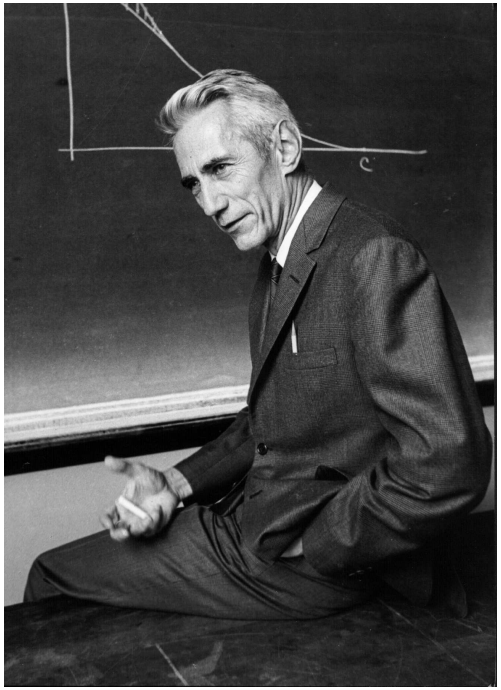
Information Theoretic Limits of Randomness Generation

Abbas El Gamal

Stanford University

Shannon Centennial, University of Michigan, September 2016

Information theory



The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem.

– A Mathematical Theory of Communication, Shannon (1948)

Information theory

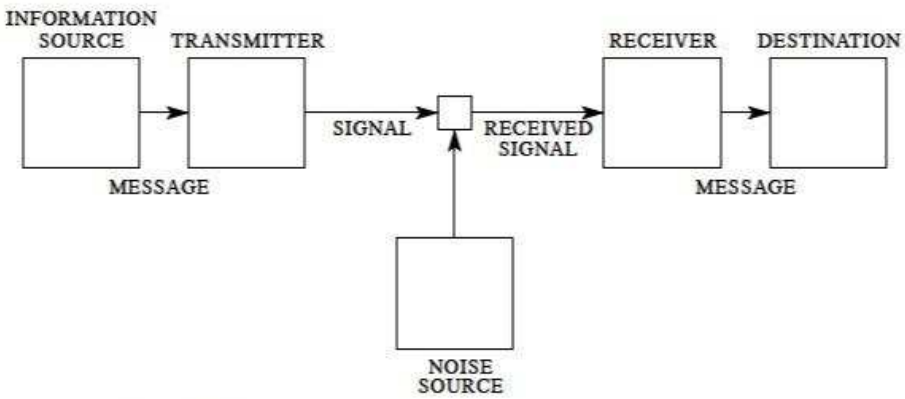


Fig. 1 — Schematic diagram of a general communication system.

- Probabilistic models for the source and the channel

Information theory

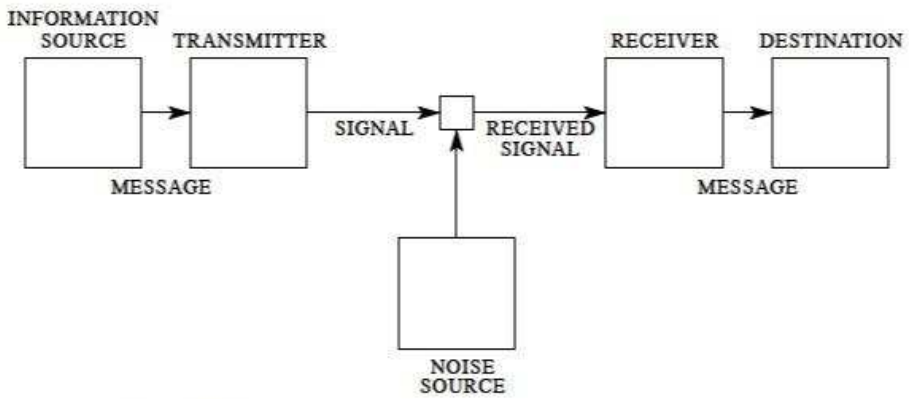


Fig. 1 — Schematic diagram of a general communication system.

- Probabilistic models for the source and the channel
- Entropy as measure of information

Information theory

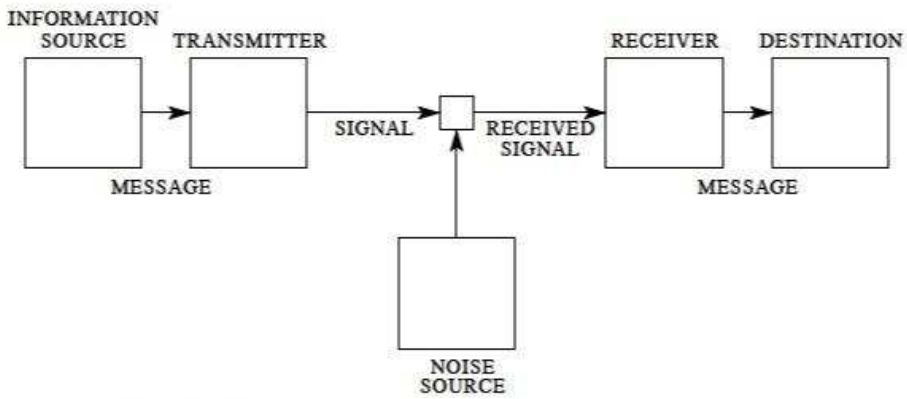


Fig. 1 — Schematic diagram of a general communication system.

- Probabilistic models for the source and the channel
- Entropy as measure of information
- Minimum rate for compression: entropy H

Information theory

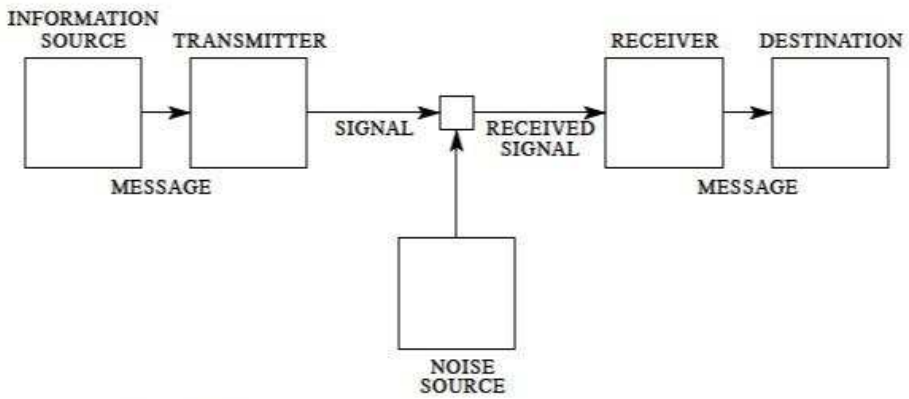


Fig. 1 — Schematic diagram of a general communication system.

- Probabilistic models for the source and the channel
- Entropy as measure of information
- Minimum rate for compression: entropy H
- Highest rate of reliable communication: capacity $C = \max$ mutual information I

Information theory

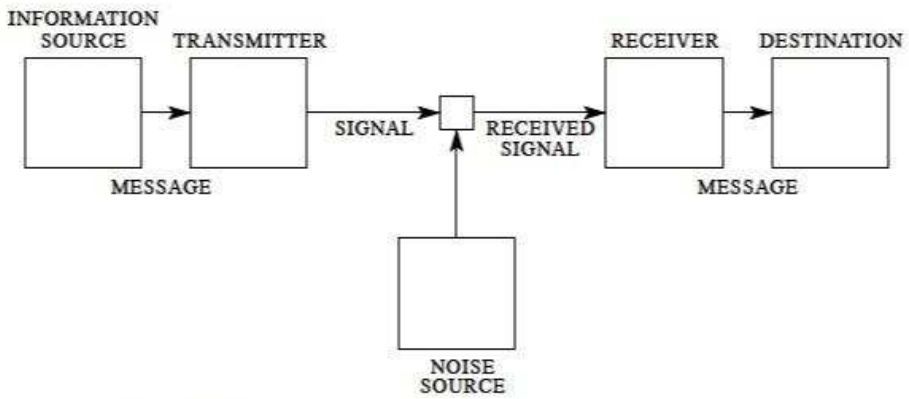


Fig. 1 — Schematic diagram of a general communication system.

- Probabilistic models for the source and the channel
- Entropy as measure of information
- Minimum rate for compression: entropy H
- Highest rate of reliable communication: capacity $C = \max$ mutual information I
- Source can communicated over channel iff $H \leq C$

Bits as a universal interface between source and channel

Profound impact on practice

Einstein looms large, and rightly so. But we'er not living in the relativity age, we're living in the information age. It's Shannon whose fingerprints are on every electronic device we own, every computer screen we gaze into, every means of digital communication. He's one of these people who so transform the world that, after the transformation, the old world is forgotten.

– James Gleick in the NEW YORKER, April 30, 2016

Information theory beyond communication

Information theory beyond communication

INFORMATION theory has, in the last few years, become something of a scientific bandwagon. Starting as a technical tool for the communication engineer, it has received an extraordinary amount of publicity in the popular as well as the scientific press. . . . As a consequence, it has perhaps been ballooned to an importance beyond its actual accomplishments. . . . Applications are being made to biology, psychology, linguistics, fundamental physics, economics, the theory of organization, and many others. . . .

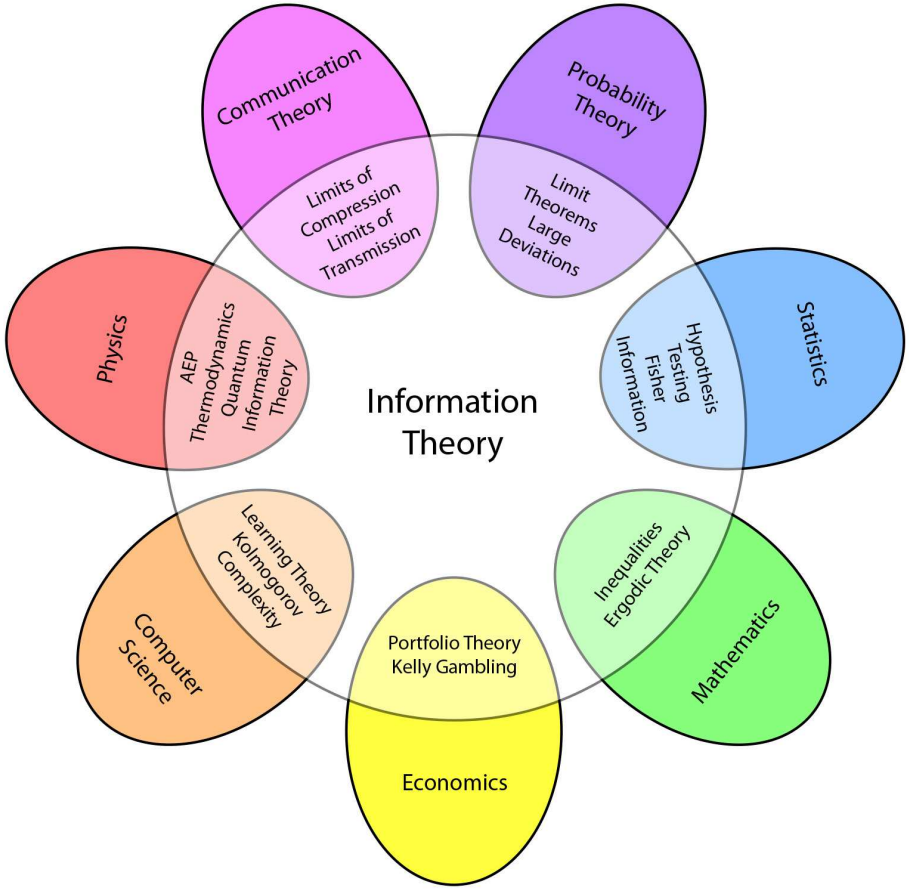
– *The Bandwagon, Shannon (1956)*

Information theory beyond communication

INFORMATION theory has, in the last few years, become something of a scientific bandwagon. Starting as a technical tool for the communication engineer, it has received an extraordinary amount of publicity in the popular as well as the scientific press. . . . As a consequence, it has perhaps been ballooned to an importance beyond its actual accomplishments. . . . Applications are being made to biology, psychology, linguistics, fundamental physics, economics, the theory of organization, and many others. . . . I personally believe that many of the concepts of information theory will prove useful in these other fields — and, indeed, some results are already quite promising — but the establishing of such applications is not a trivial matter of translating words to a new domain, but rather the slow tedious process of hypothesis and experimental verification. . . .

– The Bandwagon, Shannon (1956)

Information theory beyond communication



– *Elements of Information Theory, Cover–Thomas (1991)*

Randomness generation

- Generating random variables with prescribed distribution from coin flips

Randomness generation

- Generating random variables with prescribed distribution from coin flips
- Important applications:
  - Monte Carlo simulation
  - Randomized algorithms
  - Cryptography
- Fundamental questions:
  - Notions and measures of common information
  - Secrecy
  - Classical and quantum channel synthesis

Randomness generation

- Generating random variables with prescribed distribution from coin flips
- Important applications:
  - Monte Carlo simulation
  - Randomized algorithms
  - Cryptography
- Fundamental questions:
  - Notions and measures of common information
  - Secrecy
  - Classical and quantum channel synthesis
- What is the minimum amount of coin flips needed?
  - Entropy and mutual information arise naturally as limits/bounds
  - Proofs use information theoretic techniques

Outline

- Centralized randomness generation:
  - One shot using a variable number of coin flips
  - Asymptotic using a fixed number of coin flips

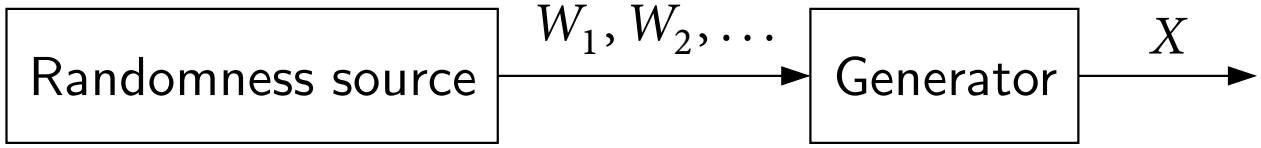
Outline

- Centralized randomness generation:
  - One shot using a variable number of coin flips
  - Asymptotic using a fixed number of coin flips
- Distributed randomness generation:
  - Asymptotic
  - One shot – recent results

Outline

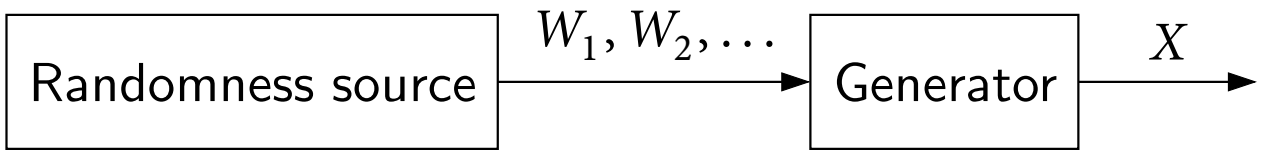
- Centralized randomness generation:
  - One shot using a variable number of coin flips
  - Asymptotic using a fixed number of coin flips
- Distributed randomness generation:
  - Asymptotic
  - One shot – recent results
- Channel synthesis (common randomness/communication tradeoff):
  - Asymptotic
  - One shot

One shot randomness generation

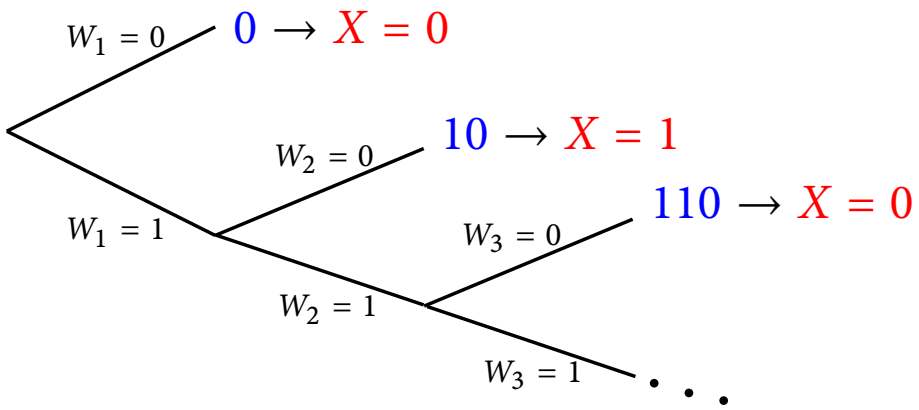


- $W_1, W_2, \ldots$ is a $\mathrm{Bern}(1/2)$ sequence; $X \sim p(x)$
- Use prefix-free code to generate X (as in zero error compression)

One shot randomness generation

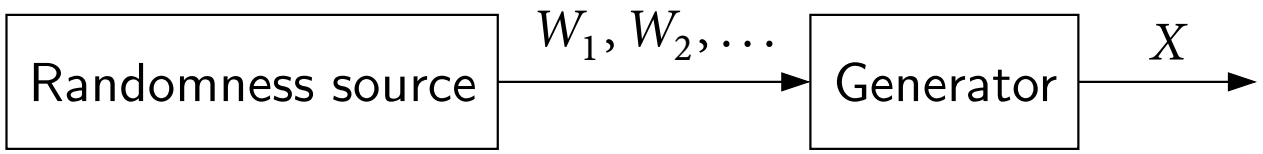


- $W_1, W_2, \ldots$ is a $\mathrm{Bern}(1/2)$ sequence; $X \sim p(x)$
- Use prefix-free code to generate X (as in zero error compression)
- Example: generating $X \sim \mathrm{Bern}(1/3)$:



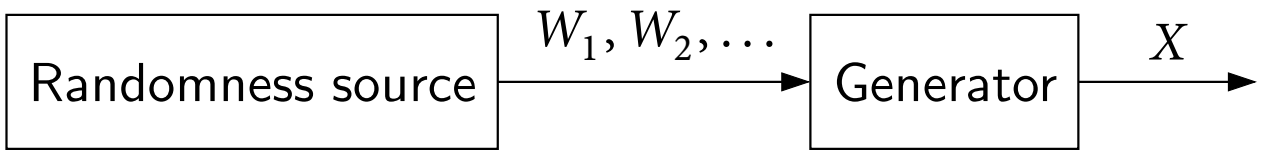
$$\mathrm{P}\{X = 1\} = \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots = \frac{1}{3}$$

One shot randomness generation



- $W_1, W_2, \ldots$ is a $\mathrm{Bern}(1/2)$ sequence; $X \sim p(x)$
- Use prefix-free code to generate X (as in zero error compression)
- Let L be the number of W_i bits used to generate X
- What is the minimum $\mathsf{E}(L)$?

