MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Huffman Coded Sphere Shaping with Short Length Reduced Complexity

Millar, D.S.; Fehenberger, T.; Yoshida, T.; Suzuki, N.; Koike-Akino, T.; Kojima, K.; Parsons, K.

TR2019-105 September 24, 2019

Abstract

A novel probabilistic shaping architecture is proposed which approximates the optimal sphere-bound while utilizing a Huffman tree and binary mapping/demapping based on look-up-tables. A 100 symbol long DM sequence achieves within 0.2 dB of the infinite length asymptote, and is implemented with a LUT of 100 kbit.

European Conference on Optical Communication (ECOC)

HUFFMAN CODED SPHERE SHAPING WITH SHORT LENGTH AND REDUCED COMPLEXITY

David S. Millar (1), Tobias Fehenberger (1), Tsuyoshi Yoshida(2), Toshiaki Koike-Akino(1), Keisuke Kojima(1), Naoki Suzuki(2), and Kieran Parsons(1)*

(1) Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA, USA.
(2) Mitsubishi Electric Corporation, 5-1-1 Ofuna, Kamakura, Kanagawa, 247-8501 Japan.
**E-mail: millar@merl.com*

Keywords: Coded modulation, Probabilistic shaping

Abstract

A novel probabilistic shaping architecture is proposed which approximates the optimal sphere-bound while utilizing a Huffman tree and binary mapping/demapping based on look-up-tables. A 100 symbol long DM sequence achieves within 0.2 dB of the infinite length asymptote, and is implemented with a LUT of 100 kbit.

1 Introduction

Since the first demonstrations of the probabilistic amplitude shaping (PAS) framework [1] in fiber-optic simulations [2, 3] and experiments [4–7], PAS has been applied to and studied in numerous different settings. A crucial building block of PAS is the distribution matcher (DM) that transforms uniformly distributed input bits into blocks of shaped amplitudes. All fixed-length DMs suffer from rate loss that directly reduces the net data rate of communication systems with probabilistic shaping, and this rate loss generally decreases with block length. In the original PAS paper [1], constant composition distribution matching (CCDM) realized via arithmetic coding [8] is considered as DM. While CCDM is simple to analyse—every output symbol sequence is a unique and equiprobable permutation of the same composition—it suffers from relatively high rate loss for short block lengths. A compounding problem is the fact that the arithmetic coding based implementation of CCDM is inherently serial in the input length. The long block lengths required for low rate loss, and correspondingly high serialism have proven prohibitive for implementation in hardware thus-far.

To reduce the required block lengths, advanced DM systems with variable compositions have been proposed. Multiset-partition distribution matching (MPDM) [9] can be viewed as layered CCDM operations. Other methods that carry out the DM task are enumerative sphere shaping (ESS) [10–12] and shell mapping (SM) [13], or variable-length DM with framing [14]. Improved architectures enabling higher throughput than nonbinary CCDM have also been proposed. In bit-level DM [15–17], the target distribution is factorized such that constituent binary DMs can be run in parallel, which supports only product compositions. The parallel-amplitude architecture of [18] by comparison imposes no constraints on the composition, but may induce a small additional rate loss. For any binary DM scheme, subset ranking (SR) has been proposed in [18] as a low-complexity method for CCDM mapping and demapping.

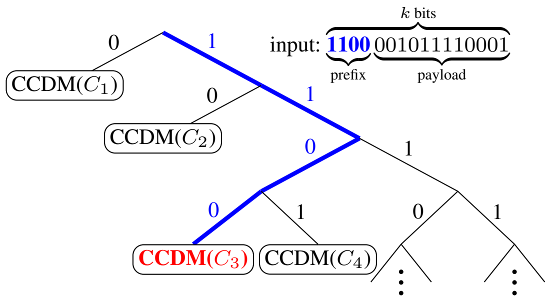

Fig. 1 Illustration of a Huffman tree for HCSS. A four-bit prefix determines the composition to be used (here C_3), and the payload is mapped with CCDM methods denoted as CCDM(C_3).

