

Distribution Matching for High Spectral Efficiency Optical Communication with Multiset Partitions

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

TR2018-192 March 09, 2019

Abstract

In this paper, we review the concept of multisetpartition based distribution matching (MPDM). We then characterize performance over short block lengths of 20, 40 and 80 symbols, and with 16QAM, 64QAM and 56QAM modulation formats. We note that a 40 symbol distribution matcher length can obtain better than uniform performance, while an 80 symbol distribution matcher can utilize most of the available gain over a wide range of signal-to-noise ratios.

IEEE Journal of Lightwave Technology

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 acknowledgment 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.

Distribution Matching for High Spectral Efficiency Optical Communication with Multiset Partitions

David S. Millar, *Member, IEEE*, Tobias Fehenberger, *Member, IEEE*, Toshiaki Koike-Akino, *Senior Member, IEEE*, Keisuke Kojima, *Senior Member, IEEE*, Kieran Parsons, *Senior Member, IEEE*

(Invited Paper)

Abstract—In this paper, we review the concept of multiset-partition based distribution matching (MPDM). We then characterize performance over short block lengths of 20, 40 and 80 symbols, and with 16QAM, 64QAM and 256QAM modulation formats. We note that a 40 symbol distribution matcher length can obtain better than uniform performance, while an 80 symbol distribution matcher can utilize most of the available gain over a wide range of signal-to-noise ratios.

Index Terms—Modulation, probabilistic shaping.

I. INTRODUCTION

PROBABILISTIC amplitude shaping (PAS) has emerged as a leading architecture for the implementation of capacity approaching constellation shaping [1]–[5]. Application of constant composition amplitude sequences to this approach can be shown to provide capacity approaching performance in the long block length regime (known as constant composition distribution matching (CCDM)) [6]. However, the long sequences required by CCDM to achieve low rate-loss are problematic for hardware implementation in high-speed optical transmission systems. This is in part due to the previously described matching and dematching algorithms (based on arithmetic coding), which are highly serial in nature [6]. Optical transmission systems generally operate in the region of 30–50 GBd, with a hardware processing clock rate on the order of 500 MHz, therefore requiring parallelism of approximately 60–100 \times . This may be contrasted with, for example, low-density parity check codes [7], which require long block lengths for good performance, but may be decoded in parallel with graph-based decoding algorithms. Significant amounts of research recently have therefore focused on low-complexity and highly parallel algorithms for PAS implementation [8]–[12].

We have previously proposed the concept of multiset-partition distribution matching (MPDM) [13]. This distribution matcher (DM) enables the use of shorter DM block lengths with lower rate loss than that achieved by CCDM by ensuring that the probability mass function (PMF) of the entire distribution matcher achieves the target PMF, while individual sequences may not. As the number of permissible sequences is expanded beyond those which satisfy the target PMF (i.e.,

those with typical composition) to include all sequences which may be generated by partitioning a multiset with the target composition, we note that the number of sequences permissible in MPDM is always greater than or equal to the number of sequences permissible in CCDM (as the set of CCDM sequences is contained within the set of MPDM sequences). In turn, this indicates that MPDM will have equal or lower rate-loss compared with CCDM for all sequence lengths and distributions. A simplified implementation of this was proposed by considering sequences which correspond pairwise to the target PMF (pairwise MPDM) [13, Sec. III-B]. We also previously showed that by rounding down the number of sequences of each composition considered and using both a power of two number of sequences per composition, and a power of two number of sequences overall, we may construct an optimal prefix code – also considered as a binary tree, which describes the selection of a particular distribution within the distribution matcher. This structure also enables the use of the previously described matching and dematching algorithms for CCDM, albeit at much reduced block lengths.

In the next section of this paper, we summarize the operation of probabilistic shaping, PAS and CCDM. We will then recap the MPDM architecture, and the design algorithm. We then show simulated signal-to-noise ratio (SNR) gain for varying block lengths, and for modulation formats varying from 16-ary quadrature amplitude modulation (QAM), to 64QAM, and 256QAM. Finally, we examine the SNR gain of MPDM compared with uniform QAM for block lengths of 20, 40 and 80 symbols.

