Basics of information theory and information complexity

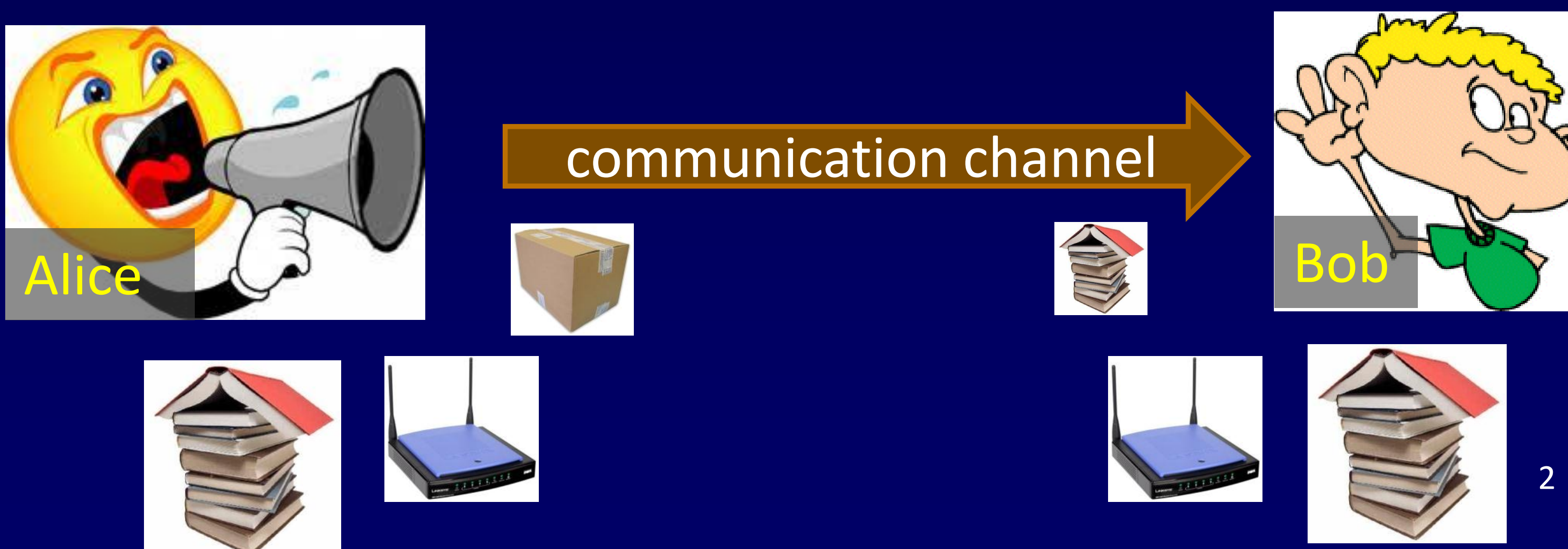
a tutorial

Mark Braverman
Princeton University

June 1, 2013

Part I: Information theory

- Information theory, in its modern format was introduced in the 1940s to study the problem of transmitting data over physical channels.



Quantifying "information"

- Information is measured in *bits.*
- The basic notion is Shannon's entropy.
- The entropy of a random variable is the (typical) number of bits needed to remove the uncertainty of the variable.
- For a discrete variable:

$$H(X) \coloneqq \sum \Pr[X = x] \log 1/\Pr[X = x]$$

Shannon's entropy

- Important examples and properties:
  - If $X = x$ is a constant, then $H(X) = 0$.
  - If X is uniform on a finite set S of possible values, then $H(X) = \log S$.
  - If X is supported on at most n values, then $H(\mathrm{X}) \leq \log n$.
  - If Y is a random variable determined by X, then $H(Y) \leq H(X)$.

Conditional entropy

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

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

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

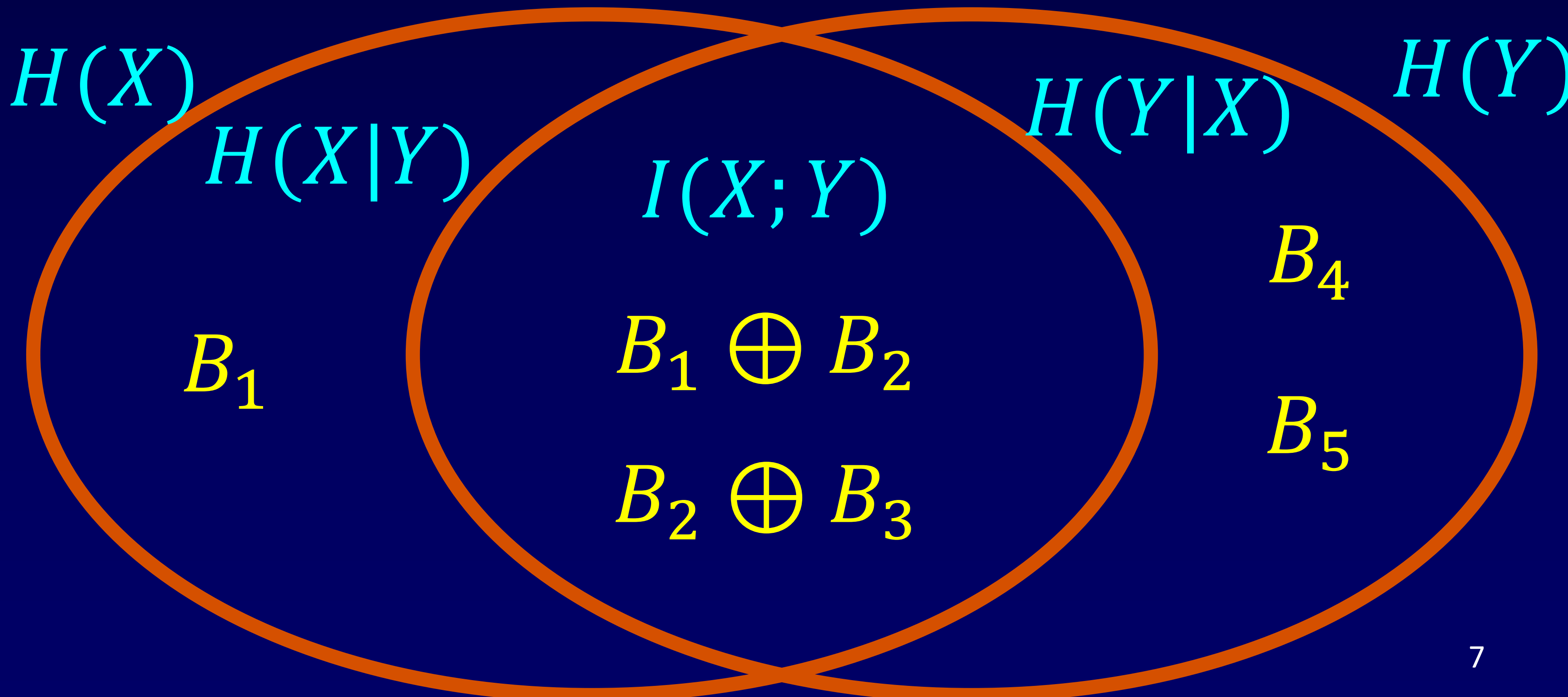
$$H(XY) = H(X) + H(Y|X) = H(X) + H(Y).$$

Example

- $X = B_1, B_2, B_3$
- $Y = (B_1 \oplus B_2), (B_2 \oplus B_4), (B_3 \oplus B_4), B_5$
- Where $B_1, B_2, B_3, B_4, B_5 \in_U \{0,1\}$.
- Then
  - $H(X) = 3; H(Y) = 4; H(XY) = 5;$
  - $H(X|Y) = 1 = H(XY) - H(Y);$
  - $H(Y|X) = 2 = H(XY) - H(X).$

Mutual information

- $X = B_1, B_2, B_3$
- $Y = (B_1 \oplus B_2), (B_2 \oplus B_4), (B_3 \oplus B_4), B_5$



Mutual information

- The mutual information is defined as

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

- "By how much does knowing X reduce the entropy of Y?"
- Always non-negative $I(X;Y) \geq 0$.
- Conditional mutual information:

$$I(X;Y|Z) \coloneqq H(X|Z) - H(X|YZ)$$

- Chain rule for mutual information:

$$I(XY;Z) = I(X;Z) + I(Y;Z|X)$$

- Simple intuitive interpretation.

Example – a biased coin

- A coin with ε-Heads or Tails bias is tossed several times.
- Let $B \in \{H, T\}$ be the bias, and suppose that *a-priori* both options are equally likely: $H(B) = 1$.
- How many tosses needed to find B?
- Let $T_1, \dots, T_k$ be a sequence of tosses.
- Start with $k = 2$.

What do we learn about B?

- $I(B;T_1T_2) = I(B;T_1) + I(B;T_2|T_1) =$
  $$I(B;T_1) + I(B\,T_1;T_2) - I(T_1;T_2)$$
  $$\leq I(B;T_1) + I(B\,T_1;T_2) =$$
  $$I(B;T_1) + I(B;T_2) + I(T_1;T_2|B)$$
  $$= I(B;T_1) + I(B;T_2) = 2 \cdot I(B;T_1).$$
- Similarly,
  $$I(B;T_1 \ldots T_k) \leq k \cdot I(B;T_1).$$
- To determine B with constant accuracy, need $0 < \mathrm{c} < I(B;T_1 \ldots T_k) \leq k \cdot I(B;T_1)$.
- $k = \Omega(1/I(B;T_1))$.

Kullback–Leibler (KL)-Divergence

- A distance metric between distributions on the same space.
- Plays a key role in information theory.

$$D(P \parallel Q) \coloneqq \sum_{x} P[x] \log \frac{P[x]}{Q[x]}.$$

- $D(P \parallel Q) \geq 0$, with equality when $P = Q$.
- Caution: $D(P \parallel Q) \neq D(Q \parallel P)$!

Properties of KL-divergence

- Connection to mutual information:

$$I(X;Y) = E_{y\sim Y} D(X_{Y=y} \parallel X).$$

- If $X \perp Y$, then $X_{Y=y} = X$, and both sides are 0.
- Pinsker's inequality:

$$|P - Q|_1 = O(\sqrt{D(P \parallel Q)}).$$

- Tight!

$$D(B_{1/2+\varepsilon} \parallel B_{1/2}) = \Theta(\varepsilon^2).$$

Back to the coin example

- $I(B;T_1) = E_{b\sim B} D(T_{1,B=b} \parallel T_1) =$
$$D\left(B_{\frac{1}{2}\pm\varepsilon} \parallel B_{\frac{1}{2}}\right) = \Theta(\varepsilon^2).$$
- $k = \Omega\left(\frac{1}{I(B;T_1)}\right) = \Omega(\frac{1}{\varepsilon^2}).$
- "Follow the information learned from the coin tosses"
- Can be done using combinatorics, but the information-theoretic language is more natural for expressing what's going on.

Back to communication

- The reason Information Theory is so important for communication is because information-theoretic quantities readily *operationalize.*
- Can attach operational meaning to Shannon's entropy: $H(X) \approx$ "the cost of transmitting X".
- Let $C(X)$ be the (expected) cost of transmitting a sample of X.