In this work, we propose three novel concepts related to distribution matching. Our previous work on MPDM [9, 19] is extended to a new architecture—Huffman coded sphere shaping (HCSS)—that approximates the optimal sphere bound, and thus has lower rate loss. Secondly, a novel CCDM method for nonbinary alphabets is proposed that has a lower number of serial operations than arithmetic coding, which is, to the best of our knowledge, the first constructive CCDM alternative to arithmetic coding for nonbinary alphabets. Finally, we demonstrate that the SR method enables multiplication-free mapping and demapping, based on a lookup table (LUT) whose size is less than 10 kbit.

2 Huffman Coded Sphere Shaping (HCSS)

In order to approximate the optimal sphere bound codebook [20], we sort all possible compositions by their energy and select the low-energy compositions first. We further constrain the number of permutations of each composition to be a power of two, which is the main conceptual difference to known techniques such as ESS and SM. This power-of-two constraint comes at the expense of a small rate loss as not the entire

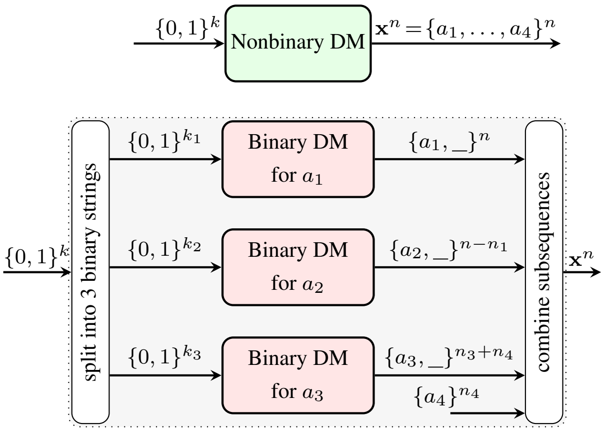

Fig. 2 Parallel-amplitude architecture: Transformation from a nonbinary DM (top) to parallel binary DMs (bottom) for 4 amplitudes. We propose the use of LUTs for the binary DM.

signal space inside an n-sphere can be used. This forces a dyadic distribution on the compositions, which allows the use of Huffman coding to determine a prefix which addresses each composition without additional rate loss, and the use of existing CCDM algorithms for mapping the payload to a sequence. Figure 1 illustrates the principle of HCSS where the prefix (bold blue) of the k-bit input selects the composition (red), and the binary remainder is mapped with CCDM methods such as those studied in Sec. 3.

3 Constant-composition Mapping Methods

3.1 Subset Ranking Based on Lookup Tables

Subset ranking (SR) has been proposed in [18] as a low-serialism CCDM algorithm for binary alphabets. To transform a nonbinary to a binary mapping operation, a parallel-amplitude architecture was also proposed in [18], which allows operation of $m-1$ DMs in parallel. A block diagram of this transformation is shown in Fig. 2 for $m = 4$ shaped amplitudes, which corresponds to 64-ary quadrature amplitude modulation (QAM).

For these variable-length binary DMs, the principle of SR is to represent a (shaped) sequence by the indices of one binary symbol, resulting in a constant-order subset. The number of preceding subsets in an ordering (e.g., lexicographical) is known as rank. By ranking a subset, a one-to-one correspondence between a shaped sequence and the binary rank is established, which carries out DM demapping. The inverse operation of ranking is DM mapping, or unranking: for a given rank (the binary input), the subsets defining each binary DM sequence are determined, and the shaped sequence is therefore mapped.

In [18], highly parallel algorithms for ranking and unranking are presented whose computational complexity lies mostly in calculating binomial coefficients. For short lengths, all required binomial coefficients may be precomputed, and stored in a LUT. For a given n, the number of binomial coefficients to be computed is $\left\lfloor \frac{n}{2} \right\rfloor - 1$ due to their symmetry around $n/2$ and since $\binom{n}{1} = n$. Hence, the number of required LUT entries for all DMs with length up to n is $\sum_{i=4}^{n} \left(\left\lfloor \frac{i}{2} \right\rfloor - 1 \right)$, where all trivial cases which result in a binomial coefficient equal to 1 or n are omitted. The size of each LUT entry is $\left\lceil \log_2 \binom{n}{w} \right\rceil$ bits, with the maximum size occurring for $w = \lfloor n/2 \rfloor$. This gives an overall LUT size of

$$\sum_{i=4}^{n} \sum_{w=2}^{\left\lfloor \frac{i}{2} \right\rfloor} \left\lceil \log_2 \binom{i}{w} \right\rceil \text{ bits.} \tag{1}$$

As an example, a LUT for all DMs up to length $n = 50$ has 14.3 kbit size, with the maximum LUT entry requiring 47 bit. A more detailed study of LUT sizes is given in Sec. 4.3.

