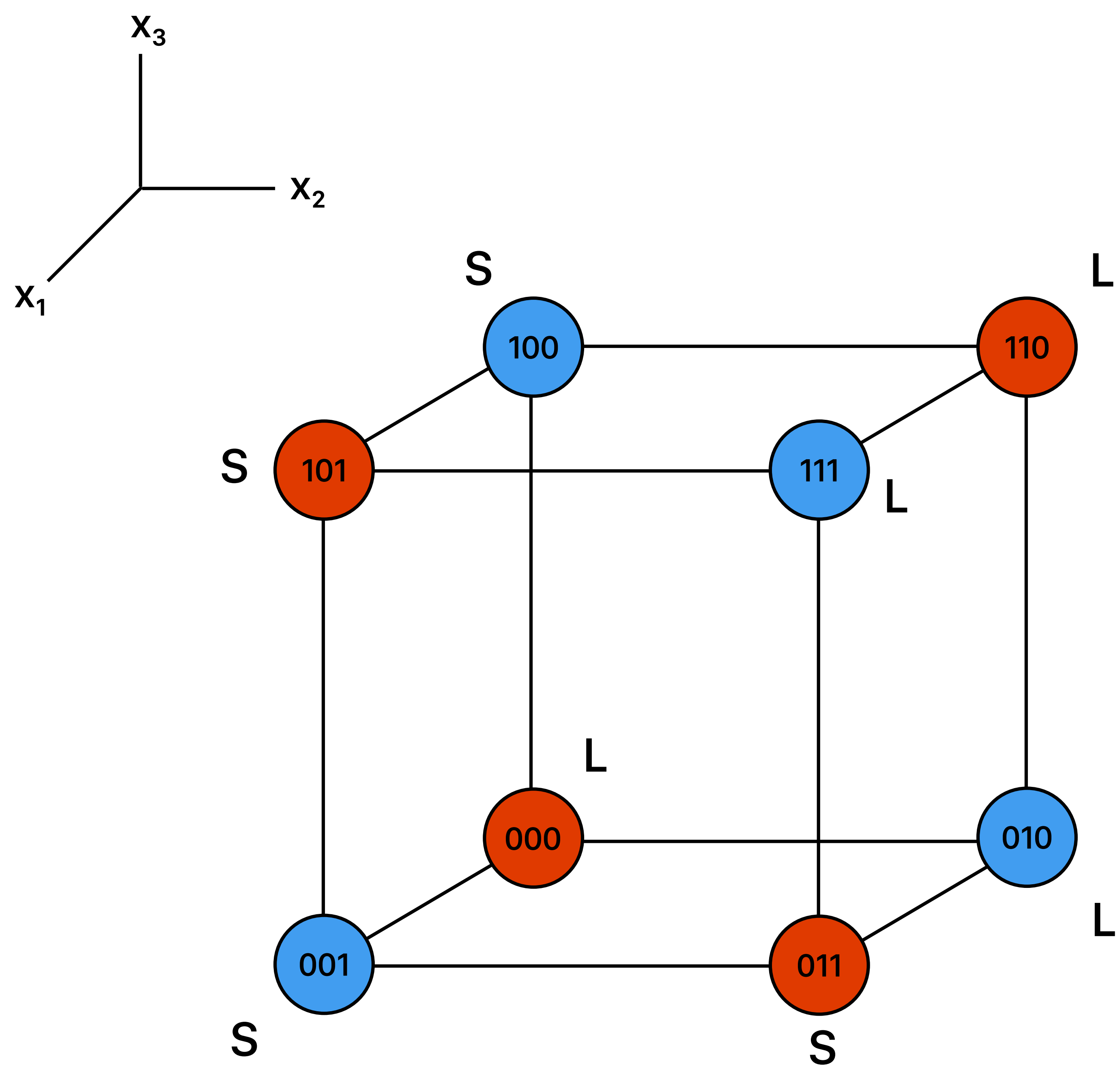
p	0 0 1	0 1 0	1 0 0	1 1 1
	ε	$\frac{1}{2} - \varepsilon$	ε	$\frac{1}{2} - \varepsilon$
q	0 0 0	0 1 1	1 0 1	1 1 0
	$\frac{1}{2} - \varepsilon$	ε	ε	$\frac{1}{2} - \varepsilon$

