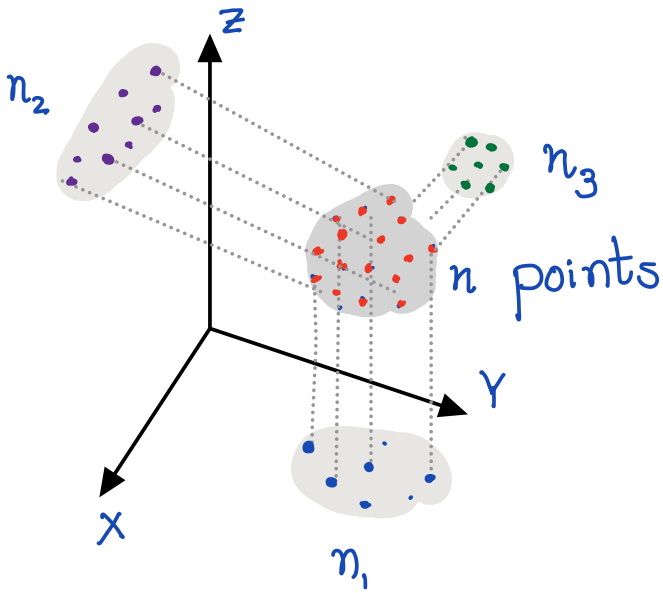
It is entropy that counts

ICTS monthly colloquium

14 November 2022

Points in three dimensions



n points in $\mathbb{R}^3$

n_1 distinct projections on XY

n_2 distinct projections on XZ

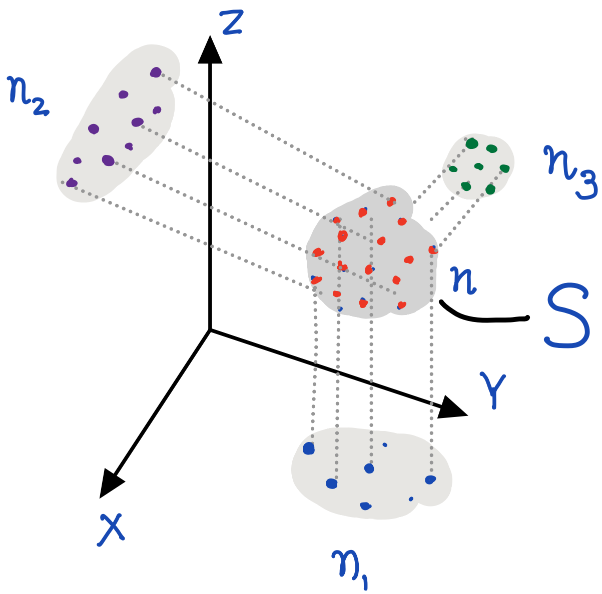
n_3 distinct projections on YZ

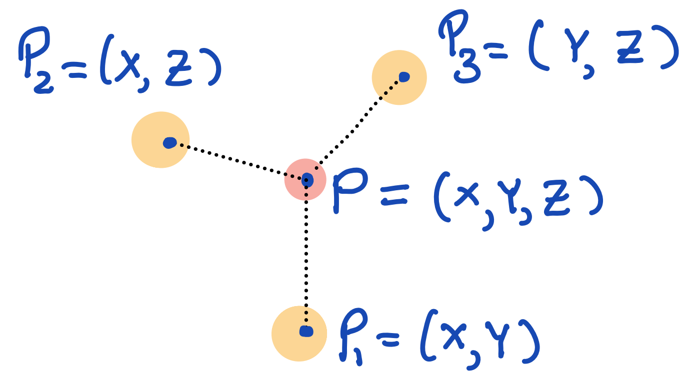
Then, $n_1 n_2 n_3 \geqslant n^2$.

Loomis-Whitney inequality, 1949

Why?

Information





To specify one among a set of n possibilities, we require $\log n$ bits of information.

Each piece of information about P is available from two sources. So, obviously ...

$$2 \log n \leq \log n_1 + \log n_2 + \log n_3$$

$$\Downarrow$$

$$n^2 \leq n_1 n_2 n_3$$

Entropy

It is entropy that counts.

Pick P uniformly at random from S.

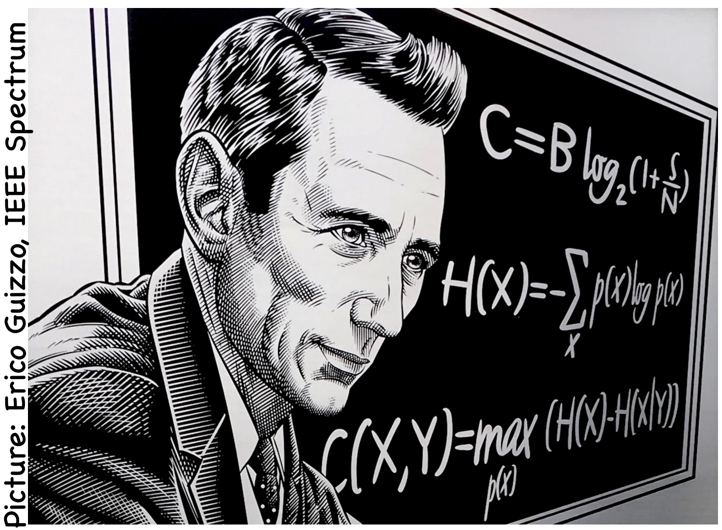
P has maximum entropy, so $H[P] = \log n$.

