# Information Complexity and Applications

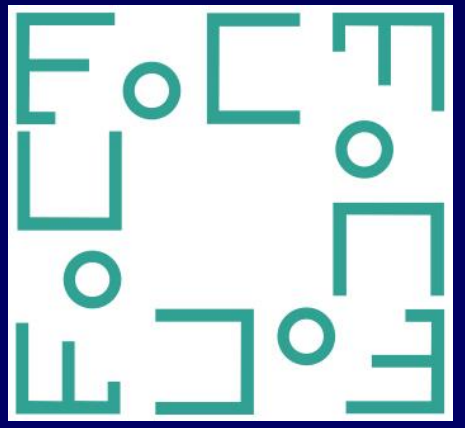
Mark Braverman

Princeton University and IAS

FoCM'17

July 17, 2017

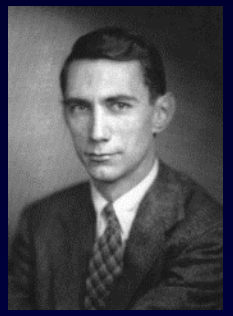




# Coding vs complexity: a tale of two theories

| Coding | Computational Complexity |
| --- | --- |
| Goal: data transmission | Goal: computation |
| Different channels | |
| "Big" questions are answered with theorems | |
| "$BSC_{1/3}$ can transmit $\approx 0.052$ trits per application" | |

# A key difference



- Information theory is a very effective language: fits many coding situations perfectly
- Shannon's channel coding theory is "continuous":
  - Turn the channel into a continuous resource;
  - Separate the communication channel from how it is used

# Theory of computation is "discrete"

- Von Neumann (~1948):

  *"…Thus formal logic is, by the nature of its approach, cut off from the best cultivated portions of mathematics, and forced onto the most difficult part of the mathematical terrain, into combinatorics.*

  *The theory of automata, … will have to share this unattractive property of formal logic. It will have to be, from the mathematical point of view, combinatorial rather than analytical."*

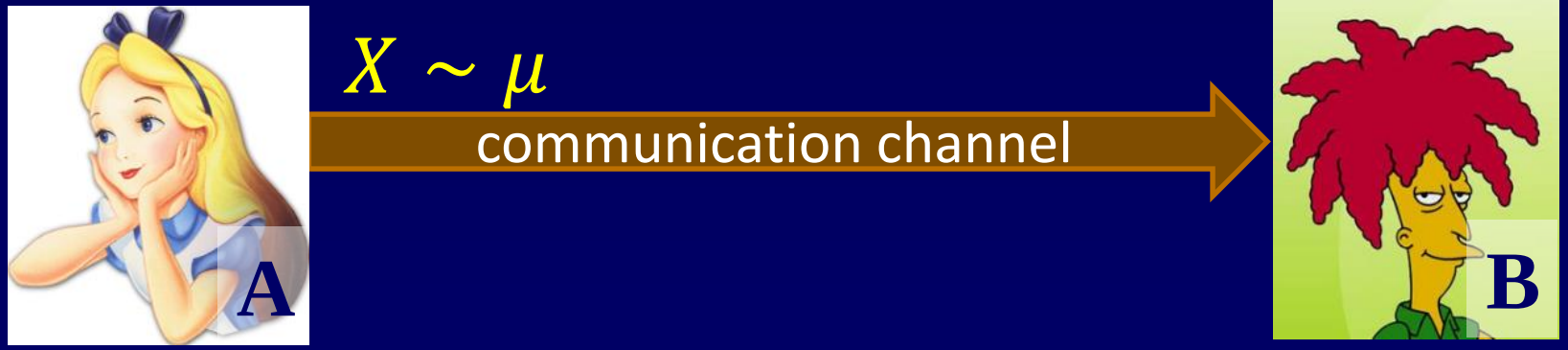
# Overview

- Today: Will discuss the extension of the information language to apply to problems in complexity theory.

# Background: Shannon's entropy

- Assume a lossless binary channel.
- A message $X$ is distributed according to some prior $\mu$.
- The *inherent* amount of bits it takes to transmit $X$ is given by its entropy

$$H(X) = \sum \mu[X = x] \log_2(1/\mu[X = x]).$$



# Shannon's Noiseless Coding Theorem

- The cost of communicating many copies of $X$ scales as $H(X)$.
- Shannon's source coding theorem:
  - Let $C_n(X)$ be the cost of transmitting $n$ independent copies of $X$. Then the *amortized transmission cost*

$$\lim_{n\to\infty} C_n(X)/n = H(X).$$

- *Operationalizes* $H(X)$.

# $H(X)$ is nicer than $C_n(X)$

- Sending a uniform trit $T$ in $\{1,2,3\}$.
- Using the prefix-free encoding $\{0,10,11\}$ sending on trit $T_1$ costs $C_1 = 5/3 \approx 1.667$ bits.
- Sending two trits $(T_1T_2)$ costs $C_2 = \frac{29}{9}$ bits using the encoding $\{000,001,010,011,100,101,110,1110,1111\}$. The cost per trit is $29/18 \approx 1.611 < C_1$.
- $C_1 + C_1 \neq C_2$.

# $H(X)$ is nicer than $C_n(X)$

- $C_1 = \frac{15}{9}, C_2 = \frac{29}{9}$
- $C_1 + C_1 \neq C_2$.
- The entropy $H(T) = \log_2 3 \approx 1.585$.
- We have $H(T_1 T_2) = \log_2 9 = H(T_1) + H(T_2)$.
- $H(T)$ is additive over independent variables.
- $C_n = n \cdot \log_2 3 \pm o(n)$.

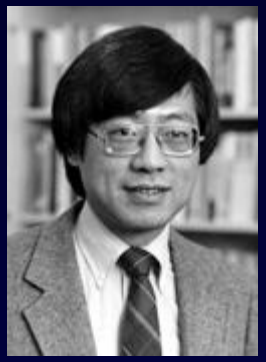
# Today

- We will discuss generalizing information and coding theory to interactive computation scenarios:

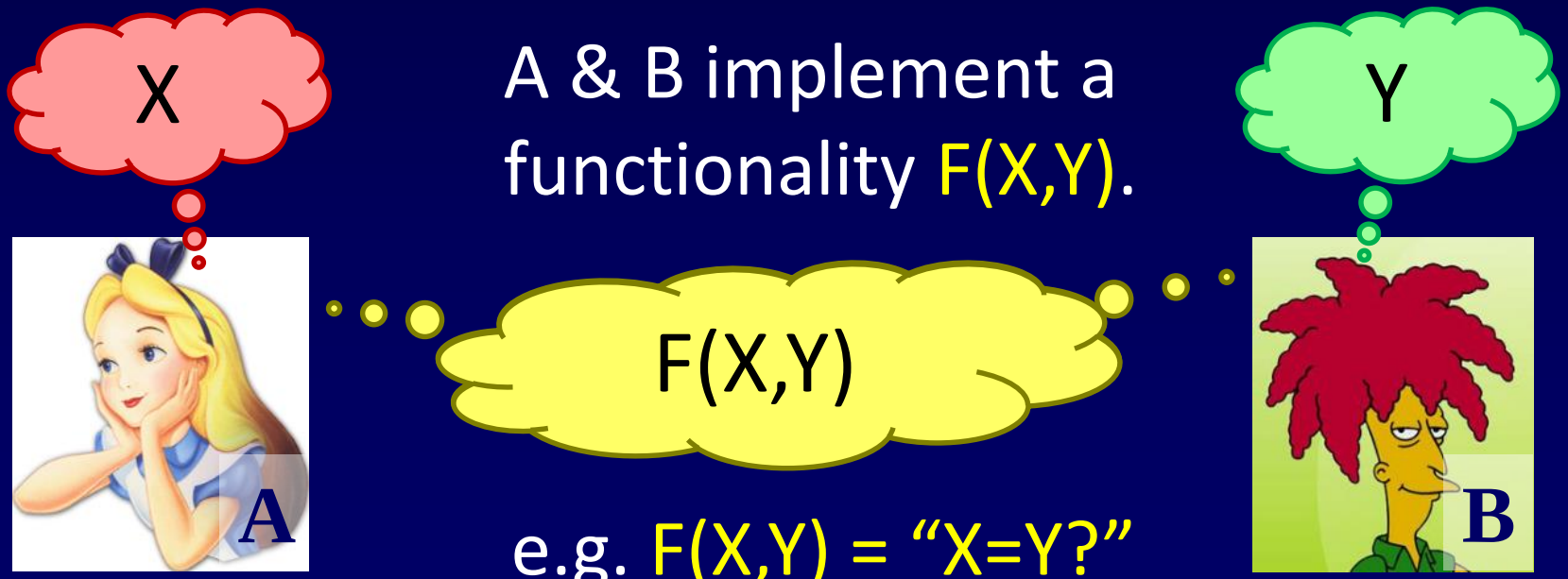
*"using interaction over a channel to solve a computational problem"*

