SHARED INFORMATION

Prakash Narayan

with

Imre Csiszár, Sirin Nitinawarat, Himanshu Tyagi, Shun Watanabe

Outline

Two-terminal model: Mutual information

Operational meaning in:

- Channel coding: channel capacity
- Lossy source coding: rate distortion function
- Binary hypothesis testing: Stein's lemma

Interactive communication and common randomness

- *Two-terminal model: Mutual information*
- *Multiterminal model: Shared information*

Applications

Outline

Two-terminal model: Mutual information

Operational meaning in:

- Channel coding: channel capacity
- Lossy source coding: rate distortion function
- Binary hypothesis testing: Stein's lemma

Interactive communication and common randomness

Applications

Mutual Information

Mutual information is a measure of mutual dependence between two rvs.

Mutual Information

Mutual information is a measure of mutual dependence between two rvs.

Let X_1 and X_2 be $\mathbb{R}$-valued rvs with joint probability distribution $P_{X_1X_2}$.

The **mutual information** between X_1 and X_2 is

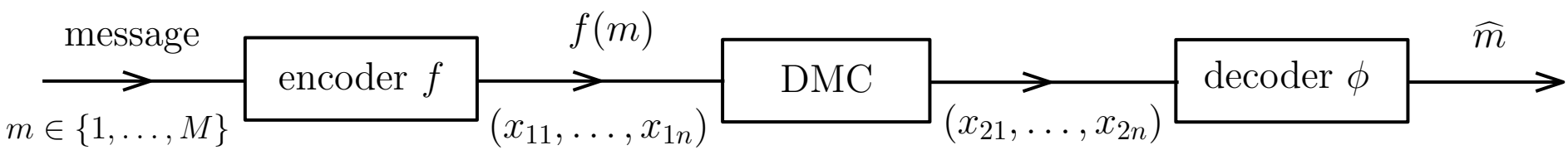
$$
\begin{aligned}
I\left(X_1 \wedge X_2\right) &= \begin{cases} \mathbb{E}_{P_{X_1X_2}}\left[\log \frac{dP_{X_1X_2}}{dP_{X_1} \times P_{X_2}}\left(X_1, X_2\right)\right], & \text{if } P_{X_1X_2} \prec P_{X_1} \times P_{X_2} \\ \infty, & \text{if } P_{X_1X_2} \not\prec P_{X_1} \times P_{X_2} \end{cases} \\
&= D\Big(P_{X_1X_2} \,||\, P_{X_1} \times P_{X_2}\Big). \quad \left(\text{Kullback} - \text{Leibler divergence}\right)
\end{aligned}
$$

When X_1 and X_2 are *finite-valued*,

$$
\begin{aligned}
I\left(X_1 \wedge X_2\right) &= H\left(X_1\right) + H\left(X_2\right) - H\left(X_1, X_2\right) \\
&= H\left(X_1\right) - H\left(X_1 \mid X_2\right) \;=\; H\left(X_2\right) - H\left(X_2 \mid X_1\right) \\
&= H\left(X_1, X_2\right) - \Big[H\left(X_1 \mid X_2\right) + H\left(X_2 \mid X_1\right)\Big].
\end{aligned}
$$

Channel Coding

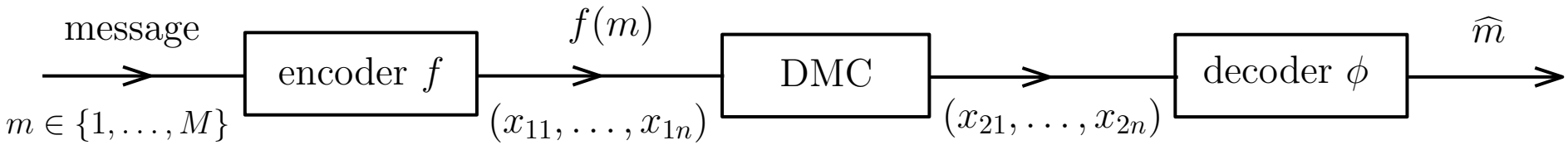
Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be finite alphabets, and $W : \mathcal{X}_1 \to \mathcal{X}_2$ be a stochastic matrix.



Discrete memoryless channel (DMC):

$$W^{(n)}\left(x_{21}, \ldots, x_{2n} \mid x_{11}, \ldots, x_{1n}\right) = \prod_{i=1}^{n} W\left(x_{2i} \mid x_{1i}\right).$$

Channel Capacity



Goal: Make code rate $\frac{1}{n} \log M$ as large as possible while keeping

$$\max_m P\big(\phi\left(X_{21}, \ldots, X_{2n}\right) \neq m \mid f(m)\big)$$

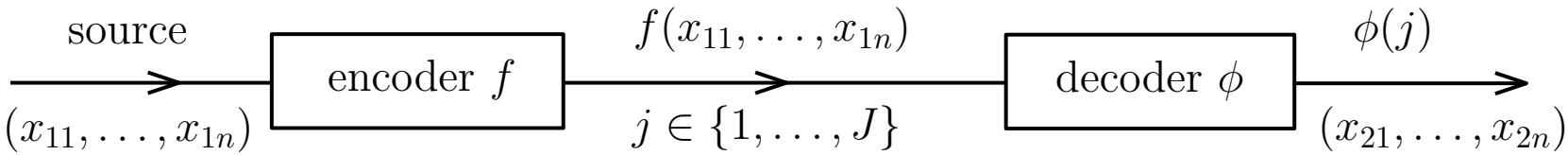
to be small, in the asymptotic sense as $n \to \infty$.

