Advances in Coded Modulation for Optical Communications

Gerhard Kramer
Department of Electrical and Computer Engineering
Technical University of Munich

OFC
Los Angeles, USA
March 22, 2017
Outline

1) Capacity and Shaping (old topic: 1960s)
2) PAS* Transmitter Architecture (seems new)
3) PAS Receiver Architecture
4) Applications to ATSC**, DVB***, and Optical Fiber
5) Product DM****

Acknowledgment: some graphics borrowed from G. Böcherer, R.-J. Essiambre and M. Magarini

* PAS = Probabilistic Amplitude Shaping, aka Probabilistic Constellation Shaping or PCS, aka RateX
** ATSC = Advanced Television Systems Committee
*** DVB = Digital Video Broadcast
**** DM = Distribution Matching
Main Goal

Review an architecture for higher-order modulation and shaping that approaches Shannon capacity and is simple and flexible.
Goal: transmit efficiently, quickly, reliably over a noisy channel

What is the “noisy channel”? Pragmatic answer*: the part of a communication system one is unable or unwilling to change

Digital and probabilistic answer:
- Quantize dimensions (time/frequency) and values into n symbols \( x = x_1, \ldots, x_n \) and \( y = y_1, \ldots, y_n \) (just think of way-oversampling)
- “Noisy channel” is a conditional distribution \( P( y | x ) \)

* Definition attributed to John L. Kelly
For large \( n \), \( I(X;Y) \) gives the number of bits one can transmit reliably for any permissible blocks \( X,Y \).

Let \( d \) be the number of “dimensions” used; usually \( d = T \cdot B \) where \( T \) and \( B \) are the respective time and bandwidth “consumed”.

Spectral efficiency is the maximum \( I(X;Y)/(T \cdot B) \) under the constraints.

Capacity is terminology that is overloaded, e.g., it can mean:

- the spectral efficiency;
- the maximum rate in bits/second;
- the maximum rate bits/channel use;
- and so on.
Consider complex-alphabet, circularly-symmetric, AWGN channel

\[ Y = \left( X_i + jX_Q \right) + \left( Z_i + jZ_Q \right) \]

with \( Z_i \) and \( Z_Q \) independent, \( \text{Var}[Z_i] = \text{Var}[Z_Q] = N/2 \), and input power constraint \( E[|X|^2] \leq P \) has \( \text{SNR} = P/N \) and

\[ I(X;Y) \leq C = \log(1 + \text{SNR}) \]

Best \( X \) is **Gaussian**, 0-mean, circularly symmetric with \( E[|X|^2] = P \)

Consequence: should mimic **Gaussian** distributions with the available discrete amplitude/phase signals

For other channels: same idea but with another distribution
The set of discrete amplitude/phase signals for X is the modulation set.

- **5 bits/symbol**
  - 32-QAM
  - 64-QAM

- **6 bits/symbol**
  - 64-QAM

- **7 bits/symbol**
  - 128-QAM

- **8 bits/symbol**
  - 256-QAM

- **9 bits/symbol**
  - 512-QAM

- **10 bits/symbol**
  - 1024-QAM

QAM=Quadrature Amplitude Modulation
Uniform QAM has 1.53 dB (≈30%) energy gap* to C at high SNR

Solutions: use geometric shaping and/or probabilistic shaping to mimic Gaussian distributions

* gap may be larger for non-linear channels
Geometric vs. Probabilistic Shaping

- **Geometric shaping**: use uniform probabilities, change point positions
- **Probabilistic shaping**: use QAM positions, change point probabilities
- We consider the latter, which is usually better (see next page);
  Probabilistic shaping is also used in “classic” modem standards
SNR gap* to C for ATSC 3.0 (2016 proposed standard) with 46 modcods vs. 1 PAS modcod with 256-QAM and DVB S2 rate 5/6 code (same n, #it)

* Figure from F. Steiner, G. Böcherer (arXiv Aug. 2016)
2) PAS Transmitter Architecture

- Coding & modulation methods (there are more): TCM*, BICM*, multilevel-coding & multi-stage decoding, direct mapping, etc.
- Shaping methods (there are more):
  1) 1960s: many-to-one mapping
  2) 1980s: trellis shaping
  3) 1990s: shell mapping (ITU-T V.34), geometric shaping
  4) 2000s: superposition and multi-stage decoding
  5) 2000s: concatenated shaping
  6) 2010s: iterative polar mappings
  7) 2010s: bootstrap scheme
  8) 2015: Probabilistic Amplitude Shaping (PAS)**

* TCM = trellis coded modulation, BICM = bit interleaved coded modulation

Layered (shaping-coding-modulation) with **systematic** encoding

- **Important features:**
  1. **Performance:** approach AWGN capacity including shaping
  2. **Flexibility:** tune rate via shaping, labeling, coding, or modulation; and layering lets you choose your favorite DM and code algorithms
  3. **Simplicity:** complexity similar to BICM
Information bits mapped to $2^{m-1}$ shaped amplitudes $A_i$ via DM*.

