# Integration of Probabilistic Shaping and Forward Error Correction

## Spectral Efficiency, Rate, Overhead

Georg Böcherer

CNRS / GdR ISIS Workshop on Coding, Modulation, and Signal Processing for Optical Communications
Telecom ParisTech
Paris, France

September 19, 2019

# Outline

# Outline

Motivation

Layered Probabilistic Shaping (PS)

Spectral Efficiency, Rate and Overhead

Achievable Rates

References

# Constellation, FEC, and Shaping Design for Optical Communications

- At the receiver side, forward error correction (FEC) must remove the residual impairments after the DSP.
- The residual impairments are modelled reasonably well as additive Gaussian noise.
- The design of constellation, FEC, and shaping is done assuming additive white Gaussian noise (AWGN).

# AWGN Capacity

- Real-valued zero mean Gaussian noise $Z$ with variance $\sigma^2$.
- Average input power $\mathbb{E}[X^2] \leq \mathcal{E}$.
- $\text{SNR} = \frac{\mathcal{E}}{\sigma^2}$.
- AWGN capacity $0.5 \log_2(1 + \text{SNR})$.
- $\mathbb{E}[\cdot]$ is the expectation operator.

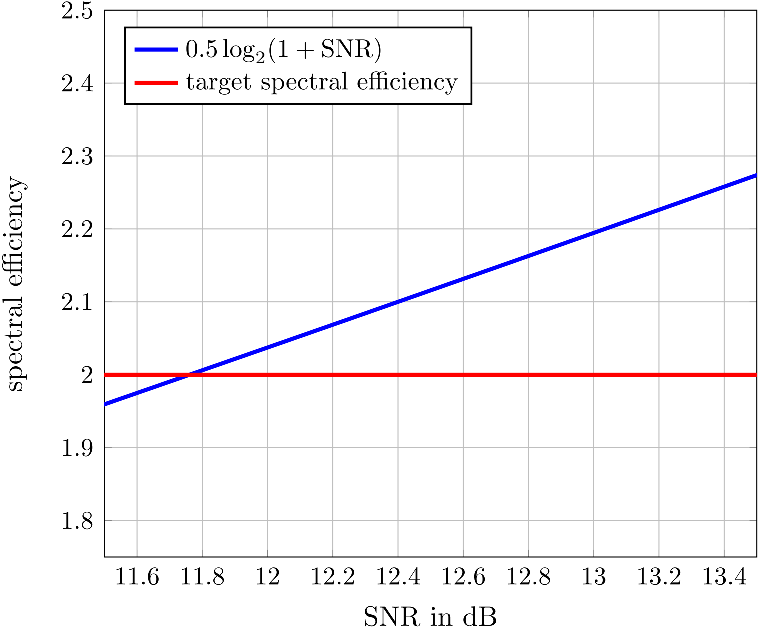
# Constrained AWGN Capacity

- Finite alphabet $\mathcal{X}$ (e.g., 4-ASK: $\mathcal{X} = \{\pm 1, \pm 3\}$).
- Input distribution $P_X$ on $\mathcal{X}$.
- Constrained capacity

$$\begin{aligned} \max_{P_X, \Delta} \quad & \mathbb{I}(X; \Delta X + Z) \\ \text{subject to} \quad & \mathbb{E}[(X\Delta)^2] \leq \mathcal{E} \end{aligned}$$

- $\mathbb{I}(X;Y)$ is the mutual information of $X$ and $Y$.

# Design Problem



# Design Problem

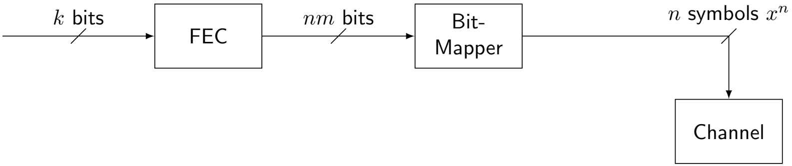
- We want to achieve a target spectral efficiency at the lowest possible SNR.
- Reformulation of the constrained capacity:

$$\begin{aligned} \min_{P_X,\Delta} \quad & \mathbb{E}[(X\Delta)^2] \\ \text{subject to} \quad & \mathbb{I}(X; \Delta X + Z) \geq \text{SE} \end{aligned}$$

- **Topic of this talk:** how to achieve it in practice.

# Outline

# Approach

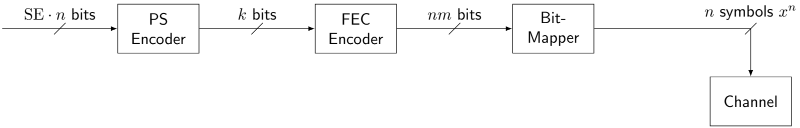



- FEC fixed and given.
- Shaping requirement: $x^n \in \mathcal{S}$ for shaping set $\mathcal{S}$.

