

## Mapping Strategies for Short-Length Probabilistic Shaping

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

TR2019-106 September 24, 2019

### Abstract

Mapping techniques for probabilistic amplitude shaping are reviewed. We focus on energy considerations, classification of methods and systems, and implementation aspects. A numerical study of required block lengths is carried out for two short-blocklength applications, namely achieving half of the maximum shaping gains and enabling rate adaptivity.

*European Conference on Optical Communication (ECOC)*

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgement of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.



# MAPPING STRATEGIES FOR SHORT-LENGTH PROBABILISTIC SHAPING

*Tobias Fehenberger\*, David S. Millar, Toshiaki Koike-Akino, Keisuke Kojima, and Kieran Parsons*

*Mitsubishi Electric Research Labs, 201 Broadway, Cambridge, MA, USA.* \*E-mail: tobias.fehenberger@ieee.org

**Keywords:** Probabilistic Shaping, Distribution Matching, Short Blocklength

## Abstract

Mapping techniques for probabilistic amplitude shaping are reviewed. We focus on energy considerations, classification of methods and systems, and implementation aspects. A numerical study of required block lengths is carried out for two short-blocklength applications, namely achieving half of the maximum shaping gains and enabling rate adaptivity.

## 1 Introduction

Constellation shaping describes the optimization of a modulation format with equidistant and equiprobable signal points towards some shape that is tailored to the transmission channel [1]. Geometric shaping, on the one hand, optimizes the location of constellation points. While this approach offers large shaping gains for symbol-wise forward error correction (FEC) decoding, the shaped constellation often does not have Gray mapping any more, which can make it challenging to realise shaping gains in bit-interleaved coded modulation (BICM) systems with binary FEC [2]. Probabilistic shaping (PS), on the other hand, modifies the probability of the constellation symbols, which remain on a square grid. Classic PS schemes are for example based on many-to-one mappings [3], trellis shaping [4], and shell mapping [5].

PS has been applied in fibre optics as early as 2012 using trellis shaping [6] and shell mapping [7]. More work on PS has been published afterwards, such as [8, 9], but a widespread interest of the fibre-optic community in the topic of PS sparked in 2015 with the proposal of probabilistic amplitude shaping (PAS) [10]. The first demonstrations [11–13] were published the same year.

Since its proposal, the PAS architecture has become somewhat ubiquitous for shaped communication systems by offering a low-complexity and flexible integration into existing BICM settings with only minor modifications to the coded modulation structure. Compared to conventional uniform quadrature amplitude modulation (QAM), it is mainly two benefits that make PAS highly attractive for optical systems. Firstly, a shaping gain of approximately 1 dB signal-to-noise ratio (SNR) can be achieved for high-order QAM [13], which can be a significant improvement especially for highly optimized fibre-optic links such as submarine systems. Maybe even more appealing to a broader market is the feature of rate adaptivity. The net throughput can be adapted seamlessly to the channel without modifying other system parameters such as symbol rate, QAM order, and FEC overhead [14], which allows to fill the spectral efficiency gaps of square QAM. An alternative approach to realise variable rates is time-hybrid modulation [15].

A key component in the PAS framework is the distribution matcher (DM), which is the main focus of this work. A DM can be considered the PAS workhorse as it carries out the mapping from uniformly distributed bits to shaped amplitudes, which are then combined with uniform bits (coming either from data or the FEC parity bits) to generate shaped QAM symbols. Various types of fixed-to-fixed-length DMs and mapping schemes (which, strictly speaking, do not match a distribution) for PAS exist. This paper aims at providing an overview of some of these schemes and provides insight into their internal structure and short-length performance. We also comment on complexity considerations for the investigated schemes.

