On Joint Design of Probabilistic Shaping and Forward Error Correction for Optical Systems

OFC 2018 / San Diego
12 March 2018

www.huawei.com

Georg Böcherer

Mathematical and Algorithmic Sciences Lab

Huawei Technologies Paris

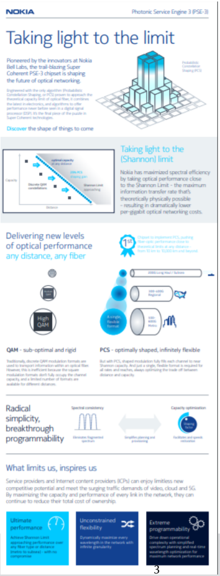
PAS History

- G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015
- P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016
- F. Buchali, G. Böcherer, W. Idler, L. Schmalen, P. Schulte, and F. Steiner, "Experimental demonstration of capacity increase and rate-adaptation by probabilistically shaped 64-QAM," in *Proc. Eur. Conf. Optical Commun. (ECOC)*, Paper PDP3.4, Valencia, Spain, 2015

PAS History

Bell Labs Prize Final 2015

PSE-3 Infographic 2018



This Tutorial

- For practical performance of PS, visit **booth 2228** and check out the **PSE-3**.
- **This talk:** A foundation of PS design tools.

Outline

- PS Achievable FEC Rates
- **Case Study:** Offline Calculation of PS Achievable FEC Rates
- PS Achievable Rates

Part 1: PS Achievable FEC Rates

Code word detection in noise

PS Code Ensemble

- Linear code

$$\mathscr{C} = \{\boldsymbol{c} \in \mathscr{X}^n : \boldsymbol{c}\boldsymbol{H}^T = \boldsymbol{0}\}. \tag{1}$$

- Transmit shaped code word $\boldsymbol{x} \in \mathscr{C}$ with empirical distribution P_X.
- Non-negative decoding metric

$$q(x, y), \quad x \in \mathscr{X}, y \in \mathscr{Y}. \tag{2}$$

- Decoding rule

$$\hat{\boldsymbol{c}} = \underset{\boldsymbol{c}:\, \boldsymbol{c}\boldsymbol{H}^T = \boldsymbol{0}}{\operatorname{argmax}} \prod_{j=1}^{n} q(c_j, y_j). \tag{3}$$

- Decoding error if $\boldsymbol{c} \neq \boldsymbol{x}$.

Question: is there a rate R code that decodes x^n correctly from y^n?

Literature

- Shannon, 1948 [4]: mutual information by typicality.
- Gallager, 1968 [5]: mutual information by error exponent.
- Kaplan & Shamai, 1993 [6]: generalized mutual information (GMI) by error exponent.
- Ganti, Lapidoth, Telatar, 2000 [7]: LM-rate and GMI by threshold decoder.

PS ensemble is NOT treated

Research on PS achievable rates since 2014, my findings:

- G. Böcherer, "Achievable rates for probabilistic shaping," *arXiv preprint*, 2017. [Online]. Available: https://arxiv.org/abs/1707.01134
- Explains why I don't use the GMI and its variations.

PS Achievable FEC Rate

- Measurement x^n, y^n: For code rates $< R_{\text{FEC}}$, there exist codes that can decode x^n from y^n using metric q where

$$\boxed{R_{\text{FEC}} = \log_2|\mathscr{X}| - \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left[-\log_2 \frac{q(x_i, y_i)}{\sum_{a\in\mathscr{X}} q(a, y_i)}\right]}_{\text{uncertainty}}}$$

- Memoryless channel $p_{Y|X}$:

$$R_{\text{FEC}} = \log_2|\mathscr{X}| - \mathbb{E}\left[-\log_2 \frac{q(X, Y)}{\sum_{a\in\mathscr{X}} q(a, Y)}\right].$$

PS Achievable FEC Rate

Powerful tool, can be directly instantiated for

- Optimal metric.
- Binary FEC: Achievable Binary Code (ABC) Rate.
- Soft-decision (SD) metric.
- Hard-decision (HD) metric.
- ...