# Shaping Set Example

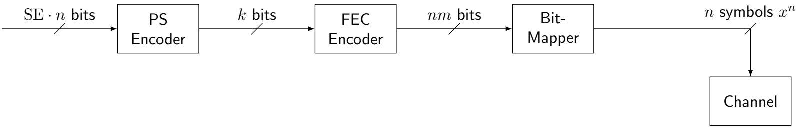
- Input distribution $P_X$.
- Shaping set $\mathcal{S}$ contains all length $n$ sequences in $\mathcal{X}^n$ that have approximately distribution $P_X$.
- Known as typical set, type class $\mathcal{T}^n(P_X)$.

# PS Encoder




1. Identify FEC encoder inputs that map to sequences $x^n \in \mathcal{S}$.
2. Let PS encoder index valid FEC encoder inputs.

# PS Encoder




| component | rate |
|---|---|
| Bit-mapper | $m$ [bits/symbol] |
| FEC encoder | $R_{\text{fec}} = \frac{k}{mn}$ |
| PS encoder | $R_{\text{ps}} = \frac{\text{SE} \cdot n}{k}$ |

- Spectral efficiency is $\text{SE} = R_{\text{ps}} \cdot R_{\text{fec}} \cdot m$.
- PS rate $R_{\text{ps}}$ depends on FEC rate $R_{\text{fec}}$.
- PS rate depends **implicitly** on shaping set $\mathcal{S}$.

# Outline

# Shaping Set Rate $R_{\text{ss}}$

$$R_{\text{ss}} = \frac{\log_2 |\mathcal{S}|}{nm}$$

# Shaping Set Rate: Examples

- **No shaping:**

$$R_{ss} = \frac{\log_2 |\mathcal{S}|}{nm} = \frac{\log_2 |\mathcal{X}|^n}{nm} = \frac{m}{m} = 1.$$

- $P_X$**-type class (**$n$ **large)**:

$$R_{ss} = \frac{\mathbb{H}(P_X)}{m}.$$

- $P_X$**-type class (**$n$ **not so large)**[1]

$$R_{ss} = \frac{\log_2 \binom{n}{n_1, n_2, \ldots, n_{|\mathcal{X}|}}}{mn}, \quad n_i = n \cdot P_X(x_i).$$

- $\mathbb{H}(P_X) = \mathbb{H}(X)$ is the entropy of $X$.

[1] Encoding into type classes can be done efficiently by Constant Composition Distribution Matching (CCDM) [1].

# Shaping Set Rate: More Examples

- Consider 1D 4-ASK constellation
$$\mathcal{X} = \{\pm 1, \pm 3\}.$$
- **Shaping set:**
  - Constrain amplitude to distribution
$$P_A(1) = \frac{n_1}{n}, \quad P_A(3) = \frac{n_3}{n}.$$
  - Leave sign unconstrained.
- **Shaping set rate** is
$$R_{\mathrm{ss}} = \frac{\log_2 |\mathcal{S}|}{nm} = \frac{\log_2 \left[\binom{n}{n_1, n_3} \cdot 2^n\right]}{nm} = \frac{\log_2 \binom{n}{n_1, n_3}}{nm} + \frac{1}{m}.$$

# Spectral Efficiency

$$\begin{aligned}\mathrm{SE} &\leq \left[\frac{\log_2|\mathcal{S}|}{n} - m(1-R_{\mathrm{fec}})\right]^+ \\ &= m\cdot[1-(1-R_{\mathrm{ss}})-(1-R_{\mathrm{fec}})]^+ \\ &= m\cdot[R_{\mathrm{ss}}+R_{\mathrm{fec}}-1]^+ = mR_{\mathrm{ps}}R_{\mathrm{fec}}\end{aligned}$$

PS and FEC are separated!!

- With "=" if there are at least $m(1-R_{\mathrm{fec}})$ unconstrained bit [2].
- With "=" asymptotically in $n$ in general [3], [4].

# Rate, Redundancy, Overhead

| | FEC | Shaping Set |
|---|---|---|
| Rate | $R_{\text{fec}} = \frac{k}{nm}$ | $R_{\text{ss}} = \frac{\log_2 \lvert\mathcal{S}\rvert}{nm}$ |
| Redundancy | $1 - R_{\text{fec}}$ | $1 - R_{\text{ss}}$ |
| Overhead in % | $100 \cdot \left(\frac{1}{R_{\text{fec}}} - 1\right)$ | $100 \cdot \left(\frac{1}{R_{\text{ss}}} - 1\right)$ |
| Total overhead in % | $100 \cdot \left(\frac{1}{R_{\text{ss}} + R_{\text{fec}} - 1} - 1\right)$ | |

# Outline

# Textbook Information Theory

- Code

$$\mathcal{C} = \left\{ x^n(1), x^n(2), \dots, x^n(2^{\mathsf{SE}n}) \right\}$$

with codeword entries iid $\sim P_X$.

- ML rule

$$\hat{w} = \underset{w \in \{1,\dots,2^{\mathsf{SE}n}\}}{\operatorname{argmax}} P^n_{Y|X}(y^n | x^n(w))$$

- Vanishing error probability for large $n$ if

$$\mathsf{SE} < \mathbb{I}(X;Y).$$

# Layered Probabilistic Shaping (Decoding)

- Code

$$\mathcal{C} = \left\{x^n(1), x^n(2), \dots, x^n(2^{R_{\text{fec}}mn})\right\}$$

with codeword entries iid uniform. NB: $R_{\text{fec}}m > \text{SE}$.