## 2 Principles of Mapping Techniques

### 2.1 General Aspects

A DM, or in general any mapping for PAS, solves the indexing problem of mapping each of the  $2^k$  binary inputs (having length  $k$ ) to exactly one out of  $M^n$  shaped sequences, where  $M$  denotes the number of shaped amplitudes. Note that this mapping is usually injective, i.e., some output sequences will never be addressed and thus, never be transmitted. An important figure of merit to consider is rate loss, which is the difference of the DM rate,  $k/n$ , to the entropy of the amplitudes. The rate loss is to be minimized for efficient DM operation and decreases with block length  $n$ , see Sec. 4 for a detailed discussion. Lookup tables for long blocks are prohibitively large, which is why constructive methods are required.

### 2.2 Mapping System and Method

Different mapping *systems* can be differentiated based on the properties of their output sequences. For a particular system, various *methods* can be used to implement the actual mapping. As an example, all output sequences of a CCDM system have identical composition (and therefore identical energy) and differ only in the order of the occurring amplitudes. A possible method for implementing such a CCDM system is arithmetic coding [16, Sec. IV]. We will further use this classification of method and system in Sec. 2.4.

### 2.3 Advanced Mapping Systems and Methods for PAS

Improving upon CCDM, advanced shaping systems produce output sequences of varying compositions. Multiset-partition distribution matching (MPDM) [17, 18] uses various CCDM-like sequences that are addressed in a Huffman tree. This yields lower rate loss than CCDM and has the additional benefit that it leverages any progress that is made for CCDM methods, such as more efficient implementations. Shell mapping (SM) [5, 19] and enumerative sphere shaping (ESS) [20] are closely related mapping systems that in principle use every sequence up to a certain maximum energy. While this is the most energy-efficient way of constructing a shaped output, the implementation complexity can be challenging, in particular for shell mapping [20, Sec. IV-D].<sup>\*</sup> Other DM approaches use framing of variable-length coding in order to achieve fixed-to-fixed length mapping [21]. For CCDM systems, bit-level transformations have been proposed [22–24] which modify the DM mapping method such that DM instances can be run in parallel in order to increase the throughput.

Recently, advanced methods that improve upon the arithmetic-coding method for CCDM have been proposed. Limited to binary-output alphabets, subset ranking [24] establishes a different mapping that has fewer serial operations and thus, allows low-latency DM operation. A similar technique for binary CCDM has also been proposed in [25].

### 2.4 Energy Considerations

Figure 1 shows an illustration of the three different shaping systems CCDM, MPDM, and ESS (highlighted by colour) for a fixed  $n$  and a fixed distribution. The binary input space on the left contains all Bernoulli- $\frac{1}{2}$  distributed sequences of length  $2^{k_{CCDM}}$ ,  $2^{k_{MPDM}}$ , and  $2^{k_{ESS}}$ , which can be mapped by the respective scheme. In general, CCDM has the smallest input size for a fixed  $n$ , while ESS has the largest. Each of the binary inputs is to be mapped to exactly one out of  $M^n$  output sequences, and how this injective mapping is established relates to the mapping method. For instance, a conventional arithmetic-coding CCDM and a parallel-amplitude CCDM with subset ranking address sequences with the same properties, yet the actual output and how the mapping is performed is different.

On the right of Fig. 1, the output signal space is depicted. The entire signal space containing  $M^n$  sequences is delimited by the black solid ellipse. Its bottom part corresponds to low-energy sequences and the top to those with high energy. CCDM uses sequences with fixed composition and thus occupies a line of fixed energy in the output space. Note that other compositions can have the same energy, so CCDM does not span an entire energy level. MPDM by contrast uses various energy levels, including the CCDM one. ESS by design is the most energy-efficient of these techniques and utilises all sequences up to a certain maximum energy level. Setting implementation

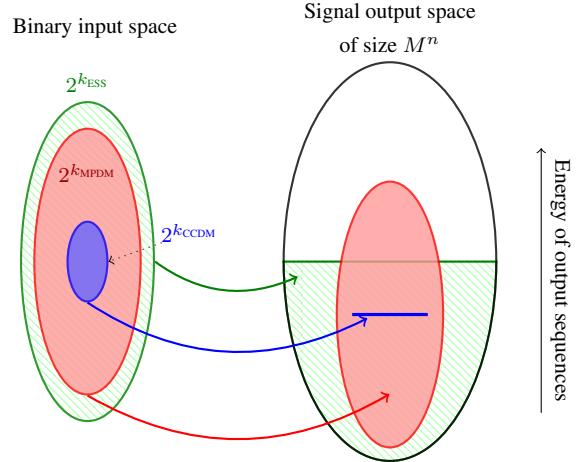


Fig. 1 Mapping from binary input sequences (left) to shaped output sequences (right) for CCDM (blue), MPDM (red), and ESS (green), for a fixed block length  $n$  and distribution.