- In Computer Science, the amount of communication needed to solve a problem is studied by the area of *communication complexity.*

# Communication complexity [Yao'79]



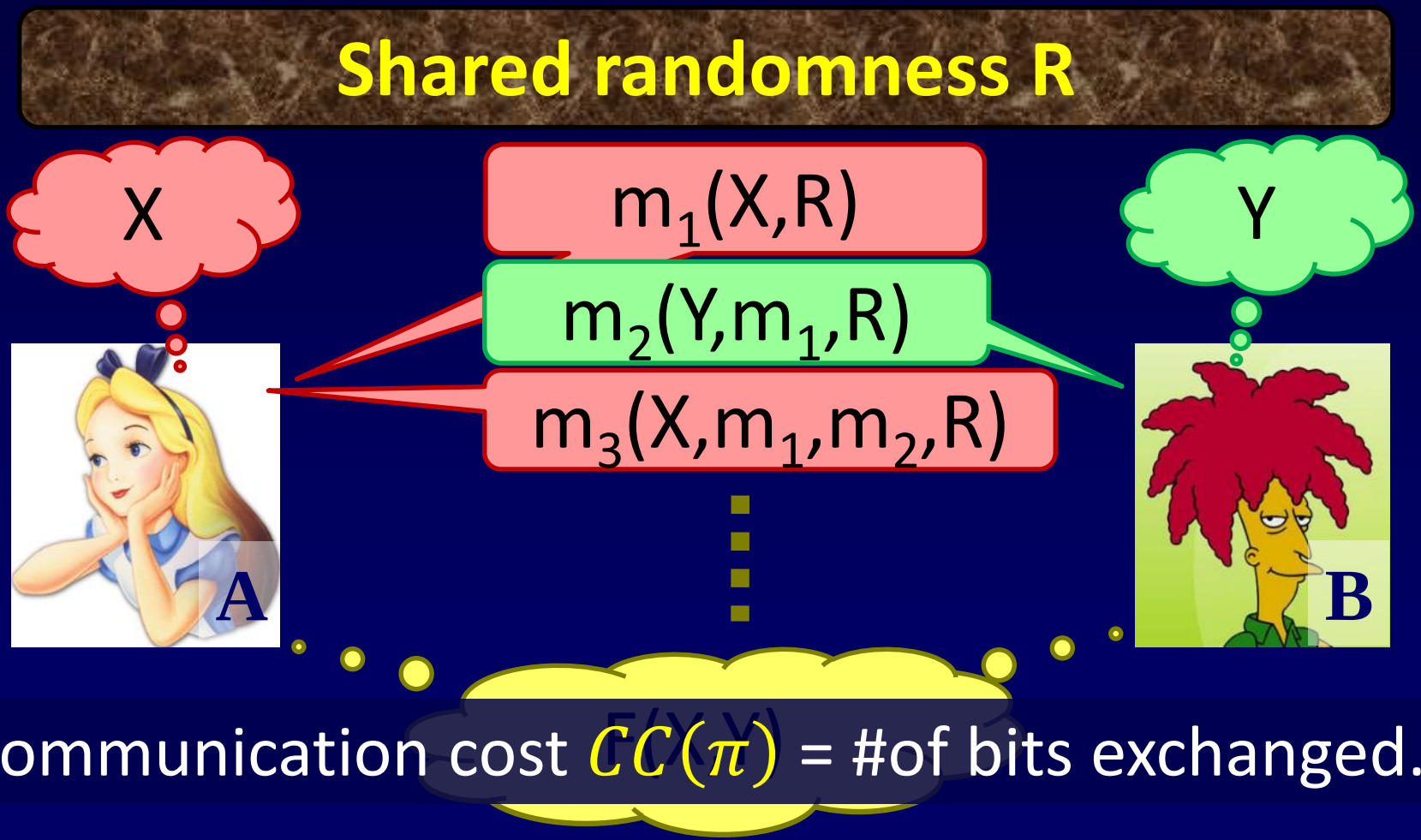
- Considers functionalities requiring *interactive* computation.
- Focus on the *two party* setting first.



# Communication complexity

Goal: implement a functionality $F(X,Y)$.

A protocol $\pi(X,Y)$ computing $F(X,Y)$:




Communication cost $CC(\pi)$ = #of bits exchanged.

# Communication complexity

- (Distributional) communication complexity with input distribution $\mu$ and error $\varepsilon$: $CC(F, \mu, \varepsilon)$. Error $\leq \varepsilon$ w.r.t. $\mu$:

$$CC(F, \mu, \varepsilon) := \min_{\pi : \mu(\pi(X,Y) \neq F(X,Y)) \leq \varepsilon} CC(\pi)$$

- (Randomized/worst-case) communication complexity: $CC(F, \varepsilon)$. Error $\leq \varepsilon$ on all inputs.
- Yao's minimax:

$$CC(F, \varepsilon) = \max_{\mu} CC(F, \mu, \varepsilon).$$

# A tool for *unconditional lower bounds* about computation

- Streaming;
- Data structures;
- Distributed computing;
- VLSI design lower bounds;
- Circuit complexity;
- One of two main tools for unconditional lower bounds.
- Connections to other problems in complexity theory (e.g. hardness amplification).

# Set disjointness and intersection

Alice and Bob each given a set $X \subseteq \{1, \ldots, n\}$, $Y \subseteq \{1, \ldots, n\}$ (can be viewed as vectors in $\{0,1\}^n$).

- *Intersection* $Int_n(X, Y) = X \cap Y$.
- *Disjointness* $Disj_n(X, Y) = 1$ *if* $X \cap Y = \emptyset$, *and* $0$ *otherwise*
- *A non-trivial theorem* [Kalyanasundaram-Schnitger'87, Razborov'92]: $CC(Disj_n, 1/4) = \Omega(n)$.
- *Exercise: Solve* $Disj_n$ *with error* $\to 0$ *(say,* $1/n$*) in* $0.9n$ *bits of communication. Can you do* $0.6n$*?* $0.4n$*?*

# Direct sum

- $Int_n$ is just $n$ times $2$-bit $AND$.
- $\neg Disj_n$ is a disjunction of $2$-bit $AND$s.
- What is the connection between the communication cost of one $AND$ and the communication cost of $n$ $AND$s?
- Understanding the connection between the hardness of a problem and the hardness of its pieces.
- A natural approach to lower bounds.

# How does CC scale with copies?

- $CC(F^n, \mu^n, \varepsilon)/n \to ?\ CC(F, \mu, \varepsilon)?$

Recall:

- $\lim_{n\to\infty} C_n(X)/n = H(X)$
- **Information complexity** is the corresponding scaling limit for $CC(F^n, \mu^n, \varepsilon)/n$.
- Helps understand problems composed of smaller problems.

