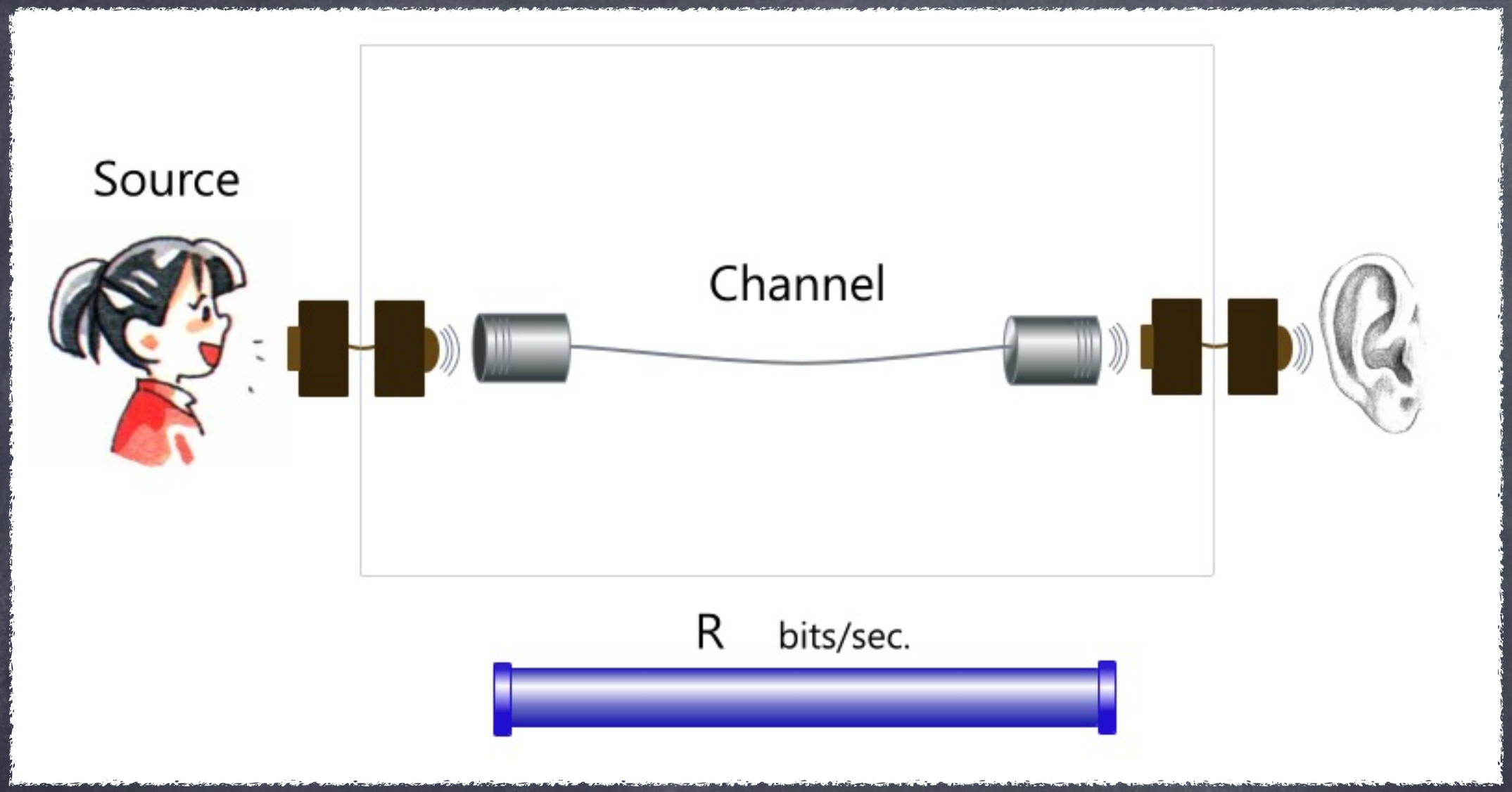
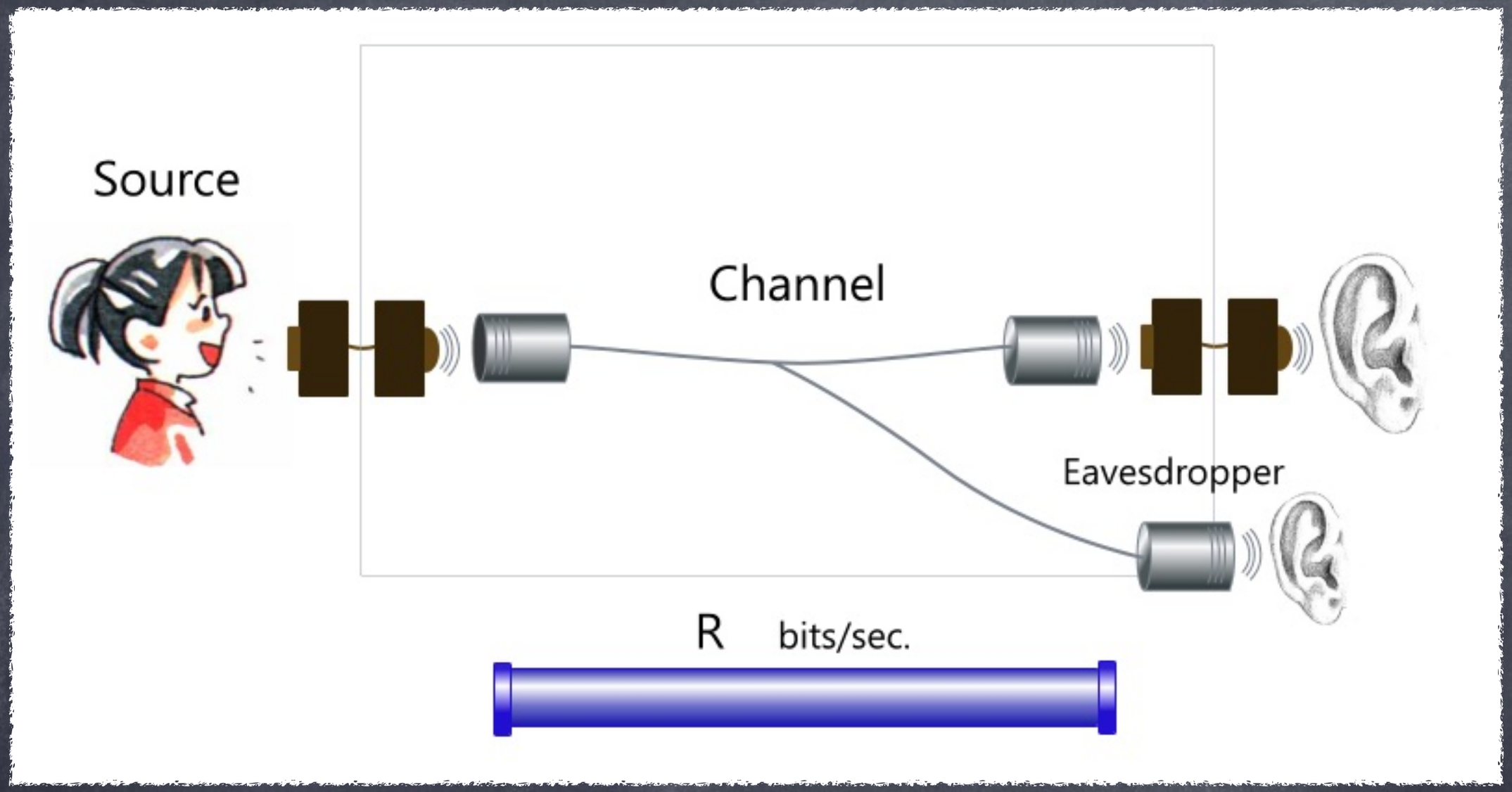
Rate-Distortion Theory for Secrecy Systems