$$\log n = H[P] = H[(X,Y,Z)] = H[X] + H[Y|X] + H[Z|XY]$$
$$\log n_1 \geq H[P_1] = H[(X,Y)] = H[X] + H[Y|X]$$
$$\log n_2 \geq H[P_2] = H[(X,Z)] = H[X] + H[Z|X]$$
$$\log n_3 \geq H[P_3] = H[(Y,Z)] = H[Y] + H[Z|Y]$$

$$\log n_1 + \log n_2 + \log n_3 \geq H[P_1] + H[P_2] + H[P_3] \geq 2H[P] = 2\log n$$

Shannon entropy

$$X \equiv \begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ p_1 & p_2 & \cdots & p_r \end{pmatrix} \begin{matrix} \leftarrow \text{outcomes} \\ \leftarrow \text{probabilities} \end{matrix}$$



Picture: Erico Guizzo, IEEE Spectrum

Claude Shannon (1916-2001)

- $H[X] = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} + \cdots + p_r \log_2 \frac{1}{p_r}$

  $\leq \log_2 r$

- $H[f(X)] \leq H[X]$

$$\begin{pmatrix} a_1 & a_2 & \cdots & & a_r \\ \underbrace{p_1 \quad p_2} & & \underbrace{\cdots} & & \underbrace{p_r} \end{pmatrix}$$

Conditional entropy, Mutual information

X, Y : Random variables with some joint distribution

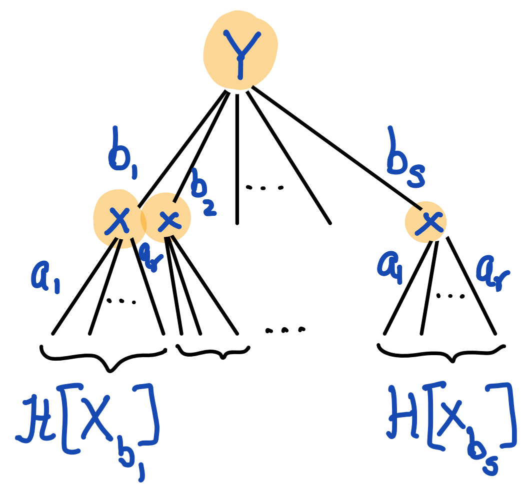
$$H[XY] = H[(X,Y)] = \sum_{ij} p_{ij} \log_2 \frac{1}{p_{ij}}$$

- Conditional entropy of X given Y

$$H[X|Y] = H[XY] - H[Y]$$

$$= \mathop{E}_{y \leftarrow Y}\big[H[X_y] \big]$$

average



(theorem!) $\leq \; H[X]$

- Mutual information $\quad I[X:Y] = H[X] - H[X|Y]$

$$= H[X] + H[Y] - H[XY]$$

Operational motivation

A
Sees X
Wants to tell B

Sends B a message
$M(X)$

B
Decodes $M(X)$
and obtains X

Suppose we know that $X \equiv \begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ p_1 & p_2 & & p_r \end{pmatrix}$.

How many **bits** must A send on average?

blue - 001
red - 00001
green =

It is entropy that counts.

$$H[X] \le T[X] \le H[X] + 1$$

Transmission cost for sending X.

Coin A

$$\begin{array}{cc} 0 & 1 \\ \end{array}$$
$$\begin{pmatrix} 0.5 & 0.5 \end{pmatrix}$$

$H[A] = 1$

$T(A) = 1$

Coin B

$$\begin{array}{cc} 0 & 1 \\ \end{array}$$
$$\begin{pmatrix} 0.25 & 0.75 \end{pmatrix}$$

$H[B] = 0.81$

$T(B) = 1$

What does entropy count?

What does entropy count?