One shot randomness generation



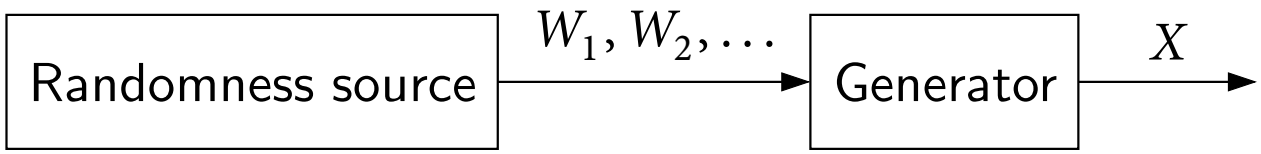
- $W_1, W_2, \ldots$ is a $\mathrm{Bern}(1/2)$ sequence; $X \sim p(x)$
- Use prefix-free code to generate X (as in zero error compression)
- Let L be the number of W_i bits used to generate X
- What is the minimum $\mathsf{E}(L)$?
- Knuth–Yao (1976) showed that:

$$H(X) \leq \min \mathsf{E}(L) < H(X) + 2 \quad \text{bits/symbol,}$$

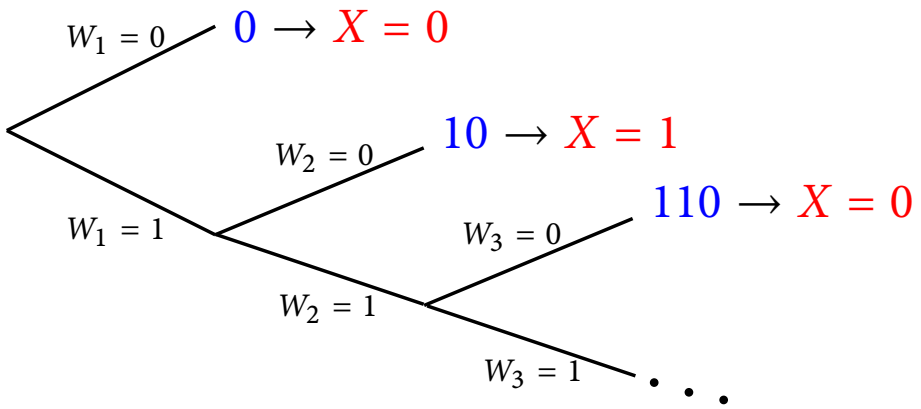
where $H(X) = \mathsf{E}[-\log p(X)]$ is the entropy of X

Measure of randomness of X

One shot randomness generation



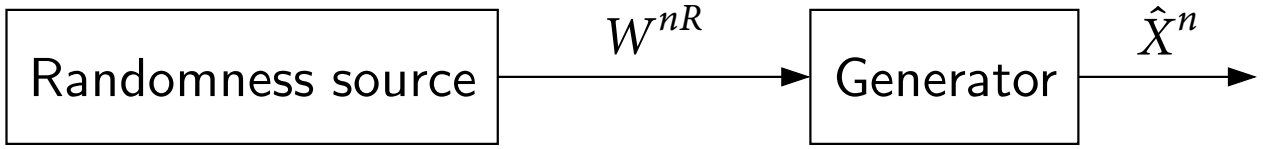
- Example: generating $X \sim \text{Bern}(1/3)$:



$$\mathsf{E}(L) = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot (2 + \mathsf{E}(L)) \;\Rightarrow\; \mathsf{E}(L) = 2,$$

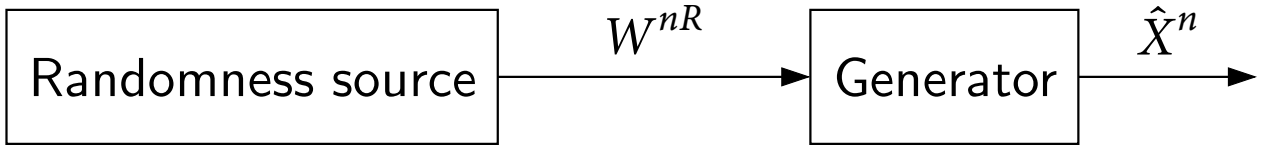
$H(1/3) = 0.9183$

Asymptotic randomness generation



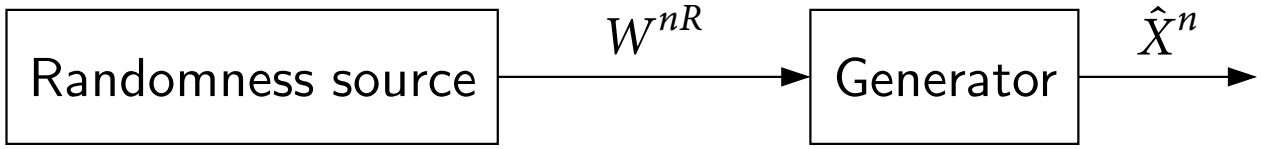
- W^{nR} is i.i.d. $\mathrm{Bern}(1/2)$ nR-bit sequence; $X \sim p(x)$

Asymptotic randomness generation



- W^{nR} is i.i.d. $\mathrm{Bern}(1/2)$ nR-bit sequence; $X \sim p(x)$
- Wish to generate $\hat{X}^n$ such that: $p_{\hat{X}^n}(x^n) \to \prod_{i=1}^n p_X(x_i)$

Asymptotic randomness generation



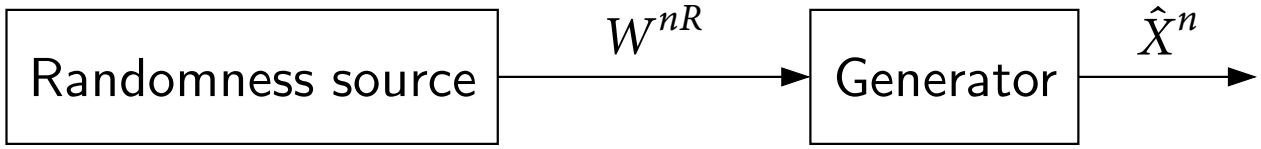
- W^{nR} is i.i.d. $\mathrm{Bern}(1/2)$ nR-bit sequence; $X \sim p(x)$
- Wish to generate $\hat{X}^n$ such that: $p_{\hat{X}^n}(x^n) \to \prod_{i=1}^n p_X(x_i)$

  More specifically, we want the total variation distance:

$$d_n = \sum_{x^n} \left| p_{\hat{X}^n}(x^n) - \prod_{i=1}^n p_X(x_i) \right| \to 0 \text{ as } n \to \infty$$

- What is the minimum randomness generation rate R?

Asymptotic randomness generation



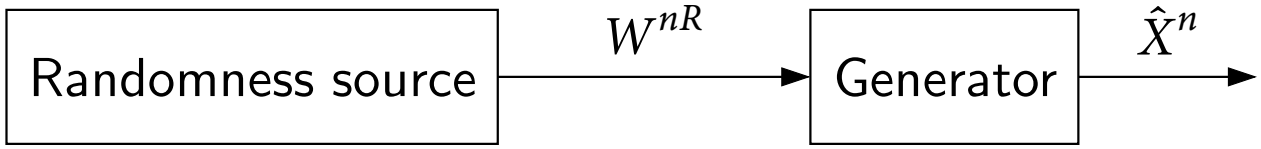
- W^{nR} is i.i.d. $\mathrm{Bern}(1/2)$ nR-bit sequence; $X \sim p(x)$
- Wish to generate $\hat{X}^n$ such that: $p_{\hat{X}^n}(x^n) \to \prod_{i=1}^n p_X(x_i)$

  More specifically, we want the total variation distance:

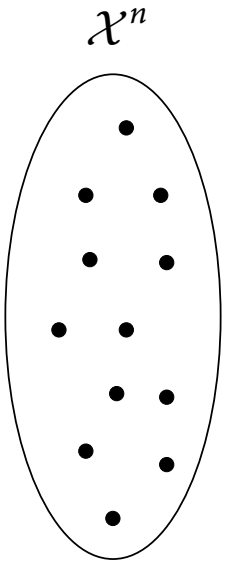
$$d_n = \sum_{x^n} \left| p_{\hat{X}^n}(x^n) - \prod_{i=1}^n p_X(x_i) \right| \to 0 \text{ as } n \to \infty$$

- What is the minimum randomness generation rate R?
- Turns out $\min R = H(X)$

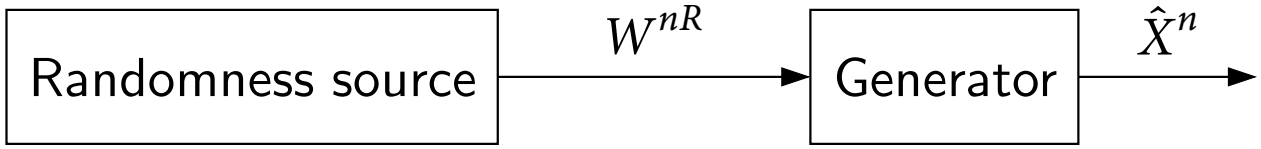
Random coding generation scheme



- Use random coding to show existence of codes (Han–Verdu 1993):
  - Randomly generate $X^n(w)$, $w \in [1 : 2^{nR}]$, $X^n(w) \sim \prod_{i=1}^n p_X(x_i)$
  - The generated code is revealed to the generator

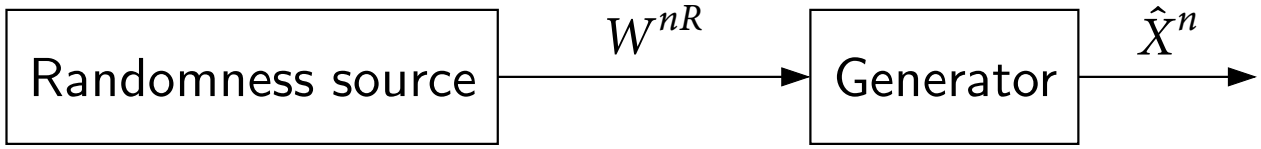


Random coding generation scheme



- Use random coding to show existence of codes (Han–Verdu 1993):
  - Randomly generate $X^n(w)$, $w \in [1 : 2^{nR}]$, $X^n(w) \sim \prod_{i=1}^n p_X(x_i)$
  - The generated code is revealed to the generator
  - Upon observing $W = w$, the generator outputs $\hat{X}^n = X^n(w)$

Random coding generation scheme



- Use random coding to show existence of codes (Han–Verdu 1993):
  - Randomly generate $X^n(w)$, $w \in [1:2^{nR}]$, $X^n(w) \sim \prod_{i=1}^n p_X(x_i)$
  - The generated code is revealed to the generator
  - Upon observing $W = w$, the generator outputs $\hat{X}^n = X^n(w)$
  - Can show that: $\mathsf{E}(D_n) \to 0$ if $R > H(X)$
  - Hence, there exists a sequence of codes with $d_n \to 0$ if $R > H(X)$
- This type of existence proof extends to distributed generation settings

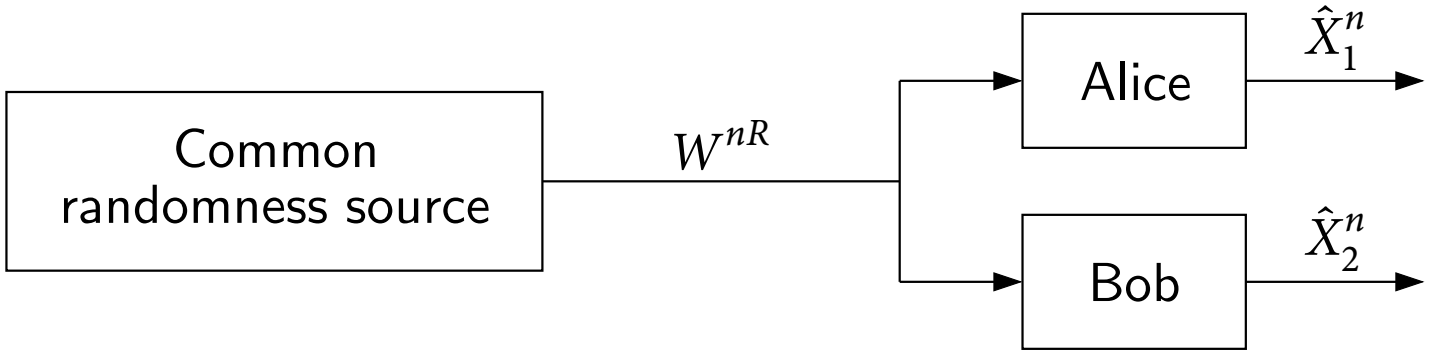
Summary of centralized randomness generation

- One shot:
  - Entropy is the limit — same as for zero error compression
  - Both use optimal prefix codes
- Asymptotic:
  - Entropy is the limit — same as for lossless compression
  - Use random coding (key technique in information theory)
- Same limit for one shot and asymptotic — a more relaxed condition

Outline

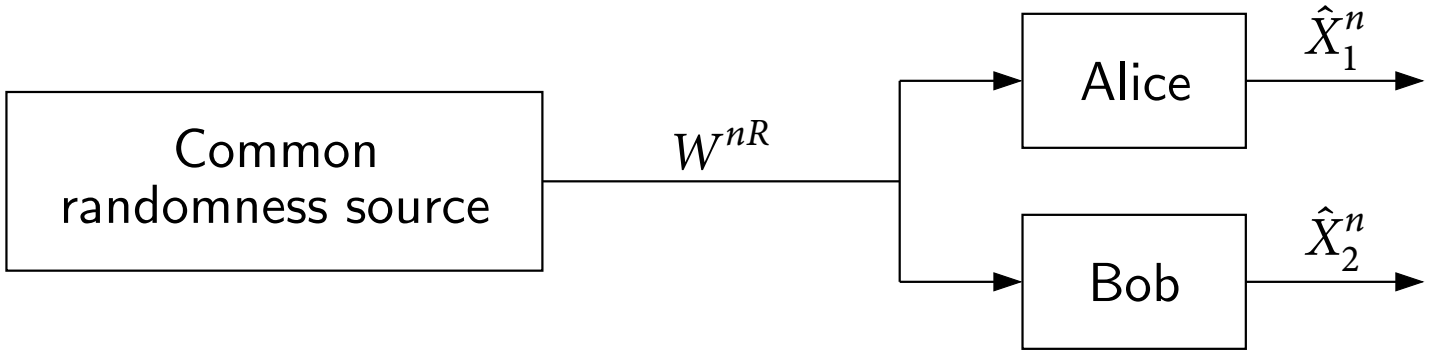
- Centralized randomness generation:
  - One shot
  - Asymptotic
- Distributed randomness generation:
  - Asymptotic
  - One shot – recent results
- Channel synthesis (common randomness/communication tradeoff):
  - Asymptotic
  - One shot

Asymptotic distributed randomness generation



- W^{nR} i.i.d. $\mathrm{Bern}(1/2)$ sequence; $(X_1, X_2) \sim p(x_1, x_2)$
- Alice and Bob each has unlimited independent local randomness

Asymptotic distributed randomness generation

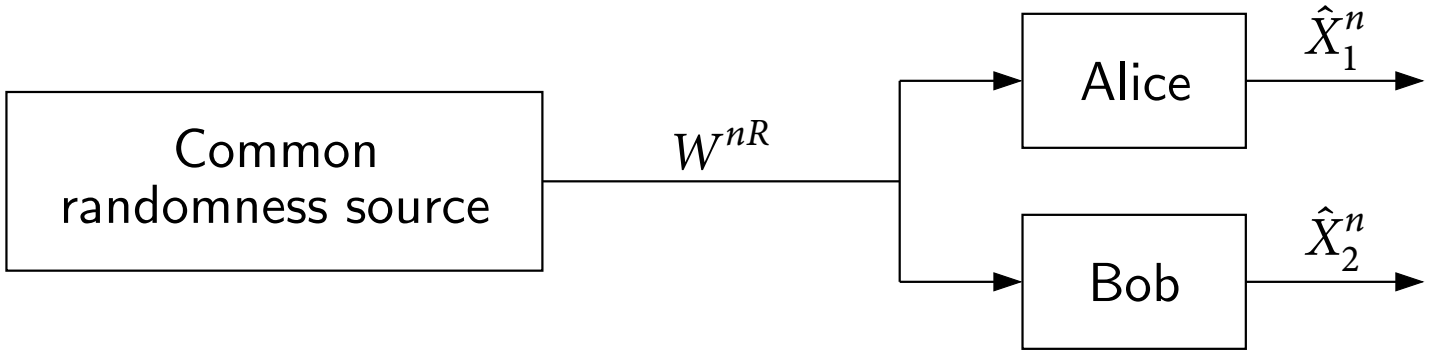


- W^{nR} i.i.d. $\mathrm{Bern}(1/2)$ sequence; $(X_1, X_2) \sim p(x_1, x_2)$
- Alice and Bob each has unlimited independent local randomness
- Alice generates $\hat{X}_1^n$ and Bob generates $\hat{X}_2^n$ such that:

$$d_n = \sum_{(x_1^n, x_2^n)} \left| p_{\hat{X}_1^n, \hat{X}_2^n}(x_1^n, x_2^n) - \prod_{i=1}^{n} p_{X_1, X_2}(x_{1i}, x_{2i}) \right| \to 0$$

- What is the minimum common randomness rate R?

Asymptotic distributed randomness generation



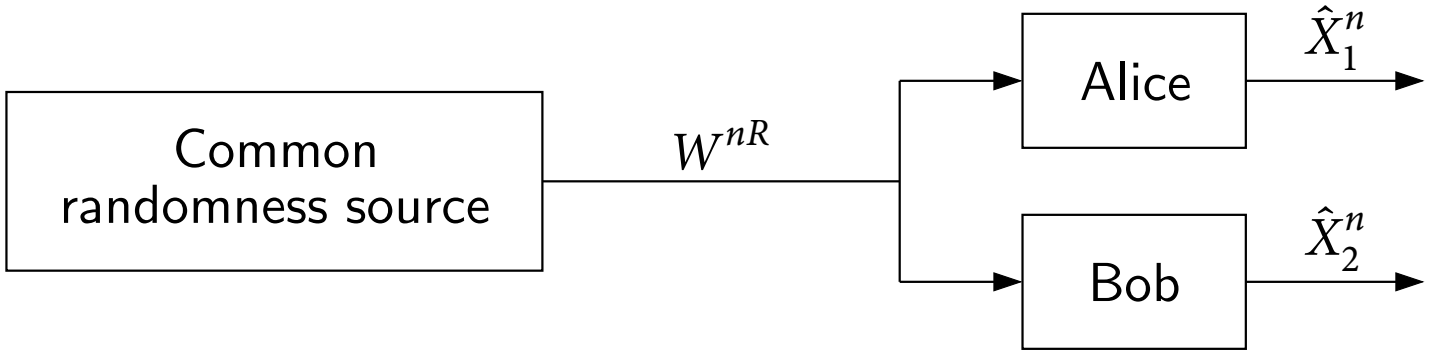
- W^{nR} i.i.d. $\mathrm{Bern}(1/2)$ sequence; $(X_1, X_2) \sim p(x_1, x_2)$
- Alice and Bob each has unlimited independent local randomness
- Alice generates $\hat{X}_1^n$ and Bob generates $\hat{X}_2^n$ such that:

$$d_n = \sum_{(x_1^n, x_2^n)} \left| p_{\hat{X}_1^n, \hat{X}_2^n}(x_1^n, x_2^n) - \prod_{i=1}^{n} p_{X_1, X_2}(x_{1i}, x_{2i}) \right| \to 0$$

- What is the minimum common randomness rate R?
- Wyner (1975) showed that:

$$\min R = J(X_1; X_2) = \min_{X_1 \to U \to X_2} I(X_1, X_2; U) \quad \text{Wyner common information}$$

Asymptotic distributed randomness generation



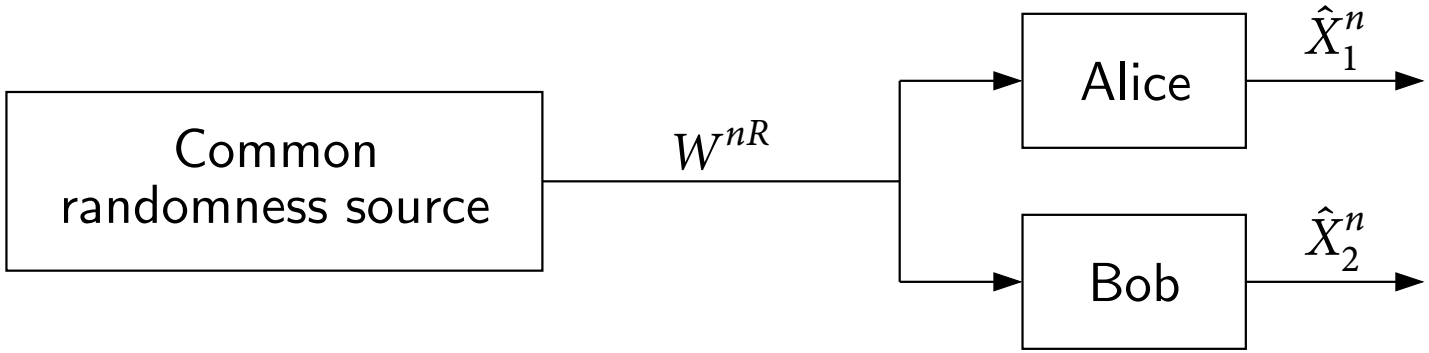
- Wyner (1975) showed that:

$$\min R = J(X_1; X_2) = \min_{X_1 \to U \to X_2} I(X_1, X_2; U) \quad \text{Wyner common information}$$

- Mutual information:

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \quad \text{Another measure of common information} \end{aligned}$$

Asymptotic distributed randomness generation



- Wyner (1975) showed that:

$$\min R = J(X_1; X_2) = \min_{X_1 \to U \to X_2} I(X_1, X_2; U) \quad \text{Wyner common information}$$

- Mutual information:

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \quad \text{Another measure of common information} \end{aligned}$$

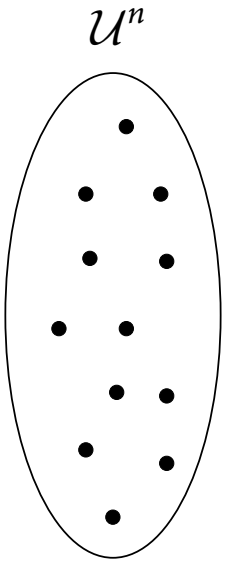
- It is not difficult to see that: $I(X_1; X_2) \le J(X_1; X_2) \le \min\{H(X_1), H(X_2)\}$