SPCOM 2014



Communication

Move a signal from one place to another

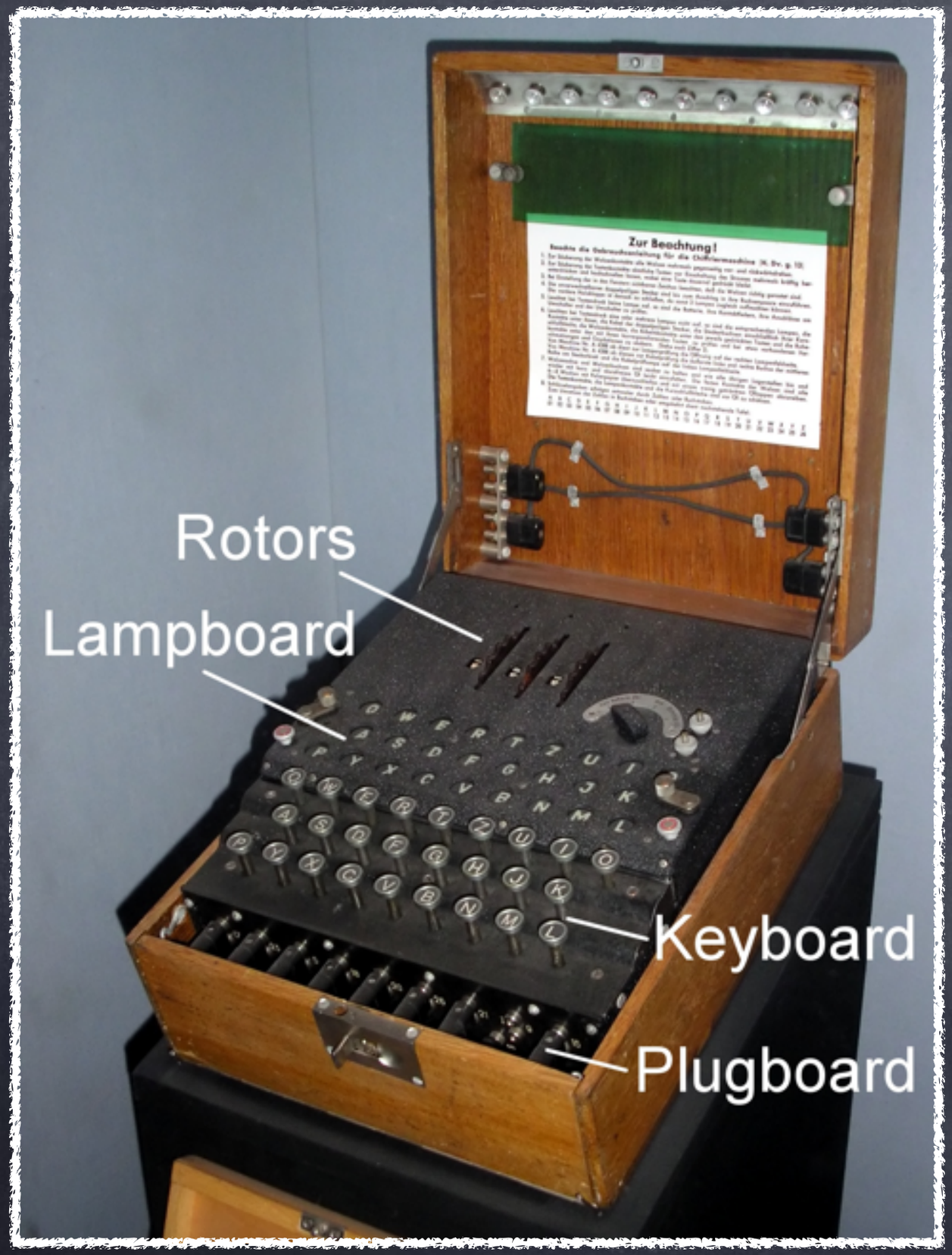


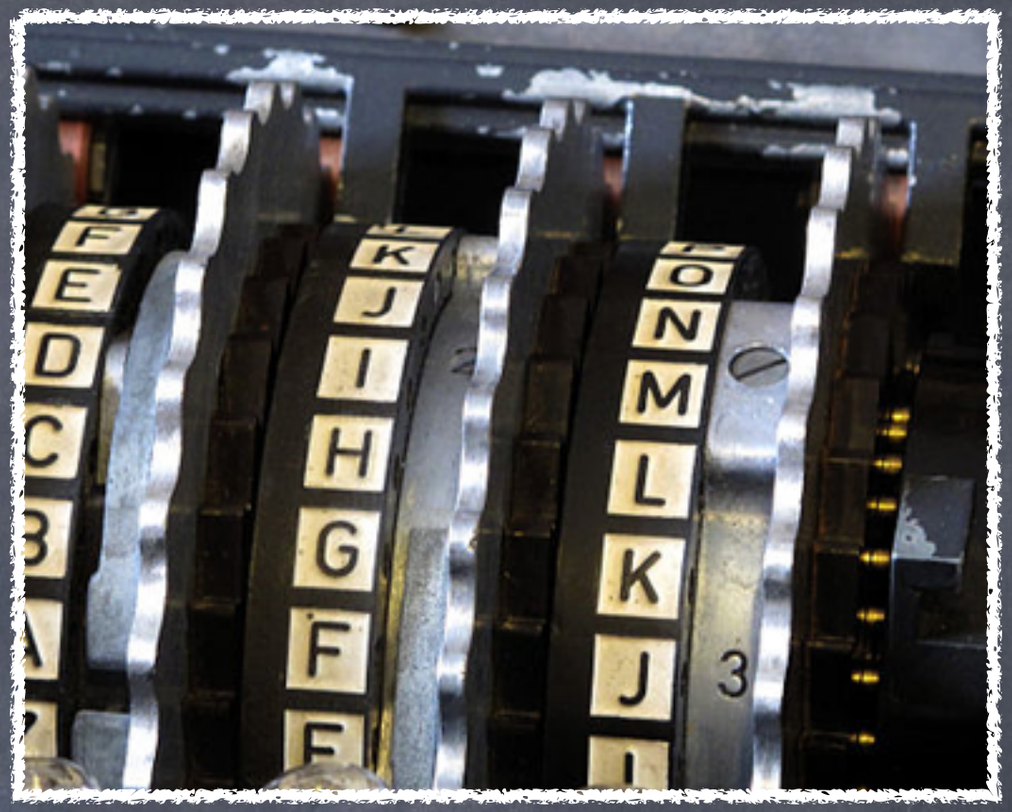
Communication

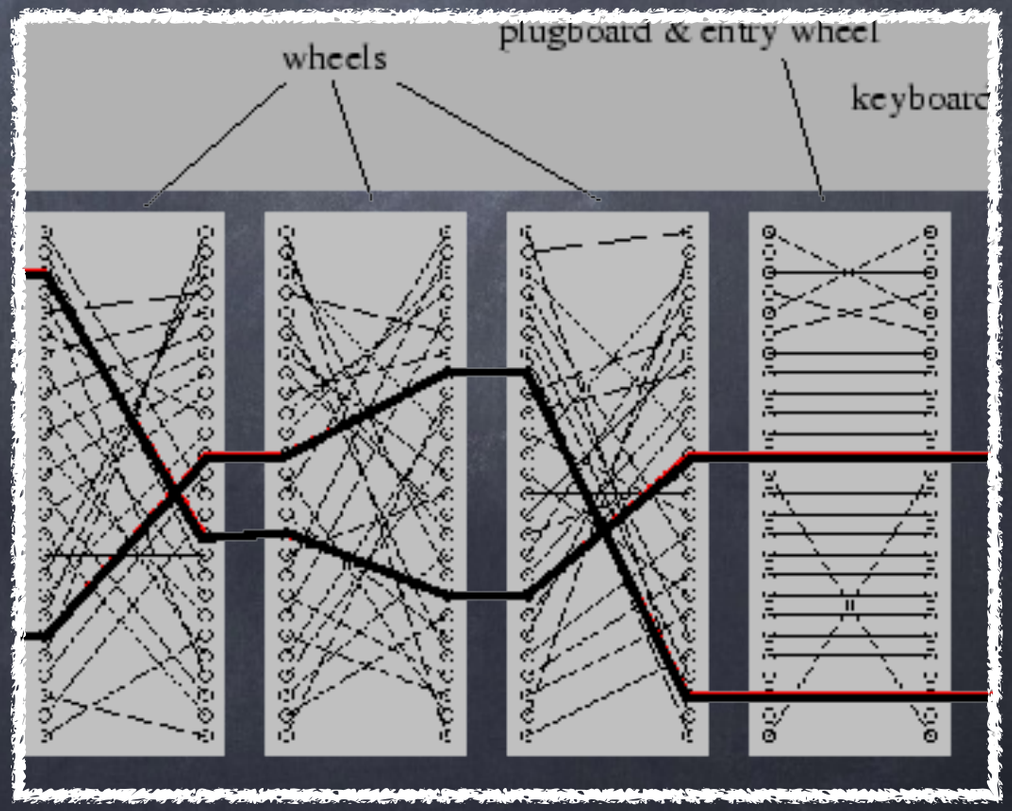
Move a signal from one place to another

Secrecy Systems

- Old Ciphers:
  - Obscure (design kept secret)
  - Complicated
- New Ciphers
  - Computationally challenging
  - Complicated

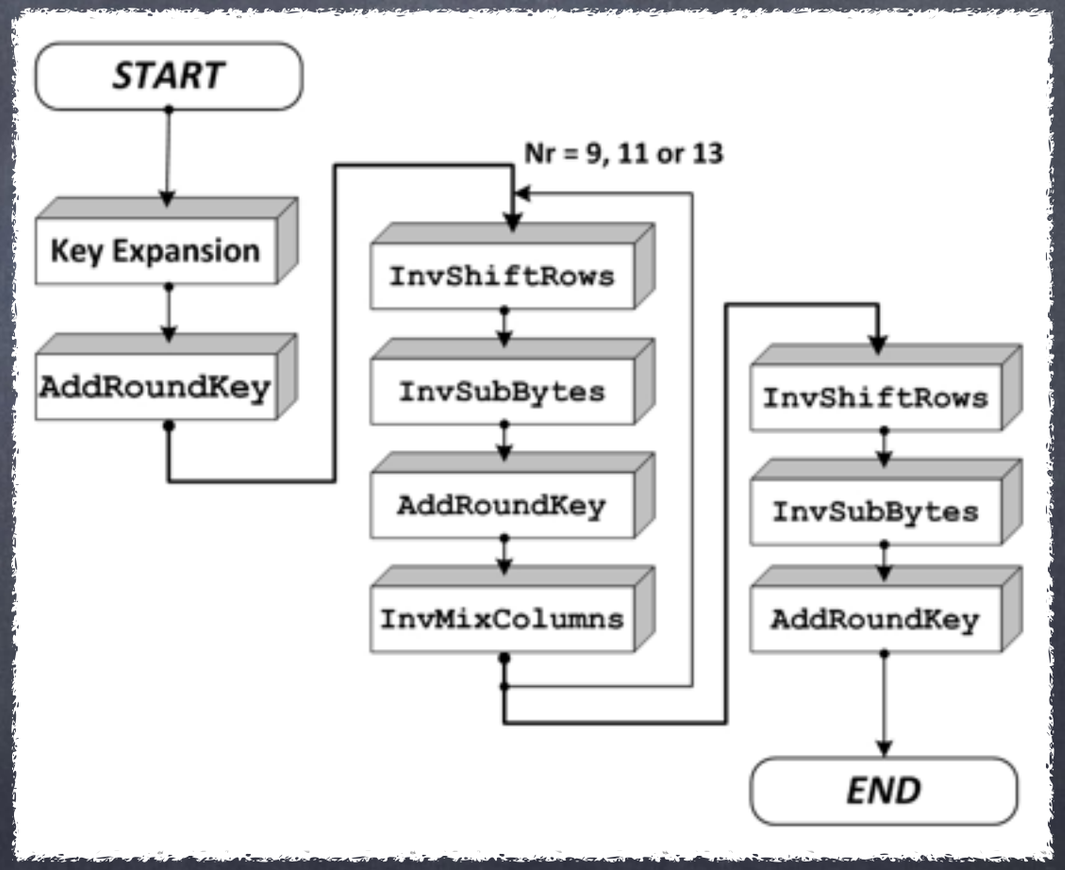

Zur Beachtung!
Rotors
Lampboard
Keyboard
Plugboard




wheels
plugboard & entry wheel
keyboard

AES

- Complicated
- Not known how to undo



Modern Cryptography

- Began in 1970's (from IT community)
- Built on fundamental mathematical problems that are believed difficult to compute
- Diffie-Hellman
- RSA

Information Theoretic Security

- Shannon's '49 paper
  - Fewer assumptions (omnipotent adversary)
  - More resources required for mathematical guarantees

Perfect Secrecy

- Independence between the transmitted message and the information
- Perfect Secrecy achieved with one-time pad (Vernam cipher).
- Shannon: This method is also necessary. (i.e. $R_K > H(X)$)

One-time Pad

Information: 0110101000010110

Key: 1101001101001 01

$\oplus$

Message: 1011100101100 11

Proof of Perfect Secrecy

Consider 1 bit: $M = X + K \bmod 2$

$P(M=0|X=0) = 1/2$

$P(M=0|X=1) = 1/2$

Therefore, $X \perp M$

Many versions of one-time pad

- Each key specifies a one-to-one mapping.
- With a little excess message rate and excess key rate, a random mapping for each key will achieve near perfect secrecy.

Necessity of one-time-pad

For each key there is a one-to-one mapping from X to M:

$$\text{support}(M) \geq \text{support}(X)$$

Perfect secrecy:

$$P(M|X) = P(M)$$

For each X, all randomness from key:

$$\text{support}(K) \geq \text{support}(M)$$

Overview

Channel Capacity

Source Coding

Secure Channel Coding

Secure Source Coding

IID Model

- Let the information signal be i.i.d. with a known distribution P_X
- Represent as: $X_1, X_2, \ldots, X_n = X^n$
- Codec functions over the block

Compression

Encoder: $f: \quad \mathcal{X}^n \to [2^{nR}]$

Decoder: $g: \quad [2^{nR}] \to \mathcal{Y}^n$

Compress and Conceal

- Compress losslessly to $R=H(X)$
- One-time-pad the result: $R_K = R$

Usual Relaxations

- Lossless compression:
  - Allow negligible error probability
- Perfect secrecy:
  - Allow negligible leakage

Lossless Compression

Lossless Compression

- R is achievable if for any $\varepsilon>0$, there exists an n, f, and g such that $P(X^n \neq Y^n)<\varepsilon$
- Minimum achievable rate is $H(X)$

Entropy

$$H(X) = \mathbf{E} \log \frac{1}{P_X(X)}$$

$H(X) \geq 0$ (deterministic)

$H(X) \leq \log|X|$ (uniform)

$H(X,Y,Z) = H(X) + H(Y|X) + H(Z|X,Y)$

Binary Entropy

- X is Bern(p)
  - $P(X=1) = p, \quad P(X=0) = 1-p$
- $H(X) = p \log(1/p) + (1-p) \log(1/(1-p))$
- Call this the binary entropy function
  - $h(p)$

Conditional Entropy

$$H(Y|X) = \mathbf{E}_X H_{P_{Y|X}}(Y)$$

$$= \mathbf{E} \log \frac{1}{P_{Y|X}(Y|X)}$$

Mutual Information

$$
\begin{aligned}
I(X;Y) &= H(X) + H(Y) - H(X,Y) \\
&= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X)
\end{aligned}
$$

