Exploring Function and Distribution Structure in Interactive Computing Through Examples

Prakash Ishwar

Joint work with Nan Ma and Piyush Gupta

Banff International Research Station for Mathematical Innovation and Discovery

BOSTON UNIVERSITY

NSF

Preliminary remarks

- Focus:
  - Lossless distributed function computation in source networks
  - Nodes connected by bidirectional rate-limited error-free bit-pipes
  - Discrete Memoryless Stationary Sources
- New degree of freedom: multi-round interaction
- Disclaimers:
  - No structured coding ensembles
  - No Gaussian Quadratic problem
  - Some theory but no proofs
  - Lots of simple but striking examples

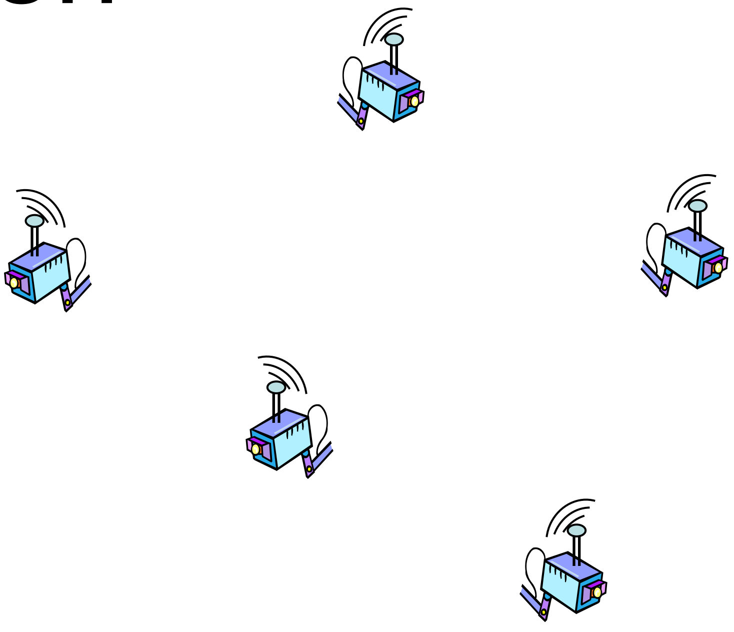
Outline

- Introduction
- General two-terminal problem
- Co-located network with independent sources
- General multi-terminal problem: some observations

BOSTON UNIVERSITY

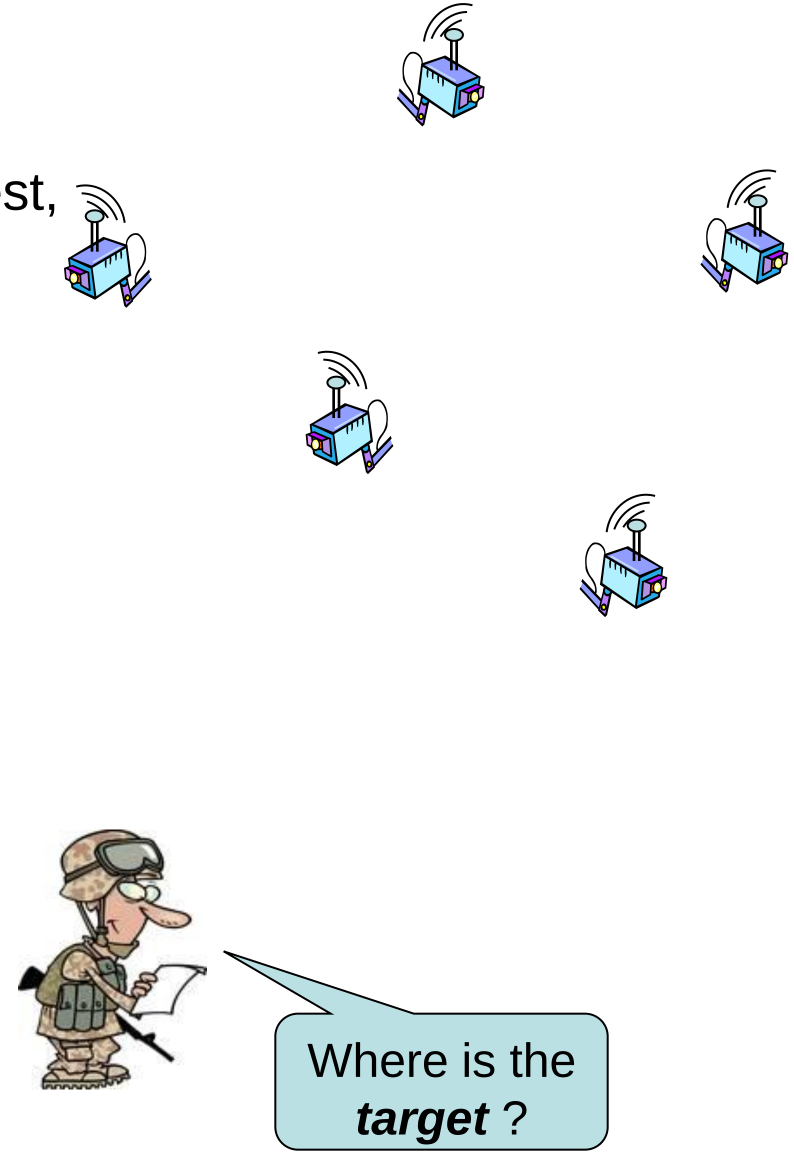
Motivation

- Wireless sensor networks:
  - Provide only information of interest, not the entire data



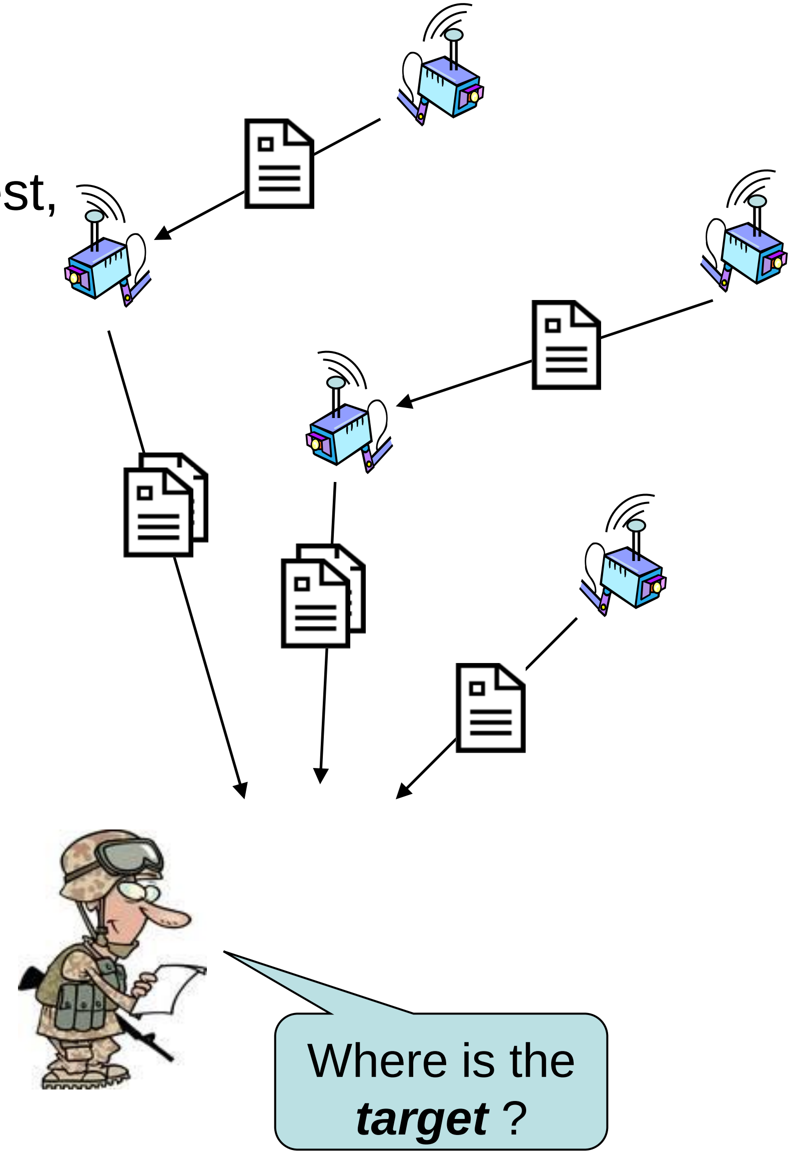
Motivation

- Wireless sensor networks:
  - Provide only information of interest, not the entire data
- Traditional data networks:



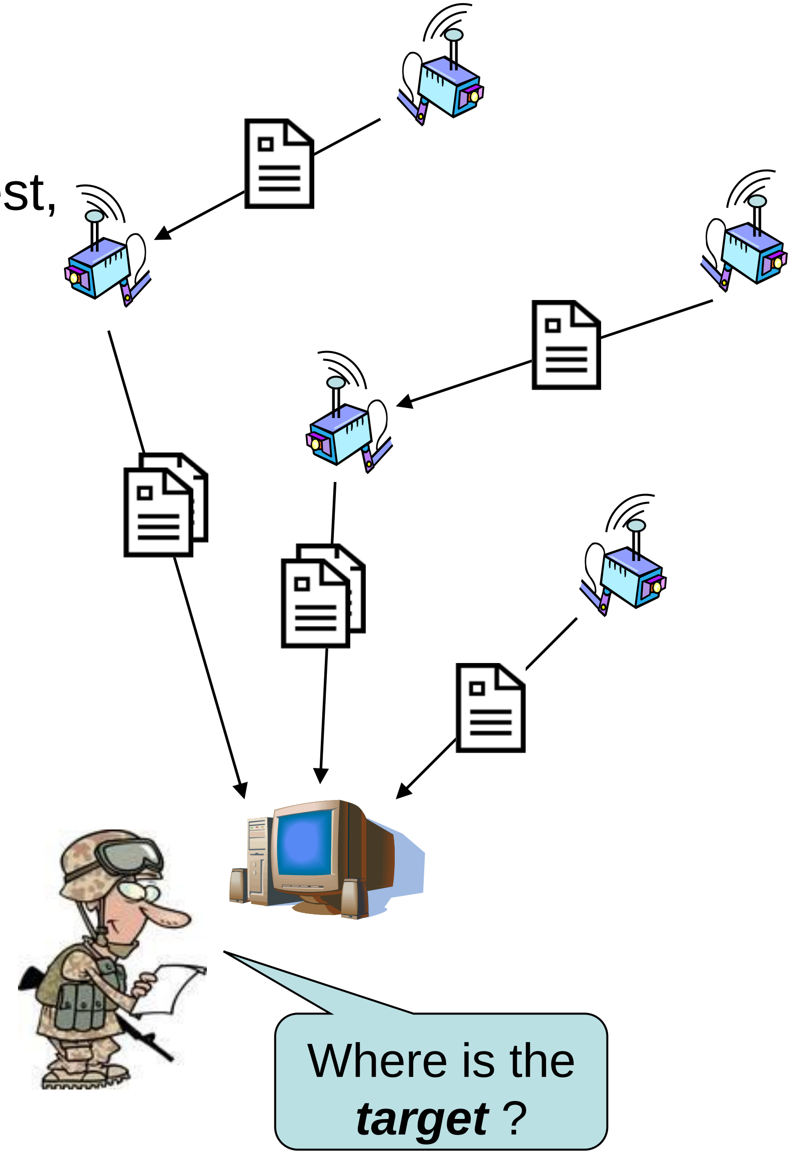
Motivation

- Wireless sensor networks:
  - Provide only information of interest, not the entire data
- Traditional data networks:
  - Move data to destination



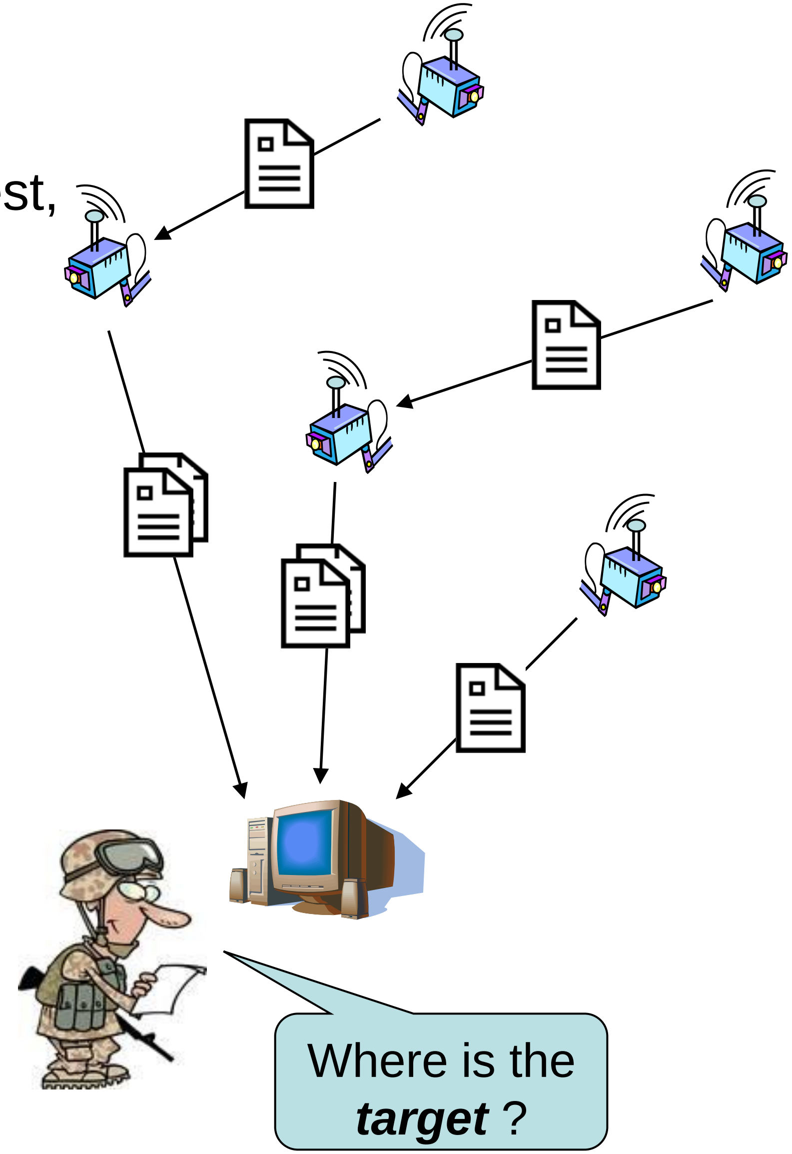
Motivation

- Wireless sensor networks:
  - Provide only information of interest, not the entire data
- Traditional data networks:
  - Move data to destination
  - Process data at destination



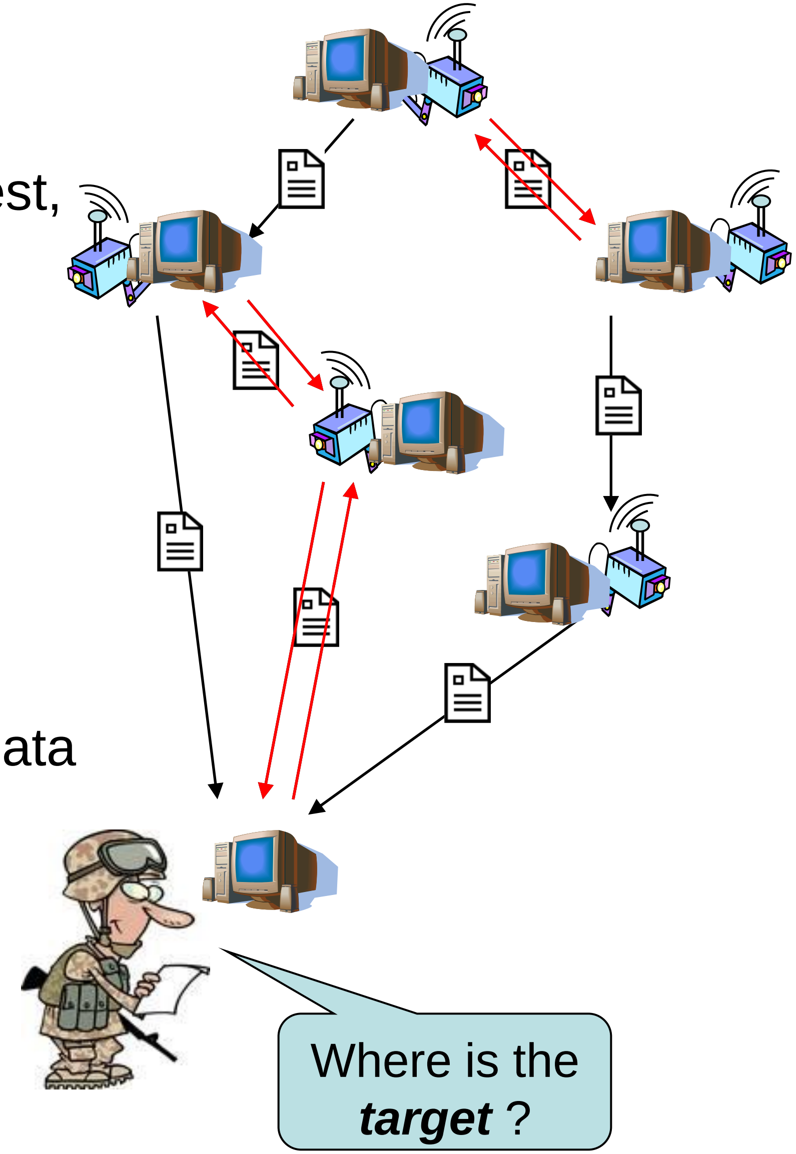
Motivation

- Wireless sensor networks:
  - Provide only information of interest, not the entire data
- Traditional data networks:
  - Move data to destination
  - Process data at destination
  - Inefficient communication