$H(X) = C(X)?$

- Not quite.
- Let trit $T \in_U \{1,2,3\}$.
- $C(T) = \frac{5}{3} \approx 1.67$.
- $H(T) = \log 3 \approx 1.58$.
- It is always the case that $C(X) \geq H(X)$.

1	0
2	10
3	11

But $H(X)$ and $C(X)$ are close

- Huffman's coding: $C(X) \leq H(X) + 1$.
- This is a *compression result*: "an uninformative message turned into a short one".
- Therefore: $H(X) \leq C(X) \leq H(X) + 1$.

Shannon's noiseless coding

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

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

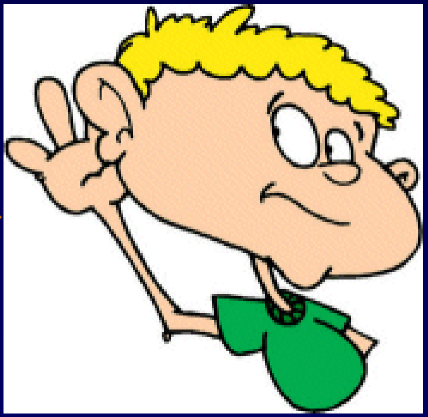
- This equation gives $H(X)$ operational meaning.

$H(X)$ *operationalized*

$X_1, \dots, X_n, \dots$

$H(X)$ per copy
to transmit X's

communication
channel



$H(X)$ is nicer than $C(X)$

- $H(X)$ is additive for independent variables.
- Let $T_1, T_2 \in_U \{1,2,3\}$ be independent trits.
- $H(T_1 T_2) = \log 9 = 2 \log 3.$
- $C(T_1 T_2) = \frac{29}{9} < C(T_1) + C(T_2) = 2 \times \frac{5}{3} = \frac{30}{9}.$
- Works well with concepts such as *channel capacity*.

“Proof” of Shannon’s noiseless coding

- $n \cdot H(X) = H(X^n) \leq C(X^n) \leq H(X^n) + 1.$

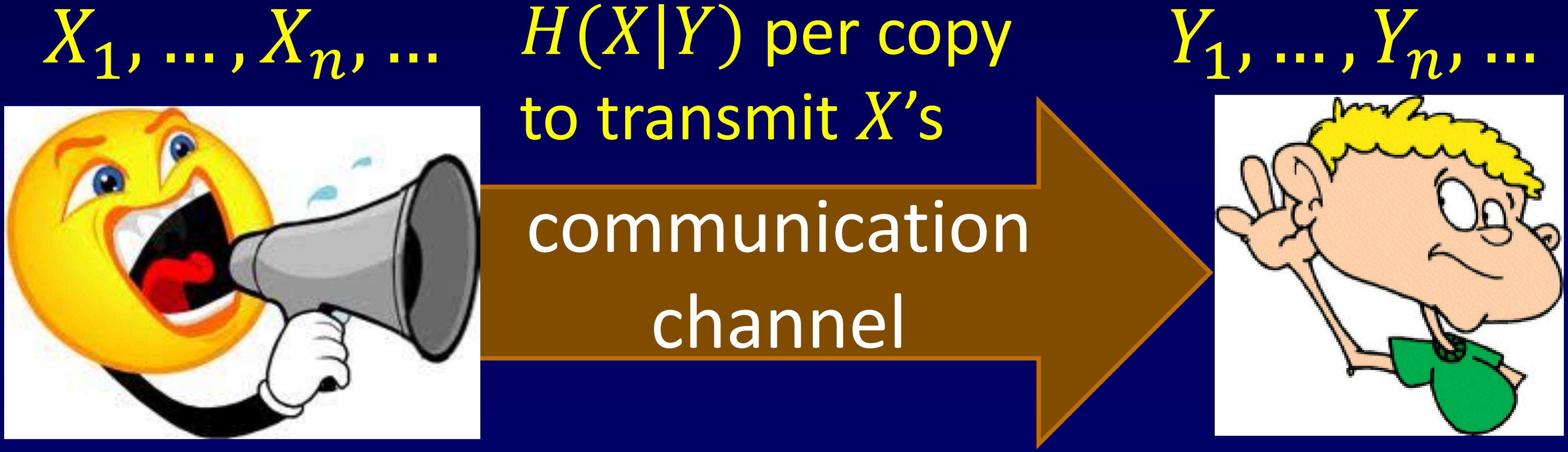
Additivity of entropy

Compression (Huffman)

- Therefore $\lim_{n \to \infty} C(X^n)/n = H(X)$.

Operationalizing other quantities

- Conditional entropy $H(X|Y)$:
- (cf. Slepian-Wolf Theorem).



Operationalizing other quantities

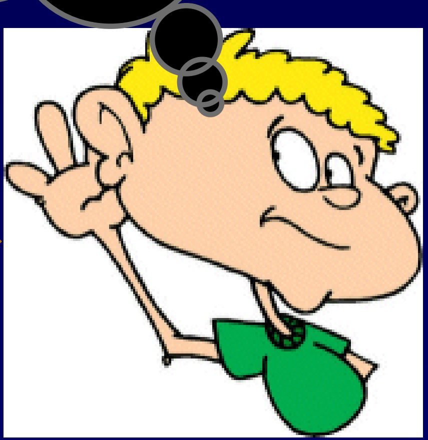
- Mutual information $I(X;Y)$:

$Y_1, \ldots, Y_n, \ldots$

$X_1, \ldots, X_n, \ldots$

$I(X;Y)$ per copy to *sample* Y's

communication channel



Information theory and entropy

- Allows us to formalize intuitive notions.
- Operationalized in the context of one-way transmission and related problems.
- Has nice properties (additivity, chain rule…)
- Next, we discuss extensions to more interesting communication scenarios.

Communication complexity

- Focus on the *two party* randomized setting.

Shared randomness R

X

Y

A & B implement a functionality $F(X,Y)$.

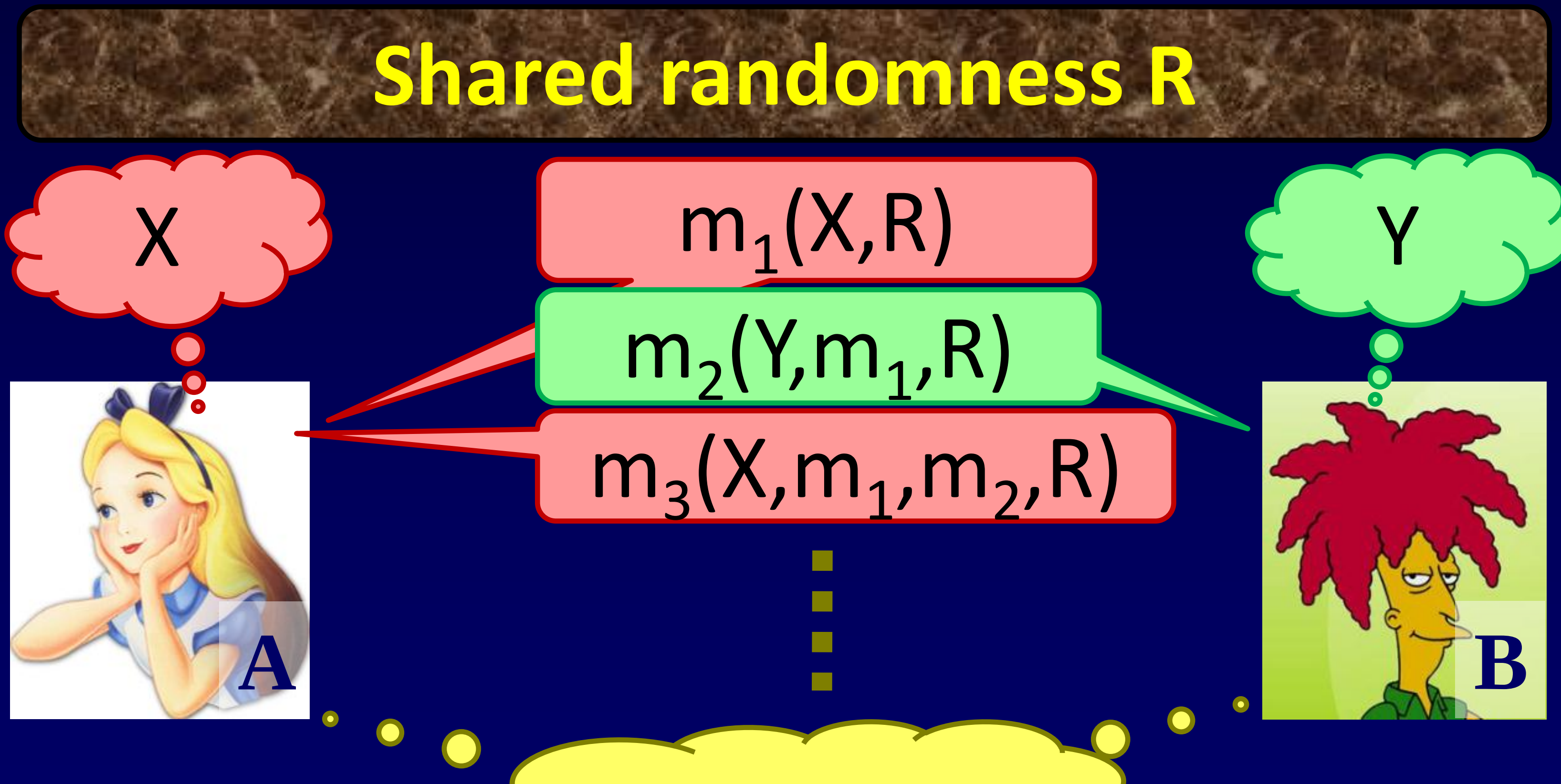
F(X,Y)

e.g. $F(X,Y) = \text{“}X = Y?\text{”}$

Communication complexity

<u>Goal</u>: implement a functionality $F(X,Y)$.

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



Communication cost = #of bits exchanged.

Communication complexity

- Numerous applications/potential applications (some will be discussed later today).
- Considerably more difficult to obtain lower bounds than transmission (still much easier than other models of computation!).