$I(X;Y) \geq 0$ (independent)

$I(X;Y) \leq H(X)$ (function)

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

Data-processing Inequality

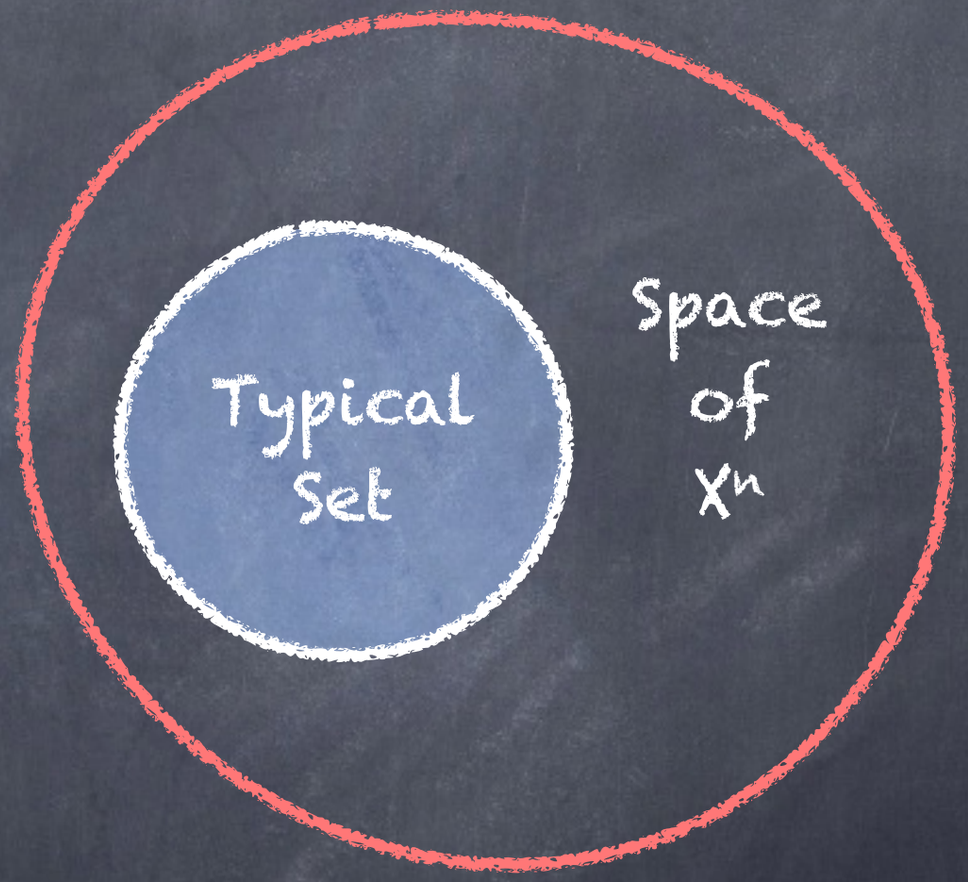
- Markov chain: $X - Y - Z$
- $I(X;Z) \leq I(X;Y)$

Typical Set and A.E.P.

- $A = \{x^n : P(x^n) \approx 2^{-nH(X)}\}$
  - exponent within ε of $H(X)$
- $P(A) \to 1$ (law of large numbers)
  - $1/n \sum_i \log(P(X_i)) \to E \log P(X_i) = -H(X)$
- $|A| \approx 2^{nH(X)}$

Lossless Compression

- Enumerate the typical set:



- Error if X^n not typical (negligible probability)
- $R \approx H(X)$

Converse

- $I(X^n;Y^n) \leq I(X^n;M) \leq H(M) \leq nR$
- $I(X^n;Y^n) = nH(X) - H(X^n|Y^n)$
- Fano's inequality:
  - $H(X^n|Y^n) \leq n\, f(P(X^n \neq Y^n))$
  - $f(x) \to 0$ as $x \to 0$, (depends on alphabet size)

Converse for Perfect Secrecy (Block coding)

Information Theoretic Argument

- $H(K) \geq I(K;X^n|M) = H(X^n|M) = H(M)$

Perfect Secrecy Metrics

Relative Entropy

- Distance between two distributions

$$D(P||Q) = \mathbf{E}_P \log \frac{P(X)}{Q(X)}$$

- $D(P||Q) \geq 0$ $(P=Q)$
- Not symmetric and no triangle inequality
- Also called
  - "divergence"
  - "K-L divergence"

M.I. as Divergence

- Mutual Information:
  - $I(X;Y) = D(P_{X,Y} \| P_X P_Y)$

Total Variation

- Another distance between distributions

$$D_{TV}(P||Q) = \sup_{A \in \mathcal{F}} P(A) - Q(A)$$

$$= \sum_x |p(x) - q(x)|$$

- $D_{TV}(P||Q) \geq 0$ (P=Q)
- True distance metric:
  - Symmetric; triangle inequality; etc.
- Directly related to hypothesis testing (used in crypto)

Asymptotic Equivalence

- Pinsker's inequality:
  - $D_{TV}(P||Q) \leq \sqrt{(1/2\ D_{KL}(P||Q))}$
- Converse: If $P \ll Q$ and Q is i.i.d. and discrete
  - If $D_{TV}(P||Q)$ decays exponentially
  - Then $D_{KL}(P||Q)$ decays with the same exponent

Perfect Secrecy Relaxations

- "Strong secrecy":
  - $I(X^n;Y^n)$ negligible
- "Weak secrecy":
  - $1/n\ I(X^n;Y^n)$ negligible
- Total variation:
  - $D_{TV}(P_{X^n,Y^n}||P_{X^n}P_{Y^n})$ negligible

Divergence

Many methods for lossless compression and perfect secrecy

- Random codebook generation
- Random binning
- Linear encoding

Lossy Compression

Rate-Distortion Theory

Average Distortion

- We need a relevant metric for lossy transmission
- Bounded distortion function: $d(x,y)$
- Average distortion:
  - $d(x^n,y^n) = 1/n \sum_i d(x_i,y_i)$

Need to encode more carefully

- Random binning and linear encoding perform poorly

Puzzle

- Given an n-bit random sequence
- 1-bit distortion:
  - How many bits of description are needed to have no more than one bit of distortion? (n - ?)
- 1-bit transmission:
  - How much distortion can be achieved with only a one-bit description? (n/2 - ?)

Definition of achievability

- Rate R is achievable for distortion D if there exist n, f, and g operating at rate R such that

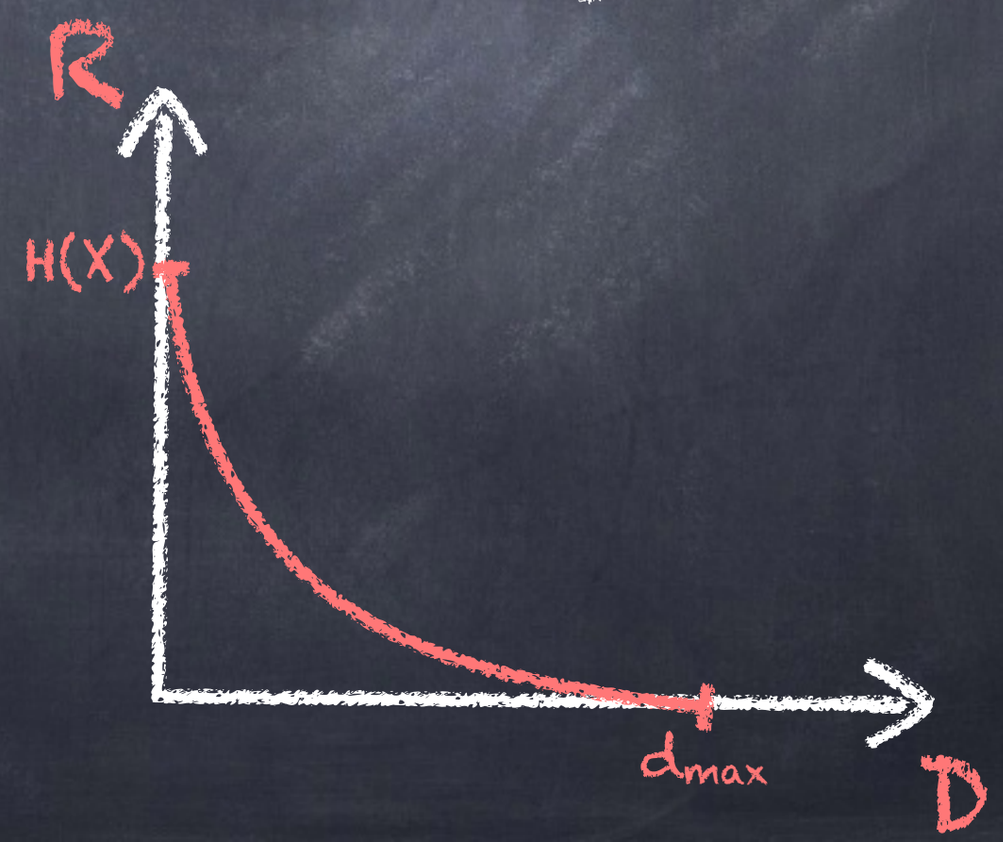
$$E\, d(X^n, Y^n) \leq D$$

Rate-Distortion Theorem [Shannon]

$$R(D) = \min_{P_{Y|X} \,:\, \mathbf{E}d < D} I(X;Y)$$

Formula is still an optimization problem

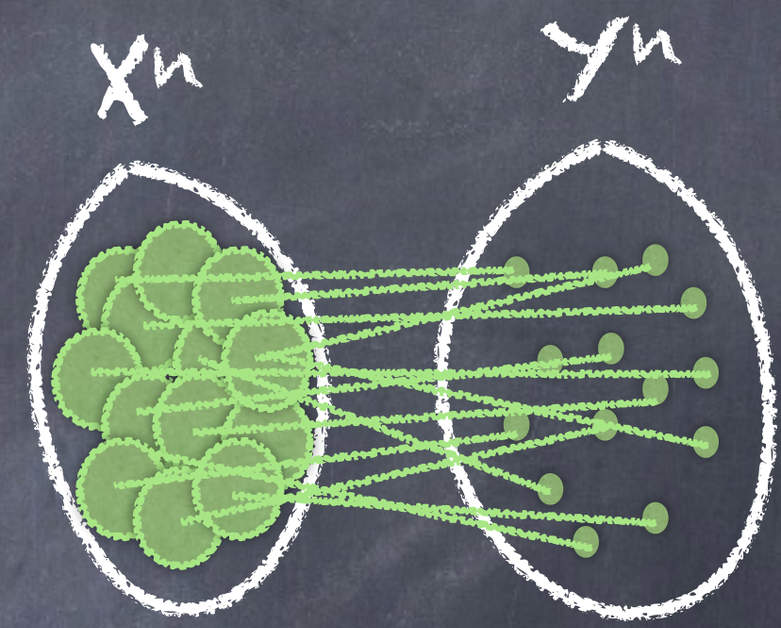
- Choose $P_{Y|X}$ (given P_X and $d(x,y)$):
  - $R > I(X;Y)$
  - $D > E\, d(X,Y)$



Examples

- Binary signal, Hamming distortion:
  - $R(D) = [h(p) - h(D)]_+$
  - Choice of $P_{Y|X}$ is s.t. $P_{X|Y}$ is BSC(D)
- Gaussian signal, squared-error:
  - $R(D) = [1/2 \log(\sigma/D)]_+$
  - Choice of $P_{Y|X}$ is s.t. $P_{X|Y}$ is AGN(D)