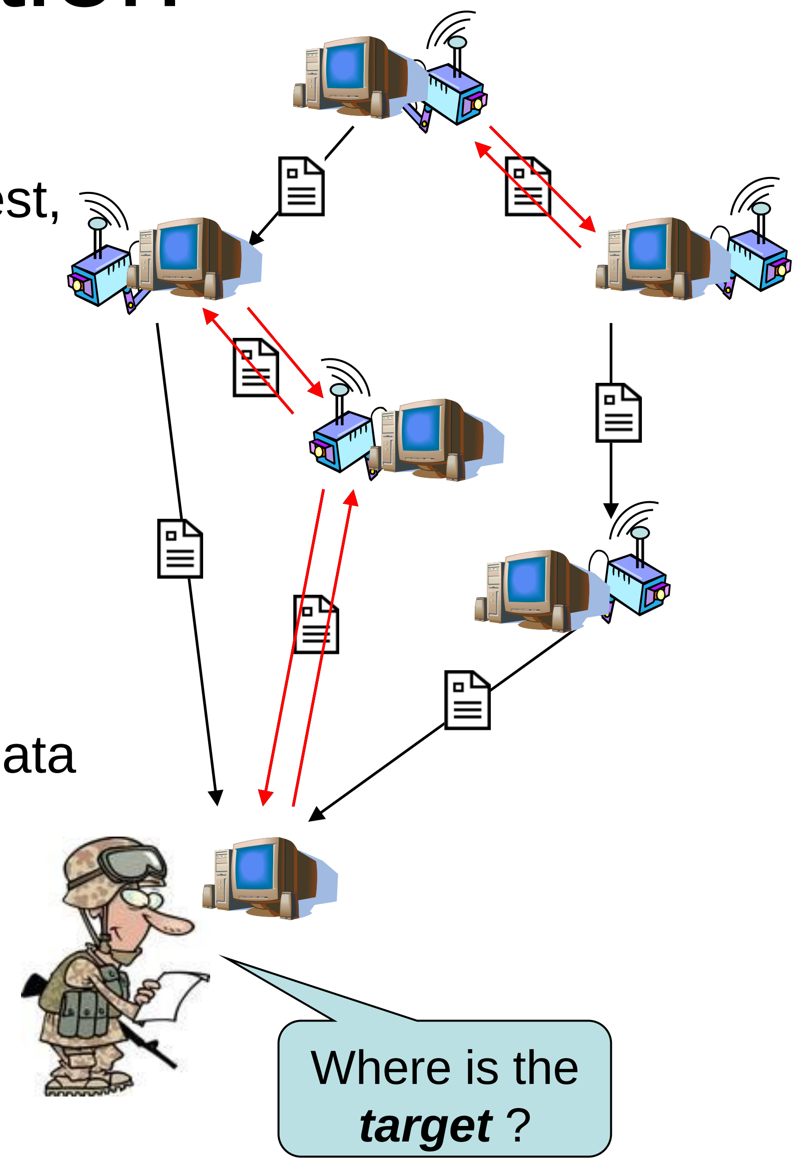
Motivation

- Wireless sensor networks:
  - Provide only information of interest, not the entire data
- Traditional data networks:
  - Move data to destination
  - Process data at destination
  - Inefficient communication
- In-network computing:
  - Distributed computing: process data as it moves



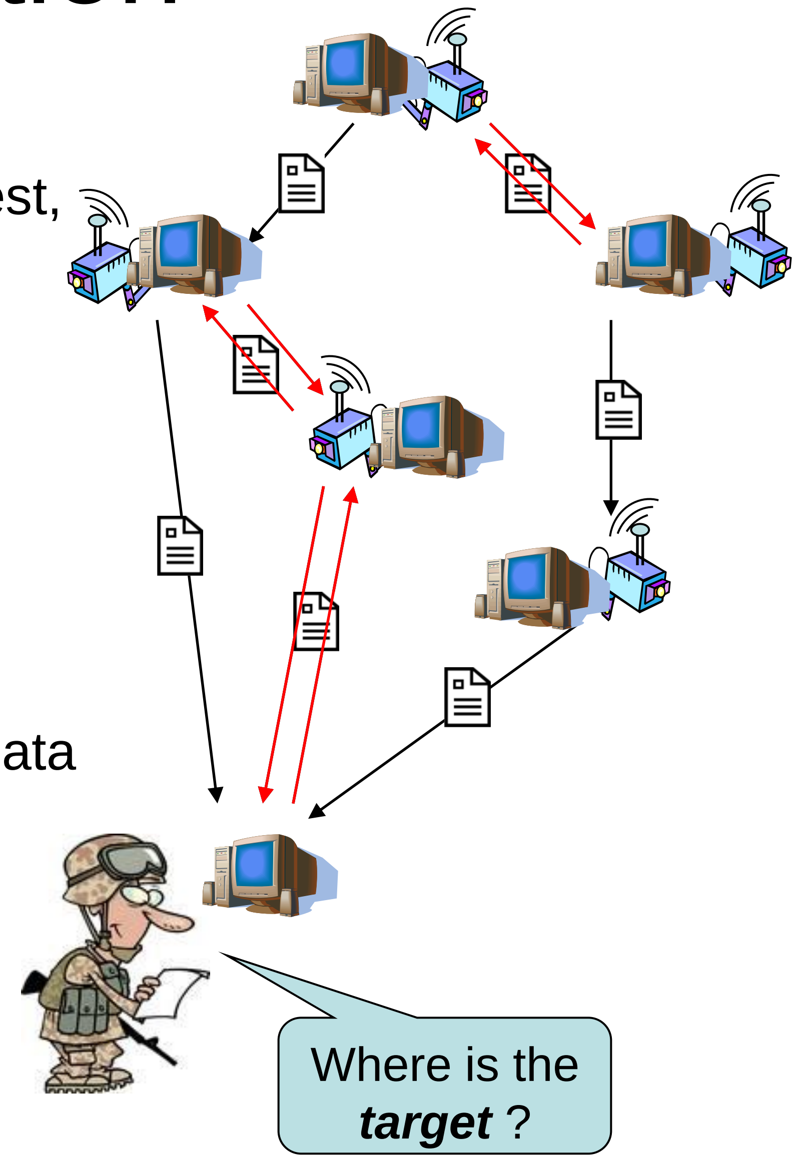
Motivation

- Wireless sensor networks:
  - Provide only information of interest, not the entire data
- Traditional data networks:
  - Move data to destination
  - Process data at destination
  - Inefficient communication
- In-network computing:
  - Distributed computing: process data as it moves
  - Efficient communication



Motivation

- Wireless sensor networks:
  - Provide only information of interest, not the entire data
- Traditional data networks:
  - Move data to destination
  - Process data at destination
  - Inefficient communication
- In-network computing:
  - Distributed computing: process data as it moves
  - Efficient communication
  - Two-way communication (interaction)



Motivation

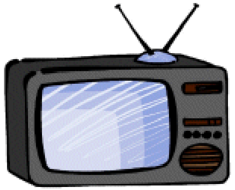
- Is interaction useful?

Motivation

- Is interaction useful? Yes!

Motivation

- Is interaction useful? Yes!



Without interaction (one-way):
Inefficient communication

Motivation

- Is interaction useful? Yes!

Without interaction (one-way):
Inefficient communication

Motivation

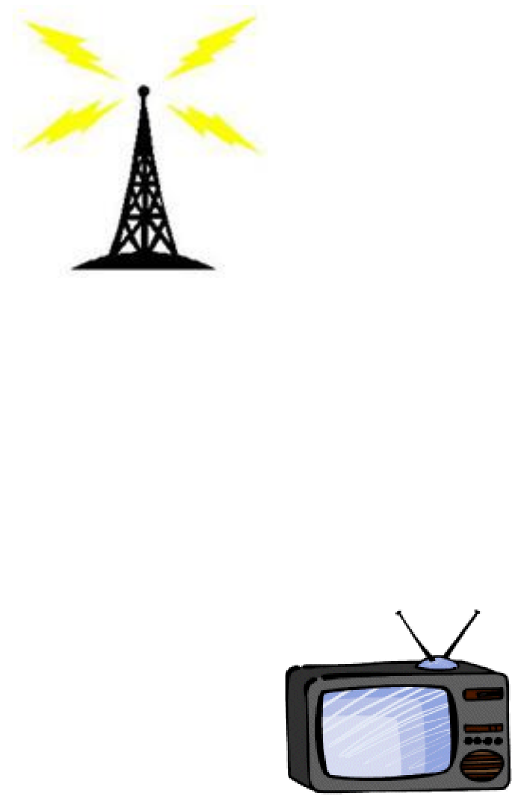
- Is interaction useful? Yes!

Without interaction (one-way):
Inefficient communication

Motivation

- Is interaction useful? Yes!

Without interaction (one-way):
Inefficient communication

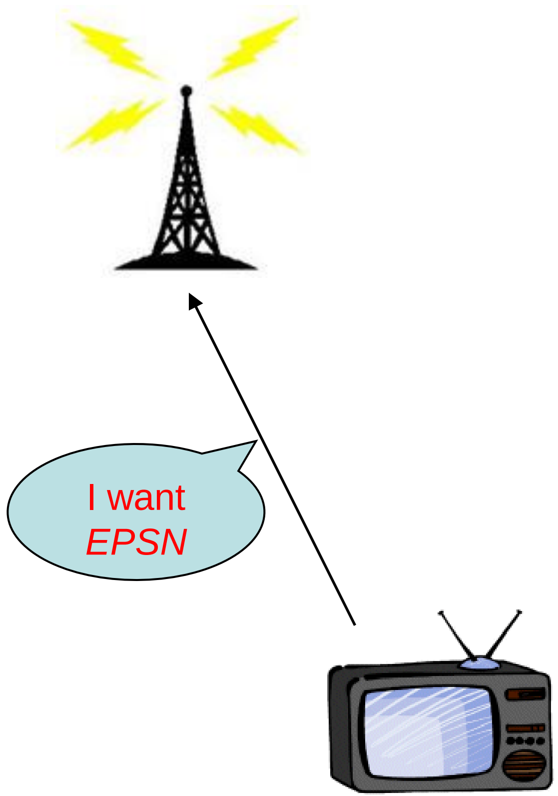


With interaction (two-way):
Efficient communication

Motivation

- Is interaction useful?   Yes!

Without interaction (one-way):
Inefficient communication



With interaction (two-way):
Efficient communication

Motivation

- Is interaction useful? Yes!

Without interaction (one-way):
Inefficient communication

With interaction (two-way):
Efficient communication

Motivation

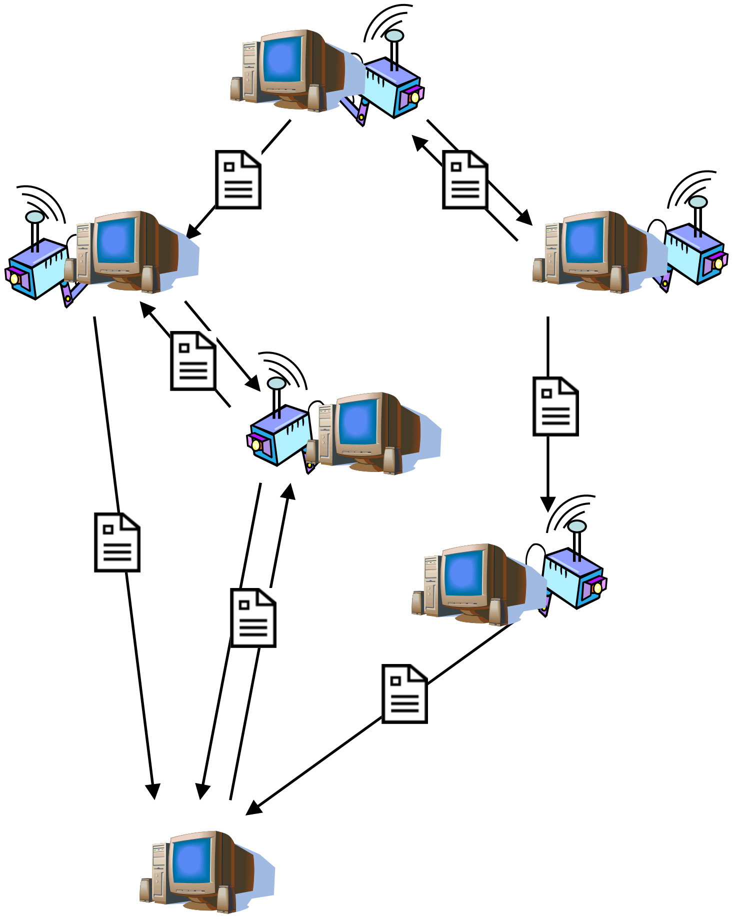
- Under what conditions is interaction useful?

Motivation

- Under what conditions is interaction useful?
- How useful is interaction?

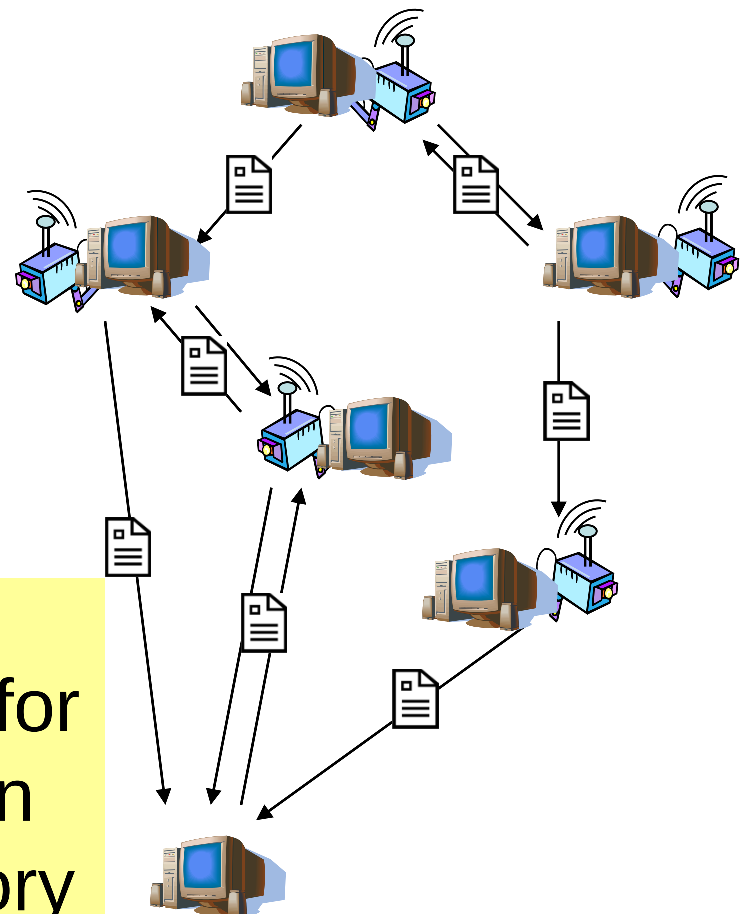
Motivation

- Under what conditions is interaction useful?
- How useful is interaction?
- What is the best way to interact?
  - At *what time*, *who* should send *what* message to *whom*?



Motivation

- Under what conditions is interaction useful?
- How useful is interaction?
- What is the best way to interact?
  - At *what time*, *who* should send *what* message to *whom*?



<u>Main Goal:</u>

Explore the benefit of *interaction* for distributed *function computation* in the framework of information theory

Related work

- Communication complexity (Yao, …, [Kushilevitz & Nisan])

There	Here
Focus on Pr(comp.error) = 0	• Pr(comp.error) $\rightarrow$ 0 as #samples $\rightarrow \infty$ • E [distortion] $\leq D$
Bits	Rate (bits per source sample)

- Two-way source coding [Kaspi'86]

There	Here
Source reproduction	Function computation
No example to show interaction useful	Many examples show interaction useful

- Coding for computing [Orlitsky & Roche'00]

There	Here
Two terminals, two messages	Multiple terminals, t messages

Related work (continued 1)

- Scaling laws of max. rate of computation [Giridhar & Kumar'05]
  - For divisible, type-sensitive, type-threshold function classes
  - In random planar and collocated networks
  - Communication complexity flavor … (Here: distributed block source coding flavor)
  - Pr(compute. error) = 0 … (Here: Pr(comput. error) $\rightarrow$ 0 as #samples $\rightarrow \infty$, and also expected distortion criteria)

- Networks of finite max degree [Subramanian, Gupta, & Shakkottai'07]
  - Further subdivision of type-sensitive function class
  - If allow Pr(sample error) < ε then for some type-sensitive functions like AVERAGE, computation rate increases to type-threshold class

- Acyclic networks [Appuswami, Franceschetti, Karamchandani, & Zeger'07]
  - No interaction over multiple rounds of communication
  - Min-cut bound, tight for divisible functions in multi-edge tree networks
  - Bound not tight in general

Related work (continued 2)

- CEO-style rate-distortion problem [Prabhakaran, Ramchandran, & Tse’04]
  - Multiple rounds of communication
  - Conditioned on desired (hidden) source, observations of agents are independent
  - Lower bound on minimum sum-rate (for given distortion)
  - Bound tight for jointly Gaussian sources and MSE

- Network coding [many refs. too long to list]
  - Mainly non interactive, focus on data dissemination than function computation

- Gossip/Consensus algorithms [many refs]
  - Single sample at each node (zero block-coding rate)
  - Real-valued message exchanges (infinite # bits); With quantization: [Kashyap, Basar, & Srikant]
  - Focus on rate of convergence

- Computation over noisy channels
  - [Nazer & Gastpar]: for specialized classes of “matched” source-channel pairs
  - [Gallager’88], [Ying,Srikant,&Dullerud’07], [Ayaso,Shah,&Dahleh’08]: single sample at each node (zero block-coding rate)

Related work (continued 3)

- Network communication problems with conferencing decoders
- Secrect key agreement problems with public discussion [U. Maurer et al., I.Csiszar, P.Narayanan et al.]
- Feedback problems
- Secure multi-party computation problems [large CS theory literature]
- …

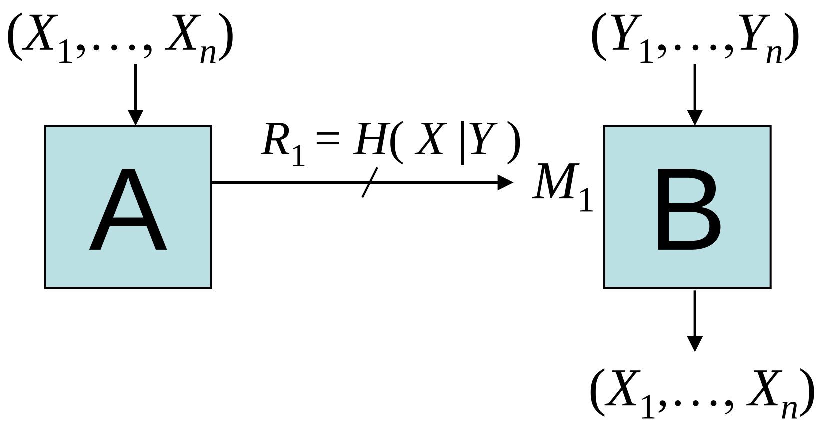
Outline

- Introduction
- General two-terminal problem
- Co-located network with independent sources
- General multi-terminal problem: some observations