# Interactive information complexity

- Information complexity ::

  communication complexity

as

- Shannon's entropy ::

  transmission cost

# Information theory in two slides

- For two (potentially correlated) variables $X, Y$, the *conditional entropy* of $X$ given $Y$ is the *amount of uncertainty left in* $X$ *given* $Y$:

$$H(X|Y) \coloneqq E_{y\sim Y}\mathrm{H}[\mathrm{X}|\mathrm{Y}=\mathrm{y}].$$

- One can show $H(XY) = H(Y) + H(X|Y)$.
- This important fact is knows as the *chain rule*.
- If $X \perp Y$, then

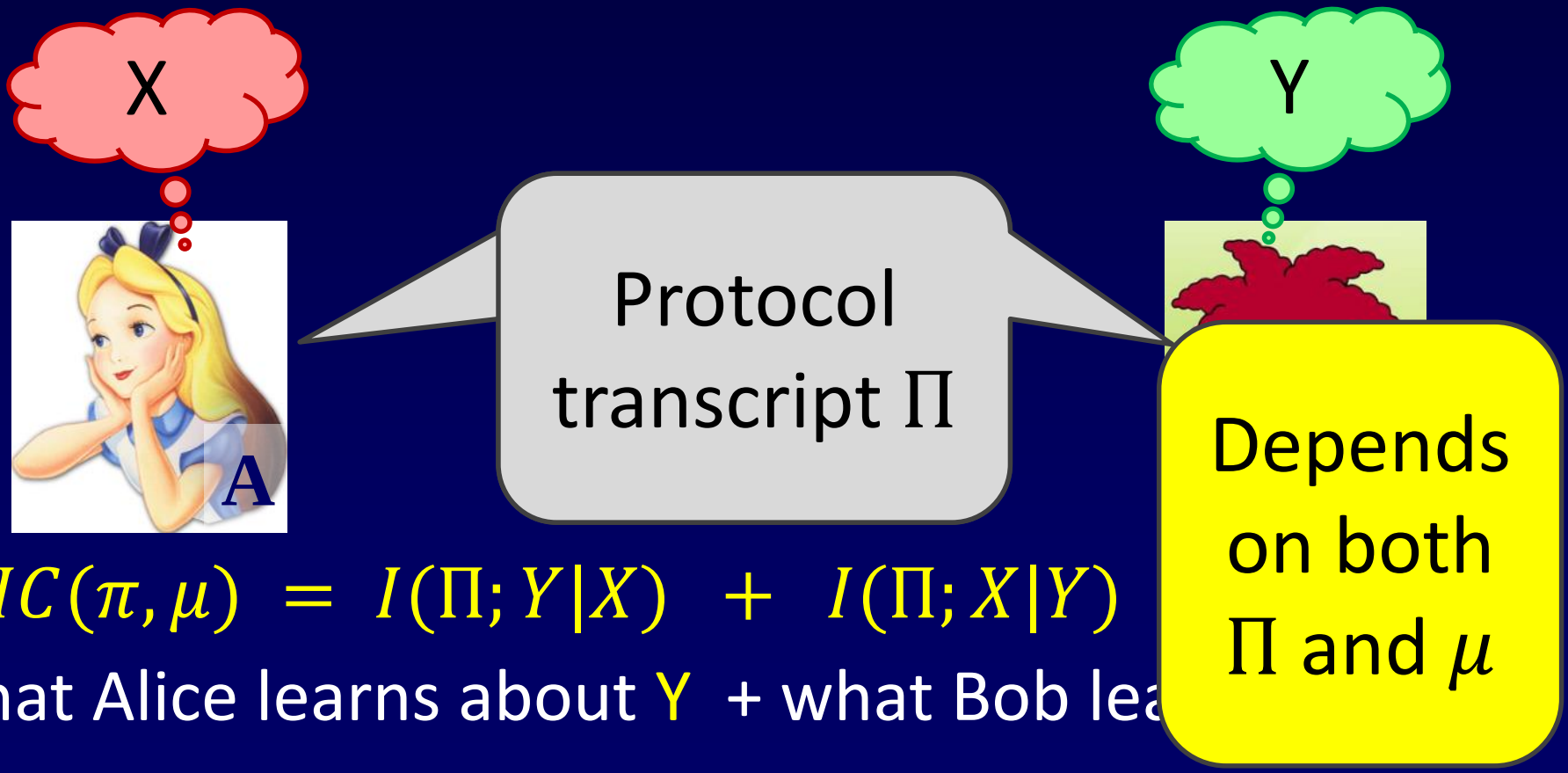
$$H(XY) = H(X) + H(Y|X) = H(X) + H(Y).$$

# Mutual information

- The mutual information is defined as
$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$
- "How much knowing $X$ reduce the uncertainty of $Y$?"
- Conditional mutual information:
$$I(X;Y|Z) \coloneqq H(X|Z) - H(X|YZ)$$
- Simple intuitive interpretation.

# The information cost of a protocol

- Prior distribution: $(X, Y) \sim \mu$.

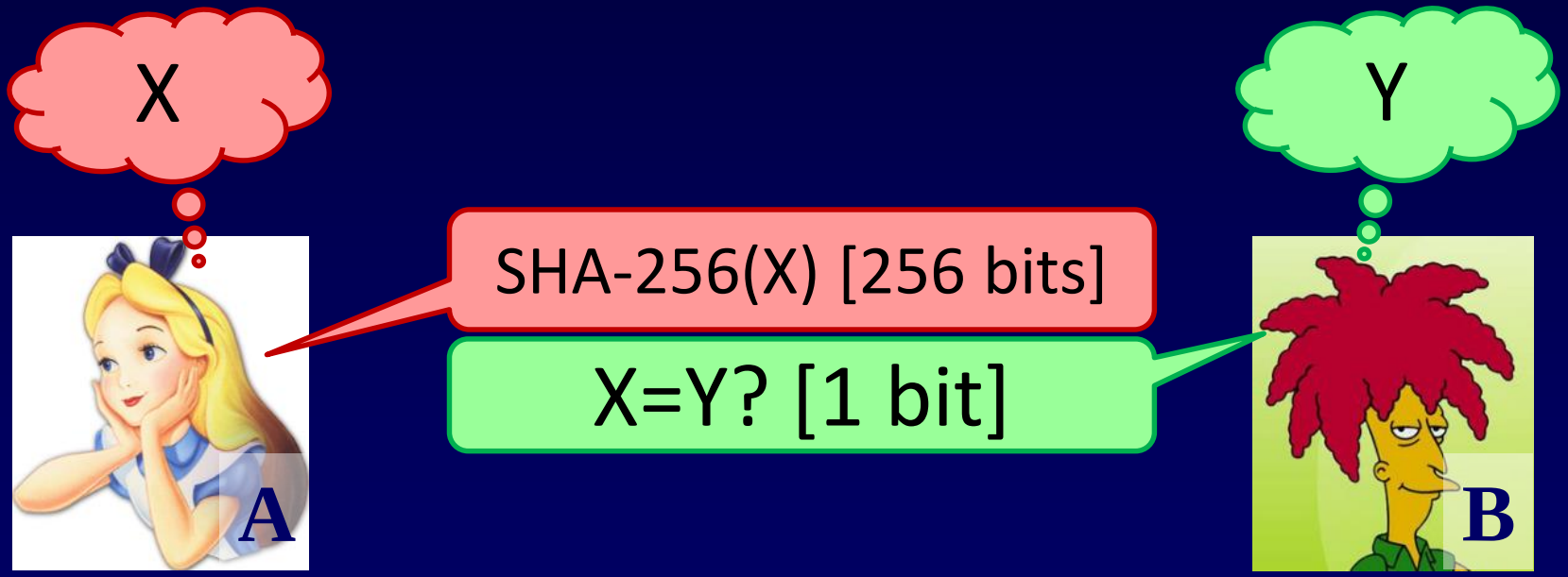



$$IC(\pi, \mu) = I(\Pi; Y|X) + I(\Pi; X|Y)$$

what Alice learns about Y + what Bob lea

# Example

- $F$ is "$X = Y?$".
- $\mu$ is a distribution where $X = Y$ w.p. ½ and $(X, Y)$ are random w.p. ½.




$$IC(\pi, \mu) = I(\Pi; Y|X) + I(\Pi; X|Y) \approx 1 + 129 = 130 \text{ bits}$$

what Alice learns about Y + what Bob learns about X

# The information complexity of a problem

- Communication complexity:

$$CC(F,\mu,\varepsilon) := \min_{\substack{\pi \text{ computes} \\ F \text{ with error} \leq \varepsilon}} CC(\pi).$$

- Analogously: Needed!

$$IC(F,\mu,\varepsilon) := \inf_{\substack{\pi \text{ computes} \\ F \text{ with error} \leq \varepsilon}} IC(\pi,\mu).$$

- (Easy) fact: $IC(F,\mu,\varepsilon) \leq CC(F,\mu,\varepsilon)$.

