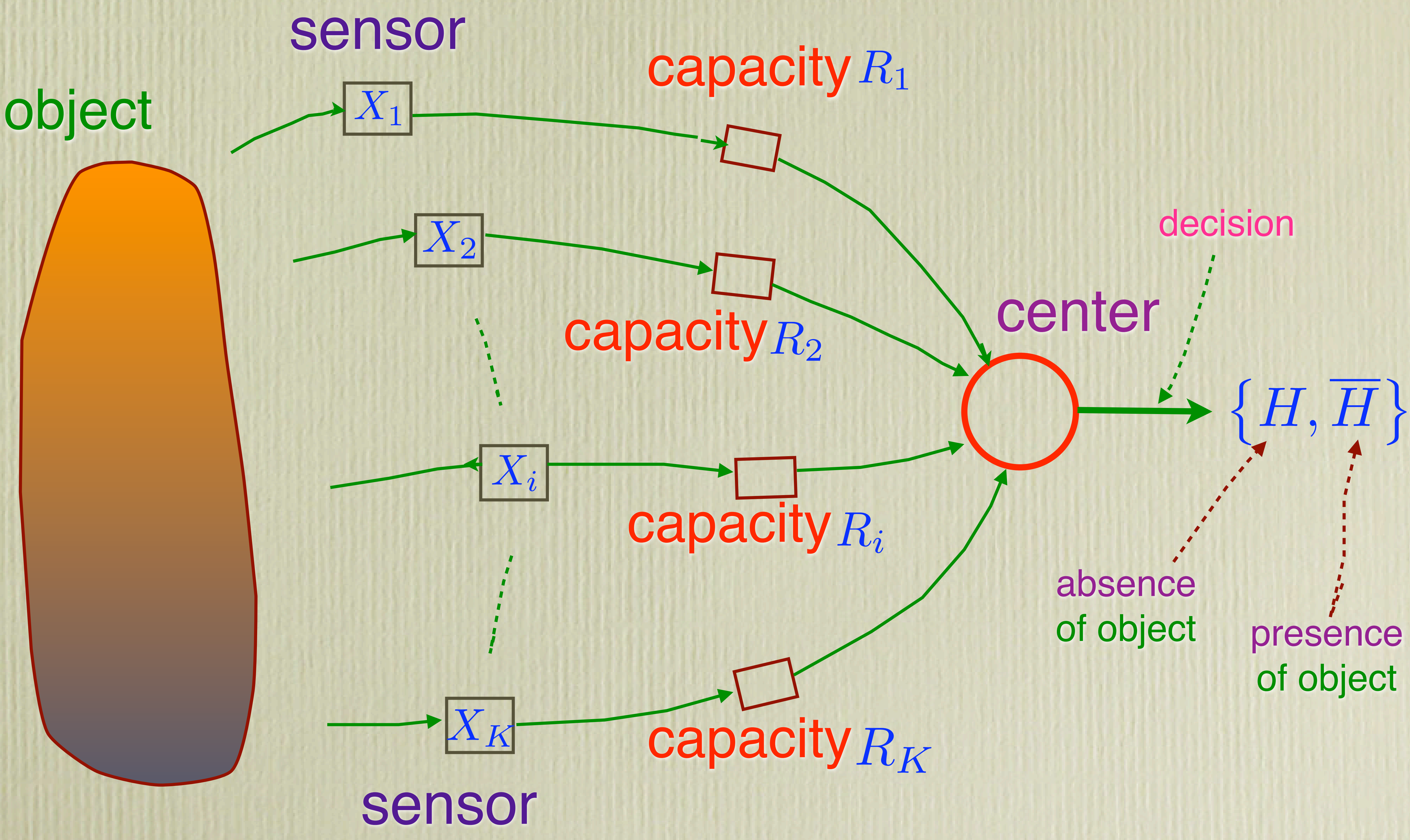
Trade-off of Data Compression and Hypothesis Testing

Te Sun Han,
jointly with H. Shimokawa and S. Amari

at BIRS, Jan. 15-20, 2012

I. What is hypothesis testing with data compression?

Distributed sensor system



sensor

object

X_1

capacity R_1

X_2

capacity R_2

decision

center

$\{H, \overline{H}\}$

X_i

capacity R_i

absence
of object

presence
of object

X_K

capacity R_K

sensor

- $X_1, X_2, \cdots, X_K$ are mutually correlated

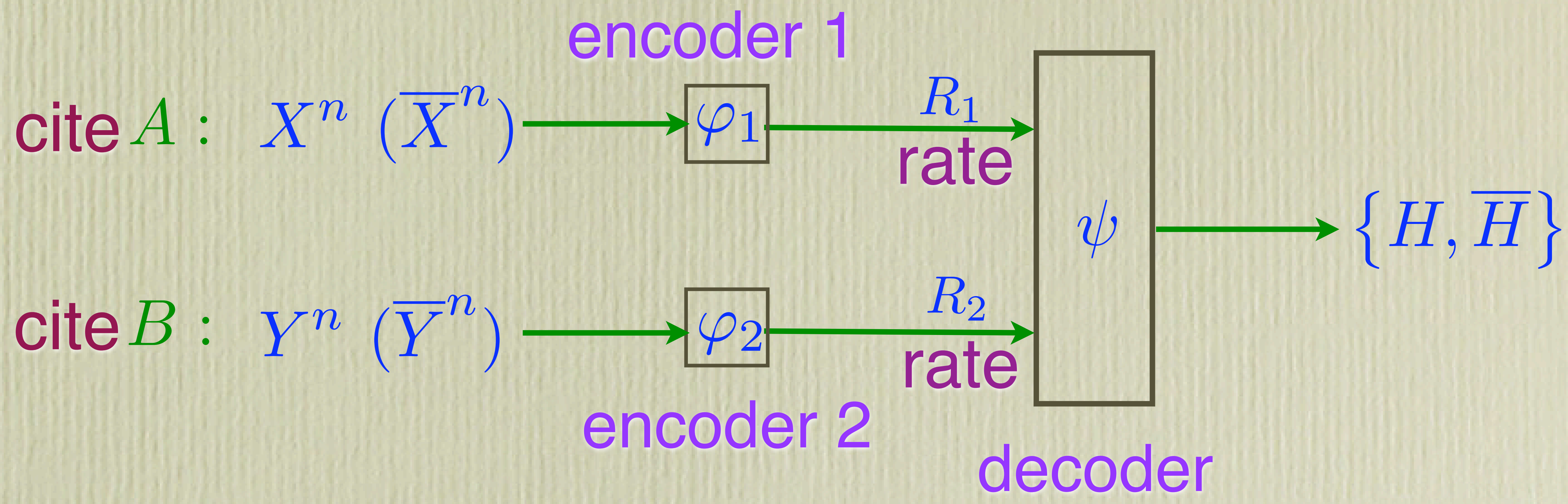
I. Formulation of hypothesis testing $(K=2)$

absence ⇢ $H : XY$ (null hypothesis)

presence ⇢ $\overline{H} : \overline{XY}$ (alternative hypothesis)

