Mismatched Decoding and Bit-Interleaved Coded Modulation

Albert Guillén i Fàbregas
ICREA & UPF
Cambridge

2015 Munich Workshop on Coding and Modulation
Munich, 30 July 2015

Joint work with

- Jonathan Scarlett (EPFL)
- Alfonso Martinez (UPF)
- Li Peng (Mathworks)
- Alex Alvarado (UCL)

Outline

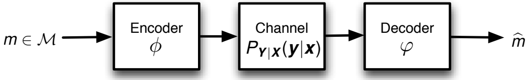
Setup

Rates

Error Exponents

BICM

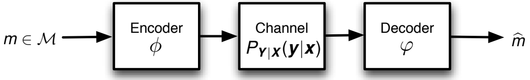
Setup



- Code $\mathcal{C}(n, |\mathcal{M}|) = \{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(|\mathcal{M}|)}\}$
  - codewords $\boldsymbol{x}^{(i)} \in \mathcal{X}^n$
  - rate $R = \frac{1}{n} \log |\mathcal{M}|$
- Discrete memoryless channel

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}^{(i)}) = \prod_{k=1}^{n} P_{Y|X}(y_k|x_k^{(i)})$$

Setup



- Error probability

$$P_e(n, |\mathcal{M}|) = \mathbb{P}\{\widehat{M} \neq M\}$$

- Maximum-Likelihood decoding

$$\widehat{m} = \underset{i=1,\ldots,|\mathcal{M}|}{\arg\max}\ P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}^{(i)}) = \underset{i=1,\ldots,|\mathcal{M}|}{\arg\max} \prod_{k=1}^{n} P_{Y|X}(y_k|x_k^{(i)})$$

Setup

- Capacity

$$C = \max_{P_X} I(X;Y)$$

- Random-coding error exponent for distribution Q

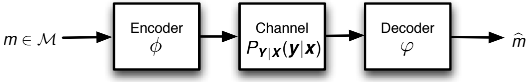
$$E_r(R, Q) = \max_{\rho \in [0,1]} E_0(\rho, Q) - \rho R$$

$$E_0(\rho, Q) = -\log \sum_{x,y} Q(x) P_{Y|X}(y|x) \left(\frac{\sum_{\bar{x}} Q(\bar{x}) P_{Y|X}(y|\bar{x})^{\frac{1}{1+\rho}}}{P_{Y|X}(y|x)^{\frac{1}{1+\rho}}} \right)^{\rho}$$

C. Shannon, "A mathematical theory of communication," Bell Syst. Tech. Journal, vol. 27, pp. 379–423, July and Oct. 1948.

R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc. New York, NY, USA, 1968.

Setup



- Maximum-Likelihood decoding

$$\widehat{m} = \underset{i=1,\dots,|\mathcal{M}|}{\arg\max}\ P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}^{(i)}) = \underset{i=1,\dots,|\mathcal{M}|}{\arg\max} \prod_{k=1}^{n} P_{Y|X}(y_k|x_k^{(i)})$$

- Maximum-Metric decoding

$$\widehat{m} = \underset{i=1,\dots,|\mathcal{M}|}{\arg\max}\ q(\boldsymbol{x}^{(i)}, \boldsymbol{y}) = \underset{i=1,\dots,|\mathcal{M}|}{\arg\max} \prod_{k=1}^{n} q(x_k^{(i)}, y_k)$$

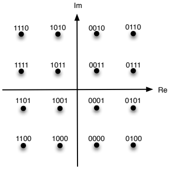
Examples

- Zero-error
- Zero undetected error
- Channel uncertainty $q(x, y) = \widehat{P}_{Y|X}(y|x)$
- Practical constraints
  - nearest-neighbor (non AWGN)
  - metric quantization

Examples

Bit-Interleaved Coded Modulation

- Pragmatic approach
- Simple decoder
- Minimal capacity penalty
- Simple and flexible design
- Used in most standards (DVB, WiFi, WiMAX, DSL, 4G...)



E. Zehavi, "8-PSK trellis codes for a Rayleigh fading channel", *IEEE Trans. Commun.*, 1992.
G. Caire, G. Taricco and E. Biglieri,"Bit-Interleaved Coded Modulation", *IEEE Trans. Inf. Theory*, 1998.