BOSTON UNIVERSITY

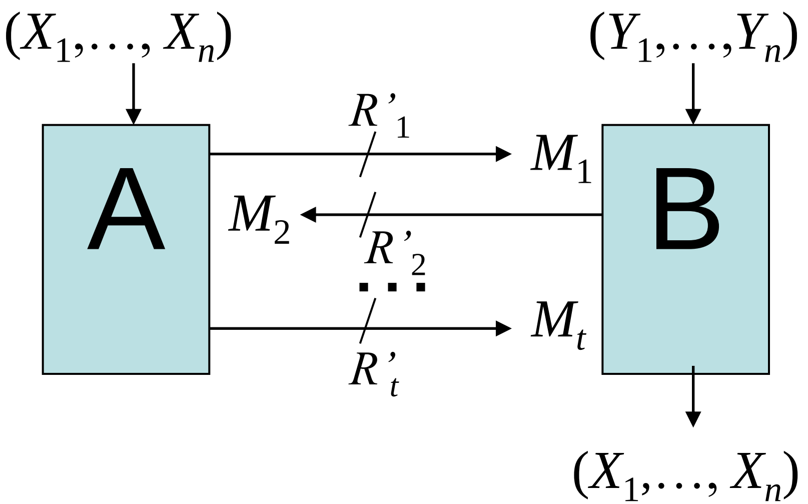
Example 1: Source reproduction

- *Goal:* only B reproduces $(X_1,\ldots,X_n):\ \ f_B(x,y)=x,\ \ f_A(x,y)=0$



Single msg (t =1)

$R_1 = H\,(\,X\,|\,Y\,)$ (Slepian-Wolf coding)



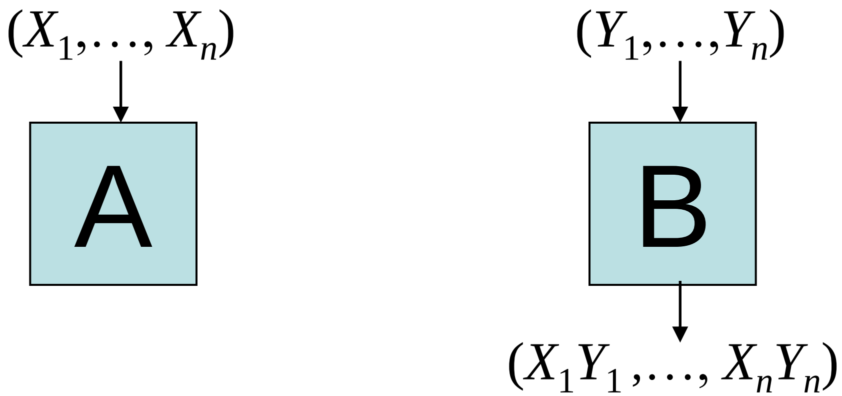
Multiple msgs ($t \geq 1$)

$R'_1+\ldots+R'_t \geq H\,(\,f_B(X,Y)\,|Y\,) = H\,(\,X\,|Y\,)$

- ***No benefit in multiple messages***
- If A also reprod. $\mathbf{Y}$, at least two msgs. No benefit to use $t > 2$ msgs
- *Caveat:* interaction still beneficial if $\Pr(error) = 0$ [Orlitsky, et al] or for faster rate of convergence and for *nonergodic* sources [Da-ke He, et al]

Example 2: Function computation

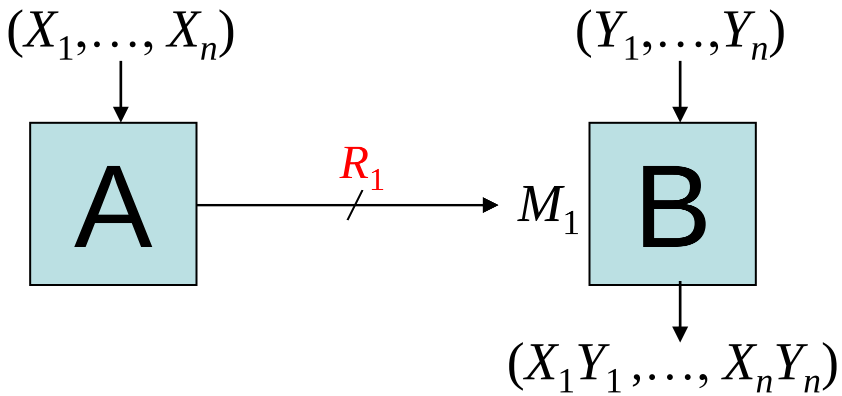
- Indep. sources: $X \sim$ Uniform$\{1, 2, \ldots, L\}$, $Y \sim$ Bernoulli (p)
- Only B reproduces XY



$f_B(x, y)$	$y = 0$	$y = 1$
$x = 1$	0	1
$x = 2$	0	2
…	…	…
$x = L$	0	L

Example 2: Function computation

- Indep. sources: $X \sim$ Uniform$\{1, 2, \ldots, L\}$, $Y \sim$ Bernoulli (p)
- Only B reproduces XY

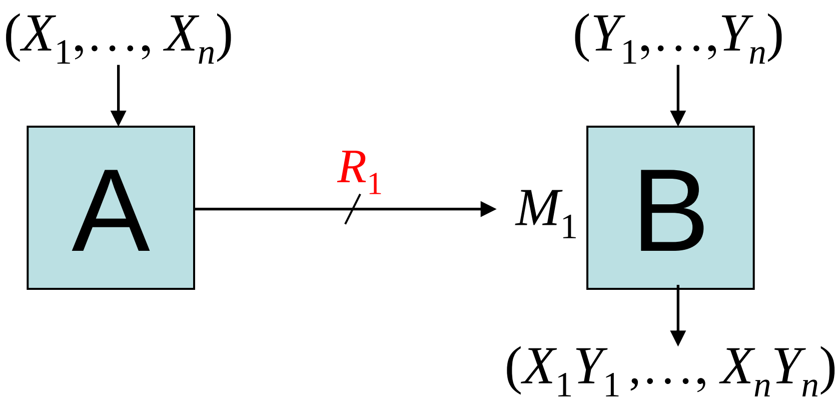


Single msg: $R_1 = ?$

$f_B(x, y)$	$y = 0$	$y = 1$
$x = 1$	0	1
$x = 2$	0	2
…	…	…
$x = L$	0	L

Example 2: Function computation

- Indep. sources: $X \sim$ Uniform$\{1, 2, \ldots, L\}$, $Y \sim$ Bernoulli (p)
- Only B reproduces XY



Single msg: $R_1 = ?$

$f_B(x, y)$	$y = 0$	$y = 1$
$x = 1$	0	1
$x = 2$	0	2
…	…	…
$x = L$	0	L

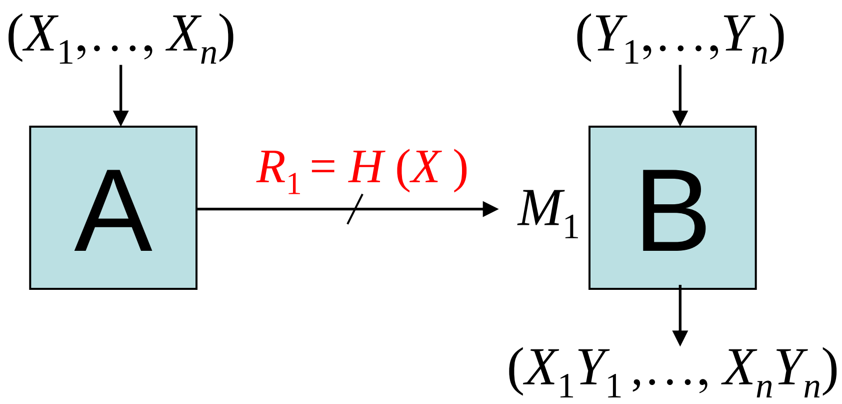
Han & Kobayashi (1987):

If for any (x, y), $p_{XY}(x, y) > 0$ and any two rows are different, then

$R_1 \geq H(X \mid Y)$ no better than sending $\mathbf{X}$ completely!

Example 2: Function computation

- Indep. sources: $X \sim$ Uniform$\{1, 2, \ldots, L\}$, $Y \sim$ Bernoulli (p)
- Only B reproduces XY



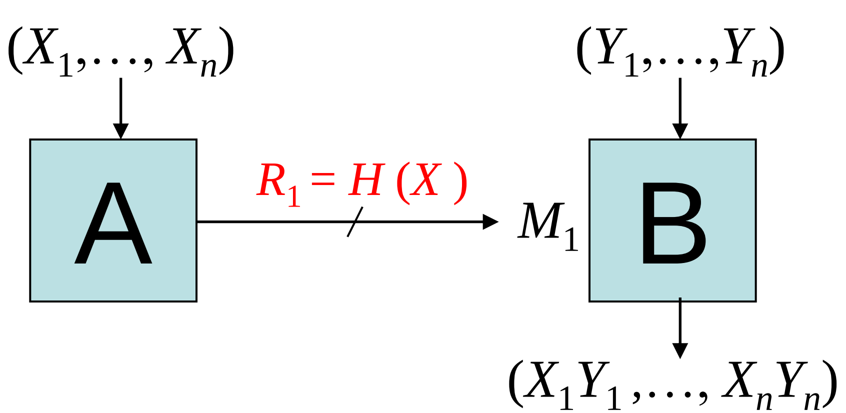
Single msg:

no better than sending **X**

[Han, Kobayashi, 1987]

Example 2: Function computation

- Indep. sources: $X \sim$ Uniform$\{1, 2, \ldots, L\}$, $Y \sim$ Bernoulli (p)
- Only B reproduces XY



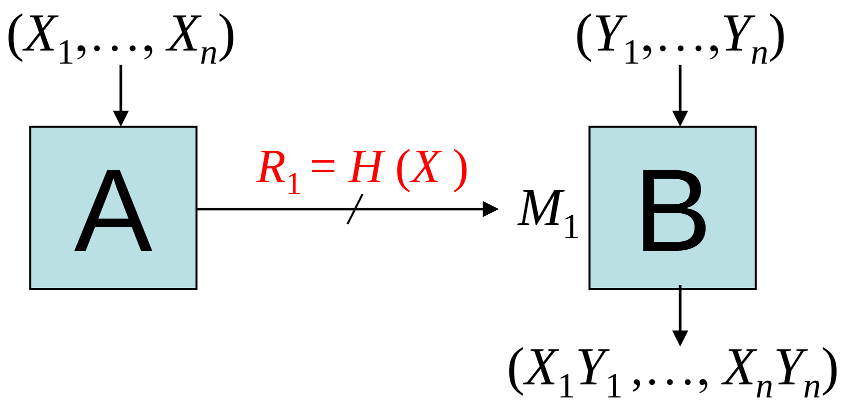
Single msg:

no better than sending **X**

[Han, Kobayashi, 1987]

Example 2: Function computation

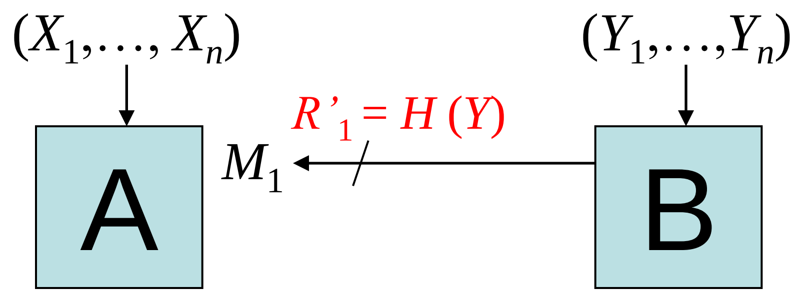
- Indep. sources: $X \sim$ Uniform$\{1, 2, \ldots, L\}$, $Y \sim$ Bernoulli (p)
- Only B reproduces XY



Single msg:

no better than sending **X**

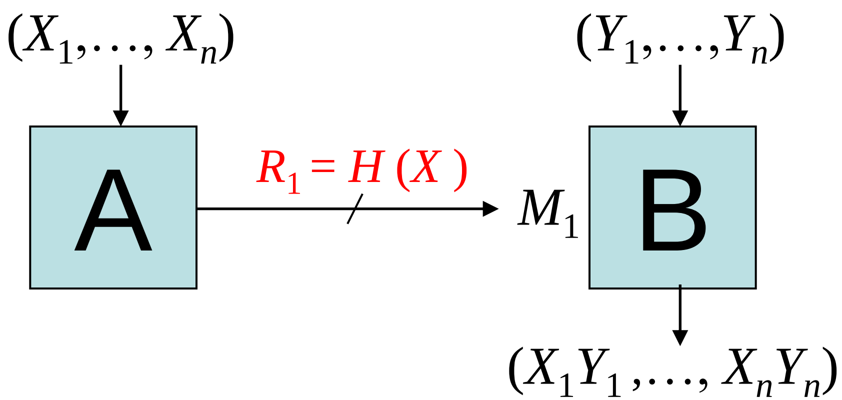
[Han, Kobayashi, 1987]



1st msg: compress **Y**

Example 2: Function computation

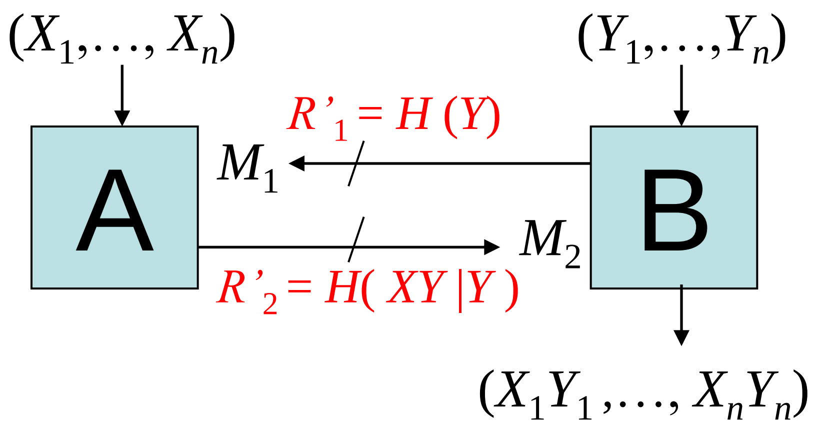
- Indep. sources: X ~ Uniform$\{1, 2, \ldots, L\}$, Y ~ Bernoulli (p)
- Only B reproduces XY



Single msg:

no better than sending **X**

[Han, Kobayashi, 1987]

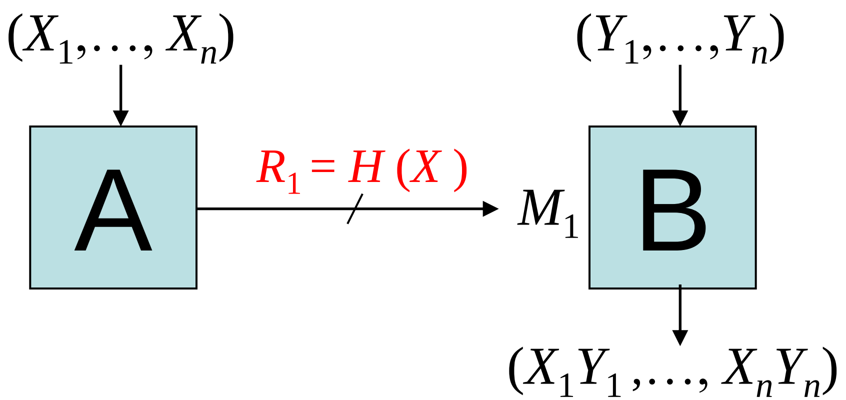


1st msg: compress **Y**

2nd msg: send **X** only if $Y = 1$

Example 2: Function computation

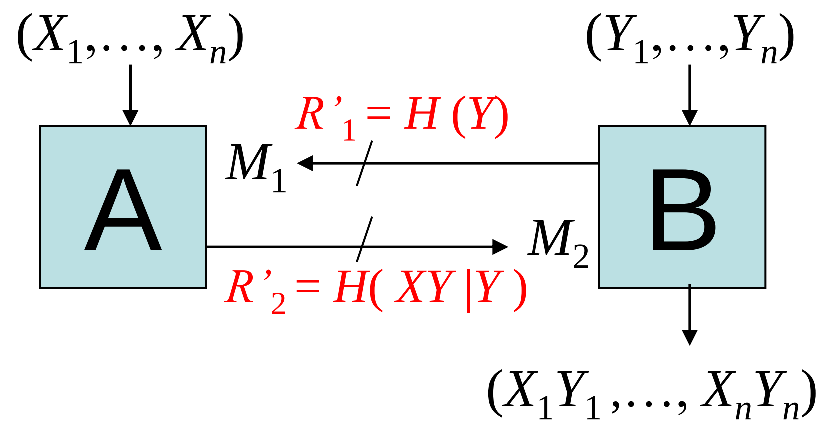
- Indep. sources: X ~ Uniform$\{1, 2, \ldots, L\}$, Y ~ Bernoulli (p)
- Only B reproduces XY



Single msg:

no better than sending **X**

[Han, Kobayashi, 1987]