aspects aside (more on this in Sec. 4), energy-efficient mapping systems are designed to occupy as much of the bottom of the output signal space as possible.

## 3 Numerical Analysis

In the following, we evaluate the performance of the previously discussed shaping schemes as a function of the block length  $n$ . The focus is put on achieving some shaping gain that is deliberately smaller than the maximum. The following results are obtained by computing the achievable information rate (AIR) for bit-metric decoding and the additive white Gaussian noise (AWGN) channel [26]. We subtract from this AIR, which is achievable only for infinite-length DM, the rate loss of the shaping schemes, see [17, Appendix] for details. Maxwell-Boltzmann distributions are used as target compositions for CCDM and MPDM at each SNR. For ESS, the target rate, and thereby the distribution, is varied as to maximize the finite-length AIR. This AIR-based evaluation allows to study a wide range of SNRs without the need to implement FEC and to carry out time-consuming Monte Carlo simulations.

### 3.1 Obtaining Shaping Gain

We first study the required block length to achieve a certain shaping gain. The targeted SNR savings over uniform signalling are set to 0.2 dB, 0.5 dB, and 0.6 dB for QAM of order {16, 64, 256}, respectively, which is approximately equal to 50% of the maximum achievable shaping gain for each format. Figure 2 shows for this setting the required block length  $n$  in amplitude symbols as a function of SNR of the AWGN channel. The block length was varied with a granularity of 5 symbols for CCDM (dotted), MPDM (solid), and ESS (dashed). As expected from the discussion in Sec. 2.4, we observe that an efficient usage of the available signal space corresponds to a small required block length. Hence, CCDM requires in all cases the longest blocks of more than 100 amplitude symbols. MPDM obtains the target gain with significantly shorter blocks than CCDM, in particular for 64QAM and

<sup>\*</sup>SM and ESS differ in which sequences are ignored due to the output space not having  $2^x$  sequences. SM discards sequences of the highest-energy shells, while ESS is marginally less energy-efficient by not using some low-energy sequences (unless modifications are made as in [20, Sec. IV-D]).

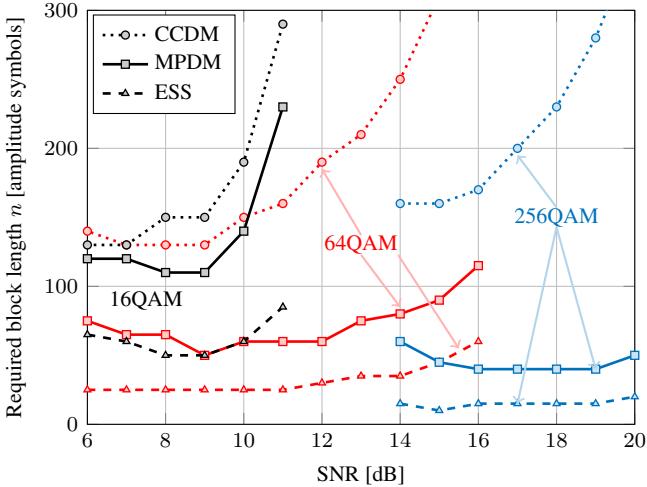


Fig. 2 Minimum block length  $n$  to achieve  $\{0.2, 0.5, 0.6\}$  dB shaping gain for  $\{16, 64, 256\}$ -QAM, respectively.