- $XY,\ \overline{XY}$: correlated random variables with values in finite sets $\mathcal{X}, \mathcal{Y}$
- $P_{XY}, P_{\overline{XY}}$: probability distributions of $XY,\ \overline{XY}$
- $X^nY^n = (X_1Y_1, \cdots, X_nY_n)$; i.i.d. variables $\sim P_{XY}$
- $\overline{X}^n\overline{Y}^n = (\overline{X}_1\overline{Y}_1 \cdots, \overline{X}_n\overline{Y}_n)$: i.i.d. variables $\sim P_{\overline{XY}}$

II. Multiterminal hypothesis testing $(K = 2)$



encoder 1

cite A : $X^n \ (\overline{X}^n)$ φ_1 R_1 rate

ψ $\{H, \overline{H}\}$

cite B : $Y^n \ (\overline{Y}^n)$ φ_2 R_2 rate

encoder 2

decoder

- encoders $\varphi_1 : \ \mathcal{X}^n \to \mathcal{M}_1 \equiv \{1, 2, \cdots, M_1\}$

  $\varphi_2 : \ \mathcal{Y}^n \to \mathcal{M}_2 \equiv \{1, 2, \cdots, M_2\}$
- rate constraints $\frac{1}{n} \log M_1 \le R_1, \quad \frac{1}{n} \log M_2 \le R_2$
- decoder $\psi : \ \mathcal{M}_1 \times \mathcal{M}_2 \to \{H, \overline{H}\}$

- Acceptance region $(\subset \mathcal{X}^n \times \mathcal{Y}^n)$:collection of cells

$$\mathcal{A}_n \equiv \{(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \psi(\varphi_1(\mathbf{x}), \varphi_2(\mathbf{y})) = H\}$$

- type 1 error (prob. of false alarm)

$\alpha_n \equiv \Pr\{X^n Y^n \in \mathcal{A}_n^c\}$ (c denotes the complement)

@error probability that absence is mistaken for presence

- type 2 error (prob. of missing)

$$\beta_n \equiv \Pr\left\{\overline{X}^n \overline{Y}^n \in \mathcal{A}_n\right\}$$

@error probability that presence is mistaken for absence

Problem formulation

- We consider the following constraint on type 1 error probability:

$$\alpha_n \le \varepsilon \quad (0 < \forall \varepsilon < 1 : \text{constant})$$

- Under this constraint, define the optimal type 2 error probability as

$$\beta_n^*(R_1, R_2, \varepsilon) \equiv \min_{\varphi_1, \varphi_2, \psi, \alpha_n \le \varepsilon} \beta_n$$

- Define the power exponent as:

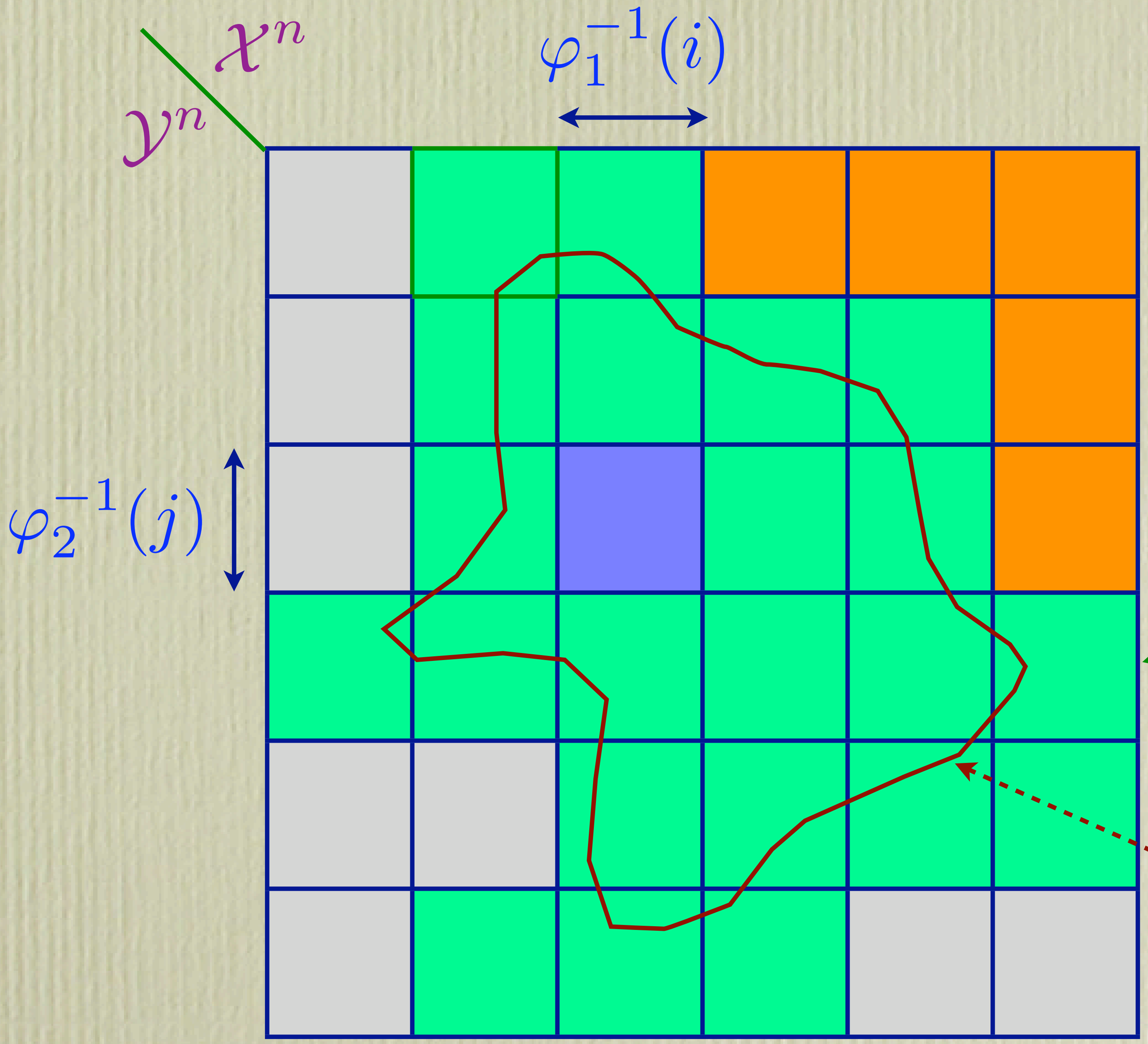
$$\theta(R_1, R_2, \varepsilon) \equiv \liminf_{n \to \infty} \left(-\frac{1}{n} \log \beta_n^*(R_1, R_2, \varepsilon) \right)$$

