Rate Loss Reduction through Look-up Table Design for Hierarchical Distribution Matcher in Probabilistic Amplitude Shaped systems

Pantea Nadimi Goki(1), Luca Potì(2)

(1) TECIP Institute Scuola Superiore Sant'Anna, 56124, Pisa, Italy, pantea.nadimigoki@santannapisa.it
(2) CNIT Photonic Networks and Technologies National Laboratory, 56124 Pisa, Italy, luca.poti@cnit.it

Abstract *We proposed a new method for the rate loss reduction in the hierarchical distribution matching (Hi-DM) approach for the constellation shaping. This method makes Hi-DM comparable with the enumerative sphere shaping (ESS) approach, at short block length, in terms of rate loss.*

Introduction

Probabilistic amplitude shaping (PAS) has attracted a lot of attention in fibre optic communication in the last few years. Its feature of rate adaptively and its ability to providing shaping gain to diminish the gap to the Shannon capacity, compared to the uniform QAM (U_QAM) constellation, make it in the center of attention in high data rate transmission[1],[2]. It also has the capability to integrate shaping into existing forward error correction (FEC)[3]. While in U_QAM, each constellation point is sent with equal probability, in PAS_QAM a Maxwell Boltzmann (MB) distribution is targeted for constellation symbols to maximize entropy for a given average energy[4], where the low energy symbols transmit more frequently than high energy ones. Applying MB distribution, known as the discrete counterpart of the Gaussian distribution, enables the rate maximization for a fixed average power for an additive white Gaussian noise (AWGN) channel [5]. Several methods have been proposed to realize the PAS constellation, which can be separated into two groups based on their approach, direct (distribution matching) and indirect (Sphere shaping) [6]. The distribution matcher (DM) is the essential component of the direct PAS system, which maps *k* uniformly distributed input bits to the *N* shaped amplitudes, concerning the target distribution, with rate *R=k/N*. The constant composition distribution matcher (CCDM)[7],[3] and Multiset partition DM (MPDM)[8] are the known shaping architectures using DM. The CCDM proposed in[3] is an invertible technique, based on an appropriate concatenation of a fixed to fixed-length DM, in which the shaping blocks precede the binary FEC coding. The MPDM uses different compositions and employs short-length DMs at a particular rate loss in comparison to the CCDM[8]. Recently an effective DM architecture has been proposed based on combination of several small look-up-table (LUT) blocks in a hierarchical structure, called hierarchical DM (Hi-DM)[9], according to this architecture, a generalized hierarchical DM approach has been proposed lately[10],[11], in which the DMs of any kind (e.g., Hi-CCDM) can be combined in a hierarchical structures. LUT-based Hi-DM allows for a fully parallelized implementation. The LUTs can be generated to minimize the average symbol energy, and a careful design may provide a reduction of rate loss and an improvement in overall performance. In this paper, we propose a method for LUT design in Hi-DM proposed in[11] capable of reducing rate loss. We show that the proposed strategy makes Hi-DM comparable with the enumerative sphere shaping (ESS) in terms of rate loss, at short block length performance. Finally, we investigate Hi-DM performance with the proposed LUT design based on required OSNR.

Hi-DM Structure

We follow the same notation and the same DM structure/algorithms proposed in[11] for LUT-based Hi-DM, in which the Hi-DM includes several layers. Each layer consists of a set of small LUTs containing disjoint sequences of amplitudes sorted by energy. However, the amplitudes from the M_QAM alphabet ($\mathcal{A}$={$1,3,\ldots,2M-1$}) are located only in the inner layer, and the set of small LUTs of each layer plays the alphabet role for the following layer (virtual alphabet). Therefore, in each upper layer, the LUTs consist of sequences of the LUTs taken from its lower layer, sorted by increasing mean energy. While the number of LUTs in each layer can be chosen based on system requirements, the last layer always consists of a single LUT. Fig. 1 shows an example of Hi-DM structure with 2 layers LUT-based DM with reference encoding method. $D_j^{(l)}$ defines the j^{th} DM (LUT) in layer 'l', with k_l input bits and N_l output amplitudes, M_{shl} is the number of LUTs in layer 'l' which corresponds to the number of virtual alphabets of the layer '$l+1$', equivalent to M_{l+1}.