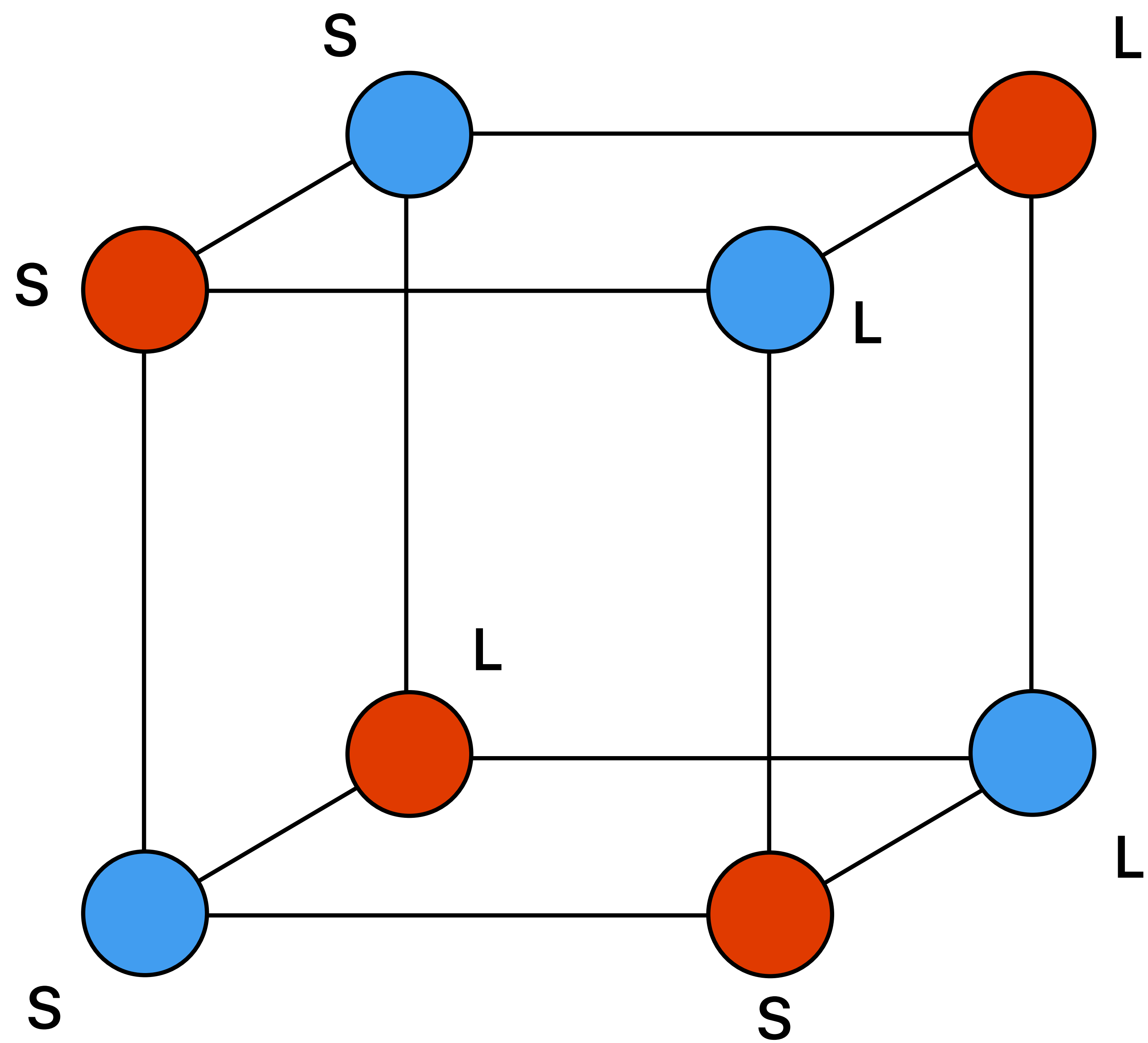
$$\mathrm{adv}_{AD} - \mathrm{adv}_{NA} = \frac{1}{2} - 2\varepsilon \qquad \frac{\mathrm{adv}_{AD}}{\mathrm{adv}_{NA}} = 2$$