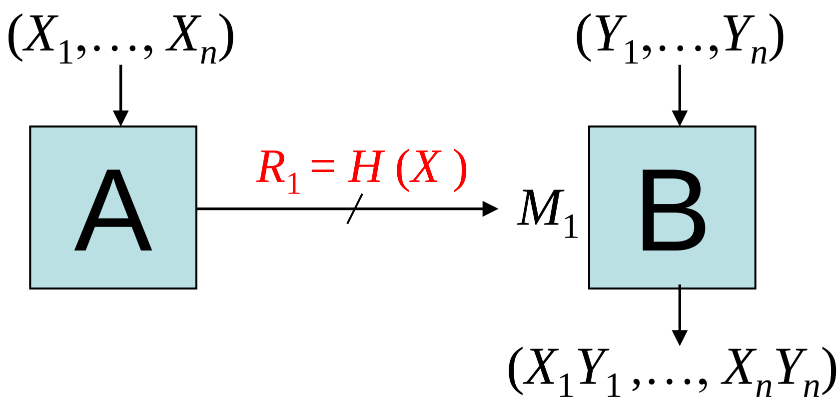
1st msg: compress **Y**

2nd msg: send **X** only if $Y = 1$

$R_1 = \log_2 L$ *strictly* > $R'_1 + R'_2 = h_2(p) + p \log_2 L$

Example 2: Function computation

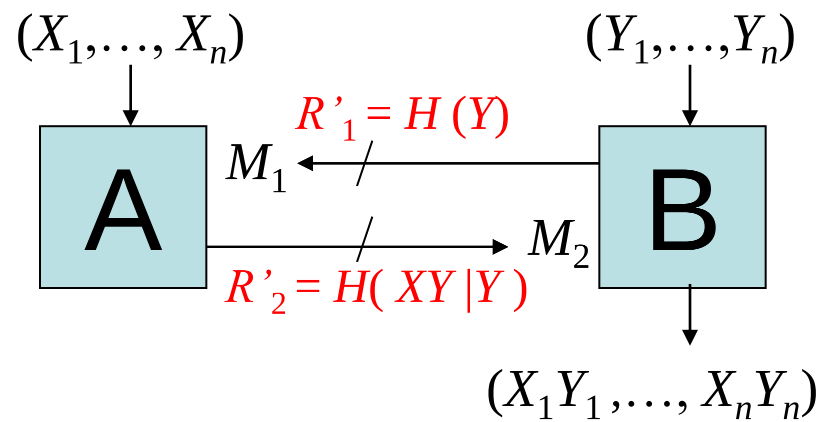
- Indep. sources: $X \sim$ Uniform$\{1, 2, \ldots, L\}$, $Y \sim$ Bernoulli (p)
- Only B reproduces XY



Single msg:

no better than sending **X**

[Han, Kobayashi, 1987]



1[st] msg: compress **Y**

2[nd] msg: send **X** only if $Y = 1$

$R_1 = \log_2 L$ *strictly* > $R'_1 + R'_2 = h_2(p) + p \log_2 L$

- ***Even for indep. sources, interaction gain can be arbitrarily large***

General two-terminal problem

- 2-component DMS source, 2 locations:

  $(X_i, Y_i) \sim \text{iid } p_{XY}$

- Samplewise function computation:

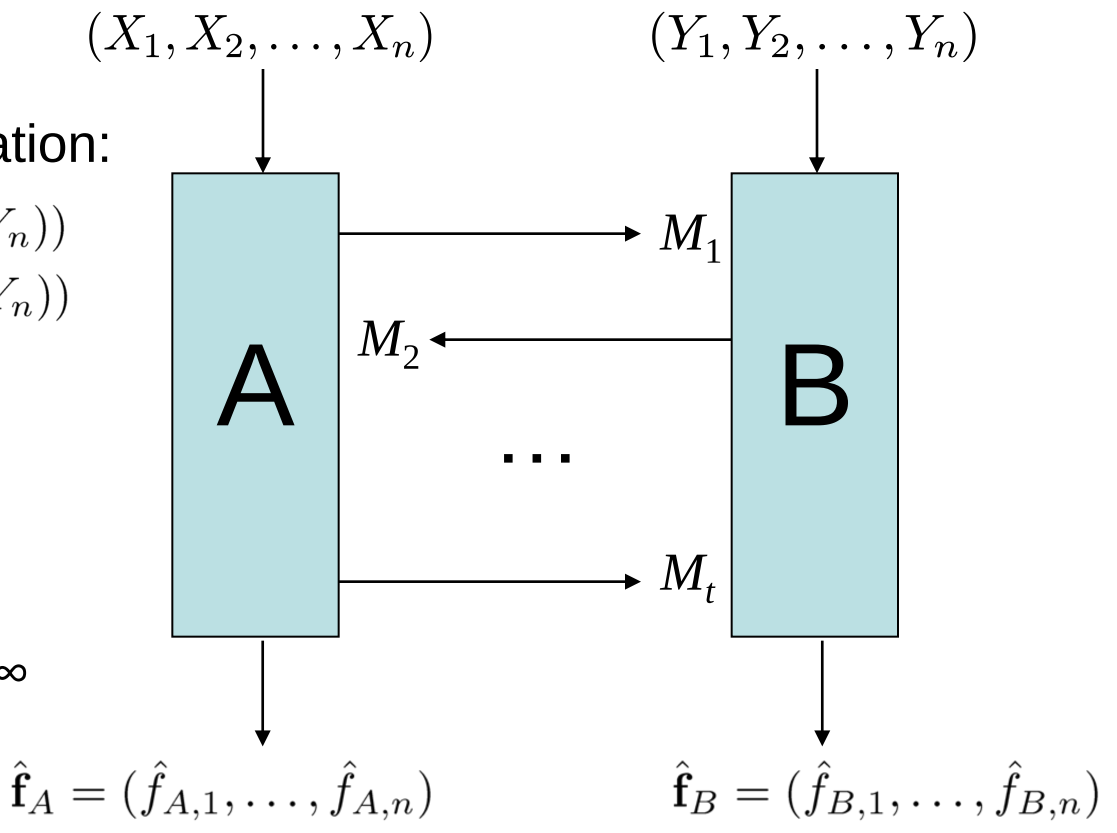
$$\begin{aligned}\mathbf{f}_A &= (f_A(X_1,Y_1),\ldots,f_A(X_n,Y_n)) \\ \mathbf{f}_B &= (f_B(X_1,Y_1),\ldots,f_B(X_n,Y_n))\end{aligned}$$

- t alternating messages

- In this talk, focus on:

  $\Pr$(comp. error) $\rightarrow 0$ as $n \rightarrow \infty$

- Can also handle coupled single-letter distortion



Two-terminal interaction

- Admissible rate-tuple $(R_1, \ldots, R_t)$:

  Exists a sequence of codes:

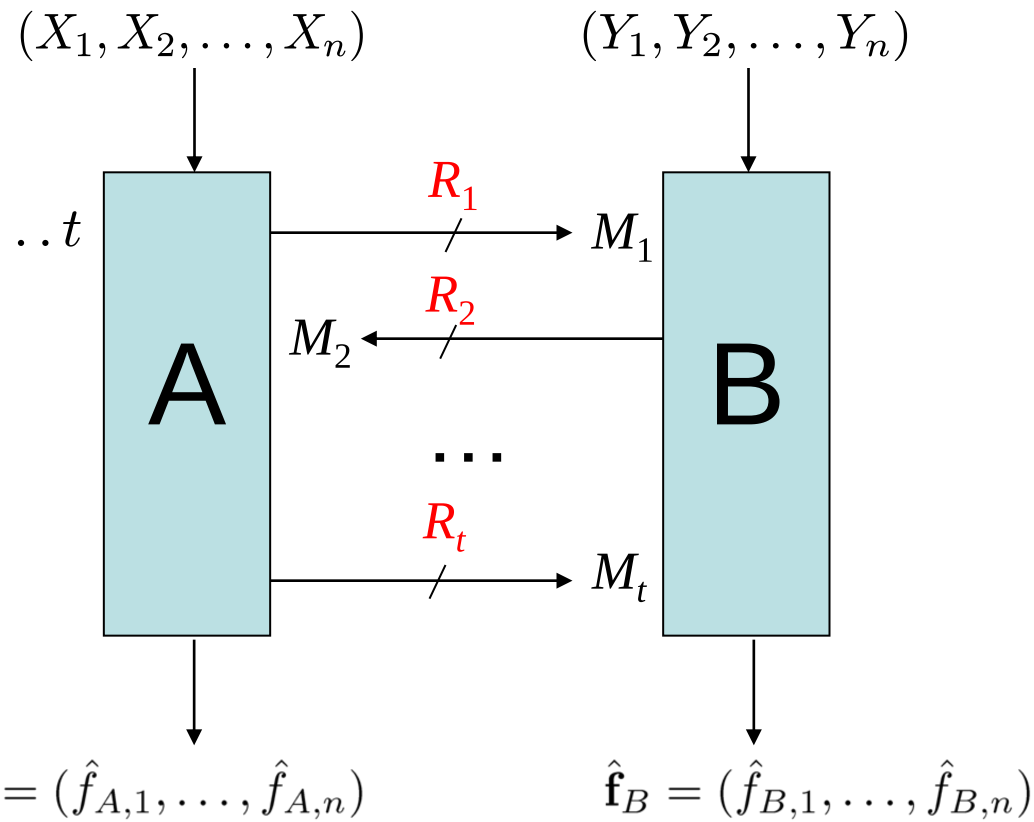
  as $n \to \infty$

$$(\#\text{ bits msg } j)/n \to R_j, \;\; j = 1 \ldots t$$

$$\Pr[(\mathbf{f}_A \neq \hat{\mathbf{f}}_A) \text{ or } (\mathbf{f}_B \neq \hat{\mathbf{f}}_B)] \to 0$$

- Rate region $\mathcal{R}_t^A$:

  set of admissible rate-tuples



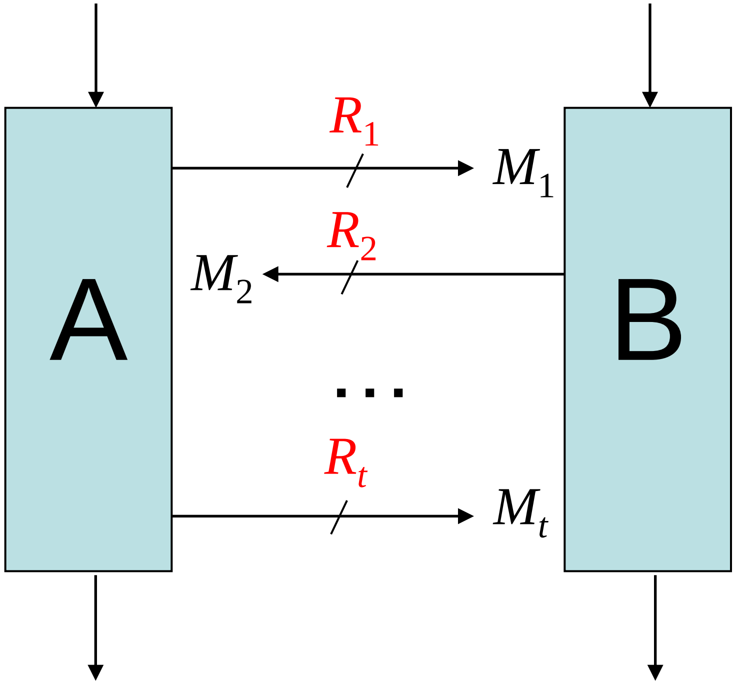
Need:

$$\mathbf{f}_A = (f_A(X_1, Y_1), \ldots, f_A(X_n, Y_n)), \quad \mathbf{f}_B = (f_B(X_1, Y_1), \ldots, f_B(X_n, Y_n))$$

M_1

Two-terminal interaction

$$(X_1, X_2, \ldots, X_n) \qquad (Y_1, Y_2, \ldots, Y_n)$$



$$\hat{\mathbf{f}}_A = (\hat{f}_{A,1}, \ldots, \hat{f}_{A,n}) \qquad \hat{\mathbf{f}}_B = (\hat{f}_{B,1}, \ldots, \hat{f}_{B,n})$$

<u>Goals:</u>

1) Obtain a computable characterization of the rate region (limit-free and independent of n)

2) Understand the benefit of interaction for different sources and functions

Need:

$$\mathbf{f}_A = (f_A(X_1, Y_1), \ldots, f_A(X_n, Y_n)), \quad \mathbf{f}_B = (f_B(X_1, Y_1), \ldots, f_B(X_n, Y_n))$$

Information-theoretic rate region

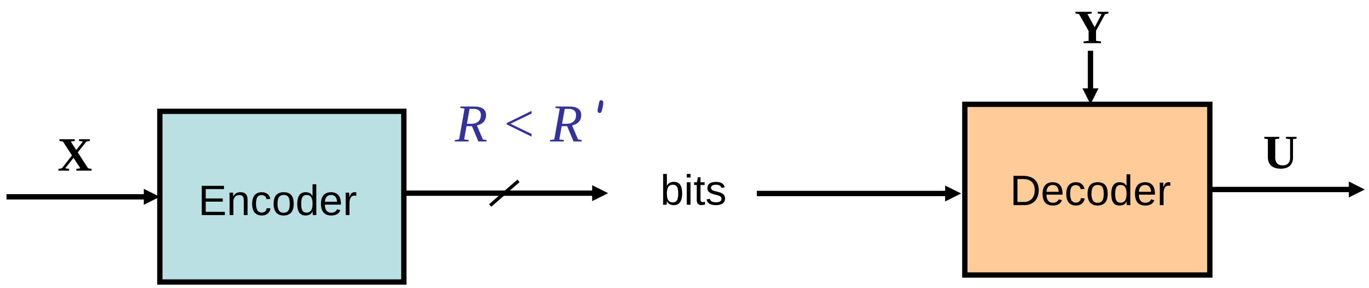
[Nan Ma & PI: ISIT'08, IT'11]

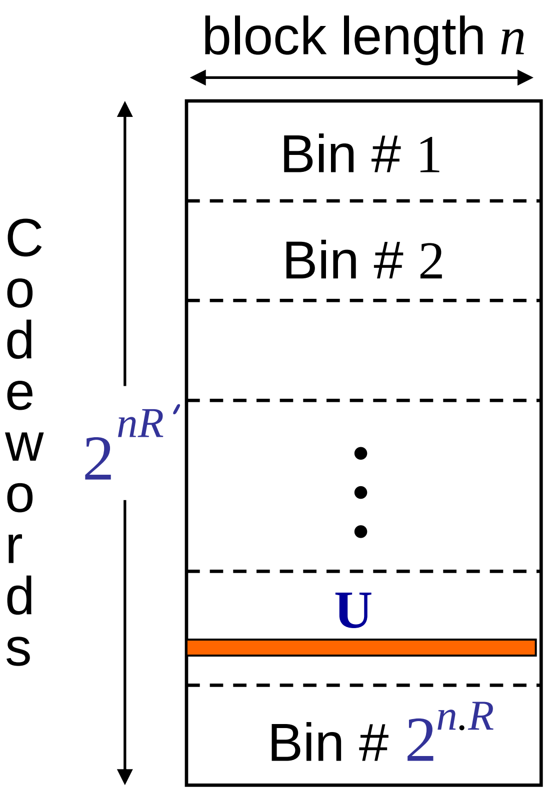
$$\mathcal{R}_t^A = \{ \ (R_1 \ldots R_t) \mid \exists \, U^t, s.t. \ |\mathcal{U}_i| \cdot \ \ g_i(|\mathcal{X}|, |\mathcal{Y}|),$$

$$R_i \geq \begin{cases} I(X;U_i|Y,U^{i-1}), \ U_i - (X,U^{i-1}) - Y, & i \text{ odd} \\ I(Y;U_i|X,U^{i-1}), \ U_i - (Y,U^{i-1}) - X, & i \text{ even} \end{cases}$$

$$H(f_A(X,Y)|X,U^t) = 0, \ H(f_B(X,Y)|Y,U^t) = 0 \ \}$$

Wyner-Ziv coding





$R' \geq I(X;U)$ $\qquad$ $R \geq I(X;U \mid Y)$, U - Y - X

- $2^{nR'}$ codewords split into 2^{nR} bins
- Encoder sends only bin-index => rate reduced from R' to R but extra confusion for decoder
- Decoder uses **Y** to disambiguate entries in bin
- Encoding and decoding: find statistically most consistent codeword
- Notion of decoding failure

Information-theoretic rate region

$$\mathcal{R}_t^A = \{\ (R_1 \ldots R_t) \mid \exists\, U^t, s.t.\ |\mathcal{U}_i| \cdot \quad g_i(|\mathcal{X}|, |\mathcal{Y}|),$$

$$R_i \geq \begin{cases} I(X; U_i | Y, U^{i-1}),\ U_i - (X, U^{i-1}) - Y, & i \text{ odd} \\ I(Y; U_i | X, U^{i-1}),\ U_i - (Y, U^{i-1}) - X, & i \text{ even} \end{cases}$$

$$H(f_A(X,Y)|X, U^t) = 0,\ H(f_B(X,Y)|Y, U^t) = 0\ \}$$

Achievability (sequence of Wyner-Ziv codes):

- 1st msg: Quantizes $\mathbf{X}$ to $\mathbf{U}_1$ with side info $\mathbf{Y}$

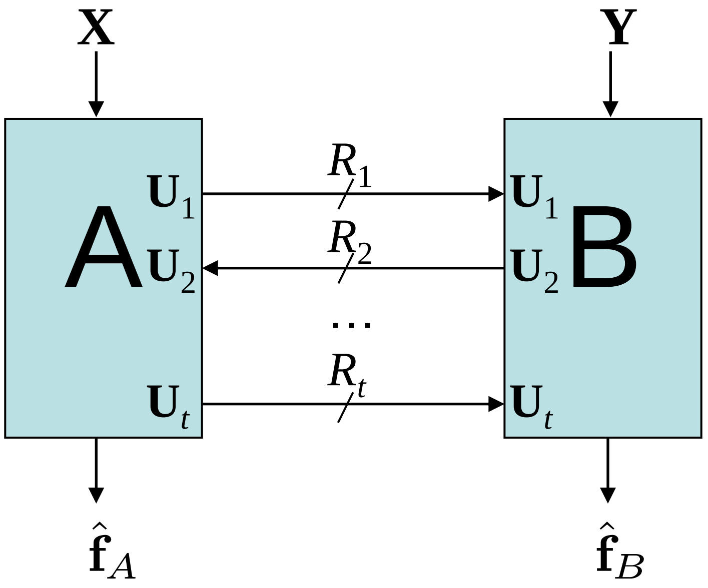
  $$R_1 = I(X; U_1|Y), \quad U_1 - X - Y$$

- 2nd msg: Quantizes $(\mathbf{Y}, \mathbf{U}_1)$ to $\mathbf{U}_2$ with side info $(\mathbf{X}, \mathbf{U}_1)$

  $$R_2 = I(Y; U_2|X, U_1), \quad U_2 - (Y, U_1) - X$$

  ....