Achievability Proof



- Proofs come later
- General idea—Covering
- New simple proof (likelihood encoder)

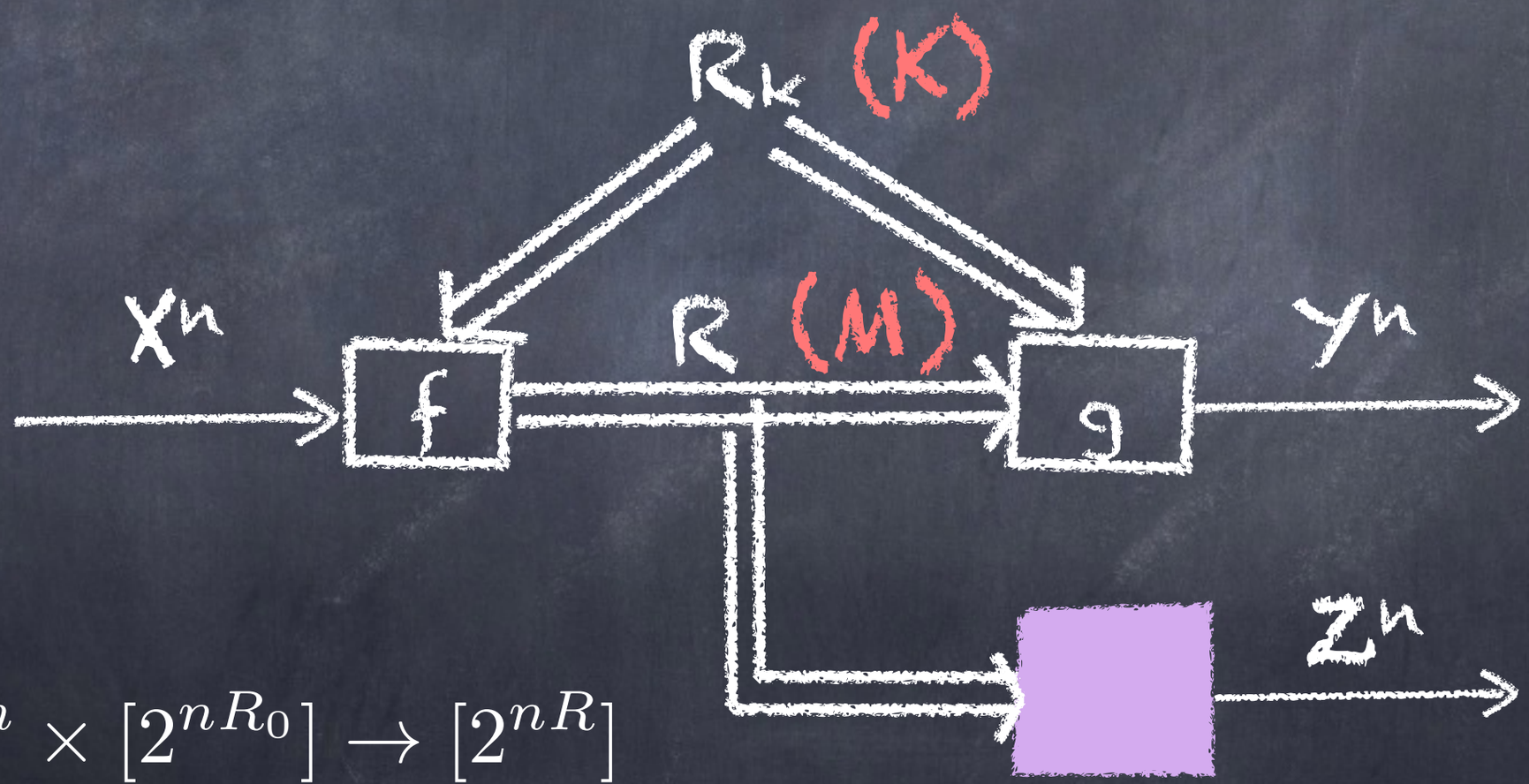
Alternative lossy compression theory

- No distortion function
- Measure quality by entropy:
  - $D = 1/n\ H(X^n|M)$ ⟵ normalized "equivocation"
- <u>Theorem:</u> $R(D) = H(X) - D$
  - Not very interesting theory (almost any encoding works)
  - Not very meaningful operationally

Source Coding for Secrecy

Setting

- Same as Shannon cipher



$$f \;:\; \mathcal{X}^n \times [2^{nR_0}] \to [2^{nR}]$$

$$g \;:\; [2^{nR}] \times [2^{nR_0}] \to \mathcal{Y}^n$$

Performance Metric

- Theory will include:
  - Lossy reconstruction
  - Imperfect secrecy
- Secrecy also measured by distortion
  - Lowest distortion achievable by the eavesdropper

Two distortion functions

- Distortion at the intended receiver:
  - $d_1(x^n,y^n) = 1/n \sum_i d_1(x_i,y_i)$
- Distortion at the eavesdropper
  - $d_2(x^n,z^n) = 1/n \sum_i d_2(x_i,z_i)$

Pessimistic

- Assume eavesdropper makes best use of the information:

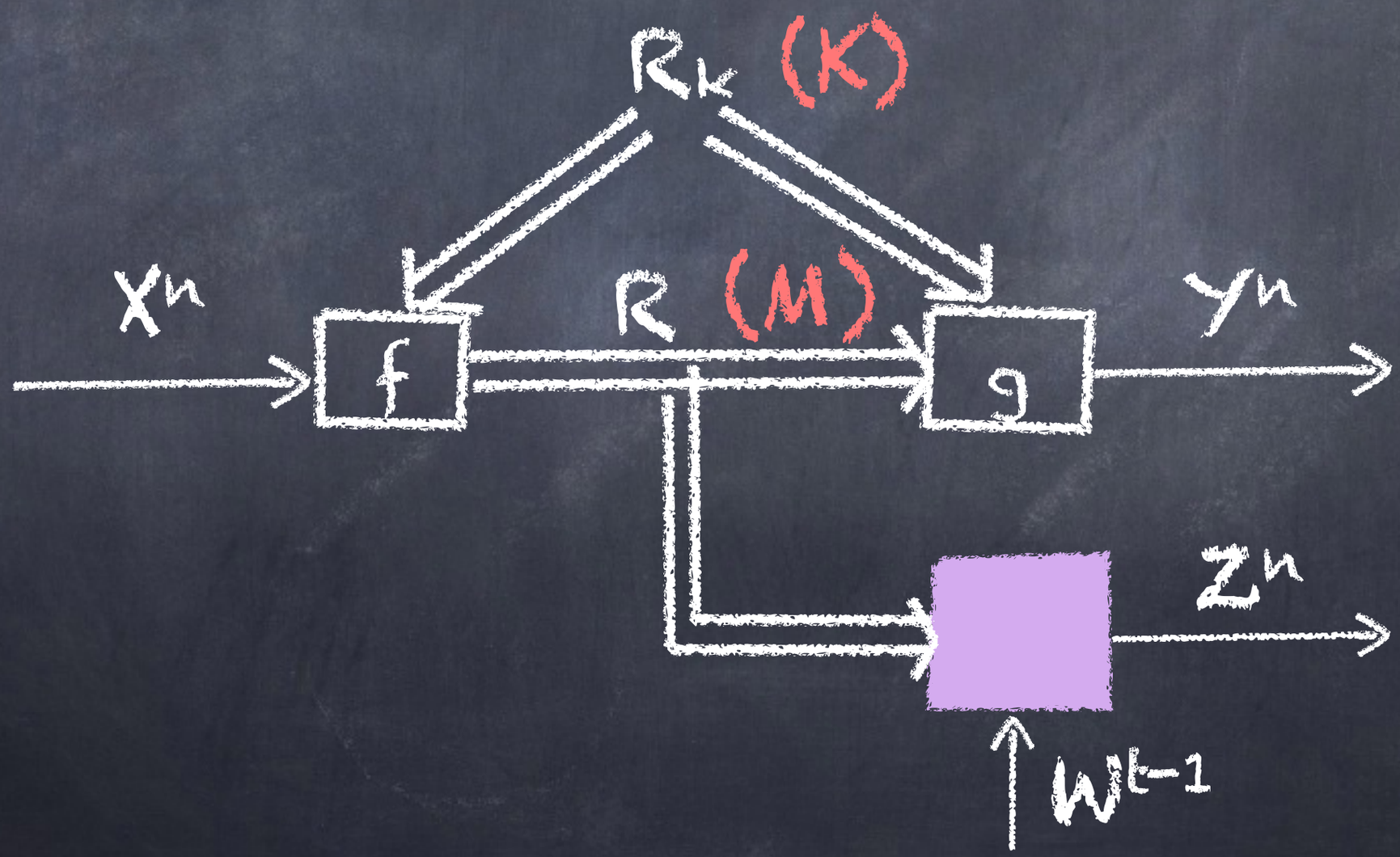
$$D_2 = \min_{Z^n = z^n(M)} \mathbf{E}\, d_2(X^n, Z^n)$$

More pessimistic

- Causal disclosure
  - The eavesdropper sees X and Y causally (potentially noisy)
  - Call this extra information W
- "Real-time" estimation

Setting

- With causal disclosure



What is W?

- Special cases: $W=X$, $W=Y$, $W=(X,Y)$, $W=\emptyset$

  No disclosure

  Full disclosure

- Noisy observations: $P_{W_1|X}$, $P_{W_2|Y}$, $W=(W_1,W_2)$

Why causal disclosure?

- More general than no disclosure
- Application to real-time systems
  - (however, the encoder is not RT)
- More satisfying and natural theory emerges
- Game theory interpretation

Non-causal disclosure

- Some signals are not "time signals"
- We will also consider non-causal disclosure
- The theory is almost the same as causal disclosure

"Payoff function"

- Allow for joint distortion function
  - $\pi(x,y,z)$
  - $\pi(x^n,y^n,z^n) = 1/n \sum_i \pi(x_i,y_i,z_i)$
- Can think of π as payoff in repeated game

Definition of achievability

- (R, R_K, Π) is achievable if there exists an n, f, and g operating at rates R and R_K such that

$$\min_{\text{(adversary)}} E\, \pi(X^n, Y^n, Z^n) \geq \Pi$$

Rate-Distortion Theory for Secrecy Systems

- (R, R_k, Π) achievable if and only if
  - There exist U, V, Y s.t.
    - $W_1 - X - (U,V) - Y - W_2$
    - $R > I(X;U,V)$
    - $R_k > I(W;V|U)$
    - $\min_{z(.)} E\, \pi(X,Y,z(U)) > \Pi$

Recall that we are given:
P_X, $P_{W1|X}$, and $P_{W2|Y}$

No Disclosure

- $W=\emptyset$
- Optimal to choose: $U=\emptyset$, $V=Y$
- (R,R_k,Π) achievable if and only if
  - There exists $P_{Y|X}$ s.t.
  - $R > I(X;Y)$
  - $R_k > 0$
  - $\min_z E\ \pi(X,Y,z) > \Pi$

Perfect secrecy with negligible key!

Example of negligible key

- Suppose X is Bern(1/2)
- We can transmit X^n losslessly
  - n bits of communication ($R=1$)
  - 1 bit of secret key ($R_K=1/n \to 0$)
- If $K=0$, $M=X^n$.
- If $K=1$, M is bit-wise-not of X^n.