$$\Downarrow$$

Problem: Determine the function $\theta(R_1, R_2, \varepsilon)$ of R_1, R_2 or its lower bounds

- It is obvious that this function is nondecreasing in each of R_1, R_2

Remark: about the acceptande region $\mathcal{A}_n$



$\mathcal{A}_n$ collection of rectangular cells

Neyman-Pearson’s optimal region

III.Lower bounds of $\theta(R_1, R_2)$

■ Auxiliary random variables U, V

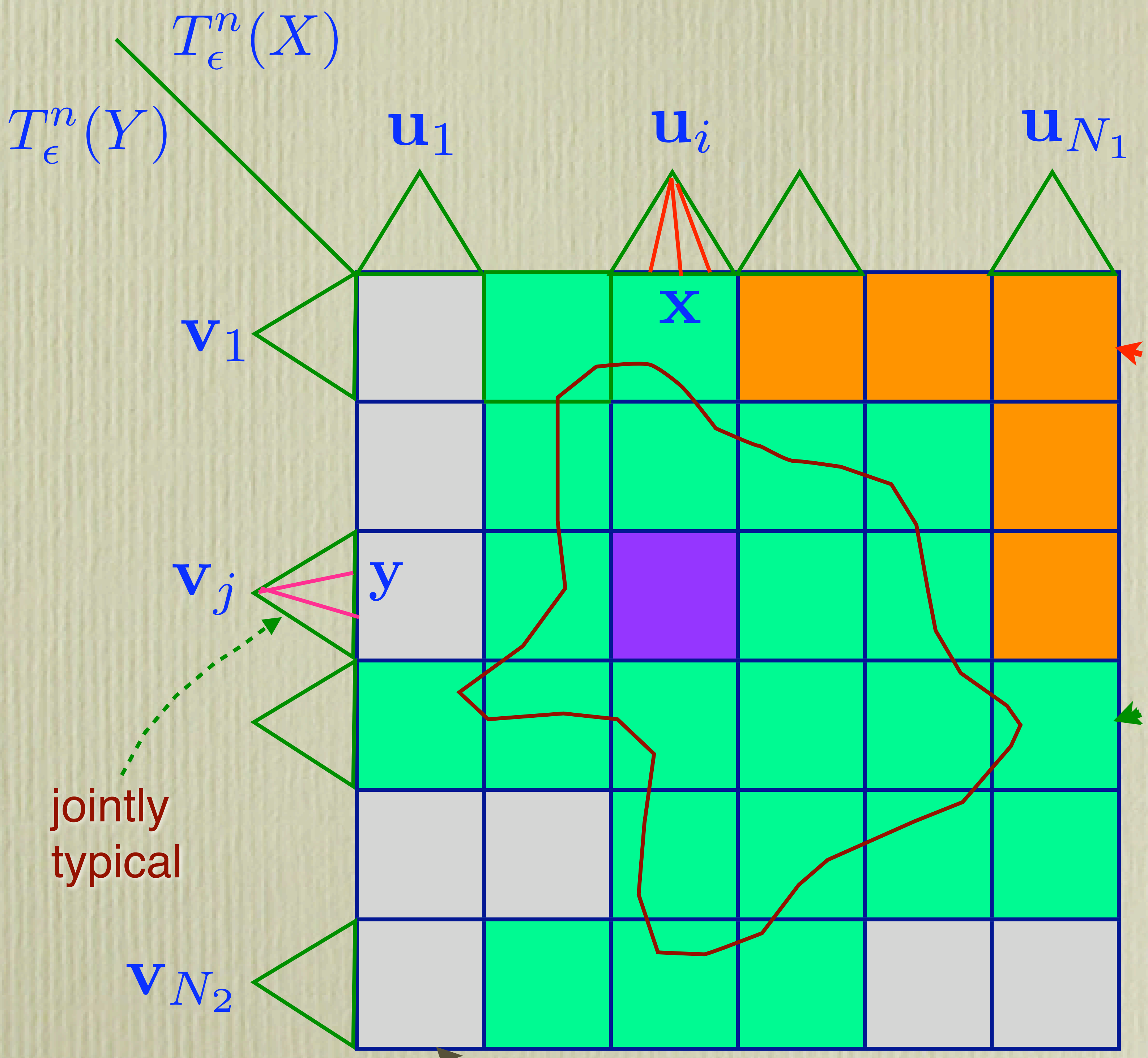
⇒for quantization

- $\mathcal{S}_{XY} = \{(P_{U|X}, P_{V|Y}) : U \to X \to Y \to V\}$
- For each $(P_{U|X}, P_{V|Y}) \in \mathcal{S}_{XY}$:

  define the Markov chain $\overline{U} \to \overline{X} \to \overline{Y} \to \overline{V}$
  by the condition $P_{\overline{U}|\overline{X}} = P_{U|X}, P_{\overline{V}|\overline{Y}} = P_{V|Y}$