In the example, the mapping process starts from

978-1-6654-3868-1/21/$31.00 ©2021 IEEE

the upper layer, in which $k_2 = 2$ input bits are

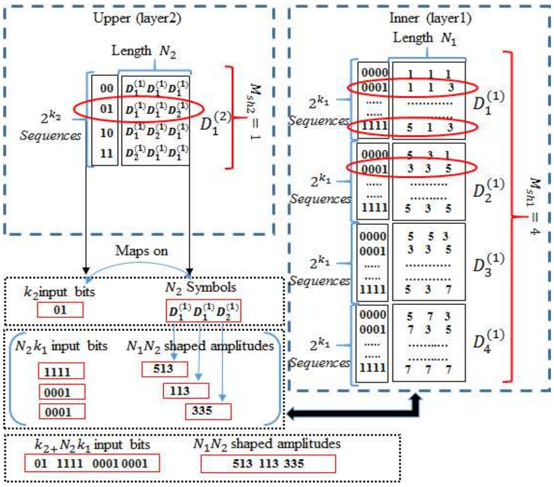

Fig. 1: Hi-DM Structure with 2layers LUT-based DM, maps (k_2+N_2k_1) input bits on (N_1N_2) shaped amplitudes.

mapped on $N_2 = 3$ output virtual amplitudes, as shown in Fig.1.

The 2 input bits (01) encode the sequence $\left\{D_1^{(1)}D_1^{(1)}D_2^{(1)}\right\}$ of $D_1^{(2)}$that will be used for mapping the rest of the input bits in the first layer. Therefore, each $k_1 = 4$ input bits will encode a specific sequence of $D_j^{(1)}$ where $j = 1,1,2$.Thus, layer one takes $N_2k_1 = 12$ bits and maps them on $N_1N_2 = 9$ shaped amplitudes. N_1N_2 gives the overall DM block length.

In this example, the 2-layers DM maps $k_2 + N_2k_1 = 14$ input bits on the $N_1N_2 = 9$ output amplitudes, with rate *R=1.55* and entropy *H= 1.68*, which gives a rate loss *R*loss=*H*-*R*= *0.13*. It can be noticed that finite-length DMs if they are more suitable for practical implementations, they still give high rate loss in particular when short block lengths are used. On the other hand, LUT design of an *'L'* layers structure may act on the following characterization vectors: K=(k_1,…,k_L) input bits, N=(N_1,…,N_L) output amplitudes, and M=(M_1,…,M_L) alphabet size of each layer, where M_1 is the size of ($\mathcal{A}$). Different combinations of these vectors, with the same rate and the same DM block length, could result in various entropies and consequently different rate losses. In the example of Fig.1: K=(*4,2*), N=(*3,3*), M=(*4,4*) giving *R*loss=*0.13,* while other combinations may reduce *R*loss. Our method returns an optimal combination of characterization vectors that reduces rate loss.

LUT design

To realize an invertible and fixed-to-fixed-length Hi-DM structure the following conditions are applied[11].

In each layer all DMs:

1. are unique;
2. have the same size, i.e., the same block length (N_l) and the same number of sequences (2^{k_l});
3. take the same number of input bits ($\mathrm{k_l}$);
4. $M_{l+1}2^{k_l} \le M_l^{N_l}$ that ensures a sufficient number of sequences for filling all DMs.

In order to design LUTs we impose the following additional constraints:

5. $M_l = 2^x = m \;\forall\; l \neq 1, m\; is\; fixed$;
6. $N_1 > N_l = n \;\forall\; l \neq 1, starting\; with\; n = 2.$

The first constraint ensures equal alphabet dimension for all layers (except the first layer). Based on the second restriction, most of the amplitudes must be reserved to the first layer and the rest of them will be equally distributed for all the other layers. Assigning long block length to the first layer ($N_1 > N_l$) and a high number of virtual alphabets to the following layer (M_2) allows to map bits on long sequences of amplitudes with low energies. Hence, those sequences with a high probability of smaller amplitudes, achieve probability closer to the target MB distribution. Consequently, it results in a reduction of the rate loss. Indeed, increasing the number of alphabets in the upper layer (M_{l+1}) causes enhancement in the number of sequences in the lower layer ($M_{l+1}2^{k_l}$). Because of the full parallelization specification of the Hi-DM structure performance, the layer *'l'* in the structure will use $\boldsymbol{T_l} = \prod_{h=l+1}^{L} N_h$ times, with $T_L = 1$, therefore the overall number of input bit is $k = \sum_{l=1}^{L} k_l \prod_{h=l+1}^{L} N_h$, and the overall block length is $N_O = \prod_{l=1}^{L} N_l$ [11]. Giving short and equal block lengths to all layers ($N_l = n \;\forall\; l \neq 1$) helps to control the speed of changing rate in the upper layers. The number of selected sequences used for mapping input bits (index sequences) from layer *'L'* to $'l = 2'$ increases from n^0 to n^{L-1} by $n^{L-(l-1)}$ (where $n = N_{l\geq 2}$). Accordingly, we can easily replace $\boldsymbol{T_l}$ with $\boldsymbol{T_l} = n^i : i = L - l,\; l = \{1,2,..,L\}$, thus the overall number of input bits is $k = \sum_{l=1,i=L-l}^{l=L} k_l n^i$. Furthermore, imposing these additional constraints make condition 4 easier, especially at short block lengths. With the mentioned constraints, it is possible to reduce *R*loss without increasing the block length. As illustrated in Fig. 2 (a), increasing the #DMs/layer (*m*) leads to a decrease in the *R*loss, while the overall block length remains constant at a given rate. For instance, a 7layers structure with a DM rate of 0.70 and M_1=2, increasing #DMs from 64 to 128 results in a 22% rate loss reduction, while block length (N=768) is kept constant. On the other side, it was shown that increasing the number of