# Information = amortized communication

- Recall: $\lim_{n\to\infty} C_n(X)/n = H(X)$

*Theorem: [B.-Rao'11]*

- $\lim_{n\to\infty} CC(F^n, \mu^n, \varepsilon)/n = IC(F, \mu, \varepsilon)$.
- Corollary:

$$\lim_{n\to\infty} CC(Int_n, 0^+)/n = IC(AND, 0)$$

# The two-bit AND

- Alice and Bob each have a bit $X, Y \in \{0,1\}$ distributed according to some $\mu$ on $\{0,1\}^2$.
- Want to compute $X \wedge Y$, while revealing to each other as little as possible to each others' inputs (w.r.t. the worst $\mu$).
- Answer $IC(AND, 0)$ is a number between $1$ and $2$.

# The two-bit AND

Results [B.-Garg-Pankratov-Weinstein'13]:

- $IC(AND, 0) \approx 1.4922$ bits.
- Find the value of $IC(AND, \mu, 0)$ for all priors $\mu$ and exhibit the information-theoretically optimal protocol for computing the $AND$ of two bits.
- Studying $IC(AND, \mu, 0)$ as a function $\mathbb{R}^{+4}/\mathbb{R}^{+} \to \mathbb{R}^{+}$ is a functional minimization problem subject to a family of constraints (cf. construction of harmonic functions).

# The two-bit AND

- Studying $IC(AND, \mu, 0)$ as a function $\mathbb{R}^{+4}/\mathbb{R}^{+} \to \mathbb{R}^{+}$ is a functional minimization problem subject to a family of constraints (cf. construction of harmonic functions).
- We adopt a "guess and verify" strategy, although the general question of computing the information complexity of a function from its truth table is a very interesting one.

# The optimal protocol for AND

$X \in \{0,1\}$


If X=1, A=1

If X=0, A=$U_{[0,1]}$

1

0

$Y \in \{0,1\}$


If Y=1, B=1

If Y=0, B=$U_{[0,1]}$

"Raise your hand when your number is reached"

$X \in \{0,1\}$


If X=1, A=1

If X=0, A=$U_{[0,1]}$

1

0

$Y \in \{0,1\}$


If Y=1, B=1

If Y=0, B=$U_{[0,1]}$

# Corollary: communication complexity of intersection

- **Corollary**:

$$\lim_{\varepsilon \to 0} CC(Int_n, \varepsilon) \approx 1.4922 \cdot n \pm o(n).$$

- Specifically, e.g.

$$CC\left(Int_n, \frac{1}{n}\right) \approx 1.4922 \cdot n \pm o(n).$$

- Note: Require $\omega(1)$ rounds of interaction.

  Using $r$ rounds results in $+\Theta\left(\frac{n}{r^2}\right)$ cost!

# Communication complexity of Disjointness

- With some additional work, obtain a tight bound on the communication complexity of $Disj_n$ with tiny error:

$$\lim_{\varepsilon \to 0} CC(Disj_n, \varepsilon) = C_{DISJ} \cdot n \pm o(n),$$

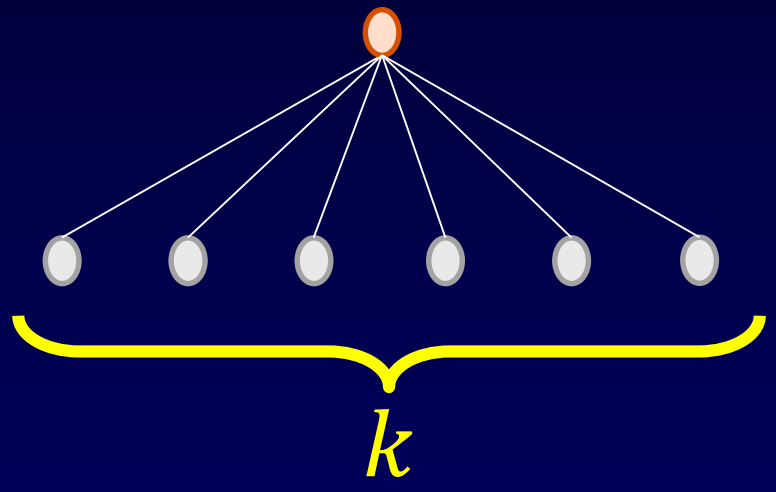
where

$$C_{DISJ} := \max_{\mu : \mu(1,1)=0} IC(AND, \mu, 0) \approx 0.4827 \ldots$$

- Intuition: $Disj_n$ is an $n$-wise repetition of $AND$, where the probability of a $(1,1)$ is very low ($\ll 1$).

# Beyond two parties

- Disjointness in the coordinator model [B.-Ellen-Oshman-Pitassi- Vaikuntanathan'13].
- $k$ players.
- Each $p_i$ holding a subset $S_i \subset \{1, \ldots, n\}$
- Want to decide whether the intersection $\bigcap S_i$ is empty.



# $Disj$ in the coordinator model

- $k$ players, input length $n$.
- Naïve protocol: $O(n \cdot k)$ communication.
- Turns out to be asymptotically optimal!
- The argument uses information complexity.
  - The hard part is to design the hard distribution and the "right" information cost measure.

# The “hard” distribution

- $S_i \subset \{1, \ldots, n\}$. Want to decide whether the intersection $\bigcap S_i$ is empty.
- Should have very few (close to $0$) intersections.

# The “hard” distribution

Coordinator keeps querying players until she finds a $0$: $\sim O(n)$ communication

- Attempt #1:
  - Plant many $0$’s (e.g. $50\%$):

$$k \left\{ \begin{matrix} 1 & 0 & 0 & \dots & 1 & 0 & 1 \\ 1 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 1 & 1 & \dots & 0 & 0 & 1 \\ 1 & 0 & 0 & \dots & 1 & 0 & 1 \end{matrix} \right.$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{n}$$

# The "hard" distribution

Each player sends its 0's: still $O(n \log n)$ communication

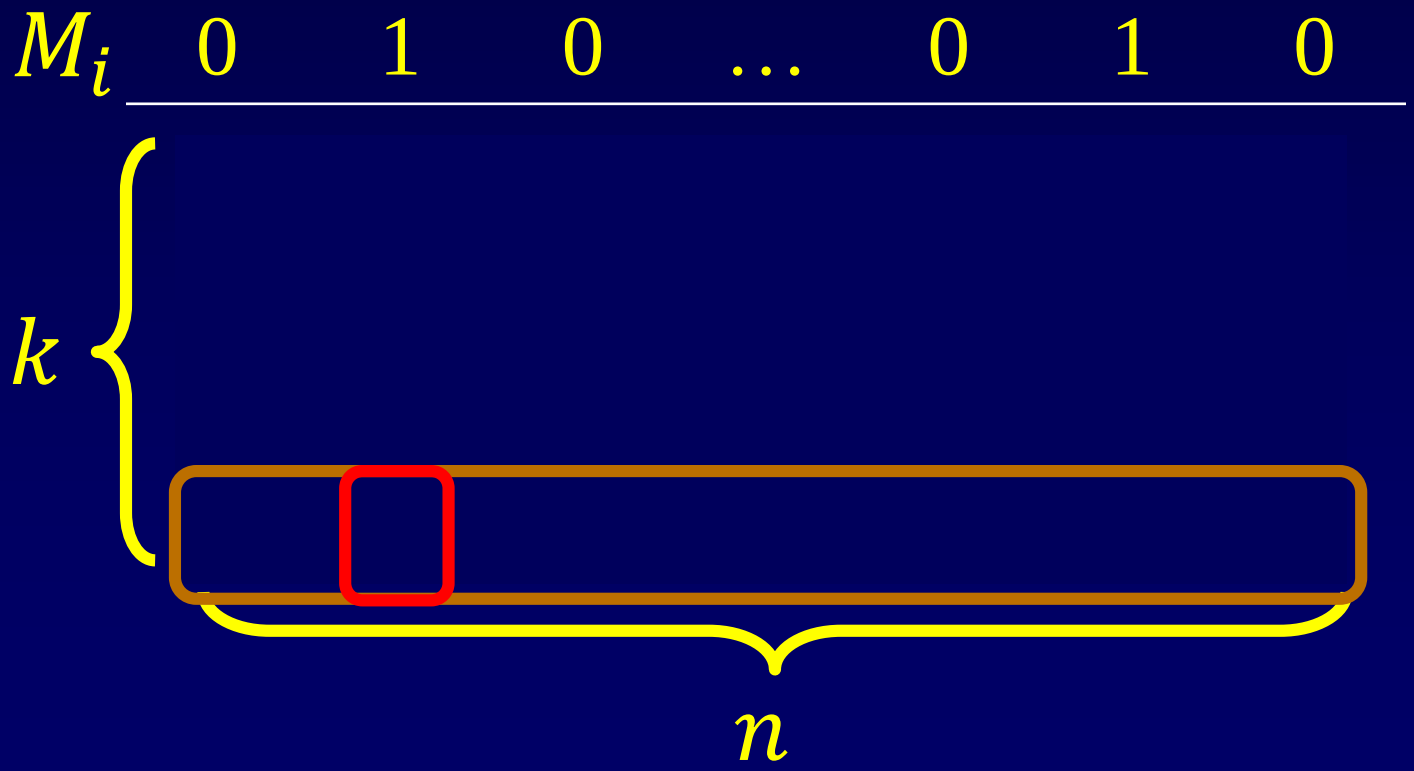
- Attempt #2:
  - Plant one zero in each coordinate