Quantization

- $(\mathbf{u}_i, \mathbf{x}) \in T_\epsilon^n(UX)$: i is unique for $\mathbf{x}$
- $(\mathbf{v}_j, \mathbf{y}) \in T_\epsilon^n(VY)$: j is unique for $\mathbf{y}$



- $N_1 \simeq \exp[nI(U;X)]$
- $N_2 \simeq \exp[nI(V;Y)]$

$\mathcal{D}_n : H(\mathbf{u}_i\mathbf{v}_j) \geq H(UV)$

- acceptance region:

$\mathcal{A}_n =$ the union of rectangular cells specified by $\mathbf{u}_i\mathbf{v}_j$ such that $\mathbf{u}_i\mathbf{v}_j \in T_\epsilon^n(UV)$

$\mathcal{F}_n : H(\mathbf{u}_i\mathbf{v}_j) \leq H(UV)$

Coding--Data compression

- Classify N_1 $\mathbf{u}_i$ $'s$ into $\exp[nR_1]$ bins
- Classify N_2 $\mathbf{v}_j$ $'s$ into $\exp[nR_2]$ bins

- Letting $\mathbf{u}_i$ correspond to $\mathbf{x}$, define the encoder 1:

$$\varphi_1(\mathbf{x}) = \text{the bin number } I(\mathbf{u}_i) \text{ of } \mathbf{u}_i$$

- Letting $\mathbf{v}_j$ correspond to $\mathbf{y}$, define the encoder 2:

$$\varphi_2(\mathbf{y}) = \text{the bin number } J(\mathbf{v}_j) \text{ of } \mathbf{v}_j$$

Minimum entropy decoding

- On receiving the bin numbers $I, J,$ let the decoder be defined by

$$\begin{aligned}\psi_n(I,J) \quad &= \quad \text{the rectangular cell specified by} \\ &\quad \tilde{\mathbf{u}}\tilde{\mathbf{v}} \text{ such that} H(\tilde{\mathbf{u}}\tilde{\mathbf{v}}) = \min_{i,j} H(\mathbf{u}_i\mathbf{v}_j) \\ &\quad \text{with } I = I(\mathbf{u}_i) \text{ and } J = J(\mathbf{v}_j)\end{aligned}$$

Error probabilities for minimum entropy coding

1) error probablity that an element of $\mathcal{D}_n$ is decoded as that of the acceptance region $\mathcal{A}_n$: $\mu_n \equiv \Pr\{\mathcal{D}_n \to \mathcal{A}_n\}$

2) error probability that an element of the acceptance region $\mathcal{A}_n$ is decoded as that of $\mathcal{F}_n$: $\lambda_n \equiv \Pr\{\mathcal{A}_n \to \mathcal{F}_n\}$

- In both cases the error probabilities are evaluated as $\mu_n \approx \lambda_n \approx \exp[-nE_b]$ where $E_b \geq 0$ is specified later.

Error probabilities for hypothesis testing under data compression

- type 1 error probability:

$$\begin{aligned} 1-\alpha_n &= (1-\lambda_n)\,\Pr\{X^nY^n \in \mathcal{A}_n\} + \mu_n \Pr\{X^nY^n \in \mathcal{D}_n\} \\ &\gtrsim (1-\exp[-nE_b])\,\Pr\{X^nY^n \in \mathcal{A}_n\} \\ &\to 1 \ (n \to \infty) \quad \text{if } E_b > 0 \end{aligned}$$

thus, $\alpha_n \to 0 \ (\leq \varepsilon)$ where we have noticed that $\Pr\{X^nY^n \in \mathcal{A}_n\} \to 1,$ because $\mathcal{A}_n$ contains the set of typical sequences $T_\epsilon^n(XY)$

- type 2 error probability:

$$\begin{aligned} \beta_n &= (1-\lambda_n)\,\Pr\{\overline{X}^n\overline{Y}^n \in \mathcal{A}_n\} + \mu_n \Pr\{\overline{X}^n\overline{Y}^n \in \mathcal{D}_n\} \\ &\lesssim \Pr\{\overline{X}^n\overline{Y}^n \in \mathcal{A}_n\} + \exp[-nE_c]\,\Pr\{\overline{X}^n\overline{Y}^n \in \mathcal{D}_n\} \end{aligned}$$

⇓

Thus, it suffices to evaluate $\Pr\{\overline{X}^n\overline{Y}^n \in \mathcal{A}_n\}$ and $\Pr\{\overline{X}^n\overline{Y}^n \in \mathcal{D}_n\}$

- Calculation shows that

$$\begin{aligned} \Pr\{\overline{X}^n\overline{Y}^n \in \mathcal{A}_n\} &\approx \exp[-n\sigma_1] \\ \Pr\{\overline{X}^n\overline{Y}^n \in \mathcal{D}_n\} &\approx \exp[-n\sigma_2] \end{aligned}$$

where

$$\begin{aligned} \sigma_1 &= \min_{\tilde{U}\tilde{X}\tilde{Y}\tilde{V} \in \mathcal{B}_1} D(\tilde{U}\tilde{X}\tilde{Y}\tilde{V} \| \overline{UXYV}) \\ \sigma_2 &= \min_{\tilde{U}\tilde{X}\tilde{Y}\tilde{V} \in \mathcal{B}_2} D(\tilde{U}\tilde{X}\tilde{Y}\tilde{V} \| \overline{UXYV}); \end{aligned}$$

divergence

$$\begin{aligned} \mathcal{B}_1 &= \{\tilde{U}\tilde{X}\tilde{Y}\tilde{V} : P_{\tilde{U}\tilde{X}} = P_{UX}, P_{\tilde{V}\tilde{Y}} = P_{VY}, P_{\tilde{U}\tilde{V}} = P_{UV}\} \\ \mathcal{B}_2 &= \{\tilde{U}\tilde{X}\tilde{Y}\tilde{V} : P_{\tilde{U}\tilde{X}} = P_{UX}, P_{\tilde{V}\tilde{Y}} = P_{VY}, H(\tilde{U}\tilde{V}) \geq H(UV)\} \end{aligned}$$