Setup Rates Error Exponents BICM

Examples

Bit-Interleaved Coded Modulation

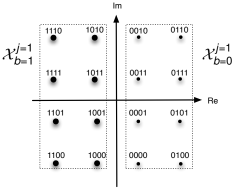
$$\hat{\mathrm{m}} = \arg\max_{i=1,\dots,|\mathcal{M}|} q(\boldsymbol{x}^{(i)}, \boldsymbol{y})$$
$$= \arg\max_{i=1,\dots,|\mathcal{M}|} \prod_{k=1}^{n} q(x_k^{(i)}, y_k)$$
$$= \arg\max_{i=1,\dots,|\mathcal{M}|} \prod_{k=1}^{n} \prod_{j=1}^{m} q_j(b_j(x_k^{(i)}), y_k)$$

Examples

Bit-Interleaved Coded Modulation

$$q_j(b_j(x) = b, y) = \sum_{x' \in \mathcal{X}_b^j} P_{Y|X}(y|x')P_X(x')$$

$$\widehat{\text{m}} = \arg\max_{i=1,\dots,|\mathcal{M}|} \prod_{k=1}^{n} \prod_{j=1}^{m} \sum_{x' \in \mathcal{X}^j_{b_j(x_k^{(i)})}} P_{Y|X}(y_k|x')P_X(x')$$



Rates

$$C_q = ?$$

Rates

i.i.d. Random Coding

$$\mathbb{P}[\boldsymbol{X} = \boldsymbol{x}] = \prod_{k=1}^{n} Q(x_k)$$

- Mismatched decoder

$$\widehat{m} = \arg\max_{i=1,\ldots,|\mathcal{M}|} q(\boldsymbol{x}^{(i)}, \boldsymbol{y})$$

- Can achieve rate

$$\sum_{x,y} Q(x) P_{Y|X}(y|x) \log \frac{q(x,y)}{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)}$$

$$I(X;Y) = \sum_{x,y} Q(x) P_{Y|X}(y|x) \log \frac{P_{Y|X}(y|x)}{\sum_{\bar{x}} Q(\bar{x}) P_{Y|X}(y|\bar{x})}$$

T. R. M. Fischer, "Some remarks on the role of inaccuracy in Shannon's theory of information transmission," in Trans. 8th Prague Conf. on Inf. Theory, 1971, pp. 211–226.

Rates

i.i.d.

- Mismatched decoder

$$\widehat{m} = \arg\max_{i=1,\ldots,|\mathcal{M}|} q(\boldsymbol{x}^{(i)}, \boldsymbol{y}) = \arg\max_{i=1,\ldots,|\mathcal{M}|} q(\boldsymbol{x}^{(i)}, \boldsymbol{y})^s$$

- Generalized Mutual Information

$$I^{\text{GMI}}(Q) = \sup_{s \geq 0} \sum_{x,y} Q(x) P_{Y|X}(y|x) \log \frac{q(x,y)^s}{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s}$$

G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," Arch. Elek. Über., vol. 47, no. 4, pp. 228–239, 1993.

Rates

Constant Composition Random Coding

- Codewords have the same empirical distribution

$$p_{\boldsymbol{x}}(x) = \frac{\text{Number of symbols } x \text{ in } \boldsymbol{x}}{n}$$

- If metric $q(x, y)$ is replaced by $q(x, y)^s e^{a(x)}$

$$\prod_{k=1}^{n} q(x_k, y_k)^s e^{a(x_k)} = \left(\prod_{k=1}^{n} q(x_k, y_k)\right)^s e^{\sum_{k=1}^{n} a(x_k)}$$

- LM rate

$$I_{\text{LM}}(Q) = \sup_{s\geq 0, a(\cdot)} \sum_{x,y} Q(x) P_{Y|X}(y|x) \log \frac{q(x, y)^s e^{a(x)}}{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{a(\bar{x})}}$$

J. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, MIT, 1983.

I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," IEEE Trans. Inf. Theory, Jan. 1981.