Wyner's generation scheme

- Show $R > J(X_1; X_2) = \min_{X_1 \to U \to X_2} I(X_1, X_2; U)$ suffices

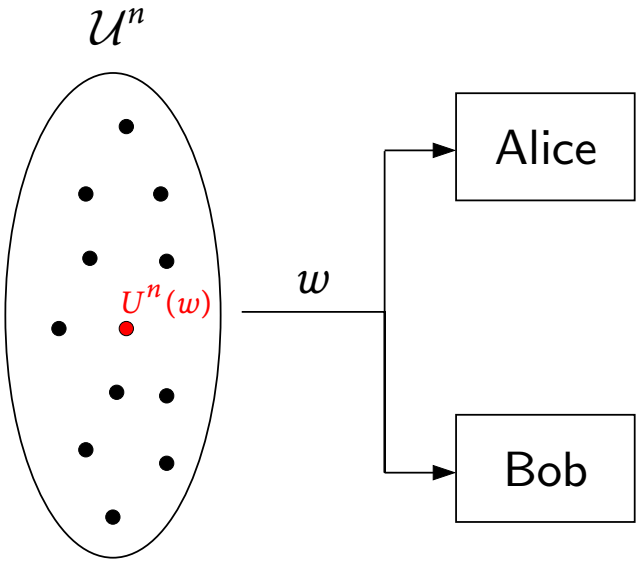
Wyner's generation scheme

- Let $p(u)p(x_1|u)p(x_2|u)$ ($p(x_1, x_2)$ given) attain $J(X_1; X_2)$
- Generate $U^n(w)$, $w \in [1 : 2^{nR}]$, each $\sim \prod_{i=1}^n p_U(u_i)$
- Codebook revealed to Alice and Bob before communication
- Source outputs $W = w$
- Alice generates $\hat{X}_1^n \sim \prod_{i=1}^n p_{X_1|U}(\hat{x}_{1i}|u_i(w))$
- Bob generates $\hat{X}_2^n \sim \prod_{i=1}^n p_{X_2|U}(\hat{x}_{1i}|u_i(w))$



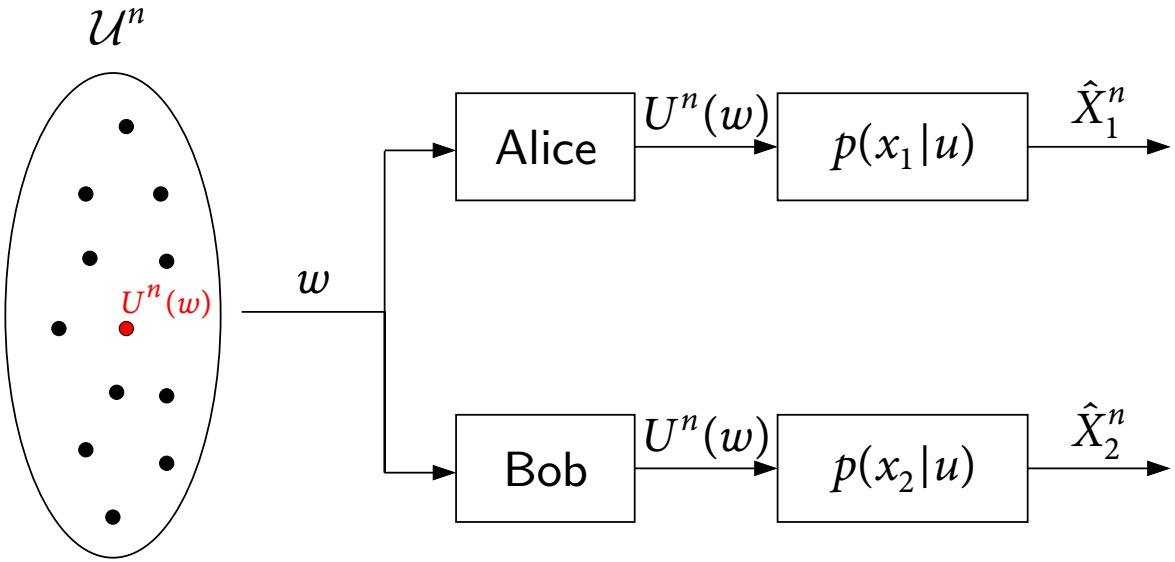
Wyner's generation scheme

- Let $p(u)p(x_1|u)p(x_2|u)$ ($p(x_1, x_2)$ given) attain $J(X_1; X_2)$
- Generate $U^n(w)$, $w \in [1 : 2^{nR}]$, each $\sim \prod_{i=1}^n p_U(u_i)$
- Codebook revealed to Alice and Bob before communication
- Source outputs $W = w$
- Alice generates $\hat{X}_1^n \sim \prod_{i=1}^n p_{X_1|U}(\hat{x}_{1i}|u_i(w))$
- Bob generates $\hat{X}_2^n \sim \prod_{i=1}^n p_{X_2|U}(\hat{x}_{1i}|u_i(w))$



Wyner's generation scheme

- Let $p(u)p(x_1|u)p(x_2|u)$ ($p(x_1, x_2)$ given) attain $J(X_1; X_2)$
- Generate $U^n(w)$, $w \in [1 : 2^{nR}]$, each $\sim \prod_{i=1}^n p_U(u_i)$
- Codebook revealed to Alice and Bob before communication
- Source outputs $W = w$ from which Alice and Bob recover $U^n(w)$
- Alice generates $\hat{X}_1^n \sim \prod_{i=1}^n p_{X_1|U}(\hat{x}_{1i}|u_i(w))$
- Bob generates $\hat{X}_2^n \sim \prod_{i=1}^n p_{X_2|U}(\hat{x}_{1i}|u_i(w))$



Wyner's generation scheme

- Let $p(u)p(x_1|u)p(x_2|u)$ ($p(x_1, x_2)$ given) attain $J(X_1; X_2)$
- Generate $U^n(w)$, $w \in [1 : 2^{nR}]$, each $\sim \prod_{i=1}^n p_U(u_i)$
- Codebook revealed to Alice and Bob before communication
- Source outputs $W = w$ from which Alice and Bob recover $U^n(w)$
- Alice generates $\hat{X}_1^n \sim \prod_{i=1}^n p_{X_1|U}(\hat{x}_{1i}|u_i(w))$
- Bob generates $\hat{X}_2^n \sim \prod_{i=1}^n p_{X_2|U}(\hat{x}_{1i}|u_i(w))$
- Using asymptotic centralized generation scheme: $R > H(U)$ suffices

Wyner's generation scheme

- Let $p(u)p(x_1|u)p(x_2|u)$ ($p(x_1, x_2)$ given) attain $J(X_1; X_2)$
- Generate $U^n(w)$, $w \in [1 : 2^{nR}]$, each $\sim \prod_{i=1}^n p_U(u_i)$
- Codebook revealed to Alice and Bob before communication
- Source outputs $W = w$ from which Alice and Bob recover $U^n(w)$
- Alice generates $\hat{X}_1^n \sim \prod_{i=1}^n p_{X_1|U}(\hat{x}_{1i}|u_i(w))$
- Bob generates $\hat{X}_2^n \sim \prod_{i=1}^n p_{X_2|U}(\hat{x}_{1i}|u_i(w))$
- Using asymptotic centralized generation scheme: $R > H(U)$ suffices
- Wyner (1975) showed that only $R > I(X_1, X_2; U)$ ($\leq H(U)$) suffices

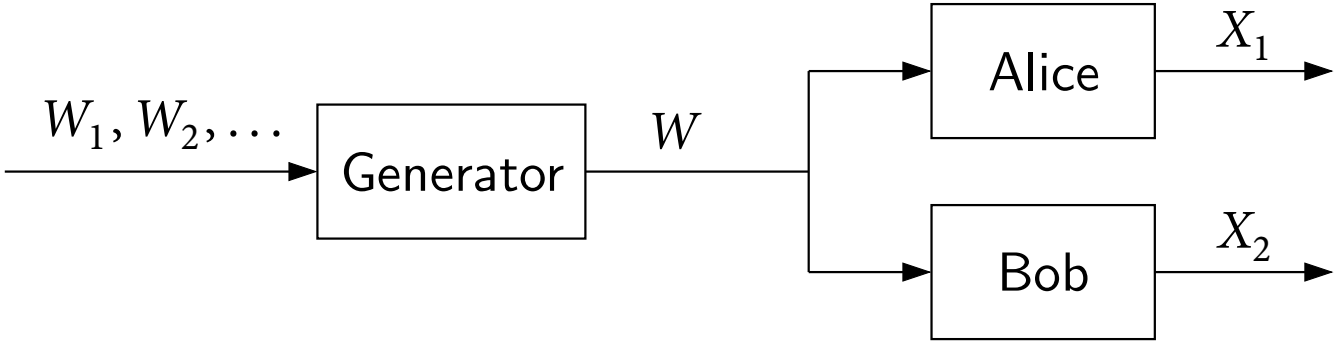
  Proof uses a covering lemma similar to lossy compression

Wyner's generation scheme

- Let $p(u)p(x_1|u)p(x_2|u)$ ($p(x_1, x_2)$ given) attain $J(X_1; X_2)$
- Generate $U^n(w)$, $w \in [1 : 2^{nR}]$, each $\sim \prod_{i=1}^n p_U(u_i)$
- Codebook revealed to Alice and Bob before communication
- Source outputs $W = w$ from which Alice and Bob recover $U^n(w)$
- Alice generates $\hat{X}_1^n \sim \prod_{i=1}^n p_{X_1|U}(\hat{x}_{1i}|u_i(w))$
- Bob generates $\hat{X}_2^n \sim \prod_{i=1}^n p_{X_2|U}(\hat{x}_{1i}|u_i(w))$
- Using asymptotic centralized generation scheme: $R > H(U)$ suffices
- Wyner (1975) showed that only $R > I(X_1, X_2; U)$ ($\leq H(U)$) suffices

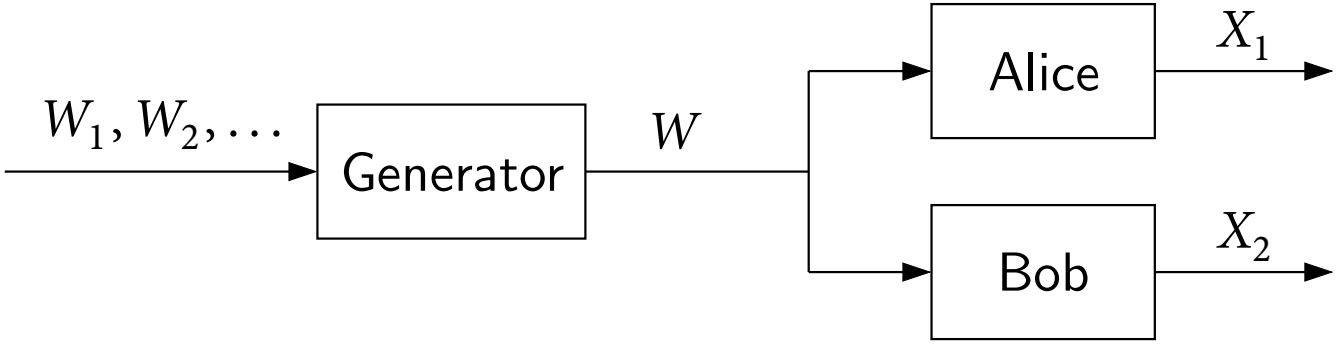
  Proof uses a covering lemma similar to lossy compression
- Converse ($\min R \geq J(X_1; X_2)$) uses information theoretic inequalities

One shot distributed randomness generation



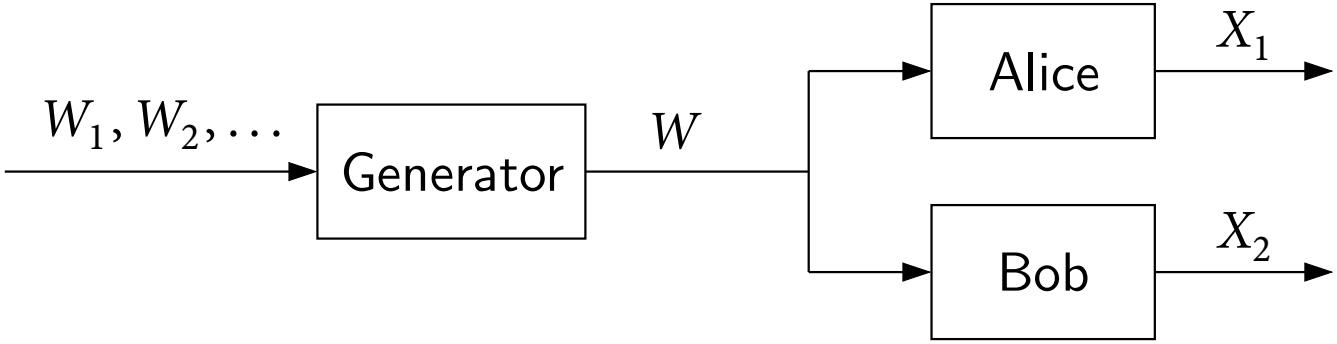
- $W_1, W_2, \ldots$ i.i.d. $\mathrm{Bern}(1/2)$; $(X_1, X_2) \sim p(x_1, x_2)$
- Generator outputs common randomness variable W using prefix code

One shot distributed randomness generation



- $W_1, W_2, \ldots$ i.i.d. $\mathrm{Bern}(1/2)$; $(X_1, X_2) \sim p(x_1, x_2)$
- Generator outputs common randomness variable W using prefix code
- Let L be number of W_i bits used to generate W
- What is the minimum $\mathsf{E}(L)$?

One shot distributed randomness generation



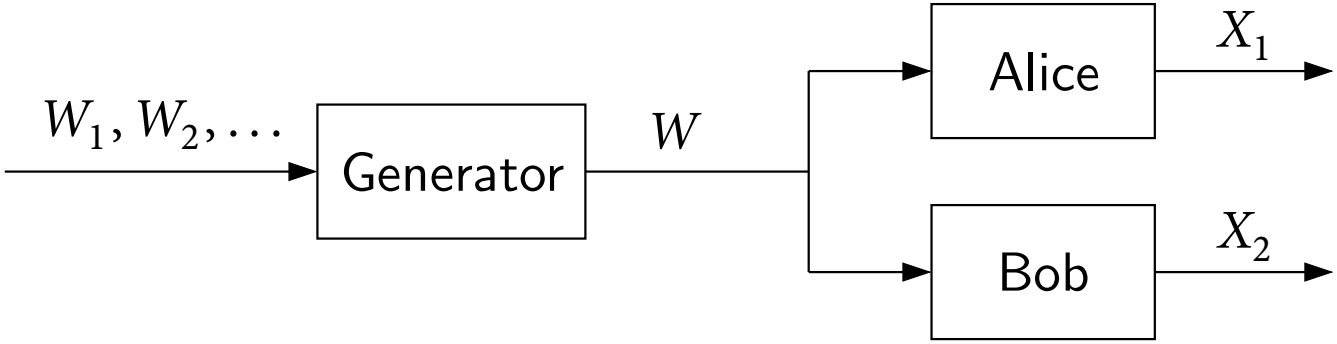
- What is the minimum $\mathsf{E}(L)$?
- By Knuth–Yao (1976) can show:

$$G(X_1; X_2) \leq \min \mathsf{E}(L) < G(X_1; X_2) + 2, \text{ where}$$

$$G(X_1; X_2) = \min_{X_1 \to W \to X_2} H(W)$$ common entropy (Kumar–Li–EG 2014)

Another measure of common information (in addition to I and J)

One shot distributed randomness generation



- What is the minimum $\mathsf{E}(L)$?
- By Knuth–Yao (1976) can show:

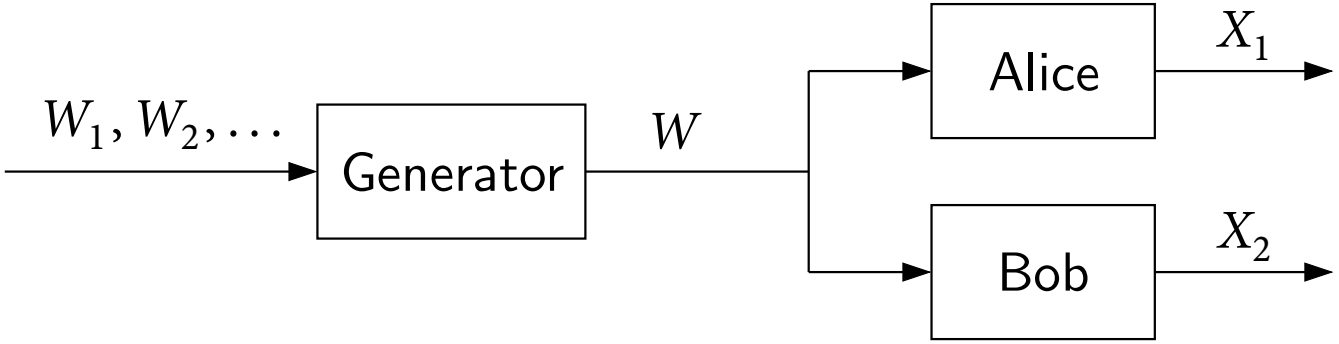
$$G(X_1; X_2) \leq \min \mathsf{E}(L) < G(X_1; X_2) + 2, \text{ where}$$

$$G(X_1; X_2) = \min_{X_1 \to W \to X_2} H(W)$$ common entropy (Kumar–Li–EG 2014)

Another measure of common information (in addition to I and J)

- $G(X_1; X_2)$ is very difficult to compute except for simple cases

(X_1, X_2) discrete

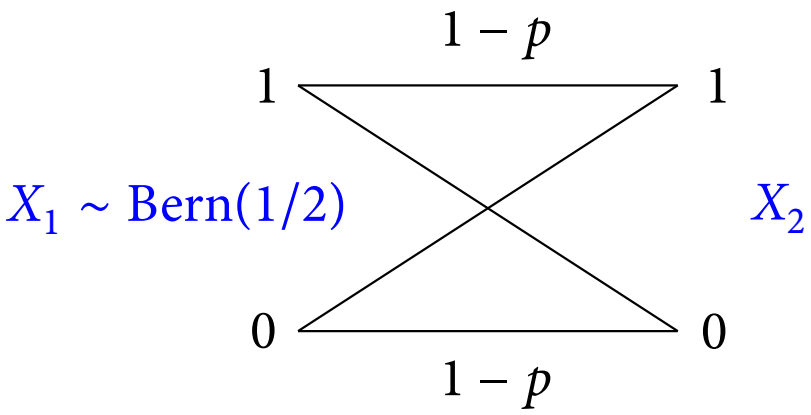


$$I(X_1; X_2) \le J(X_1; X_2) \le G(X_1; X_2) \le \min\{H(X_1), H(X_2)\}$$

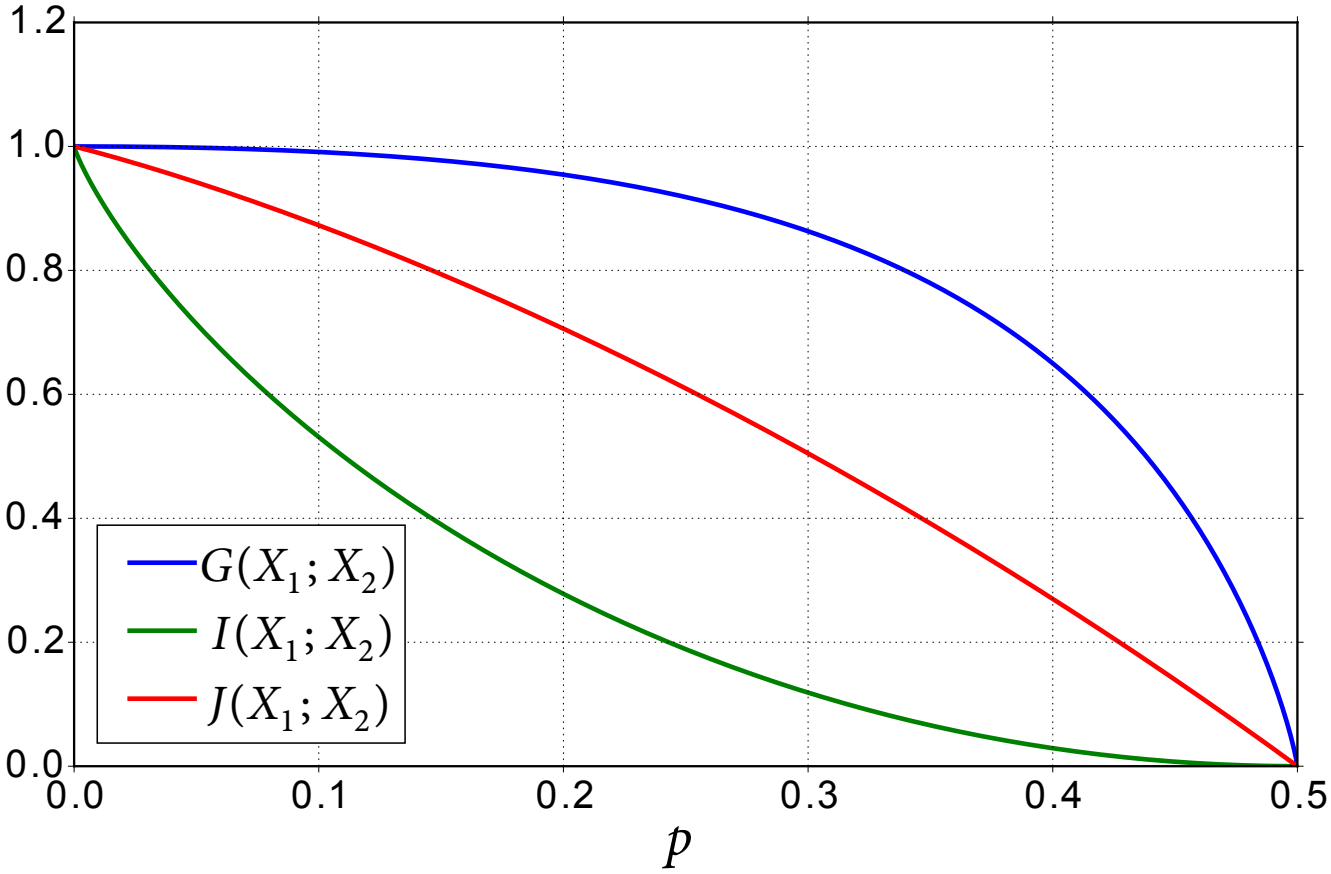
$I(X_1; X_2)$:	mutual information
$J(X_1; X_2) = \min_{X_1 \to U \to X_2} I(X_1, X_2; U)$:	Wyner common information
$G(X_1; X_2) = \min_{X_1 \to W \to X_2} H(W)$:	common entropy