[C.E. Shannon, 1948]

$$\text{Channel capacity } C \;=\; \max_{P_{X_1}: P_{X_2|X_1} = W} I\left(X_1 \wedge X_2\right).$$

Lossy Source Coding

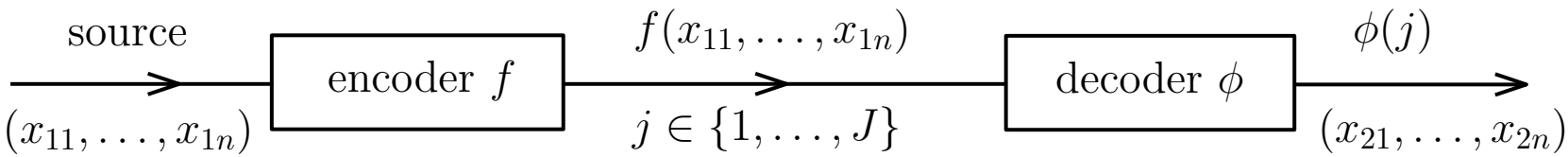
Let $\{X_{1t}\}_{t=1}^{\infty}$ be an $\mathcal{X}_1$-valued i.i.d. source.



Distortion measure:

$$d\big((x_{11}, \ldots, x_{1n}), (x_{21}, \ldots, x_{2n})\big) \;=\; \frac{1}{n} \sum_{i=1}^{n} d\big(x_{1i}, x_{2i}\big).$$

Rate Distortion Function



Goal: Make (compression) code rate $\frac{1}{n} \log J$ as small as possible while keeping

$$P\left(\frac{1}{n} \sum_{i=1}^{n} d\left(X_{1i}, X_{2i}\right) \ \leq \ \Delta\right)$$

to be large, in the asymptotic sense as $n \to \infty$.

[Shannon, 1948, 1959]

$$\text{Rate distortion function } R\left(\Delta\right) \ = \min_{P_{X_2|X_1}:\ \mathbb{E}[d(X_1,X_2)] \leq \Delta} I\left(X_1 \wedge X_2\right).$$

Simple Binary Hypothesis Testing

Let $\{(X_{1t}, X_{2t})\}_{t=1}^{\infty}$ be an $\mathcal{X}_1 \times \mathcal{X}_2$-valued i.i.d. process generated according to

$$H_0 : \; P_{X_1 X_2} \qquad \text{or} \qquad H_1 : \; P_{X_1} \times P_{X_2} .$$

Test:

$$\begin{aligned} \text{Decides } H_0 \text{ w.p. } & T\left(0 \mid x_{11}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n}\right), \\ H_1 \text{ w.p. } & T\left(1 \mid x_{11}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n}\right) = 1 - T\left(0 \mid \ldots\right). \end{aligned}$$

Stein's lemma [H. Chernoff, 1956]: For *every* $0 < \epsilon < 1$,

$$\lim_n -\frac{1}{n} \log \inf_{T:\; P_{H_0}(T \text{ says } H_0) \geq 1-\epsilon} P_{H_1}\left(T \text{ says } H_0\right)$$

$$= \; D\left(P_{X_1 X_2} \,||\, P_{X_1} \times P_{X_2}\right) \; = \; I\left(X_1 \wedge X_2\right).$$

Outline

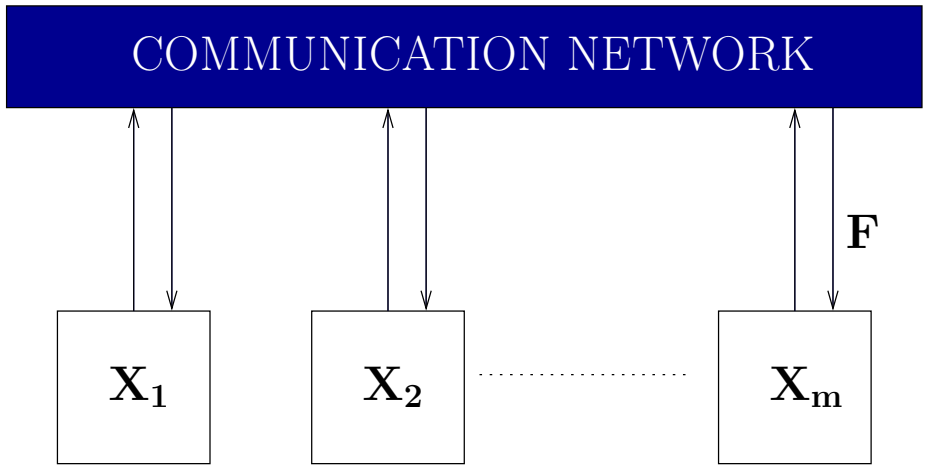
Two-terminal model: Mutual information

Interactive communication and common randomness

- *Two-terminal model: Mutual information*
- *Multiterminal model: Shared information*

Applications

Multiterminal Model



- Set of terminals $= \mathcal{M} = \{1, \ldots, m\}$.
- $X_1, \ldots, X_m$ are finite-valued rvs with known joint distribution $P_{X_1 \ldots X_m}$ on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_m$.
- Terminal $i \in \mathcal{M}$ observes data X_i.
- Multiple rounds of *interactive communication* on a *noiseless channel* of *unlimited capacity*; all terminals hear *all communication*.