3.2 Multiset Ranking

Multiset ranking (MR) is a generalization of SR to nonbinary alphabets, which removes the parallel-amplitude constraint, and thus achieves the same rate loss as conventional arithmetic-coding CCDM. The number of preceding sequences, which is referred to as relative rank, is computed for all amplitudes assuming that this amplitude were to be used at the current position. Note that this assumes a fixed sorting of the shaped sequences, such as lexicographical. Next, the cumulative relative ranks are compared to the target rank (i.e., the binary DM input), and the last amplitude whose relative rank does not exceed the target rank is chosen. Afterwards, the relative ranks are updated based on the chosen amplitude, and the above steps are carried out until the shaped sequence of length n is constructed. In contrast to arithmetic coding, which is serial in k for mapping and n for demapping, MR is serial in n for mapping and demapping, and thus requires fewer serial operations while not incurring additional rate loss.

4 Numerical Results

4.1 Rate Loss Performance

In Fig. 3, rate loss is shown as a function of block length n for various shaping methods. Rate loss is defined as the difference between the asymptotic rate (i.e., the entropy) and the actual rate of the DM scheme. CCDM and MPDM achieve the distribution $[0.4, 0.3, 0.2, 0.1]$. The ideal sphere bound achieves the lowest possible rate loss given a cardinality and length [20]. For each n, HCSS and the ideal n-sphere (which has similar performance to ESS and SM) are set to operate at the rate of MPDM. We observe that the rate loss of the proposed HCSS is significantly lower than CCDM and lies in between the rate loss of an ideal, fully populated sphere and MPDM. From the inset figure, we note that the rate loss penalty compared with the ideal sphere bound is consistently low for multiset ranking HCSS (MR-HCSS), while subset ranking HCSS (SR-HCSS) exhibits a significant additional penalty only when the block length is very small.

4.2 AWGN Performance with Forward Error Correction

To evaluate the performance of the presented shaping schemes, Monte-Carlo simulations of 64QAM symbols over the additive white Gaussian noise (AWGN) channel were performed. In

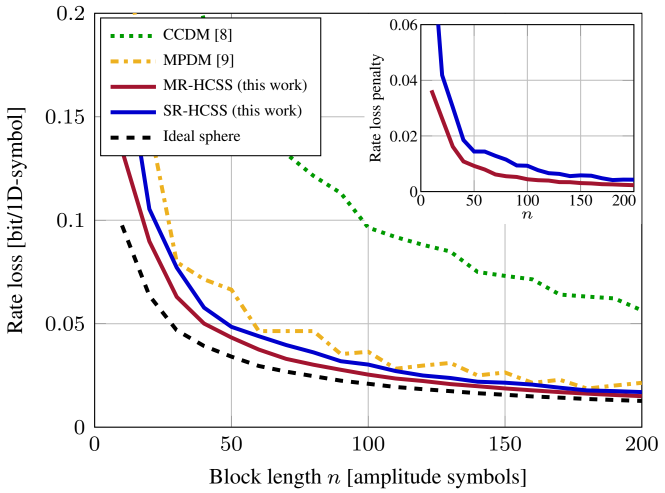

Fig. 3 Rate loss versus block length for various DMs, including the proposed MR-HCSS and SR-HCSS. Inset is the relative rate loss penalty in bits per 1D symbol for MR-HCSS and SR-HCSS compared to the ideal sphere bound.

Fig. 4, the frame error rate (FER) of low-density parity-check (LDPC) codes from the DVB-S2 standard (length 64800 bits) is shown versus signal-to-noise ratio (SNR) in dB. The throughput is set to 4.5 bits per 2D-symbol (bit/2D-sym), which is achieved with a rate-3/4 code for uniform and rate-4/5 for shaped signalling. For $n = 100$, MR-HCSS and with SR-HCSS give a shaping gain of approximately 0.77 dB and 0.73 dB, respectively. CCDM with $n = 100$ (not shown in Fig. 4) achieves the performance of uniform 64QAM. MPDM at $n = 100$ performs slightly worse than the HCSS schemes due to the limited use of the signal space, but the penalty compared to MR-HCSS is less than 0.1 dB. For $n = 20$, however, the performance of the schemes differs significantly, with MPDM being worse than uniform. MR-HCSS and SR-HCSS are 0.35 dB and 0.15 dB more power-efficient than uniform 64QAM. The reason for the greater variation in performance is that the rate loss decreases by $1/n$ for each additional input bit that can be addressed.