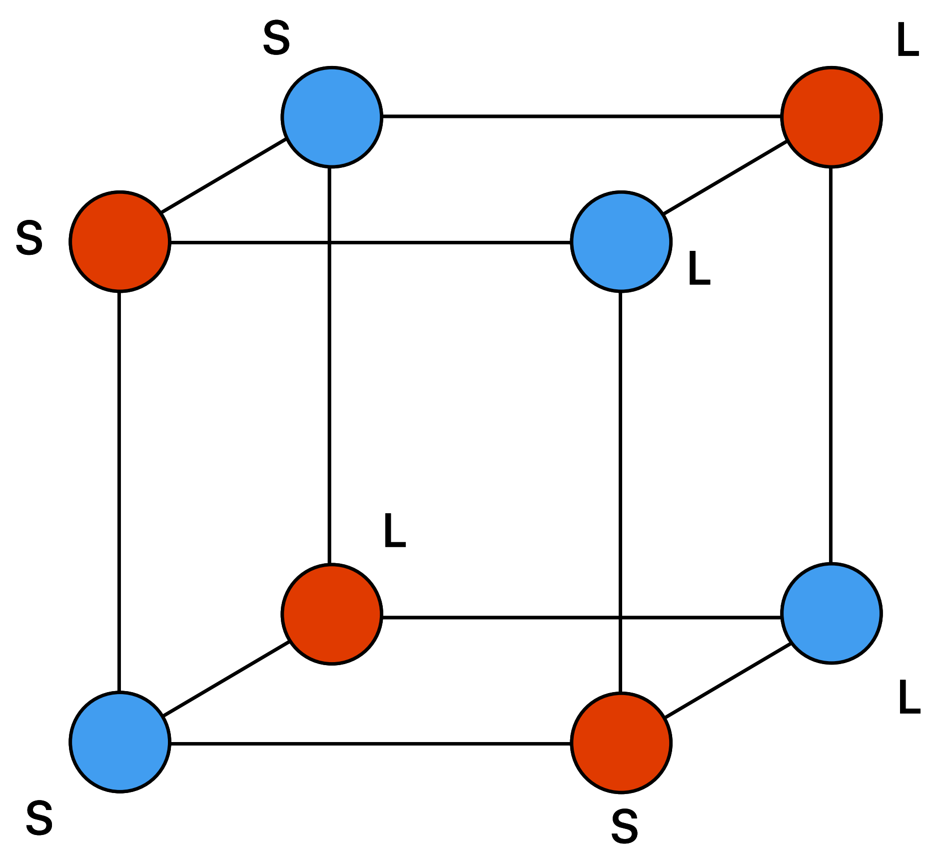
Let $S = \varepsilon,\ L = \dfrac{1}{2} - \varepsilon.$