Interactive Communication

Interactive communication

- Assume: Communication occurs in consecutive time slots in r rounds.
- Communication is described in terms of the mappings

$$f_{11}, \ldots, f_{1m}, \; f_{21}, \ldots, f_{2m}, \; \ldots, \; f_{r1}, \ldots, f_{rm}$$

  - f_{ji} : message in round j from terminal i, $1 \leq j \leq r$, $1 \leq i \leq m$
  - f_{ji} is any function of X_i and of all previous communication.

Interactive Communication

Interactive communication

- Assume: Communication occurs in consecutive time slots in r rounds.
- Communication is described in terms of the mappings

$$f_{11}, \ldots, f_{1m},\ f_{21}, \ldots, f_{2m},\ \ldots,\ f_{r1}, \ldots, f_{rm}$$

  - f_{ji} : message in round j from terminal i, $1 \le j \le r$, $1 \le i \le m$
  - f_{ji} is any function of X_i and of all previous communication.

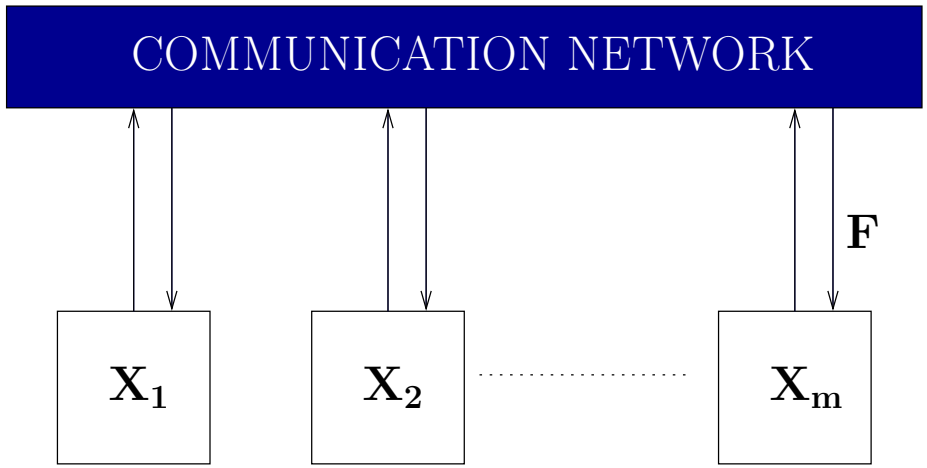
- The corresponding rvs representing the communication are

$$\mathbf{F} = \mathbf{F}\big(X_1, \ldots, X_m\big) = \Big(F_{11}, \ldots, F_{1m},\ F_{21}, \ldots, F_{2m},\ \ldots,\ F_{r1}, \ldots, F_{rm}\Big)$$

  - $F_{11} = f_{11}\big(X_1\big),\ F_{12} = f_{12}\big(X_2, F_{11}\big),\ \ldots$
  - $F_{ji} = f_{ji}\big(X_i;\ \text{all previous communication}\big)$.

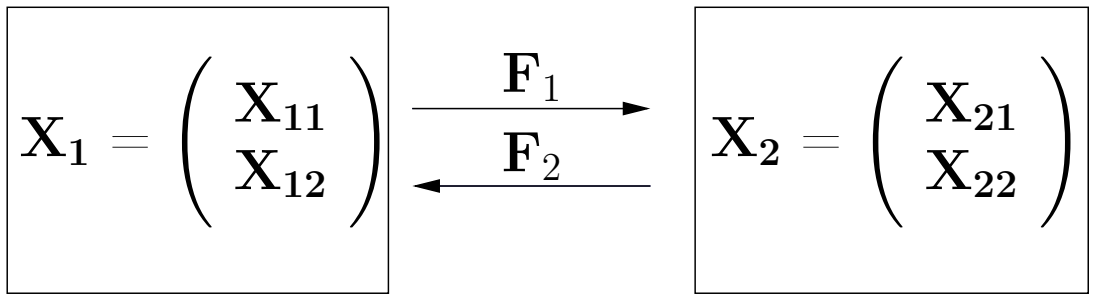
Simple communication: $\mathbf{F} = \big(F_1, \ldots, F_m\big),\quad F_i = f_i\big(X_i\big),\ 1 \le i \le m.$

Applications



- *Data exchange: Omniscience*.
- *Signal recovery: Data compression*.
- *Function computation*.
- *Cryptography: Secret key generation*.

Example: Function Computation



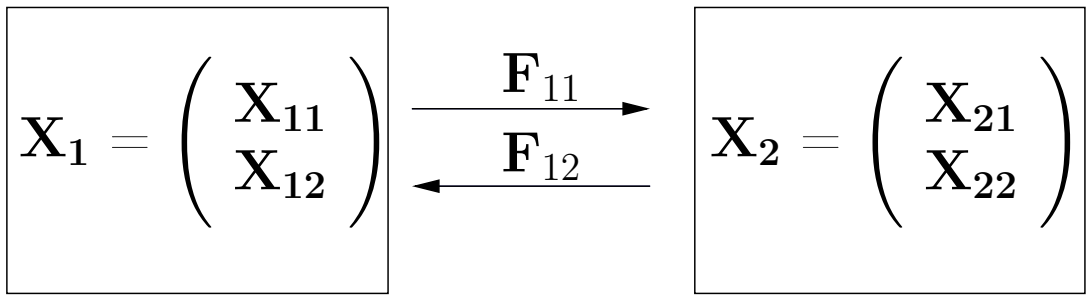
[S. Watanabe]

- $X_{11}, X_{12}, X_{21}, X_{22}$ are mutually independent $(0.5, 0.5)$ bits.
- Terminals 1 and 2 wish to compute:

$$G = g(X_1, X_2) = \mathbb{1}\Big((X_{11}, X_{12}) = (X_{21}, X_{22})\Big).$$

- *Simple communication*: $\mathbf{F} = \Big(F_1 = (X_{11}, X_{12}),\ F_2 = (X_{21}, X_{22})\Big)$.
  - Communication complexity: $H(\mathbf{F}) = 4$ bits.
  - No privacy: Terminal 1 or 2, or an observer of $\mathbf{F}$, learns all the data X_1, X_2.