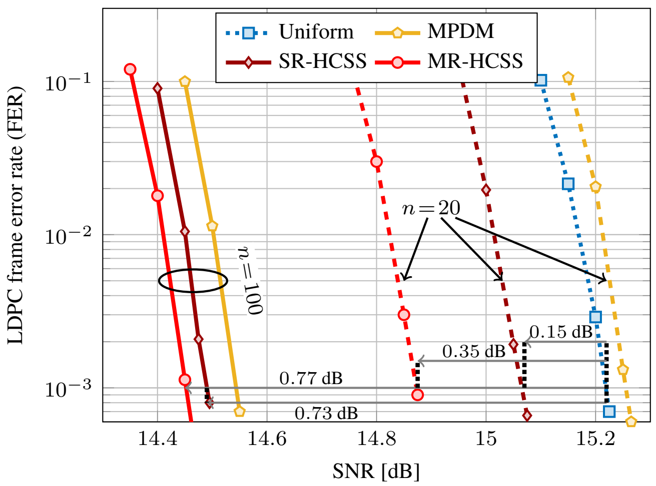

Fig. 4 Post-LDPC FER for 64QAM at an information rate of 4.5 bit/2D-sym. All shaping gains are evaluated at 10^{-3} FER.

4.3 *LUT Size*

For the same setup as studied above, Fig. 5 shows the SNR gain over uniform 64QAM, both at a FER of 10^{-3}, as a function of LUT size for SR-HCSS. With less than 1 kbit size, a small shaping gain of 0.15 dB is achieved and rate adaptivity can be realized. When increasing the LUT to approximately 7 kbit, more than 0.5 dB gain are obtained, achieving around 0.4 dB gain less than the infinite length DM. We observe that the gain increase slows down for longer block lengths, with a LUT with more than 100 kbit required for 0.73 dB shaping gain, which is within 0.2 dB of the infinite length DM gain. Increasing the LUT size to around 1 Mbit improves the shaping gain to only 0.82 dB. We note that these values are many orders of magnitude smaller than if CCDM were carried out directly with a LUT. A CCDM with, for example, $n = 20$ and $k = 32$ would require a LUT size of $2^{32} \cdot 20 \cdot \log_2 4 \approx 171.8$ Gbit size.

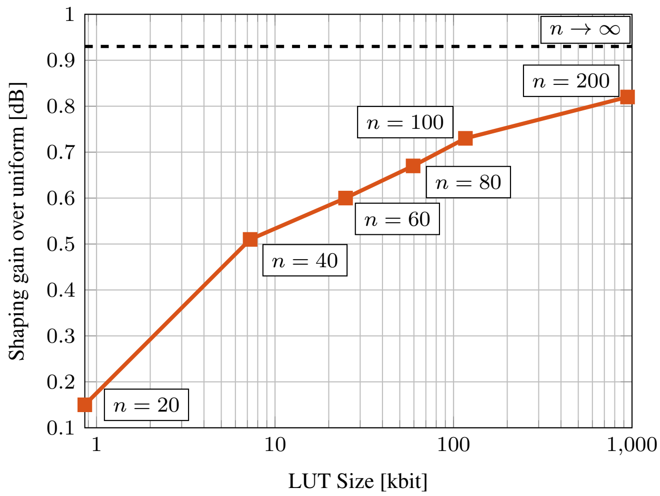

Fig. 5 SNR gain in dB with SR-HCSS over uniform 64QAM versus LUT size in kbit at a target rate of 4.5 bit/2D-sym.

5 Conclusions

We have demonstrated new techniques and implementations for high-throughput distribution matching. The performance of the proposed HCSS is close to the optimum, and its structure allows to use CCDM algorithms. The performance improvement over uniform 64QAM is numerically found to be more than 0.7 dB for a block length of $n = 100$, requiring a LUT of around 100 kbit when implemented with subset ranking, a penalty of only 0.2 dB compared with an infinite length DM. For ultra-short lengths of $n = 20$ where rate adaptivity is the main objective, shaping gains of up to 0.35 dB are feasible. For CCDM, two new algorithms are proposed. Multiset ranking is an alternative to arithmetic coding that requires fewer serial operations. We further show a LUT-based implementation of subset ranking that may be regarded as a trade-off between performance and LUT size, with significant shaping gains demonstrated for LUT sizes of less than 10 kbit.