$$k \left\{ \begin{matrix} 1 & 0 & 0 & \dots & 1 & 1 & 1 \\ 1 & 1 & 1 & \dots & 0 & 1 & 0 \\ 0 & 1 & 1 & \dots & 1 & 1 & 1 \\ 1 & 1 & 1 & \dots & 1 & 0 & 1 \end{matrix} \right.$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{n}$$

# The “hard” distribution

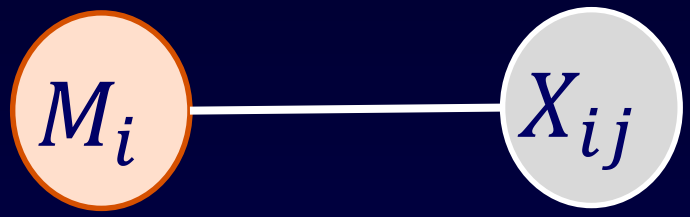
- “Mix” the two attempts.
  - Each coordinate has an RV $M_i \sim B_{1/3}$.
  - If $M_i = 0$, plant many $0$’s (e.g. $50\%$).
  - If $M_i = 1$, plant a single $0$.



# The information cost notion

- Assume the coordinator knows the $M_i$'s, player $j$ knows the $X_{ij}$.
- The information cost:
  - what the coordinator learns about the $X_{ij}$'s+
  - what each player learns about the $M_i$'s
- Proving that the sum total is $\Omega(n \cdot k)$ requires some work, but the hardest part are the definitions.

# Intuition for hardness




- Focus on a single $(i, j)$ pair: $i$'th coordinate, $j$'th player.
- $(M_i, X_{ij})$ are equally likely to be $(0,0)$, $(0,1)$ and $(1,1)$
- If $\mathrm{M_i} = 1$, then the coordinator *needs* to know $\mathrm{X_{ij}}$ (which is almost certainly $1$ in this case).
- Either $P_i$ will learn about $M_i$, or will reveal too much about $X_{ij}$ when $M_i = 0$.

# Multiparty information complexity

- We don’t have a multiparty information complexity theory for general distributions.
- There is a fair chance that the difficulty is conceptual.
- One key difference between 2 players and 3+ players is the existence of secure multiparty computation.

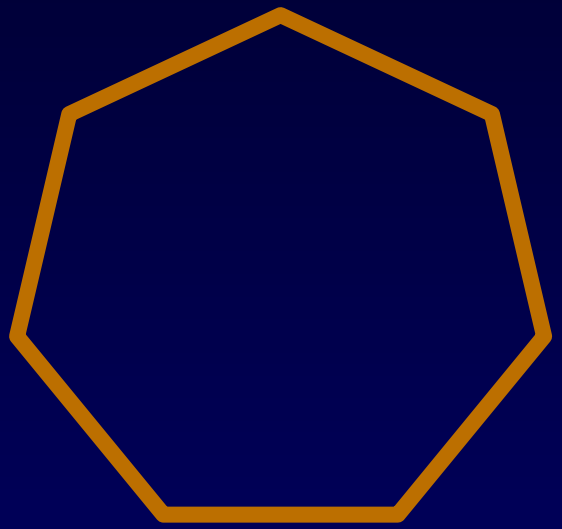
# Beyond communication

- Many applications to other interactive communication regimes:
  - Distributed joint computation & estimation;
  - Streaming;
  - Noisy coding…
- We will briefly discuss a non-communication application: two prover games.

# Two-prover games

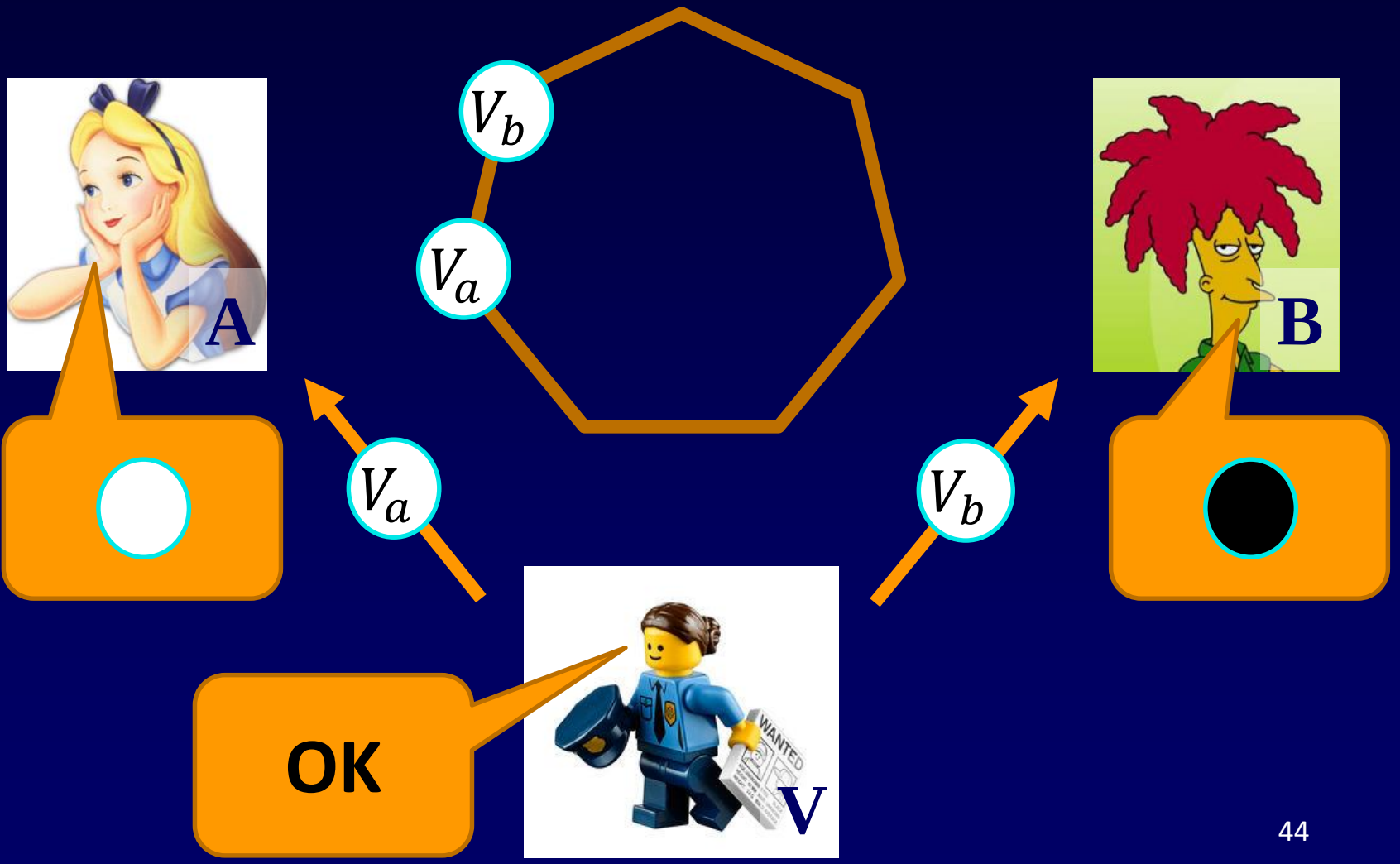
- Closely connected to hardness of approximation:
  - Probabilistically Checkable Proofs and the Unique Games Conjecture.
- A nice way of looking at constraint satisfaction problems.