Eavesdropper cannot successfully estimate any bit of X^n

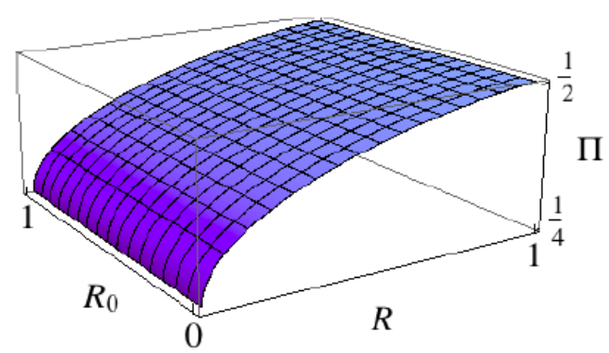
Full Causal Disclosure

- (R, R_k, Π) achievable if and only if
  - There exist U,V,Y s.t.
    - $X - (U,V) - Y$
    - $R > I(X;U,V)$
    - $R_k > I(X,Y;V|U)$
    - $\min_{z(.)} E\ \pi(X,Y,z(U)) > \Pi$

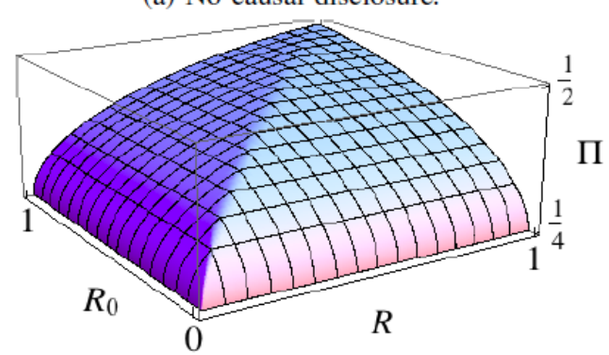
Forced to achieve very robust secrecy

Binary Jamming Example

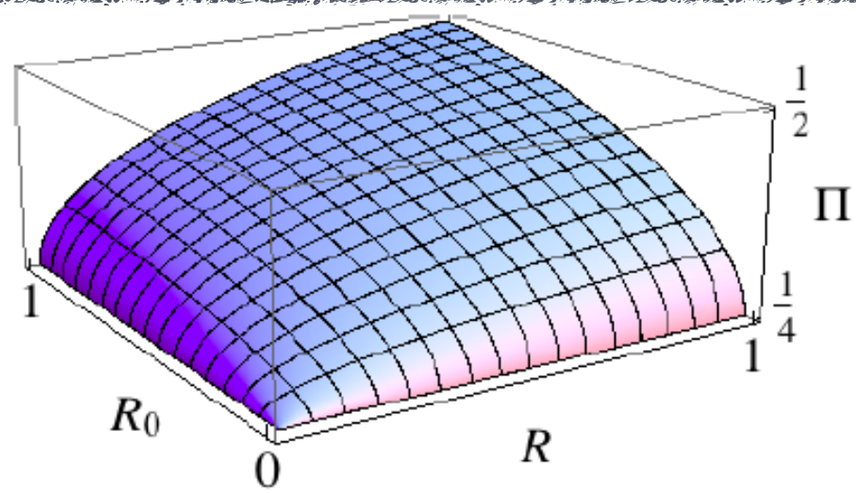
- X is Bern(1/2)
- $\pi(x,y,z) = \delta\{x=y\neq z\}$



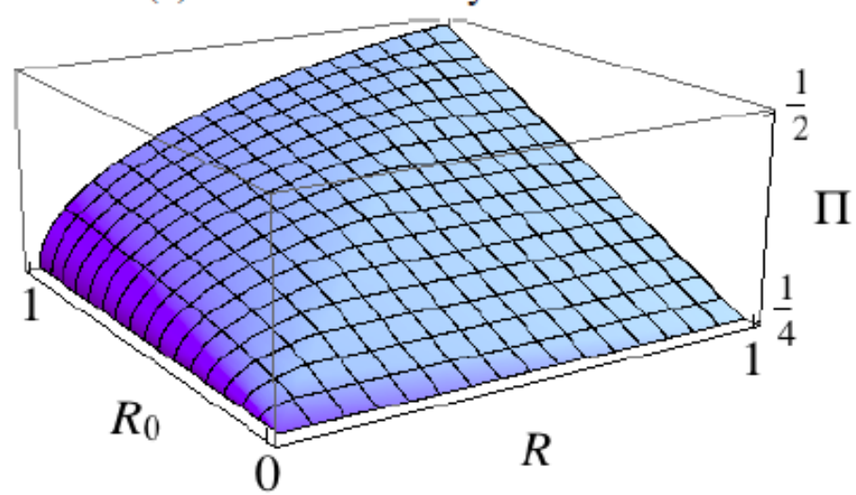
(a) No causal disclosure.



(b) Node A causally disclosed.



(c) Node B causally disclosed.



(d) Nodes A and B causally disclosed.

Lossless Transmission (Imperfect Secrecy)

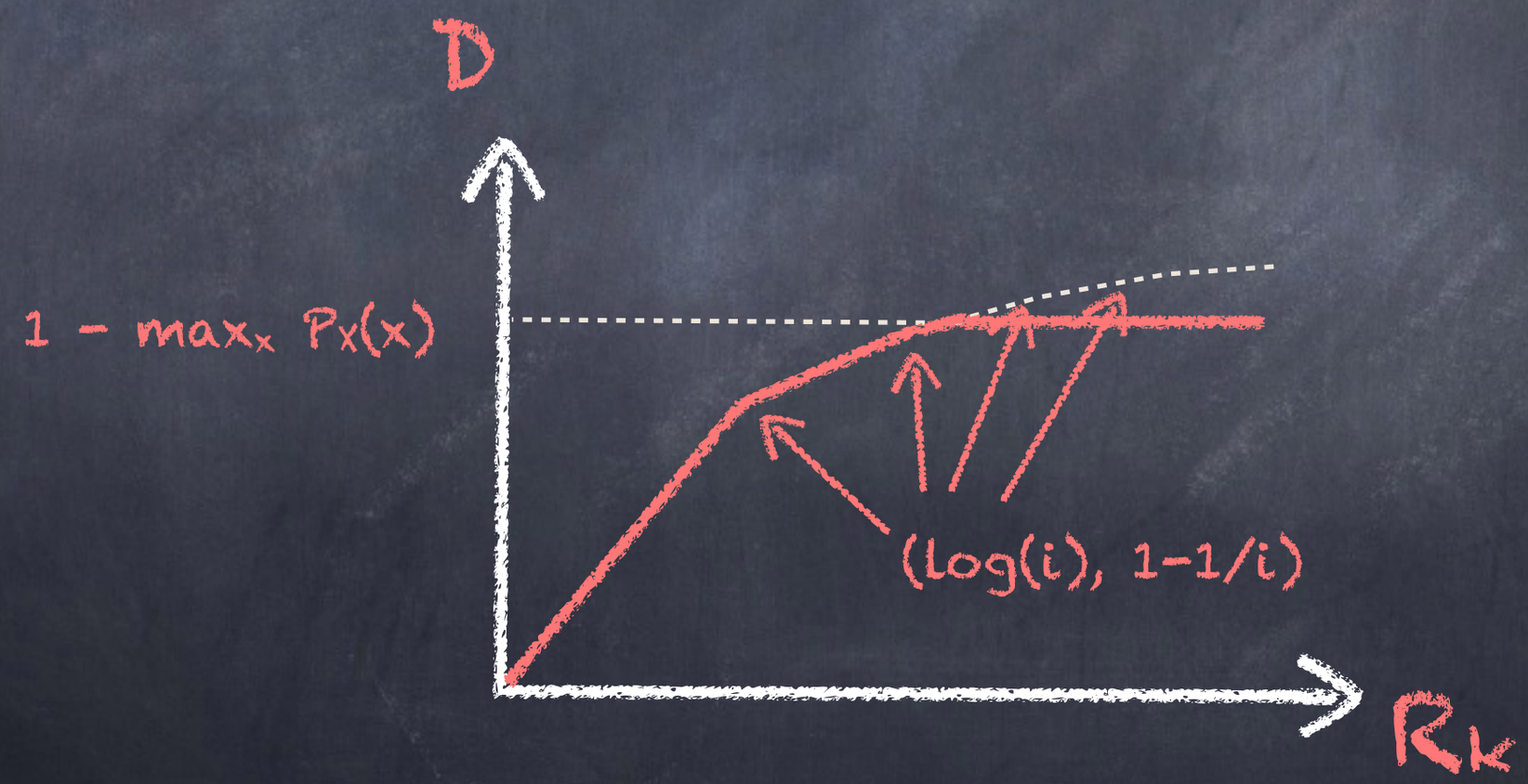
- Set $Y=X$ in previous theorem
- Must have $V=X$ also
- (R,R_k,Π) achievable if and only if
  - There exists U s.t.
  - $W-X-U$
  - $R > H(X)$ ← Not affected by optimization
  - $R_k > I(W;X|U)$ ← Reduces to a linear program
  - $\min_{z(.)} E\ \pi(X,z(U)) > \Pi$

Lossless Transmission (Imperfect Secrecy)

- Set $Y=X$ in previous theorem
- Must have $V=X$ also
- (R,R_k,D) achievable if and only if
  - There exists U s.t.
  - $W-X-U$
  - $R > H(X)$ ← Not affected by optimization
  - $R_k > I(W;X|U)$ ← Reduces to a linear program
  - $\min_{z(.)} E\ \pi(X,z(U)) > D$

Hamming Distortion

- For any source distribution and full causal disclosure



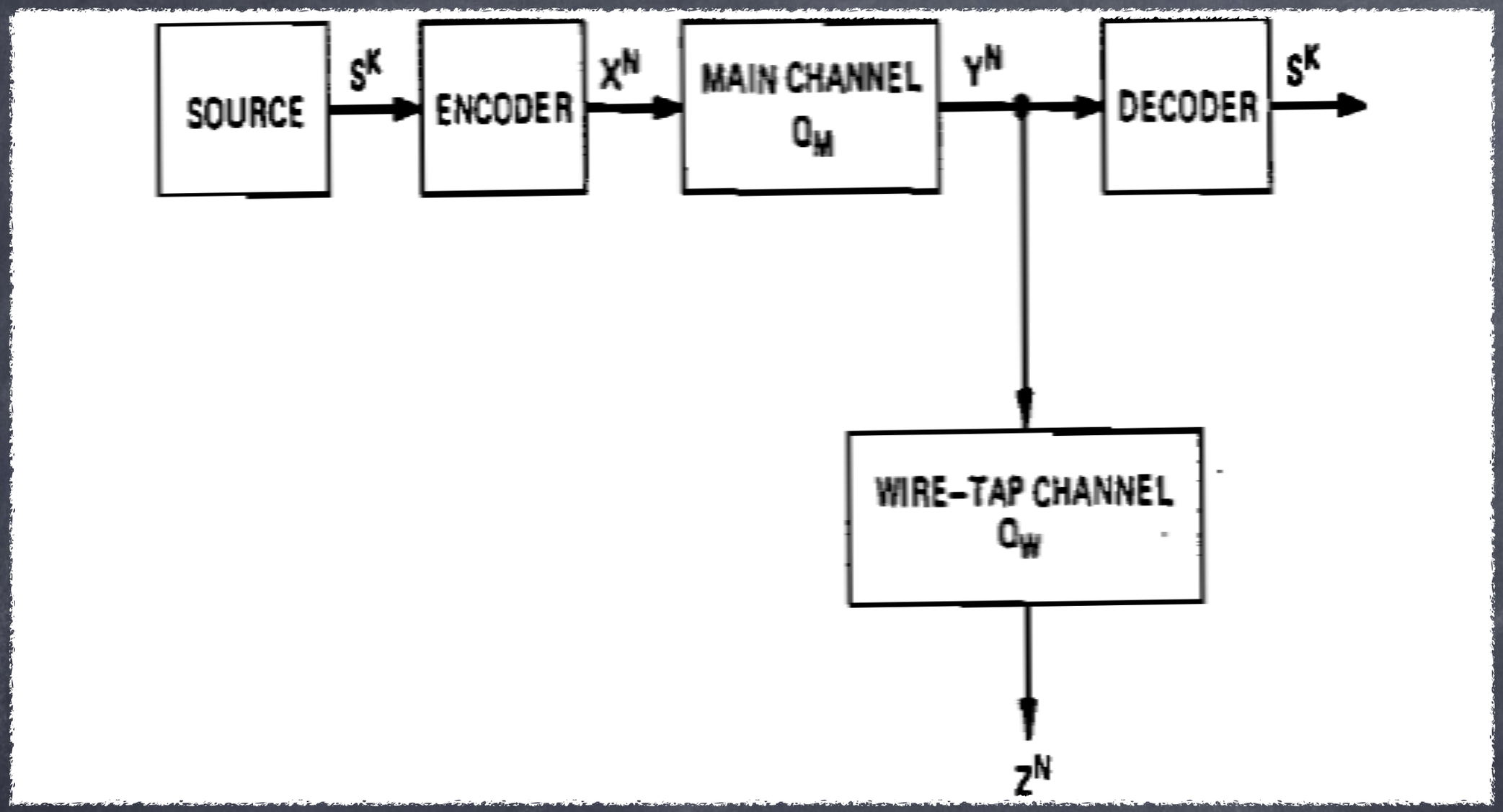
Non-causal disclosure

- Disclose all times except the one being estimated
  - Same achievable region
- Disclose all times including the one being estimated
  - Eavesdropper uses U and W