- Clearly, $\mathcal{B}_1 \subset \mathcal{B}_2$ and hence $\sigma_1 \geq \sigma_2$

- Furthermore, let with $[x]^+ = \max(x, 0)$:

$$E_1 = \begin{cases} +\infty & \text{if } R_1 \geq I(U;X), \\ [R_1 - I(U;X|V)]^+ & \text{if } R_1 < I(U;X); \end{cases}$$

$$E_2 = \begin{cases} +\infty & \text{if } R_2 \geq I(V;Y), \\ [R_2 - I(V;Y|U)]^+ & \text{if } R_2 < I(V;Y); \end{cases}$$

$$E_{12} = \begin{cases} +\infty & \text{if } R_1 \geq I(U;V), R_2 \geq I(V;Y); \\ [R_1 + R_2 - I(UV;XY)]^+ & \text{otherwise} \end{cases}$$

then, we have

$$E_b = \min(E_1, E_2, E_{12})$$

- Summarizing, set

$$E_{SHA}(R_1, R_2) = \max_{\substack{P_{U|X}, P_{V|Y} : \\ R_1 \geq I(U; X|V) \\ R_2 \geq I(V; Y|U) \\ R_1 + R_2 \geq I(UV; XY)}} \min(\sigma_1, \sigma_2 + E_b)$$

power exponent
with coding error

power exponent
without coding error

- then,

Theorem 1 (lower bound):

$$\theta(R_1, R_2, \varepsilon) \geq E_{SHA}(R_1, R_2)$$

Corollary 1 (Stein's lemma: without coding error)

Suppose that $R_1 \geq H(X), \;\; R_2 \geq H(Y)$ holds,

then,

$$\theta(R_1, R_2, \varepsilon) = E_{SHA}(R_1, R_2) = D(XY||\overline{XY}).$$

Remark 1: This coincides with that of Han (1987, Corollary 5).

Proof: Put $U = X, V = Y$ in Theorem 1.

Corollary 2 (Stein's lemma: with coding error):

Suppose that

$$\begin{aligned} R_1 &\geq D + H(X|Y), \\ R_2 &\geq D + H(Y|X), \\ R_1 + R_2 &\geq D + H(XY), \\ D &\equiv D(XY||\overline{XY}) - \min_{\tilde{X}\tilde{Y} \in \mathcal{B}_3} D(\tilde{X}\tilde{Y}||\overline{XY}), \\ \mathcal{B}_3 &\equiv \{\tilde{X}\tilde{Y} : P_{\tilde{X}} = P_X, P_{\tilde{Y}} = P_Y, H(\tilde{X}\tilde{Y}) \geq H(XY)\}. \end{aligned}$$

then, we have

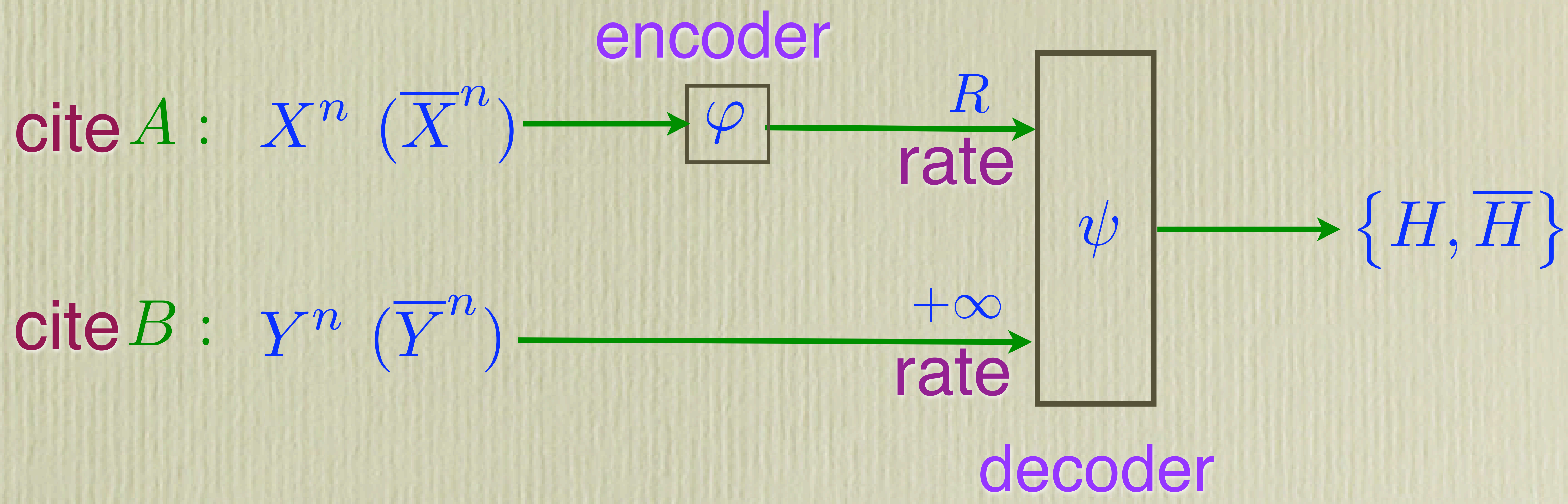
$$\theta(R_1, R_2, \varepsilon) = E_{SHA}(R_1, R_2) = D(XY||\overline{XY}).$$

reduces to Corollary 1

Remark: The case with $D \geq I(X;Y)$, error coding not help, but the case with $D < I(X;Y)$, error coding helps.