Example: Optimal Metric

- Optimal metric

$$q(x,y) = P_{X|Y}(x|y).$$

- Uncertainty

$$\mathbb{H}(X|Y).$$

- Achievable FEC Rate

$$R_{\mathrm{FEC}} = \log_2|\mathscr{X}| - \mathbb{H}(X|Y).$$

Example: ABC Rate

- m-bit constellation label $\boldsymbol{b} = b_1 \cdots b_m$.
- Binary metric

$$q(\boldsymbol{b}, y) = \prod_{i=1}^{m} q_i(b_i, y).$$

- ABC rate

$$\boxed{R_{\text{abc}} = 1 - \frac{1}{m} \sum_{i=1}^{m} \mathbb{E}\left[-\log_2 \frac{q_i(B_i, Y)}{\sum_{b \in \{0,1\}} q_i(b, Y)} \right]}$$

Example: SD Decoding

- Bitwise demapper calculates

$$\ell_i = \log \frac{P_{B_i|Y}(0|y)}{P_{B_i|Y}(1|y)}.$$

log domain	probability domain
$q_{\log}(\boldsymbol{b}, \ell) = \sum_{i=1}^{m}(1-2b_i)\ell_i$	$q(\boldsymbol{b}, \ell) = \prod_{i=1}^{m} e^{s(1-2b_i)\ell_i}$

- Optimal for channel $P_{\boldsymbol{B}|Y}$, achieving

$$\boxed{R_{\text{abc}} = 1 - \frac{1}{m}\sum_{i=1}^{m} \mathbb{H}(B_i|Y).}$$

Example: HD Decoding

- Demapper calculates

$$\hat{b}_i = \omega_i(y).$$

- Hamming metric

$$q(b, \hat{b}_i) = \mathbb{1}(b, \hat{b}_i) = \begin{cases} 1, & b = \hat{b}_i \\ 0, & \text{otherwise} \end{cases}$$

log domain	probability domain
$q_{\log}(\boldsymbol{b}, \hat{\boldsymbol{b}}) = \sum_{i=1}^{m} \mathbb{1}(b_i, \hat{b}_i)$	$q(\boldsymbol{b}, \hat{\boldsymbol{b}}) = \prod_{i=1}^{m} e^{s\mathbb{1}(b_i, \hat{b}_i)}$

- Achieves

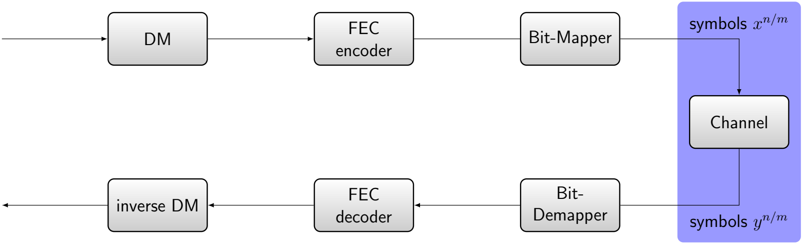
$$\boxed{R_{\text{abc}} = 1 - \mathbb{H}_2(\varepsilon)}$$

where ε is the preFEC-BER.

Part 2: Case Study

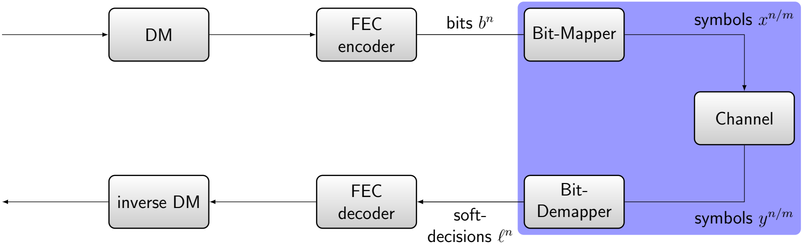
Offline Calculation of Achievable FEC Rates from Measurements

16-QAM Experiment



- Gray labelled 16-QAM constellation $\Rightarrow m = 4$.
- $n/m = 64\,800/4 = 16\,200$ quadrature amplitude modulation (QAM) symbols $x^{n/m}$.
- Noisy measurement $y^{n/m}$.