# The odd cycle game

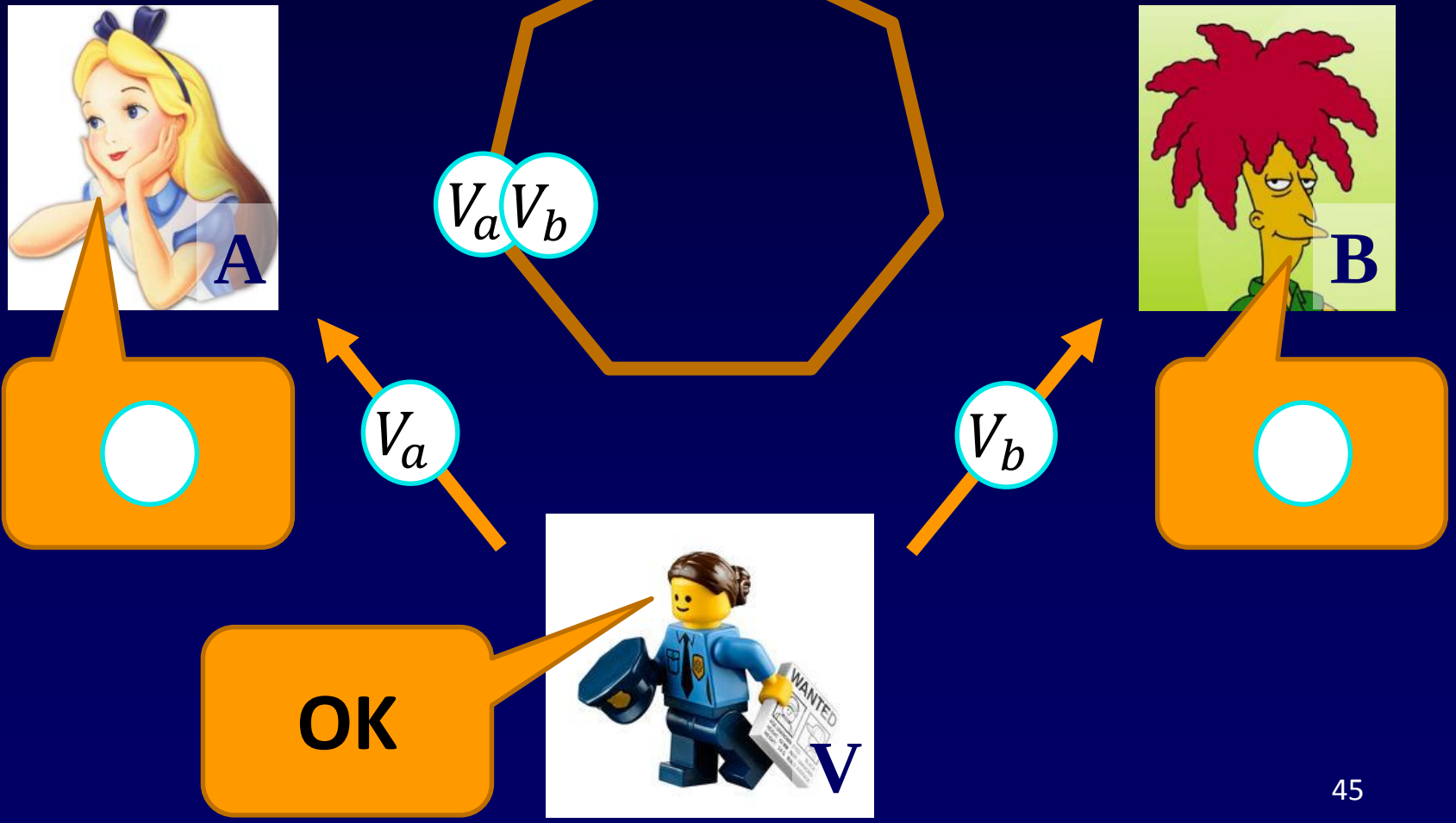




- Alice and Bob want to convince Victoria that the 7-cycle is 2-colorable.
- Asks them to color the same or adjacent vertices. Accepts if consistent.