- All inequalities can be strict

Example: (X_1, X_2) DSBS



Example: (X_1, X_2) DSBS

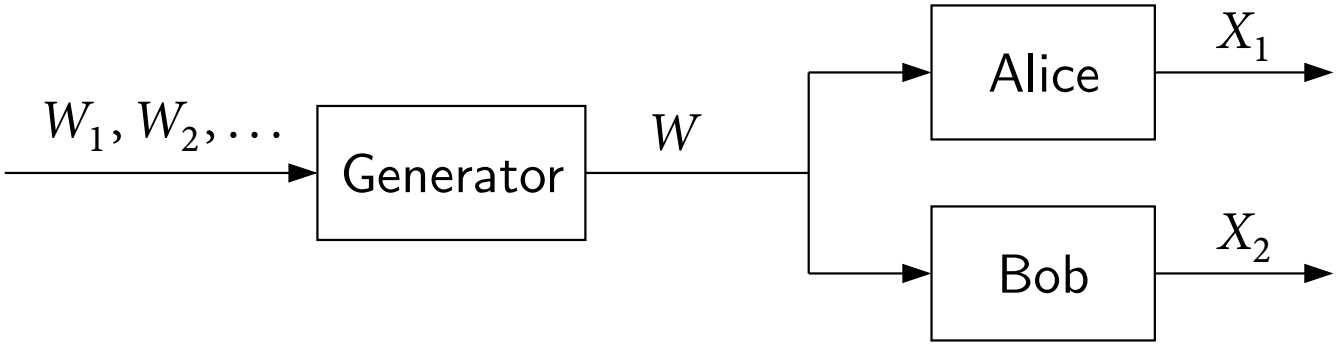


$$G(X_1; X_2) = H((1/2 - p)/(1 - p)) \text{ for } p \leq 1/2$$

$$J(X_1; X_2) = 1 + H(p) - 2H(\alpha), \text{ where } 2\alpha(1 - \alpha) = p$$

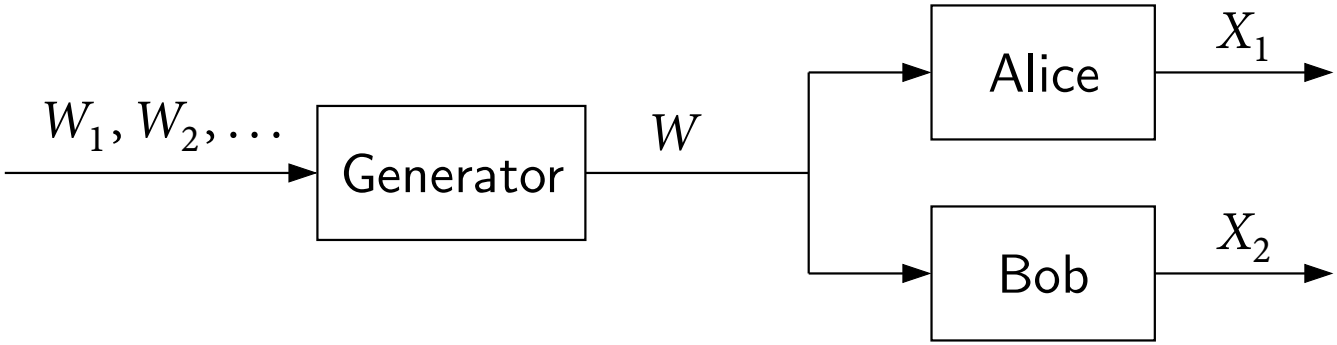
$$I(X_1; X_2) = 1 - H(p); \; H(X_1) = H(X_2) = 1$$

(X_1, X_2) continuous (Li–EG 2016)



- We still have: $I(X_1; X_2) \leq J(X_1; X_2) \leq G(X_1; X_2)$
- No known general upper bound on G or J $\big(H(X_1) = H(X_2) = \infty\big)$

(X_1, X_2) continuous (Li–EG 2016)



- We still have: $I(X_1; X_2) \le J(X_1; X_2) \le G(X_1; X_2)$
- No known general upper bound on G or J $\big(H(X_1) = H(X_2) = \infty\big)$
- Example: (X_1, X_2) Gaussian with $\sigma_1 = \sigma_2 = 1$, $0 \le \rho < 1$

$$I(X_1; X_2) = \frac{1}{2} \log\Big(\frac{1}{1-\rho^2}\Big)$$

$$J(X_1; X_2) = \frac{1}{2} \log\Big(\frac{1+\rho}{1-\rho}\Big), \; U \sim \mathrm{N}(0, \rho), \; X_j = U + Z_j, \; Z_j \sim \mathrm{N}(0, 1-\rho)$$

- Is $G(X_1; X_2)$ also finite?

Upper bound on G

- For (X_1, X_2) with log-concave pdf:

Theorem (Li–EG 2016)

$$I(X_1; X_2) \leq J(X_1; X_2) \leq G(X_1; X_2) \leq I(X_1; X_2) + 24$$

Upper bound on G

- For (X_1, X_2) with log-concave pdf:

Theorem (Li–EG 2016)

$$I(X_1; X_2) \leq J(X_1; X_2) \leq G(X_1; X_2) \leq I(X_1; X_2) + 24$$

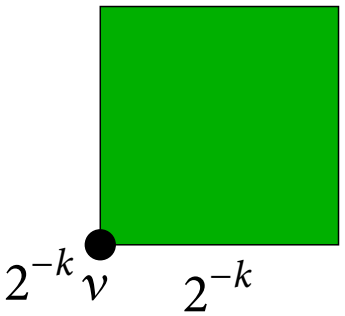
- Result extends to generation of n continuous random variables

Outline of proof

- (X_1, X_2) uniformly distributed over a set in $\mathbb{R}^2$:
  - Construct W using dyadic decomposition scheme
  - Bound $H(W)$ by $I(X_1; X_2) + \text{const.}$ for S convex using erosion entropy
- Extend proof to general pdf:
  - Apply scheme to uniform over positive hypograph of pdf
  - Bound $G(X_1; X_2)$ by $I(X_1; X_2) + 24$ for log-concave pdf

$(X_1, X_2) \sim \mathrm{Unif}(S)$

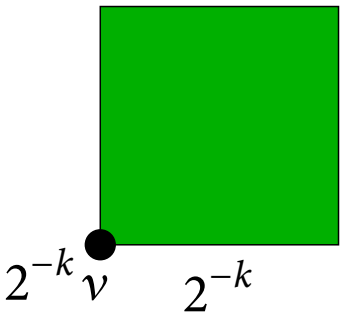
- Dyadic square:



$k \in \mathbb{Z}, \ v \in \mathbb{Z}^2$

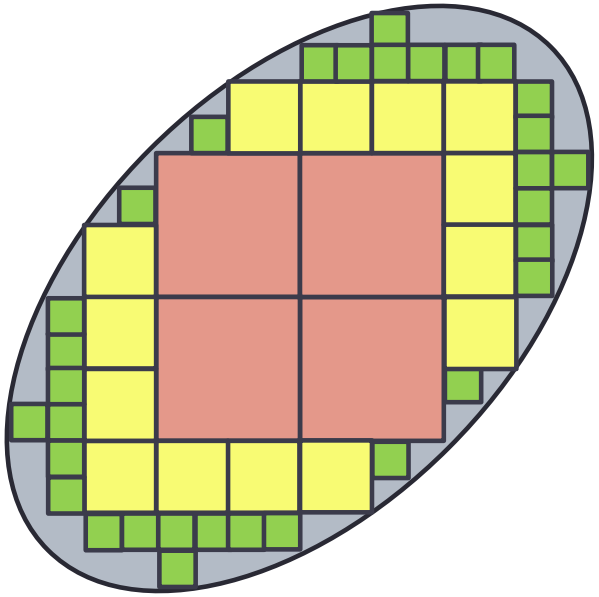
$(X_1, X_2) \sim \mathrm{Unif}(S)$

- Dyadic square:



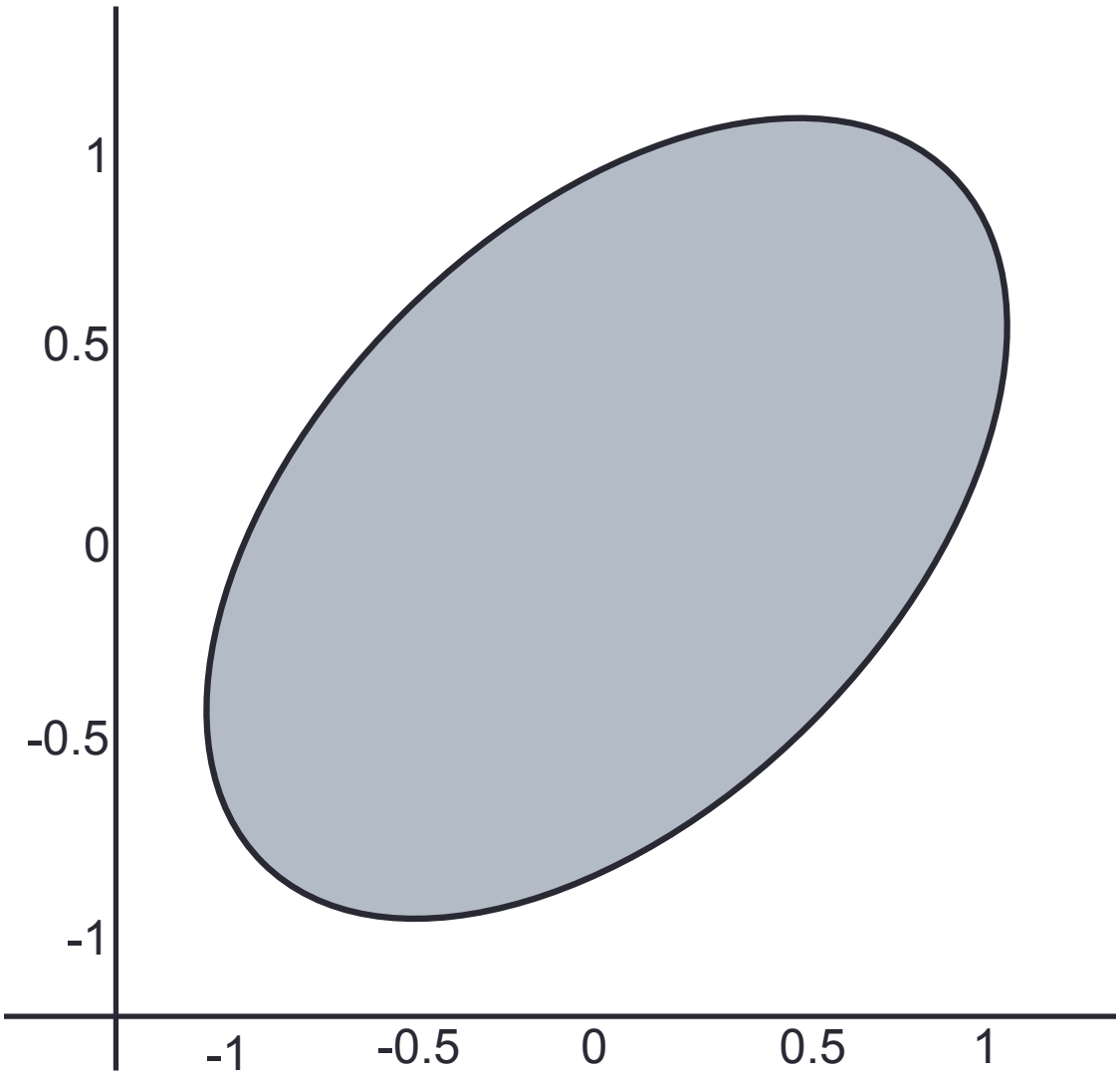
$k \in \mathbb{Z}, \ v \in \mathbb{Z}^2$