Rates

Cost Constrained Random Coding

- Codewords meet a cost function

$$\mathbb{P}\{\boldsymbol{X} = \boldsymbol{x}\} = \frac{1}{\mu_n} \prod_{k=1}^{n} Q(x_k) \mathbb{1}\left\{\left|\frac{1}{n}\sum_{k=1}^{n} a(x_k) - \mathbb{E}_Q[a(X)]\right| \leq \frac{\delta}{n}\right\}$$

- LM rate is also achieved

$$I_{\text{LM}}(Q) = \sup_{s\geq 0, a(\cdot)} \sum_{x,y} Q(x) P_{Y|X}(y|x) \log \frac{q(x, y)^s e^{a(x)}}{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{a(\bar{x})}}$$

A. Ganti, A. Lapidoth, and E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," IEEE Trans. Inf. Theory, Nov. 2000.

S. Shamai and I. Sason, "Variations on the Gallager bounds, connections, and applications," IEEE Trans. Inf. Theory, Dec. 2002

Rates

Properties

- Different ensembles achieve different rates
- Ensemble tightness: $\overline{P}_e \to 1$ when
  - $R > I_{\text{GMI}}(Q)$ (i.i.d. random coding)
  - $R > I_{\text{LM}}(Q)$ (constant-composition random coding)
  - $R > I_{\text{LM}}(Q)$ (cost-constrained random coding)
- Data Processing Inequality: $I_{\text{LM}}(Q) \le I(X;Y)$ with equality iff

$$\log q(x,y) = \alpha(x) + \beta(y) + c \log P_{Y|X}(y|x)$$

  for some $\alpha(x)$, $\beta(y)$, $c > 0$
- $I_{\text{LM}}(Q)$ can be *non-convex* in Q

N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," IEEE Trans. Inf. Theory, Nov. 1994.

Error Exponents

- Upper bound

$$\overline{P}_e \le \mathbb{E}\left[\min\left\{1, (|\mathcal{M}|-1)\mathbb{P}\{q(\boldsymbol{X}', \boldsymbol{Y}) \ge q(\boldsymbol{X}, \boldsymbol{Y}) | \boldsymbol{X}, \boldsymbol{Y}\}\right\}\right]$$

- Lower bound

$$\overline{P}_e \ge \frac{1}{4}\mathbb{E}\left[\min\left\{1, (|\mathcal{M}|-1)\mathbb{P}\{q(\boldsymbol{X}', \boldsymbol{Y}) \ge q(\boldsymbol{X}, \boldsymbol{Y}) | \boldsymbol{X}, \boldsymbol{Y}\}\right\}\right]$$

- Ensemble tight exponent given by

$$\liminf_{n\to\infty} -\frac{1}{n} \log \mathbb{E}\left[\min\left\{1, (|\mathcal{M}|-1)\mathbb{P}\{q(\boldsymbol{X}', \boldsymbol{Y}) \ge q(\boldsymbol{X}, \boldsymbol{Y}) | \boldsymbol{X}, \boldsymbol{Y}\}\right\}\right]$$

- Method of types can be used to find exponent (primal form)

J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched Decoding: Error Exponents, Second-Order Rates and Saddlepoint Approximations", IEEE Trans. Inf. Theory, vol. 60, no. 5, pp. 2647-2666, May 2014.

I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, 2nd ed. Cambridge University Press, 2011.

R. Gallager, "Fixed composition arguments and lower bounds to error probability," http://web.mit.edu/gallager/www/notes/notes5.pdf.

Error Exponents

Code Ensembles

- i.i.d.

$$\mathbb{P}\{\boldsymbol{X}=\boldsymbol{x}\}=\prod_{k=1}^{n} Q(x_k)$$

- Constant composition

$$\mathbb{P}\{\boldsymbol{X}=\boldsymbol{x}\}=\begin{cases}\frac{1}{|T(Q)|} & \boldsymbol{x}\in T(Q)\\ 0 & \text{otherwise}\end{cases}$$

- Multiple cost constraints, $\phi_l=\mathbb{E}_Q[a_l(X)]$

$$\mathbb{P}\{\boldsymbol{X}=\boldsymbol{x}\}=\frac{1}{\mu_n}\prod_{k=1}^{n} Q(x_i)\mathbb{1}\left\{\left|\frac{1}{n}\sum_{k=1}^{n} a_l(x_k)-\phi_l\right|\le\frac{\delta}{n},\ l=1,\dots,L\right\}$$

J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched Decoding: Error Exponents, Second-Order Rates and Saddlepoint Approximations", IEEE Trans. Inf. Theory, vol. 60, no. 5, pp. 2647-2666, May 2014.

