(article starts on next page)

Technologies toward Implementation of Probabilistic Constellation Shaping

Tsuyoshi Yoshida[1], Magnus Karlsson[2], and Erik Agrell[2]

[1] Mitsubishi Electric Corporation, Ofuna 5-1-1, Kamakura, Kanagawa, 247-8501 Japan
   Yoshida.Tsuyoshi@ah.MitsubishiElectric.co.jp
[2] Chalmers University of Technology, SE412-96, Gothenburg, Sweden

Abstract *Reverse concatenation of forward error correction and distribution matching significantly improves the implementation capability of probabilistic constellation shaping. However, to take full advantage of the benefits, one should carefully understand the practical aspects and trade-offs.*

Introduction

To handle a growing traffic, high throughput and high-spectral efficiency are required in optical communications. Bit-metric decoding and probabilistic constellation shaping (PCS)[1] with reverse concatenation, meaning that forward error correction (FEC) encoding/decoding is inside the distribution matching/dematching (DM/invDM), have shown good performance close to the Shannon limit. This scheme does not need iterative demapping or nonbinary soft FEC coding, which requires too complex circuit resources that are not widely employed in practice. The reverse concatenation simplifies invDM by using binary operations instead of multi-bit-resolution log-ratio of likelihoods or probabilities (L-value) operations. Very good performance is achieved in offline evaluations with high-gain FEC and rate-loss-less DM for PCS. However, in practice we will face complexity and power consumption issues. Thus there is a tradeoff between implementation capabilities and performance. In this paper, we review the practical boundary of achievable rates and complexity of the DM and FEC. A particularly serious problem is error rate increase in the reverse concatenation architecture compared with the non-shaped case, and its management by high throughput DM will be explained.

System model

We consider the system model and the corresponding performance metrics/monitors shown in Fig. 1. For ease of explanation, we here assume pulse amplitude modulation (PAM) and its extensions to in-phase/quadrature modulation and dual polarization. At the transmitter side, the incoming client signal is converted to a framed signal, e.g., following the $n \times 100$G optical transport units (OTUCn)[2] of a modern optical transport network (OTN) standard protocol[2,3]. The output bit sequence is $A \in \mathcal{B}^{N_{\mathrm{OTU}}}$, where $\mathcal{B} \in \{0,1\}$ and N_{OTU} is the block length of OTUCn, i.e., $130560 \times n$. Its length is converted (in a bullet shown in Fig. 1) and a uniformly distributed bit
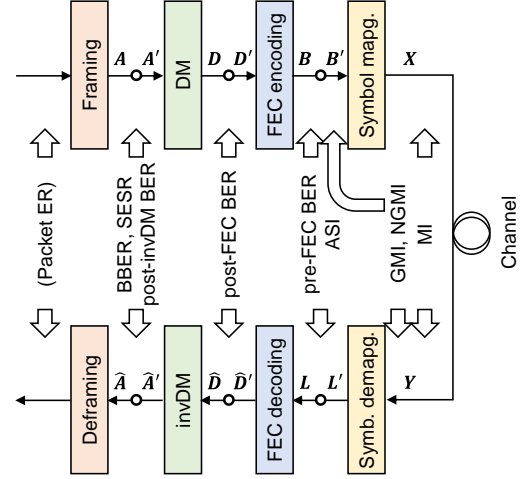

Fig. 1. System model of a communication system with reverse concatenation based PCS, including key performance metrics.

sequence $A' \in \mathcal{B}^{\sum_{i=1}^m N_i}$ is changed to a shaped symbol sequence $X \in \mathcal{X}^{N_s}$ inside the DM and a corresponding binary output sequence $D \in \mathcal{B}^{mN_s}$, where $\mathcal{X}$ denotes a 2^m-PAM symbol (including placeholders for FEC parity bits to be determined in the next block) and N_i is the DM input information length for bit level $i \in \{1,2,\cdots,m\}$ of the PAM symbol per DM-word. N_s is the DM-word length (or DM output block length). A systematic binary FEC encoder generates $n_c - k$ parity bits from the incoming k payload bits $D' \in \mathcal{B}^k$, and outputs the FEC codeword $B \in \mathcal{B}^{n_c}$. The bits $B' \in \mathcal{B}^m$ are converted to a transmitted symbol $X \in \mathcal{X}$.