Bitwise Demapping



- Define **offline** a label $\{0,1\} \to \mathscr{X}$ on the input alphabet $\mathscr{X}$.
- Represent the n/m input symbols $x^{n/m}$ by n bits b^n according to the label.
- Demapper **assumes** Gaussian noise.
- For each bit b_{ji}, the demapper outputs

$$\ell_{ji} = \log \frac{P_{B_i|Y}(0|y_j)}{P_{B_i|Y}(1|y_j)}. \tag{4}$$

ABC Rate

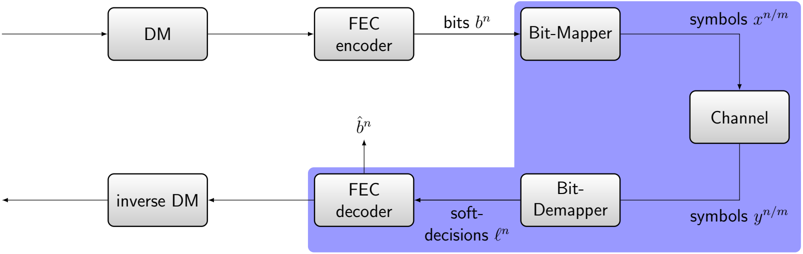
- For channel measurement $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n/m}, \ell_1, \ldots, \ell_{n/m}$, ABC rate is

$$R_{\text{abc}} = 1 - \frac{1}{\frac{n}{m}} \sum_{j=1}^{n/m} \frac{1}{m} \sum_{i=1}^{m} \left(-\log_2 \frac{e^{(1-2b_{ji})\frac{\ell_{ji}}{2}}}{e^{-\frac{\ell_{ji}}{2}} + e^{\frac{\ell_{ji}}{2}}} \right) \tag{5}$$

$$= 0.6156\,\text{bit}. \tag{6}$$

$\Rightarrow$ For code rates < 0.6156 bit, there exist FEC codes that can recover b^n from ℓ^n.

Offline Evaluation of FEC Codes



- **Objective:** Check if actual forward error correction (FEC) decoders can recover b^n from ℓ^n so that $\hat{b}^n = b^n$.

Offline Evaluation of FEC Codes

- MATLAB implements the length $n = 64\,800$ DVB-S2 LDPC codes of rates

$$R_{\text{FEC}} = 1/4,\ 1/3,\ 2/5,\ 1/2,\ 3/5,\ 2/3,\ 3/4,\ 4/5,\ 5/6,\ 8/9,\ 9/10 \tag{7}$$

- **Objective:** use ABC rate to predict which of these FEC Rates are achievable for our 16-QAM measurement b^n, ℓ^n.
- We check this by passing ℓ^n to the respective decoders and check if for the output we have $\hat{b}^n = b^n$.
- **Problem:** we transmitted b^n before choosing a code and b^n may not be a code word in any of the codes of interest.

Offline Evaluation of FEC Codes

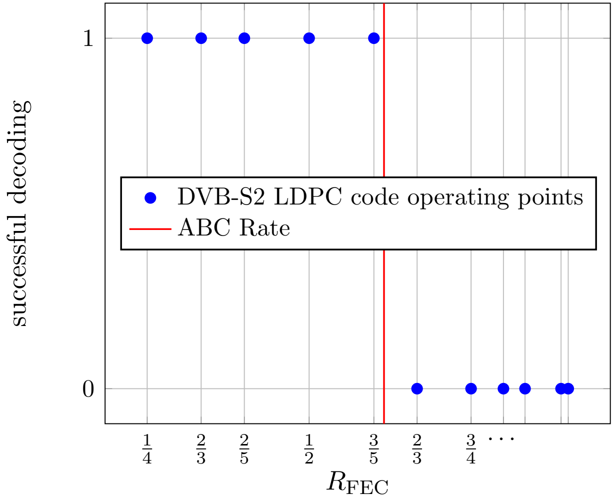
The following procedure solves the problem of b^n not being a code word.