- Dyadic decomposition: Partition S into largest possible dyadic squares

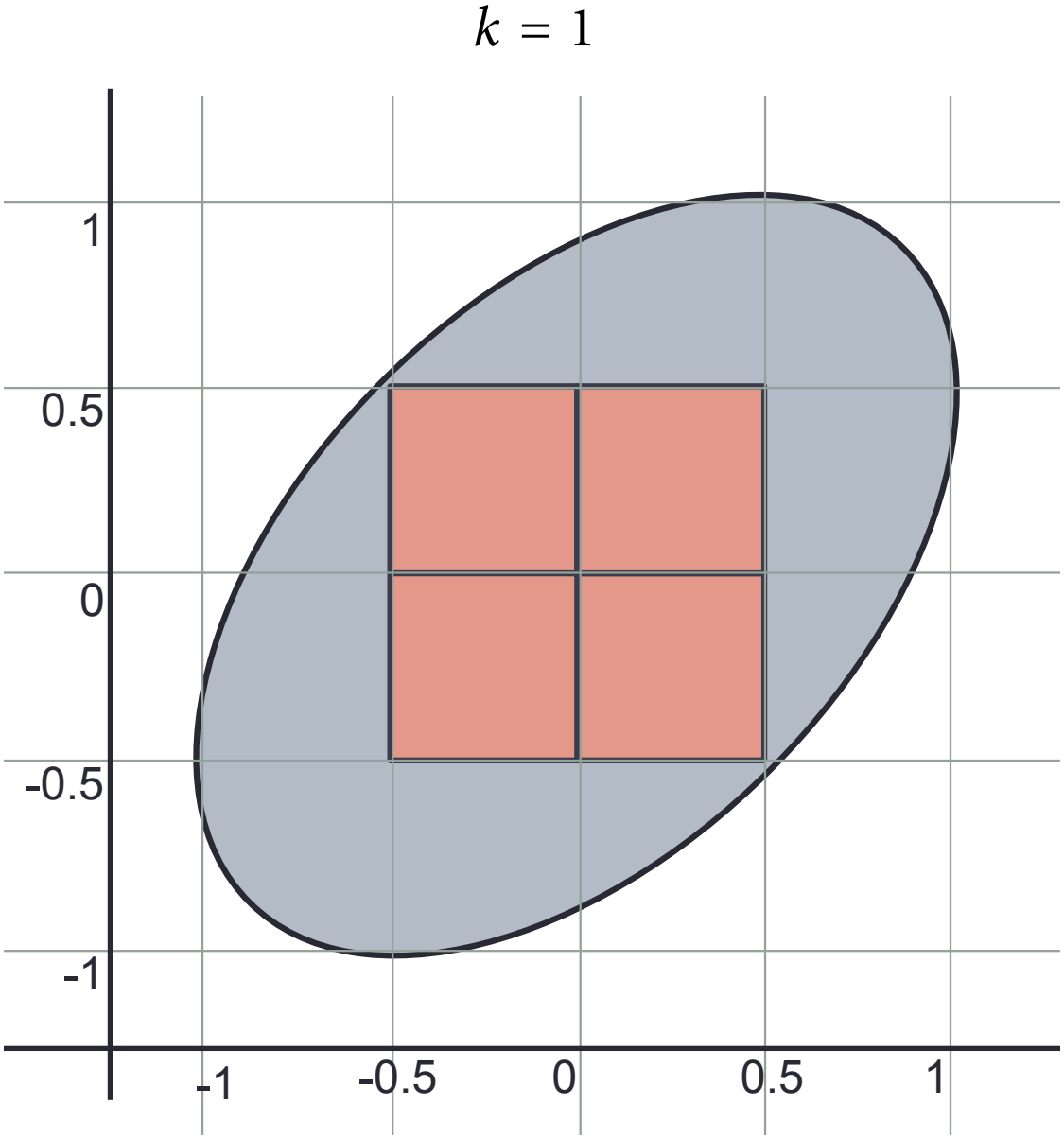


Dyadic decomposition for an ellipse

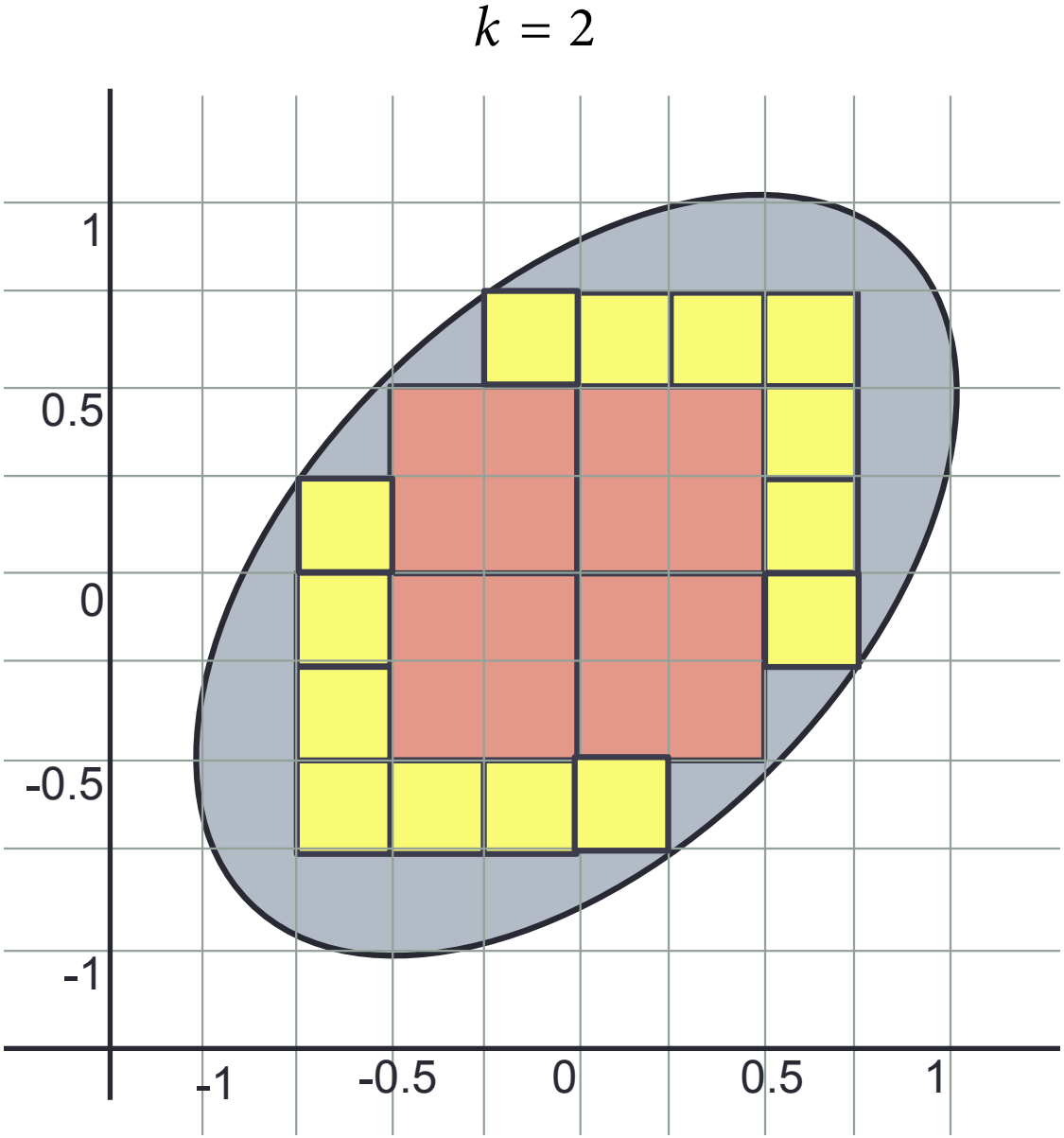
$k < 1$: no dyadic squares inside S



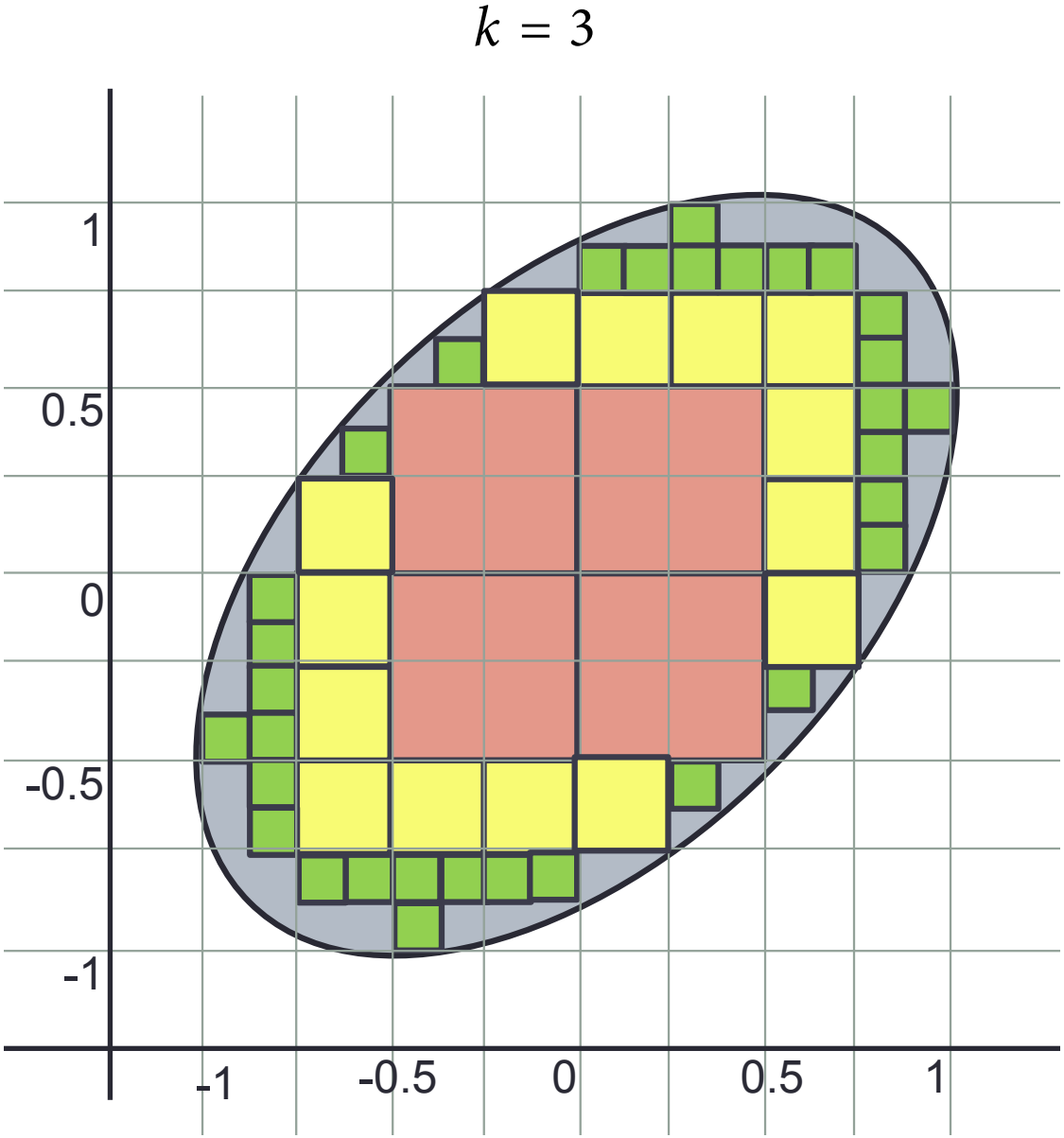
Dyadic decomposition for an ellipse



Dyadic decomposition for an ellipse



Dyadic decomposition for an ellipse

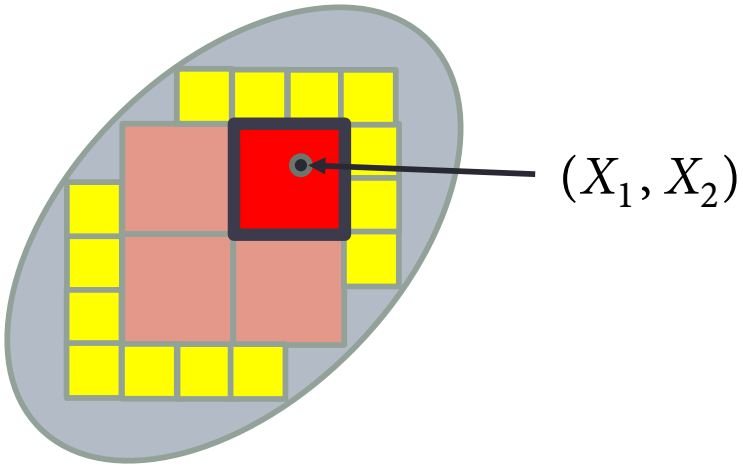


Constructing W from dyadic decomposition

- W_{D} is the square containing (X_1, X_2)

  Conditioned on W_{D}, X_1 and X_2 are uniform over dyadic square

  Hence, $X_1 \to W_{\mathrm{D}} \to X_2$

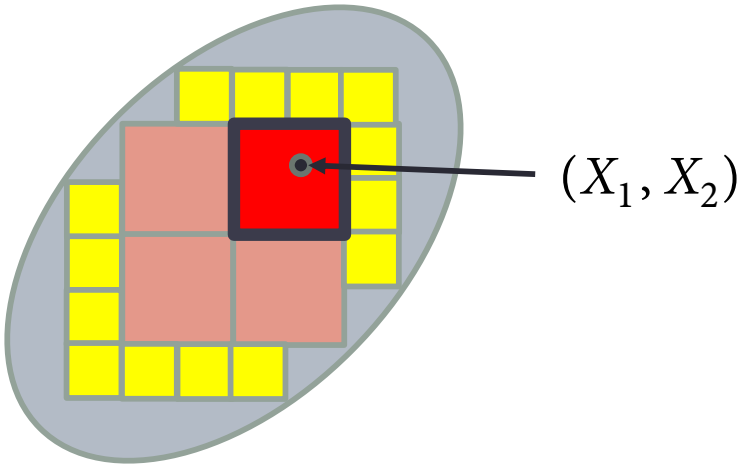


Constructing W from dyadic decomposition

- W_{D} is the square containing (X_1, X_2)

  Conditioned on W_{D}, X_1 and X_2 are uniform over dyadic square

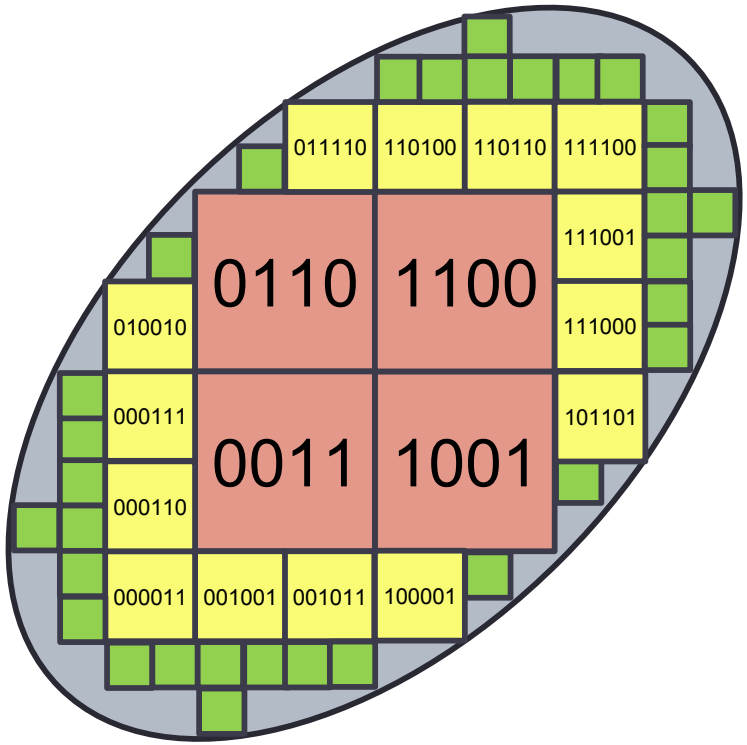
  Hence, $X_1 \to W_{\mathrm{D}} \to X_2$



- And, $H(W_{\mathrm{D}}) \geq G(X_1; X_2) = \min_{X_1 \to W \to X_2} H(W)$

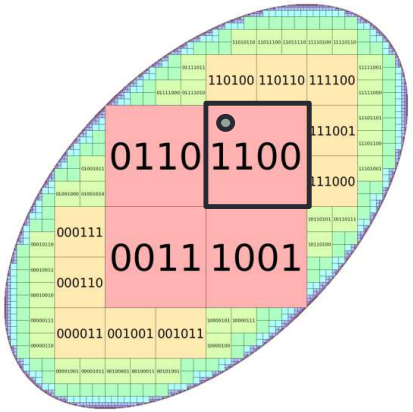
Generating W_{D} from coin flips

- Generate W_{D} using a prefix code



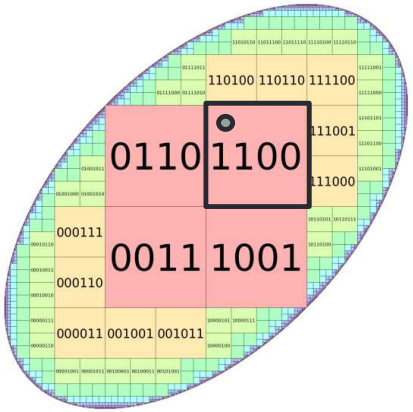
- Dyadic decomposition and code revealed to all parties

Generation scheme for $(X_1, X_2) \sim \mathrm{Unif}(S)$

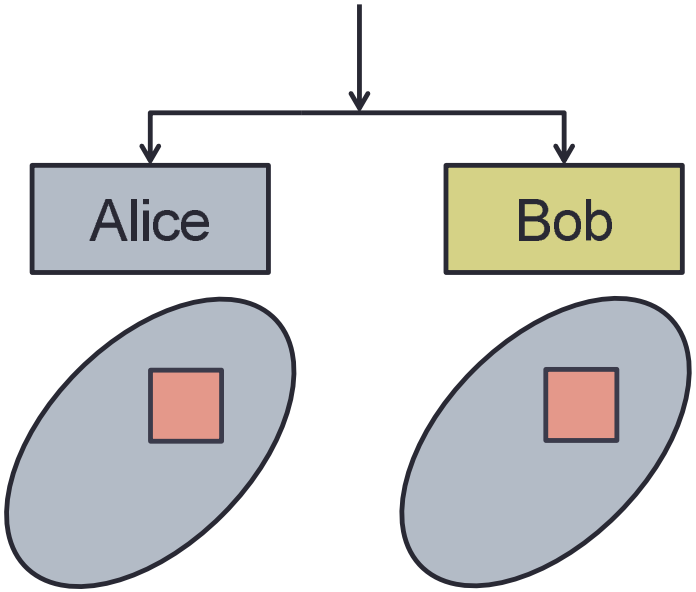


$$1100 \to W_{\mathrm{D}} = w$$

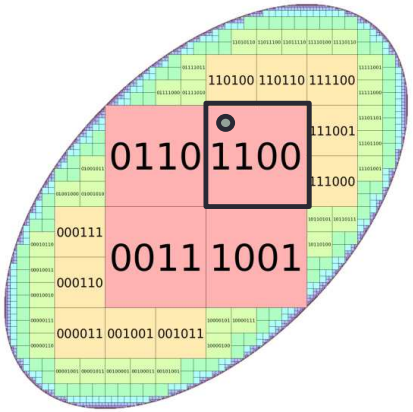
Generation scheme for $(X_1, X_2) \sim \mathrm{Unif}(S)$



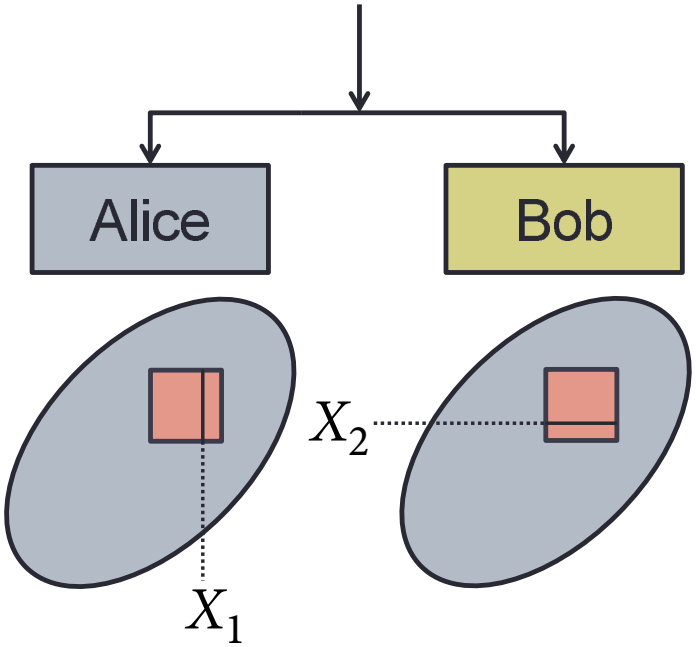
$$1100 \rightarrow W_{\mathrm{D}} = w$$



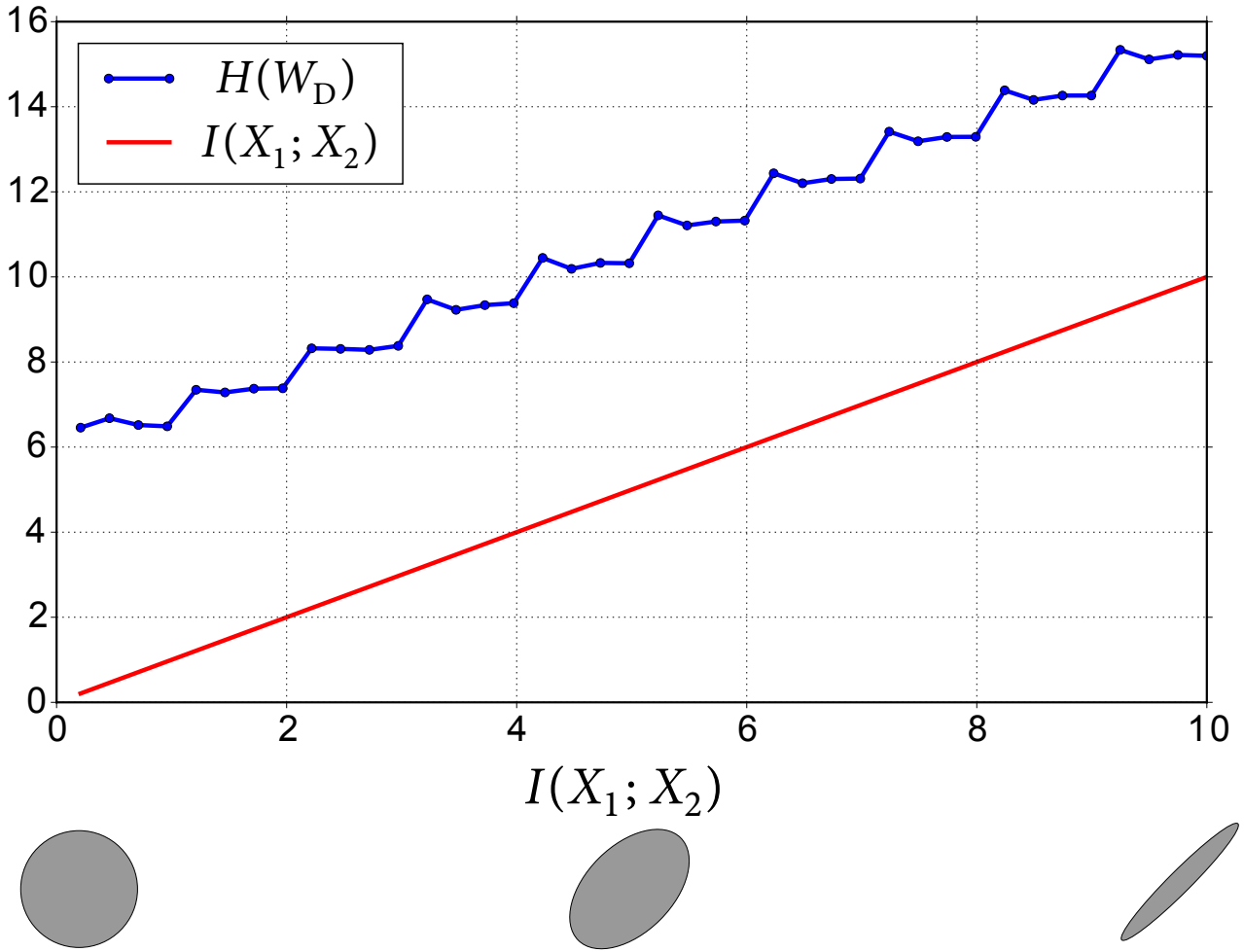
Generation scheme for $(X_1, X_2) \sim \mathrm{Unif}(S)$



$$1100 \rightarrow W_{\mathrm{D}} = w$$



(X_1, X_2) uniform over ellipse



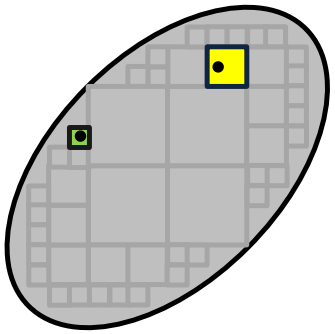
- Gap between I and G of ≤ 6 bits (vs. 24 for general log-concave pdf)

Bounding $H(W_{\mathrm{D}})$

- $H(W_{\mathrm{D}}) = \mathsf{E}\Big[-\log(L_{\mathrm{D}}^2/A(S))\Big]$, $A(S)$ is area of S

  L_{D}: side length of largest dyadic square (X_1, X_2) in

  $\Rightarrow (1/2)(H(W_{\mathrm{D}}) - \log A(S)) = \mathsf{E}\Big[-\log L_{\mathrm{D}}\Big]$

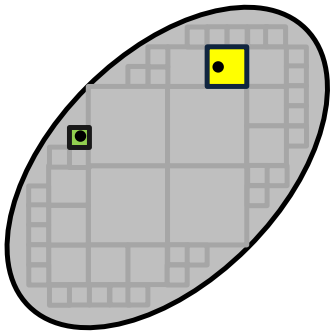


Bounding $H(W_{\mathrm{D}})$

- $H(W_{\mathrm{D}}) = \mathsf{E}\Big[-\log(L_{\mathrm{D}}^2/A(S))\Big]$, $A(S)$ is area of S

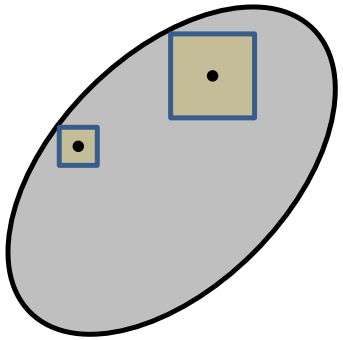
  L_{D}: side length of largest dyadic square (X_1, X_2) in

  $\Rightarrow (1/2)(H(W_{\mathrm{D}}) - \log A(S)) = \mathsf{E}\Big[-\log L_{\mathrm{D}}\Big]$



- Erosion entropy: $h_{\ominus}(S) = \mathsf{E}\Big[-\log L_{\mathrm{C}}\Big]$

  L_{C}: side length of largest square centered at (X_1, X_2)

