# Universal Decoding for Source–Channel Coding With Side Information

Neri Merhav

Department of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 3200004, Israel

ISIT 2016, July 10–15, 2016.

# Some Related Earlier Work

Universal channel decoding

Goppa (1975); Csiszár & Körner (1981); Csiszár (1982); Ziv (1985); Merhav (1993); Liu & Hughes (1996); Feder & Lapidoth (1998); Lapidoth & Ziv (1998); Feder & Merhav (2002); Lomnitz & Feder (2012); Merhav (2013); ...

Universal source decoding (Slepian–Wolf)

Csiszár & Körner (1981); Oohama & Han (1994); Draper (2004); Chen, He, Jagmohan & Lastras–Montaño (2008); Sarvotham, Baron & Baraniuk (2005); ...

Universal joint source–channel decoding

Csiszár (1980).

# Background: Csiszár's Universal JSC Decoder

A DMS $\boldsymbol{u} \in \mathcal{U}^n$ is directly mapped into a channel input vector $\boldsymbol{x}[\boldsymbol{u}] \in \mathcal{X}^n$ and fed into a DMS whose output $\boldsymbol{y}$ is used to decode $\hat{\boldsymbol{u}}$.

The error exponent is upper bounded by

$$E^{\mathsf{jsc}} \leq \min_R \left[ E^{\mathsf{s}}(R) + E^{\mathsf{c}}(R) \right],$$

where

$E^{\mathsf{s}}(R) =$ the source coding exponent and

$E^{\mathsf{c}}(R) =$ the channel reliability function.

The error exponent is lower bounded by

$$E^{\mathsf{jsc}} \geq \min_R \left[ E^{\mathsf{s}}(R) + E^{\mathsf{c}}_{\mathsf{r}}(R) \right],$$

where $E^{\mathsf{c}}_{\mathsf{r}}(R) =$ the random coding exponent.

# Background: Csiszár's Universal JSC Decoder (Cont'd)

Some comments:

- An equivalent lower bound

$$E^{\text{jsc}} \geq \max_{Q} \min_{P',W'} \{D(P'\|P) + D(W'\|W|Q) + [I(X;Y') - H(U')]_+\},$$

  where: $P =$ source, $Q =$ random coding distribution, and $W =$ channel.

- Idea of upper bound: Each source type $P'$ induces a sub–code at rate $R = H(P')$. Counting errors only within the worst sub–code.
- The bounds coincide when $R^*$ of the former is above $R_{\text{crit}}$.

# Background: Csiszár's Universal JSC Decoder (Cont'd)

Comments (continued):

- $\Rightarrow$ for $R^* \geq R_{\text{crit}}$, errors within sub–codes dominate.
- Universal decoder – "generalized MMI": asymptotically equivalent to

$$\hat{\boldsymbol{u}} = \arg\max_{\boldsymbol{u}} [\hat{I}_{\boldsymbol{x}[\boldsymbol{u}]\boldsymbol{y}}(X;Y) - \hat{H}_{\boldsymbol{u}}(U)].$$