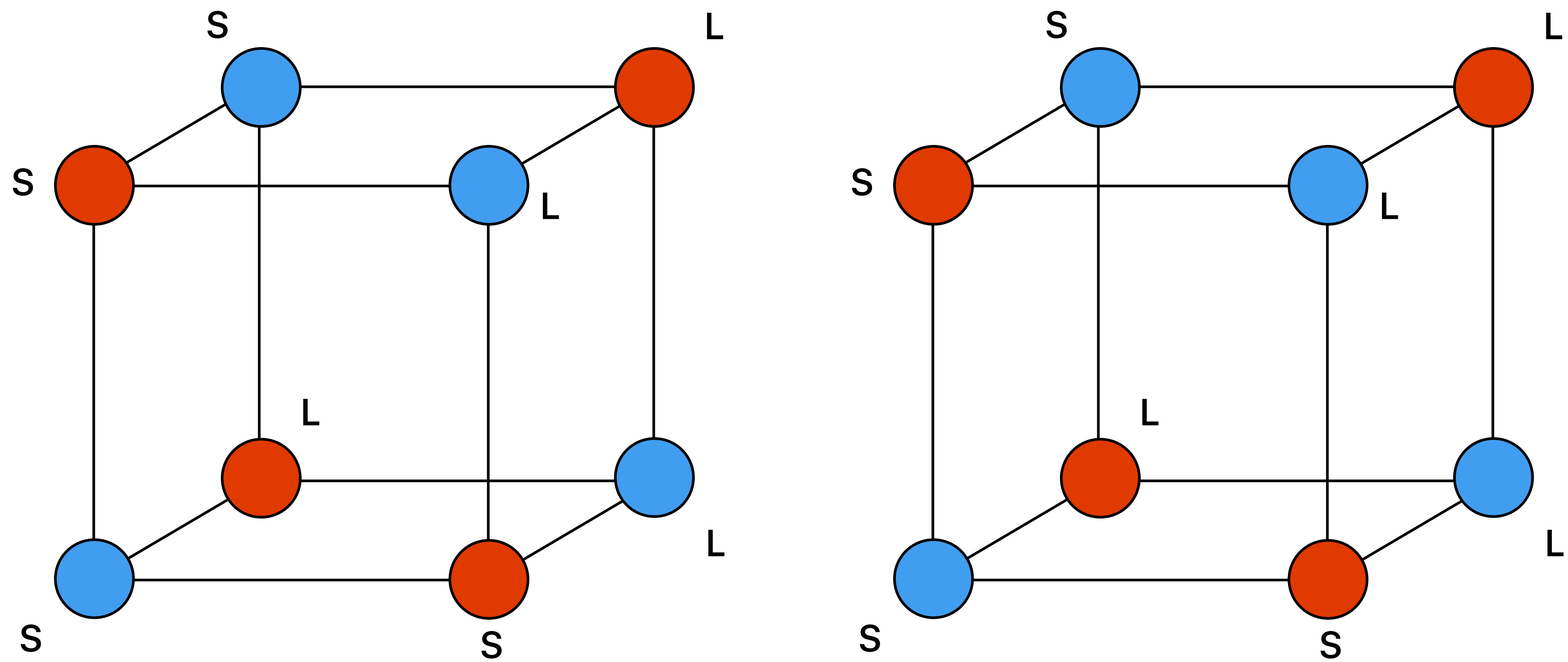


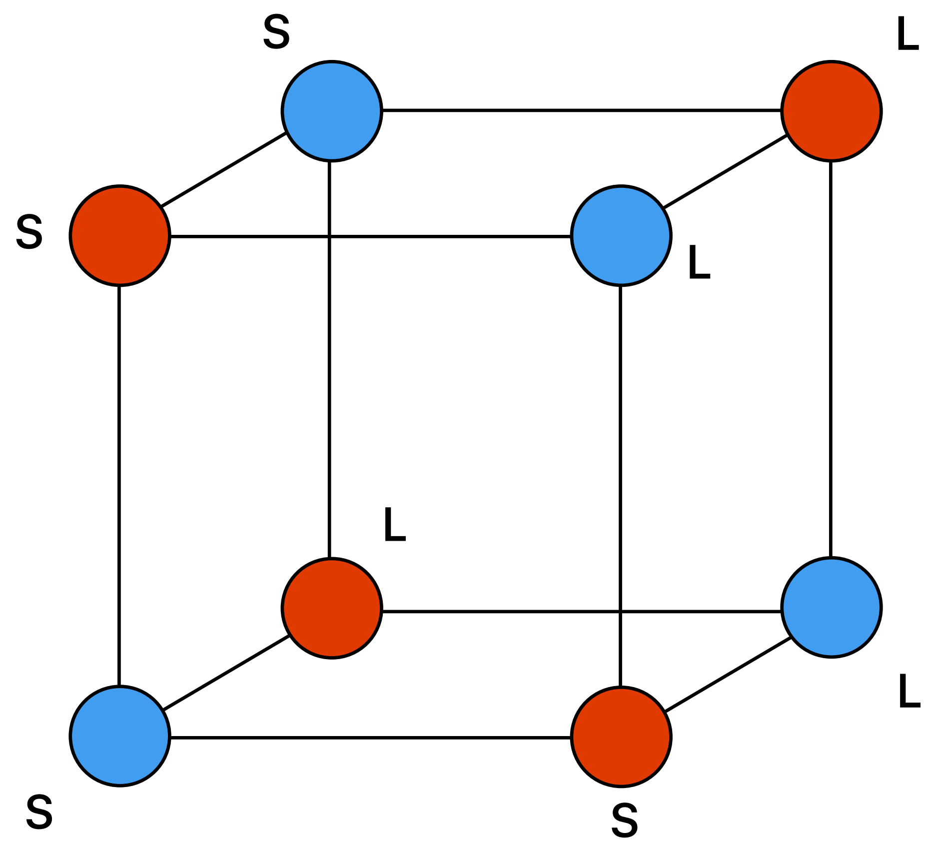
Let $S = \varepsilon,\ L = \dfrac{1}{2} - \varepsilon.$