Coin A

$\begin{pmatrix} 0 & 1 \\ 0.5 & 0.5 \end{pmatrix}$

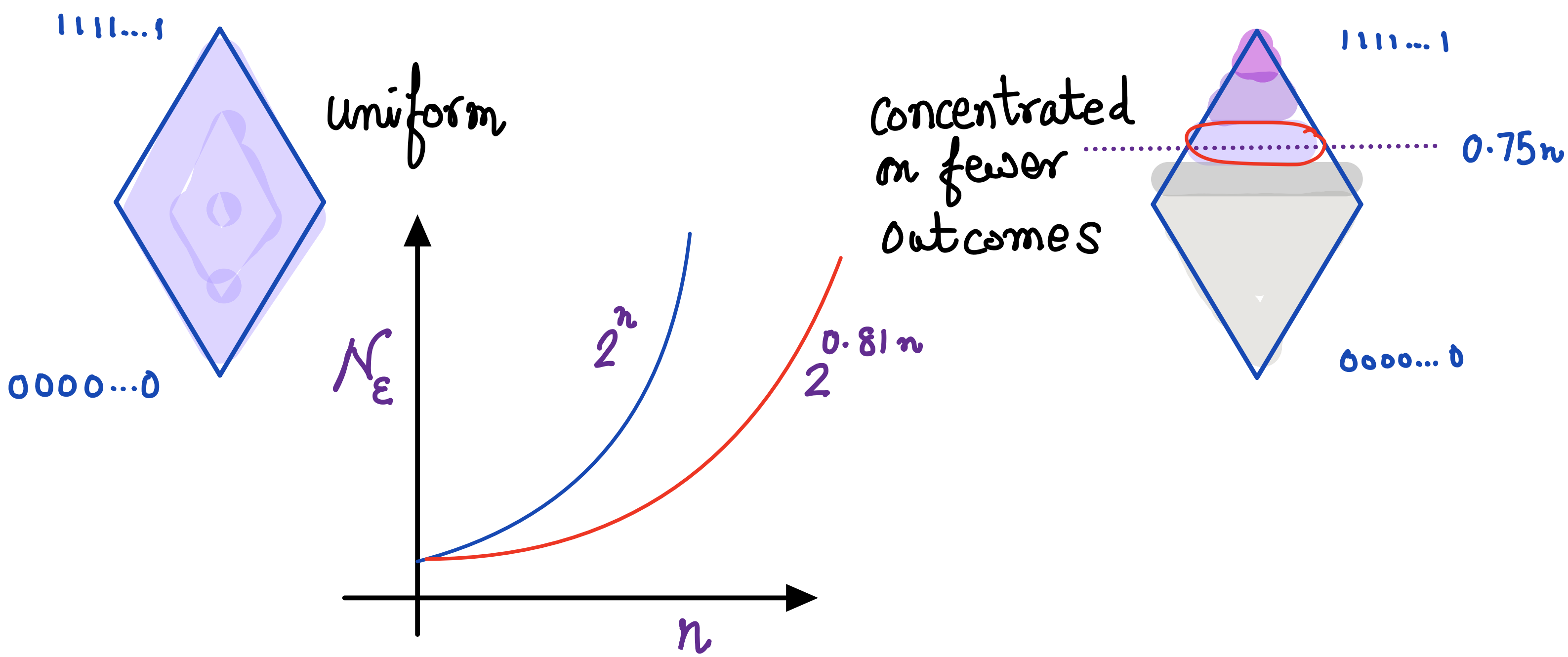
#outcomes = 2^n

Toss the coin n times, independently.

Coin B

$\begin{pmatrix} 0 & 1 \\ 0.25 & 0.75 \end{pmatrix}$

#outcomes = 2^n



Asymptotics

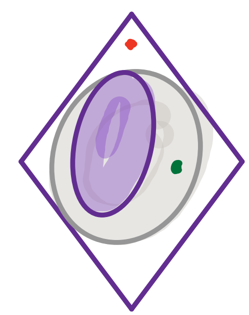
$$X \equiv \begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ p_1 & p_2 & & p_r \end{pmatrix}; \quad X^{(n)} = \underbrace{X_1 X_2 \cdots X_n}_{}$$

n independent samples

#outcomes= r^n

(some impossible, some unlikely)

$$\boxed{\begin{array}{l} N_\varepsilon(n) = \min |S| \\ \qquad \text{Set of outcomes with total probability} \geq \varepsilon > 0 \\ \forall \varepsilon > 0 \quad \lim_{n \to \infty} \dfrac{\log N_\varepsilon(n)}{n} = H[X] \end{array}}$$



It is entropy that

Not wholly

Or in full measure

But substantially

- Not all outcomes
- Not even all outcomes with positive probabilities
- But enough outcomes with total probability at least ε

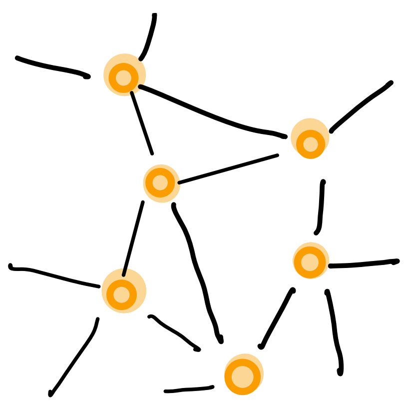
Back to combinatorics

G: a graph

n: number of nodes

d: degree of each vertex

d=4



The number of walks of length $= nd^r$

What if all the vertices don't have the same degree?

Claim: $\#\text{walks} \geq n\bar{d}^r$ average degree

- Pick a random vertex v_0 with prob. proportional its to degree.
- Perform a random walk

$$\log \#\text{walks} \geq H[v_0, v_1, \dots, v_r]$$

$$= H[v_0] + H[v_1 | v_0] + \dots + H[v_r | v_0 \dots v_{r-1}]$$

$$\underset{(!)}{=} \log n + \sum_{i=1}^{r} E[\log \deg(v_i)]$$

The v_i are identically distributed!

$$\underset{(!)}{\geq} \log n + r \log \bar{d}$$

A better formulation

- Pick one of the $n\bar{d}$ edges at random, say,

$$\vec{e}_1 = (v_0, v_1)$$

- From v_1, perform a random walk to obtain

$$v_0 \xrightarrow{\vec{e}_1} v_1 \xrightarrow{\vec{e}_2} v_2 \xrightarrow{\vec{e}_3} v_3 \cdots v_{r-1} \xrightarrow{\vec{e}_r} v_r$$