Communication complexity

- (Distributional) communication complexity with input distribution μ and error ε: $CC(F, \mu, \varepsilon)$. Error $\leq \varepsilon$ w.r.t. μ.
- (Randomized/worst-case) communication complexity: $CC(F, \varepsilon)$. Error $\leq \varepsilon$ on all inputs.
- Yao's minimax:

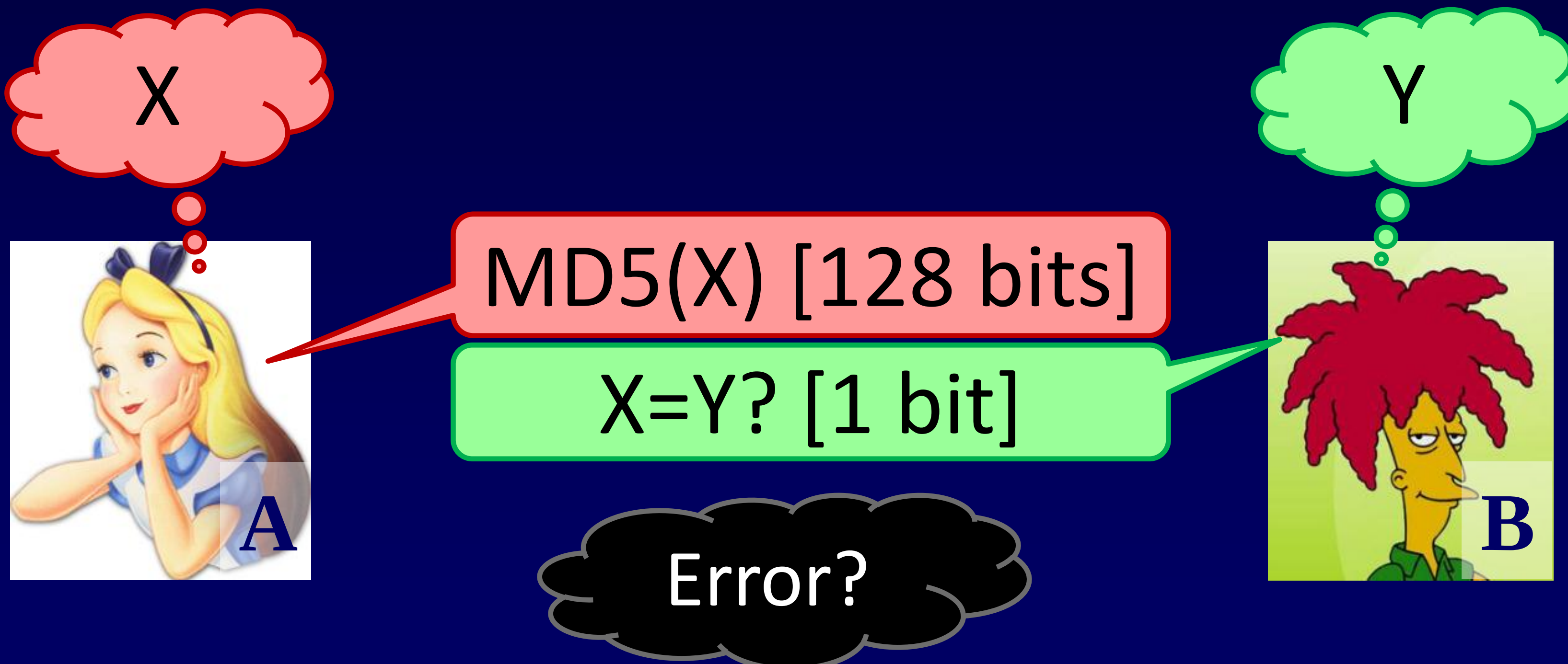
$$CC(F, \varepsilon) = \max_{\mu} CC(F, \mu, \varepsilon).$$

Examples

- $X, Y \in \{0,1\}^n$.
- Equality $EQ(X,Y) \coloneqq 1_{X=Y}$.
- $CC(EQ, \varepsilon) \approx \log\frac{1}{\varepsilon}$.
- $CC(EQ, 0) \approx n$.

Equality

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



- Shows that $CC(EQ, \mu, 2^{-129}) \leq 129$.

Examples

- $X, Y \in \{0,1\}^n$.
- Inner product $IP(X,Y) \coloneqq \sum_i X_i \cdot Y_i \pmod 2$.
- $CC(IP, 0) = n - o(n)$.

In fact, using information complexity:

- $CC(IP, \varepsilon) = n - o_\varepsilon(n)$.

Information complexity

- **Information complexity** $IC(F,\varepsilon)$::

  communication complexity $CC(F,\varepsilon)$

as

- Shannon's entropy $H(X)$ ::

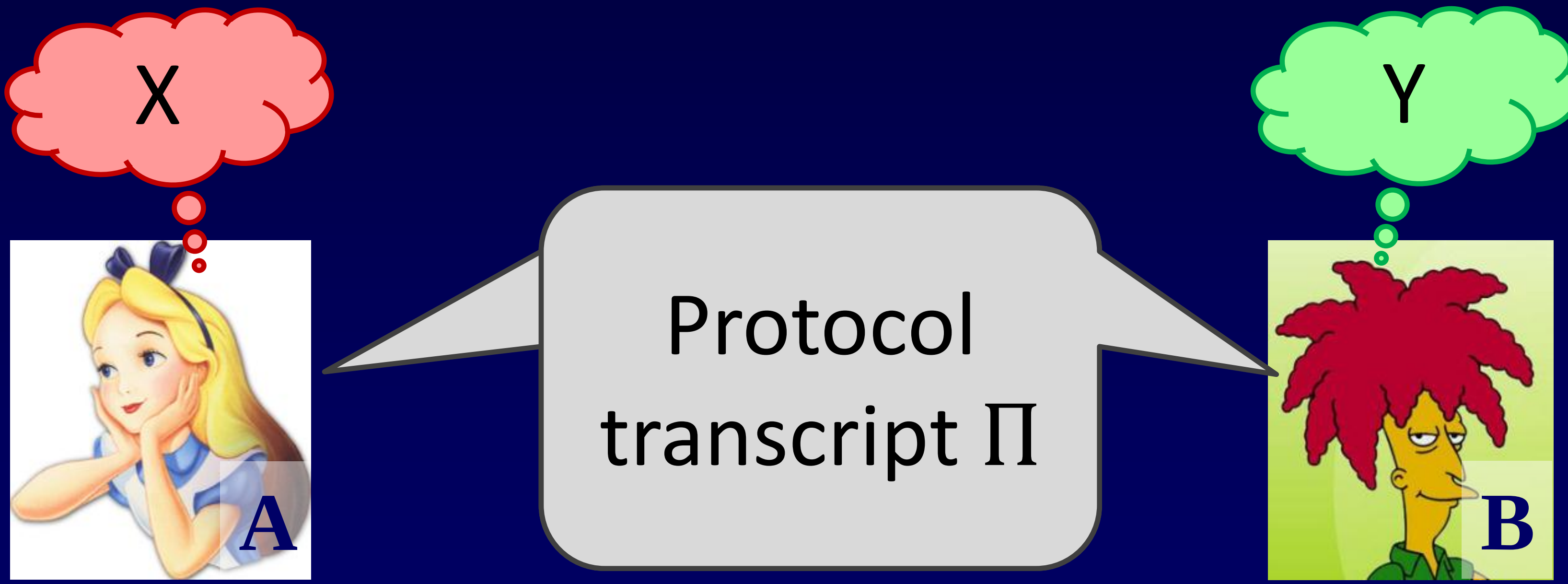
  transmission cost $C(X)$

Information complexity

- The *smallest* amount of *information* Alice and Bob need to exchange to solve F.
- How is information measured?
- Communication cost of a protocol?
  - Number of bits exchanged.
- Information cost of a protocol?
  - Amount of information revealed.