- Recover $\mathbf{f}_A$ based on $(\mathbf{X}, \mathbf{U}_1 \ldots \mathbf{U}_t)$: $H(f_A|X, U_1 \ldots U_t) = 0$
- Recover $\mathbf{f}_B$ based on $(\mathbf{Y}, \mathbf{U}_1 \ldots \mathbf{U}_t)$: $H(f_B|Y, U_1 \ldots U_t) = 0$



Information-theoretic rate region

$$\mathcal{R}_t^A = \{ \ (R_1 \dots R_t) \mid \exists\, U^t, s.t.\ |\mathcal{U}_i| \cdot \ \ g_i(|\mathcal{X}|, |\mathcal{Y}|),$$
$$R_i \geq \begin{cases} I(X; U_i | Y, U^{i-1}),\ U_i - (X, U^{i-1}) - Y, & i \text{ odd} \\ I(Y; U_i | X, U^{i-1}),\ U_i - (Y, U^{i-1}) - X, & i \text{ even} \end{cases}$$
$$H(f_A(X,Y)|X,U^t) = 0,\ H(f_B(X,Y)|Y,U^t) = 0 \ \}$$

Converse (impossible to do better):

- Standard information inequalities
- Auxiliary random variables

$$U_{1,i} = (M_1, X_1, \dots, X_{i-1}, Y_{i+1}, \dots, Y_n)$$
$$U_2 = M_2,\ \dots,\ U_t = M_t$$

- Cardinality bounds on alphabets of auxiliary random variables

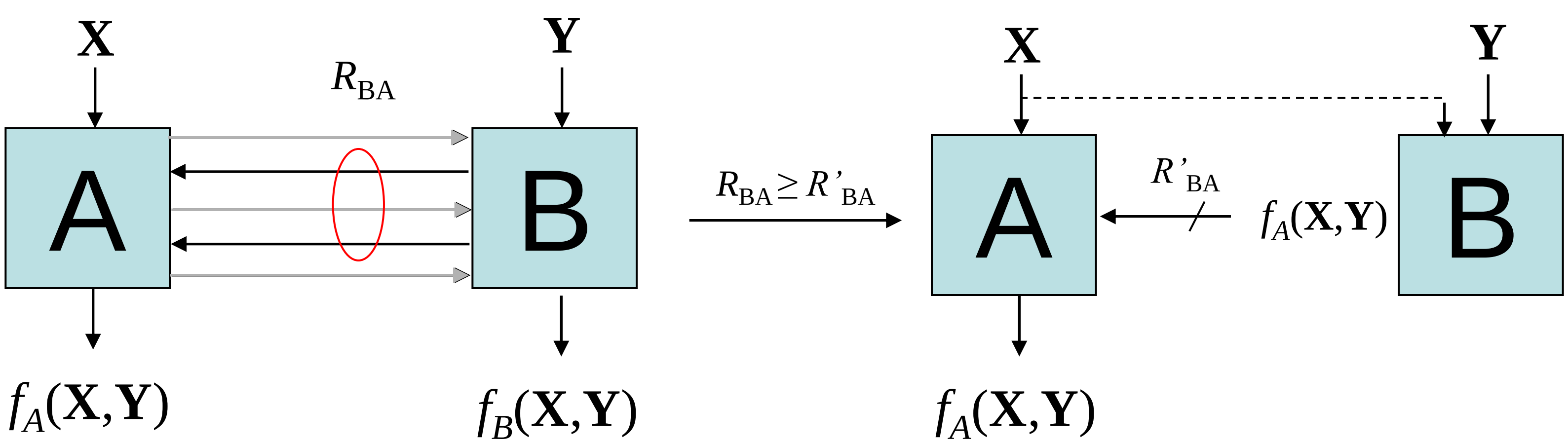
Minimum sum-rate

- *t*-msg min sum-rate: $R_{sum,t} = \min I(X;U^t|Y) + I(Y;U^t|X)$

aux. r. v. subject to

$$\begin{aligned} & U_i - (X, U^{i-1}) - Y, \ \ i \text{ odd} \\ & U_i - (Y, U^{i-1}) - X, \ \ i \text{ even} \\ & H(f_A(X,Y)|X,U^t) = 0 \\ & H(f_B(X,Y)|Y,U^t) = 0 \end{aligned}$$

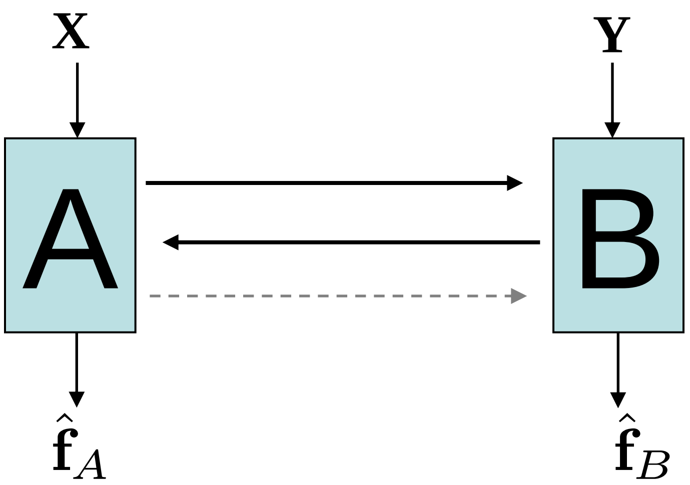
- <u>*Genie lower-bound:*</u> $R_{sum,t} \geq H(f_A(X,Y)|X) + H(f_B(X,Y)|Y)$



$$R_{BA} \geq R'_{BA} \geq H(f_A(X,Y)|X)$$

Minimum sum-rate

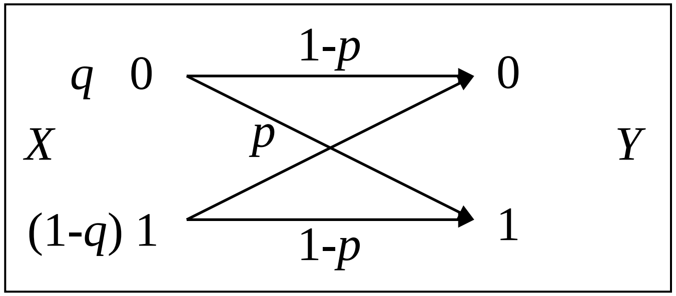
- $R_{sum,1} \geq R_{sum,2} \geq R_{sum,3} \geq \ldots \geq R_{sum,\infty}$
  - Each message could be a null message

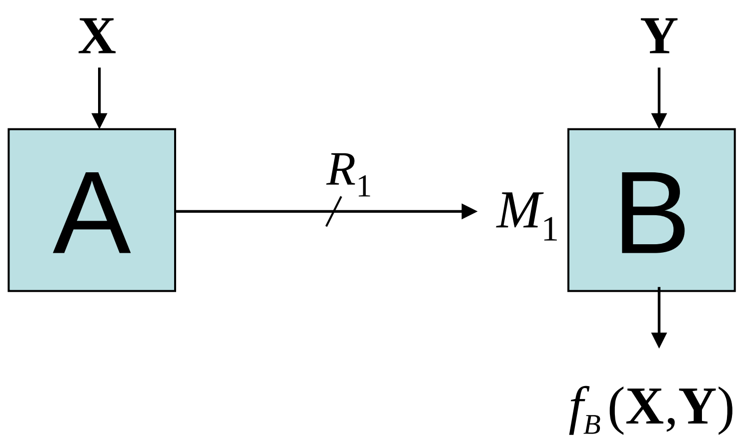


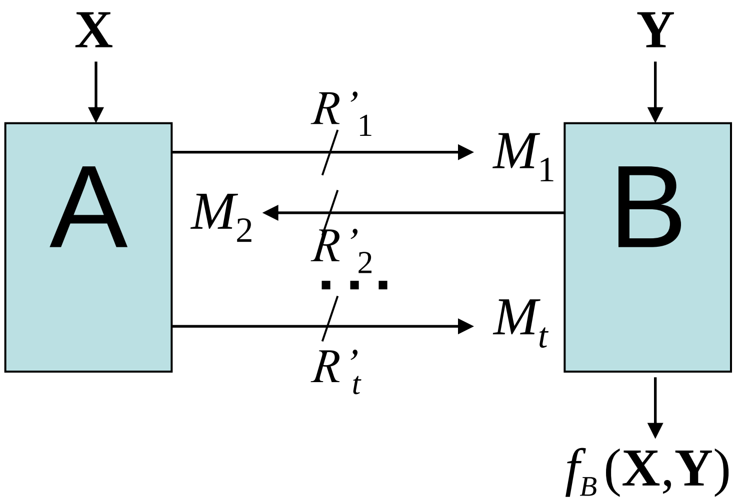
- For all finite t, $R_{sum,t}$ computable; $R_{sum,\infty}$ not.
- *Recent result:* a new functional characterization of $R_{sum,\infty}$
- *Opens new dimension of investigation:* message asymptotics with infinitesimal rate messages

Example 3: Effect of Distribution

- Correlated binary sources:
- Only B reprod. samplewise $f_B(\mathbf{X},\mathbf{Y})$







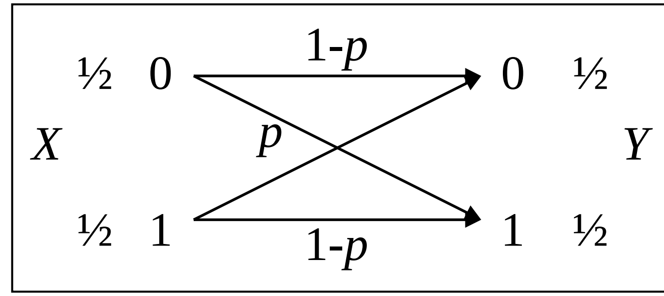
<u>Theorem:</u> *for <u>any</u> function* $f_B(x, y)$:

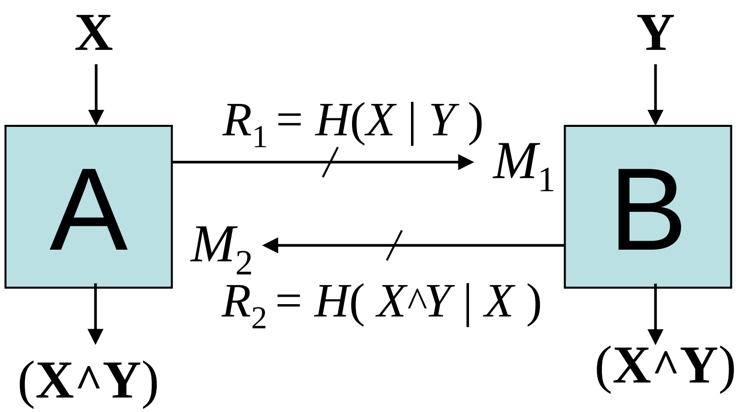
$$\min R_1 \quad = \quad \min\,(R'_1 + \ldots + R'_t)$$

- ***<u>Even ∞-msg interaction is not better than one-msg. comm.</u>***

Example 4: Effect of Demand

- Doubly symmetric binary sources ($q = 1/2$)
- Both sides reproduce **X^Y** (Boolean AND)

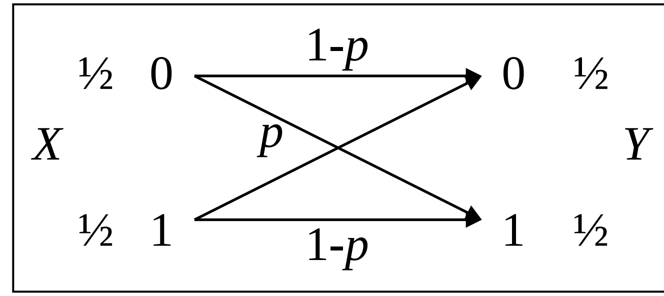


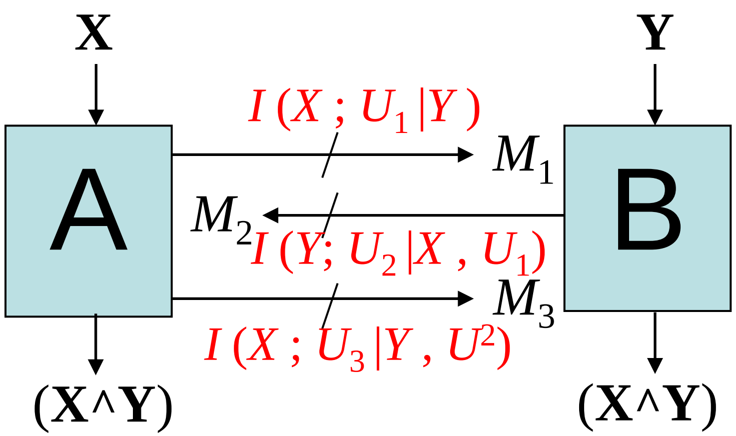


2 msgs

Example 4: Effect of Demand

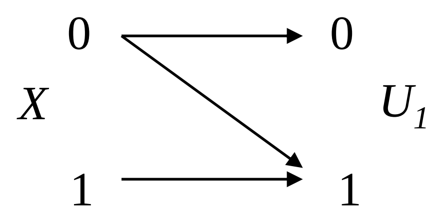
- Doubly symmetric binary sources ($q = 1/2$)
- Both sides reproduce **X^Y** (Boolean AND)





3 msgs

$\mathbf{U_1}$: part of zeros of **X**



$\mathbf{U_2}$: all of zeros of **Y**

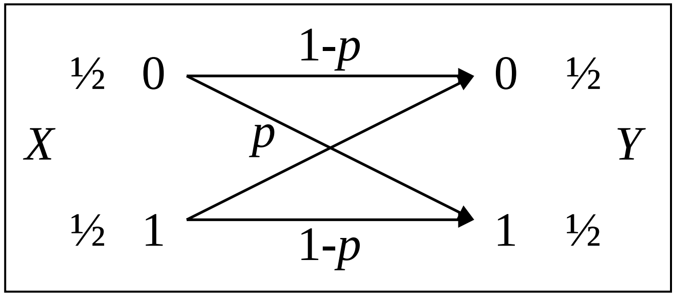
$$U_2 = Y \wedge U_1$$

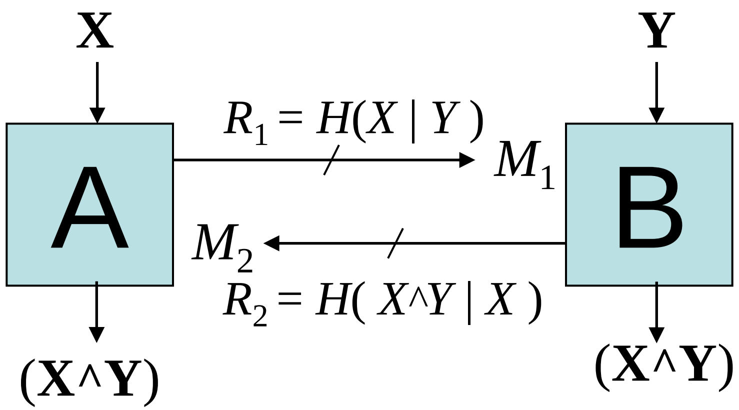
$\mathbf{U_3}$: **X^Y**

$$U_3 = X \wedge U_2 = X \wedge Y$$

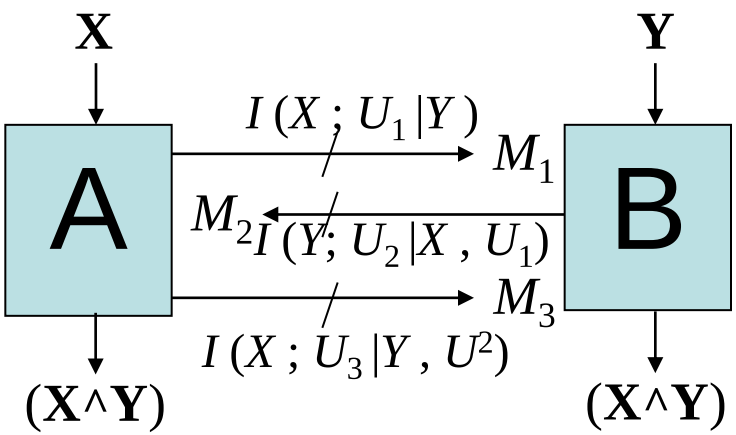
Example 4: Effect of Demand

- Doubly symmetric binary sources ($q = 1/2$)
- Both sides reproduce **X^Y** (Boolean AND)





2 msgs



3 msgs: strictly better than 2 msgs

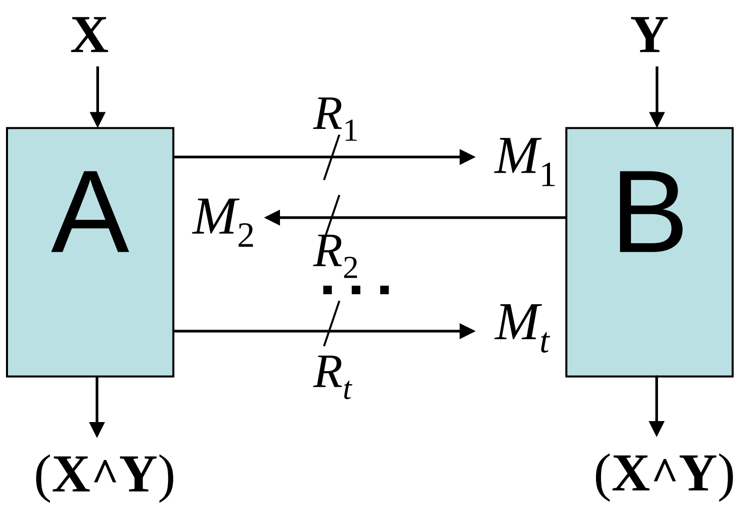
- E.g, $p = ½$, X, $Y \sim$ iid Ber (½), 2-msg: 1.5 vs 3-msg: 1.406
- ***<u>3 messages are better than 2 (interaction does help here)</u>***

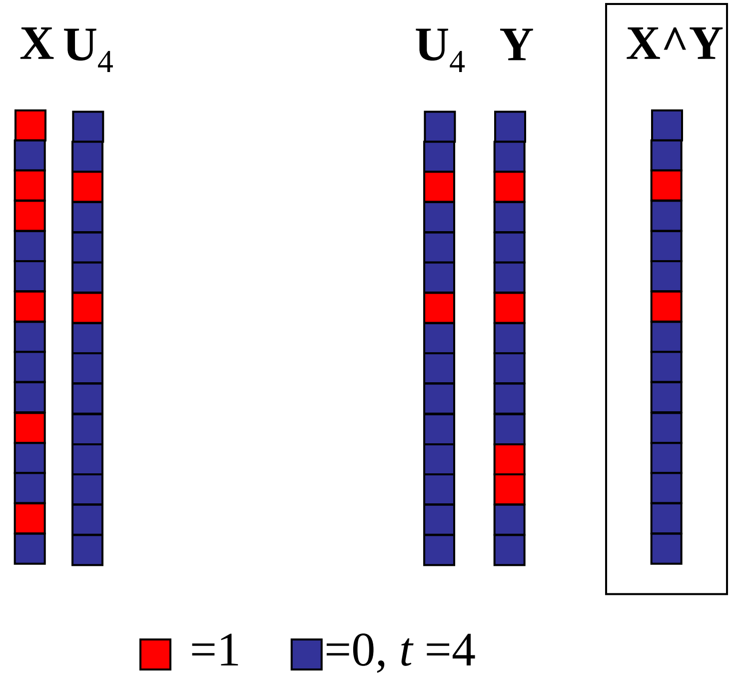
Example 5: ∞-msg interaction

- Independent (X , Y), $X \sim$ Ber (p), $Y \sim$ Ber (q)
- Both sides reproduce **X^Y**
- ∞-msg minimum sum-rate:

For $p = q = ½$, X, $Y \sim$ iid Ber(½),
∞-msg: 1.36 vs. 2-msg: 1.5 and 3-msg: 1.41

$$h_2(p) + p\, h_2(q) + p \log_2 q + p\,(1-q) \log_2 e \qquad (q > p)$$





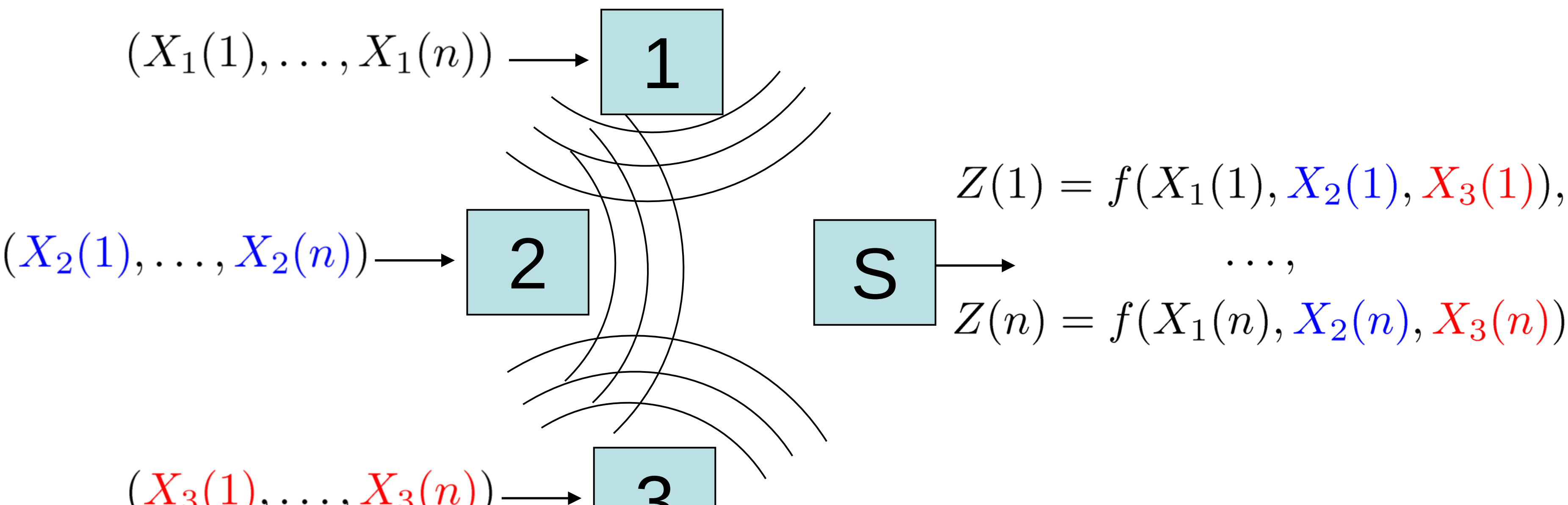
Outline

- Introduction
- General two-terminal problem
- Co-located network with independent sources
- General multi-terminal problem: some observations

BOSTON UNIVERSITY

Collocated network

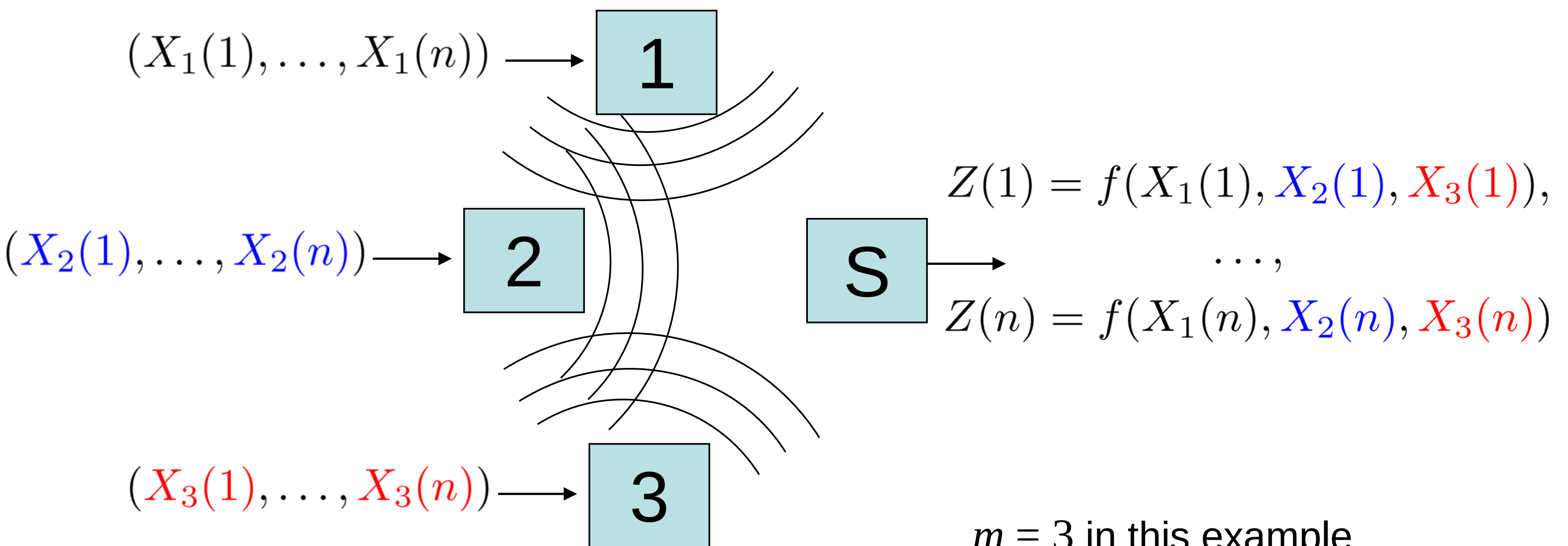
- Consider:
  - m sensors. Each observes n samples of a source
  - A sink needs to compute a samplewise function
  - A sequence of (noiseless) broadcasts for r rounds
  - How many bits/sample for each message? In total?



m = 3 in this example

Collocated network

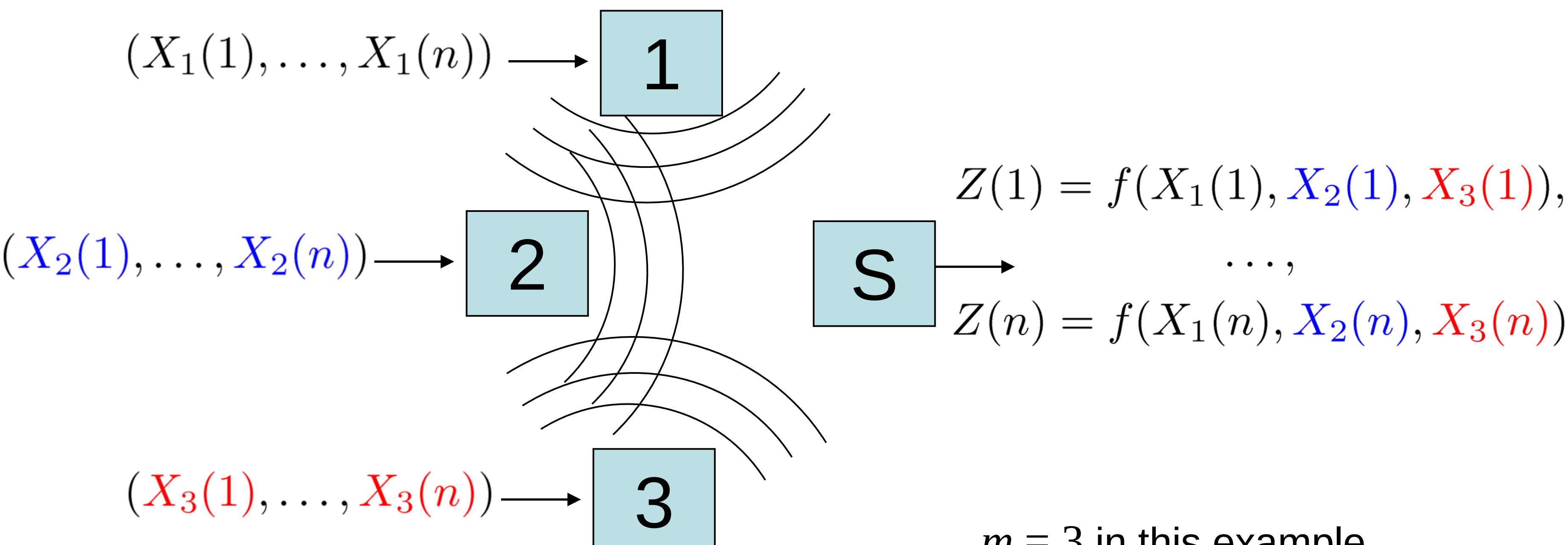
- Sources: iid across samples ~ p_{X_i}, independent across sensors
- Samplewise function: $Z(i) = f(X_1(i), \ldots, X_m(i)),\ i = 1, \ldots, n$
- Broadcasting for r rounds, $t = mr$ msgs: $(1,\ldots,m,\ \ 1,\ldots,m,\ \ \ldots,\ \ 1,\ldots,m)$
- Pr(computation result vector ≠ correct function vector) → 0 as $n \to \infty$



$m = 3$ in this example

Collocated network

- Operational rate R_i (in bits/sample): (# bits msg. i)$/n \to R_i$, as $n \to \infty$
- Rate region $\mathcal{R}_r$: set of all operational $(R_1, R_2, \ldots, R_t)$
- Minimum sum-rate: $R_{sum,r} = \min \left(R_1 + \ldots + R_t\right)$

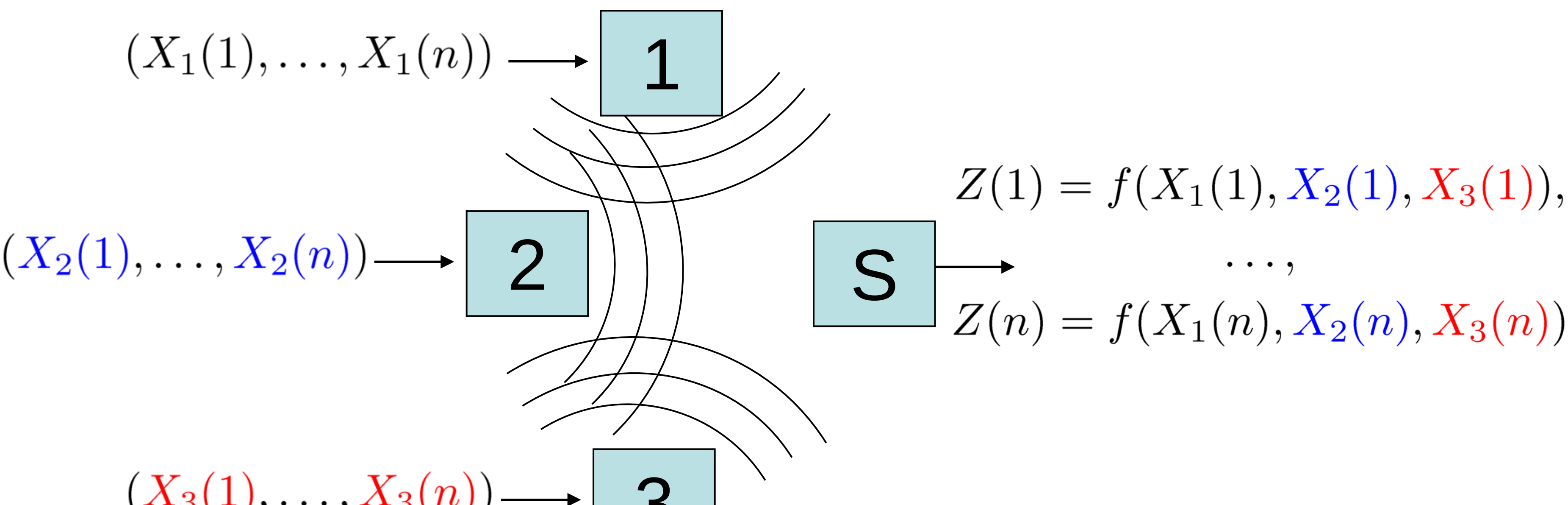


m = 3 in this example

Collocated network

Goals:

- Obtain a computable characterization of $\mathcal{R}_r$ (independent of *n*)
- Scaling behavior of $R_{sum,r}$ w.r.t. *m* (# sensors)
- Understand the benefit of interaction for different sources and functions



m = 3 in this example

Information-theoretic rate region

A characterization independent of n :

$\{1, 2, \ldots, t\}$

$$
\begin{aligned}
\mathcal{R}_r \quad = \quad & \{ \ (R_1 \ldots R_t) \mid \exists \ U^t, s.t. \ \forall j \in [1, t], k = (j \bmod m), \\
& R_j \geq I(X_k; U_j | U^{j-1}), \\
& U_j - (U^{j-1}, X_k) - (X^{k-1}, X_{k+1}^m), \\
& H(f(X^m)|U^t) = 0 \ \}
\end{aligned}
$$

[Nan Ma, PI, and P.Gupta: ISIT'09]

Information-theoretic rate region

$$
\begin{aligned}
\mathcal{R}_r \quad = \quad & \{\ (R_1 \ldots R_t) \mid \exists\, U^t, s.t.\ \forall j \in [1,t], k = (j \bmod m), \\
& R_j \geq I(X_k; U_j | U^{j-1}), \\
& U_j - (U^{j-1}, X_k) - (X^{k-1}, X_{k+1}^m), \\
& H(f(X^m)|U^t) = 0\ \}
\end{aligned}
$$

Achievability:

- 1st msg: Sensor-1 quantizes $\mathbf{X}_1$ to $\mathbf{U}_1$ and broadcasts $\mathbf{U}_1$

$$R_1 = I(X_1; U_1), \quad U_1 - X_1 - X_2^m$$

- 2nd msg: Sensor-2 quantizes $\mathbf{X}_2$ to $\mathbf{U}_2$ with side info $\mathbf{U}_1$ available to every node, and broadcasts $\mathbf{U}_2$ (conditional coding)

$$R_2 = I(X_2; U_2 | U_1), \quad U_2 - (U_1, X_2) - (X_1, X_3^m)$$

....

- Recover $\mathbf{Z}$ based on $(\mathbf{U}_1 \ldots \mathbf{U}_t)$: $\quad H(f(X^m)|U_1 \ldots U_t) = 0$

Information-theoretic rate region

$$
\begin{aligned}
\mathcal{R}_r \quad = \quad & \{\ (R_1 \ldots R_t) \mid \exists\ U^t, s.t.\ \forall j \in [1,t], k = (j \bmod m), \\
& R_j \geq I(X_k; U_j | U^{j-1}), \\
& U_j - (U^{j-1}, X_k) - (X^{k-1}, X_{k+1}^m), \\
& H(f(X^m)|U^t) = 0\ \}
\end{aligned}
$$

Converse *(impossible to do better)*:

- Standard information inequalities
- Auxiliary random variables

$$U_1(i) = (M_1, X_1(1), \ldots, X_1(i-1), \ldots, X_m(1), \ldots, X_m(i-1))$$
$$U_2 = M_2,\ \ldots,\ U_t = M_t$$

- Cardinality bounds on alphabets of auxiliary random variables

Minimum sum-rate

$$\mathcal{R}_r = \{ (R_1 \dots R_t) \mid \exists\, U^t, s.t.\ \forall j \in [1,t], k = (j \bmod m),$$
$$R_j \geq I(X_k; U_j | U^{j-1}),$$
$$U_j - (U^{j-1}, X_k) - (X^{k-1}, X_{k+1}^m),$$
$$H(f(X^m)|U^t) = 0 \}$$

Minimum sum-rate:

$$R_{sum,r} = \min_{U^t} I(X^m; U^t)$$

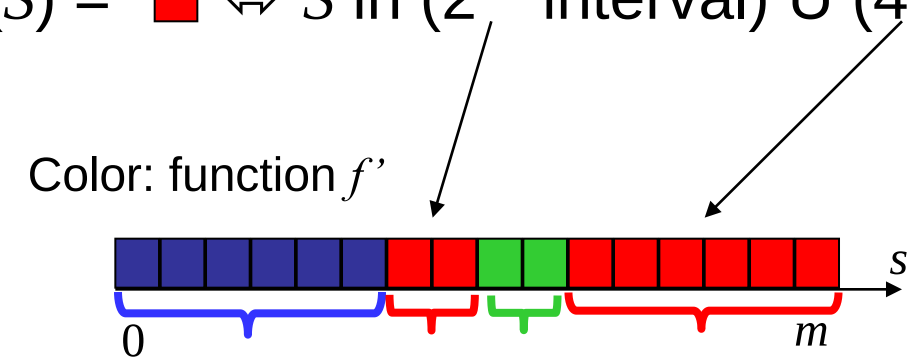
aux. r.v. subject to

$$\forall j \in [1,t], k = (j \bmod m),$$
$$U_j - (U^{j-1}, X_k) - (X^{k-1}, X_{k+1}^m),$$
$$H(f(X^m)|U^t) = 0 \}$$