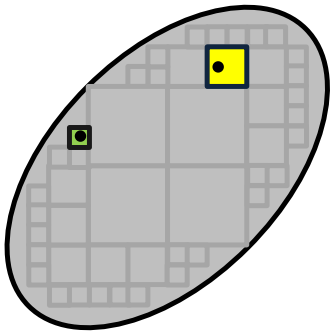


Bounding $H(W_{\mathrm{D}})$

- $H(W_{\mathrm{D}}) = \mathsf{E}\Big[-\log(L_{\mathrm{D}}^2/A(S))\Big]$, $A(S)$ is area of S

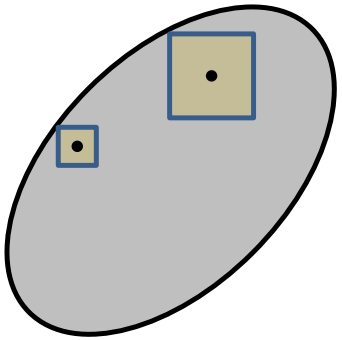
  L_{D}: side length of largest dyadic square (X_1, X_2) in

  $\Rightarrow (1/2)(H(W_{\mathrm{D}}) - \log A(S)) = \mathsf{E}\Big[-\log L_{\mathrm{D}}\Big]$



- Erosion entropy: $h_{\ominus}(S) = \mathsf{E}\Big[-\log L_{\mathrm{C}}\Big]$

  L_{C}: side length of largest square centered at (X_1, X_2)



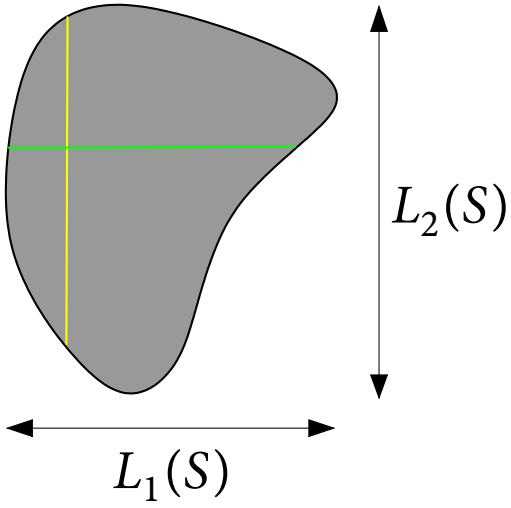
Lemma 1

$$\frac{1}{2}(H(W_{\mathrm{D}}) - \log A(S)) \le h_{\ominus}(S) + 2$$

Bounding erosion entropy for orthogonally convex set

- Suppose S is orthogonally convex, i.e.,

  intersection with each axis-aligned line is connected



$A(S)$: area of S

- Then it can be shown that:

Lemma 2

$$h_\ominus = \mathsf{E}[-\log L_{\mathrm{C}}] \leq \log\left(\frac{L_1(S) + L_2(S)}{A(S)}\right) + \log e$$

Bounding $H(W_{\mathrm{D}})$

- Substituting from Lemma 2 into Lemma 1, we have:

$$H(W_{\mathrm{D}}) \leq \log\left(\frac{(L_1(S) + L_2(S))^2}{A(S)}\right) + 4 + 2\log e$$

Depends on perimeter to area ratio of S

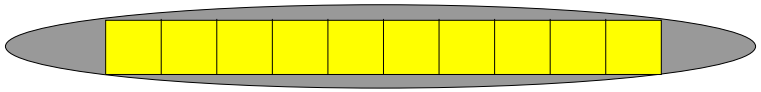
Bounding $H(W_{\mathrm{D}})$

- Substituting from Lemma 2 into Lemma 1, we have:

$$H(W_{\mathrm{D}}) \leq \log\left(\frac{(L_1(S) + L_2(S))^2}{A(S)}\right) + 4 + 2\log e$$

  Depends on perimeter to area ratio of S

- A flat S has high perimeter to area ratio $\Rightarrow$ high $H(W_{\mathrm{D}})$



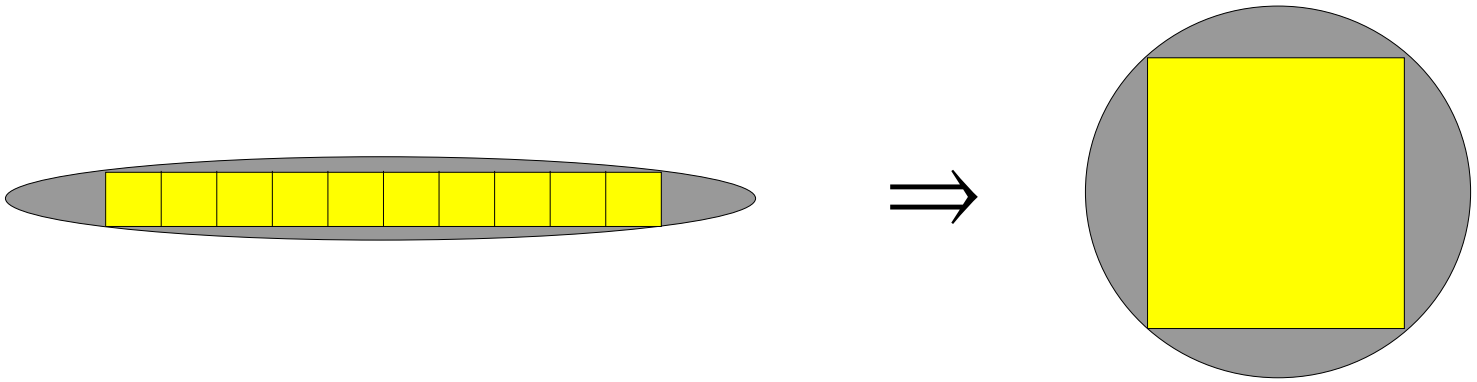
Bounding $H(W_{\mathrm{D}})$

- Substituting from Lemma 2 into Lemma 1, we have:

$$H(W_{\mathrm{D}}) \leq \log\left(\frac{(L_1(S) + L_2(S))^2}{A(S)}\right) + 4 + 2\log e$$

  Depends on perimeter to area ratio of S

- A flat S has high perimeter to area ratio $\Rightarrow$ high $H(W_{\mathrm{D}})$
- Pre-scale (X_1, X_2) to $(\tilde{X}_1, \tilde{X}_2) \sim \mathrm{Unif}(\tilde{S})$ such that $L_1(\tilde{S}) = L_2(\tilde{S})$



Bounding $H(W_{\mathrm{D}})$

- Substituting from Lemma 2 into Lemma 1, we have:

$$H(W_{\mathrm{D}}) \leq \log\left(\frac{(L_1(S) + L_2(S))^2}{A(S)}\right) + 4 + 2\log e$$

  Depends on perimeter to area ratio of S

- A flat S has high perimeter to area ratio $\Rightarrow$ high $H(W_{\mathrm{D}})$
- Pre-scale (X_1, X_2) to $(\tilde{X}_1, \tilde{X}_2) \sim \mathrm{Unif}(\tilde{S})$ such that $L_1(\tilde{S}) = L_2(\tilde{S})$
- This yields:

Proposition

$$H(\tilde{W}_{\mathrm{D}}) \leq \log\left(\frac{4L_1(S) \cdot L_2(S)}{A(S)}\right) + 4 + 2\log e$$

Bounding $H(\tilde{W}_{\mathrm{D}})$ by $I(X_1; X_2)$ for S convex

- We have

$$H(\tilde{W}_{\mathrm{D}}) \leq \log\left(\frac{L_1(S) \cdot L_2(S)}{A(S)}\right) + 6 + 2\log e$$

Bounding $H(\tilde{W}_{\mathrm{D}})$ by $I(X_1; X_2)$ for S convex

- We have

$$H(\tilde{W}_{\mathrm{D}}) \leq \log\left(\frac{L_1(S) \cdot L_2(S)}{A(S)}\right) + 6 + 2\log e$$

- Now

$$I(X_1; X_2) = h(X_1) + h(X_2) - \log A(S)$$

Bounding $H(\tilde{W}_{\mathrm{D}})$ by $I(X_1; X_2)$ for S convex

- We have

$$H(\tilde{W}_{\mathrm{D}}) \leq \log\left(\frac{L_1(S) \cdot L_2(S)}{A(S)}\right) + 6 + 2\log e$$

- Now

$$I(X_1; X_2) = h(X_1) + h(X_2) - \log A(S)$$

- Combining, we have

$$H(\tilde{W}_{\mathrm{D}}) \leq I(X_1; X_2) + (\log L_1(S) - h(X_1)) + (\log L_2(S) - h(X_2)) + 6 + 2\log e$$

Bounding $H(\tilde{W}_{\mathrm{D}})$ by $I(X_1; X_2)$ for S convex

- We have

$$H(\tilde{W}_{\mathrm{D}}) \leq \log\left(\frac{L_1(S) \cdot L_2(S)}{A(S)}\right) + 6 + 2\log e$$

- Now

$$I(X_1; X_2) = h(X_1) + h(X_2) - \log A(S)$$

- Combining, we have

$$H(\tilde{W}_{\mathrm{D}}) \leq I(X_1; X_2) + (\log L_1(S) - h(X_1)) + (\log L_2(S) - h(X_2)) + 6 + 2\log e$$

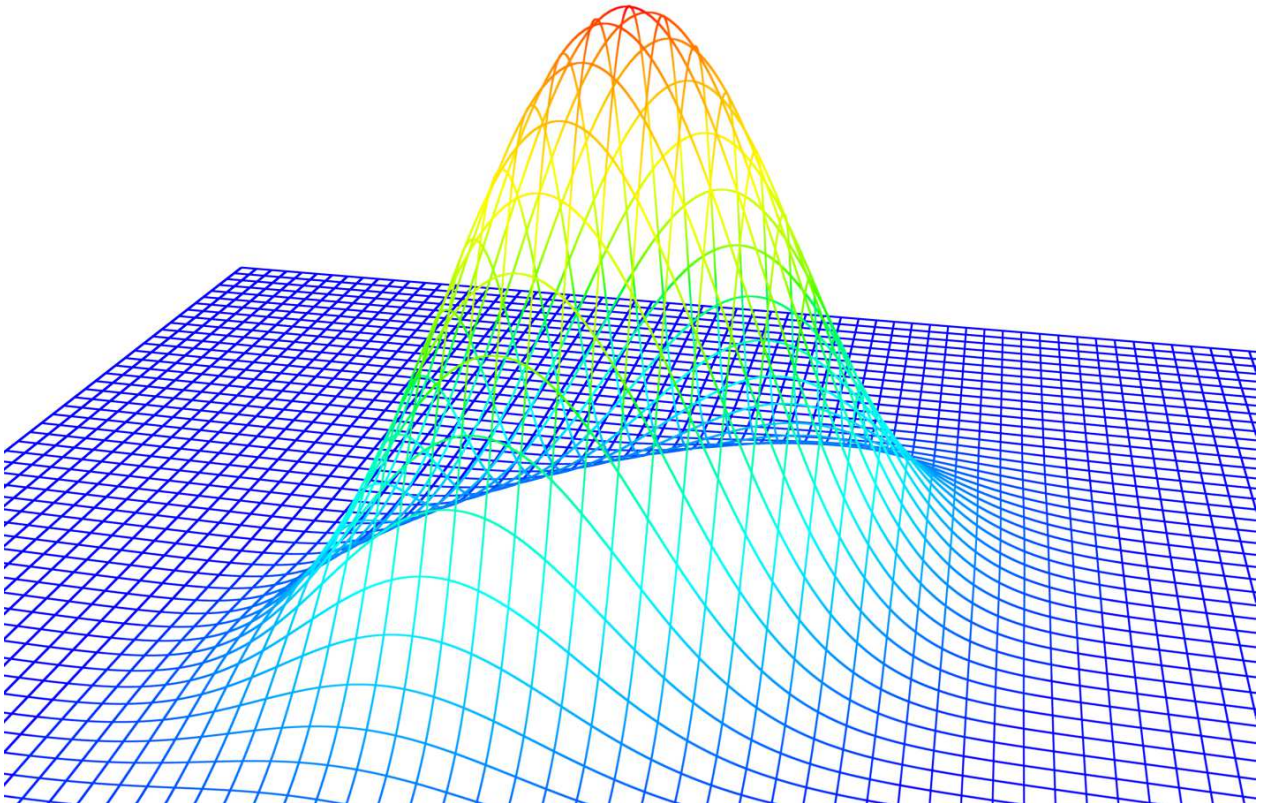
- If S is convex, marginals close to uniform

  And, we obtain constant gap between $H(\tilde{W}_{\mathrm{D}})$ and $I(X_1; X_2)$

Generation scheme for $(X_1, X_2) \sim f$

- Consider positive part of hypograph of the pdf

$$S = \{(x_1, x_2, z) : \ 0 \le z \le f(x_1, x_2)\} \subset \mathbb{R}^3$$

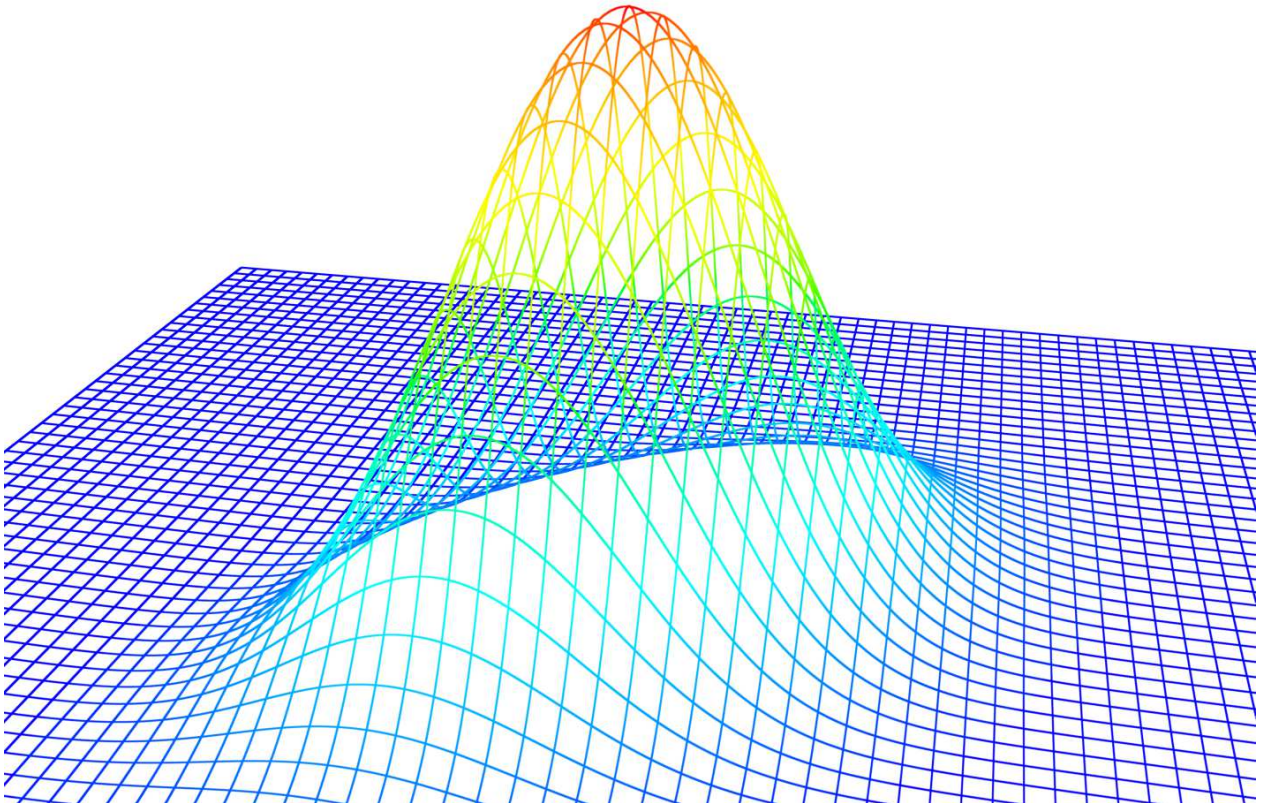


Generation scheme for $(X_1, X_2) \sim f$

- Consider positive part of hypograph of the pdf

$$S = \{(x_1, x_2, z) : \ 0 \le z \le f(x_1, x_2)\} \subset \mathbb{R}^3$$

- Key observation: If we let $(X_1, X_2, Z) \sim \mathrm{Unif}(S)$, then $(X_1, X_2) \sim f$

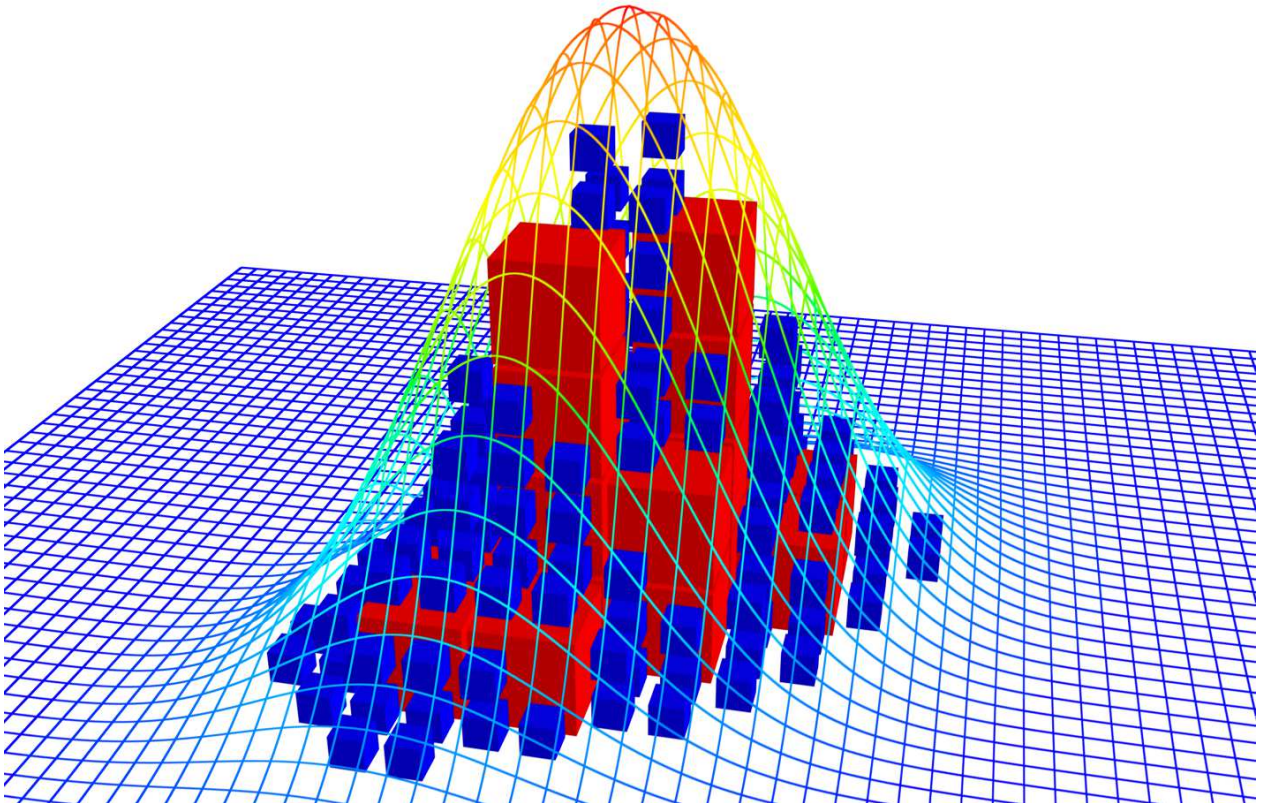


Generation scheme for $(X_1, X_2) \sim f$

- Consider positive part of hypograph of the pdf

$$S = \{(x_1, x_2, z) : \ 0 \le z \le f(x_1, x_2)\} \subset \mathbb{R}^3$$

- Key observation: If we let $(X_1, X_2, Z) \sim \mathrm{Unif}(S)$, then $(X_1, X_2) \sim f$
- Apply dyadic decomposition scheme to uniform over $S \subset \mathbb{R}^3$



Bound on G for log-concave pdf

- f log-concave $\Rightarrow$ positive part of hypograph orthogonally convex
- Using erosion entropy, scaling and additional nontrivial steps, we obtain:

Theorem

$$I(X_1; X_2) \leq J(X_1; X_2) \leq G(X_1; X_2) \leq I(X_1; X_2) + 24$$

Bound on G for log-concave pdf

- f log-concave $\Rightarrow$ positive part of hypograph orthogonally convex
- Using erosion entropy, scaling and additional nontrivial steps, we obtain:

Theorem

$$I(X_1; X_2) \leq J(X_1; X_2) \leq G(X_1; X_2) \leq I(X_1; X_2) + 24$$

- Result can be extended to n agents:

Theorem

$$I_D \leq J(X_1; \ldots; X_n) \leq G(X_1; \ldots; X_n) \leq I_D + n^2 \log e + 9n \log n$$

$I_D(X_1; \ldots; X_n) = h(X_1, \ldots, X_n) - \sum_{i=1}^{n} h(X_i | X^{i-1}, X_{i+1}^n)$ is the Dual total correlation

Summary of distributed randomness generation

- Asymptotic distributed randomness generation:
  - Wyner common information J is the limit
  - Use random coding and similar covering lemma to lossy compression

Summary of distributed randomness generation

- Asymptotic distributed randomness generation:
  - Wyner common information J is the limit
  - Use random coding and similar covering lemma to lossy compression
- One shot distributed randomness generation:
  - Common entropy is the limit
  - Common entropy within constant of mutual information for log concave pdfs

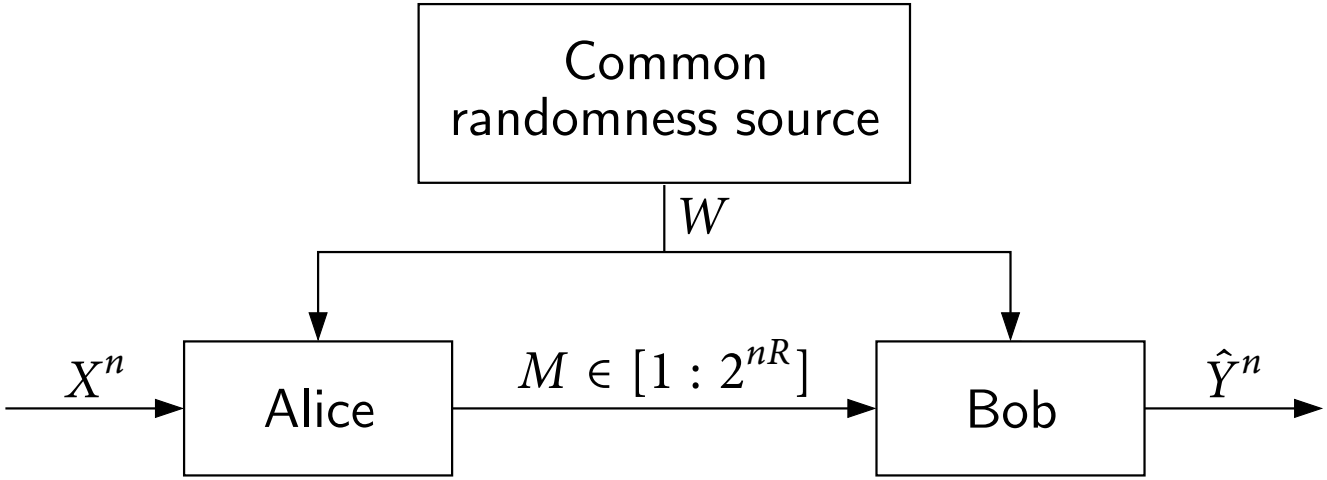
Summary of distributed randomness generation

- Asymptotic distributed randomness generation:
  - Wyner common information J is the limit
  - Use random coding and similar covering lemma to lossy compression
- One shot distributed randomness generation:
  - Common entropy is the limit
  - Common entropy within constant of mutual information for log concave pdfs
- Limit can be larger for one shot than asymptotic $(G > J)$
- Common information has several different measures (I, J, G)

Outline

- Centralized randomness generation:
  - One shot
  - Asymptotic
- Distributed randomness generation:
  - Asymptotic
  - One shot – recent results
- Channel synthesis (common randomness/communication tradeoff):
  - Asymptotic
  - One shot

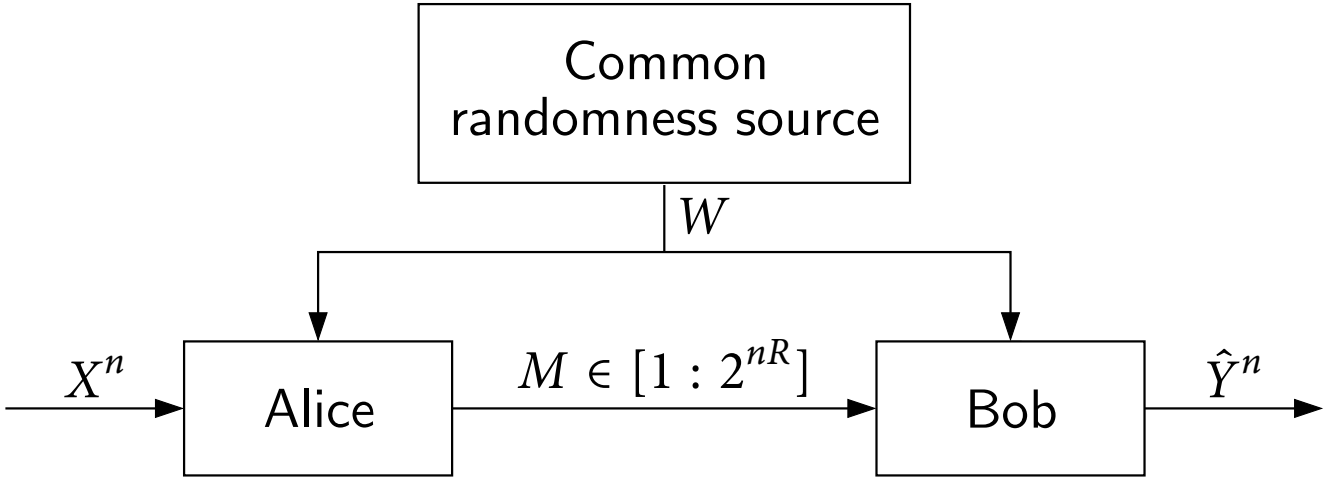
Asymptotic channel simulation with common randomness



- $(X, Y) \sim p(x, y)$; $\hat{\mathcal{Y}} = \mathcal{Y}$; $X^n \sim \prod_{i=1}^n p_X(x_i)$
- W: common randomness
- Bob wishes to generate $\hat{Y}^n$ such that:

  $d_n = \sum_{(x^n, y^n)} \left| \prod_{i=1}^n p_X(x_i) p_{\hat{Y}^n|X^n}(y^n|x^n) - \prod_{i=1}^n p_{X,Y}(x_i, y_i) \right| \to 0$
- What is the minimum communication rate R?

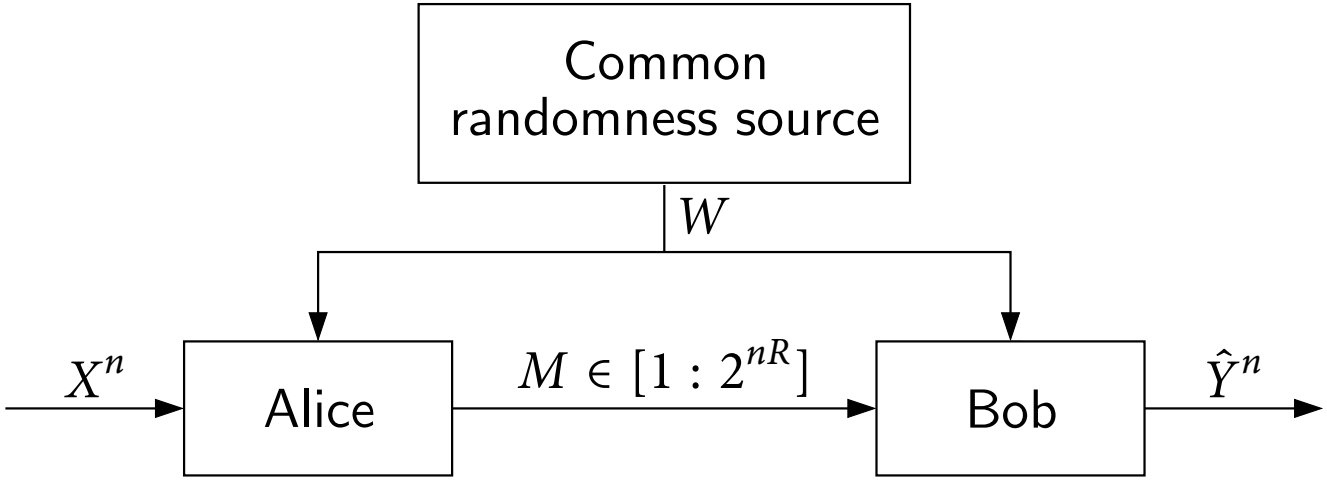
Asymptotic channel simulation with common randomness



- $(X, Y) \sim p(x, y)$; $\hat{\mathcal{Y}} = \mathcal{Y}$; $X^n \sim \prod_{i=1}^n p_X(x_i)$
- W: common randomness
- Bob wishes to generate $\hat{Y}^n$ such that:

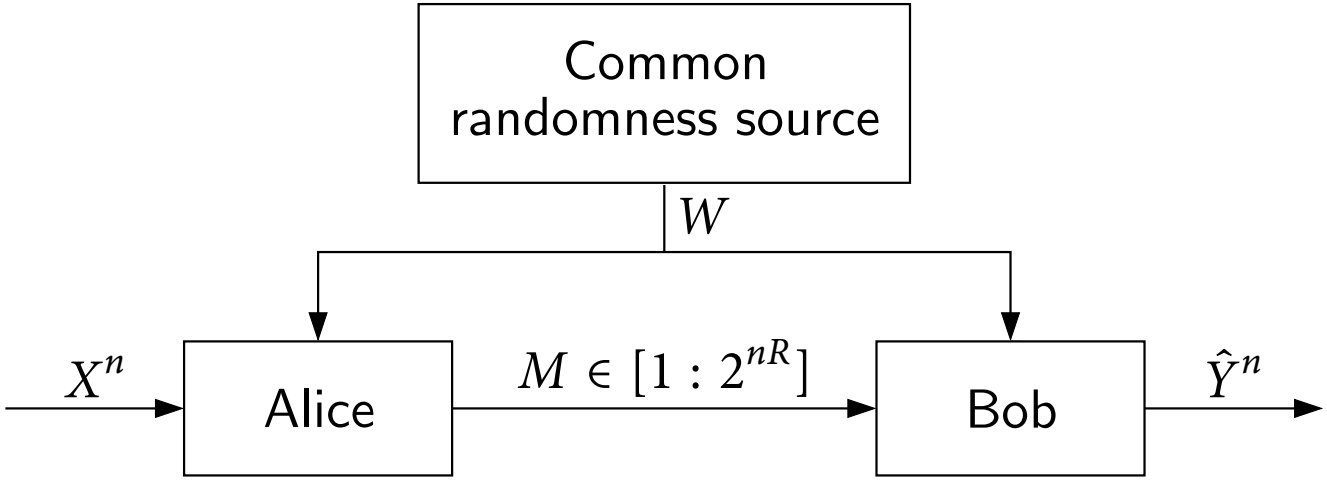
  $d_n = \sum_{(x^n, y^n)} \left| \prod_{i=1}^n p_X(x_i) p_{\hat{Y}^n|X^n}(y^n|x^n) - \prod_{i=1}^n p_{X,Y}(x_i, y_i) \right| \to 0$
- What is the minimum communication rate R?
- Bennett–Shor–Smolin–Thapliyal (2002) showed that: $\min R = I(X; Y)$

Asymptotic channel simulation with common randomness



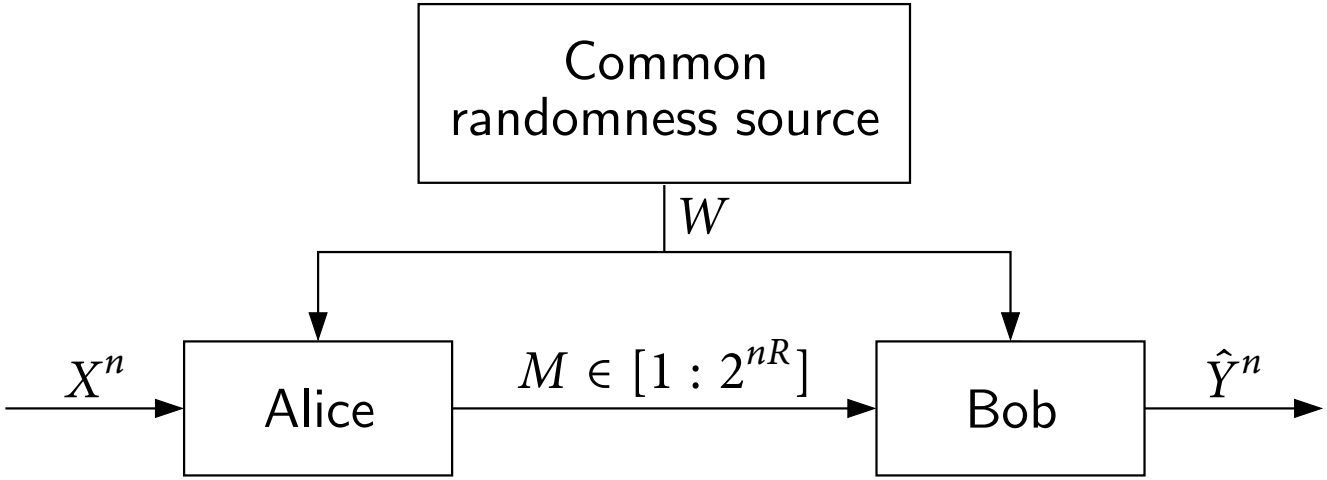
- For X^n arbitrary: $\min R = \max_{p(x)} I(X;Y) = C$, capacity of channel $p(y|x)$

Asymptotic channel simulation with common randomness



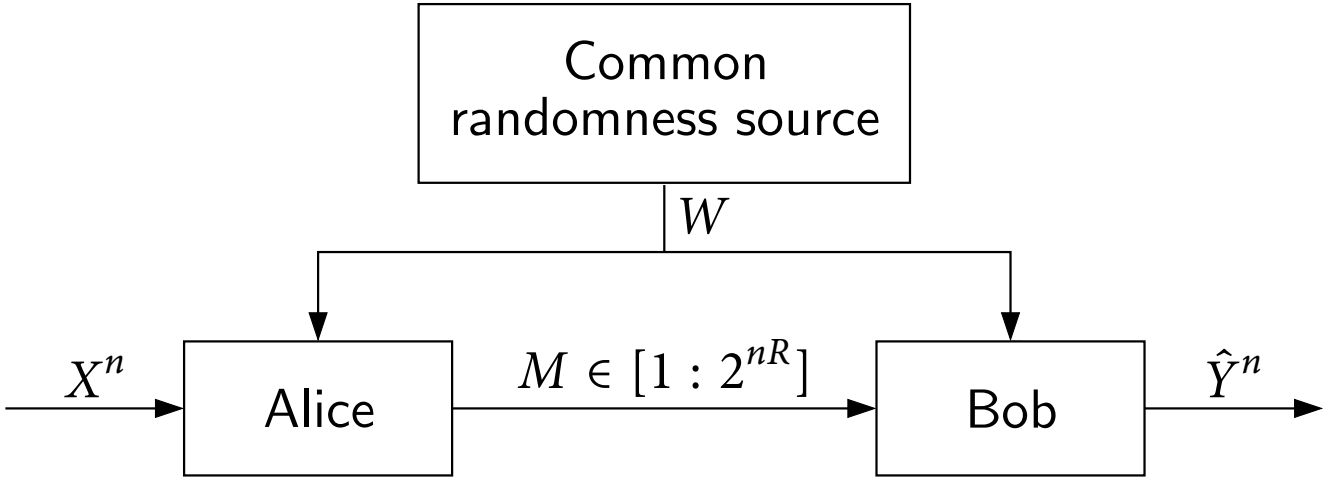
- For X^n arbitrary: $\min R = \max_{p(x)} I(X;Y) = C$, capacity of channel $p(y|x)$
- Reverse Shannon theorem: Noiseless channel with capacity C and unlimited common randomness can simulate any noisy channel with capacity C

Asymptotic channel simulation with common randomness



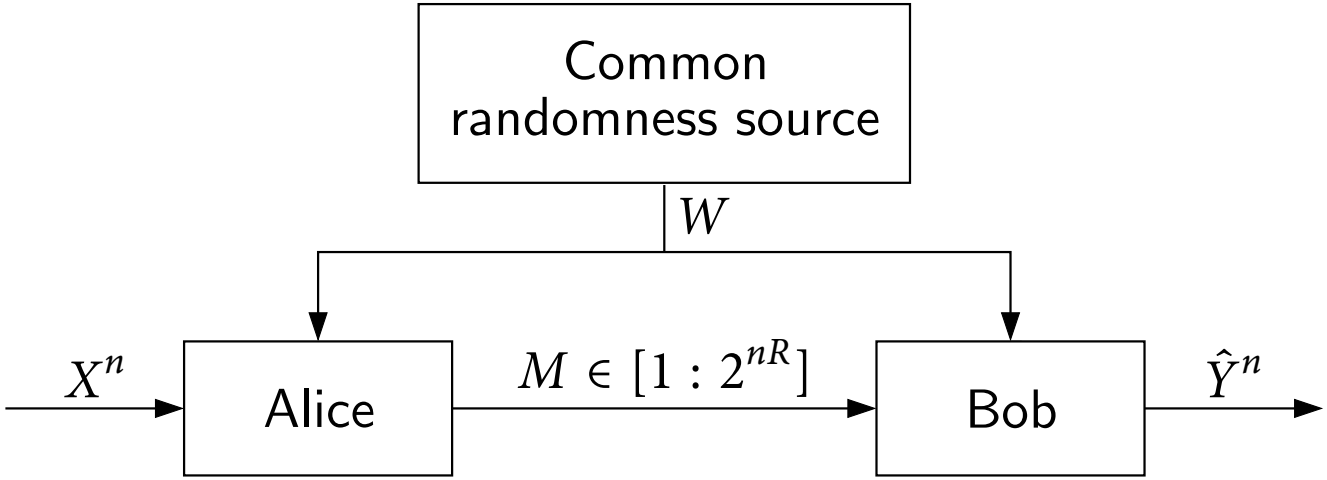
- For X^n arbitrary: $\min R = \max_{p(x)} I(X;Y) = C$, capacity of channel $p(y|x)$
- Reverse Shannon theorem: Noiseless channel with capacity C and unlimited common randomness can simulate any noisy channel with capacity C
- Shannon theorem: Noiseless channel with capacity C can be simulated by any noisy channel with capacity C

Asymptotic channel simulation with common randomness



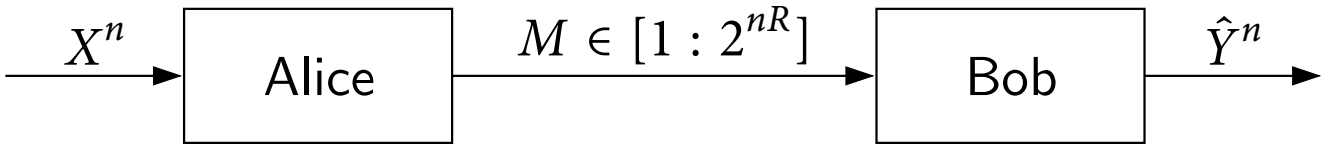
- For X^n arbitrary: $\min R = \max_{p(x)} I(X;Y) = C$, capacity of channel $p(y|x)$
- Reverse Shannon theorem: Noiseless channel with capacity C and unlimited common randomness can simulate any noisy channel with capacity C
- Shannon theorem: Noiseless channel with capacity C can be simulated by any noisy channel with capacity C
- Hence any channel p can simulate a channel q iff $C(p) \geq C(q)$

Asymptotic channel simulation with common randomness



- For X^n arbitrary: $\min R = \max_{p(x)} I(X;Y) = C$, capacity of channel $p(y|x)$
- Reverse Shannon theorem: Noiseless channel with capacity C and unlimited common randomness can simulate any noisy channel with capacity C
- Shannon theorem: Noiseless channel with capacity C can be simulated by any noisy channel with capacity C
- Hence any channel p can simulate a channel q iff $C(p) \geq C(q)$
- Does the same hold for entanglement-assisted quantum channels?