Basic definition 1: The information cost of a protocol

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

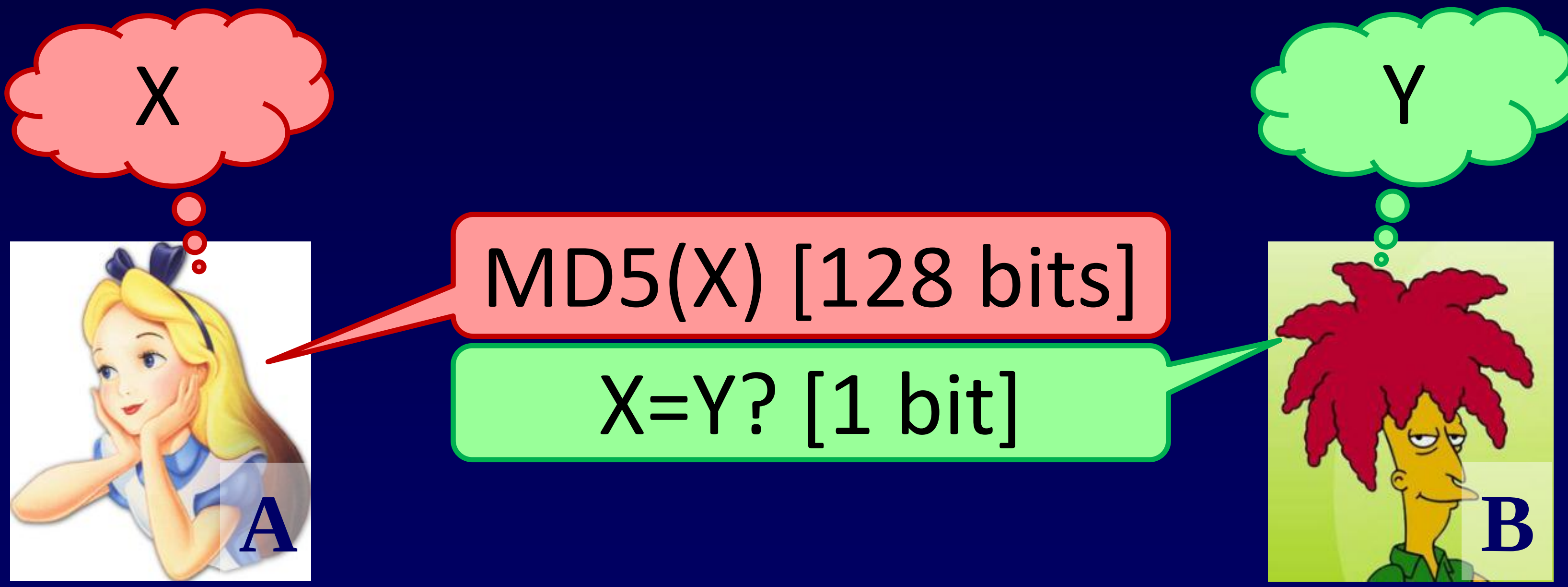


$$IC(\pi, \mu) = I(\Pi; Y|X) + I(\Pi; X|Y)$$

what Alice learns about Y + what Bob learns about X

Example

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



$IC(\pi, \mu) = I(\Pi; Y|X) + I(\Pi; X|Y) \approx 1 + 64.5 = 65.5$ bits

what Alice learns about Y + what Bob learns about X

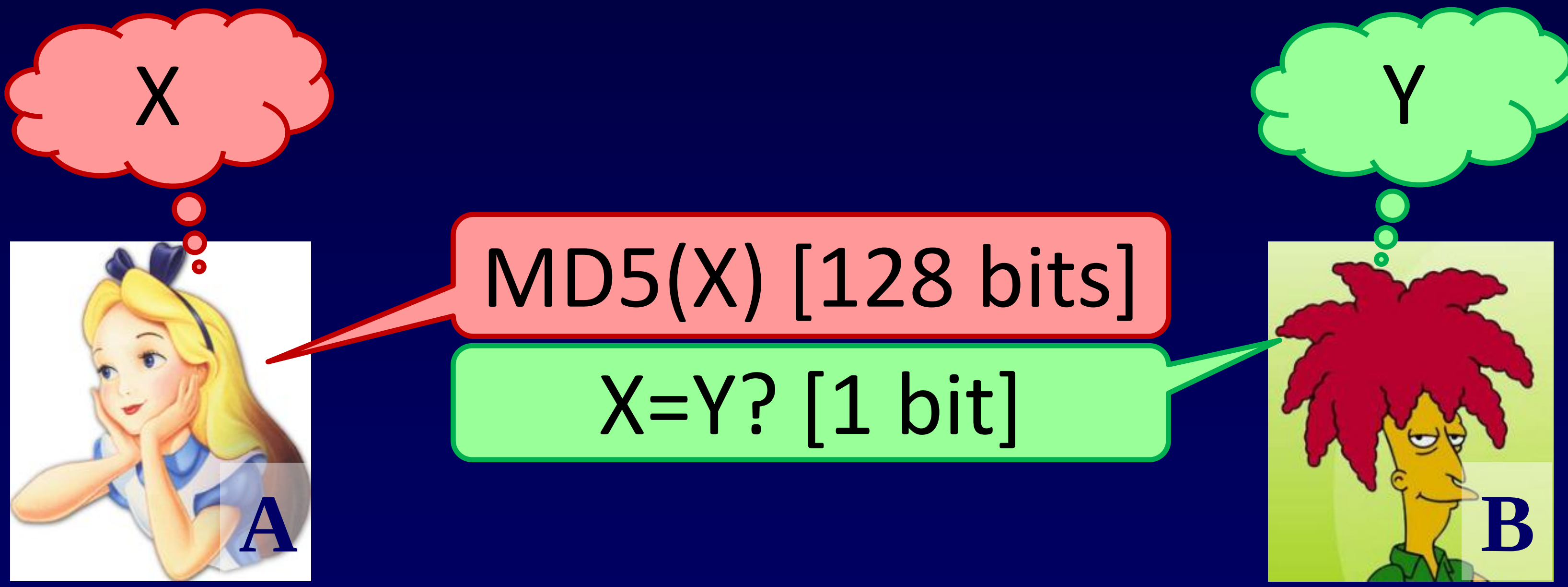
Prior μ matters a lot for information cost!

- If $\mu = 1_{(x,y)}$ a singleton,

$$IC(\pi, \mu) = 0.$$

Example

- F is "$X = Y?$".
- μ is a distribution where (X, Y) are just uniformly random.



$IC(\pi, \mu) = I(\Pi; Y|X) + I(\Pi; X|Y) \approx 0 + 128 = 128$ bits

what Alice learns about Y + what Bob learns about X

Basic definition 2: Information complexity

- Communication complexity:

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

- Analogously:

Needed!

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

Prior-free information complexity

- Using minimax can get rid of the prior.
- For communication, we had:

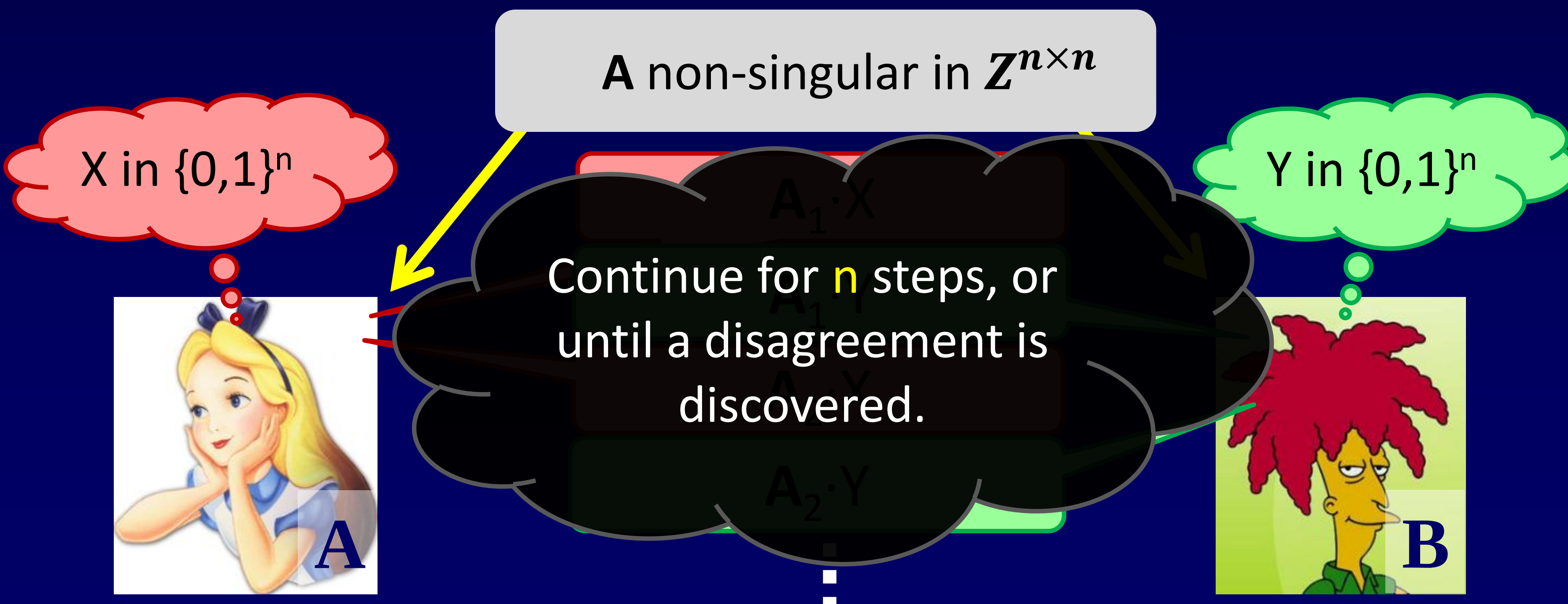
$$CC(F,\varepsilon) = \max_{\mu} CC(F,\mu,\varepsilon).$$

- For information

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

Ex: The information complexity of Equality

- What is $IC(EQ, 0)$?
- Consider the following protocol.



Analysis (sketch)

- If $X \neq Y$, the protocol will terminate in $O(1)$ rounds on average, and thus reveal $O(1)$ information.
- If $X=Y$… the players only learn the fact that $X=Y$ (≤ 1 bit of information).
- Thus the protocol has $O(1)$ information complexity for any prior μ.

Operationalizing IC: Information equals amortized communication

- Recall [Shannon]: $\lim_{n\to\infty} C(X^n)/n = H(X).$
- Turns out: $\lim_{n\to\infty} CC(F^n,\mu^n,\varepsilon)/n = IC(F,\mu,\varepsilon),$ *for* $\varepsilon > 0.$ [Error ε allowed on each copy]
- For $\varepsilon = 0$: $\lim_{n\to\infty} CC(F^n,\mu^n,0^+)/n = IC(F,\mu,0).$
- [$\lim_{n\to\infty} CC(F^n,\mu^n,0)/n$ an interesting open problem.]

Information = amortized communication

- $\lim_{n\to\infty} CC(F^n, \mu^n, \varepsilon)/n = IC(F, \mu, \varepsilon)$.
- Two directions: "$\leq$" and "$\geq$".

- $n \cdot \underbrace{H(X) = H(X^n)}_{\text{Additivity of entropy}} \leq \underbrace{C(X^n) \leq H(X^n)}_{\text{Compression (Huffman)}} + 1$.

The "$\leq$" direction

- $\lim_{n\to\infty} CC(F^n, \mu^n, \varepsilon)/n \leq IC(F, \mu, \varepsilon)$.
- Start with a protocol π solving F, whose $IC(\pi, \mu)$ is close to $IC(F, \mu, \varepsilon)$.
- Show how to *compress* many copies of π into a protocol whose communication cost is close to its information cost.
- More on compression later.

The "$\geq$" direction

- $\lim_{n\to\infty} CC(F^n,\mu^n,\varepsilon)/n \geq IC(F,\mu,\varepsilon)$.
- Use the fact that $\frac{CC(F^n,\mu^n,\varepsilon)}{n} \geq \frac{IC(F^n,\mu^n,\varepsilon)}{n}$.
- *Additivity* of information complexity:

$$\frac{IC(F^n,\mu^n,\varepsilon)}{n} = IC(F,\mu,\varepsilon).$$

Proof: Additivity of information complexity

- Let $T_1(X_1, Y_1)$ and $T_2(X_2, Y_2)$ be two two-party tasks.
- E.g. "Solve $F(X, Y)$ with error $\leq \varepsilon$ w.r.t. μ"
- Then

$$IC(T_1 \times T_2, \mu_1 \times \mu_2) = IC(T_1, \mu_1) + IC(T_2, \mu_2)$$

- "$\leq$" is easy.
- "$\geq$" is the interesting direction.

$$IC(T_1,\mu_1) + IC(T_2,\mu_2) \leq IC(T_1 \times T_2, \mu_1 \times \mu_2)$$

- Start from a protocol π for $T_1 \times T_2$ with prior $\mu_1 \times \mu_2$, whose information cost is I.
- Show how to construct two protocols π_1 for T_1 with prior μ_1 and π_2 for T_2 with prior μ_2, with information costs I_1 and I_2, respectively, such that $I_1 + I_2 = I$.

$\pi((X_1, X_2), (Y_1, Y_2))$

$\pi_1(X_1, Y_1)$

- Publicly sample $X_2 \sim \mu_2$
- Bob privately samples $Y_2 \sim \mu_2|_{X_2}$
- Run $\pi((X_1, X_2), (Y_1, Y_2))$

$\pi_2(X_2, Y_2)$

- Publicly sample $Y_1 \sim \mu_1$
- Alice privately samples $X_1 \sim \mu_1|_{Y_1}$
- Run $\pi((X_1, X_2), (Y_1, Y_2))$

Analysis - π_1

$$\pi_1(X_1, Y_1)$$

- Publicly sample $X_2 \sim \mu_2$
- Bob privately samples $Y_2 \sim \mu_2|_{\mathrm{X}_2}$
- Run $\pi((X_1, X_2), (Y_1, Y_2))$

- Alice learns about Y_1:
$$I(\Pi; Y_1 | \mathrm{X}_1 \mathrm{X}_2)$$
- Bob learns about X_1:
$$I(\Pi; X_1 | Y_1 Y_2 X_2).$$
- $I_1 = I(\Pi; Y_1 | \mathrm{X}_1 \mathrm{X}_2) + I(\Pi; X_1 | Y_1 Y_2 X_2).$

Analysis - π_2

$\pi_2(X_2, Y_2)$
- Publicly sample $Y_1 \sim \mu_1$
- Alice privately samples $X_1 \sim \mu_1|_{\mathrm{Y}_1}$
- Run $\pi((X_1, X_2), (Y_1, Y_2))$

- Alice learns about Y_2:

$$I(\Pi; Y_2 | \mathrm{X}_1 \mathrm{X}_2 \mathrm{Y}_1)$$

- Bob learns about X_2:

$$I(\Pi; X_2 | Y_1 Y_2).$$

- $I_2 = I(\Pi; Y_2 | \mathrm{X}_1 \mathrm{X}_2 \mathrm{Y}_1) + I(\Pi; X_2 | Y_1 Y_2).$

Adding I_1 and I_2

$$
\begin{aligned}
I_1 + I_2 &\\
&= I(\Pi; Y_1 | \mathrm{X}_1 \mathrm{X}_2) + I(\Pi; X_1 | Y_1 Y_2 X_2) \\
&\qquad + I(\Pi; Y_2 | \mathrm{X}_1 \mathrm{X}_2 \mathrm{Y}_1) + I(\Pi; X_2 | Y_1 Y_2)
\end{aligned}
$$

$= I(\Pi; Y_1 | \mathrm{X}_1 \mathrm{X}_2) + I(\Pi; Y_2 | \mathrm{X}_1 \mathrm{X}_2 \mathrm{Y}_1) +$
$I(\Pi; X_2 | Y_1 Y_2) + I(\Pi; X_1 | Y_1 Y_2 X_2) =$

$I(\Pi; Y_1 Y_2 | X_1 X_2) + I(\Pi; X_2 X_1 | Y_1 Y_2) = I.$

Summary

- Information complexity is additive.
- Operationalized via “Information = amortized communication”.
- $\lim_{n\to\infty} CC(F^n, \mu^n, \varepsilon)/n = IC(F, \mu, \varepsilon)$.
- Seems to be the “right” analogue of entropy for interactive computation.