- $A_i$ mapped to bits $b(A_i)$ and encoded with a binary matrix $P$: parity-forming part of a rate $(m-1)/m$ systematic generator matrix of a (good) code, e.g., turbo, LDPC, polar, spatially-coupled.

- Coded bits mapped to sign bits $S_i$.

- To increase rate, some inform. bits may be encoded without shaping.

* E.g., constant-composition DM (Schulte-Böcherer, 2016), shell mapper, etc.

More generally, the numbers may represent not only (1D) amplitudes but (xD) points.
- Encoder matrix \( \mathbf{P} \)
  - gets \((m-1)n\) bits from amplitude rail and \(\gamma n\) bits from sign rail
  - puts out \((1-\gamma)n\) sign bits with approximately uniform distribution
- Code rate: \((m-1+\gamma)/m\) with \((m-1+\gamma)n\) “systematic” and \((1-\gamma)n\) coded bits
- Overall rate including shaped modulation: \(H(A) + \gamma\)
Matching Code to Channel

Notes
- overall rate $H(A) + \gamma$ must be less than $I(X ; Y) = I(AS ; Y)$
- e.g.: 8-ASK, $m=3$ (results in 64 QAM)
- example curves: $\gamma=0$ and $\gamma=1/4$
- rates modified by changing $\gamma$ and $P_A$, which changes $H(A)$

Figure courtesy of G. Böcherer
Goals:

- same decoding complexity as BICM: must use per-bit processing or bit-metric decoding (BMD) that loses information rate
- approach capacity: must carefully design modulation labeling

BMD:

```
Channel       Demod.       Decoder
```

1 real number per bit

a “natural” labeling:  000  001  010  011  111  110  101  100
BRGC* labeling:       000  001  011  010  110  111  101  100

-7 -5 -3 -1 +1 +3 +5 +7

* BRGC = binary reflected Gray code
BMD Rates

- BMD processing: for each bit $B_i$ of a modulation symbol $Y$, compute the log-likelihood ratio (LLR):

$$L_i = \log \frac{P_{B_i}(0)}{P_{B_i}(1)} + \log \frac{P_{Y|B_i}(Y|0)}{P_{Y|B_i}(Y|1)}$$

- An achievable BMD rate is

$$R_{BMD} = \max \left[ 0, H(B_1 \ldots B_{R_m}) - \sum_{i=1}^{R_m} H(B_i|Y) \right]$$

which is (at most but usually) less than $I(X;Y)$

- $R_{BMD}$ becomes the “BICM capacity” if bit levels are independent

- Dependent bit levels reduce rate but can reduce symbol energy. Dependence is generally beneficial.

* G. Böcherer (arXiv 2014), a “twice-generalized” mutual information: GMI 2.0
4) Applications (ATSC, DVB, Optical Fiber)

- **DVB**: digital video broadcast codes, **RateX**: PAS with one code

![Graph showing SNR vs. Rate with various FER values for different modulation schemes.](image)

- Shannon’s log\((1+\text{SNR})\) bound
- **DVB-S2X**: 116 modcods with 8 modulations, 55 codes
- **RateX**: 8/9 DVB-S2 LDPC code, 4096 QAM

*Figure courtesy of G. Böcherer*
Application to Optical Fiber

- **Experimental rates** for DP64-QAM; rate/distance gains small/large

Figure courtesy of G. Böcherer

* F. Buchali et al. (JLT, April 2016)
With code: distance gain of 2 – 4 loops of 240 km each
5) Product DM

- **Idea**: generate target distributions over large alphabets by combining outputs of parallel DMs with smaller (e.g., binary) output alphabets

- **Advantages:**
  - high-throughput architecture
  - enables PAS for OFDM:
    share DMs for lower bit-levels among different sub-carriers; this improves the power efficiency significantly as compared with shaping per sub-carrier

6) Conclusions

- PAS architecture delivers
  - **Performance**: approach AWGN capacity (other methods work too)
  - **Flexibility**: can finely tune rate via distribution matcher
  - **Simplicity**: layered & systematic encoding; bit metric decoding

- Architecture is already part of industrial communication systems:
  - PAS/PCS is on the roadmap for submarine link upgrades
  - 5G: **flexibility** is key to support a **rich diversity** of channels & devices

- Credits to the LNT Coding Team:
  - “Old”: Georg Böcherer, Patrick Schulte, Fabian Steiner
  - New: Peihong Yuan, Marcin Pikus, Tobias Prinz
  - Newer: Antonia Wachter-Zeh (TT Prof), Thomas Jerkovits
  - Support: Gianluigi Liva & his DLR Team
Notes
- BMD with BRGC is within 0.01 dB of C
- similar gaps for 4-ASK to 64-ASK

Figure courtesy of G. Böcherer

\[ \frac{1}{2} \log_2(1 + P/1) \]

- BRGC
- natural-based labeling
64-QAM Shaping Distributions

Figures courtesy of G. Böcherer