- Pick an arbitrary code word c^n from a code of interest.
- Calculate the scrambling sequence $s^n = c^n \oplus b^n$.
- Calculate the modified demapper output $\tilde{\ell}^n$ with

$$\tilde{\ell}_i = (1 - 2s_i)\ell_i. \tag{8}$$

- Pass $\tilde{\ell}^n$ to the decoder and check if it decides for c^n.

Offline Evaluation of FEC Codes



Part 3: PS Achievable Rates

Mapping to shaped sequences

From Achievable FEC Rates to Achievable Rates

- **Recall:** Measurement x^n, y^n, achievable FEC Rate

$$Rac = \log_2 |\mathscr{X}| - \underbrace{\sum_{i=1}^{n} \left[-\log_2 \frac{q(x_i, y_i)}{\sum_{a \in \mathscr{X}} q(a, y_i)} \right]}_{\text{uncertainty } u_s} \tag{9}$$

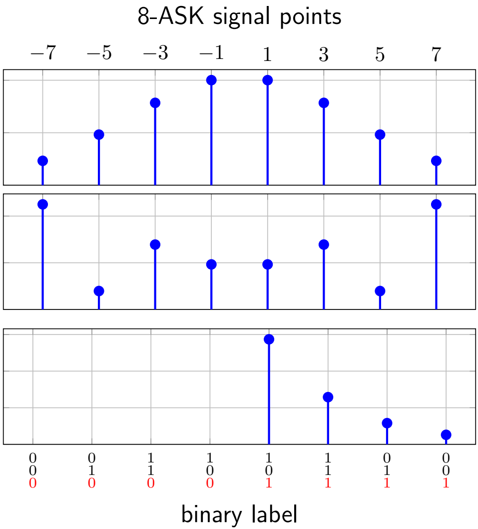
- Let $\mathscr{S} \subseteq \mathscr{C}$ be the subset of code word achieving uncertainty $\leq u_s$.
- Achievable rate is

$$\boxed{R = \left[\frac{\log_2 |\mathscr{S}|}{n} - u_s \right]^+.} \tag{10}$$

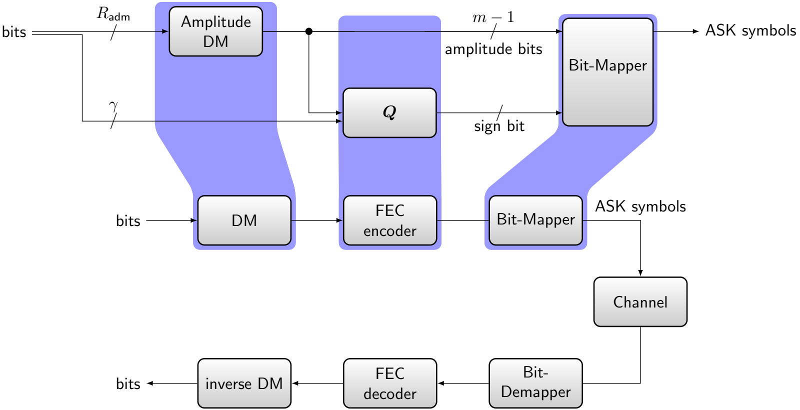
- **Challenge 1:** identify $\mathscr{S}$ and $|\mathscr{S}|$.
- **Challenge 2:** encode into $\mathscr{S}$.

Example: Shaping for Coherent Transmission

Good 1D Input Distributions



probabilistic amplitude shaping (PAS) [1]



- How many shaped code words can this architecture index?

Intermezzo: Types [9]

- Let x^n be a sequence with symbols in $\mathscr{X}$.
- Let P_{x^n} be the empirical distribution of x^n, i.e.,

$$P_{x^n}(a) = \frac{\text{number of occurences of } a \text{ in } x^n}{n} = \frac{n_a}{n}, \quad a \in \mathscr{X}. \tag{11}$$

- $P_X = P_{x^n}$ is a distribution on $\mathscr{X}$ and is called an n-type.
- All permutations of x^n also have the n-type P_X.
- Let $\mathscr{T}^n(P_X)$ be the set of all permutations of x^n.