At the receiver side, the received symbol Y is demapped by bit-metric decoding to bit-wise *a posteriori* L-values $L' \in \mathcal{L}^m$. The FEC then decodes from the L-values $L \in \mathcal{L}^n$ to $\widehat{D'} \in \mathcal{B}^k$. The decoded bit sequence $\widehat{D} \in \mathcal{B}^{N_s}$ is dematched to $\widehat{A'} \in \mathcal{B}^{\sum_{i=1}^m N_i}$, and $\widehat{A} \in \mathcal{B}^{N_{\mathrm{OTU}}}$ is deframed.

Performance metrics and monitoring

The optical physical layer historically used a target bit error rate (BER) down to 10^{-15} following the standard[4]. There are several achievable information rates (AIRs) that describe capacity bounds. The generalized mutual information (GMI) is an AIR for bit-interleaved coded modulation with uniform (non-shaped) and

independent signalling. Here normalized GMI (NGMI, GMI/m) works as a good FEC estimate[5] for error-free operation such as a post-FEC BER of <10^{-15}. There is a rate loss due to the non-ideal FEC; the required NGMI must therefore be larger than the FEC code rate R_c[5, Tab. III].

For reverse concatenation PCS, the AIR is $R_{\text{bmd}}^{\text{ps}}$ for probabilistic amplitude shaping (PAS)[1,6]. $R_{\text{bmd}}^{\text{ps}}$ takes the same value as the GMI calculated for the PAS architecture[6-8] $\mathbb{H}(X) - \sum_{i=1}^{m} \mathbb{H}(B_i|Y)$ or $\mathbb{H}(X) - \sum_{i=1}^{m} \mathbb{H}(B_i|L_i)$. The information rate is $\mathbb{H}(X) - (1 - R_c)m$ with ideal DM, where $\mathbb{H}(X)$ is the symbol entropy calculated from the probability mass function of the symbol X. The true information rate is decreased by the DM rate-loss $\Delta = \mathbb{H}(X) - \sum_{i=1}^{m} N_i/N_s - (1 - R_c)m$. As FEC threshold metrics, the NGMI for PAS[7] or asymmetric information (ASI)[8] are useful.

While the received symbol Y is usually used for the calculation of GMI or NGMI, the L-values can be used instead when the channel assumed in soft symbol demapping (auxiliary channel) is matched to the true channel. Then ASI is an equivalent metric to NGMI[8]. The practical circuits will have a minor information loss in the soft demapping due to quantization and other approximations. GMI or NGMI cannot take the loss into account, but ASI can because it is defined by the L-value just before FEC decoding. The post-FEC BER can be estimated by using the metric; however, the post-invDM BER requirement must be <10^{-15} in contrast with the post-FEC BER, due to a potential error rate increase in the invDM.

Another point to note is that OTN framing does not evaluate post-FEC BER but background block error rate (BBER) and severely errored second rate (SESR)[2,3]. BBER and SESR are calculated from block errors, and the observation period is different. We will discuss only BBER from now for simplicity. Typical BBER requirement is on the order of 10^{-7}. For reverse concatenation PCS, a large N_s can cause a long error burst resulting from the residual error at the FEC decoder output, though the large N_s reduces the DM rate-loss Δ. Usually the BER after the invDM is higher than that after the FEC decoder as stated above. The behavior of block performance monitors is critical for system design, and we will show simulation results for them below. The upper-layer packets have smaller or larger lengths than the OTN frame block length, so BBER is not sufficient, but both post-invDM BER and BBER should be considered.

Issues on hardware implementation
A state-of-the-art DM performance is given by m-out-of-n codes[9] or constant composition DM (CCDM)[10]. When N_s is approximately infinite, the rate loss Δ approaches zero, but the digital signal processing requirements will be prohibitively complex. Recently, complexity-reduced implementations have been studied[11–13].