Entropy vs. Information Complexity

	Entropy	IC
Additive?	Yes	Yes
Operationalized	$\lim_{n\to\infty} C(X^n)/n$	$\lim_{n\to\infty} \frac{CC(F^n, \mu^n, \varepsilon)}{n}$
Compression?	Huffman: $C(X) \leq H(X) + 1$	**???!**

Can interactive communication be compressed?

- Is it true that $CC(F, \mu, \varepsilon) \leq IC(F, \mu, \varepsilon) + O(1)$?
- Less ambitiously:

$$CC(F, \mu, O(\varepsilon)) = O\big(IC(F, \mu, \varepsilon)\big)?$$

- (Almost) equivalently: Given a protocol π with $IC(\pi, \mu) = I$, can Alice and Bob simulate π using $O(I)$ communication?
- Not known in general…

Direct sum theorems

- Let F be any functionality.
- Let $C(F)$ be the cost of implementing F.
- Let F^n be the functionality of implementing n independent copies of F.
- The direct sum problem:

  "Does $C(F^n) \approx n \cdot C(F)$?"
- In most cases it is obvious that $C(F^n) \leq n \cdot C(F)$.

Direct sum – randomized communication complexity

- Is it true that
$$CC(F^n, \mu^n, \varepsilon) = \Omega(n \cdot CC(F, \mu, \varepsilon))?$$
- Is it true that $CC(F^n, \varepsilon) = \Omega(n \cdot CC(F, \varepsilon))$?

Direct product – randomized communication complexity

- Direct sum

$$CC(F^n, \mu^n, \varepsilon) \;=\; \Omega(n \cdot CC(F, \mu, \varepsilon))?$$

- Direct product

$$CC(F^n, \mu^n, (1-\varepsilon)^n) \;=\; \Omega(n \cdot CC(F, \mu, \varepsilon))?$$

Direct sum for randomized CC and interactive compression

Direct sum:

- $CC(F^n, \mu^n, \varepsilon) = \Omega(n \cdot CC(F, \mu, \varepsilon))$?

In the limit:

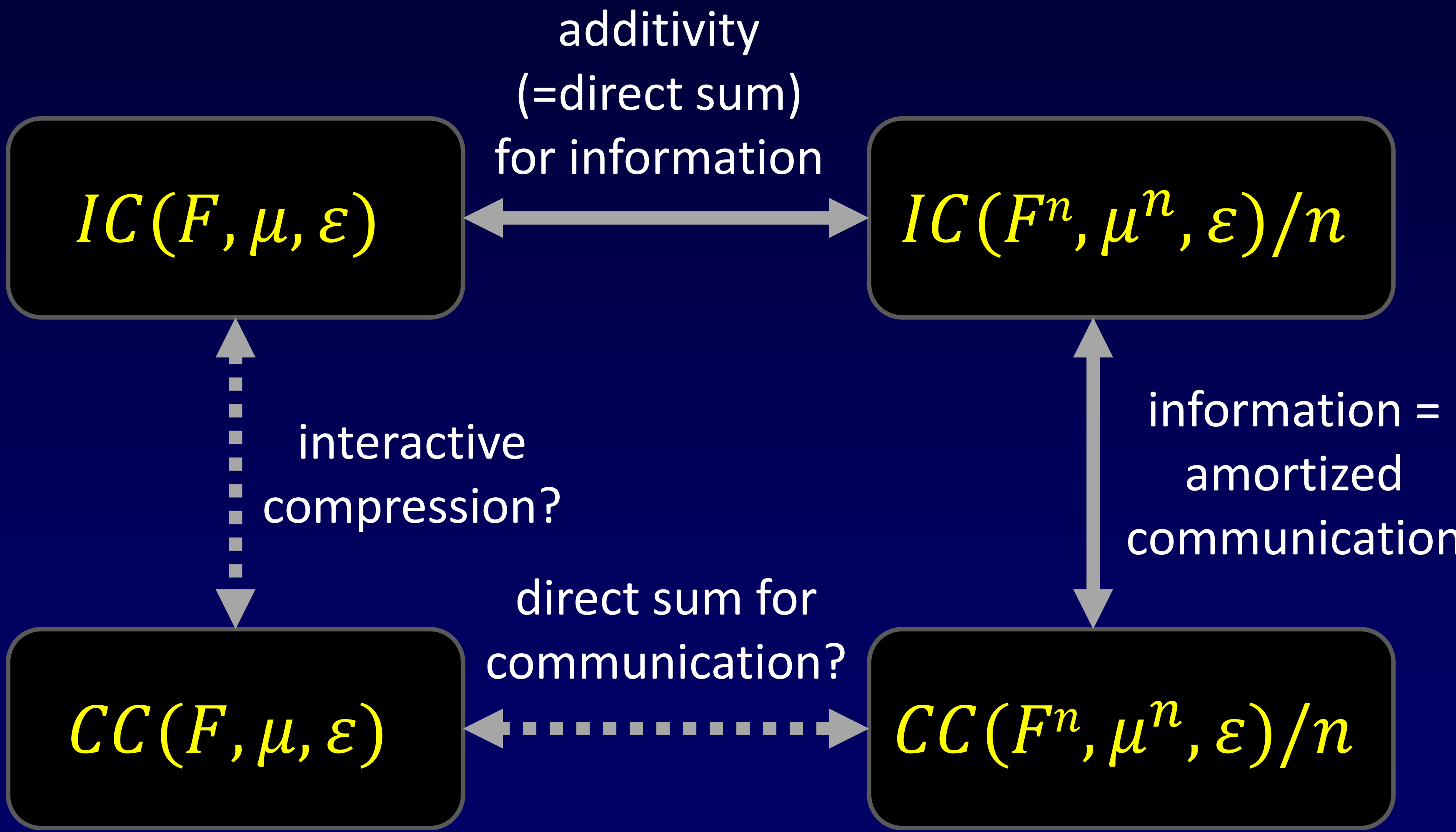
- $n \cdot IC(F, \mu, \varepsilon) = \Omega(n \cdot CC(F, \mu, \varepsilon))$?

Interactive compression:

- $CC(F, \mu, \varepsilon) = O(IC(F, \mu, \varepsilon))$?

Same question!

The big picture



Current results for compression

A protocol π that has C bits of communication, conveys I bits of information over prior μ, and works in r rounds can be simulated:

- Using $\tilde{O}(I + r)$ bits of communication.
- Using $\tilde{O}\left(\sqrt{I \cdot C}\right)$ bits of communication.
- Using $2^{O(I)}$ bits of communication.
- If $\mu = \mu_X \times \mu_Y$, then using $O(I \operatorname{polylog} C)$ bits of communication.

Their direct sum counterparts

- $CC(F^n, \mu^n, \varepsilon) = \widetilde{\Omega}(n^{1/2} \cdot CC(F, \mu, \varepsilon))$.
- $CC(F^n, \varepsilon) = \widetilde{\Omega}(n^{1/2} \cdot CC(F, \varepsilon))$.

For product distributions $\mu = \mu_X \times \mu_Y$,

- $CC(F^n, \mu^n, \varepsilon) = \widetilde{\Omega}(n \cdot CC(F, \mu, \varepsilon))$.

When the number of rounds is bounded by $r \ll n$, a direct sum theorem holds.

Direct product

- The best one can hope for is a statement of the type:
  $CC(F^n, \mu^n, 1 - 2^{-O(n)}) = \Omega(n \cdot IC(F, \mu, 1/3)).$
- Can prove:

$$CC\left(F^n, \mu^n, 1 - 2^{-O(n)}\right) = \tilde{\Omega}(n^{1/2} \cdot CC(F, \mu, 1/3)).$$

Proof 2: Compressing a one-round protocol

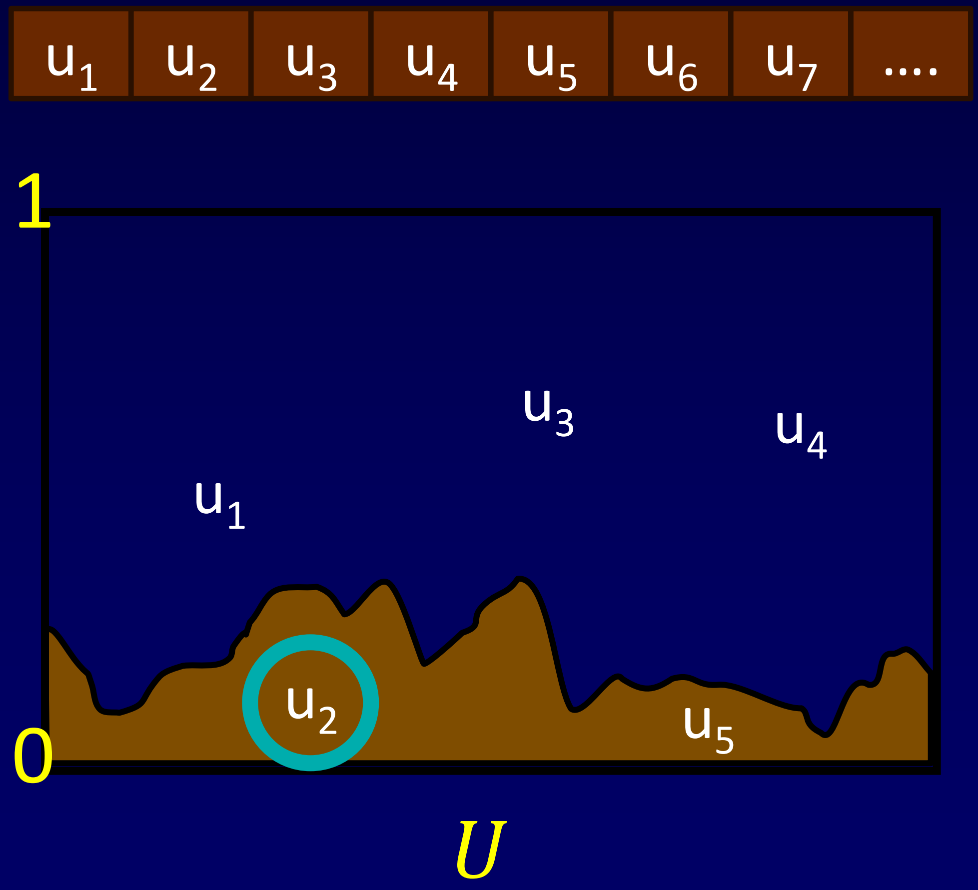
- Say Alice speaks: $IC(\pi, \mu) = I(M; X|Y)$.
- Recall KL-divergence:

$$I(M; X|Y) = E_Y\, D(M_{XY} \parallel M_Y) = E_Y\, D(M_X \parallel M_Y)$$

- Bottom line:
  - Alice has M_X; Bob has M_Y;
  - <u>Goal</u>: sample from M_X using $\sim D(M_X \parallel M_Y)$ communication.