Example: Function Computation

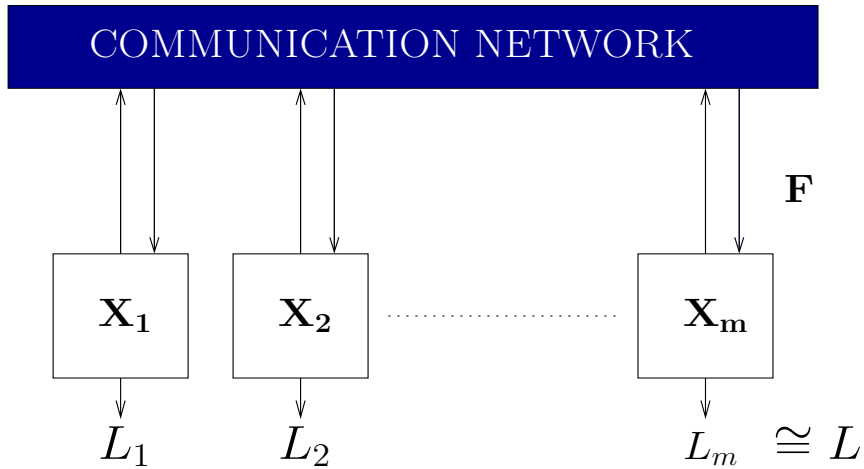


- *Interactive communication protocol*:
  - $\mathbf{F} = \Big(F_{11} = (X_{11}, X_{12}),\ F_{12} = G\Big)$.
  - Complexity: $H(\mathbf{F}) = 2.81$ bits.
  - Some privacy: Terminal 2, or an observer of $\mathbf{F}$, learns X_1; terminal 1, or an observer of $\mathbf{F}$, either learns X_2 w.p. 0.25 or w.p. 0.75 that X_2 differs from X_1.

Related Work

- Exact function computation
  - Yao '79: Communication complexity.
  - Gallager '88: Algorithm for parity computation in a network.
  - Giridhar-Kumar '05: Algorithms for computing functions over sensor networks.
  - Freris-Kowshik-Kumar '10: Survey: Connectivity, capacity, clocks, computation in large sensor networks.
  - Orlitsky-El Gamal '84: Communication complexity with secrecy.
- Information theoretic function computation
  - Körner-Marton '79: Minimum rate for computing parity.
  - Orlitsky-Roche '01: Two terminal function computation.
  - Nazer-Gastpar '07: Computation over noisy channels.
  - Ma-Ishwar '08: Distributed source coding for interactive computing.
  - Ma-Ishwar-Gupta '09: Multiround function computation in colocated networks.
  - Tyagi-Gupta-Narayan '11: Secure function computation.
  - Tyagi-Watanabe '13, '14 Secrecy generation, secure computing.
- Compressing interactive communication
  - Schulman '92: Coding for interactive communication.
  - Braverman-Rao '10: Information complexity of communication.
  - Kol-Raz '13, Heupler '14: Interactive communication over noisy channels.

Common Randomness



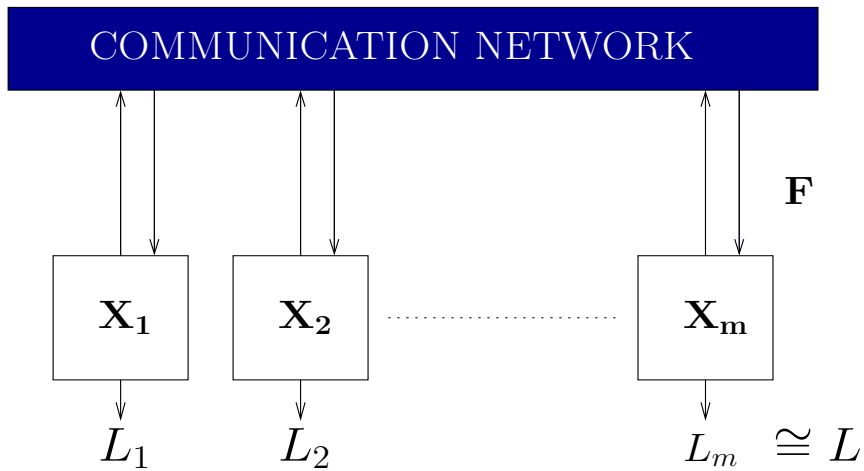
For $0 \leq \epsilon < 1$, given interactive communication $\mathbf{F}$, an rv $L = L(X_1, \ldots, X_m)$ is ϵ-CR for the terminals in $\mathcal{M}$ using $\mathbf{F}$, if there exist *local estimates*