II. PROBABILISTIC SHAPING

In the contemporary literature for optical communications, certain characteristics of probabilistic shaping systems are near universal. The constellation points are normally considered to be square QAM; constellations are assumed to be shaped in the amplitude domain only (on a per quadrature basis), and therefore have reflective symmetry about the real and imaginary axes. This allows for the forward error correction (FEC) code to be applied to the stream of shaped bits, and the uniformly distributed parity bits assigned to the sign bits of the constellation – this configuration is also known as the “reverse concatenation” of FEC and distribution matching. A further common feature is that each shaped symbol sequence which corresponds to a block of uniformly distributed bits has exactly the desired amplitude probability mass function (PMF) – this is known as CCDM [6].

D. S. Millar, T. Koike-Akino, K. Kojima and K. Parsons are with Mitsubishi Electric Research Laboratories, Cambridge, MA 02139, USA.

T. Fehenberger was with the Technical University of Munich and Mitsubishi Electric Research Laboratories. He is currently a visiting researcher at the Eindhoven University of Technology, Netherlands.

Manuscript received XXX.

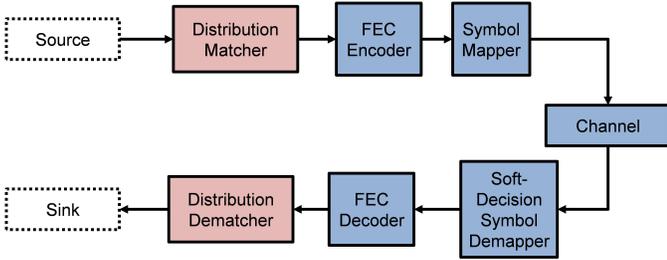


Fig. 1. A schematic of a communication system employing probabilistic amplitude shaping. The information bits are first shaped by a distribution matcher, then encoded by an FEC encoder, before being mapped to symbols and transmitted over the channel. The parity bits have uniform distribution, and are mapped to the sign bits in the desired constellation. At the receiver, the symbols are demapped, before the FEC code is decoded. The noise-free shaped bits are then passed to a distribution dematcher to recover the original information bits.

A. Probabilistic Amplitude Shaping (PAS)

The reverse concatenation approach to PAS is critical to realizable implementations, and is described in Fig. 1. Uniformly distributed bits from a source are mapped by a distribution matcher to symbol sequences, which have the desired distribution of amplitudes. The bits representing the shaped amplitudes are then encoded with an FEC encoder, and the parity bits (which have a uniform distribution) are assigned to the sign bits of the constellation. At the receiver side, the FEC decoder operates on the noisy received symbols in the conventional manner, and the noise- and error-free amplitude sequences are passed to a distribution dematcher, which outputs uniformly distributed data bits. This architecture provides several important benefits – firstly, the AWGN capacity may be approached with a bit-interleaved coded modulation (BICM) scheme, which is an important condition for practicality; secondly, the distribution dematcher can operate on noise- and error-free symbol sequence inputs. This reduces the complexity of the dematching algorithm requirements. By varying the entropy of the target distribution and the rate of the distribution matcher, this also enables the transmission rate to be varied while maintaining a constant code rate and symbol rate [1].

While these advances have pointed to possible implementations of probabilistic shaping in the near future, there are significant hurdles which remain. While CCDM remains the most widely studied technique at the time of writing, the high degree of serialism required in the distribution matcher and dematcher, combined with long block lengths required for good performance mean that it is highly challenging to implement CCDM in hardware. There is therefore significant impetus in the development of alternative architectures for PAS [8]–[11], [14]–[16], which do not suffer the problems of CCDM.

B. Distribution Matching

A distribution matcher (DM) is an algorithm which provides a one-to-one mapping of words of uniformly distributed bits onto sequences of symbols with a specified probability distribution. Conversely, distribution dematching is an algorithm providing the inverse operation, i.e. a one-to-one mapping of

sequences of symbols onto words of uniformly distributed bits. More specifically, in the case of PAS, we may say that the DM maps an input bit sequence onto an output sequence of amplitudes, with some specified target distribution P_A . For finite-length DM, the target distribution P_A must be quantized to $P_{\tilde{A}}$ such that the number of occurrences of each amplitude is integer. For a number of possible DM output sequences $N_{\text{Seq}} = 2^k$ where k is the number of input bits addressed by the DM, where each sequence consists of n symbols, we may quantify the rate loss of the DM as

$$R_{\text{loss}} = \mathbb{H}(\tilde{A}) - \frac{k}{n}, \quad (1)$$

where the entropy of the quantized distribution $\mathbb{H}(\tilde{A})$ is given by

$$\mathbb{H}(\tilde{A}) = - \sum_i P_{\tilde{A}}(i) \log_2(P_{\tilde{A}}(i)). \quad (2)$$