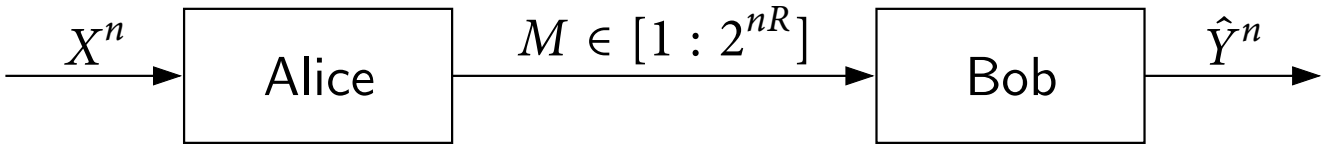
Channel simulation without common randomness



- $(X, Y) \sim p(x, y)$; $X^n \sim \prod_{i=1}^n p_X(x_i)$
- Bob wishes to generate $\hat{Y}^n$ such that:

  $d_n = \sum_{(x^n, y^n)} \left| \prod_{i=1}^n p_X(x_i) p_{\hat{Y}^n|X^n}(y^n|x^n) - \prod_{i=1}^n p_{X,Y}(x_i, y_i) \right| \to 0$
- What is the minimum communication rate R?

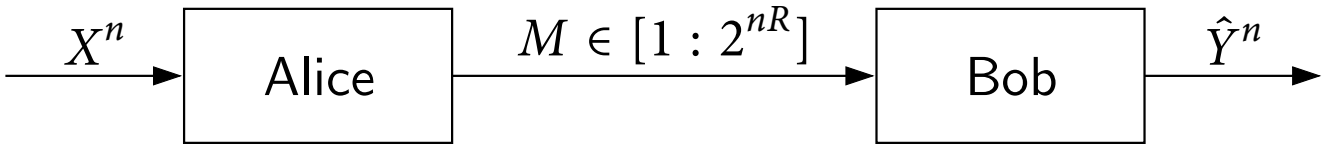
Channel simulation without common randomness



- $(X, Y) \sim p(x, y)$; $X^n \sim \prod_{i=1}^n p_X(x_i)$
- Bob wishes to generate $\hat{Y}^n$ such that:

  $d_n = \sum_{(x^n, y^n)} \left| \prod_{i=1}^n p_X(x_i) p_{\hat{Y}^n|X^n}(y^n|x^n) - \prod_{i=1}^n p_{X,Y}(x_i, y_i) \right| \to 0$
- What is the minimum communication rate R?
- Cuff (2013) showed that: $\min R = J(X; Y)$

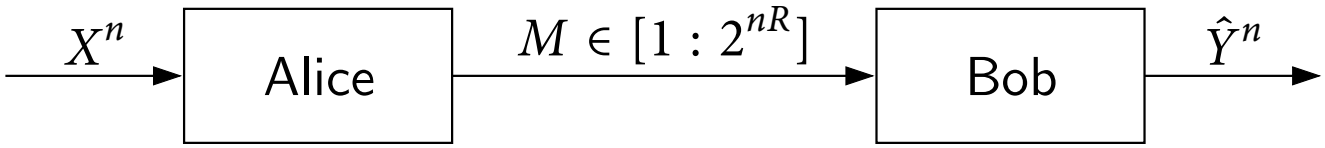
Channel simulation without common randomness



- $(X, Y) \sim p(x, y)$; $X^n \sim \prod_{i=1}^n p_X(x_i)$
- Bob wishes to generate $\hat{Y}^n$ such that:

  $d_n = \sum_{(x^n, y^n)} \left| \prod_{i=1}^n p_X(x_i) p_{\hat{Y}^n|X^n}(y^n|x^n) - \prod_{i=1}^n p_{X,Y}(x_i, y_i) \right| \to 0$
- What is the minimum communication rate R?
- Cuff (2013) showed that: $\min R = J(X; Y)$
- Operationally shows that: $J(X; Y) \geq I(X, Y)$

Channel simulation without common randomness



- $(X, Y) \sim p(x, y)$; $X^n \sim \prod_{i=1}^n p_X(x_i)$
- Bob wishes to generate $\hat{Y}^n$ such that:

  $d_n = \sum_{(x^n, y^n)} \left| \prod_{i=1}^n p_X(x_i) p_{\hat{Y}^n|X^n}(y^n|x^n) - \prod_{i=1}^n p_{X,Y}(x_i, y_i) \right| \to 0$
- What is the minimum communication rate R?
- Cuff (2013) showed that: $\min R = J(X; Y)$
- Operationally shows that: $J(X; Y) \geq I(X, Y)$
- For $X = Y$, problem reduces to lossless compression with limit $H(X)$

Tradeoff between communication and common randomness

- Suppose $W \in [1 : 2^{nR_C}]$ (still have $M \in [1 : 2^{nR}]$)

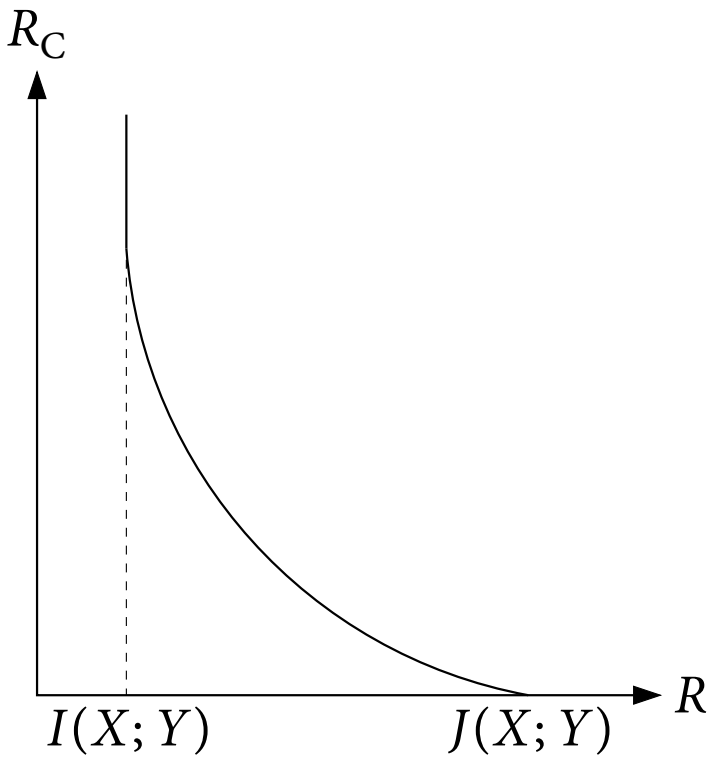
Tradeoff between communication and common randomness

- Suppose $W \in [1 : 2^{nR_C}]$ (still have $M \in [1 : 2^{nR}]$)
- Cuff (2013), Bennett–Devetak–Harrow–Shor–Winter (2014) established optimal tradeoff between R and common randomness rate R_C

Set of (R, R_C) such that:

$$R \geq I(X;U),$$
$$R_C + R \geq I(X,Y;U)$$

for some $X \to U \to Y$



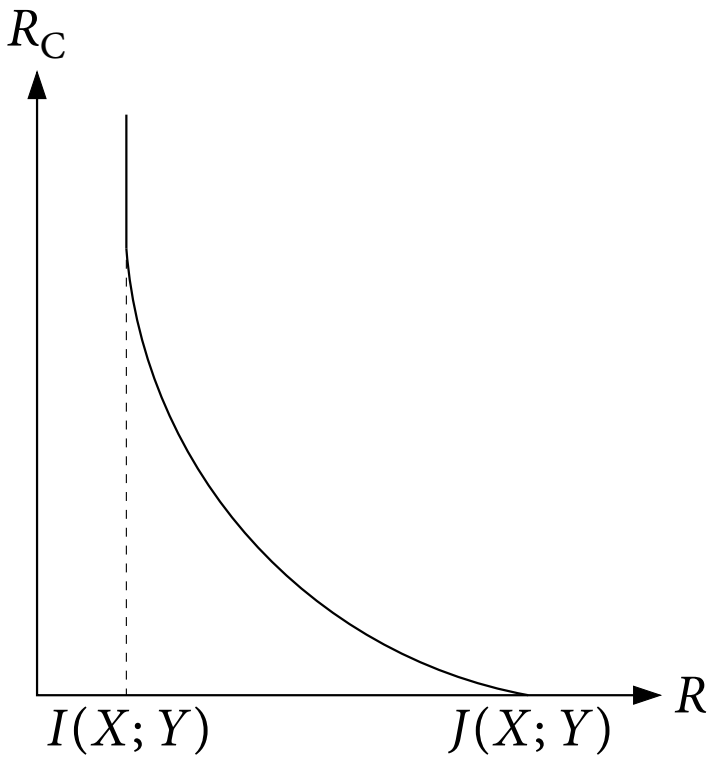
Tradeoff between communication and common randomness

- Suppose $W \in [1 : 2^{nR_C}]$ (still have $M \in [1 : 2^{nR}]$)
- Cuff (2013), Bennett–Devetak–Harrow–Shor–Winter (2014) established optimal tradeoff between R and common randomness rate R_C

Set of (R, R_C) such that:

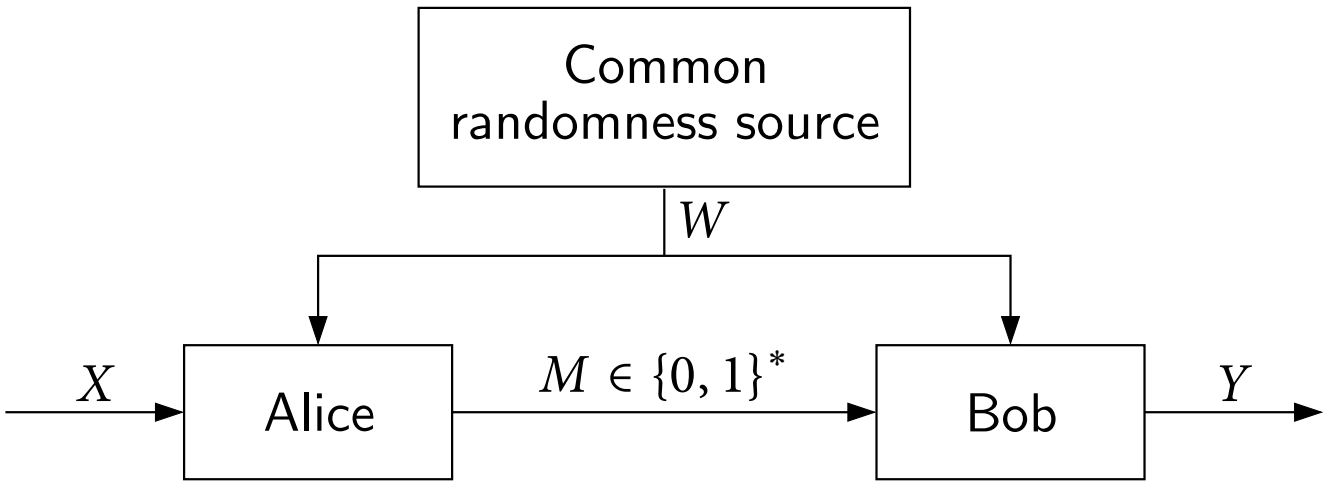
$$R \geq I(X;U),$$
$$R_C + R \geq I(X,Y;U)$$

for some $X \to U \to Y$



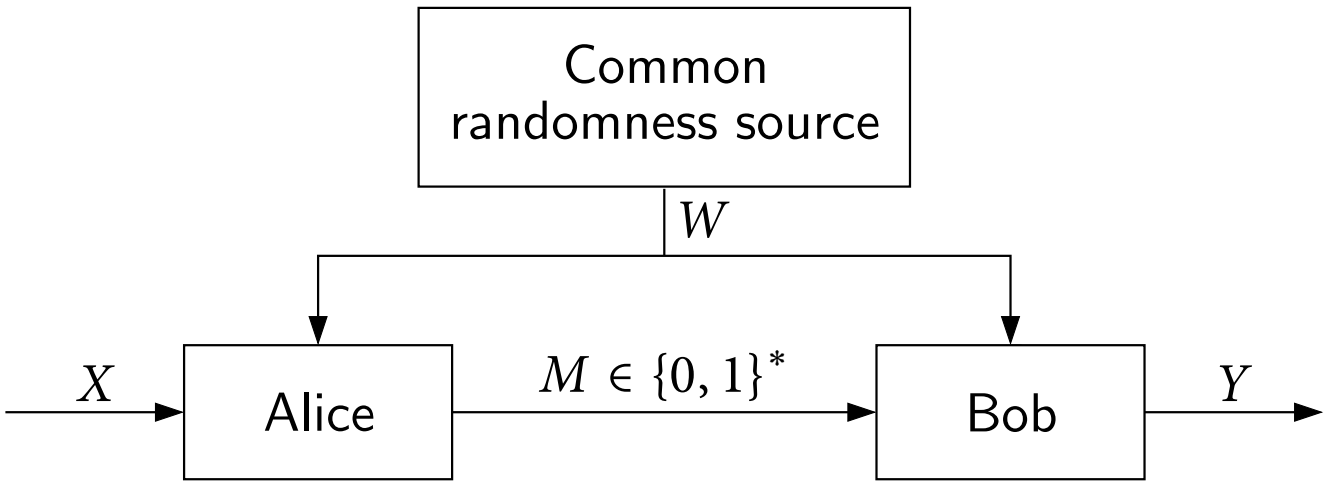
- Application to information theoretic secrecy (Cuff 2013)

One shot channel simulation with common randomness



- $(X, Y) \sim p(x, y)$; L length of prefix code for M
- Bob wishes to generate $Y \sim p_{Y|X}(y|x)$
- What is the minimum $\mathsf{E}(L)$?

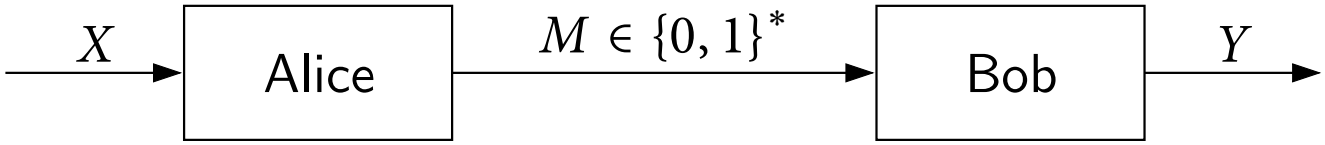
One shot channel simulation with common randomness



- $(X, Y) \sim p(x, y)$; L length of prefix code for M
- Bob wishes to generate $Y \sim p_{Y|X}(y|x)$
- What is the minimum $\mathsf{E}(L)$?
- Harsha–Jain–McAllester–Radhakrishnan (2010) showed that:

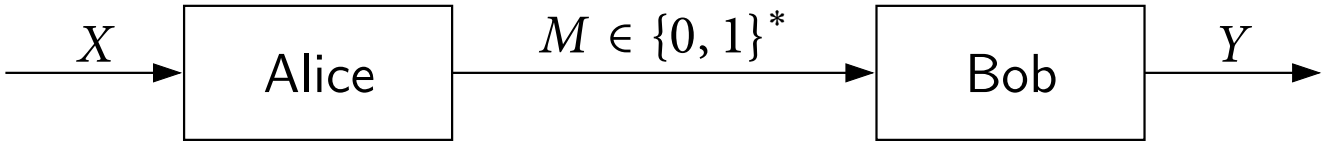
$$I(X;Y) \leq \min \mathsf{E}(L) \leq I(X;Y) + 2\log(I(X;Y)+1) + O(1)$$

One shot channel simulation without common randomness

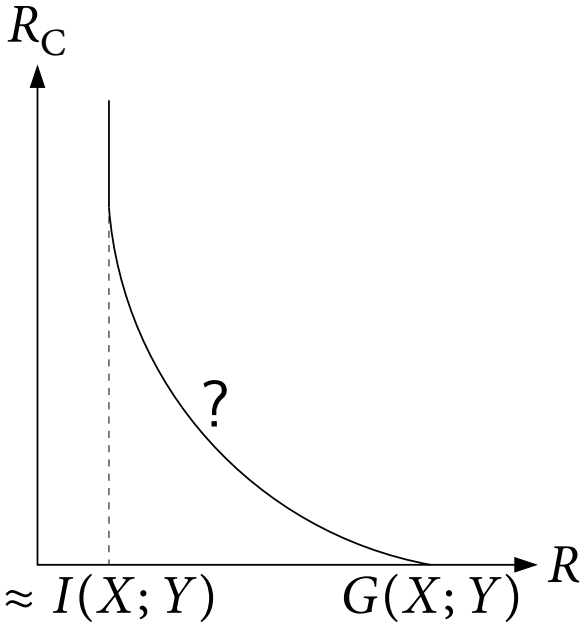


- By Kumar–Li–EG (2014): $G(X;Y) \leq \min \mathsf{E}(L) < G(X;Y) + 1$
- For $X = Y$, problem reduces to zero error compression with limit $H(X)$

One shot channel simulation without common randomness



- By Kumar–Li–EG (2014): $G(X;Y) \leq \min \mathsf{E}(L) < G(X;Y) + 1$
- For $X = Y$, problem reduces to zero error compression with limit $H(X)$
- Optimal tradeoff between R and R_{C} is not known



Conclusion

- Limits of randomness generation and communication closely related

Conclusion

- Limits of randomness generation and communication closely related
- Entropy is the limit on compression and randomness generation

Conclusion

- Limits of randomness generation and communication closely related
- Entropy is the limit on compression and randomness generation
- Mutual information is the limit on:
  - Channel capacity and lossy compression
  - Channel simulation with common randomness

Conclusion

- Limits of randomness generation and communication closely related
- Entropy is the limit on compression and randomness generation
- Mutual information is the limit on:
  - Channel capacity and lossy compression
  - Channel simulation with common randomness
- Wyner common information is the limit on:
  - Asymptotic distributed randomness generation
  - Asymptotic channel simulation without common randomness

Conclusion

- Limits of randomness generation and communication closely related
- Entropy is the limit on compression and randomness generation
- Mutual information is the limit on:
  - Channel capacity and lossy compression
  - Channel simulation with common randomness
- Wyner common information is the limit on:
  - Asymptotic distributed randomness generation
  - Asymptotic channel simulation without common randomness
- Common entropy is the limit on:
  - One shot distributed randomness generation
  - One shot channel simulation without common randomness

Conclusion

- Limits of randomness generation and communication closely related
- Entropy is the limit on compression and randomness generation
- Mutual information is the limit on:
  - Channel capacity and lossy compression
  - Channel simulation with common randomness
- Wyner common information is the limit on:
  - Asymptotic distributed randomness generation
  - Asymptotic channel simulation without common randomness
- Common entropy is the limit on:
  - One shot distributed randomness generation
  - One shot channel simulation without common randomness
- Natural application of information theoretic ideas and techniques

Thank you!

References

Bennett, C. H., Devetak, I., Harrow, A. W., Shor, P. W., and Winter, A. (2014). The quantum reverse shannon theorem and resource tradeoffs for simulating quantum channels. *IEEE Trans. Info. Theory*, 60(5), 2926–2959.

Bennett, C. H., Shor, P. W., Smolin, J., and Thapliyal, A. V. (2002). Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Trans. Info. Theory*, 48(10), 2637–2655.

Cover, T. M. (1972). Broadcast channels. *IEEE Trans. Inf. Theory*, 18(1), 2–14.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. Wiley, New York.

Cuff, P. (2013). Distributed channel synthesis. *IEEE Trans. Info. Theory*, 59(11), 7071–7096.

El Gamal, A. and Kim, Y.-H. (2011). *Network Information Theory*. Cambridge, Cambridge.

Han, T. S. and Verdu, S. (1993). Approximation theory of output statistics. *IEEE Trans. Info. Theory*, 39(3), 752–772.

Harsha, P., Jain, R., McAllester, D., and Radhakrishnan, J. (2010). The communication complexity of correlation. *IEEE Trans. Info. Theory*, 56(1), 438–449.

References (cont.)

Knuth, D. E. and Yao, A. C. (1976). The complexity of random number generation. In *Algorithms and Complexity (Proc. Symp., Carnegie Mellon Univ., Pittsburgh, Pa., 1976)*, pp. 357–428. Academic Press, New York.

Kumar, G., Li, C. T., and EG, A. (2014). Exact common information. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pp. 161–165.

Li, C. T. and El Gamal(2016). Distributed simulation of continuous random variables. In *Proc. IEEE Symp. Info. Theory*, pp. 565–569.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3), 379–423, 27(4), 623–656.

Shannon, C. E. (1956). The zero error capacity of a noisy channel. *IRE Trans. Inf. Theory*, 2(3), 8–19.

Shannon, C. E. (1961). Two-way communication channels. In *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, vol. I, pp. 611–644. University of California Press, Berkeley.

Slepian, D. and Wolf, J. K. (1973). Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, 19(4), 471–480.

Wyner, A. D. (1975). The common information of two dependent random variables. *IEEE Trans. Inf. Theory*, 21(2), 163–179.