- MAP rule

$$\hat{w} = \underset{w\in\{1,\dots,2^{R_{\text{fec}}mn}\}}{\operatorname{argmax}} P^n_{X|Y}(x^n(w)|y^n)$$

- [3, Theorem 2]: Vanishing error probability for large $n$ if

$$m(1 - R_{\text{fec}}) > \mathbb{H}(X|Y).$$

# Layered Probabilistic Shaping (Encoding)

- Divide the codebook into $2^{\mathsf{SE}n}$ partitions.
- Map message $w$ to a codeword in $\mathcal{C} \cap \mathcal{S}$ in the $w$th partition.
- If no such codeword exists, declare an encoding error.
- [3, Theorem 1]: Vanishing error probability for large $n$ if

$$\mathsf{SE} < \frac{\log_2 |\mathcal{S}|}{n} - m(1 - R_{\mathsf{fec}}).$$

# Layered Probabilistic Shaping: Achievable SE

We have

$$\begin{aligned}\mathbb{I}(X;Y) &= \mathbb{H}(X) - \mathbb{H}(X|Y)\\ &\geq [\mathbb{H}(X) - m(1-R_{\text{fec}})]^+\\ &\geq \left[\frac{\log_2|\mathcal{S}|}{n} - m(1-R_{\text{fec}})\right]^+\end{aligned}$$

- By the two theorems above, we can approach equality for large $n$.

$\Rightarrow$ Layered probabilistic shaping is capacity-achieving.

# Practical Decoding Metrics

- Practical systems use a sub-optimal decoding metric $q(x, y)$ instead of $P_{X|Y}(x|y)$.
- The decoding rule becomes

$$\hat{w} = \operatorname*{argmax}_{w \in \{1,\ldots,2^{R_{\text{fec}} mn}\}} q^n(x^n(w), y^n).$$

- Achievable FEC rate generalizes to

$$m(1 - R_{\text{fec}}) \geq \mathbb{E}\left[-\log_2 \frac{q(X,Y)}{\sum_{a \in \mathcal{X}} q(a, Y)}\right] = \mathbb{U}(q, X, Y) \geq \mathbb{H}(X|Y)$$

- $\mathbb{U}(q, X, Y)$ is the uncertainty at the receiver about the input, which needs to be resolved by the FEC decoder.

# Uncertainty Examples

| | $\mathbb{U}(q, X, Y)$ |
|---|---|
| non-binary soft-decision FEC | $\mathbb{H}(X \mid Y)$ |
| binary soft-decision FEC | $\sum_{i=1}^{m} \mathbb{H}(B_i \mid Y)$ |
| binary hard-decision FEC | $m\,\mathbb{H}_2(\epsilon)$ |

- $B_1 B_2 \ldots B_m$, $B_i \in \{0, 1\}$, is the $m$-bit binary label of $X$ used by the bit-mapper.
- $\mathbb{H}_2(\epsilon) = -\epsilon \log_2 \epsilon - (1-\epsilon)\log_2(1-\epsilon)$ is the binary entropy function.
- $\epsilon$ is the BER at the FEC decoder input.

# Summary

- We can exactly quantify FEC, PS, and decoding metric penalties (assume $\mathsf{SE} > 0$)

$$
\begin{aligned}
\mathsf{SE} &= m\left[R_{\mathsf{ss}} - (1 - R_{\mathsf{fec}})\right] \\
&= mR_{\mathsf{ss}} - \mathbb{U}(q, X, Y) - m\Delta_{\mathsf{FEC}} \\
&= \mathbb{H}(X) - \mathbb{U}(q, X, Y) - m\Delta_{\mathsf{FEC}} - m\Delta_{\mathsf{PS}} \\
&= \underbrace{\mathbb{H}(X) - \mathbb{H}(X|Y)}_{\mathbb{I}(X;Y)} - m\Delta_{\mathsf{FEC}} - m\Delta_{\mathsf{PS}} - m\Delta_{\text{decoding metric}}
\end{aligned}
$$

- Textbook information theory:

$$\mathsf{SE} \leq \mathbb{I}(X;Y).$$

(Mutual information is generalized to arbitrary decoding metrics in [5]).

# Conclusions

**Discussed topics**

- Layered probabilistic shaping architecture.
- Spectral efficiency, rate, overhead.
- Information theory for component-wise benchmarking.

**Outlook**

- Use presented framework to design better systems.

# Outline

# References I


- P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.
- 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.
- G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic shaping and forward error correction for fiber-optic communication systems," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 230–244, Jan. 2019.
- G. Böcherer, D. Lentner, A. Cirino, and F. Steiner, "Probabilistic parity shaping for linear codes," *arXiv preprint*, 2019. [Online]. Available: https://arxiv.org/abs/1902.10648.

# References II

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.