$$L_i \ = L_i\big(X_i,\ \mathbf{F}\big), \quad i \in \mathcal{M},$$

of L satisfying

$$P\Big(L_i = L, \quad i \in \mathcal{M}\Big) \ \geq \ 1 - \epsilon.$$

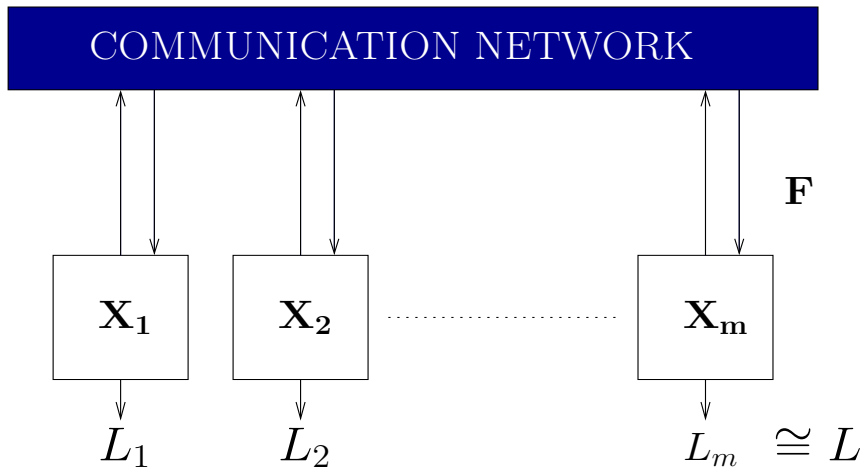
Common Randomness



Examples:

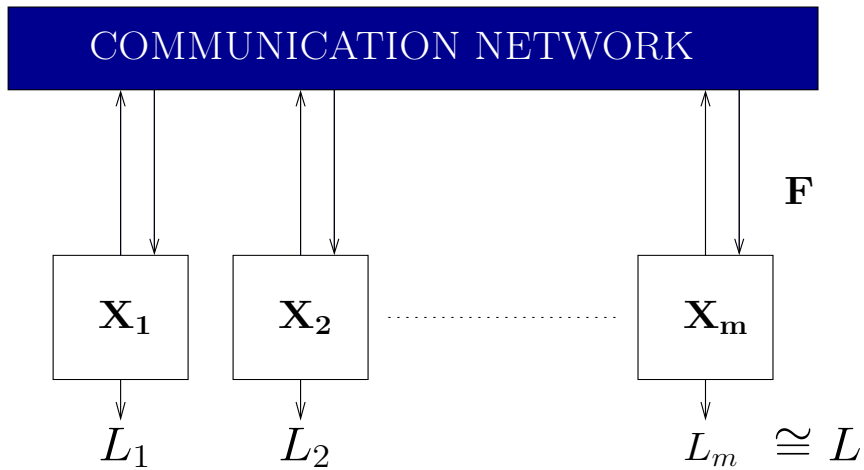
- *Omniscience*: $L = \big(X_1, \ldots, X_m\big)$.
- *Single signal*: $L = X_{i^*}$, for some fixed $i^* \in \mathcal{M}$.
- *Function computation*: $L = g\big(X_1, \ldots, X_m\big)$ for a given g.
- *Secret CR, i.e., secret key*: L with $I(L \ \wedge \ \mathbf{F}) \cong 0$.

A Basic Operational Question



¿ What is the *maximal* CR, as measured by $H\big(L|\mathbf{F}\big)$, that can be generated by a *given* interactive communication $\mathbf{F}$ for a distributed processing task **?**

A Basic Operational Question



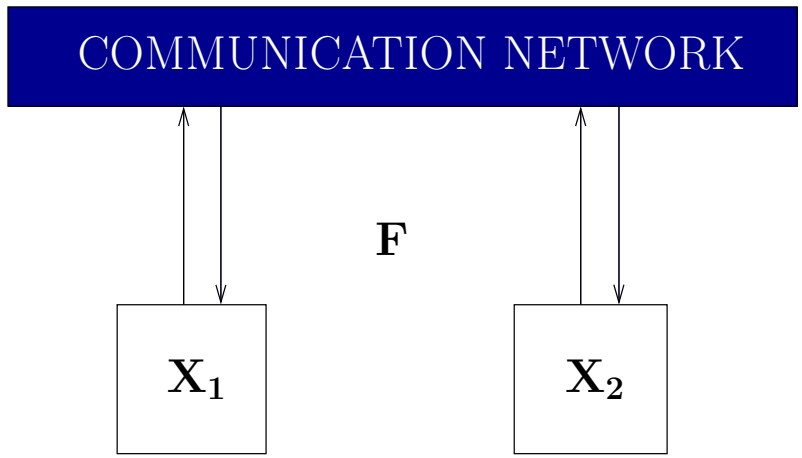
¿ What is the *maximal* CR, as measured by $H\big(L|\mathbf{F}\big)$, that can be generated by a *given* interactive communication $\mathbf{F}$ for a distributed processing task **?**

Answer in two steps:

- Fundamental property of interactive communication
- Upper bound on amount of CR achievable with interactive communication.

Shall start with the case of $m = 2$ terminals.

Fundamental Property of Interactive Communication



Lemma: [U. Maurer], [R. Ahlswede - I. Csiszár]
For interactive communication $\mathbf{F}$ of the terminals $i \in \mathcal{M} = \{1, 2\}$, with terminal i possessing "initial" data X_i,

$$I\big(X_1 \wedge X_2 \big| \mathbf{F}\big) \;\leq\; I\big(X_1 \wedge X_2\big).$$

In particular, independent rvs X_1, X_2 remain so upon conditioning on an interactive communication.