Proof: we can set $U = X, V = Y$ in Theorem 1

IV. Full side information system



encoder

cite A : X^n $(\overline{X}^n)$ φ

R

rate

ψ $\{H, \overline{H}\}$

cite B : Y^n $(\overline{Y}^n)$

$+\infty$

rate

decoder

- encoder $\varphi : \mathcal{X}^n \to \mathcal{M} \equiv \{1, 2, \cdots, M\}$
- rate constraint $\frac{1}{n} \log M \leq R,$
- decoder $\psi : \mathcal{M} \times \mathcal{Y}^n \to \{H, \overline{H}\}$

- In this full side information case, we define $\tau_1, \tau_2, \mathcal{C}_1, \mathcal{C}_2$ instead of $\sigma_1, \sigma_2, \mathcal{B}_1, \mathcal{B}_2$ as below (set $V = Y$ in all the previous formulas):

$$\begin{aligned}
\tau_1 &= \min_{\tilde{U}\tilde{X}\tilde{Y} \in \mathcal{C}_1} D(\tilde{U}\tilde{X}\tilde{Y} \| \overline{UXY}) \\
\tau_2 &= \min_{\tilde{U}\tilde{X}\tilde{Y} \in \mathcal{C}_2} D(\tilde{U}\tilde{X}\tilde{Y} \| \overline{UXY}); \\
\mathcal{C}_1 &= \{\tilde{U}\tilde{X}\tilde{Y} : P_{\tilde{U}\tilde{X}} = P_{UX}, P_{\tilde{U}\tilde{Y}} = P_{UY}\}, \\
\mathcal{C}_2 &= \{\tilde{U}\tilde{X}\tilde{Y} : P_{\tilde{U}\tilde{X}} = P_{UX}, P_{\tilde{Y}} = P_Y, H(\tilde{U}\tilde{Y}) \geq H(UY)\}
\end{aligned}$$

- Clearly, $\mathcal{C}_1 \subset \mathcal{C}_2$ and hence $\tau_1 \geq \tau_2$, moreover

$$E_c = \begin{cases} +\infty & \text{if } R \geq I(U : X), \\ [R - I(U; X|Y)]^+ & \text{if } R < I(U; X); \end{cases}$$

- Furthermore, define

$$\begin{aligned} E_{SHA}(R) &\equiv E_{SHA}(R,+\infty) \\ &= \max_{\substack{P_{U|X}: \\ R \geq I(U;X|Y)}} \min(\tau_1, \tau_2 + E_c) \end{aligned}$$

power exponent
without coding error

power exponent
with coding error

then, we have

Theorem 2 (full side lower bound):

$$\theta(R,\varepsilon) \geq E_{SHA}(R)$$

where $\theta(R,\varepsilon) \equiv \theta(R,+\infty,\varepsilon)$

Simokawa-Han-Amari (1994)

- Remark 2:

1) The value of $\tau_1 \quad = \quad \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_1} D(\tilde{U}\tilde{X}\tilde{Y}||\overline{UXY})$ remains invariant if we impose the cardinality constraint $||U|| \leq |\mathcal{X}| + 1$ in $\mathcal{C}_1 \quad = \quad \{\tilde{U}\tilde{X}\tilde{Y} : P_{\tilde{U}\tilde{X}} = P_{UX}, P_{\tilde{U}\tilde{Y}} = P_{UY}\}$

2) The value of $\tau_2 \quad = \quad \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_2} D(\tilde{U}\tilde{X}\tilde{Y}||\overline{UXY})$ remains invariant if we impose the cardinality constraint $||U|| \leq |\mathcal{X}| + |\mathcal{Y}|$ in $\mathcal{C}_2 \quad = \quad \{\tilde{U}\tilde{X}\tilde{Y} : P_{\tilde{U}\tilde{X}} = P_{UX}, P_{\tilde{Y}} = P_Y, H(\tilde{U}\tilde{Y}) \geq H(UY)\}$

Test against independence: tight bound

- Consider the case with $P_{\overline{XY}} = P_X P_Y$, then

$$\begin{aligned}
\tau_1 &= \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_1} D(\tilde{U}\tilde{X}\tilde{Y}||\overline{UXY}) \\
&\geq \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_1} D(\tilde{U}\tilde{Y}||\overline{UY}) \\
&= \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_1} D(UY||\overline{UY}) \\
&= D(UY||\overline{UY}) \\
&= D(P_{UY}||P_U P_Y) \\
&= I(U;Y).
\end{aligned}$$

optimality is due to Ahlswede-Csiszar (1986)

Thus,

Theorem 3:

$$\theta(R,\varepsilon) = E_{SHA}(R) = \max_{P_{U|X}:I(U;X)\leq R} I(U;Y).$$

$\Longleftarrow$ error coding does not help

Test for independence: lower bounds

- Consider the case with $P_{XY} = P_X P_Y$ then

$$
\begin{aligned}
\tau_1 &= \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_1} D(\tilde{U}\tilde{X}\tilde{Y}\|\overline{UXY}) & \tau_2 &= \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_2} D(\tilde{U}\tilde{X}\tilde{Y}\|\overline{UXY}) \\
&\geq \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_1} D(\tilde{U}\tilde{Y}\|\overline{UY}) & &\geq \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_2} D(\tilde{U}\tilde{Y}\|\overline{UY}) \\
&= \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_1} D(UY\|\overline{UY}) & &= \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_2} D(UY\|\overline{UY}) \\
&= D(UY\|\overline{UY}) & &= D(UY\|\overline{UY})
\end{aligned}
$$

$H(\tilde{U}\tilde{Y}) \geq H(UY)$
$= H(U) + H(Y)$

- Thus, we have

Theorem 4 :

$$\theta(R,\varepsilon) \geq \max_{P_{U|X}:I(U;X)\leq R} D(P_U P_Y \| P_{\overline{UY}})$$

⇐ error coding does not help

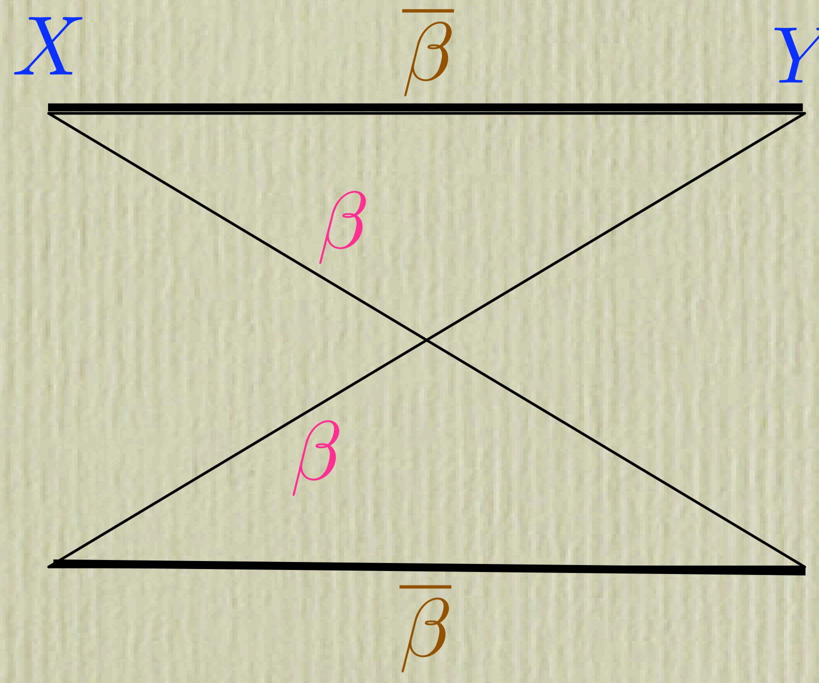
Conjecture:

If $P_{XY} = P_{\overline{X}} P_{\overline{Y}}$, then

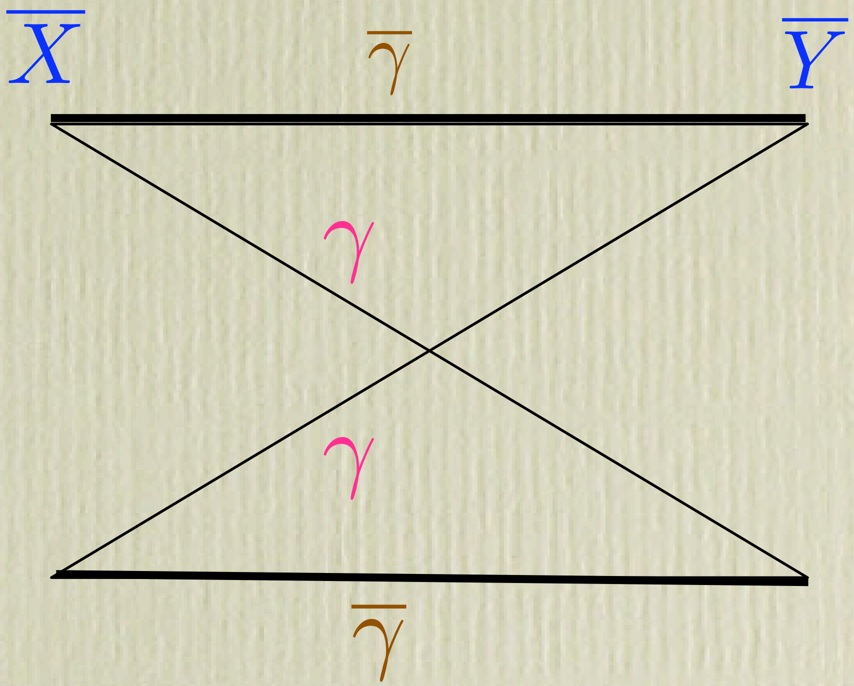
$$\theta(R, \varepsilon) = \max_{P_{U|X} : I(U;X) \leq R} D(P_{\overline{U}} P_{\overline{Y}} || P_{\overline{UY}})$$

V. Test for Binary Symmetric Sources: lower bounds

null hypothesis



alternative hypothesis



$0 < \gamma < \beta < 1/2$ (crossover probabilities)

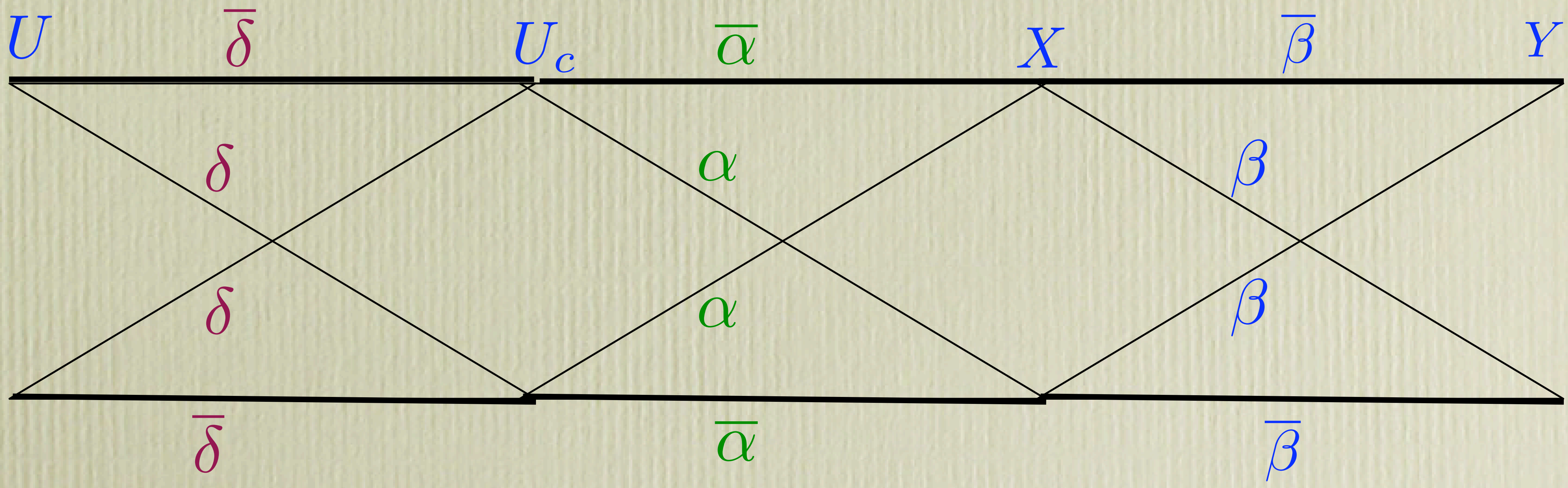
$\overline{x} \equiv 1 - x; \quad x * y \equiv x\overline{y} + \overline{x}y$