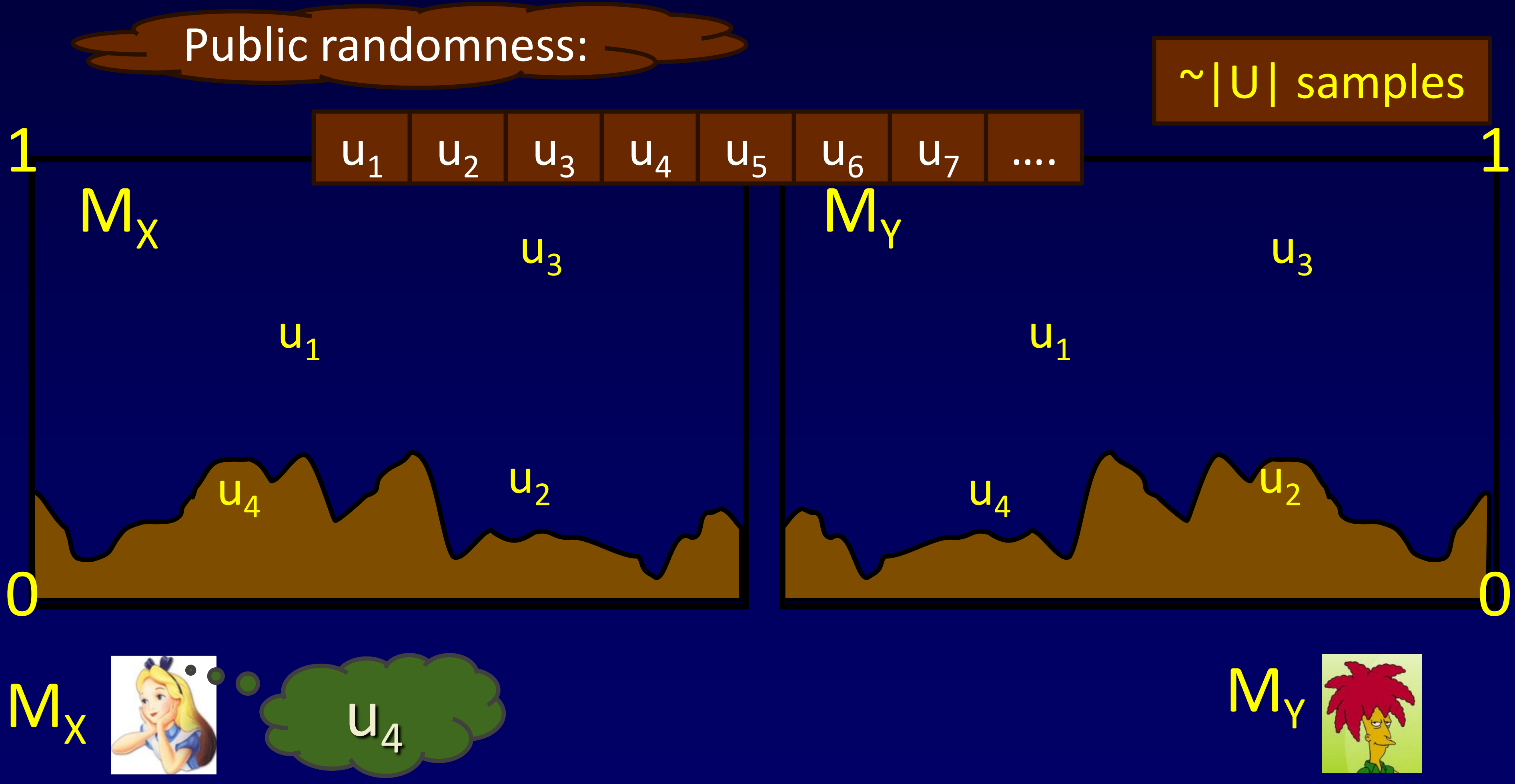
The dart board

- Interpret the public randomness as random points in $U \times [0,1]$, where U is the universe of all possible messages.
- First message under the histogram of M is distributed $\sim M$.



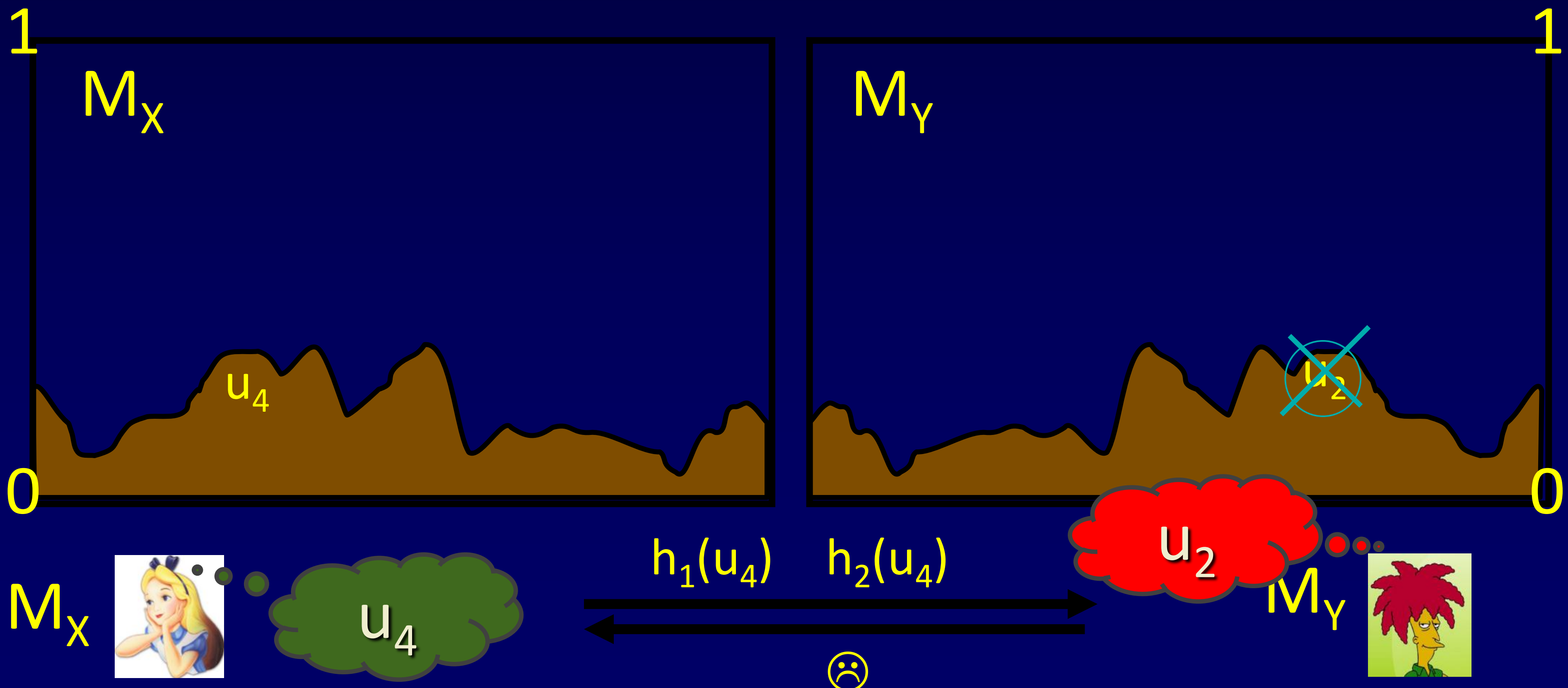
Proof Idea

- Sample using $O(\log 1/\varepsilon + D(M_X \parallel M_Y))$ communication with statistical error ε.



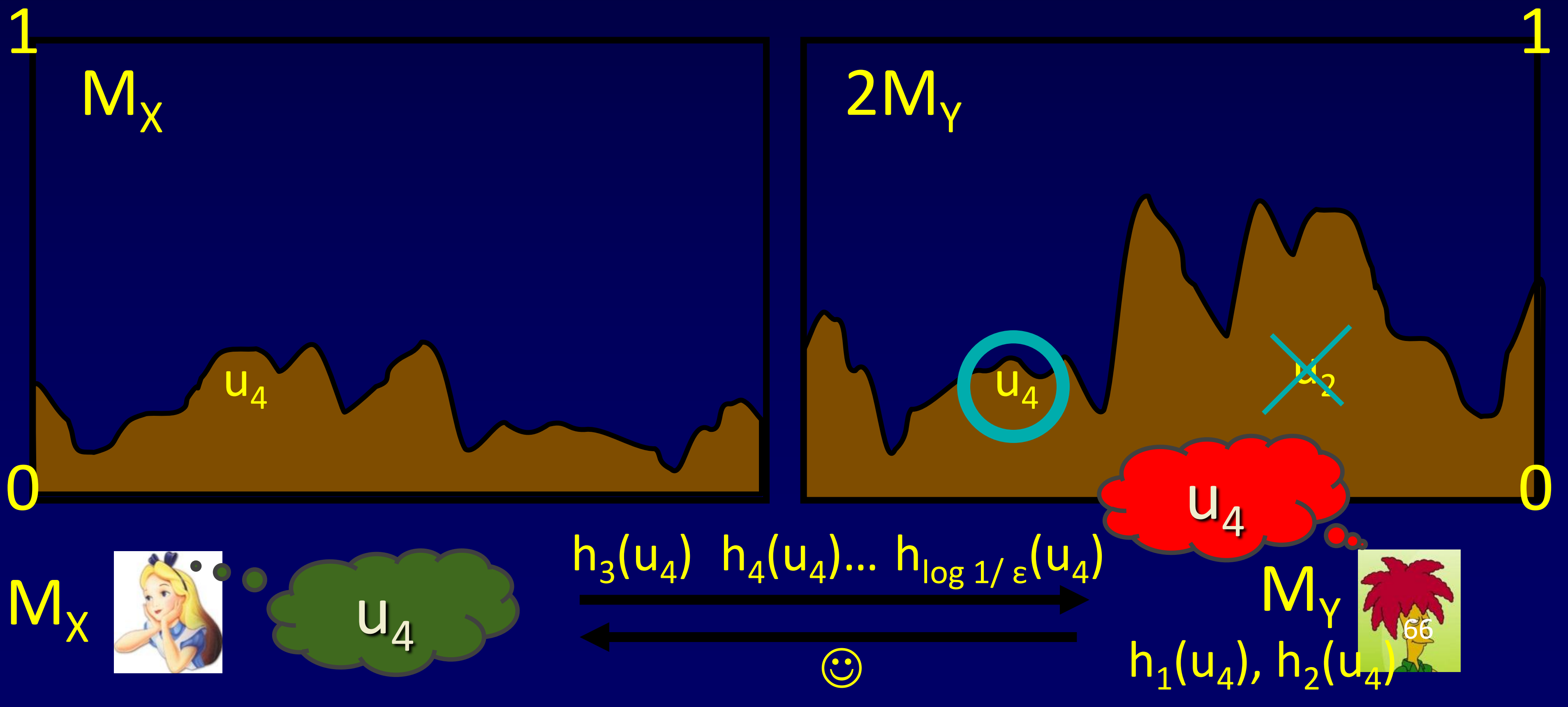
Proof Idea

- Sample using $O(\log 1/\varepsilon + D(M_X \parallel M_Y))$ communication with statistical error ε.