- $H[e_1 e_2 \ldots e_r] = H[e_1] + H[e_2 | e_1] + \cdots + H[e_r | e_1 e_2 \ldots e_{r-1}]$

$$= \log n\bar{d} + \sum_{i=2}^{r} H[e_i | v_{i-1}]$$

$$= \log n\bar{d} + (r-1) \mathop{E}_{v}[\log d_v]$$

$$\geq \log n\bar{d} + (r-1)\log \bar{d}$$

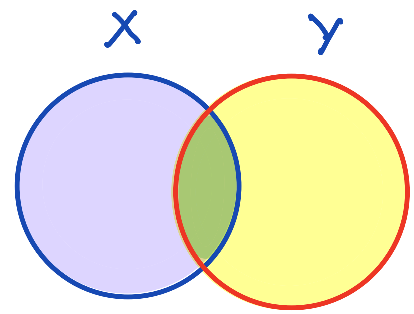
$$\geq \log n\bar{d}^{r}.$$

Because the v_i are identically distributed according to the stationary distribution of the walk.

Mutual information

X, Y: random variables with some joint distribution

$$I[X:Y] = H[X] + H[Y] - H[XY]$$
$$= H[X] - H[X|Y]$$
$$= H[Y] - H[Y|X]$$



Operational interpretation

A **B**

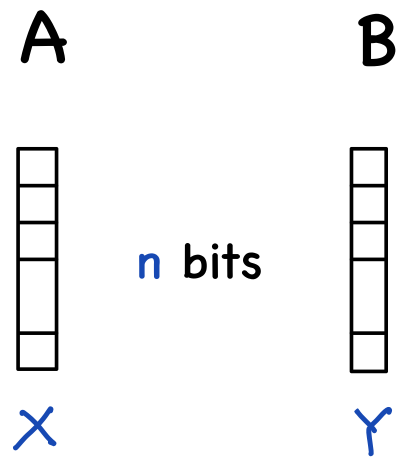
Receives X Wants to generate Y

How many bits must A send B?

Today

A communication complexity problem

Communication complexity



Determine if $X = Y$

How many bits must they exchange?

- Deterministically at least n bits.

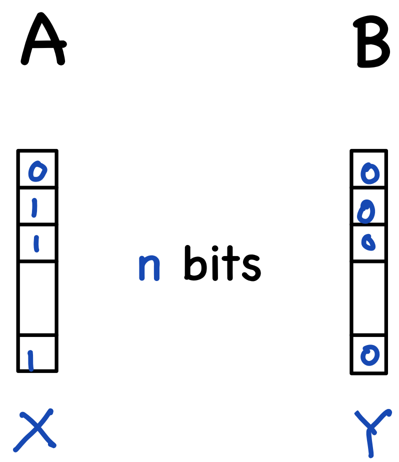
  Every input of the form (x,x) requires a different $\underbrace{\text{communication pattern}}_{\text{transcript}}$.

- With randomness,

  $O(\log n)$ bits are enough!

Communication complexity

Set disjointness



Determine if there is an i such that X[i]=Y[i]=1

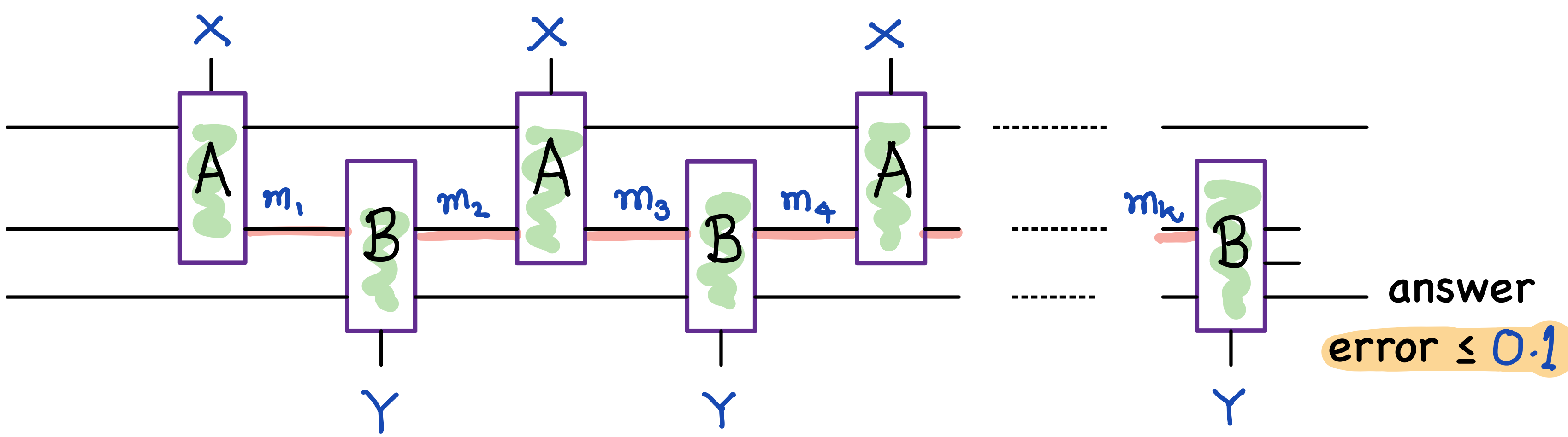
How many bits must they exchange?