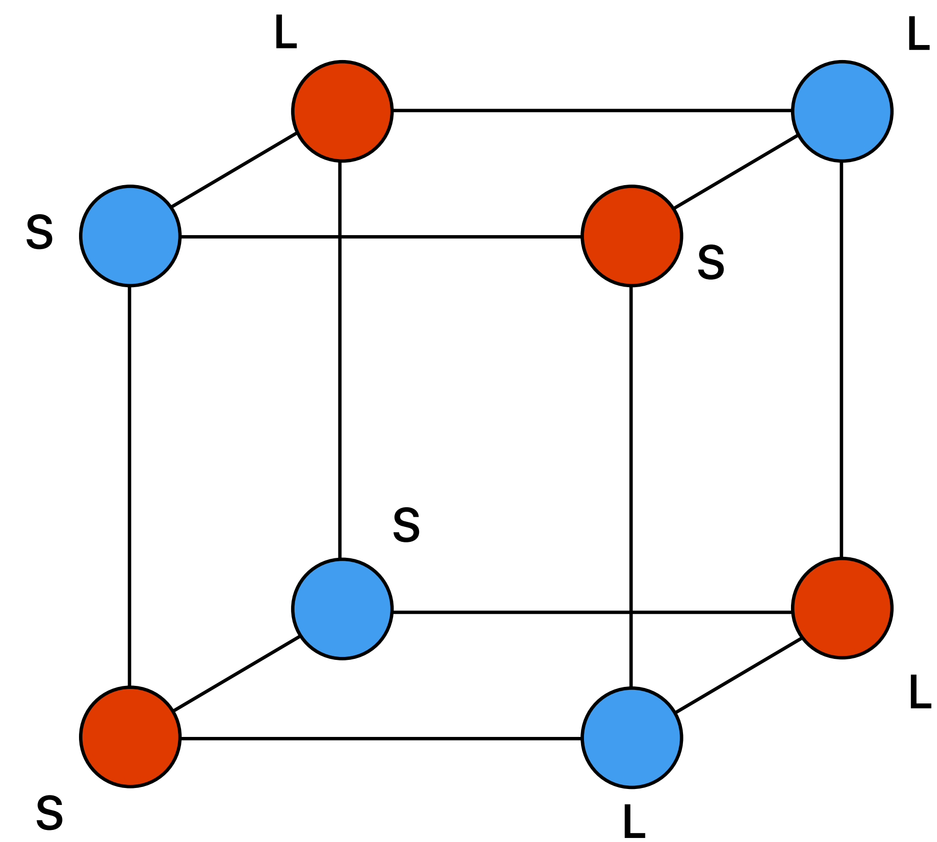


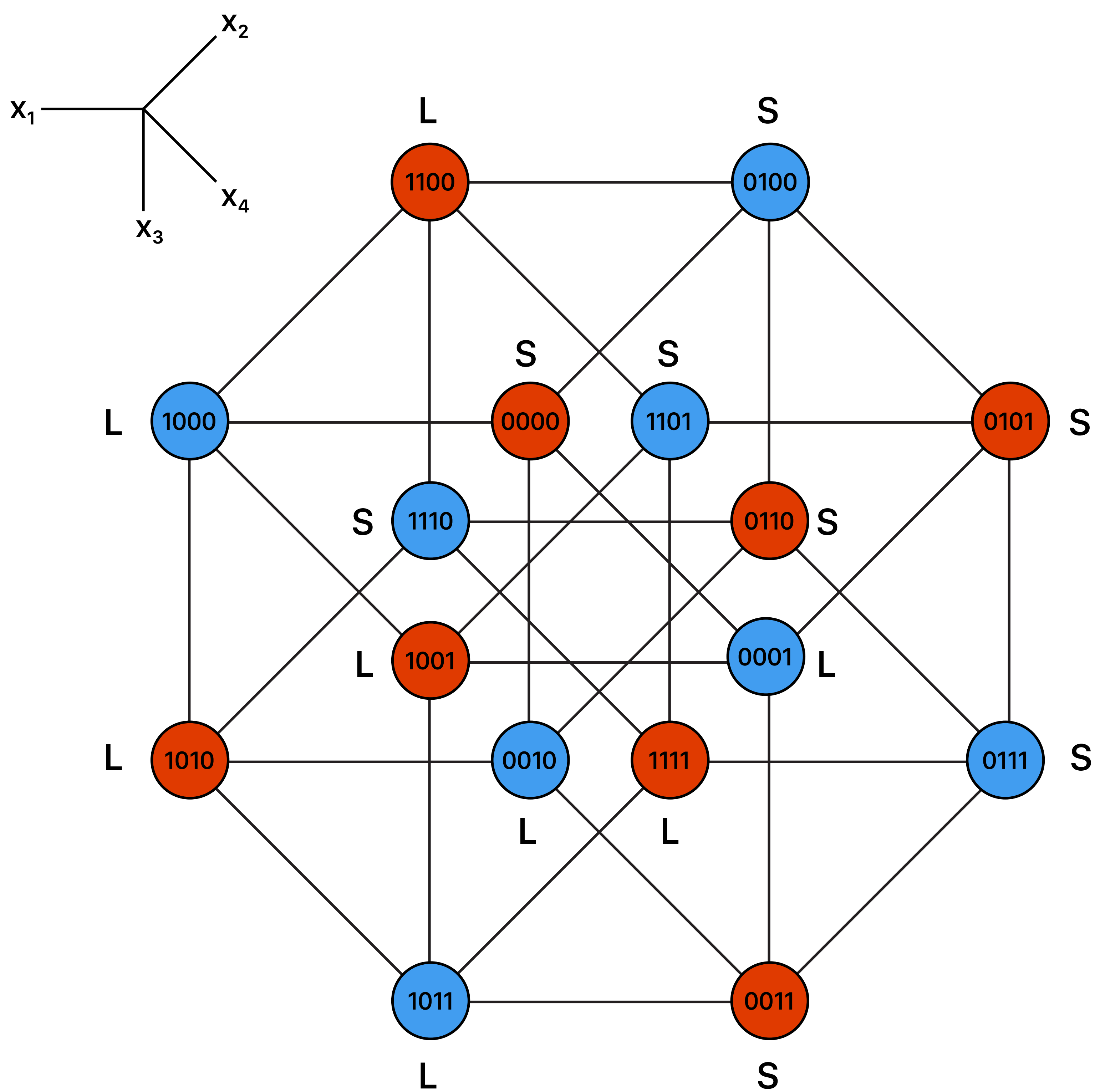

S
L
S
L
L
L
S
S

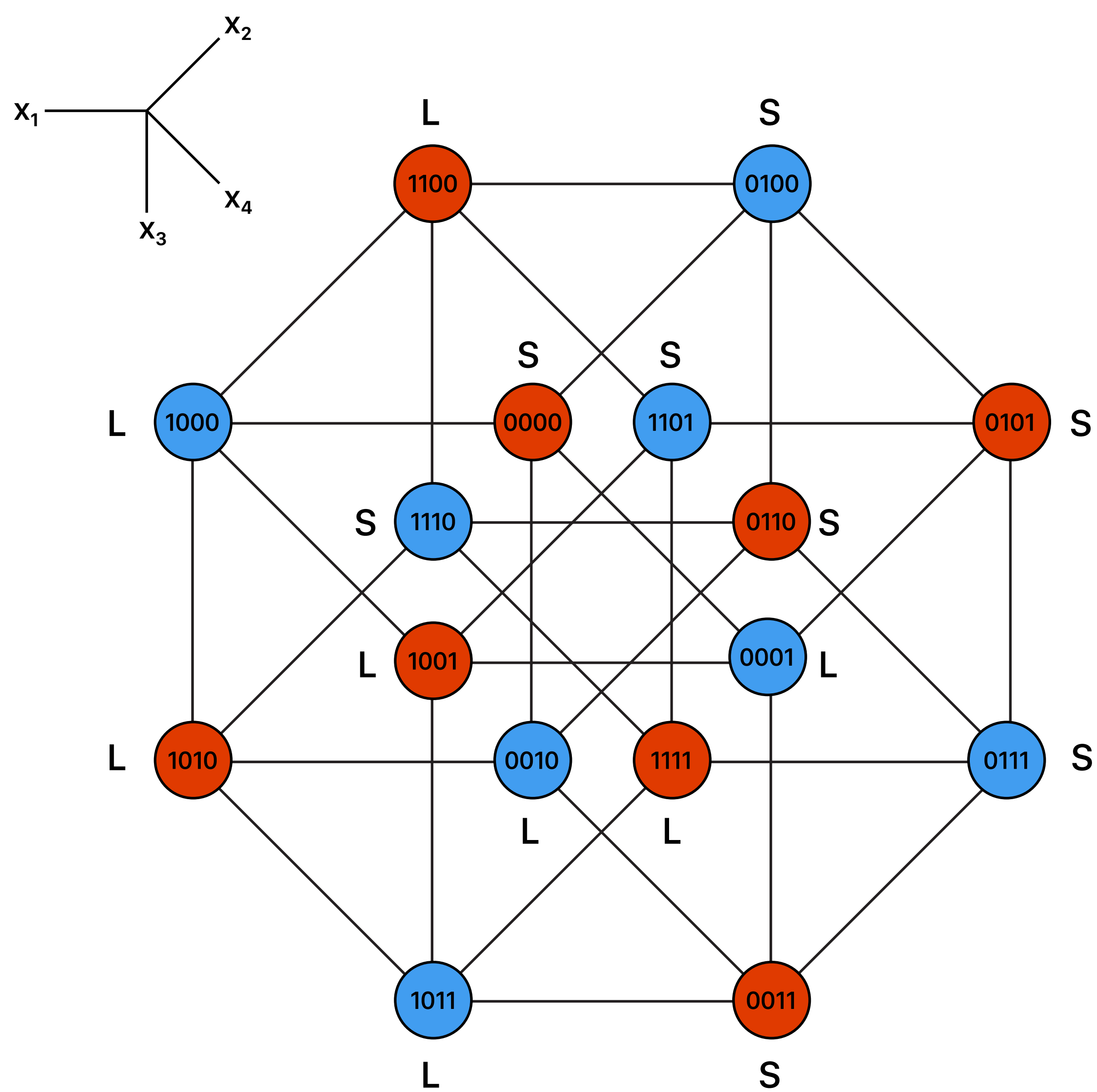

S
L
S
L
L
L
S
S


S
L
S
L
L
L
S
S
S
L
S
L
L
L
S
S


S
L
S
L
L
L
S
S


L
L
S
S
S
L
S
L


x1
x2
x3
x4
L
S
1100
0100
S
S
L
1000
0000
1101
0101
S
S
1110
0110
S
L
1001
0001
L
L
1010
0010
1111
0111
S
L
L
1011
0011
L
S



$$\mathrm{adv}_{AD} - \mathrm{adv}_{NA} = \frac{1}{2} - 2^{n-2}\varepsilon$$

$$\frac{\mathrm{adv}_{AD}}{\mathrm{adv}_{NA}} = 2$$

For any pair of distributions (p, q),

$$\frac{\mathrm{adv}_{AD}}{\mathrm{adv}_{NA}} \leq 2^{k-1}$$

For any adaptive distinguisher D_{AD}, there are 2^{k-1} permutations of bit positions that it may look at. Let Ω be the set of such permutations.

For any $\omega \in \Omega$, denote the non-adaptive distinguisher whose querying positions are ω by $D_{NA}(\omega)$. Let $\mathcal{N} = \{D_{NA}(\omega) \mid \omega \in \Omega\}$.