Equivocation Region

Alternative Problem

- No distortion at the eavesdropper
- Measure secrecy by equivocation rate
  - $\Delta_X = 1/n\ H(X^n|M)$
- Or, for lossy compression
  - $\Delta_{X,Y} = 1/n\ H(X^n,Y^n|M)$



Wyner's Wiretap Channel

'75

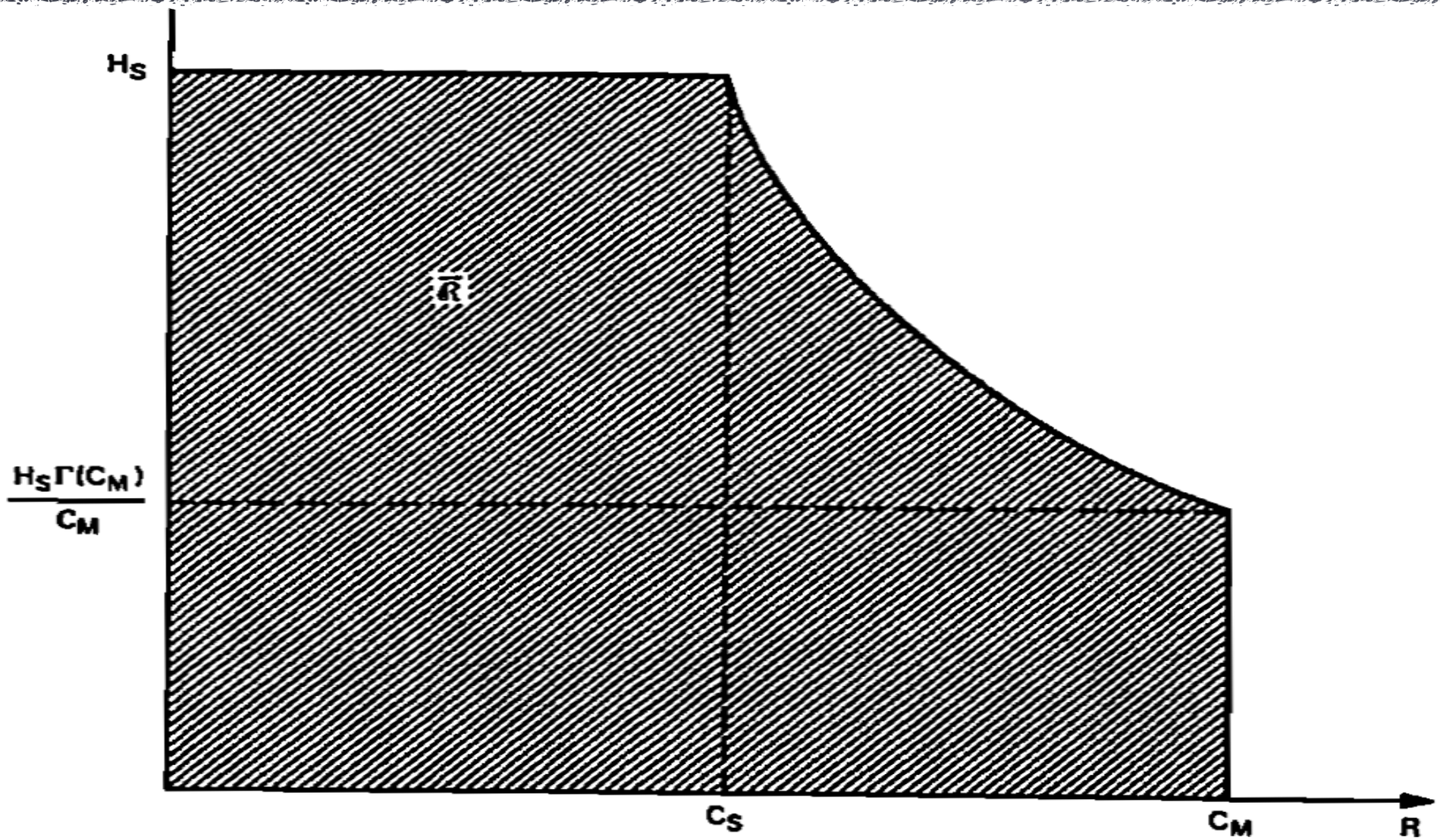


Fig. 3—Region $\overline{\mathcal{R}}$.

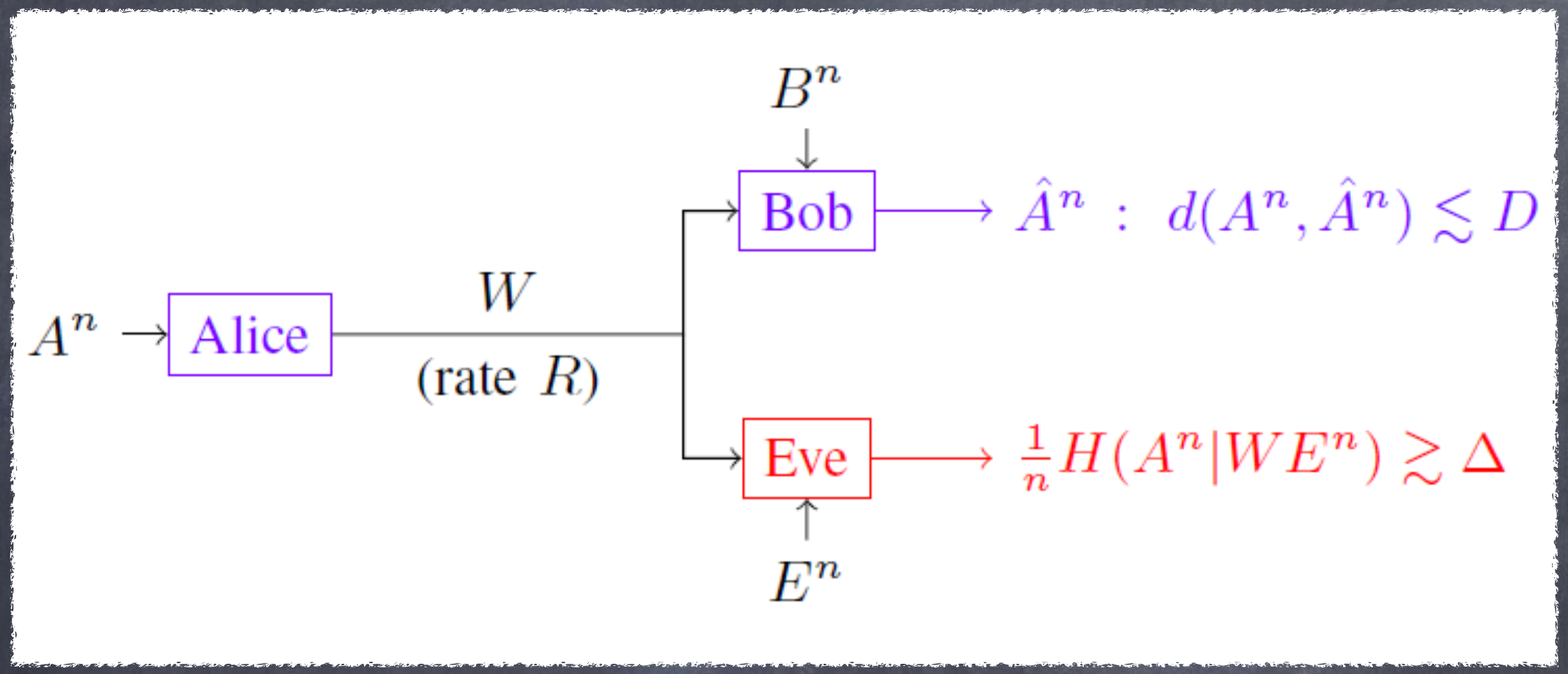
Equivocation rate

Perfect secrecy until C_S

Following Wyner [8], we shall measure confidentiality by equivocation. Our main result is a single-letter characterization of the set of triples (R_1, R_e, R_0) such that, in addition to a common message at rate R_0, a private message can be sent reliably at rate R_1 to receiver 1 with equivocation at least R_e per channel use at receiver 2. This con-

Csiszar-Korner

'78



Villard-Piantanida

'10

Recover equivocation from RD Theory

- Equivocation rate is one special case of rate-distortion theory
- Causal disclosure is necessary

Log-loss function

- Distortion function
  - $d(x,z) = \log 1/z(x)$
  - z is a distribution

- Notice: $\min_{z(m)} E\, d(X,z(M)) = H(X|M)$

  $\min_{z(x,m)} E \log 1/z(x,m)$

Equivocation if X

- Log-loss function
- Causal disclosure of X (i.e. W=X)

$$\min_{\{z(m,x_{i+1})\}} E\ 1/n\ \Sigma_i \log 1/z(X_i,M,X^{i-1})$$
$$= 1/n\ \Sigma_i\ H(X_i|M,X^{i-1})$$
$$= 1/n\ H(X^n|M)$$

- Distortion metric is now equivocation

Rate-Distortion Theorem Applies

- (R, R_k, Π) achievable if and only if
  - There exist U, V, Y s.t.
    - $W_1 - X - (U,V) - Y - W_2$
    - $R > I(X;U,V)$
    - $R_k > I(W;V|U)$
    - $\min_{z(.)} E\, \pi(X,Y,z(U)) > \Pi$

Rate-Distortion Theorem Applies

- (R, R_k, Δ) achievable if and only if
  - There exist U, V, Y s.t.
    - $X - (U,V) - Y$
    - $R > I(X;U,V)$
    - $R_k > I(X;V|U)$
    - $\min_{z(.)} E \log(1/z(x,u)) > \Delta$

Rate-Distortion Theorem Applies

- (R, R_K, Δ) achievable if and only if
  - There exist U, V, Y s.t.
    - $X - (U,V) - Y$
    - $R > I(X;U,V)$
    - $R_K > I(X;V|U)$
    - $H(X|U) > \Delta$