256QAM. ESS performs best among the considered schemes and achieves significant shaping gains for as little as  $n = 15$  with 256QAM. Furthermore, we observe that all schemes tend to require larger  $n$  in order to achieve the targeted shaping gain at high SNR. The reason for this is that at high SNR, the absolute SNR shaping gain is small, which translates into a more stringent requirement on having low rate loss and thus, demands longer blocks lengths.

### 3.2 Enabling Rate Adaptivity

In Fig. 3, we evaluate the minimum block length  $n$  that is required to achieve the same throughput as uniform QAM. The motivation behind this is to investigate the length requirement for realising rate adaptivity. The three considered QAM formats are 16QAM (black), 64QAM (red), and 256QAM (blue). Similar to Fig. 2, we observe for all considered cases that CCDM requires the longest blocklength, while very short blocks of only a few symbols are sufficient for ESS. In fact, ESS performs best at these short lengths because in this case, it is extremely important to utilise the limited signal space as efficiently as possible. By using all points inside an  $n$ -sphere, ESS obtains the lowest required block length to match uniform performance and thus, to enable rate adaptivity. For 256QAM, ESS requires only  $n = 5$  symbols, and the corresponding input length is  $k = 15$ . To realise this mapping, an alternative method to the trellis structure (which ESS uses internally) could be a lookup table whose size is  $2^{15} \cdot 5 \cdot \log_2(\sqrt{256}/2) \approx 492$  kbit.

## 4 Implementation Aspects

In the preceding section, we followed the conventional approach of comparing different schemes by studying the block length that is required for a certain shaping gain. While this is certainly a natural choice for analysing and comparing shaped systems, it comes with the caveat that this approach inherently assumes that shorter blocks are always better, for instance because they have advantages for implementation.

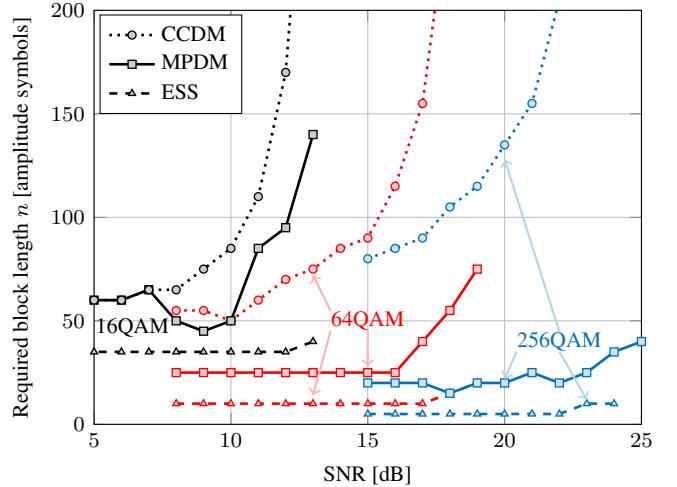


Fig. 3 Minimum block length  $n$  required to achieve the performance of uniform QAM versus SNR in dB.

In the following, we comment on this aspect by considering computational complexity, i.e., the number of additions or multiplications, latency, and storage requirements.

An example where slightly longer block lengths can be beneficial is the parallel-amplitude architecture [24]. By allowing a small additional rate loss, the throughput is increased significantly by using  $M - 1$  DMs in parallel. Furthermore, the serialism (and thus, the latency) of the subset-ranking method is smaller than arithmetic-coding CCDM. It can thus be beneficial to make the blocks slightly larger than for conventional CCDM in order to facilitate implementation. An example for reduced storage requirements is the tree structure of MPDM [17, Sec. III-C], which can introduce some rate loss but allows to use existing CCDM methods for a variable-composition mapping problem that otherwise might only be solvable with huge lookup tables. Approximate ESS [27] has slightly smaller rate than conventional ESS, yet significantly reduced storage requirements. Finally, we note that analysing computational complexity by counting additions and multiplications might give an approximate indication of complexity, but it omits many practical considerations concerning real-time operation.