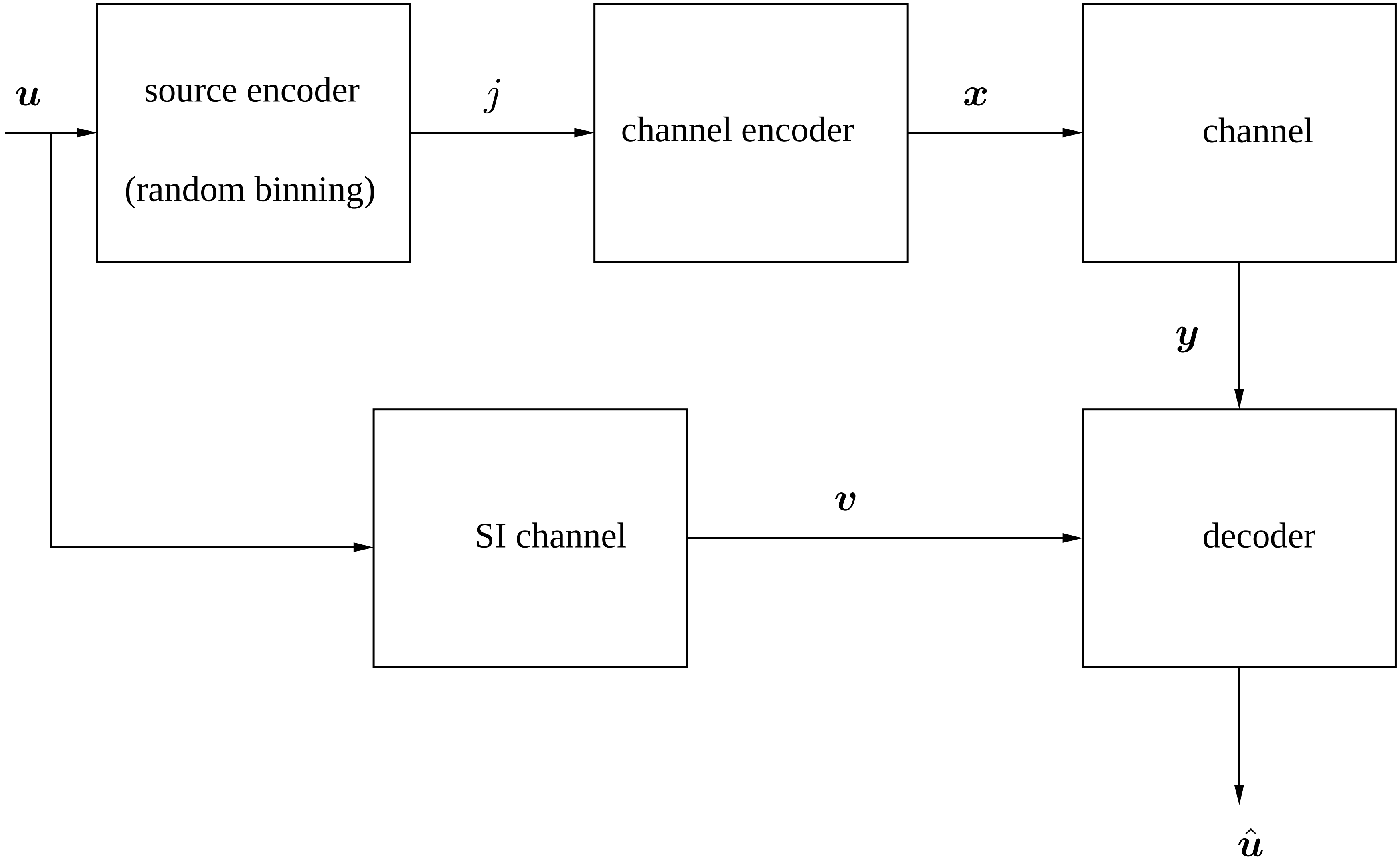
- GMMI is proved optimal for $R \geq R_{\text{crit}}$.
- Everything generalizes easily to bandwidth expansion factor $\neq 1$.

# Our Settings

We extend Csiszár's setting in several directions:

- Setting 1:
  - Availability of source side info @ decoder.
  - Slepian–Wolf source coding and channel coding.
- Setting 2: separate encodings + joint decoding of correlated sources.
- Further extensions:
  - FS sources/channels – universal decoding based on LZ coding.
  - Arbitrary sources/channels – universality w.r.t. classes of decoders.

# Setting 1



$\boldsymbol{u}$
source encoder
(random binning)
$j$
channel encoder
$\boldsymbol{x}$
channel
$\boldsymbol{y}$
SI channel
$\boldsymbol{v}$
decoder
$\hat{\boldsymbol{u}}$

# Motivations

- Separate S–W source coding and channel coding:
  - A general framework – includes JSC as a special case.
  - May be dictated by system constraints: different units/locations, modularity, etc.
- A common framework for many important special cases:
  - Separate source– and channel coding without SI: $\boldsymbol{v}$ – degenerate.
  - Pure SW source coding: clean, large–alphabet channel $\boldsymbol{y} \equiv \boldsymbol{x}$.
  - Pure channel coding: source = BSS, $\boldsymbol{v}$ – degenerate.
  - Joint source–channel coding with/out SI: high binning rate.
  - Systematic coding.

# Formulation

We are given:

- A memoryless source pair $\{(U_i, V_i)\}$, with $(U_i, V_i) \sim P$:
  - $\boldsymbol{u} = (u_1, \ldots, u_n) =$ source to be communicated.
  - $\boldsymbol{v} = (v_1, \ldots, v_n) =$ side information @ decoder.
- A memoryless channel $W(\boldsymbol{y}|\boldsymbol{x}) = \prod_t W(y_t|x_t)$.
- Bandwidth expansion factor $= 1$ without essential loss of generality.