Rate-Distortion Theorem Applies

- (R, R_K, Δ) achievable if and only if
  - There exist Y s.t.
    - $R > I(X;Y)$
    - $H(X|Y) + R_K > \Delta$
    - $H(X) > \Delta$

implicit distortion constraint on reconstruction

Summary

- Equivocation of X

$$\max_{\{n,f,g\}} 1/n\ H(X^n|M) = \min \{H(X),\ H(X|Y) + R_K\}$$

- Equivocation of Y

$$\max_{\{n,f,g\}} 1/n\ H(Y^n|M) = \max_{\{X-U-Y\}} \min \{H(Y),\ H(Y|U) + R_K\}$$

- Equivocation of (X,Y)

$$\max_{\{n,f,g\}} 1/n\ H(X^n,Y^n|M) = \max_{\{X-U-Y\}} \min \{H(X,Y),\ H(X,Y|U) + R_K\}$$

Achievability Proofs

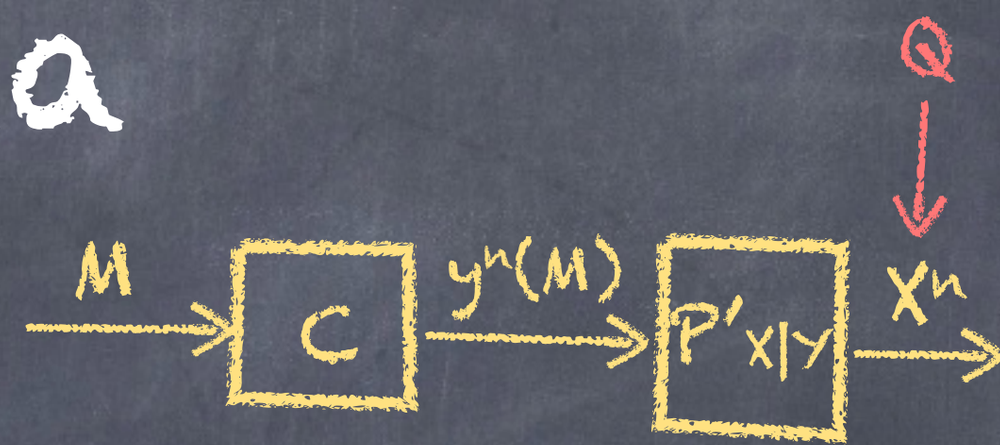
Likelihood Encoder

- Codebook: $C = \{y^n(m)\}$
- Test Channel: $P'_{Y|X}$
- Let $L(m,x^n)$ be the likelihood of message m upon observing $X^n=x^n$ through a memoryless channel $P'_{X|Y}$

$$L(m,x^n) = \prod_i P'_{X|Y}(x_i|y_i(m))$$

- Encoder: $P_{M|X} \propto L(m,x^n)$

Soft-covering Lemma



- Codebook: $C = \{y^n(m)\}$
- Memoryless Channel: $P'_{Y|X}$
- Distribution induced by uniformly sampling from the codebook and passing through the channel:

$$Q(x^n) = \frac{1}{|C|}\sum_m \prod_i P'_{X|Y}(x_i|y_i(m))$$

- If $R>I(X;Y)$ and codebook is generated randomly:

$$E_C\, d_{TV}(Q^n, P^n) \to 0$$

Properties of TV

- Continuity of expectation:
  - $|E_P f(X) - E_Q f(X)| \leq d_{TV}(P||Q)\ 2\ \max_x |f(x)|$
- Triangle inequality
- Marginal:
  - $d_{TV}(P_X||Q_X) \leq d_{TV}(P_{X,Y}||Q_{X,Y})$
- Same channel:
  - $d_{TV}(P_X||Q_X) = d_{TV}(P_X P_{Y|X}||Q_X P_{Y|X})$

Proof of RD Theorem

- Choose $P'_{Y|X}$ s.t. $R > I(X;Y)$ and $E\, d(X,Y) < D$
- Likelihood encoder (random codebook)
- Induced distribution
  - $P_X(x^n) P_{LE}(m|x^n) P_D(y^n|m)$
- Idealized distribution
  - $Q_{X,M}(x^n,m) P_D(y^n|m)$

Q is uniform on M and memoryless channel from $y^n(M)$ to X^n according to $P'_{X|Y}$

Approximating Distribution

- Same conditional distribution:
  - $P(m, y^n | x^n) = Q(m, y^n | x^n)$
- Soft-covering lemma applies to Q:
  - $E_C \, d_{TV}(P(x^n) \| Q(x^n)) \to 0$

Distortion of Q

$$E_C \, E_Q \, d(X^n, Y^n) = E_{P'} \, d(X, Y)$$

Final Step of Proof

$$E_C\, E_P\, d(X^n, Y^n) \leq E_C\, E_P\, d(X^n, Y^n) + E_C\, d_{TV}(P \| Q)\; 2\, d_{max}$$

$$= E_{P'}\, d(X, Y) + \varepsilon_n$$

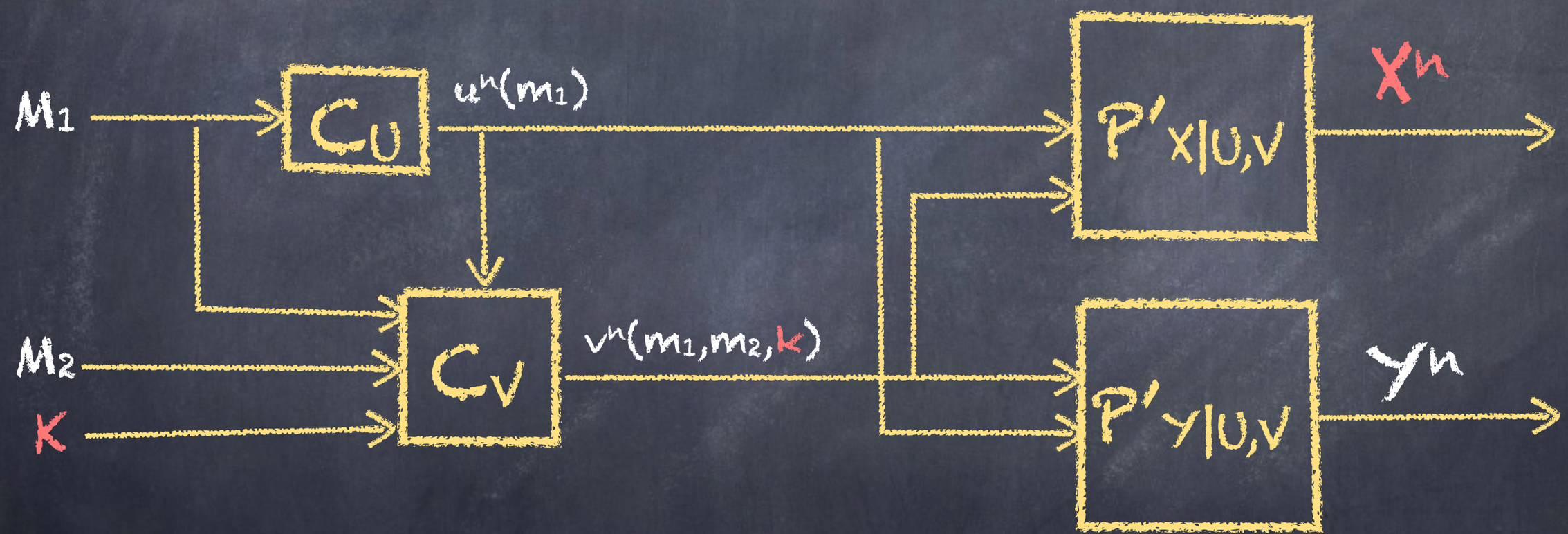
$$< D$$

Proof of RD for secrecy systems

- Full disclosure: $W = (X,Y)$
- Choose U, V, Y s.t. $X-(U,V)-Y$, $R>I(X;U,V)$, $R_k>(X,Y;V|U)$

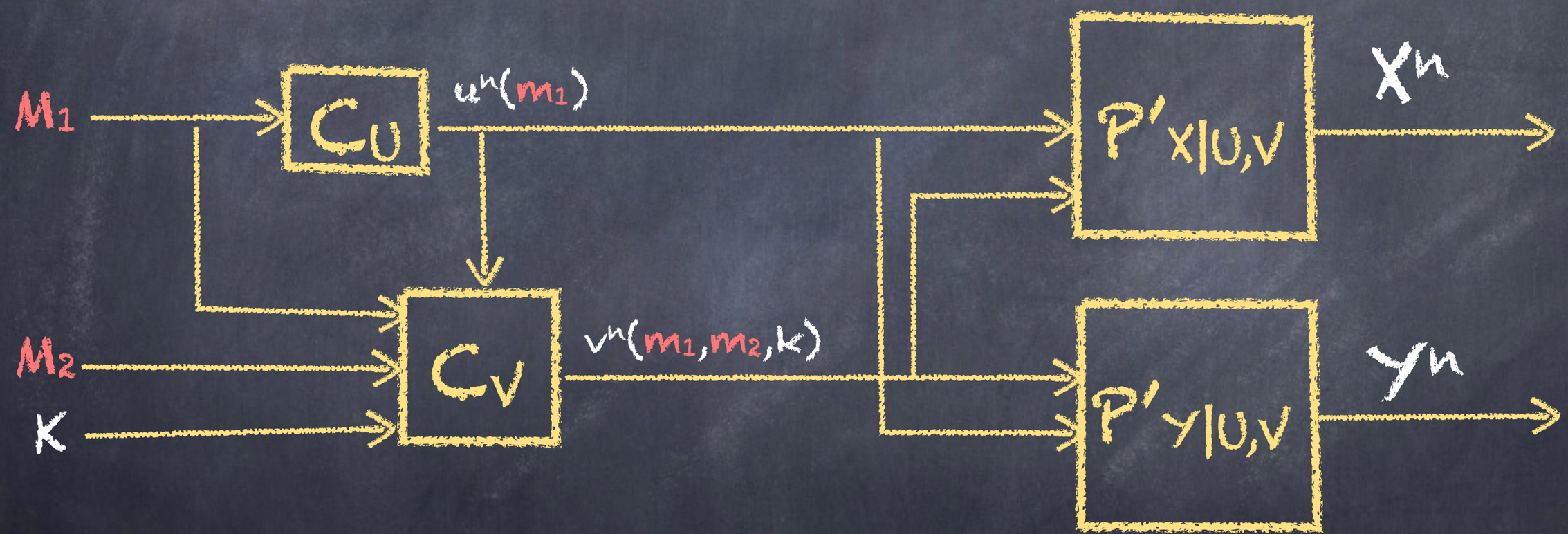
  $R_1>I(X;U)$ $R_2>I(X;V|U)$
- Superposition code:
  - $C_U = \{u^n(m_1)\}$, $C_V = \{v^n(m_1,m_2,k)\}$
- Likelihood Encoder
- Decoder looks up u^n and v^n then produces y^n according to $P'_{Y|U,V}$