Computing symmetric functions of binary sources

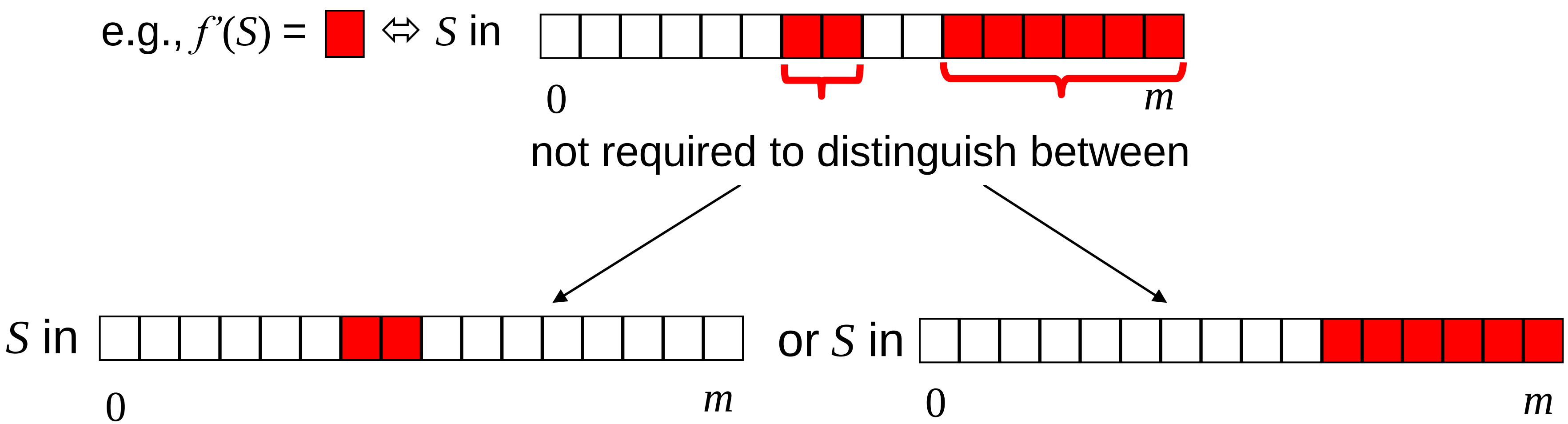
- Indep. Bernoulli sources: $Pr(X_i = 1) = p_i \in (0,1), Pr(X_i = 0) = 1 - p_i$
- Symmetric functions:
  - Invariant to permutations of arguments
  - Functions of $S = \sum_{i=1}^{m} X_i$ for binary sources $f(X^m) = f'(S)$
  - Maximal f'-monochromatic intervals: $\{ [a, b] \}$
  - Computing f ⇔ Locating S in a union of max f'-monochromatic intervals

e.g. $f'(S)$ = [red] ⇔ S in (2nd interval) U (4th interval)



Computing symmetric functions of binary sources

Computing f ⇔ Locating S in a union of several max f'-monochromatic intervals



However, due to the structure of the multiround code,
can inevitably distinguish between these cases

Lemma 2(i):

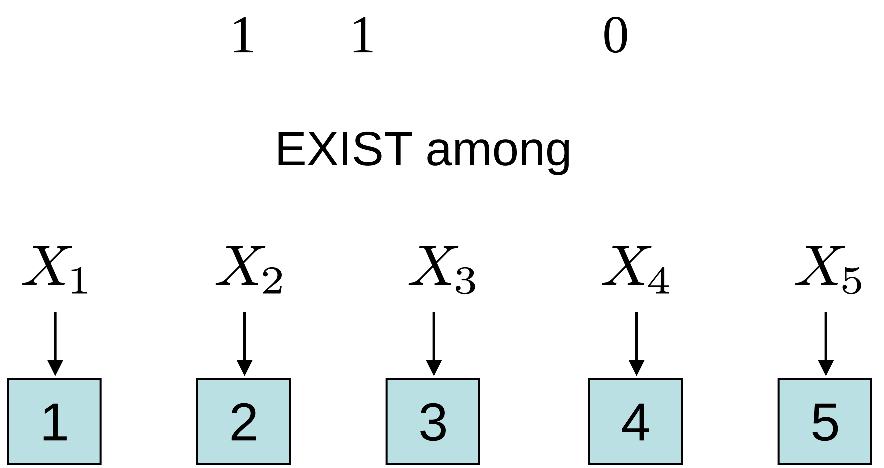
Given U^t, with probability one, there exists a single max f'-monochromatic interval to which S belongs.

Computing symmetric functions of binary sources

S in $[a, b]$ ⇔ existence of a 1's and $(m\text{-}b)$ 0's in X^m

e.g., if $m = 5$, then S in $[2, 4]$ ⇔ at least two 1's and one 0 in X^5

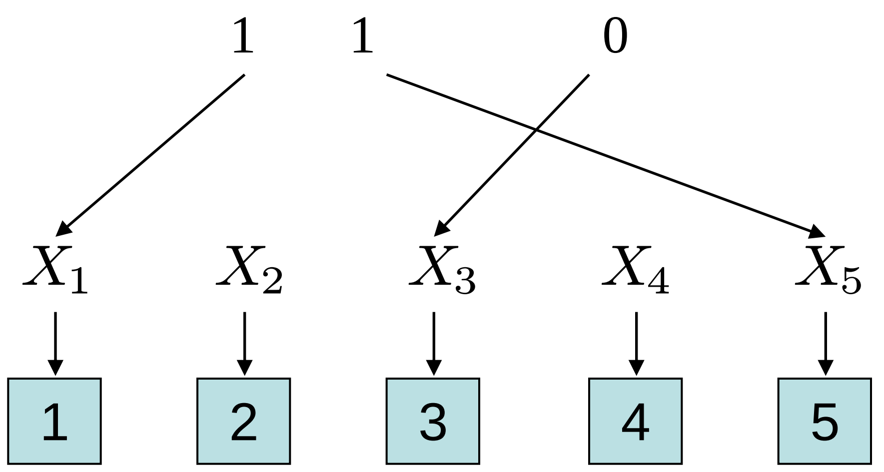
Not required to learn which X's are 1's and which are 0's



Computing symmetric functions of binary sources

S in $[a, b]$ ⇔ existence of a 1's and $(m\text{-}b)$ 0's in X^m

e.g., if $m = 5$, then S in $[2, 4]$ ⇔ at least two 1's and one 0 in X^5



However, due to the structure of the multiround code, will inevitably learn a X's which are 1 and $(m\text{-}b)$ X's which are 0.

Lemma 2(ii):

Given U^t, with probability one, can identify a X's which are 1 and $(m\text{-}b)$ X's which are 0.

Computing symmetric functions of binary sources

Lemma for single-letter characterization

(holds with Prob = 1)

Lemma 2: *Given* U^t,

(i) S *is in a single interval* $[a, b]$,

(ii) Can identify a X*'s which are* 1 *and* $(m\text{-}b)$ X*'s which are* 0.

Operational block-coding counterpart

(holds with Prob > 1 – Pr(blk. error))

Lemma 3: *Given any message sequence, for each sample* i,

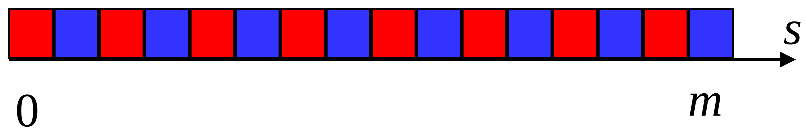
(i) Sink can identify $S(i)$ *within a single interval* $[a_i, b_i]$,

(ii) Sink can identify a_i *sensors observing* 1 *and* $(m\text{-}b_i)$ *sensors observing* 0.

Example: PARITY

m max monochromatic intervals $\{ [0, 0], [1, 1], \ldots [m, m] \}$

Color: function f'



For any zero-error code (Pr(blk. error) = 0), for each sample i:

1. Given the messages, the sink can identify $S(i)$ within a single monochromatic interval ⇔ The sink knows $S(i)$ exactly

2. If $S(i)$ in $[a_i , a_i]$, the sink knows that a_i sensors observe 1's and $(m\text{-}a_i)$ sensors observe 0's ⇔ ***The sink has to learn all the sources, in order to compute their PARITY!***

Other Implications

- Lemma 2 leads to a new lower bound for the minimum sum-rate

  "Colocated Lower Bound":

$$R_{sum,r} \geq mh(p) - \sum_{v=1, a_v \neq b_v}^{v_{\max}} (b_v - a_v) h\left(\frac{E(S|S \in [a_v, b_v]) - a_v}{b_v - a_v}\right) Pr(S \in [a_v, b_v])$$

- For any symm fn of iid Ber(1/2) srcs, $\frac{1}{2} R_{sum,1} \leq R_{sum,r} \leq R_{sum,1}$

- For any type-threshold function (e.g., MIN, MAX) of iid Ber(p) sources $R_{sum,r}(m) = \Theta(1)$ (for zero-error computation $R_{sum,r}(m) = \Theta(\log m)$)

- *"Colocated Lower Bounds" for $R_{sum,r}$ could be order-wise better than cut-set bounds*, e.g., for MIN, iid Ber(1/2), cut-set bound → 0 but new bound = Θ(1) (tight-scaling)

- Implications for secure multi-party computation

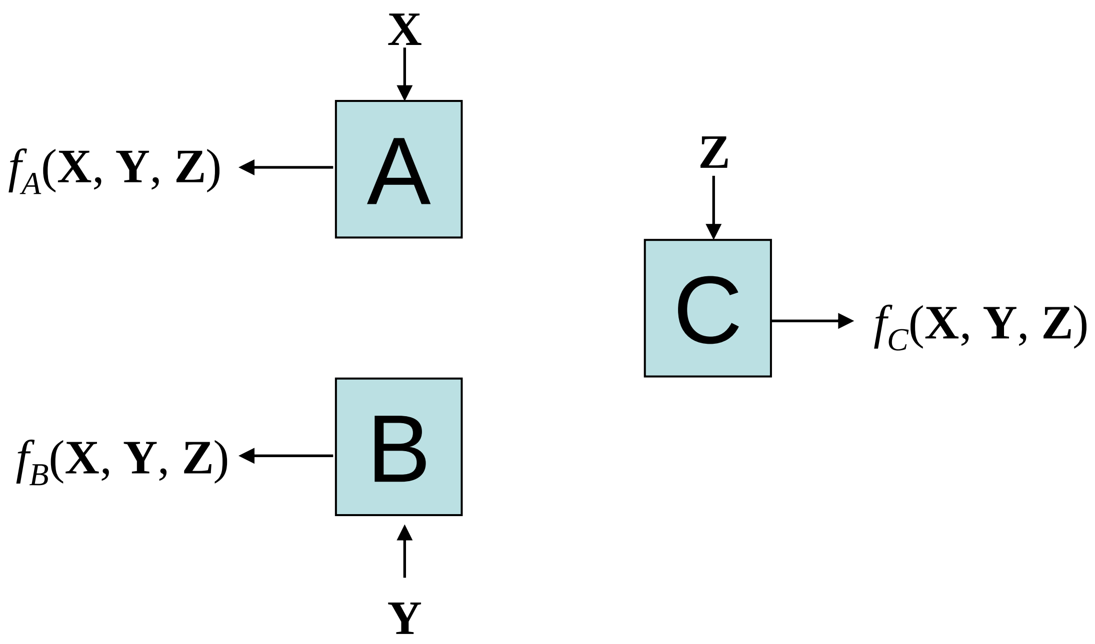
Outline

- Introduction
- General two-terminal problem
- Co-located network with independent sources
- General multi-terminal problem: some observations

BOSTON UNIVERSITY

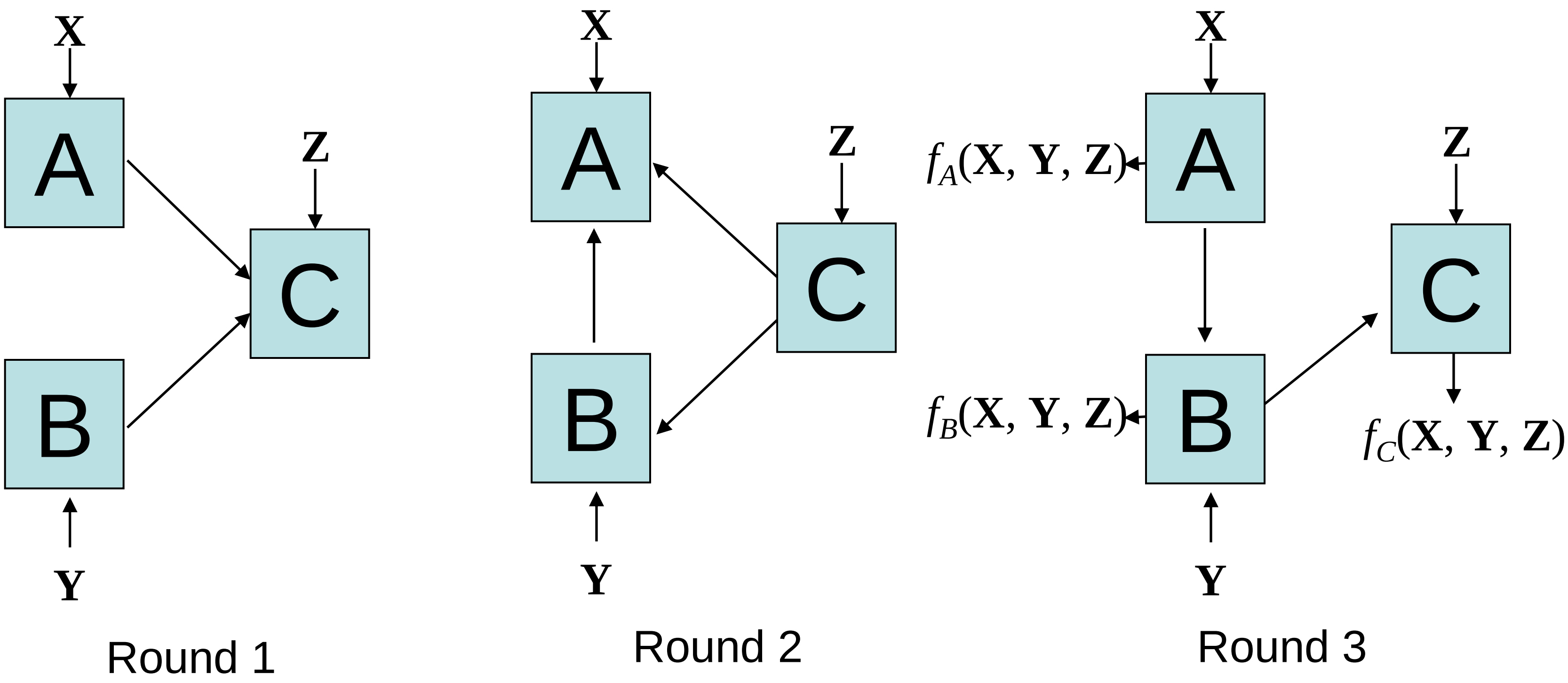
Multiterminal interaction

- *<u>m-terminal problem:</u>* m sources, m samplewise functions
- *<u>message exchanges:</u>* t rounds
- *<u>each round:</u>* concurrent message transfers.
- can switch among many non-interactive configurations



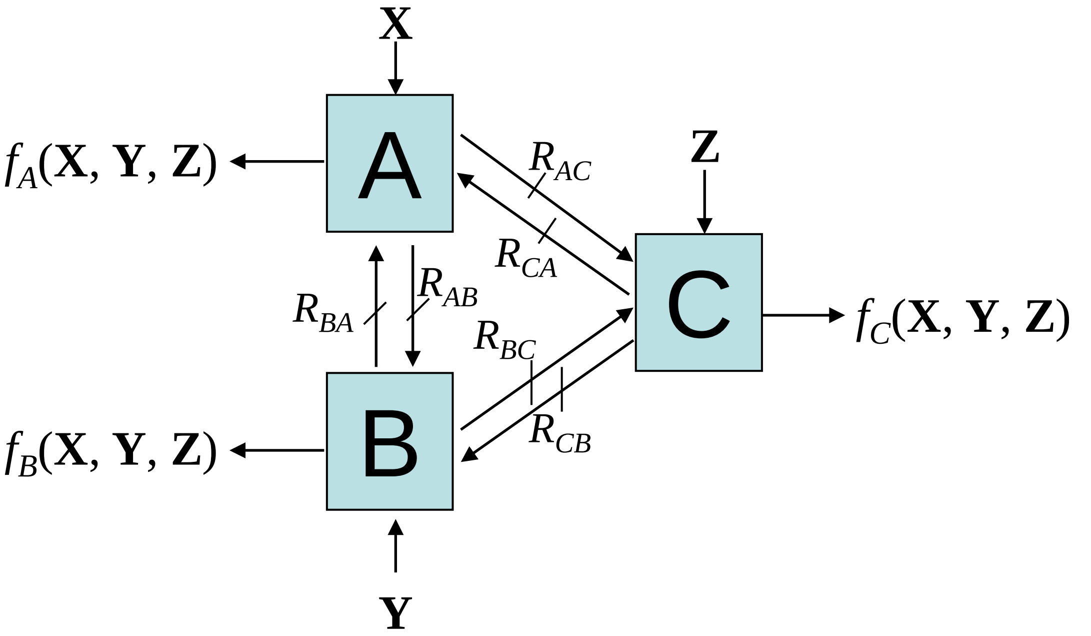
Multiterminal interaction

- *<u>m-terminal problem:</u>* m sources, m samplewise functions
- *<u>message exchanges:</u>* t rounds
- *<u>each round:</u>* concurrent message transfers.
- can switch among many non-interactive configurations



Multiterminal interaction

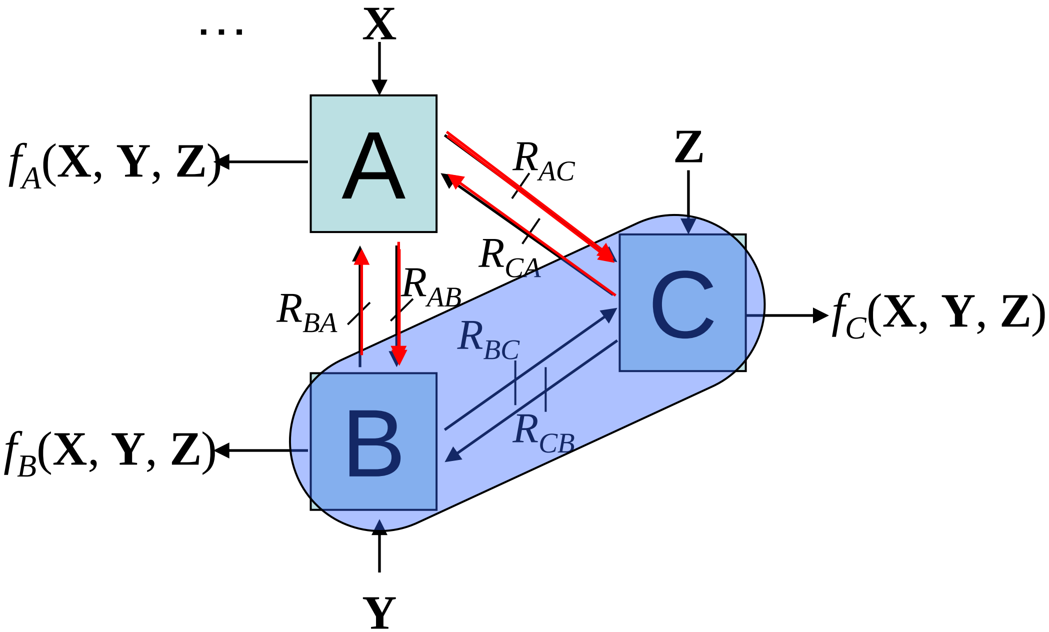
- *Complete rate region:* # bits/sample in each link, each round
- *Sum-rate region:* # bits/sample in each link, sum over all rounds
  - Region of admissible tuples $(R_{AB}, R_{BA}, R_{AC}, R_{BC}, R_{CA}, R_{CB})$
- *Minimum sum-rate:* min # bits/sample, sum over all links & all rounds
  - $R_{sum} = \min R_{AB} + R_{BA} + R_{AC} + R_{BC} + R_{CA} + R_{CB}$
- Does interaction help?