C. Constant Composition Distribution Matching

CCDM was proposed recently [6] as a method for mapping uniformly distributed bit sequences to shaped amplitude sequences and vice-versa. While this scheme suffers from highly serial matching and dematching algorithms, it is extremely widely studied [1], [2], and can provide arbitrarily low rate loss in the long sequence length regime. The CCDM algorithm specifies that each possible output sequence achieves the target PMF. In this case, we can describe a single CCDM output sequence as $C_{\text{typ}} = \{n P_{\tilde{A}}(a_1), \dots, n P_{\tilde{A}}(a_{|\mathcal{A}|})\}$. The constant-composition mapping of a uniform input sequence to a sequence that has C_{typ} is denoted as $f_{\text{ccdm}}(C_{\text{typ}})$ and can, for example, be carried out via arithmetic coding [6, Sec. IV]. The set of unique permutations of \underline{x}^n for a given C_{typ} is referred to as type class [17, Sec. 11.1], and its size is the multinomial coefficient [17, Eq. (11.17)]

$$M(C) = \binom{n}{n_1, n_2, \dots, n_{|\mathcal{A}|}} = \frac{n!}{n_1! n_2! \dots n_{|\mathcal{A}|}!}. \quad (3)$$

The number of input bits for CCDM of a particular C_{typ} is given by

$$k = \lfloor \log_2 M(C_{\text{typ}}) \rfloor, \quad (4)$$

where $\lfloor \cdot \rfloor$ denotes rounding down to the closest integer.

III. MULTISSET PARTITION DISTRIBUTION MATCHING

Multiset partition distribution matching (MPDM) is based on a pair of insights. Firstly, if DM output sequences \underline{x}^n which are not of typical composition C are included in the distribution matcher, the rate loss may be decreased, thus improving the performance of short length distribution matchers compared with CCDM. Secondly, we may consider the distribution matcher compositions as several equally sized partitions of the desired typical composition – while each composition itself need not be equal to the typical composition, the average over all compositions in a given partition is, by definition, equal to the typical composition.

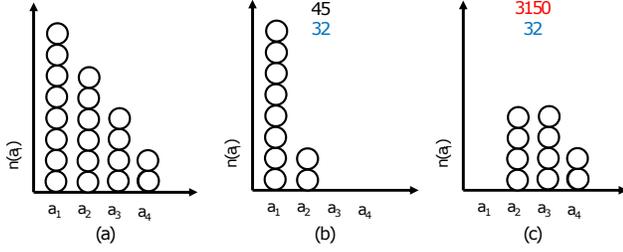


Fig. 2. Illustration of a partitioning of the typical composition C_{typ} shown leftmost in (a) into two non-typical compositions in (b) and (c).

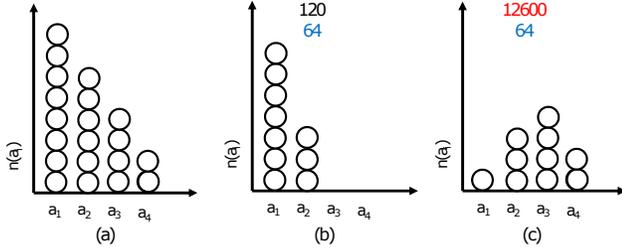


Fig. 3. Illustration of an alternative pairwise partitioning of the typical composition C_{typ} shown leftmost in (a) into two non-typical compositions in (b) and (c).

Thus, a general MPDM uses those output sequences whose compositions C_l satisfy

$$\frac{\sum_l^{N_{\text{comp}}} c_l \cdot C_l}{\sum_l^{N_{\text{comp}}} c_l} \stackrel{!}{=} C_{\text{typ}}, \quad (5)$$

where l indexes the N_{comp} possible compositions of the MPDM output sequences and c_l is the number of occurrences of C_l at the MPDM output, with $0 \leq c_l \leq M(C_l)$. The possible compositions C_l can be obtained by exhaustive search, and the choice of c_l depends on the partitioning constraint. The general partitioning problem (5) states that the *average* type of all sequences that are the output of a DM must be $P_{\bar{A}}$. The number of distinct compositions is given by

$$N_{\text{comp}} = \binom{n + |\mathcal{A}| - 1}{n}, \quad (6)$$

which can be proven, for example, with the stars-and-bars technique [18, Sec. II-5].