FEC has a big drawback of throughput increase due to the PCS. Here we compare two cases, non-shaped and reverse concatenation PCS, assuming the same client rates, FEC code rates R_c, and symbol rates. We define m_u and m_s as m for the non-shaped (uniform) and the shaped cases, resp. The shaped case has m_s/m_u times larger FEC throughput than the non-shaped case, because the bit rate is increased by DM. Thus the FEC circuit size (and power consumption) will be higher for the non-shaped case. Further, the code rate with PAS must satisfy $R_c \geq (m_s - 1)/m_s$, because the parity bit is uniform and degrades the probability mass function if put on a shaped bit level. Generalization of PAS[13] can place the parity outside the sign bit, which relaxes the R_c constraint.

Efficient DM implementation example
We propose a low-power DM (and invDM) based on a novel multi-layer symbol selection method. In this scheme, DM having N_s of 100~1000 can be realized with feasible look-up table sizes, with DM and invDM having maximum address size less than 10k, rather than 2^{100} addresses. We do not need complex operations like integer additions or multiplications. Advantages of using the look-up table include making the circuit low power and flexible. Either bit-wise[9] or symbol-wise[10] DMs are selectable, and the information rate is adjustable with high granularity. This DM achieves a high throughput because the architecture consists of a fully parallelized input/output configuration, as well as a bit-scramble selector or permutation mapper[14]. The layered operation is fully pipelined, so just a small number of instances of it is required, and the latency is <10 clock cycles. On the contrary, m-out-of-n codes or CCDM require high-precision integer multiplications, so its throughput is smaller, and a larger number of instances (at least ~N_s) is required. Then the equivalent output block length $N_{s,\text{eq}}$ becomes ~N_s^2.

Simulation
We simulated PCS-256-quadrature amplitude modulation (QAM) transmission over the Gaussian channel. A DVB-S2 low-density parity check FEC code[14] was used, whose R_c, codeword length, and decoding iteration were 5/6, 64800, and 20, resp. The soft-demapping input/output interfaces were quantized using 7 and 4 bits, resp., and the ASI reduction due to this

means <0.1 dB required SNR increase. The most significant ($i = 1$, sign-bit) and least significant ($i = 4$) bit levels for the 16-PAM symbols were not shaped. CCDM having $\sum_{i=2}^{3} N_i/N_s$ of 1014/640 or the proposed DM having $\sum_{i=2}^{3} N_i/N_s$ of 507/320 was applied for bit levels $i = 2, 3$. The information rate was 2(2+507/320) – (1 – 5/6)8 ≥ 5.83 b/channel use, which was set to the equivalent of 128-QAM with R_c = 5/6. $N_{s,eq}$ for CCDM and the proposed DM are 518400 ($> N_s^2$) and 320, resp. While 810 DM words are mapped to 32 FEC codewords in parallel in the case of CCDM, 50 or 51 DM words are mapped to a FEC codeword sequentially in the proposed DM.

Fig. 2 shows the error rates as a function of SNR. Interestingly, the post-FEC BER for low BERs does not depend on the bit level so much due to the bit-level mapping optimization of FEC codewords[1]. CCDM shows 0.13 dB lower required SNR at the post-FEC BER of 10^{-6} than the proposed scheme. The error rate increase by invDM can be characterized by the post-invDM BER to post-FEC BER ratio, which are 210 and 8 for CCDM and the proposed scheme, resp. These ratios are assumed to prevail at lower BERs like 10^{-15}, because the values well correspond to the value when assuming there is a single error in each DM word with some probability. The expected post-invDM BERs for the erroneous DM word are 0.5 over 1014 bits for CCDM and <0.04 over 507 bits for the proposed DM. To satisfy the post-invDM BER of $<10^{-15}$, the post-FEC BER must be 5×10^{-18} and 1.2×10^{-16}, resp. The FEC design and evaluation for the BER 5×10^{-18} would be tough and simulation may be impossible. If an error floor exists between 10^{-15} and 5×10^{-18}, the required SNR will be significantly larger for post-invDM BER 10^{-15}. The proposed scheme is >10 times better in this viewpoint. To quantify the burst error after invDM, the ratio of BBER to post-invDM BER is useful, which are 2000 and 19000 for CCDM and the proposed DM, resp. If this ratio is smaller, the erroneous frame has a larger number of errors, so the error burst is larger. The proposed DM's error distribution is ~10 times more random. With the proposed DM, there is a possibility to concatenate hard-decision FEC outside the DM/invDM due to this random error feature.