## 5 Conclusion

We have reviewed mapping schemes for probabilistic amplitude shaping. All considered schemes solve a similar indexing problem, yet they differ greatly in how their mapping methods operate and which shaped output sequences are addressed. Energy considerations suggest that two advanced systems, ESS and MPDM, significantly outperform CCDM. This is confirmed in numerical simulations for the cases of i) achieving approximately 50% of the maximum shaping gain for high-order QAM formats, and ii) having performance similar to uniform signalling, which enables rate adaptivity. Finally, we have discussed and presented examples that comparing shaping schemes only by considering block length does not tell the whole story when implementation aspects are taken into account.

## References

- [1] G. D. Forney, Jr., R. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, “Efficient modulation for band-limited channels,” *IEEE J. Sel. Areas Commun.*, vol. 2, no. 5, pp. 632–647, Sep. 1984.
- [2] D. S. Millar, T. Fehenberger, T. Koike-Akino, K. Kojima, and K. Parsons, “Coded modulation for next-generation optical communications,” in *Proc. Optical Fiber Communication Conference (OFC)*. San Diego, CA, USA: Paper Tu3C.3, Mar. 2018.
- [3] R. G. Gallager, *Information theory and reliable communication*. New York, NY, USA: John Wiley & Sons, 1968.
- [4] G. D. Forney, “Trellis shaping,” *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 281–300, Mar 1992.
- [5] R. Laroia, N. Farvardin, and S. A. Treter, “On optimal shaping of multidimensional constellations,” *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1044–1056, 1994.
- [6] B. P. Smith and F. R. Kschischang, “A pragmatic coded modulation scheme for high-spectral-efficiency fiber-optic communications,” *J. Lightw. Technol.*, vol. 30, no. 13, pp. 2047–2053, Jul. 2012.
- [7] L. Beygi, E. Agrell, and M. Karlsson, “Adaptive coded modulation for nonlinear fiber-optical channels,” in *IEEE GLOBECOM*, 2012, pp. 331–335.
- [8] M. P. Yankov, D. Zibar, K. J. Larsen, L. P. Christensen, and S. Forchhammer, “Constellation shaping for fiber-optic channels with QAM and high spectral efficiency,” *IEEE Photon. Technol. Lett.*, vol. 26, no. 23, pp. 2407–2410, Dec. 2014.
- [9] L. Beygi, E. Agrell, J. M. Kahn, and M. Karlsson, “Rate-adaptive coded modulation for fiber-optic communications,” *J. Lightw. Technol.*, vol. 32, no. 2, pp. 333–343, Jan. 2014.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, “Partition-based probabilistic shaping for fiber-optic communication systems,” in *Proc. Optical Fiber Communication Conference (OFC)*, San Diego, CA, USA, Mar. 2019.
- [15] X. Zhou, L. E. Nelson, and P. Magill, “Rate-adaptable optics for next generation long-haul transport networks,” *IEEE Communications Magazine*, vol. 51, no. 3, pp. 41–49, 2013.
- [16] P. Schulte and G. Böcherer, “Constant composition distribution matching,” *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.
- [17] 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.
- [18] 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.
- [19] P. Schulte and F. Steiner, “Shell mapping for distribution matching,” *arXiv:1803.03614*, Mar. 2018.
- [20] 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.
- [21] J. Cho, “Prefix-free code distribution matching for probabilistic constellation shaping,” *arXiv:1810.02411*, Mar. 2019.
- [22] G. Böcherer, P. Schulte, and F. Steiner, “High throughput probabilistic shaping with product distribution matching,” *arXiv:1702.07510*, Feb. 2017.
- [23] M. Pikus and W. Xu, “Bit-level probabilistically shaped coded modulation,” *IEEE Communications Letters*, vol. 21, no. 9, pp. 1929–1932, Sep. 2017.
- [24] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, “Parallel-amplitude architecture and subset ranking for fast distribution matching,” *arXiv:1902.08556*, Feb. 2019.
- [25] 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.
- [26] A. Alvarado, T. Fehenberger, B. Chen, and F. M. J. Willems, “Achievable information rates for fiber optics: applications and computations,” *J. Lightw. Technol.*, vol. 36, no. 2, pp. 424–439, Jan. 2018.
- [27] 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.