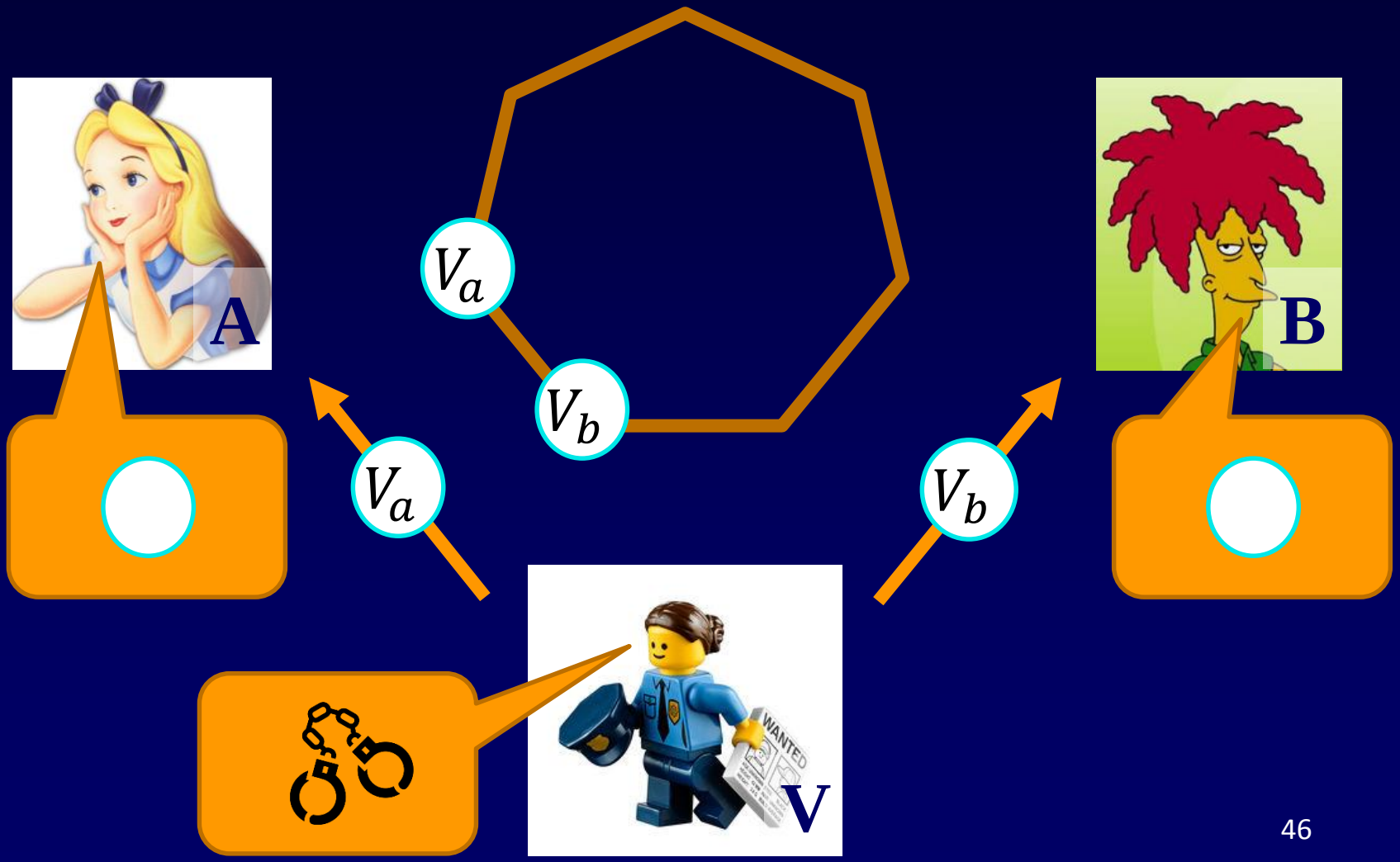
# The odd cycle game



# The odd cycle game



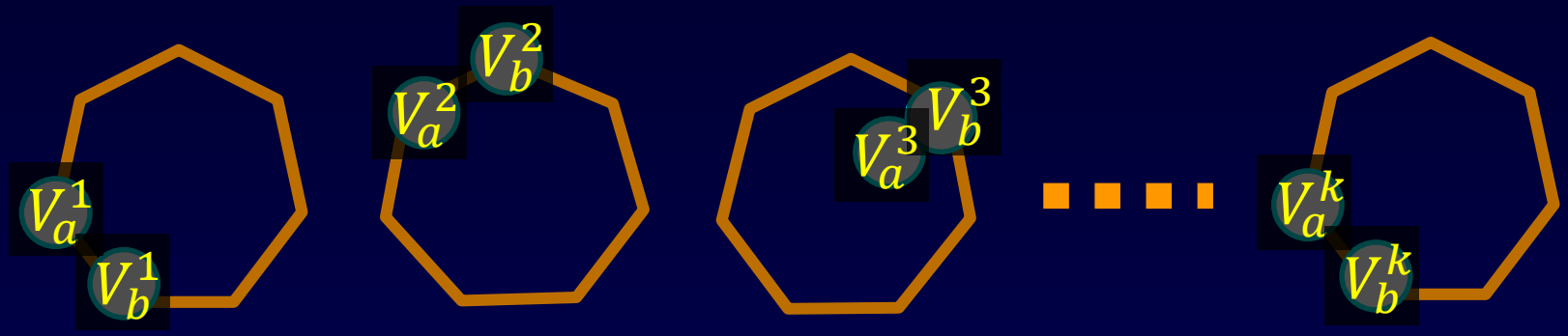
# The odd cycle game



# The odd cycle game

- An example of a “unique game”.
- If the cycle is even: Alice and Bob win with probability $1$.
- For odd cycle of length $m$, win with probability $p_1 = 1 - \frac{1}{2m}$.
- What about winning many copies of the game simultaneously?

# Simultaneous challenges




- Alice gets $V_a^1, V_a^2, V_a^3, \ldots, V_a^k$ and returns a vector of colors.
- Bob gets $V_b^1, V_b^2, V_b^3, \ldots, V_b^k$ and returns a vector of colors.
- Avoid jail if all color pairs are consistent.

# Parallel repetition

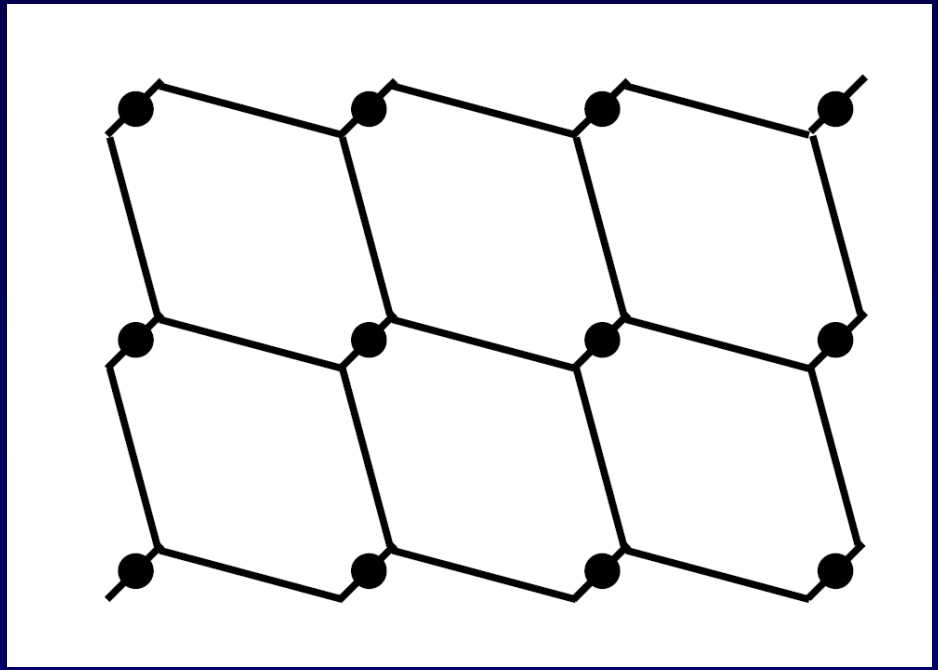
- Play $k = m^2$ copies.
- A naïve strategy: $\left(1 - \frac{1}{2m}\right)^{m^2} = e^{-\Theta(m)} \ll 1$
- Can one do better?

# Parallel repetition

- Play $k = m^2$ copies.
- A naïve strategy: $\left(1 - \frac{1}{2m}\right)^{m^2} = e^{-\Theta(m)} \ll 1$
- It turns out that one can win $m^2$ copies of the odd cycle game with a constant probability [Raz'08].
- Proof by exhibiting a strategy.

# Connection to foams

- Connected to "foams": tilings of $\mathbb{R}^d$ with a shape $A$ so that $A + \mathbb{Z}^d = \mathbb{R}^d$.
- What can the smallest surface area of A be?

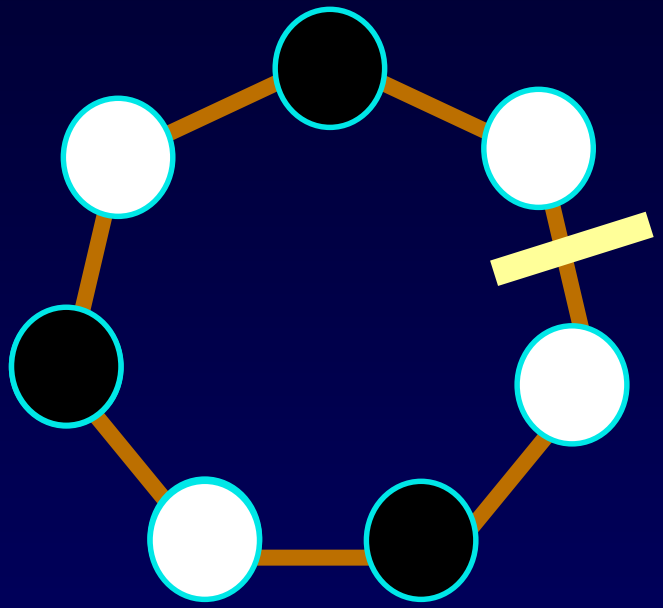


[Feige-Kindler-O'Donnell'07]