Let N be a random variable uniformly distributed on $\mathcal{N}$. Then

$$\frac{\mathrm{adv}(D_{AD})}{2^{k-1}} \leq \mathbb{E}[\mathrm{adv}(N)] = \sum_{n \in \mathcal{N}} \Pr[N = n]\mathrm{adv}(n) = \frac{1}{2^{k-1}} \sum_{n \in \mathcal{N}} \mathrm{adv}(n)$$

Hence $\sum_{n \in \mathcal{N}} \mathrm{adv}(n) \geq \mathrm{adv}(D_{AD})$. Since $|\mathcal{N}| = 2^{k-1}$, we conclude that

$$\exists n_0 \in \mathcal{N} \text{ s.t. } \mathrm{adv}(n_0) \geq \frac{\mathrm{adv}(D_{AD})}{2^{k-1}} \implies \frac{\mathrm{adv}_{AD}}{\mathrm{adv}_{NA}} \leq 2^{k-1} \quad \forall p, q$$

Consider the following type of bit strings of length $n = m + 2^m$:

$$\underbrace{a_1 a_2 \cdots a_m}_{\text{addr}} \underbrace{x_1 x_2 x_3 x_4 \cdots x_{2^m - 1} x_{2^m}}_{\text{data}}$$

Let $x_a = x_{(a_1 \cdots a_m)_2 + 1}$. A pair of **Address-Data Distributions** is defined as follows:

p contains 2^m strings of the following type:

$$a_1 \cdots a_m \quad \underbrace{x_1 \cdots x_{a-1}}_{\text{Bernoulli}(1/2)} \quad 1 \quad \underbrace{x_{a+1} \cdots x_{2^m}}_{\text{Bernoulli}(1/2)}$$

Where:

- $a_1 a_2 \cdots a_m$ ranges over all 2^m permutations
- $x_a = 1$ and $\forall i \neq a, x_i = \begin{cases} 0 & \text{w.p. } 1/2 \\ 1 & \text{w.p. } 1/2 \end{cases}$
- All strings are uniformly distributed

Replacing $x_a = 1$ by $x_a = 0$ gives q

Let $k = m + 1$. There exists an perfectly-distinguishing adaptive distinguisher with the following strategy:

- Query the first m address bits and calculate an index $a = (a_1 \cdots a_m)_2 + 1$
- Query the $(m + a)$-th bit, which is the a-th bit in the data section
- If the result is 1, then the sample is from p; otherwise it is from q

Hence $\mathrm{adv}_{AD} = 1$.