References

[1] G. Böcherer, P. Schulte, and F. Steiner, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.

[2] T. Fehenberger, G. Böcherer, A. Alvarado, and N. Hanik, "LDPC coded modulation with probabilistic shaping for optical fiber systems," in *Proc. Optical Fiber Communication Conference (OFC)*. Los Angeles, CA, USA: Paper Th.2.A.23, Mar. 2015.

[3] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, "On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel," *J. Lightw. Technol.*, vol. 34, no. 22, pp. 5063–5073, Nov. 2016.

[4] F. Buchali, G. Böcherer, W. Idler, L. Schmalen, P. Schulte, and F. Steiner, "Experimental demonstration of capacity increase and rate-adaptation by probabilistically shaped 64-QAM," in *Proc. European Conference and Exhibition on Optical Communication (ECOC)*. Valencia, Spain: Paper PDP.3.4, Sep. 2015.

[5] 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. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.

[6] T. Fehenberger, D. Lavery, R. Maher, A. Alvarado, P. Bayvel, and N. Hanik, "Sensitivity gains by mismatched probabilistic shaping for optical communication systems," *IEEE Photon. Technol. Lett.*, vol. 28, no. 7, pp. 786–789, Apr. 2016.

[7] E. Ip, Y.-K. Huang, S. Zhang, T. J. Xia, D. L. Peterson, G. Wellbrock, G. Milione, T. Wang, Y. Aono, and T. Tajima, "41.5-Tb/s transmission over 549 km of field deployed fiber using throughput optimized probabilistic-shaped 144QAM," *J. Lightw. Technol.*, vol. 37, no. 1, pp. 148–155, Aug. 2019.

[8] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.

[9] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Multiset-partition distribution matching," *IEEE Transactions on Communications*, vol. 67, no. 3, pp. 1885–1893, Mar. 2019.

[10] Y. C. Gültekin, W. van Houtum, S. Serbetli, and F. M. J. Willems, "Constellation shaping for IEEE 802.11," in *Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Montreal, QB, Canada, Oct. 2017.

[11] Y. C. Gültekin, F. M. Willems, W. van Houtum, and S. Serbetli, "Approximate enumerative sphere shaping," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Vail, CO, USA, Jun. 2018.

[12] Y. C. Gültekin, W. J. van Houtum, A. Koppelaar, and F. M. J. Willems, "Enumerative sphere shaping for wireless communications with short packets," *arXiv:1903.10244*, Mar. 2019.

[13] P. Schulte and F. Steiner, "Shell mapping for distribution matching," *arXiv preprint arXiv:1803.03614*, Mar. 2018.

[14] J. Cho, "Prefix-free code distribution matching for probabilistic constellation shaping," *arXiv:1810.02411*, Mar. 2019.

[15] M. Pikus and W. Xu, "Bit-level probabilistically shaped coded modulation," *IEEE Communications Letters*, vol. 21, no. 9, pp. 1929–1932, Sep. 2017.

[16] G. Böcherer, P. Schulte, and F. Steiner, "High throughput probabilistic shaping with product distribution matching," *arXiv:1702.07510*, Feb. 2017.

[17] Y. Koganei, K. Sugitani, H. Nakashima, and T. Hoshida, "Optimum bit-level distribution matching with at most $\mathcal{O}(N^3)$ implementation complexity," in *Proc. Optical Fiber Conference (OFC)*. Paper M4B.4, Mar. 2019.

[18] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Parallel-amplitude architecture and subset ranking for fast distribution matching," *arXiv*, Feb. 2019. [Online]. Available: https://arxiv.org/abs/1902.08556

[19] D. S. Millar, T. Fehenberger, T. Koike-Akino, K. Kojima, and K. Parsons, "Distribution matching for high spectral efficiency optical communication with multiset partitions," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 517–523, Jan. 2019.

[20] R. F. H. Fischer, *Precoding and signal shaping for digital transmission*. New York, NY, USA: John Wiley & Sons, 2005.