1D PAS Achievable Rate

- Shaping set $\mathscr{S} = \mathscr{T}^w(P_A) \times \{-1,1\}^w$, where P_A is an amplitude distribution.
- A constant composition distribution matcher (CCDM) [2] can index $2^{\lfloor \log_2 |\mathscr{T}^w(P_A)| \rfloor}$ sequences in $\mathscr{T}^w(P_A)$.
- There are 2^w sign sequences in $\{-1,1\}^w$.
- 1D PAS Achievable rate is

$$\mathrm{R_{PAS}} = \left[\frac{\lfloor \log_2 |\mathscr{T}^w(P_A)| \rfloor}{w} + 1 - u_s \right]^+ \qquad (12)$$

- We have

$$|\mathscr{T}^w(P_A)| = \binom{w}{w_1, w_2, \ldots, w_M} \qquad (13)$$

where $w_i = w \cdot P_A(a_i)$ and where M is the number of distinct amplitudes.

Asymptotic Rate on Memoryless Channel

- Suppose $P_{\boldsymbol{B}}, p_{Y|\boldsymbol{B}}$ assumed by the demapper are correct so that the uncertainty is

$$u_s = \sum_{i=1}^{m} \mathbb{H}(B_i|Y). \tag{14}$$

- Suppose further that the length w of the CCDM output is large so that

$$\frac{\log_2 |\mathscr{T}^w(P_A)|}{w} = \mathbb{H}(P_A).$$

- In this case, we have

$$\boxed{\mathrm{R}_{\mathrm{PAS}} = \left[\mathbb{H}(P_A) + 1 - \sum_{i=1}^{m} \mathbb{H}(B_i|Y)\right]^+ = \left[\mathbb{H}(\boldsymbol{B}) - \sum_{i=1}^{m} \mathbb{H}(B_i|Y)\right]^+.} \tag{15}$$

Discussion

- For **memoryless channels**, we can choose w large, e.g.,$w = n/m$ and all sequences in $\mathscr{T}^w(P_A)$ result in the same uncertainty u_s.
- **In general not true for the optical fiber.** We may therefore need to choose $w \ll n/m$. This must be accounted for when calculating the achievable rate and it may be smaller than $\left[\mathbb{H}(\boldsymbol{B}) - \sum_{i=1}^{m} \mathbb{H}(B_i|Y)\right]^+$.

Conclusions

- We learned how to determine PS achievable FEC rates **offline** from measurements.
- We learned how to determine PS rates achievable by practical systems.
- Key tools are **uncertainty**, **ABC rate**, and **counting sequences**.

Details in

- G. Böcherer, "Achievable rates for probabilistic shaping," *arXiv preprint*, 2017. [Online]. Available: https://arxiv.org/abs/1707.01134

References I

[1] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.

[2] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.

[3] F. Buchali, G. Böcherer, W. Idler, L. Schmalen, P. Schulte, and F. Steiner, "Experimental demonstration of capacity increase and rate-adaptation by probabilistically shaped 64-QAM," in *Proc. Eur. Conf. Optical Commun. (ECOC)*, Paper PDP3.4, Valencia, Spain, 2015.

[4] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, 379–423 and 623–656, 1948.

[5] R. G. Gallager, *Information Theory and Reliable Communication.* John Wiley & Sons, Inc., 1968.

[6] G. Kaplan and S. Shamai (Shitz), "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEÜ*, vol. 47, no. 4, pp. 228–239, 1993.

References II

[7] A. Ganti, A. Lapidoth, and E. Telatar, “Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit,” *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[8] G. Böcherer, “Achievable rates for probabilistic shaping,” *arXiv preprint*, 2017. [Online]. Available: `https://arxiv.org/abs/1707.01134`.

[9] I. Csiszár and P. C. Shields, “Information theory and statistics: A tutorial,” *Found. Trends Comm. Inf. Theory*, vol. 1, no. 4, pp. 417–528, 2004.

Acronyms

FEC	forward error correction
PAS	probabilistic amplitude shaping
QAM	quadrature amplitude modulation
LDPC	low-density parity-check
CCDM	constant composition distribution matcher