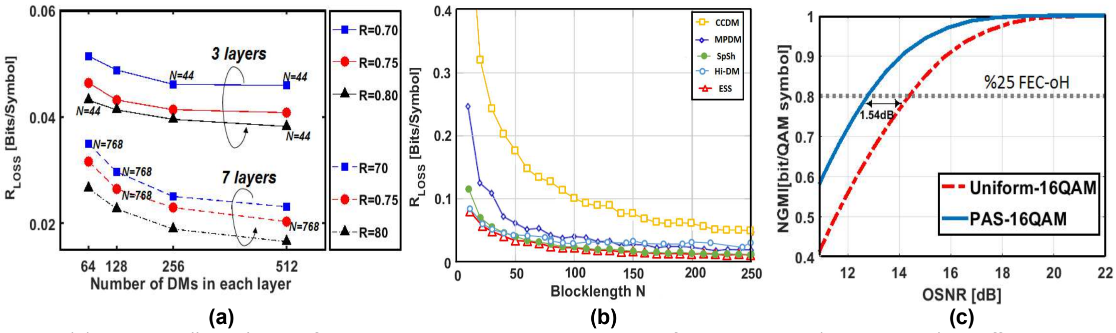

Fig. 2: (a) R_{loss} vs. #DMs/layers for 3and7 layers with some example of unchanged N (block length) at different R_{loss}. (b) R_{loss} vs. #block length (N) for different shaping Architectures. (c) NGMI vs. OSNR for U-16QAM and PAS-16QAM

layers causes enhancement in the overall block length and helps to obtain the lower R_{loss}. For instance, for DM rate of 0.70 and #DMs/layer =64, increasing the number of layers from 3 to 7 results in a 32% rate loss reduction. With this proposed method, it is also possible to find the optimum number of overall input bits and overall block length at a given target rate to achieve low rate loss by choosing m,n, and #layers. K is then obtained based on the mentioned parameters. Fig.2 (a) shows the impact of #DMs/layer on the R_{loss} for the 3 and 7layers structures of PAS-16QAM, with $N_1 \leq 12$ and $N_{l\geq 2} = 2$. Using this method, the R_{loss} of 7 layers structure defined in sec.6.3 of [11], reduces by about 10%, where the optimum overall block length for the given rate (R=1.58) found to be $N_O = 384$, with the following combination:
M=(4,64,64,64,64,64,64), **N**=(6,2,2,2,2,2,2), and **K**=(5,5,4,4,4,4,8). Noteworthily, one can calculate the overall block length (N_O) for any $L > 3$ starting from $N_O|_{L=3}$:

$$N_O = N_O\big|_{L=3} \times N_{l\geq 2}^{(L-3)} \quad (1)$$

This is true since $N_O|_l = N_O|_{l-1} \times N_{l\geq 2}$. The DM rate approximation is limited to $\pm e$ (e.g.,± 0.01), providing minimum R_{loss}. Moreover, by having the optimal combination of only 3layers structure, the *'L'* layers structure can be generated by using the *'loop of the second layer'* strategy, in which the overall rate remains constant only if $N_{l\geq 2} = 2$ and $k_2 = k_3/2$, which ends in $R_2 = R_3$, where R_l is the rate of layer *'l'*. Therefore the new combination is, **M**=(M ,m,...,m), **N**=(N1,2,…,2), **K**=(k1,k2,…,2k2) with the overall block length (N_O) obtain by (1). This method can be evaluated by the following series, which converges to one by adding the following fraction, which is defined based on the #layers (L):

$$\sum_{j=1}^{L-2}\left(\frac{1}{2}\right)^j + \frac{1}{2^{L-2}} = 1 \quad (2)$$

Using a loop of the second layer causes the overall DM rate for *'L'* layers to end in $R = \frac{k_1}{N_1} + \frac{k_2}{N_2}\left(\frac{1}{N_2} + \frac{1}{N_2^2} + \cdots + \frac{1}{N_2^L} + \frac{1}{N_2^L}\right)$, where the series in the parenthesis is equivalent to (2) hence the rate remains constant. For instance, the 3layers structure with the **M**=(2,128,128), **N**=(12,2,2), **K**=(3,5,10), and R=*0.66* can easily extend to the 7layers structure with the same rate by 36% rate loss reduction, where all the middle layers use 128,2,5 for the *M, N, k* respectively.

