Hierarchical Distribution Matching: a Versatile Tool for Probabilistic Shaping

Stella Civelli[1,2]*, **Marco Secondini**[1,2]

[1] *TeCIP Institute, Scuola Superiore Sant'Anna, Via G. Moruzzi 1, 56124, Pisa, Italy*
[2] *PNTLab, CNIT, Via G. Moruzzi 1, 56124, Pisa, Italy*
**stella.civelli@santannapisa.it*

Abstract: The hierarchical distribution matching (Hi-DM) approach for probabilistic shaping is described. The potential of Hi-DM in terms of trade-off between performance, complexity, and memory is illustrated through three case studies.

OCIS codes: 060.1660, 060.2330, 060.4080

1. Introduction

Recently, probabilistic shaping (PS) techniques have been widely investigated to improve the performance and the flexibility of optical fiber networks. By assigning different probabilities to the constellation symbols—e.g., trying to approximate the capacity-achieving Gaussian distribution for the additive white Gaussian noise (AWGN) channel—PS allows both to finely adapt the information rate (IR) to the available signal to noise ratio (SNR) and to reduce the gap between the IR achievable with uniform QAM constellations and the channel capacity [1, 2].

An effective PS approach, named probabilistic amplitude shaping (PAS) and based on a proper concatenation of a fixed-to-fixed-length distribution matcher (DM) and a systematic forward error correction (FEC) encoder, has been proposed in [3]. The key element of PAS is the DM, which maps k uniformly distributed bits on N shaped (according to a desired distribution) amplitudes from the alphabet $\mathscr{A} = \{1,3,\ldots,2M-1\}$ with rate $R = k/N$. This map induces a specific structure on the output sequence, whose elements are, therefore, not independent. Consequently, the rate R of the DM is lower than the entropy rate $\mathscr{H}$ that would be obtained with a sequence of i.i.d. amplitudes with the same target distribution, yielding the rate loss $R_{loss} = \mathscr{H} - R \geq 0$.

Different methods to realize a DM with a low rate loss have been recently proposed [4–7]. While the rate loss usually tends to zero when the block length N increases (but with different convergence speed for different DMs), this happens at the expense of increased computational cost, memory, and/or latency. To address this issue, an effective solution based on a hierarchical DM (Hi-DM) structure based on look up table (LUT)s has been recently proposed [6, 8]. Elaborating on the latter idea, our work investigates a generalized Hi-DM approach in which several short DMs are hierarchically combined to obtain a longer DM with reduced rate loss and limited complexity, memory, and latency. The proposed approach can be tailored to look for the best trade-off between the mentioned parameters, depending on the system requirements. To show the potential of Hi-DM, we illustrate three different structures.

2. Hierarchical distribution matcher

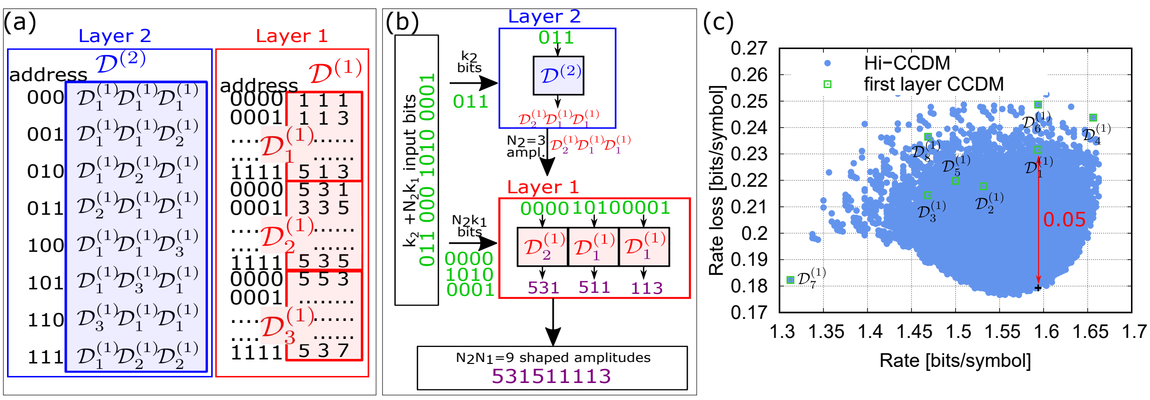

Fig. 1. Example of (a) Hi-DM structure and (b) encoding with 2 layers of LUT, $\mathbf{N} = (3,3)$, $\mathbf{k} = (4,3)$, and $\mathbf{M} = (4,3)$. (c) Rate loss versus rate for a 2-layer Hi-DM using CCDM on both layers.

The Hi-DM has a layered structure. The lowest layer consists of a set of short *disjoint* (i.e., without common output sequences) DMs that encode the input bits on the output levels. In turn, this set of DMs is seen by the upper layer as a *virtual alphabet*. Thus, in the upper layer, another DM is used to encode some additional input bits on the particular sequence of DMs that will be used in the lower layer. If more layers are desired, a set of disjoint DMs (rather than a single DM) is considered also in the upper layer and an additional layer is added. Compared to the single short DMs, the overall structure has a longer overall block length and more degrees of freedom in the generation of the output sequences with the desired target distribution, yielding a reduction of the rate loss with a moderate increase of computational and memory resources.