Concatination of BSS's: auxiliary random variables

Null hypothesis

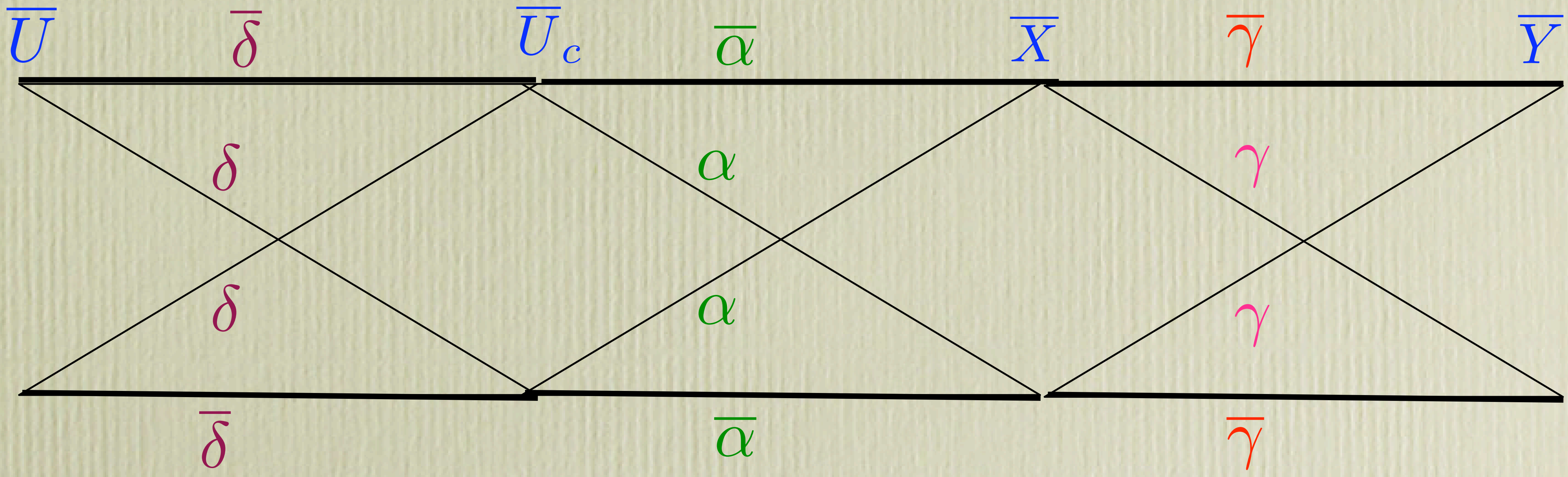
U, U_c :specified later

$U \to U_c \to X \to Y$: Markov chain:



Alternative hypothesis

$$\overline{U} \to \overline{U}_c \to \overline{X} \to \overline{Y} \quad (P_{\overline{X}|\overline{U}_c} = P_{X|U_c})$$



- Given a rate R, specify U and U_c by

$$R = I(U;X) = 1 - h(\delta * \alpha),$$
$$R = I(U_c;X|Y) = h(\alpha * \beta) - h(\alpha)$$

binary entropy function

- Thus, $\delta = \delta_0 < 1/2$ is specified by

$$h(\delta * \alpha) = 1 - h(\alpha * \beta) + h(\alpha)$$

$\delta_0 = 0$ if $\beta = 1/2$

- Equation $R = I(U;X)\ (\delta = \delta_0)$ specifies coding without error $\rightarrow \tau_1|_{\delta=\delta_0}$

- Equation $R = I(U_c;X|Y)\ (\delta = 0)$ specifies coding with error $\rightarrow \tau_2|_{\delta=0}$

- Then, the lower bounds are given as:

$$\begin{aligned}
\tau_1|_{\delta=\delta_0} &= \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_1} D(\tilde{U}\tilde{X}\tilde{Y}||\overline{UXY}) \\
&\geq \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_1} D(\tilde{U}\tilde{Y}||\overline{UY}) \\
&= \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{C}_1} D(UY||\overline{UY}) \\
&= D(UY||\overline{UY})|_{\delta=\delta_0}
\end{aligned}$$

$$\begin{aligned}
\tau_2|_{\delta=0} &= \min_{\tilde{U}_c\tilde{X}\tilde{Y}\in\mathcal{C}_2} D(\tilde{U}_c\tilde{X}\tilde{Y}||\overline{U}_c\overline{XY}) \\
&\geq \min_{\tilde{U}_c\tilde{X}\tilde{Y}\in\mathcal{C}_2} D(\tilde{U}_c\tilde{Y}||\overline{U}_c\overline{Y}) \\
&= D(U_cY||\overline{U}_c\overline{Y}) \\
&= D(UY||\overline{UY})|_{\delta=0}
\end{aligned}$$

$H(\tilde{U}_c\tilde{Y}) \geq H(U_cY) > H(\overline{U}_c\overline{Y})$ hence,

min is attained at $\tilde{U}_c\tilde{Y} = U_cY$ (BSS)

- Evaluation of the lower bounds:

$$\begin{aligned} g(\delta) &\equiv D(UY\|\overline{UY}) \\ &= -1 - h(\delta * \alpha * \beta) \\ &\quad -(1-\delta * \alpha * \beta)\log(1-\delta * \alpha * \gamma) \\ &\quad -\delta * \alpha * \beta \log(\delta * \alpha * \gamma). \end{aligned}$$

- then, $$\begin{aligned} g'(\delta)|_{\beta=1/2} &< 0 \quad (0 \le \delta < 1/2) \\ &= 0 \quad (\delta = 1/2) \end{aligned}$$

$$\Downarrow$$

There exists a $\gamma < \beta_0 = 1/2 - \epsilon_0$ (ϵ_0 is small enough)

such that

$$g'(\delta)|_{\beta=\beta_0} < 0 \quad (0 \le \delta \le \delta_0 < 1/2)$$

$$\Downarrow$$

$$g(\delta)|_{\beta=\beta_0} \text{ is dereasing in } \delta \quad (0 \le \delta \le \delta_0)$$

$$\Downarrow$$ with $\beta = \beta_0$

$$\begin{aligned} D(U_c Y || \overline{U}_c \overline{Y}) &= D(UY || \overline{UY})|_{\delta=0} \\ &> D(UY || \overline{UY})|_{\delta=\delta_0} \end{aligned}$$

error coding

no error coding

$$\Downarrow$$

Conclusion: Error coding helps!