# Formulation (Cont'd)

Communication system:

- $\boldsymbol{u}$ is mapped into a bin $j = f(\boldsymbol{u})$ at rate $R$ (selected at random).
- The bin is channel–coded into $\boldsymbol{x}[\boldsymbol{u}] = \boldsymbol{x}[f(\boldsymbol{u})]$ (random coding $\sim \mathcal{T}(Q)$).
- The message $\boldsymbol{u}$ is estimated by $\hat{\boldsymbol{u}} = g(\boldsymbol{y})$.
- The optimal MAP decoder

$$\hat{\boldsymbol{u}} = \text{arg} \max_{\boldsymbol{u}} P_{UV}(\boldsymbol{u}, \boldsymbol{v}) W(\boldsymbol{y}|\boldsymbol{x}[\boldsymbol{u}]).$$

Our goals are:

- Find the random–coding exponent of the MAP decoder, $E(R, Q)$.
- Find a universal decoder (ignorant of $P$ and $W$) that achieves $E(R, Q)$.

# Basic Result

Theorem 1

1. The random–coding error exponent of the MAP decoder is given by (see also Chang 2011):

$$E(R,Q)=\min_{P_{U'V'},W'}\{D(P_{U'V'}\|P_{UV})+D(W'\|W|Q)+[R\wedge I(X;Y')-H(U'|V')]_+\}.$$

2. The universal decoders,

$$\hat{\boldsymbol{u}}=\arg\max_{\boldsymbol{u}}[\hat{I}_{\boldsymbol{x}[\boldsymbol{u}]\boldsymbol{y}}(X;Y)-\hat{H}_{\boldsymbol{u}\boldsymbol{v}}(U|V)]$$

and

$$\hat{\boldsymbol{u}}=\arg\max_{\boldsymbol{u}}[R\wedge\hat{I}_{\boldsymbol{x}[\boldsymbol{u}]\boldsymbol{y}}(X;Y)-\hat{H}_{\boldsymbol{u}\boldsymbol{v}}(U|V)]$$

both achieve $E(R,Q)$.

# Discussion

$$E(R,Q) = \min_{P_{U'V'},W'} \{D(P_{U'V'}\|P_{UV}) + D(W'\|W|Q) + [R \wedge I(X;Y') - H(U'|V')]_+\}$$

- $E(R,Q)$ is monotonically non–decreasing, which is not trivial.
- When $R$ is large enough, there is saturation – equivalence to JSC.
- For a clean, large alphabet channel, we recover the SW error exponent.

$$\hat{\boldsymbol{u}} = \text{arg}\max_{\boldsymbol{u}} [\hat{I}_{\boldsymbol{x}[\boldsymbol{u}]\boldsymbol{y}}(X;Y) - \hat{H}_{\boldsymbol{u}\boldsymbol{v}}(U|V)]$$

$$\hat{\boldsymbol{u}} = \text{arg}\max_{\boldsymbol{u}} [R \wedge \hat{I}_{\boldsymbol{x}[\boldsymbol{u}]\boldsymbol{y}}(X;Y) - \hat{H}_{\boldsymbol{u}\boldsymbol{v}}(U|V)]$$

- Decoder 1 is a natural extension of Csiszár's GMMI decoder.
- Decoder 2 has no apparent advantage, but later results will be related.

# A Word About the Analysis

- Exponentially tight analysis based on new techniques.
- Averaging for both random binning and random coding (one at a time).
- Lower bounding $\overline{P_e}$ for the MAP decoder.
- Upper bounding $\overline{P_e}$ for the universal decoders.
- Distinguishing between errors within the same bin and errors across bins.
- Error events are dominated by pairwise errors within the types.

# Extension – Finite–State Sources and Channels

Let the source be given by

$$P(\boldsymbol{u}, \boldsymbol{v}) = \prod_{t=1}^{n} P(u_t, v_t | s_t), \quad s_t = g(s_{t-1}, u_{t-1}, v_{t-1}).$$

Similarly, let the channel be given by

$$W(\boldsymbol{y} | \boldsymbol{x}) = \prod_{t=1}^{n} W(y_t | x_t, z_t), \qquad z_t = h(z_{t-1}, x_{t-1}, y_{t-1}),$$

We have some tecnhical assumptions on the random coding distribution $Q(\boldsymbol{x})$.