The working principle is illustrated by considering the 2-layer Hi-DM example in Fig. 1(a). The upper layer (layer 2) is characterized by a DM $\mathscr{D}^{(2)}$ that takes $k_2 = 3$ input bits and maps them to $N_2 = 3$ output symbols. Here, the output symbols are not "conventional" amplitude levels, but are taken from an alphabet of $M_2 = 3$ DMs, $\mathscr{D}^{(1)} = \{\mathscr{D}_1^{(1)}, \mathscr{D}_2^{(1)}, \mathscr{D}_3^{(1)}\}$. The output of layer 2 corresponds to the specific sequence of DMs to be used in layer 1. In turn, layer 1 is characterized by the set of DMs $\{\mathscr{D}_1^{(1)}, \mathscr{D}_2^{(1)}, \mathscr{D}_3^{(1)}\}$, each with $k_1 = 4$ input bits and $N_1 = 3$ output levels taken from the M_1-ary alphabet $\{1, 3, \ldots 2M_1 - 1\}$. Thus, layer 1 takes $N_2 k_1 = 12$ input bits and eventually generates $N_2 N_1 = 9$ output levels. An example of encoding is shown in Fig. 1(b). The use of disjoint DMs in layer 1 is required to ensure that the encoding process can be inverted at the decoder.

With respect to considering a predetermined sequence of DMs in layer 1, which would simply result in a rate $R_1 = k_1/N_1$, the proposed structure allows to encode "for free" k_2 additional bits on the same number of output levels, increasing the overall rate to $R = (k_2 + N_2 k_1)/(N_2 N_1) > R_1$ and potentially reducing the rate loss.

The Hi-DM structure can be extended to L layers and is characterized by the vectors $\mathbf{N} = (N_1, \ldots N_L)$, $\mathbf{k} = (k_1, \ldots, k_L)$, and $\mathbf{M} = (M_1, \ldots, M_L)$. Each layer can use any kind of DM (e.g., constant composition distribution matching (CCDM) [5], enumerative sphere shaping (ESS) [7], or LUT) and, given the properties of the lower layer, can be designed to target the desired distribution. In the following examples, the Hi-DMs structures are designed to minimize the mean energy per symbol for a given rate or, equivalently, to target a Maxwell–Boltzmann distribution [2].

3. Three case studies: structure and performance

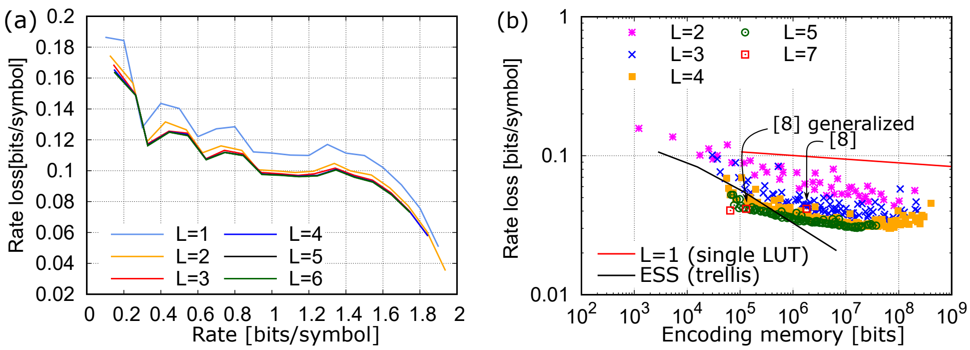

Fig. 2. Rate loss (a) versus rate for Hi-DM using PPM on all the layers except the first, and (b) versus encoding memory for Hi-DM using LUTs on all the layers.

The Hi-DM structure can be used to combine different CCDM on several layers. In this case, the Hi-DM remains fixed-to-fixed length even if the DMs on the same layer have a different number of input bits. The potential of this method is illustrated in Fig. 1(c), which shows the rate loss versus rate that can be obtained with a 2-layer Hi-DM combining $M_2 = 8$ CCDMs on layer 1 (block length $N_1 = 32$, $M_1 = 4$ amplitudes, different rates; $\mathscr{D}_1^{(1)}$ optimized for the target rate of 1.59, the others for different rates)—whose rate loss is reported with green squares—with an external CCDM, with block length $N_2 = 10$. The blue dots show the rate loss versus rate that can be obtained with different CCDM compositions on the second layer. At the target rate, the rate loss decreases of about 23%, from 0.2316 for the 1-layer CCDM to 0.1793 for the 2-layer Hi-DM.