How many bits must they exchange?

- Deterministically, at least n bits.
- Randomness does not help!
  A and B still need to exchange almost n bits.
  Kalyanasundaram and Schnitger, 1987
  Razborov, 1991
- Quantum communication helps!

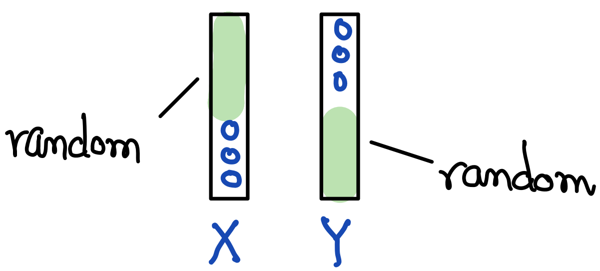
  A and B need exchange only $\sqrt{n}$ qubits.
  Buhrman, Clare and Wigderson, 1998
  Aaronson & Ambainis, 2005
  Razborov, 2003

Randomised communication protocols



Transcript $= \tau(X, Y) = (m_1, m_2, \ldots, m_k)$ ⟵ total length $\leq \frac{n}{100}$ (say)

Distributions



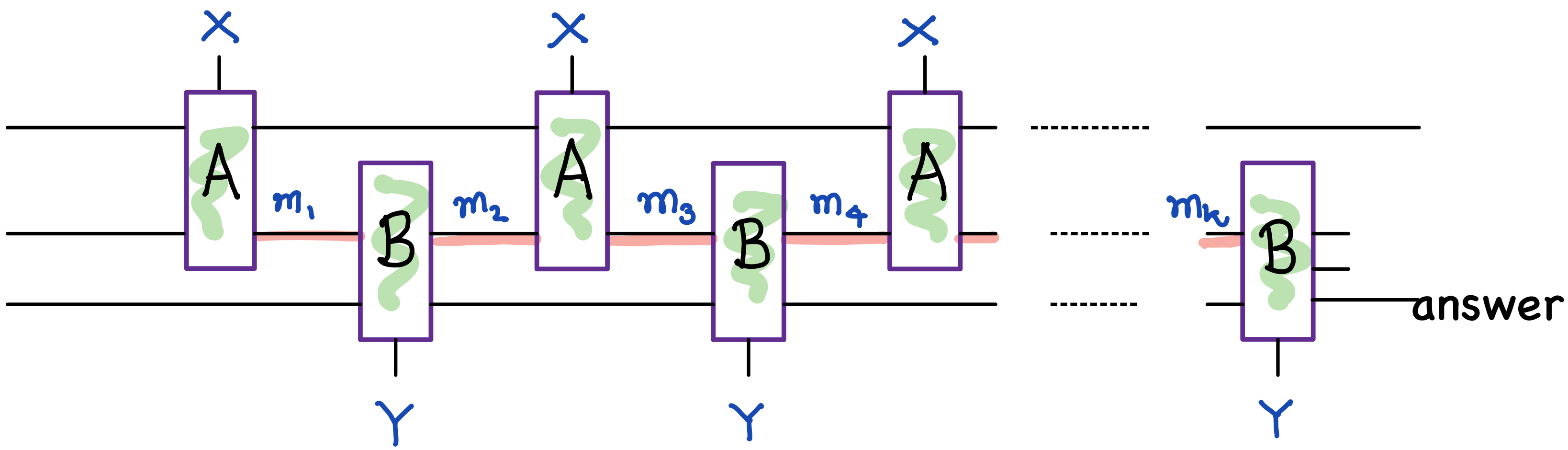
2^n such distributions. For each such distribution

$$I[X_1 \ldots X_n : \tau] \leq H[\tau] \leq n/100$$

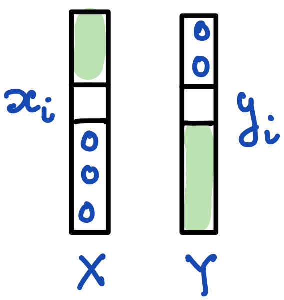
$$I[Y_1 \ldots Y_n : \tau] \leq H[\tau] \leq n/100$$