An Equivalent Form

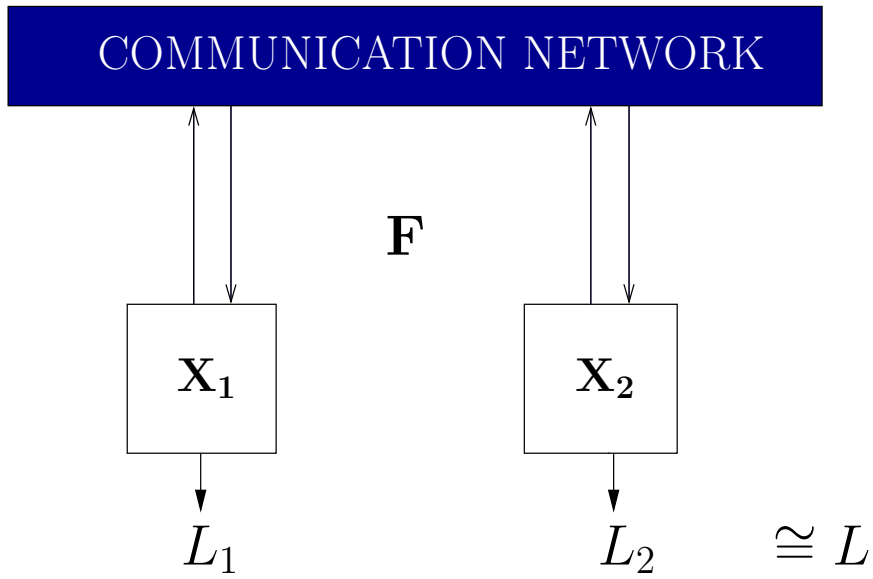
For interactive communication $\mathbf{F}$ of terminals 1 and 2:

$$I\big(X_1 \wedge X_2 \big| \mathbf{F}\big) \;\leq\; I\big(X_1 \wedge X_2\big)$$

$$\Updownarrow$$

$$H(\mathbf{F}) \;\geq\; H(\mathbf{F}|X_1) + H(\mathbf{F}|X_2).$$

Upper Bound on CR for Two Terminals



Using

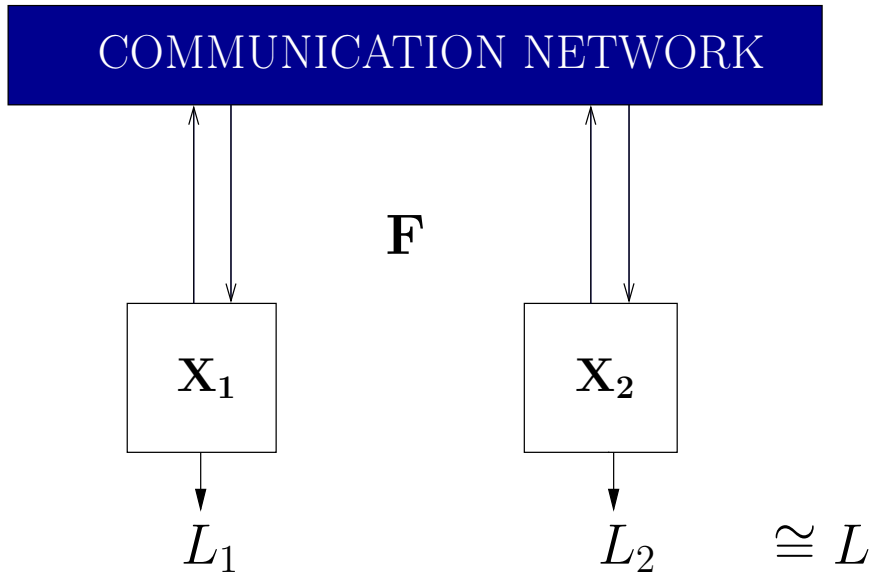
- L is ϵ-CR for $\mathcal{M} = \{1, 2\}$ with interactive $\mathbf{F}$; and
- $H(\mathbf{F}) \geq H(\mathbf{F}|X_1) + H(\mathbf{F}|X_2)$,

we get

$$H\big(L|\mathbf{F}\big) \leq H\big(X_1, X_2\big) - \Big[H\big(X_1|X_2\big) + H\big(X_2|X_1\big)\Big] + 2\nu(\epsilon),$$

where $\lim_{\epsilon \to 0} \nu(\epsilon) = 0$.

Upper Bound on CR for Two Terminals



Lemma: [I. Csiszár - P. Narayan] Let L be *any* ϵ-CR for the terminals $i \in \mathcal{M} = \{1, 2\}$ with terminal i possessing "initial" data X_i, achievable with interactive communication $\mathbf{F}$. Then

$$H\big(L|\mathbf{F}\big) \lessapprox I\big(X_1 \wedge X_2\big) = D\big(P_{X_1X_2}||P_{X_1} \times P_{X_2}\big).$$

Remark: When $\{(X_{1t}, X_{2t})\}_{t=1}^{\infty}$ is an $\mathcal{X}_1 \times \mathcal{X}_2$-valued i.i.d. process, the upper bound is attained.