Conclusions

We have summarized the practical aspects of probabilistic constellation shaping. The proposed DM is good for hardware implementation, at the expense of 0.13 dB larger required SNR than CCDM, for a post-FEC BER of 10^{-6}. When the FEC decoder misconverges, the proposed invDM outputs smaller and more randomly distributed
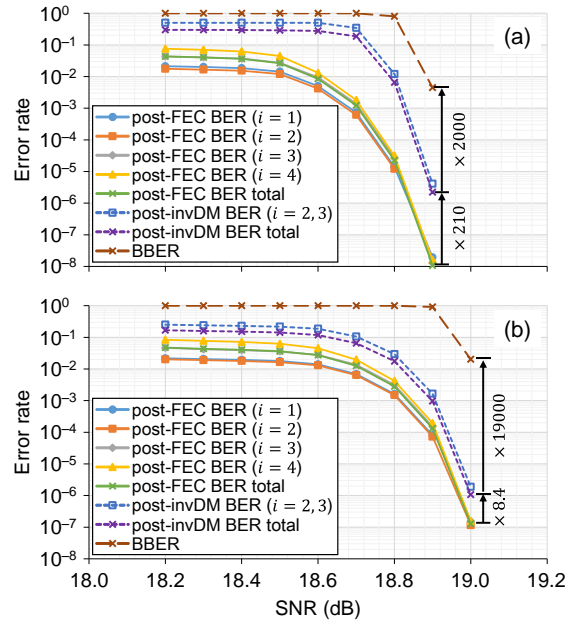

Fig. 2. Simulated error rates as a function of SNR for PCS-256QAM with (a) CCDM or (b) the proposed DM.

errors compared with the CCDM.

References

[1] G. Böcherer et al., "Bandwidth efficient and …," IEEE Trans. Commun., **63**(12), p. 4651, 2015.

[2] ITU-T, "Interfaces for the …," 2016. [Online]. Available: www.itu.int/rec/T-REC-G.709

[3] ITU-T, "Error performance parameters …," 2011. [Online]. Available: www.itu.int/rec/T-REC-G.8201

[4] ITU-T, "Forward error correction …," 2000. [Online]. Available: www.itu.int/rec/T-REC-G.975

[5] A. Alvarado et al., "Replacing the soft …," IEEE/OSA, J. Lightw. Technol., **33**(20), p. 4338, 2015.

[6] G. Böcherer, "Achievable rates for …," [Online]. Available: www.arxiv.org/pdf/1707.01134v4, 2017.

[7] J. Cho, et al., "Normalized generalized mutual …," ECOC, 2017, p. M.2.D.2.

[8] T. Yoshida et al. "Performance metrics for …," IEEE Photon. Technol. Lett., **29**(23), p. 2111, 2017.

[9] T. V. Ramabadran, "A coding scheme …," IEEE Trans Commun., **38**(8), p. 1156, 1990.

[10] P. Shulte and G. Böcherer, "Constant composition …," IEEE Trans. Inf. Theory, **62**(1), p. 430, 2016.

[11] M. Pikus and W. Xu, "Bit-level probabilistically …", IEEE Commun. Lett., **21**(9), p. 1929, 2017.

[12] G. Böcherer et al., "Fast probabilistic shaping …," ECOC, 2017, Tu.2.D.3.

[13] T. Yoshida et al., "Short-block-length shaping by …," ECOC, 2017, p. Tu.2.D.2.

[14] ETSI, EN 302 307-1 V1.4.1, 2014. [Online]. Available: www.dvb.org/standards