Proof Idea

- Sample using $O(\log 1/\varepsilon + D(M_X \parallel M_Y))$ communication with statistical error ε.



Analysis

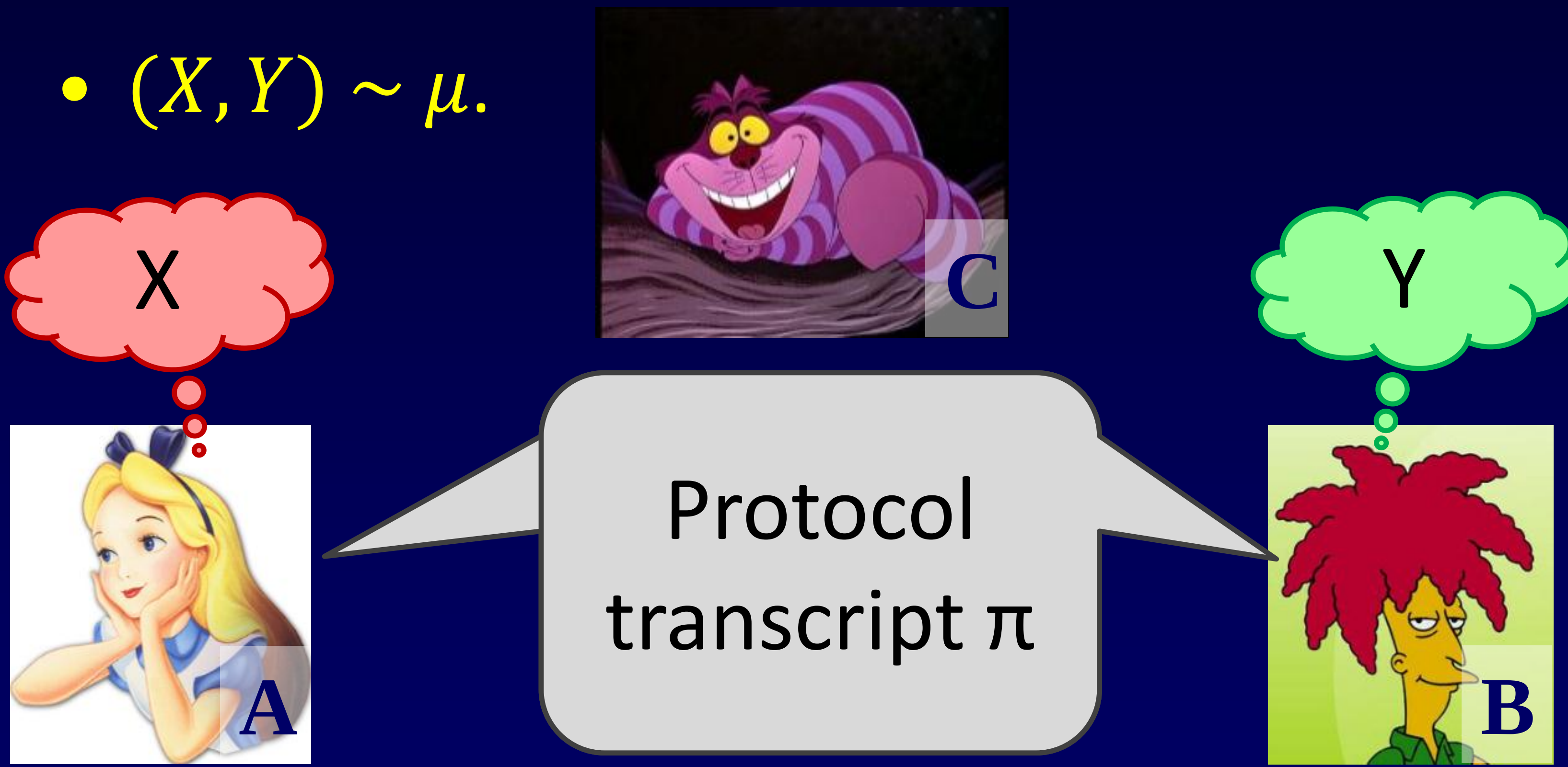
- If $M_X(u_4) \approx 2^k M_Y(u_4)$, then the protocol will reach round k of doubling.
- There will be $\approx 2^k$ candidates.
- About $k + \log 1/\varepsilon$ hashes to narrow to one.
- The contribution of u_4 to cost:
  - $M_X(u_4)\,(\log M_X(u_4)/M_Y(u_4) + \log 1/\varepsilon)$.

$$\boxed{D(M_X \parallel M_Y) := \sum_u M_X(u) \log \frac{M_X(u)}{M_Y(u)}.}$$

Done!

External information cost

- $(X, Y) \sim \mu$.

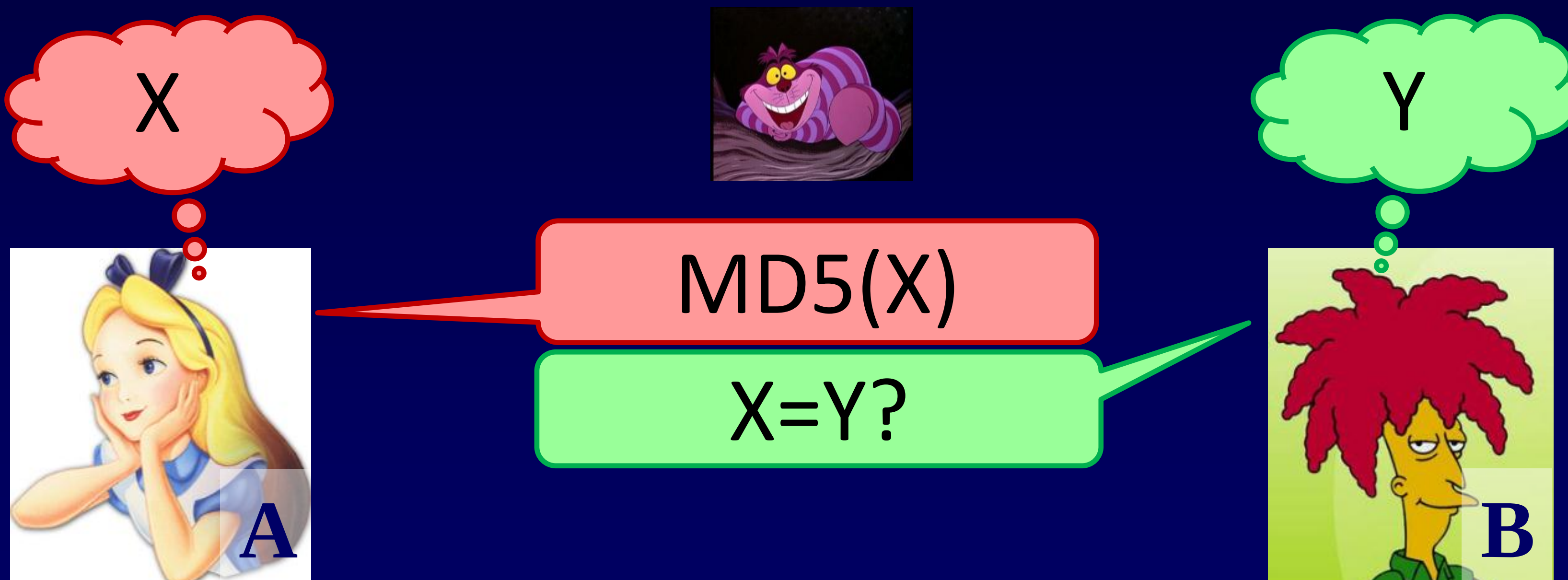


$$IC_{ext}(\pi, \mu) = I(\Pi; XY)$$

what Charlie learns about (X,Y)

Example

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



$$IC_{ext}(\pi,\mu) \ = \ I(\Pi; XY) \ = \ 129 \ bits$$

what Charlie learns about (X,Y)

External information cost

- It is always the case that
$$IC_{ext}(\pi, \mu) \geq IC(\pi, \mu).$$
- If $\mu = \mu_X \times \mu_Y$ is a product distribution, then
$$IC_{ext}(\pi, \mu) = IC(\pi, \mu).$$

External information complexity

- $IC_{ext}(F, \mu, \varepsilon) := \inf_{\substack{\pi\ computes \\ F\ with\ error\ \leq \varepsilon}} IC_{ext}(\pi, \mu)$.
- Can it be operationalized?

Operational meaning of IC_{ext}?

- <u>Conjecture</u>: Zero-error communication scales like external information:

$$\lim_{n\to\infty} \frac{CC(F^n, \mu^n, 0)}{n} = IC_{ext}(F, \mu, 0)?$$

- Recall:

$$\lim_{n\to\infty} \frac{CC(F^n, \mu^n, 0^+)}{n} = IC(F, \mu, 0).$$

Example – transmission with a strong prior

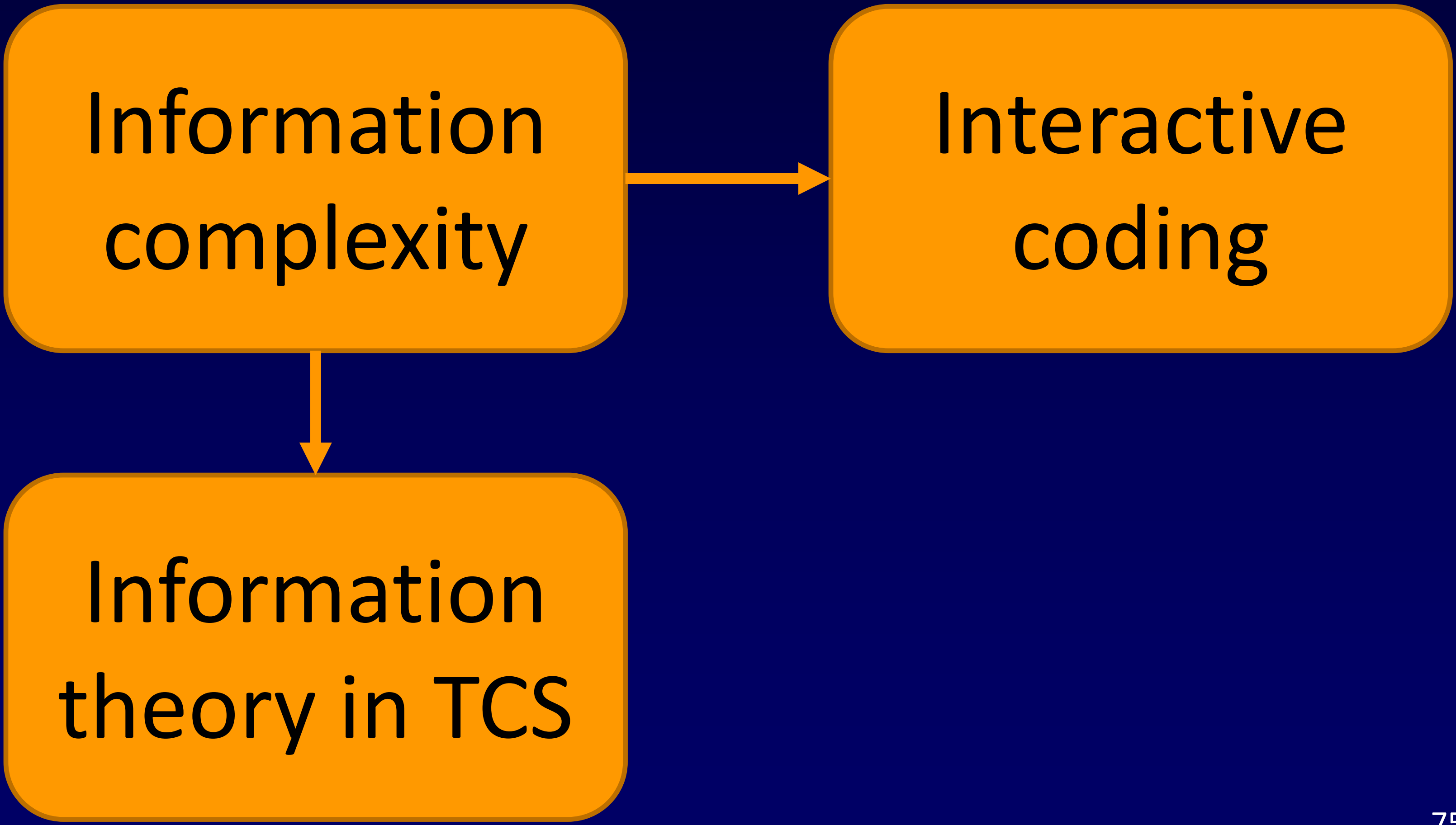
- $X, Y \in \{0,1\}$
- μ is such that $X \in_U \{0,1\}$, and $X = Y$ with a very high probability (say $1 - 1/\sqrt{n}$).
- $F(X,Y) = X$ is just the "transmit X" function.
- Clearly, π should just have Alice send X to Bob.
- $IC(F, \mu, 0) = IC(\pi, \mu) = H\left(\frac{1}{\sqrt{n}}\right) = o(1)$.
- $IC_{ext}(F, \mu, 0) = IC_{ext}(\pi, \mu) = 1$.

Example – transmission with a strong prior

- $IC(F, \mu, 0) = IC(\pi, \mu) = H\left(\frac{1}{\sqrt{n}}\right) = \mathrm{o}(1).$
- $IC_{ext}(F, \mu, 0) = IC_{ext}(\pi, \mu) = 1.$
- $CC(F^n, \mu^n, 0^+) = o(n).$
- $CC(F^n, \mu^n, 0) = \Omega(n).$

Other examples, e.g. the two-bit AND function fit into this picture.

Additional directions



Interactive coding theory

- So far focused the discussion on *noiseless coding*.
- What if the channel has noise?
- [What kind of noise?]
- In the non-interactive case, each channel has a capacity C.

Channel capacity

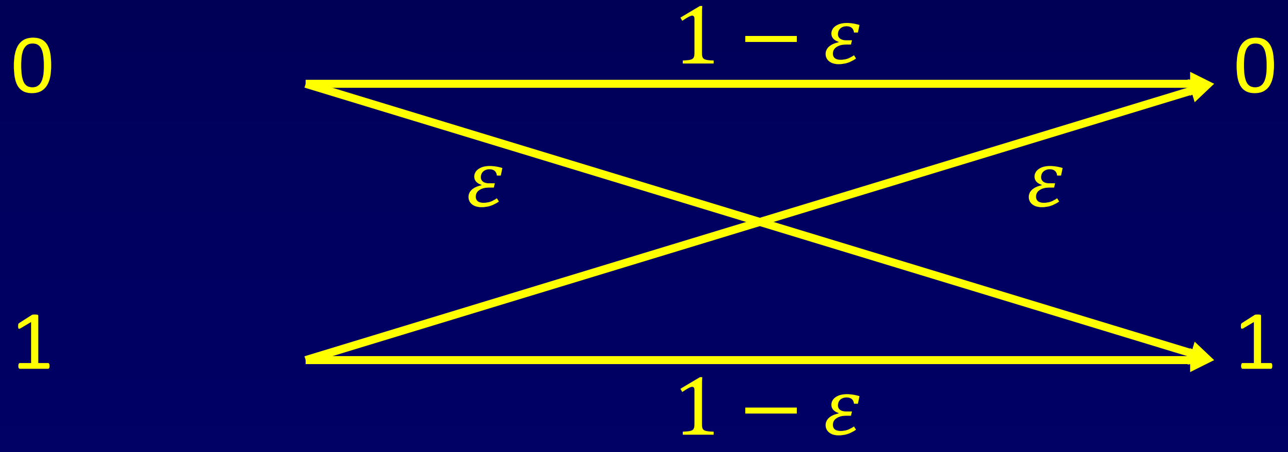
- The amortized number of channel uses needed to send X over a noisy channel of capacity C is

$$\frac{H(X)}{C}$$

- Decouples the task from the channel!

Example: Binary Symmetric Channel

- Each bit gets independently flipped with probability $\varepsilon < 1/2$.
- One way capacity $1 - H(\varepsilon)$.

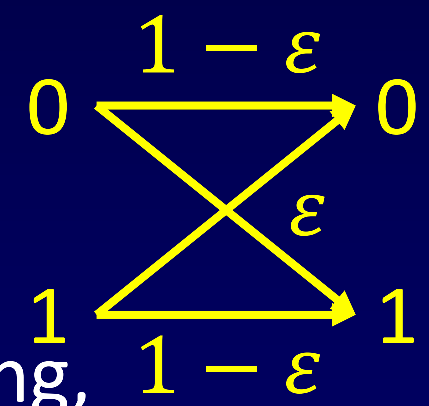


Interactive channel capacity

- Not clear one can decouple channel from task in such a clean way.
- Capacity much harder to calculate/reason about.
- Example: Binary symmetric channel.
- One way capacity $1 - H(\varepsilon)$.
- Interactive (for simple pointer jumping,

[Kol-Raz'13]):

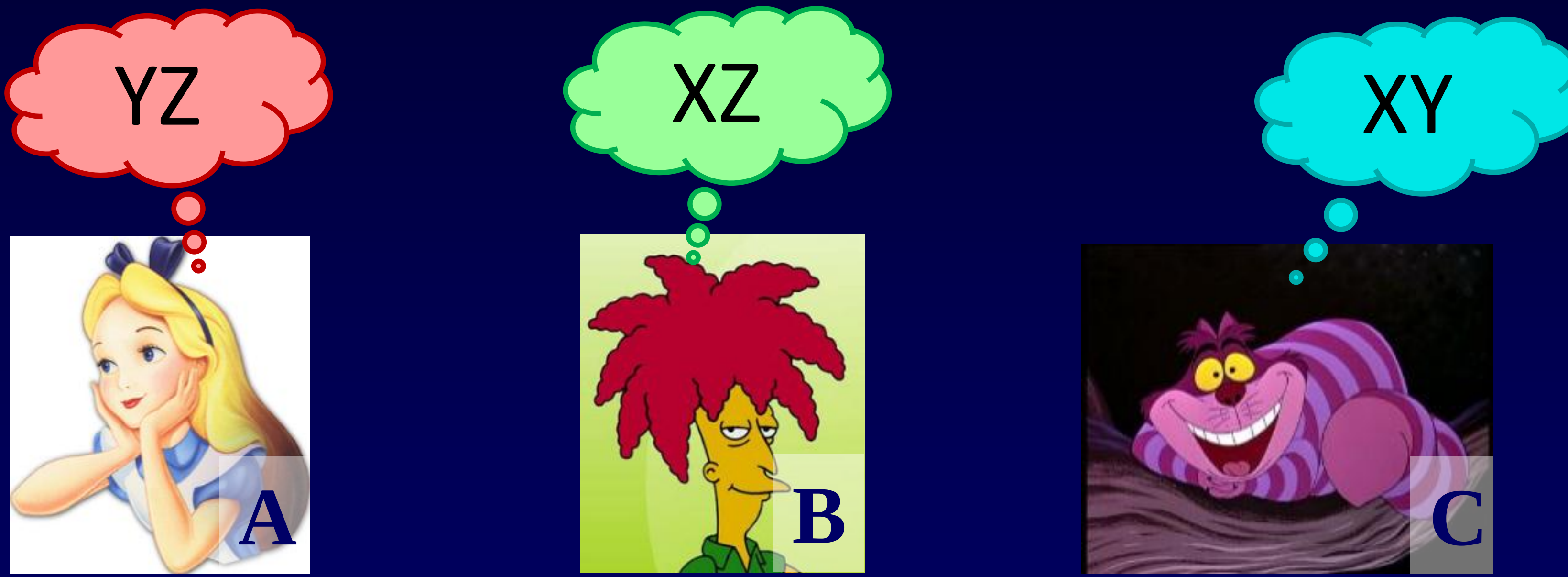
$$1 - \Theta\left(\sqrt{H(\varepsilon)}\right).$$



Information theory in communication complexity and beyond

- A natural extension would be to multi-party communication complexity.
- Some success in the number-in-hand case.
- What about the number-on-forehead?
- Explicit bounds for $\geq \log n$ players would imply explicit ACC^0 circuit lower bounds.

Naïve multi-party information cost



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

Naïve multi-party information cost

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

- Doesn't seem to work.
- Secure multi-party computation [Ben-Or,Goldwasser, Wigderson], means that anything can be computed at near-zero information cost.
- Although, these construction require the players to share private channels/randomness.

Communication and beyond…

- The rest of today:
  - Data structures;
  - Streaming;
  - Distributed computing;
  - Privacy.
- Exact communication complexity bounds.
- Extended formulations lower bounds.
- Parallel repetition?
- …

Thank You!

Open problem: Computability of IC

- Given the truth table of $F(X,Y)$, μ and ε, compute $IC(F,\mu,\varepsilon)$.
- Via $IC(F,\mu,\varepsilon) = \lim_{n\to\infty} CC(F^n,\mu^n,\varepsilon)/n$ can compute a sequence of upper bounds.
- But the rate of convergence as a function of n is unknown.

Open problem: Computability of IC

- Can compute the r-round $IC_r(F, \mu, \varepsilon)$ information complexity of F.
- But the rate of convergence as a function of r is unknown.
- Conjecture:

$$IC_r(F, \mu, \varepsilon) - IC(F, \mu, \varepsilon) = O_{F,\mu,\varepsilon}\left(\frac{1}{r^2}\right).$$

- This is the relationship for the two-bit AND.