# Extension (Cont'd)

We use the notion of conditional LZ compressibility [Ziv 1985]: Given $(\boldsymbol{x}, \boldsymbol{y}) = [(x_1, y_1), \ldots, (x_n, y_n)]$, apply LZ parsing to this sequence pair. Let

- $c(\boldsymbol{x}, \boldsymbol{y}) =$ number of phrases.
- $c(\boldsymbol{y}) =$ number of distinct phrases of $\boldsymbol{y}$.
- $\boldsymbol{y}(l) =$ the $l$th distinct $\boldsymbol{y}$–phrase, $l = 1, 2, ..., c(\boldsymbol{y})$.
- $c_l(\boldsymbol{x}|\boldsymbol{y}) =$ number of $\boldsymbol{y}(l)$ in parsing of $\boldsymbol{y}$.

$$\hat{H}_{LZ}(\boldsymbol{x}|\boldsymbol{y}) = \frac{1}{n} \sum_{l=1}^{c(\boldsymbol{y})} c_l(\boldsymbol{x}|\boldsymbol{y}) \log c_l(\boldsymbol{x}|\boldsymbol{y}).$$

For example, $n = 6$ and

$$\begin{aligned} \boldsymbol{x} &= 0 \,|\, 1 \,|\, 0\,0 \,|\, 0\,1| \\ \boldsymbol{y} &= 0 \,|\, 1 \,|\, 0\,1 \,|\, 0\,1| \end{aligned}$$

then

$$c(\boldsymbol{x}, \boldsymbol{y}) = 4, \;\; c(\boldsymbol{y}) = 3, \;\; \boldsymbol{y}(1) = 0, \; \boldsymbol{y}(2) = 1, \; \boldsymbol{y}(3) = 01,$$

$$c_1(\boldsymbol{x}|\boldsymbol{y}) = c_2(\boldsymbol{x}|\boldsymbol{y}) = 1, \;\; c_3(\boldsymbol{x}|\boldsymbol{y}) = 2.$$

# Extension (Cont'd)

We also define

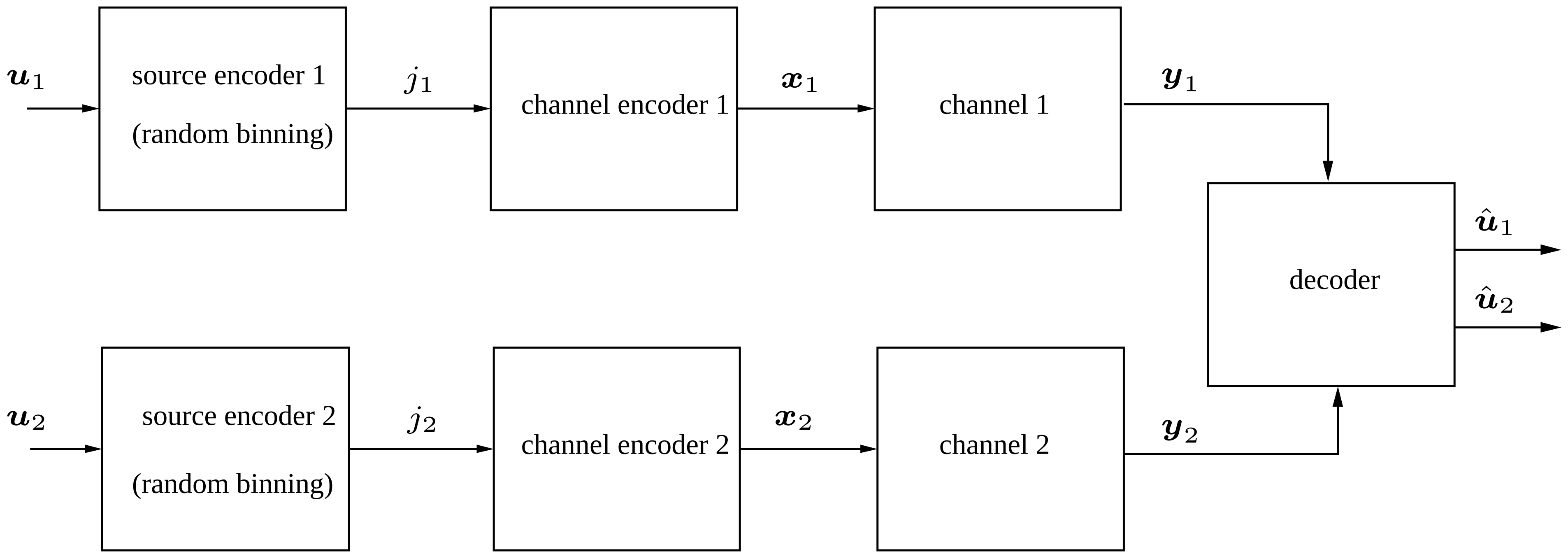
$$\hat{I}_{\mathsf{LZ}}(\boldsymbol{x};\boldsymbol{y}) = -\frac{\log Q(\boldsymbol{x})}{n} - \hat{H}_{\mathsf{LZ}}(\boldsymbol{x}|\boldsymbol{y}),$$

and finally, the universal decoding metric

$$\tilde{\boldsymbol{u}} = \text{arg}\max_{\boldsymbol{u}} \left[\hat{I}_{\mathsf{LZ}}(\boldsymbol{x}[\boldsymbol{u}];\boldsymbol{y}) - \hat{H}_{\mathsf{LZ}}(\boldsymbol{u}|\boldsymbol{v})\right].$$

The extension of Theorem 1 to FS sources/channels asserts that the above universal decoder achieves the same random–coding error exponent as the MAP decoder, in the spirit of [Ziv 1985].