Approximating Distribution



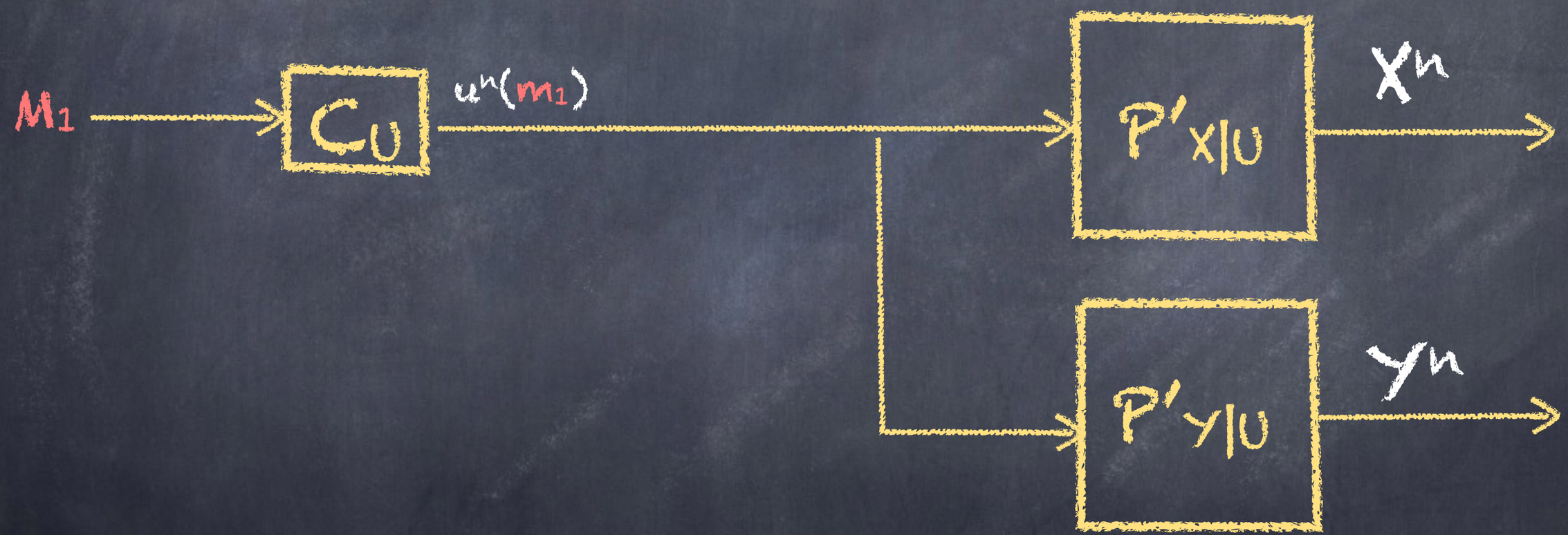
Superposition soft-covering: $P_{X|K} \approx Q_{X|K}$

Secrecy Analysis



Conditional soft-covering: $P_{X,Y|M1,M2} = P_{X,Y|u(m)}$

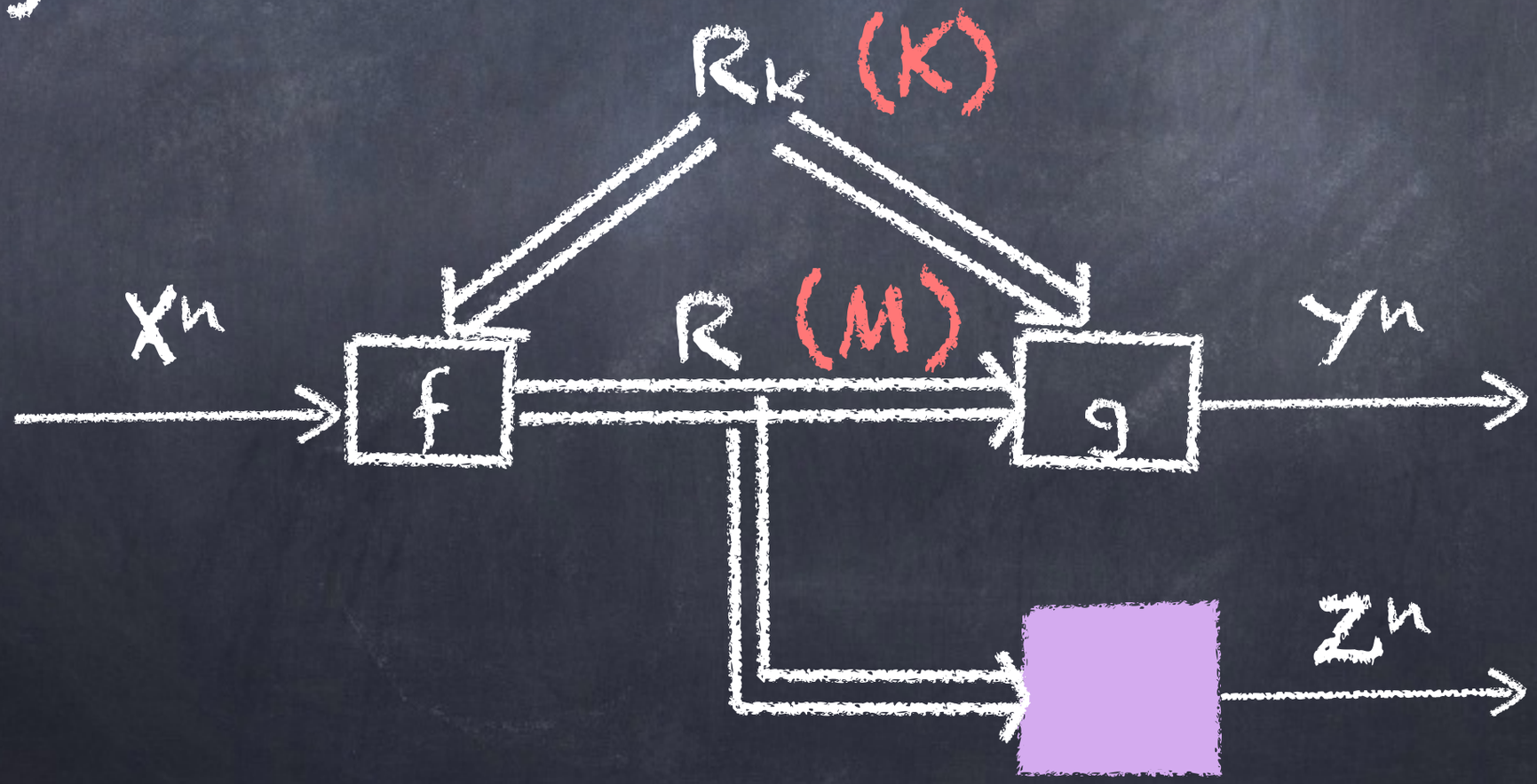
Ideal Distribution



Memoryless channel: Adversary's best z_i only depends on $u_i(m_1)$

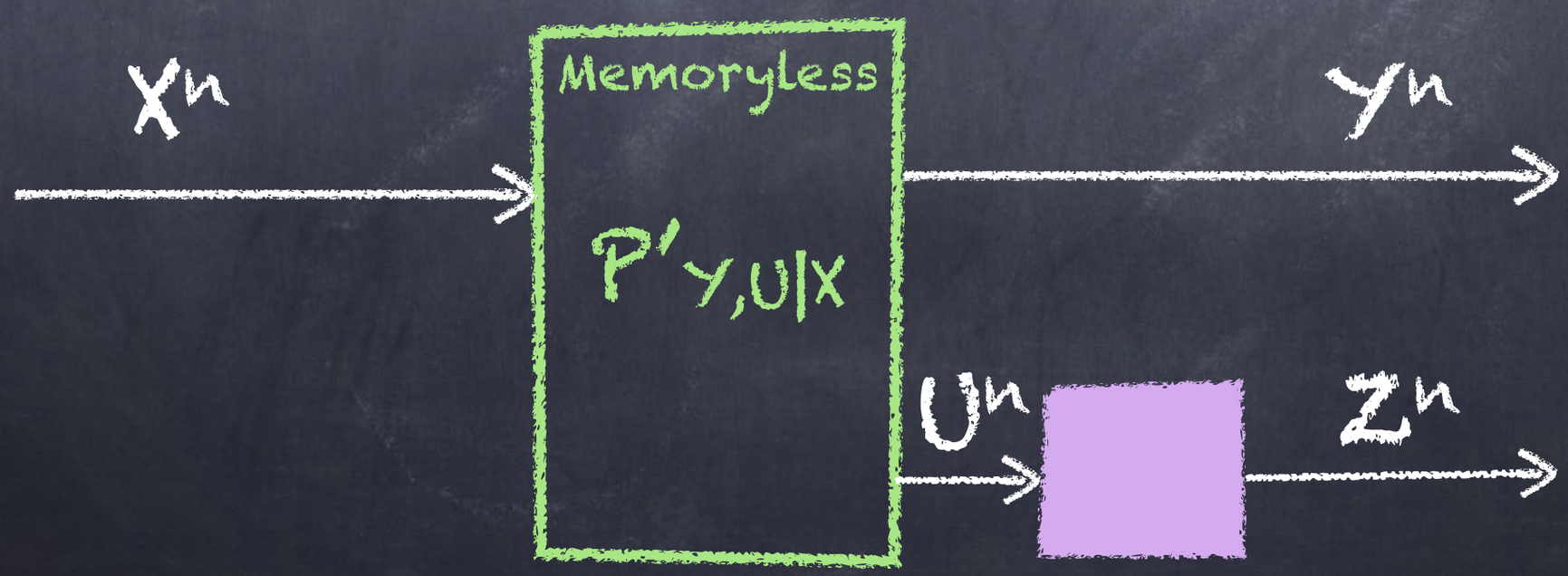
Synthetic noise

- RDSS with full causal disclosure extracts an intuitive communication system



Synthetic noise

- RDSS with full causal disclosure extracts an intuitive communication system



Other Proofs

- Wyner-Ziv
- Villard-Piantanida

Converse Proof

Rate-Distortion Theory for Secrecy Systems

- Suppose an (R, R_f, Π) scheme exists:
  - $H(M) \leq nR$
  - $H(K) \leq nR_K$
  - $X^n \perp K$
  - $W_1^n - X^n - (M,K) - Y^n - W_2^n$ (Markov chain)
- Define: $U_i = (M, W_1^{i-1}, W_2^{i-1})$ and $V = K$
- Let $Q \sim \mathrm{Unif}\{1,\ldots,n\}$, independent of all above variables
- Define: $U = (U_Q, Q)$, $X = X_Q$, $Y = Y_Q$, $W = (W_{1,Q}, W_{2,Q})$

Rate Bound

$$
\begin{aligned}
nR &\geq H(M) \\
&\geq I(M;X^n|K) \\
&= I(M,K;X^n) \\
&= \Sigma_q\ I(M,K;X_q|X^{q-1}) \\
&= \Sigma_q\ I(M,K,X^{q-1};X_q) \\
&= \Sigma_q\ I(M,K,X^{q-1},W_1^{q-1},W_2^{q-1};X_q) \\
&\geq \Sigma_q\ I(M,K,W_1^{q-1},W_2^{q-1};X_q) \\
&\geq \Sigma_q\ I(U_q,V;X_q) \\
&= n\ I(U_Q,V;X_Q|Q) \\
&= n\ I(U_Q,V,Q;X_Q) \\
&= n\ I(U,V;X)
\end{aligned}
$$

Key Rate Bound

$$
\begin{aligned}
nR_K &\geq H(K) \\
&\geq I(K;W_1^n,W_2^n|M) \\
&= \Sigma_q\ I(K;W_{1,q},W_{2,q}|M,W_1^{q-1},W_2^{q-1}) \\
&= \Sigma_q\ I(K;W_{1,q},W_{2,q}|U_q) \\
&= n\ I(K;W_{1,Q},W_{2,Q}|U_Q,Q) \\
&= n\ I(V;W|U)
\end{aligned}
$$

Performance Bound

$$\Pi \leq \min_{(\text{adversary})} E\ 1/n\ \Sigma_i\ \pi(X_i, Y_i, Z_i)$$
$$= \min_{(\text{adversary})} E\ E\ (\pi(X_Q, Y_Q, Z_Q) | Q)$$
$$= \min_{(\text{adversary})} E\ \pi(X, Y, Z)$$
$$= \min_{z(u)} E\ \pi(X, Y, z(U))$$