Rate loss and performance results

Fig.2 (b) shows the rate loss vs. overall block length for CCDM, MPDM, sphere shaping (SpSh), ESS[11], and LUT-based Hi-DM designed with our proposed method. For the sake of fair comparison, we duplicated the rate loss results of Example 8 in[6]. We consider the same target rate (1.73) as in[6] for different block lengths of Hi-DM with m_{max}=*64*, L_{max}=5, and M_1=4. Fig.2 (b) shows for the given target rate, at short block length the Hi-DM performance is better than other shaping architectures in[6], using either DMs or trellis, while it is slightly worse than the ESS. Take into account that ESS is referred to as optimized ESS in [11], and SpSh duplicated using a non-optimized ESS implementation [11]. However, at long block length, MPDM and SpSh provide slightly better results than Hi-DM.

We evaluated the performance of Hi-DM structures for the PAS implementation of an AWGN channel in terms of OSNR dependency of NGMI. The Hi-DM structure follows the proposed LUT design with the rate of 0.75 and R_{loss}=0.031. Fig.2 (c) shows the NGMI for the uniform and shaped 16QAM (constellation rate 3.5) with 25% FEC overhead for 211.3Gb/s net data rate. The 1.54dB OSNR gain is obtained with a 38Gbaud symbol rate.

Conclusions

We proposed a method for optimization LUT design in Hi-DM structures, concerning rate loss. We have shown the possibility of R_{loss} reduction without changing block length, constructing large layers structure from a 3layers system, and finding optimal block length based on a given target rate achieving low R_{loss}. We investigated the effect of Hi_DM implementation on an AWGN channel performance in terms of OSNR gain.

References

[1] A. Sheikh, A. G. Amat, and A. Alvarado, “On product Codes with Probabilistic Amplitude Shaping for High-Throughput Fiber-Optic Systems,” *IEEE Commun. Lett*. vol. 24, no. 11, pp. 2406–2410, Nov. 2020.

[2] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, “Rate Adaptation and Reach Increase by Probabilistically Shaped 64-QAM: An Experimental Demonstration”, *J. Light. Technol*. vol. 34, no. 7, pp. 1599–1609, Nov. 2016.

[3] G. Böcherer, F. Steiner, and P. Schulte, “Bandwidth Efficient and Rate-Matched Low-Density Parity-Check Coded Modulation,” *IEEE Trans. Communication.,* vol. 63, no. 12, pp. 4651–4665, Dec. 2015.

[4] T. M. Cover, and J. A. Thomas, “Elements of Information Theory”, Hoboken, NJ, U.S.A.: John Wiley & Sons, 2006.

[5] Y. C. Gültekin, W. J. V. Houtumy, A. Koppelaar, and F, M. J. Willems, “Partial Enumerative Sphere Shaping”, in *Proc. Conf. IEEE 90th Vehicular Technology* (*VTC2019-Fall*)*.,* Sept. 2019.
DOI: 10.1109/VTCFall.2019.8891195

[6] Y. C. Gültekin, T. Fehenberger, A. Alvarado,and F. M. J.Willems, “Probabilistic Shaping for Finite Blocklengths: Distribution Matching and Sphere Shaping”, *Entropy*, vol. 22, no. 5, p.581, May. 2020.

[7] P. Schulte, and G. Böcherer, "Constant Composition Distribution Matching," *IEEE Trans. Info. Theory*, vol. 62, no. 1, pp. 430-434, Jan. 2016.

[8] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Multiset-Partition Distribution Matching," *IEEE Trans. Communication*, vol. 67, no. 3, pp. 1885-1893, March 2019.

[9] T. Yoshida, M. Karlsson, and E. Agrell, "Hierarchical Distribution Matching for Probabilistically Shaped Coded Modulation," *J. Lightw. Technol*, vol. 37, no. 6, pp. 1579-1589, 15 March15, 2019.

[10] S. Civelli, and M. Secondini, "Hierarchical Distribution Matching: A Versatile Tool for Probabilistic Shaping," in *Proc. Conf. Opt. Fiber Commun. (OFC),* pp.1-3*,* 2020.

[11] S. Civelli, and M. Secondini, "Hierarchical Distribution Matching for Probabilistic Amplitude Shaping" *Entropy*, vol. 22, no. 9, p.958, Aug. 2020.