# Connection to foams

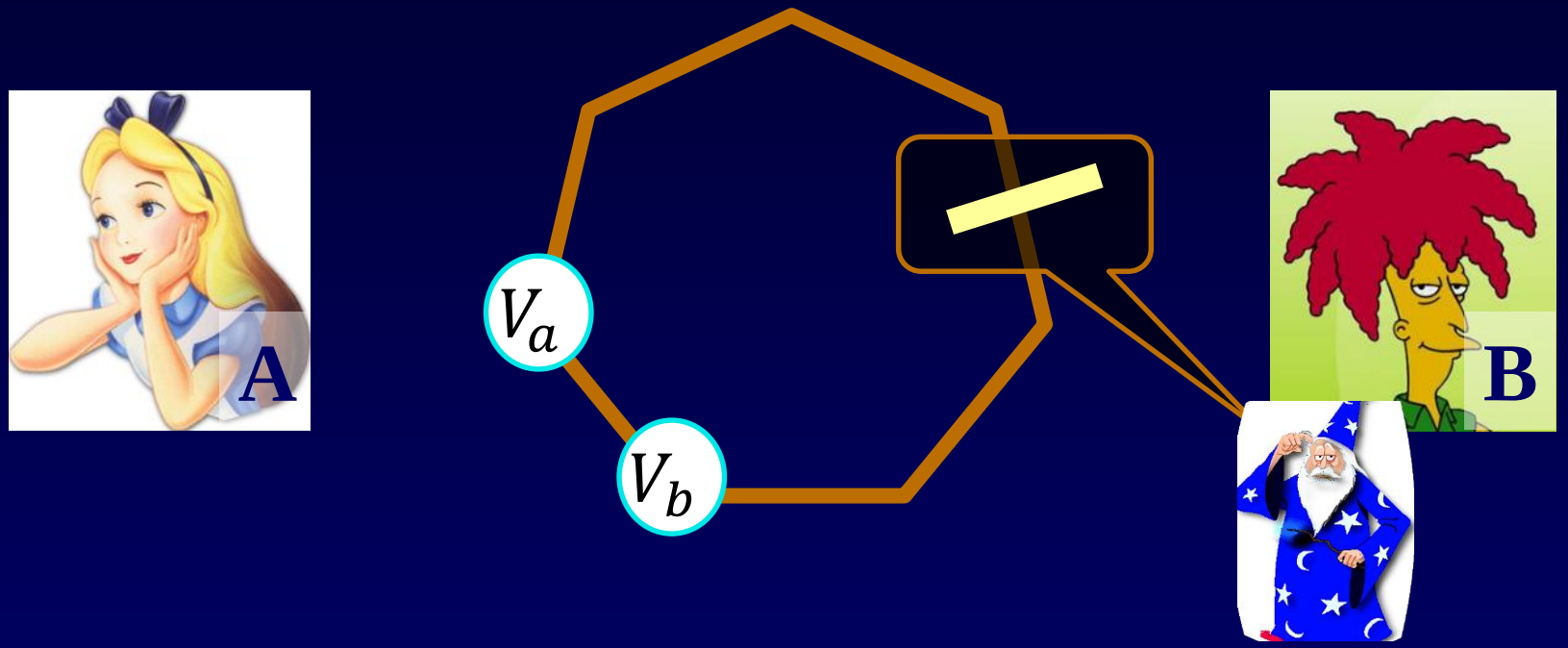
- Obvious upper bound: $O(d)$.
- Obvious lower bound (sphere of volume 1): $\Omega(\sqrt{d})$.
- [Feige-Kindler-O'Donnell'07]: Noticed a connection between the problems.
- [Kindler-O'Donnel-Rao-Wigderson'08]: a construction of foams of surface area $O(\sqrt{d})$ based on Raz's strategy.

# An information-theoretic view

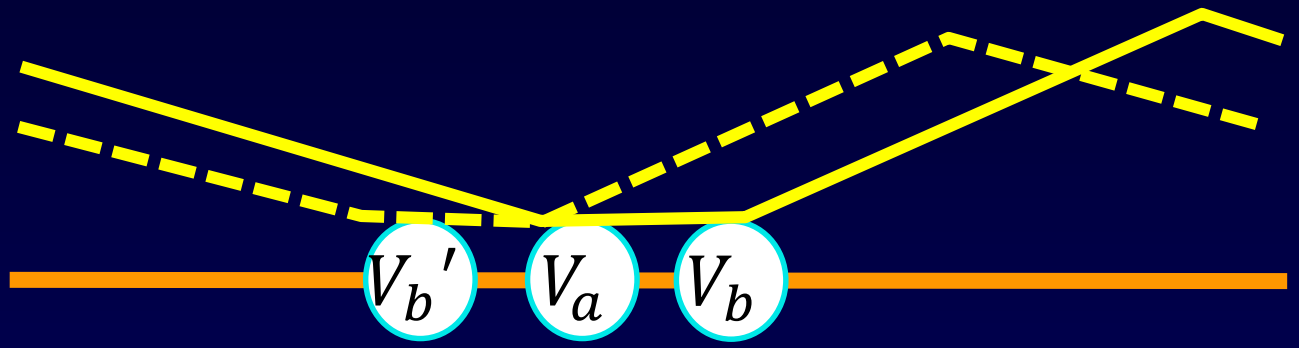

- “Advice” on where to cut the cycle wins the game with probability 1 if the cut does not pass through the challenge edge.

# An information-theoretic view



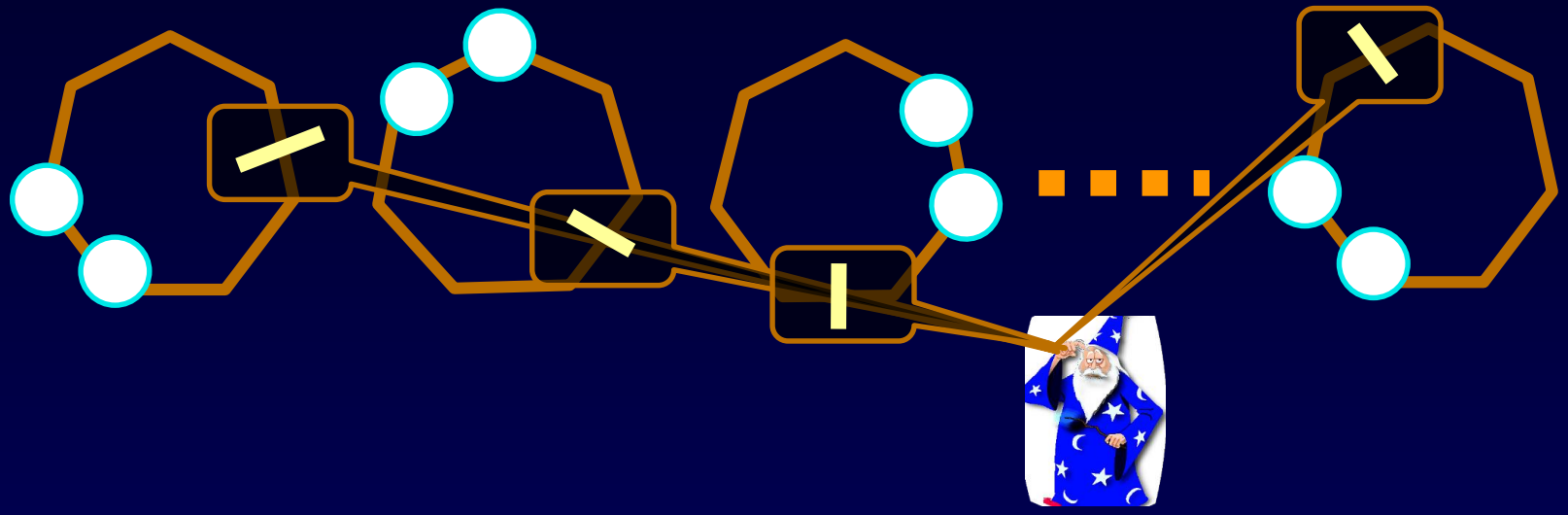

- Merlin can give such advice at "information cost" $O\left(\frac{1}{m^2}\right)$.

# The distribution




- KL-divergence between the two distributions is $\Theta\left(\frac{1}{m^2}\right)$
- Statistical distance is $\Theta\left(\frac{1}{m}\right)$

# Taking $m^2$ copies



- Total information revealed by Merlin:

$$m^2 \cdot O\left(\frac{1}{m^2}\right) = O(1).$$

- Can be simulated successfully with $O(1)$ communication, or with no communication with probability $\Omega(1)$.

# Parallel repetition

- Using similar intuition (but more technical work) can obtain a general tight parallel repetition problem in the "small value" regime. [B.-Garg'15]
- If one copy of a game $G$ has success probability $\delta < 1/2$, then $m$ copies have success probability $< \delta^{\Omega(m)}$ (*for "projection games"; for general games tight bound a bit more complicated)
- [Dinur-Steurer'14] obtained the result for projection games using spectral techniques.

# Challenges

- Information complexity beyond two communicating parties.
- Continuous measures of complexity.

# Thank You!