Information about a typical coordinate $\leq \frac{1}{100}$

Randomised communication protocols

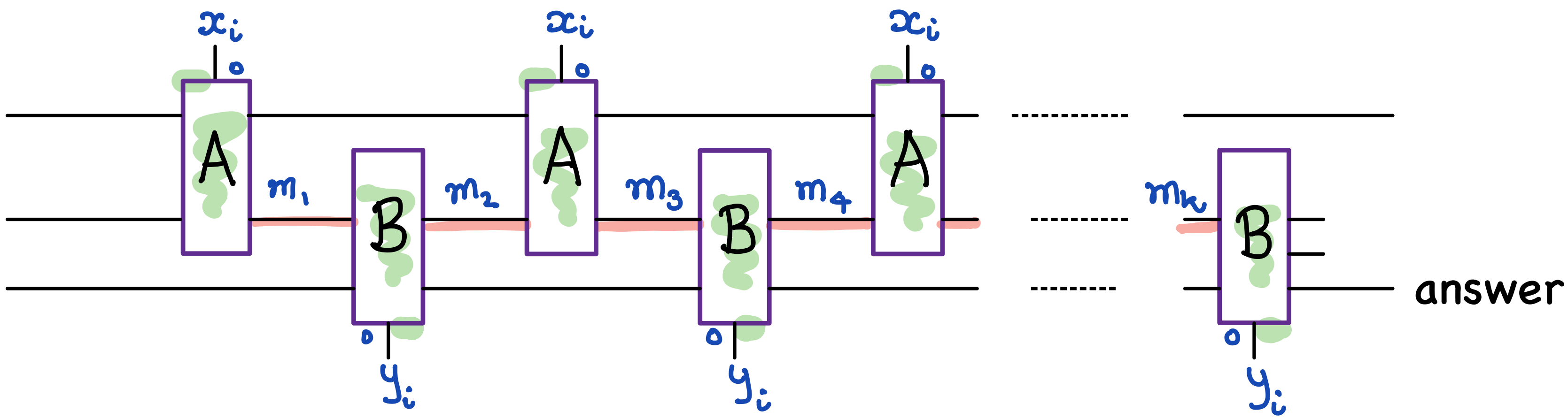


There is a coordinate i such that for a distribution of the form



- if $y_i = 0$, then $I[x_i : \tau] \leq 1/50$
- if $x_i = 0$, then $I[y_i : \tau] \leq 1/50$

Randomised communication protocols



- if $y_i = 0$, then $I[x_i : \tau] \leq 1/50$
- if $x_i = 0$, then $I[y_i : \tau] \leq 1/50$

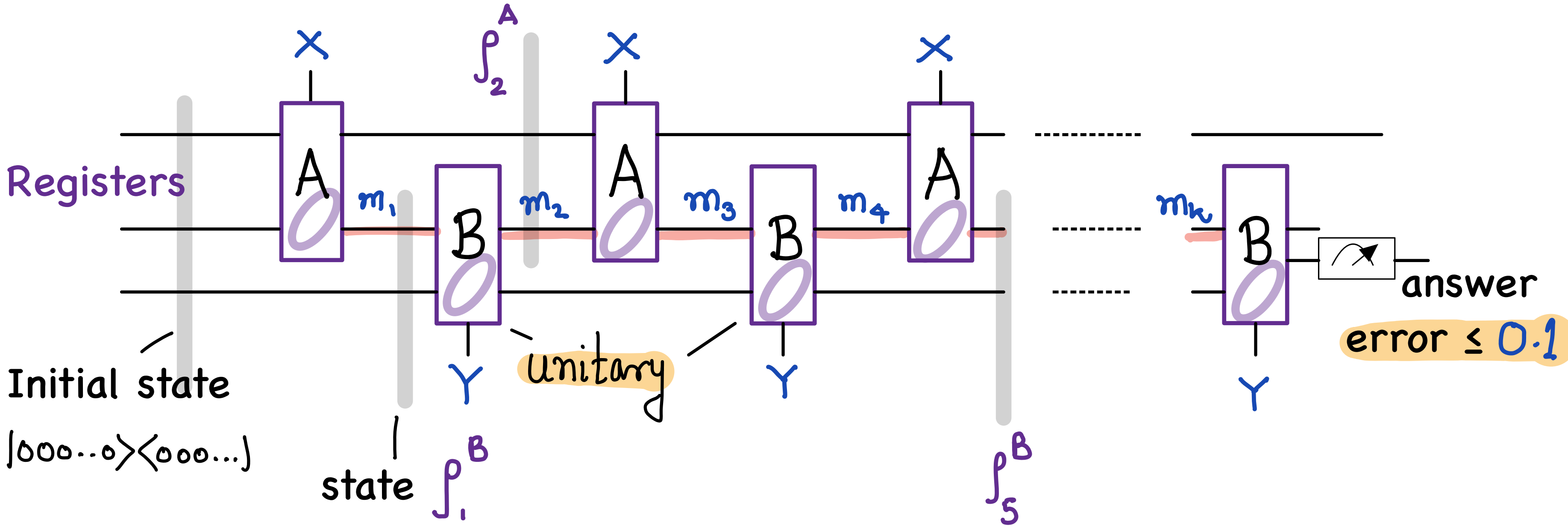
Neither party is willing to reveal much for they are afraid that the other party might have 0.

↓

The protocol must err with probability $\geq 1/4$.

The communication was long but fruitless.
The length does not count.