Error Exponents

Ensemble Tightness

$$E_r(R,Q)=\max_{0\le\rho\le 1} E_0(\rho,Q)-\rho R$$

$$E_0^{\text{iid}}(\rho,Q)=\sup_{s\ge 0}-\log\sum_{x,y} Q(x)P_{Y|X}(y|x)\left(\frac{\sum_{\overline{x}} Q(\overline{x})q(\overline{x},y)^s}{q(x,y)^s}\right)^{\rho}$$

$$E_0^{\text{cc}}(\rho,Q)=\sup_{s\ge 0,a(\cdot)}-\sum_{x} Q(x)\log\sum_{y} P_{Y|X}(y|x)\left(\frac{\sum_{\overline{x}} Q(\overline{x})q(\overline{x},y)^s e^{a(\overline{x})}}{q(x,y)^s e^{a(x)}}\right)^{\rho}$$

$$E_0^{\text{cost}}(\rho,Q,\{a_l(\cdot)\})$$

$$=\sup_{\substack{s\ge 0\\ \{r_l\},\{\overline{r}_l\}}}-\log\sum_{x,y} Q(x)P_{Y|X}(y|x)\left(\frac{\sum_{\overline{x}} Q(\overline{x})q(\overline{x},y)^s e^{\sum_{l=1}^{L}\overline{r}_l(a_l(\overline{x})-\phi_l)}}{q(x,y)^s e^{\sum_{l=1}^{L} r_l(a_l(x)-\phi_l)}}\right)^{\rho}$$

J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched Decoding: Error Exponents, Second-Order Rates and Saddlepoint Approximations", IEEE Trans. Inf. Theory, vol. 60, no. 5, pp. 2647-2666, May 2014.

Error Exponents

Proposition

$$E_r^{\text{iid}}(R,Q) \le E_r^{\text{cost}}(R,Q,\{a_l\}) \le E_r^{\text{cc}}(R,Q)$$

$$\sup_{a_1(\cdot),a_2(\cdot)} E_r^{\text{cost}}(R,Q,\{a_1,a_2\}) = E_r^{\text{cc}}(R,Q)$$

J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched Decoding: Error Exponents, Second-Order Rates and Saddlepoint Approximations", IEEE Trans. Inf. Theory, vol. 60, no. 5, pp. 2647-2666, May 2014.

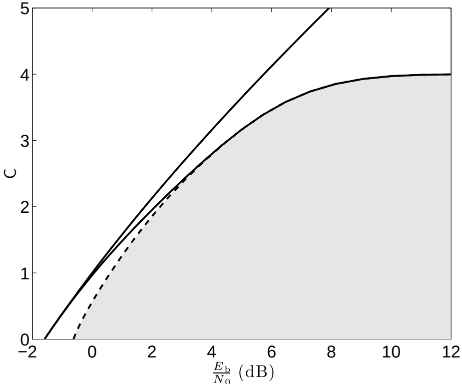
BICM

Theorem

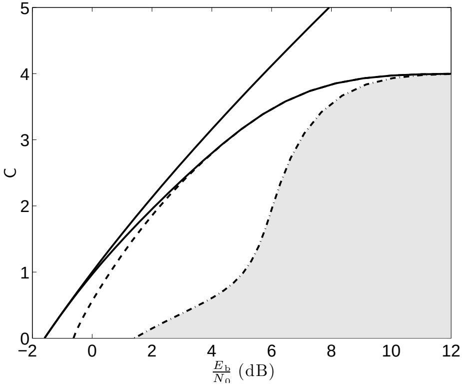
$$I_{GMI}(Q) = \sum_{j=1}^{m} I(B_j; Y)$$

A. Martinez, A. Guillén i Fàbregas, G. Caire and F. Willems, "Bit-Interleaved Coded Modulation Revisited: A Mismatched Decoding Perspective", *IEEE Trans. Inf. Theory*, June 2009.