# Setting 2




The optimal (MAP) decoder:

$$(\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2) = \text{arg} \max_{\boldsymbol{u}_1, \boldsymbol{u}_2} P(\boldsymbol{u}_1, \boldsymbol{u}_2) W_1(\boldsymbol{y}_1 | \boldsymbol{x}_1[\boldsymbol{u}_1]) W_2(\boldsymbol{y}_2 | \boldsymbol{x}_2[\boldsymbol{u}_2]).$$

# Setting 2 (Cont'd)

Here, five types of pairwise errors should be handled differently:

- $\boldsymbol{u}_1' \neq \boldsymbol{u}_1$ and $\boldsymbol{u}_2' = \boldsymbol{u}_2$.
- $\boldsymbol{u}_2' \neq \boldsymbol{u}_2$ and $\boldsymbol{u}_1' = \boldsymbol{u}_1$.
- $\boldsymbol{u}_1' \neq \boldsymbol{u}_1$ and $\boldsymbol{u}_2' \neq \boldsymbol{u}_2$, but $\boldsymbol{u}_2'$ is in the bin of $\boldsymbol{u}_2$.
- $\boldsymbol{u}_1' \neq \boldsymbol{u}_1$ and $\boldsymbol{u}_2' \neq \boldsymbol{u}_2$, but $\boldsymbol{u}_1'$ is in the bin of $\boldsymbol{u}_1$.
- $\boldsymbol{u}_1' \neq \boldsymbol{u}_1$, $\boldsymbol{u}_2' \neq \boldsymbol{u}_2$, and neither $\boldsymbol{u}_1'$ nor $\boldsymbol{u}_2'$ are in the correct bins.

# Setting 2 (Cont'd)

Each type of error "demands" a different universal decoding metric:

- Errors of types 1 and 2 are handled the same as before:

$$\begin{aligned} f_1(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2) &= R_1 \wedge \hat{I}(X_1; Y_1) - \hat{H}(U_1|U_2) \\ f_2(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2) &= R_2 \wedge \hat{I}(X_2; Y_2) - \hat{H}(U_2|U_1). \end{aligned}$$

- Errors of types 3 and 4 "would like to be handled" by:

$$\begin{aligned} f_3(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2) &= R_1 \wedge \hat{I}(X_1; Y_1) + R_2 - \hat{H}(U_1, U_2) \\ f_4(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2) &= R_2 \wedge \hat{I}(X_2; Y_2) + R_1 - \hat{H}(U_1, U_2). \end{aligned}$$

- Error of type 5:

$$\begin{aligned} & f_5(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2) \\ = \; & [R_1 \wedge \hat{I}(X_1; Y_1) + R_2 \wedge \hat{I}(X_2; Y_2) - \hat{H}(U_1, U_2)]_+ + \\ & [\hat{I}(X_1; Y_1) - R_1]_+ + [\hat{I}(X_2; Y_2) - R_2]_+. \end{aligned}$$

# Setting 2 (Cont'd)

But we need a single decoding metric and we have to confront all five types of errors at the same time!

Q: How can we integrate all these decoding metrics into one metric?

# Setting 2 (Cont'd)

But we need a single decoding metric and we have to confront all five types of errors at the same time!

Q: How can we integrate all these decoding metrics into one metric?

A: It turns out that this is achieved by taking the minimum of them.
Define

$$f_0(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2) = \min_{1 \le i \le 5} f_i(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_1, \boldsymbol{y}_2).$$

Then, the universal decoder

$$(\tilde{\boldsymbol{u}}_1, \tilde{\boldsymbol{u}}_2) = \text{arg} \max_{\boldsymbol{u}_1, \boldsymbol{u}_2} f_0(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{x}_1[\boldsymbol{u}_1], \boldsymbol{x}_2[\boldsymbol{u}_2], \boldsymbol{y}_1, \boldsymbol{y}_2)$$

achieves the same error exponent as the MAP decoder for Setting 2.

# Summary

- We extended Csiszár's universal JSC decoder in various directions:
  - Availability of decoder side information.
  - Separate source binning and channel coding.
  - Finite–state sources and channels.
- The extension to FS sources/channels can be applied also to Setting 2.
- Another extension: arbitrary sources/channels + given family of decoders.