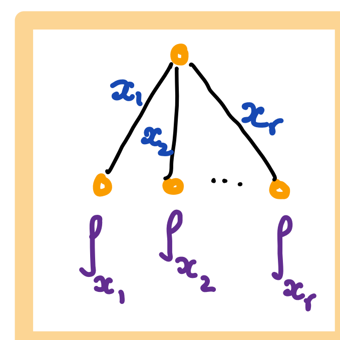
Quantum communication protocols



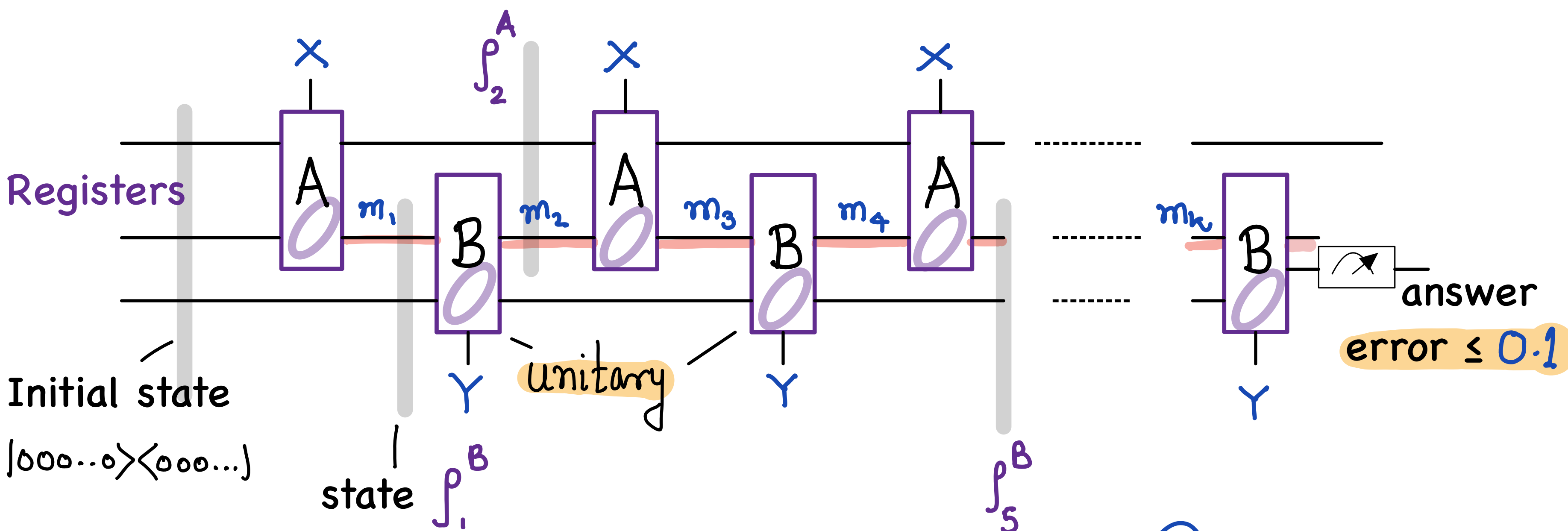
von Neumann entropy $S(\rho) = -\mathrm{Tr}\,\rho \log_2 \rho$

Mutual information $I[X:\rho] = S(\rho) - \mathop{E}_{X}[S(\rho_x)]$

(Classical $H[Y] - H[Y|X]$)



There is trouble in paradise



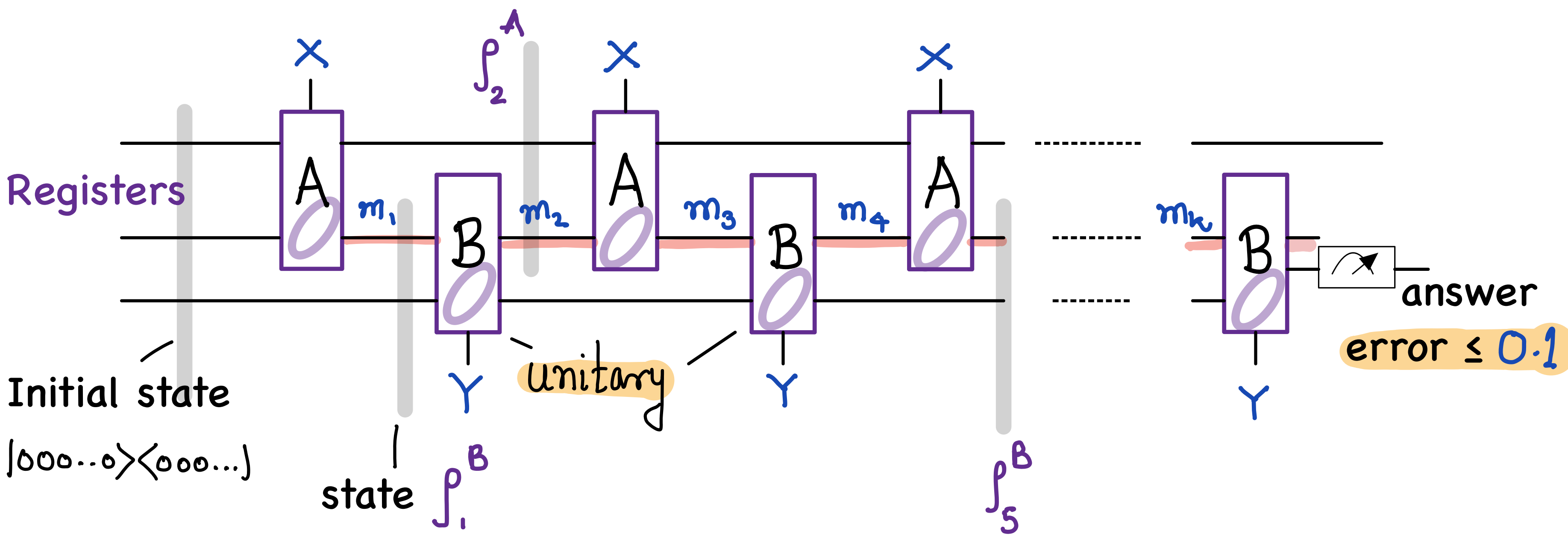
What is a transcript?