For example, Fig. 2 shows a pair of compositions which comprise a partition of the typical composition C_{typ} , shown leftmost in part (a). The total number of available sequences is determined by the lower number of available permutations for each composition in the partition – 45 for the composition shown in Fig. 2 (b), which is then rounded down to the nearest power of two – 32. There are therefore 64 total sequences in this partition which satisfy the typical composition on average, and can be addressed in a binary manner.

An alternative pairwise partitioning is shown in Fig. 3. A pair of new compositions is shown, comprising a new partition of the typical composition C_{typ} , leftmost in part (a). The total number of available sequences is again determined by the

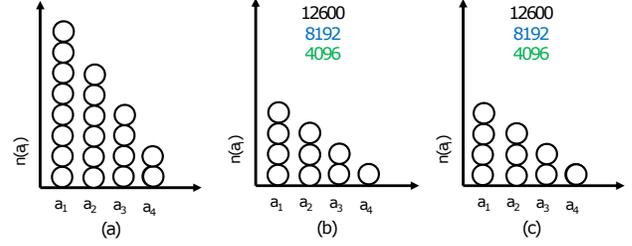


Fig. 4. Illustration of a degenerate pairwise partitioning of the typical composition C_{typ} shown leftmost in (a) into two identical compositions in (b) and (c).

lower number of available permutations for each composition in the partition – 120 for the composition shown in (b), which is then rounded down to the nearest power of two – 64. There are therefore 128 total sequences in this partition which satisfy the typical composition on average, and can be addressed in a binary manner.

Another alternative partitioning is shown in Fig. 4. A pair of identical compositions are shown, comprising degenerate partitioning of the typical composition C_{typ} , leftmost in part (a). The total number of available sequences is again determined by the lower number of available permutations for each composition in the partition – 12600 for both compositions shown, which is then rounded down to the nearest power of two – 8192. There are 8192 total sequences in this partition which satisfy the typical composition on average, which may be considered as taking 4096 from each side of the partition.

In total, considering a target PMF of $P_{\bar{A}} = [0.4, 0.3, 0.2, 0.1]$, which has $\mathbb{H}(\bar{A}) = 1.85$ bits and $C_{\text{typ}} = \{4, 3, 2, 1\}$ for $n = 10$, CCDM will allow for 8192 permutations with a binary address, or $k = 13$. This leads to 1.3 bits/amplitude, and a rate loss of 0.55 bits/amplitude. For pairwise MPDM, there are a total of 49 pairs (including the degenerate one) that satisfy the typical composition. The new total permutation count is 164214, which increases the number of input bits to $k = 17$ and thus reduces the rate loss to 0.15 bits.

While the above procedure gives a method for constructing a list of sequences for use in the distribution matcher, the large number of sequences required to achieve low rate loss even at short sequence lengths of tens of symbols ensures that additional structure is required such that existing algorithms may be used for matching and dematching. We therefore proposed [13] constructing a binary tree of compositions such that a variable length prefix code will specify a composition, while a variable length word can specify the permutation within a composition, such that the number of bits for each DM input is a constant k .

An example of this tree structure is shown in Fig. 5. The partitions are sorted according to their number of permutations, and smaller partitions are discarded or rounded down such that the total number of partitions is an integer power of two. A Huffman code is then formed describing the variable length prefix – partitions with fewer permutations are relatively less frequent, and therefore have more bits of prefix. As

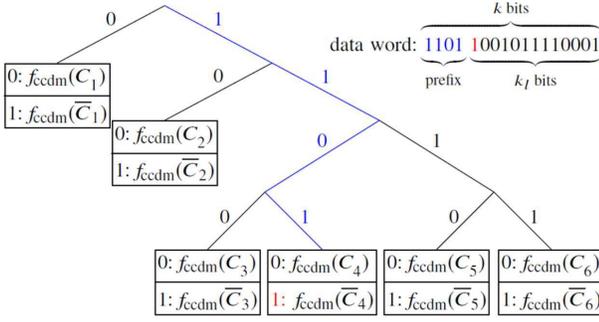


Fig. 5. Example of a binary tree structure enabling fixed length distribution matching with non-constant composition.

previously described in the literature [19] as the distribution of compositions is dyadic, the Huffman code will achieve zero rate loss. In the case of MPDM, we are able to form the prefix code without additional loss (and therefore a fixed-length to fixed-length mapping), although the rate-loss as defined in Eq. (1) remains.