In a second example, we employ the pulse position modulation (PPM) concept to decrease the rate loss with a negligible complexity. Given a set of L disjoint DMs with increasing energy on the first layer, the external layers $\ell = 2, \ldots, L$ are used according to a PPM scheme as follows. The second layer selects $N_2 - 1$ times the lower energy DM $\mathscr{D}_1^{(1)}$, and once $\mathscr{D}_2^{(1)}$, encoding $\log_2(N_2)$ bits on the *position* of the latter. The third layer applies $N_3 - 1$ times the second layer, and once the same structure but replacing $\mathscr{D}_2^{(1)}$ with $\mathscr{D}_3^{(1)}$, encoding $\log_2(N_3)$ bits on its position. This method increases the rate, with only a slight increase of the energy and a negligible additional complexity. The Hi-DM with PPM is investigated considering on the first layer L LUTs with block length $N_1 = 10$, variable

k_1, and $M_1 = 4$ amplitudes. The first LUT $\mathscr{D}_1^{(1)}$ contains the first 2^{k_1} minimum energy sequences, the second LUT $\mathscr{D}_2^{(1)}$ contains the second 2^{k_1} sequences with minimum energy, and so on. Fig. 2(a) shows the rate loss (the minimum for each k_1) for different rates and a different number of layers ($L = 1$ means that only the LUT $\mathscr{D}_1^{(1)}$ is used, without Hi-DM). For a given rate, the rate loss decreases when increasing the number of PPM layers, converging to the minimum value with approximately $L = 3$ layers. As expected, the rate loss vanishes when the rate tends to 2 (uniform constellation), for which no shaping is performed. At most, the rate loss reduction with respect to the single LUT slightly exceeds 0.015, with an improvement of about 17% at rate 1.6.

The proposed Hi-DM structure can use LUTs on all the layers, designing each LUT to minimize the mean output energy for a given rate. In this case, it becomes practically equivalent to the structure considered in [6, 8]. With LUTs, computational complexity is usually negligible and memory becomes the most important resource. The Hi-DM with LUTs is investigated at the target rate $R = 507/320 \approx 1.58$ as in [6, 8]. Fig. 2(b) shows the rate loss versus encoding memory with a different number of layers. The memory is evaluated as the number of bits required to store all the LUTs. A possible parallelization of the whole structure or of some of the inner LUTs (as implicitly done in [6, 8]) may be considered, depending on the system requirements. The figure also reports the memory required by a single LUT approach (red line), and the memory required to store the trellis structure for the ESS approach [7] (black line) which, however, has a relevant computational complexity. Fig. 2(b) shows some of the values with the best trade-off between rate loss and memory for $L \leq 5$, while for $L = 7$ only three values are shown, two of which correspond to the structure proposed in [6, 8], without the implicit parallelization (i.e., implemented according to our structure and labeled " [8] generalized") and as given by the authors (labeled " [8]"). The figure shows that increasing the number of layers, the rate loss significantly decreases with the same overall memory, or, equivalently, the memory required for a certain rate loss can be substantially diminished. Finally, the figure shows that the performance (in terms of rate loss versus memory) of the structure proposed in [6, 8] can be even improved with some structures.

4. Conclusion

We presented a general Hi-DM structure for the effective implementation of PAS. The proposed Hi-DM structure can be tailored according to the system requirements, including target distribution, available hardware resources, and rate loss. We have shown the potential of Hi-DM through three simple examples: i) a two-layer combination of short CCDMs, which reduces the rate loss by approximately 23% compared to a single CCDM with same block length; ii) a combination of a LUT-based sphere shaping layer with several PPM-like layers, which reduces the rate loss of the LUT-based DM by 10% with a negligible additional complexity; and iii) an L-layer combination of LUTs, in which the rate loss can be reduced by increasing the number of layers while keeping the required LUT memory small. In the last example, the LUT-based Hi-DM achieves a rate loss of about 0.04 (for the rate 1.58 on 4 amplitude levels) with less than 100 Kbit of memory.

Acknowledgment

This work was supported in part by Fiberhome. We thank T. Yoshida for sharing the details of the structure in [6,8].

References

1. 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. **34**, 1599–1609 (2016).
2. T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, "On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel," J. Light. Technol. **34**, 5063–5073 (2016).
3. G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," IEEE Transactions on Commun. **63**, 4651–4665 (2015).
4. T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Multiset-partition distribution matching," IEEE Transactions on Commun. **67**, 1885–1893 (2018).
5. P. Schulte and G. Böcherer, "Constant composition distribution matching," IEEE Transactions on Inf. Theory **62**, 430–434 (2016).
6. T. Yoshida, M. Karlsson, and E. Agrell, "Hierarchical distribution matching for probabilistically shaped coded modulation," J. Light. Technol. **37**, 1579–1589 (2019).
7. Y. Gultekin, W. van Houtum, and F. Willems, "On constellation shaping for short block lengths," in *2018 Symposium on Information Theory and Signal Processing in the Benelux (SITB 2018),* (University of Twente, 2018), pp. 86–96.
8. T. Yoshida, M. Karlsson, and E. Agrell, "Preferred design of hierarchical distribution matching," arXiv preprint arXiv:1906.05122 (2019).