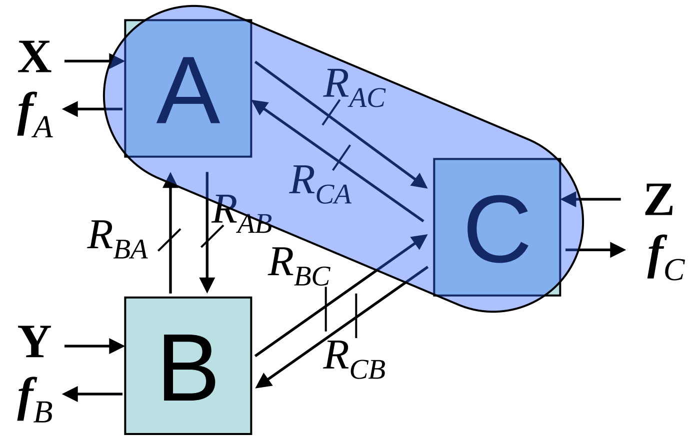


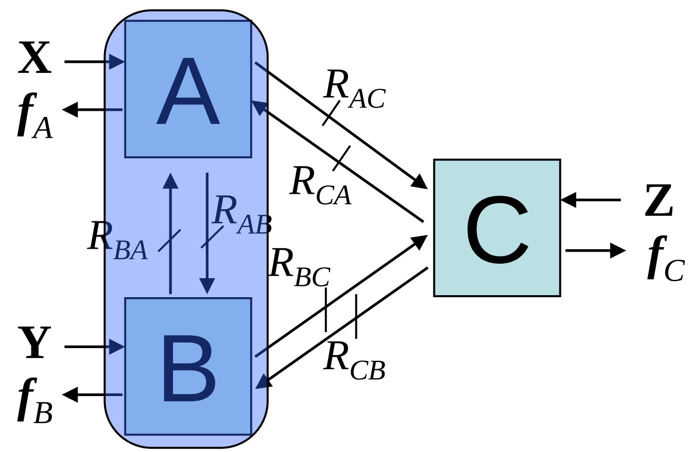
Cut-set bounds

- *Simple cut-set bounds:*
  - $R_{AB} + R_{BA} + R_{AC} + R_{CA} \geq$ 2-term min sum-rate
  - $R_{AB} + R_{AC} \geq H(f_B(X, Y, Z), f_C(X, Y, Z) | Y, Z)$

…







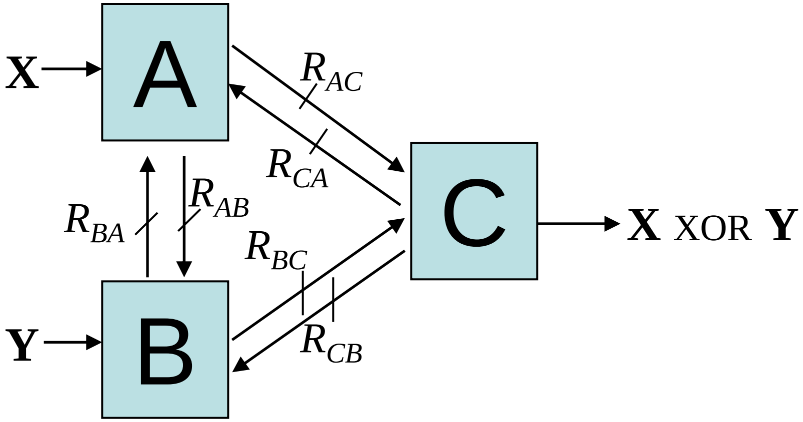
Goal: min $(R_{AB} + R_{BA} + R_{AC} + R_{BC} + R_{CA} + R_{CB})$ s.t.

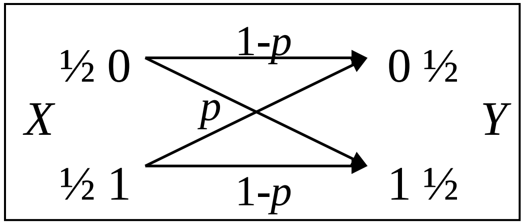
a) all sum rate lower bounds for each cut: *Linear program*

b) rates consistent with rate-regions for each cut: *Convex program*

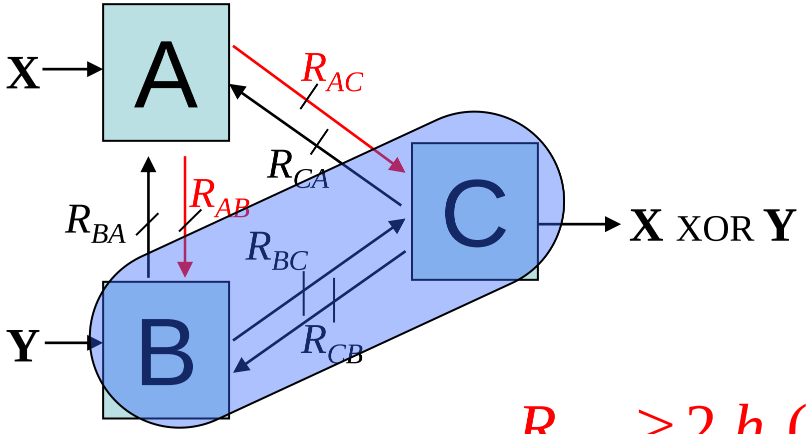
Example 1: interactive Körner-Marton

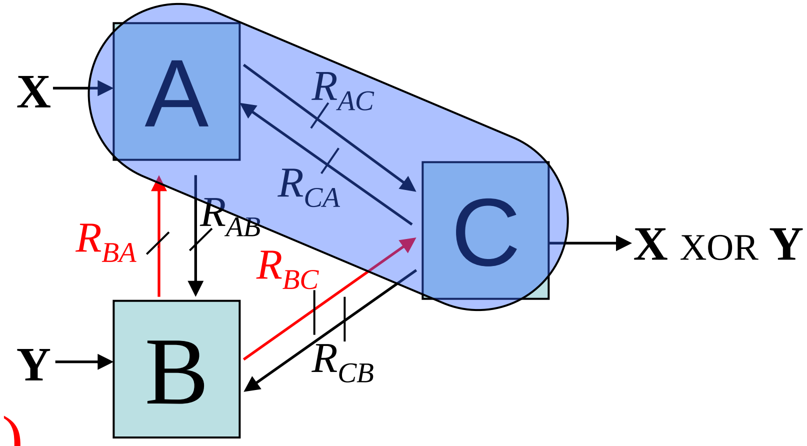
- Interactive communication allowing all possible links:





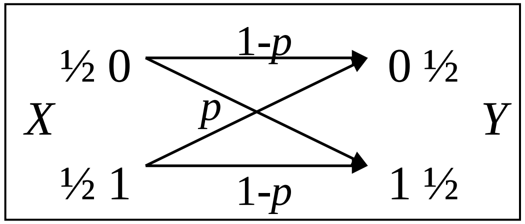
- $R_{AB} + R_{AC} \geq H(f_C(X, Y)\,|Y) = h_2(p)$ $\qquad R_{BA} + R_{BC} \geq H(f_C(X, Y)\,|\,X) = h_2(p)$



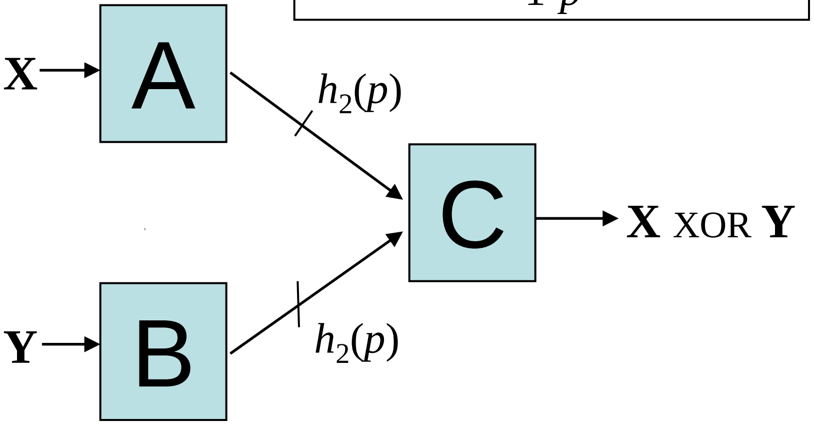


$R_{\text{sum}} \geq 2\, h_2(p)$

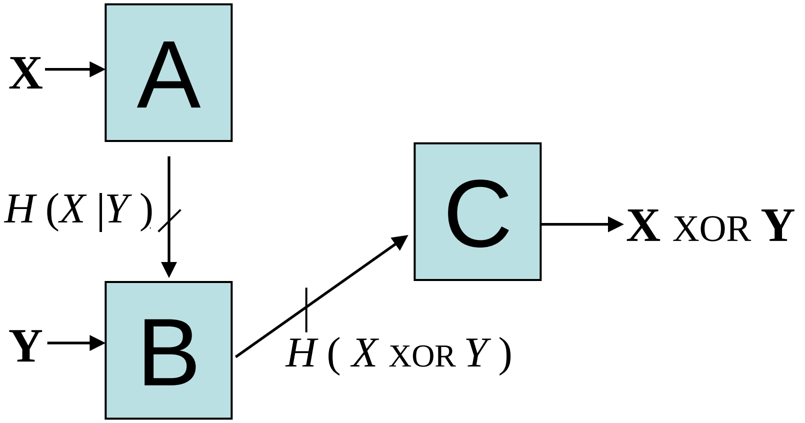
Example 1: interactive Körner-Marton



- *Non-interactive Körner-Marton:*
  - $(X, Y) \sim$ DSBS(p); $f_C(x, y) = x$ XOR y
  - Many-to-one scheme
  - $R_{AC} = R_{BC} = h_2(p)$ by linear codes
  - $R_{\text{sum}} = 2\, h_2(p)$



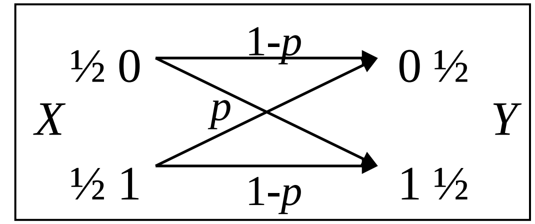
- *Relay scheme:*
  - $R_{AB} = H(X \mid Y) = h_2(p)$
  - $R_{BC} = H(f_C(X, Y)) = h_2(p)$
  - $R_{\text{sum}} = 2\, h_2(p)$

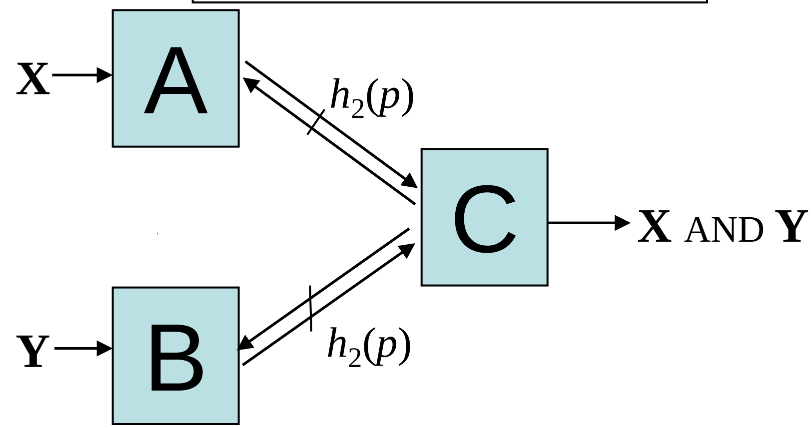


May be possible to "bypass" difficult configurations

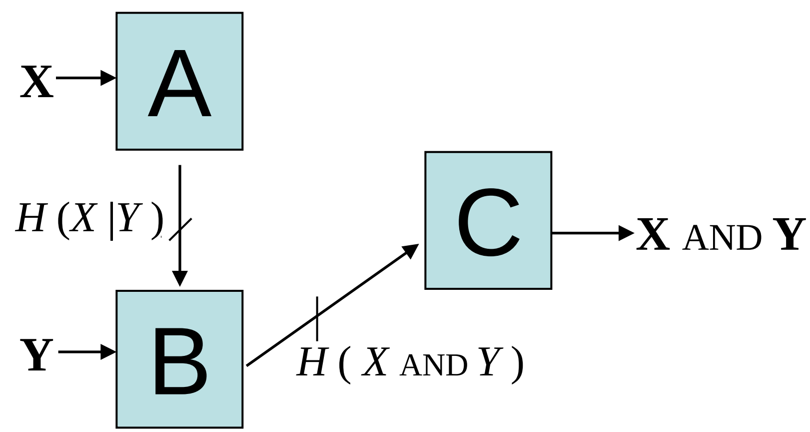
Example 2: Körner-Marton "AND"

- ***Körner-Marton problem (interactive):***
  - $(X, Y) \sim$ DSBS(p); $f_C(x, y) = x$ AND y
  - Many-to-one interactive scheme
  - $R_{\text{sum}} \geq h_2(p) + h_2(p)$
  - Min rate is unknown



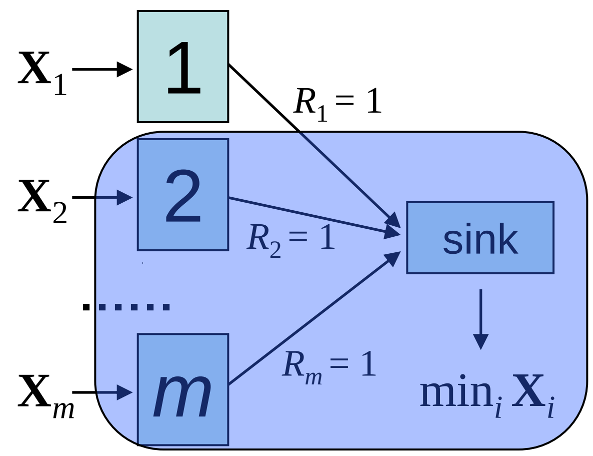


- ***Relay scheme (noninteractive)***
  - $R_{AB} = H(X \mid Y) = h_2(p)$
  - $R_{BC} = H(X \text{ AND } Y) = h_2(0.5(1\text{-}p))$
  - $R_{\text{sum}} = h_2(p) + h_2(0.5(1\text{-}p))$
  - $< 2\, h_2(p)$ for $p > 1/3$
  - Can compare configs. even if optimum is unknown

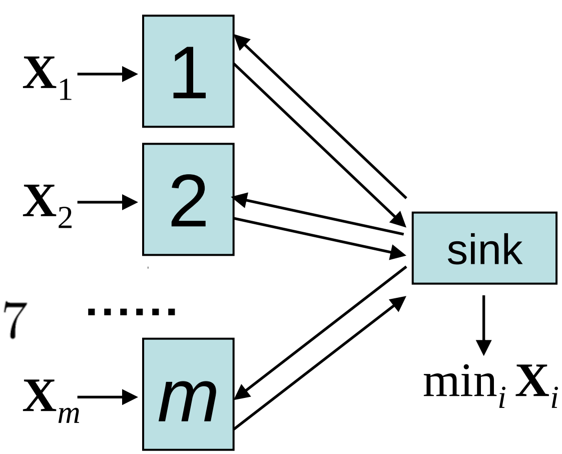


Example 3: Star networks

- $X_i \sim$ iid Ber(1/2), $f(x^m) = \min_i x_i$
- Noninteractive star network
  - Cut-set bounds: by using [Han & Kobayashi]
    Each rate ≥ 1 bit/sample
  - $R_{sum}(m) = m$



- Interactive star network
  - 1→s: send $\mathbf{X}_1$: $h_2(1/2)$
  - s→2: send $\mathbf{X}_1$: $h_2(1/2)$
  - 2→s: send $\min\{\mathbf{X}_1, \mathbf{X}_2\}$: $h_2(1/4)$
  - s→3: send $\min\{\mathbf{X}_1, \mathbf{X}_2\}$: $h_2(1/4)$
  - ……
  - Sum-rate = $2h_2(1/2) + 2\,h_2(1/4) + 2\,h_2(1/8) + \ldots < 7$
  - Using colocated lower bound: $1 \leq R_{sum}(m) < 7$
- ***Interaction changes scaling law!***



Concluding remarks

- General two-terminal problem:
  - “completely solved”
  - no benefit of interaction for data downloading;
  - benefit can be huge for computing non-trivial functions;
  - benefit depends on the structure of the functions and correlation
  - new unexplored dimension: infinite, infinitesimal-rate messages

- Colocated networks:
  - “completely solved” for independent sources
  - comm. structure reveals more information than demanded
  - cut-set bounds can be order-wise loose
  - “colocated lower bounds” order-wise tight

Questions

- Is it possible to *“bypass”* open problems in multiterminal “non-interactive” source coding by enlarging the space of strategies to include interactive ones?

- Are structured codes needed for interactive source coding?

- What are the channel coding duals of interactive source coding?

- How do distortion structure, distribution structure, and network structure influence efficiency limits in interactive source coding problems?

Thank you!