(m_1, m_2)

We cannot talk about all the messages together!

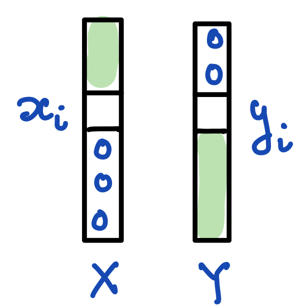
Old messages are destroyed when new messages are generated.

Paradise (partly) regained?

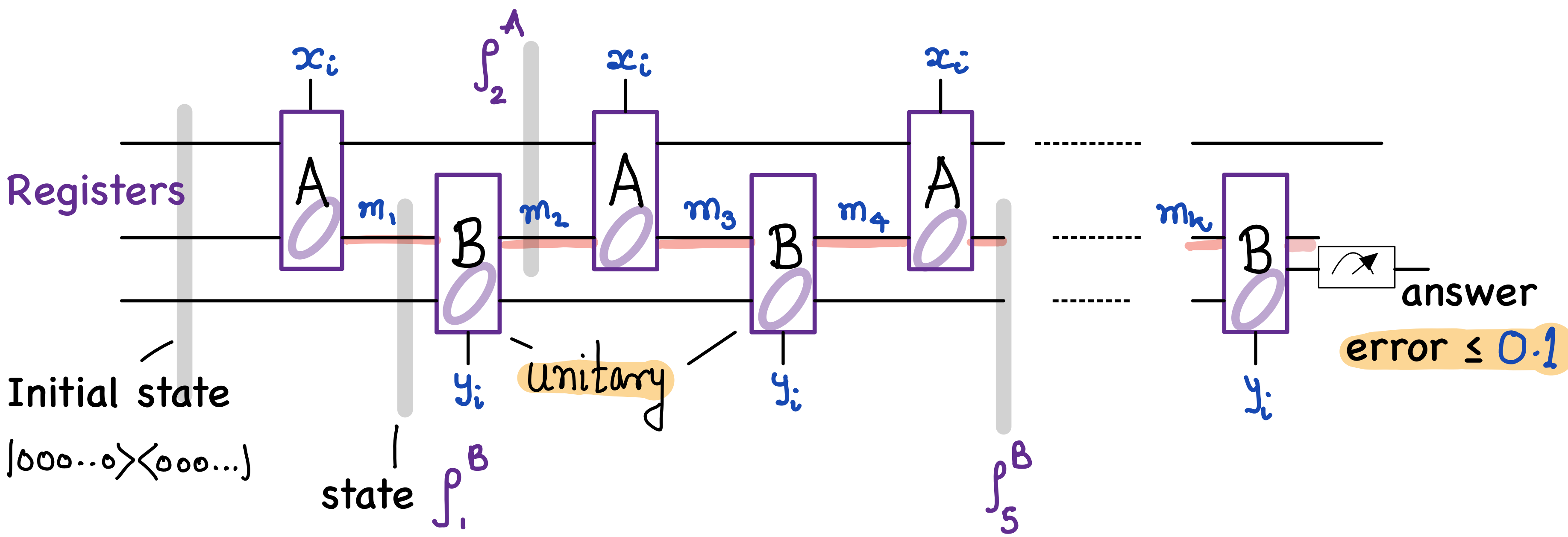


We work with: $$\sum_i I[X:\rho_i^B] + \sum_j I[Y:\rho_j^A]$$

As before, if the total communication is small, then the protocol must neglect some coordinate.



Paradise (partly) regained?



The messages maybe long and many but they do not carry much information about (x_i, y_i).

↓

In a k-round quantum protocol, the parties must exchange at least n/k^2 qubits.

Better bounds are known.

Summary

- Shannon entropy and counting
- Entropy and the number of typical sequences
- Communication complexity of Boolean functions
- Quantum communication and von Neumann entropy

Introduction to quantum information

Abhishek Dhar
Jaikumar Radhakrishnan
Samuel Joseph

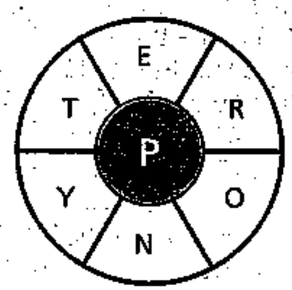
PHY 437.5 Spring 2023

… and more

Jumble

BULL'SEYE

How many words of four or more letters can you make from the letters shown? Every word must contain the central letter. There should be one seven-letter word. British English Dictionary is used as a reference.



14 Average; **16** Good; **18** Outstanding

It is entropy that counts!

Thank you!