Interactive Communication for $m \geq 2$ Terminals

Theorem 1: [I. Csiszár-P. Narayan]
For interactive communication $\mathbf{F}$ of the terminals $i \in \mathcal{M} = \{1, \ldots, m\}$, with terminal i possessing "initial" data X_i,

$$H(\mathbf{F}) \geq \sum_{B \in \mathcal{B}} \lambda_B H(\mathbf{F}|X_{B^c})$$

for every family $\mathcal{B} = \{B \subsetneq \mathcal{M}, B \neq \emptyset\}$ and set of weights ("*fractional partition*")

$$\lambda \triangleq \left\{ 0 \leq \lambda_B \leq 1, \ B \in \mathcal{B}, \ \text{ satisfying } \sum_{B \in \mathcal{B}: B \ni i} \lambda_B = 1 \ \forall \ i \in \mathcal{M} \right\}.$$

Equality holds if $X_1, \ldots, X_m$ are mutually independent.

Special case of:
M. Madiman and P. Tetali, "Information inequalities for joint distributions, with interpretations and applications," IEEE Trans. Inform. Theory, June 2010.

CR for $m \geq 2$ Terminals: Shared Information

Theorem 2: [I. Csiszár-P. Narayan, C. Chan-L. Zheng]

Given $0 \leq \epsilon < 1$, for an ϵ-CR L for $\mathcal{M}$ achieved with interactive communication $\mathbf{F}$,

$$H\big(L|\mathbf{F}\big) \lessapprox H\big(X_1,\ldots,X_m\big) - \max_{\lambda} \sum_{B\in\mathcal{B}} \lambda_B H\left(X_B|X_{B^c}\right)$$

$$= \min_{2\leq k\leq m} \min_{\mathcal{A}_k=(A_1,\ldots,A_k)} \frac{1}{k-1} D\Big(P_{X_1\ldots X_m}||\prod_{i=1}^{k} P_{X_{A_i}}\Big)$$

$$\triangleq SI\big(X_1,\ldots,X_m\big).$$

CR for $m \geq 2$ Terminals: Shared Information

Theorem 2: [I. Csiszár-P. Narayan, C. Chan-L. Zheng]

Given $0 \leq \epsilon < 1$, for an ϵ-CR L for $\mathcal{M}$ achieved with interactive communication $\mathbf{F}$,

$$H\big(L|\mathbf{F}\big) \lessapprox H\big(X_1,\ldots,X_m\big) - \max_{\lambda} \sum_{B\in\mathcal{B}} \lambda_B H\left(X_B|X_{B^c}\right)$$

$$= \min_{2\leq k\leq m} \min_{\mathcal{A}_k=(A_1,\ldots,A_k)} \frac{1}{k-1} D\Big(P_{X_1\ldots X_m}\big|\big|\prod_{i=1}^{k} P_{X_{A_i}}\Big)$$

$$\triangleq SI\big(X_1,\ldots,X_m\big).$$

Remarks:

- The proof of Theorem 2 relies on Theorem 1.
- When $\{(X_{1t},\ldots,X_{mt})\}_{t=1}^{\infty}$ is an i.i.d. process, the upper bound is attained.

Shared Information

$$SI(X_1,\ldots,X_m) \;=\; \min_{2\le k\le m} \;\; \min_{\mathcal{A}_k=(A_1,\ldots,A_k)} \;\; \frac{1}{k-1}\, D\Big(P_{X_1\ldots X_m}\,||\prod_{i=1}^{k} P_{X_{A_i}}\Big)$$

and equals 0 iff $P_{X_1\ldots X_m} \;=\; P_{X_A} P_{X_{A^c}}$ for some $A \subsetneq \mathcal{M}$.

¿ Does *shared information* have an operational significance as a measure of the mutual dependence among the rvs $X_1,\ldots,X_m$?

Extensions

Theorems 1 and 2 extend to:

- random variables with densities [S. Nitinawarat-P. Narayan]
- a larger class of probability measures [H.Tyagi-P. Narayan].

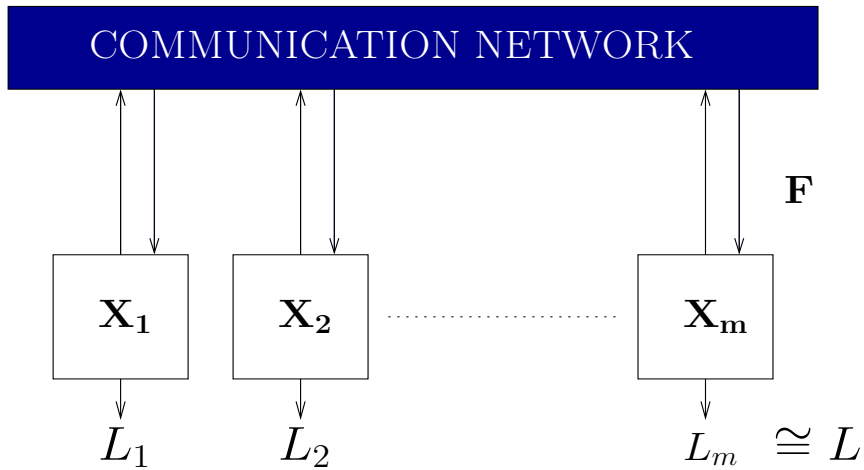
Outline

Two-terminal model: Mutual information

Interactive communication and common randomness

Applications

Omniscience



[I. Csiszár-P. Narayan]