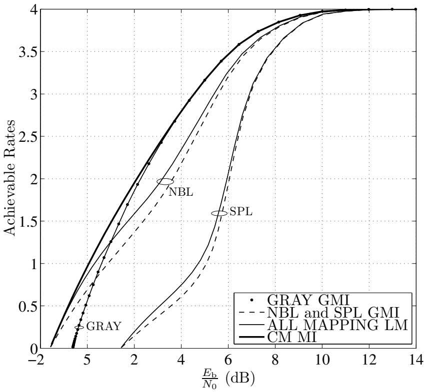
BICM


5
4
3
2
1
0
C
−2
0
2
4
6
8
10
12
$\frac{E_b}{N_0}$ (dB)

BICM


5
4
3
2
1
0
C
−2
0
2
4
6
8
10
12
$\frac{E_b}{N_0}$ (dB)

BICM



BICM

Wideband Regime

- Information rates can be expanded as

$$C(\mathrm{SNR}) = c_1 \mathrm{SNR} + c_2 \mathrm{SNR}^2 + o(\mathrm{SNR}^2)$$

- c_1, c_2 depend on the transmission scheme
- c_1 and c_2 trade off power and bandwidth (Verdú 2002)
  - $c_1 = 1$ for many modulation formats
  - $c_2 = -\frac{1}{2}$ for proper complex modulations (Prelov, Verdú 2004)

BICM

Wideband Regime

Theorem (Wideband Regime BICM)

The low-SNR expansions of BICM achievable rates satisfy

$$c_{1,GMI} = c_{1,LM} = \sum_{j=1}^{m} \frac{1}{2} \sum_{b=0}^{1} \left| \mathbb{E}[X_b^j] \right|^2 \tag{1}$$

$$c_{2,GMI} = -\frac{1}{2} \left(m\,\kappa(\mathcal{X}) - \sum_{i=1}^{m} \sum_{b} \frac{1}{2} \kappa(\mathcal{X}_b^j) \right) \tag{2}$$

$$c_{2,LM} = c_{2,GMI} + \frac{1}{2} \sum_{x} \frac{1}{2^m} \left(\sum_{i,j=1; i \neq j}^{m} r\left(\overline{\mathcal{X}}^i_{b_i(x)}, \overline{\mathcal{X}}^j_{b_j(x)}\right) \right)^2 \tag{3}$$

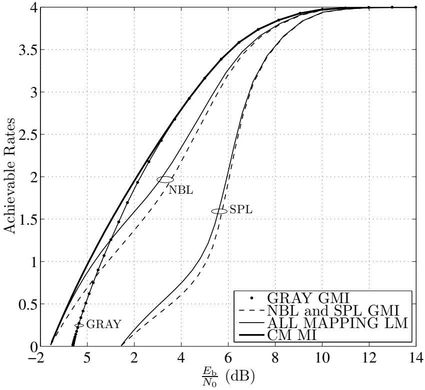
A. Martinez, L. Peng, A. Alvarado, A. Guillén i Fàbregas, “Improved Information Rates for Bit-Interleaved Coded Modulation," 2015 IEEE Int. Symp. Inf. Theory, Hong Kong.

BICM

Wideband Regime: Square 2^m-QAM

	c_1		c_2	
	GMI	LM	GMI	LM
BRGL	$\frac{3 \cdot 2^{2m}}{4 \cdot (2^{2m} - 1)} < 1$	$\frac{3 \cdot 2^{2m}}{4 \cdot (2^{2m} - 1)} < 1$	(2)	(3)
NBL	1	1	(2) $(< -\frac{1}{2})$	$-\frac{1}{2}$

BICM



Achievable Rates
4
3.5
3
2.5
2
1.5
1
0.5
0
−2
0
2
4
6
8
10
12
14
$\frac{E_b}{N_0}$ (dB)
NBL
SPL
GRAY
GRAY GMI
NBL and SPL GMI
ALL MAPPING LM
CM MI