For non-adaptive distinguishers that query $k = 2m + 1$ bits, the best possible is to query all m address bits and query $m + 1$ bits out of the 2^m data bits.

Furthermore, it must hit x_a to gain some advantage.

$$\begin{aligned}\Pr[\text{hit } x_a] &= 1 - \Pr[\text{does not hit } x_a] \\ &= 1 - \prod_{i=0}^{m} (1 - \frac{1}{2^m - i}) \\ &= 1 - \frac{2^m - m - 1}{2^m} = \frac{m+1}{2^m}\end{aligned}$$

Hence $\Pr[\text{hit } x_a] \to 0$ as $m \to \infty$.

Since $\mathrm{adv}_{NA} \leq \Pr[\text{hit } x_a]$, we know that $\mathrm{adv}_{NA} \to 0$ as $m \to \infty$.

Since an NA distinguisher with $k = m + 1$ observes less bits than one with $k = 2m + 1$, it also holds for the former.

Since $\mathrm{adv}_{AD} = 1$, it follows that for $k = m + 1$,

$$\lim_{m\to\infty} (\mathrm{adv}_{AD} - \mathrm{adv}_{NA}) = 1 \text{ and } \lim_{m\to\infty} \frac{\mathrm{adv}_{AD}}{\mathrm{adv}_{NA}} = \infty$$

Because $\mathrm{adv}_{NA} \leq \Pr[\text{hit } x_a] = \frac{m+1}{2^m}$, we have $\mathrm{adv}_{NA} = O(\frac{m}{2^m})$.

Recall that $n = m + 2^m$. Hence $\mathrm{adv}_{NA} = O(\frac{\log n}{n})$.

A quantum distinguisher is a quantum circuit, with one or more oracles U_x , where x denotes the hidden string. They are defined as follows:

$$|i\rangle|r\rangle \mapsto |i\rangle|r \oplus x_i\rangle$$

In the circuit, querying U_x up to k times is allowed.

Finally, QD outputs 0 or 1 according to its judgement for x.

Through a simple construction inspired by the Deustch-Jozsa algorithm, we can show a quantum distinguisher can be more powerful than classical ones.

Assume $n = 2,\ k = 1$, consider the following example:

p: 00, 11

q: 01, 10

Obviously, classical distinguisher cannot achieve any nonzero advantage.

However, there exists a quantum distinguisher with advantage 1.

Consider the following transformation: (ignoring normalization factors)

$$|i\rangle(|0\rangle - |1\rangle) \overset{U_x}{\mapsto} |i\rangle(|x_i\rangle - |1 \oplus x_i\rangle) = (-1)^{x_i}|i\rangle(|0\rangle - |1\rangle)$$

If we ignore the auxiliary qubit, what we are implementing is a gate $|i\rangle \mapsto (-1)^{x_i}|i\rangle$.

We prepare a state $|1\rangle + |2\rangle$ and apply the gate above:

If the sample is from p, then the state is $\pm(|1\rangle + |2\rangle)$

If the sample is from q, then the state is $\pm(|1\rangle - |2\rangle)$

They are orthogonal to each other and thus can be distinguished perfectly.

First, note that our circuit can actually calculate the XOR sum of two bits using a single query. Therefore, our construction can be generalized to larger n easily.

> **Theorem 1**: for any k, we can construct a pair of distributions on $n = 2k$ bit string. Where a quantum distinguisher can achieve perfect distinguishing, but a classical one can not gain non-zero advantage by no more than $2k - 1$ queries.

The construction is simple. Let p be uniformly distributed on all length-$2m$ strings with parity 0, and q be uniformly distributed on all strings with parity 1. While a quantum distinguisher can calculate the parity of a string with only m queries, a classical one can not do this by using less than $2m$ queries.

Secondly, people have also showed:

> **Theorem 2**: If a classical distinguisher cannot obtain nonzero advantages using $2m$ queries, then a quantum distinguisher using m queries cannot do so either.

In this sense, our construction is optimal.

The proof of the second theorem follows from the fact that the accepting probability of a k-query quantum algorithm is a polynomial of degree $2k$. [BBC+01]

[BBC+01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. J. ACM, 48(4):778–797, July 2001.

Thank You!