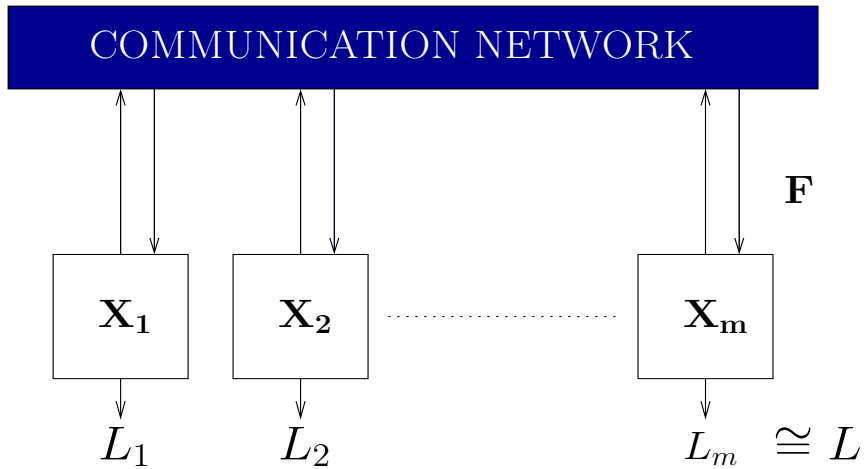
For $L = (X_1, \ldots, X_m)$, Theorem 2 gives

$$H(\mathbf{F}) \gtrapprox H(X_1, \ldots, X_m) - SI(X_1, \ldots, X_m),$$

which, for $m = 2$, is

$$H(\mathbf{F}) \gtrapprox H(X_1|X_2) + H(X_2|X_1). \qquad [\text{Slepian} - \text{Wolf}]$$

Recovery of a Single Signal



[S. Nitinawarat-P. Narayan]

With $L = X_1$, by Theorem 2

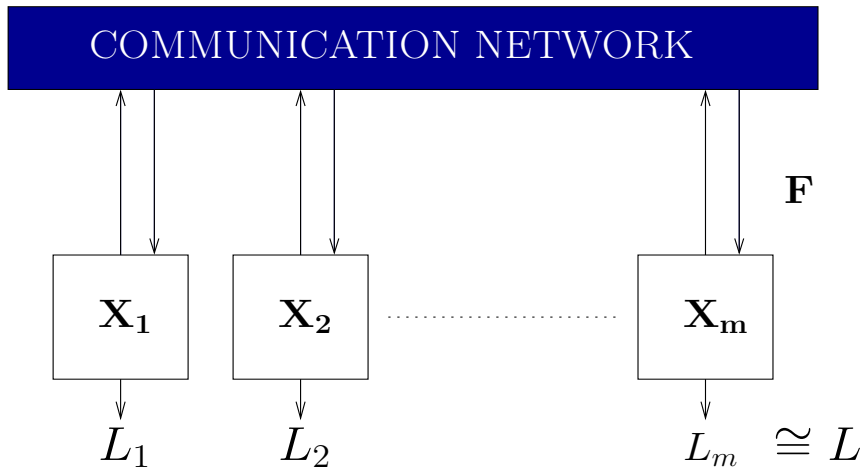
$$H(\mathbf{F}) \gtrapprox H(X_1) - SI(X_1, \ldots, X_m),$$

which, for $m = 2$, gives

$$H(\mathbf{F}) \gtrapprox H(X_1|X_2).$$

[Slepian-Wolf]

Secret Common Randomness



Terminals $1, \ldots, m$ generate CR L satisfying the *secrecy condition*

$$I\big(L \wedge \mathbf{F}\big) \;\cong\; 0.$$

By Theorem 2,

$$H\big(L\big) \cong H\big(L|\mathbf{F}\big) \;\lessapprox\; SI\big(X_1, \ldots, X_m\big).$$

- Secret key generation [I. Csiszár-P. Narayan]
- Secure function computation [H. Tyagi-P. Narayan]

Shared information and a Hypothesis Testing Problem

$$SI\big(X_1,\ldots,X_m\big) \;=\; \min_{2\le k\le m} \;\; \min_{\mathcal{A}_k=(A_1,\ldots,A_k)} \;\; \frac{1}{k-1}\, D\Big(P_{X_1\ldots X_m}\big|\big|\prod_{i=1}^{k} P_{X_{A_i}}\Big)$$

- Related to exponent of "P_e-second kind" for an appropriate binary composite hypothesis testing problem, involving restricted CR L and communication $\mathbf{F}$.

H. Tyagi and S. Watanabe, "Converses for secret key agreement and secure computing," *IEEE Trans. Information Theory*, September 2015.

In Closing ...

¿ How useful is the concept of *shared information* **?**

A: Operational meaning in specific cases of distributed processing ...

In Closing ...

¿ How useful is the concept of *shared information* ?

A: Operational meaning in specific cases of distributed processing ...

For instance

- Consider n i.i.d. repetitions (say, in time) of the rvs $X_1, \ldots, X_m$.
- Data at time instant t is $X_{1t}, \ldots, X_{mt}, \quad t = 1, \ldots, n$.
- Terminal i observes the i.i.d. data $\left(X_{i1}, \ldots, X_{in}\right), \quad i \in \mathcal{M}$.
- Shared information-based results are asymptotically tight (in n):
  - *Minimum rate* of communication for omniscience.
  - *Maximum rate* of a secret key.
  - *Necessary condition* for secure function computation.
  - Several problems in information theoretic cryptography.

Shared Information: Many Many Open Questions ...

- Significance in network source and channel coding?
- Interactive communication over noisy channels?
- Continuous-time models?

$\vdots$

$\vdots$