A. MPDM Design Algorithm

The MPDM design procedure described in the preceding sections can be summarized with the following algorithm (assuming a pairwise, binary distribution matcher):

- 1) Optimize the Maxwell-Boltzmann coefficient for the specified SNR and modulation cardinality.
- 2) Quantize the optimized PMF, for instance by minimizing Kullback-Leibler divergence, according to the specified length of the distribution matcher n .
- 3) Generate the multiset by multiplying the elements in the PMF by twice the distribution matcher length $2n$.
- 4) Determine all possible unique partitions of the multiset, whereby each partition results in a pair of multisets with size n .
- 5) Determine the number of unique permutations of each sub-multiset. The maximum number of usable sequences from each sub-multiset is the minimum of these numbers over a partition pair. The number of usable sequences for a binary distribution matcher is determined for each sub-multiset is then obtained by rounding this number down to the nearest power of two, 2^{k_i} where i is the sub-multiset index.
- 6) Sum the number of usable sequences for a binary distribution matcher over all possible partitions. Round the resulting number of partitions down to the nearest power of two to determine the total number of sequences 2^k used by the final distribution matcher.
- 7) Sort the list of partitions according to their number of usable sequences. Starting with the largest numbers first, incorporate entire partitions into the final distribution matching table, such that the total number of sequences is strictly less than 2^k .
- 8) For the final partition, take an equal number of sequences from each sub-multiset such that the total number of partitions is 2^k .

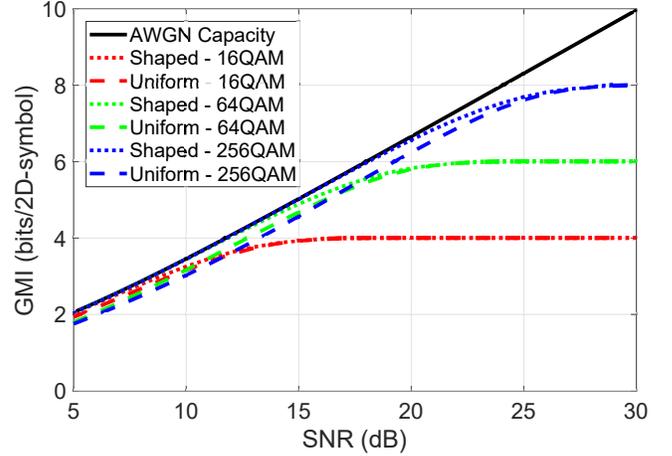


Fig. 6. Generalized mutual information (GMI) plotted against SNR in dB for 16QAM, 64QAM and 256QAM with and without probabilistic shaping. The use of bit-interleaved coded modulation (BICM) is assumed in all cases. The simulations of shaped QAM assume the use of an ideal distribution matcher, with no rate loss and an optimized Maxwell-Boltzmann distribution.

- 9) By calculating the frequency of each partition of each sub-multiset as 2^{-k_i} , it is possible to determine a Huffman code without rate loss to determine the prefix of length $k - k_i$ for each sub-multiset.
- 10) The remaining k_i bits of each distribution matcher word are mapped to a single sub-multiset according to a CCDDM mapper/demapper algorithm, such as arithmetic coding [6].

IV. PERFORMANCE OVER AWGN CHANNEL

In this section, we illustrate the performance of an optimized pairwise MPDM distribution matcher over the additive white Gaussian noise (AWGN) channel. The distribution matcher design was performed according to the previously described procedure. In this section, we consider an achievable rate for the MPDM distribution matcher design, assuming the use of BICM and FEC codes that are both capacity-achieving and of infinite length. By [20] (part III), the achievable rate is for 2^{2m} -QAM is calculated by

$$R_{\text{AIR}} = 2 \cdot \left[R_{\text{MPDM}}(n) + 1 - \sum_{i=1}^m \mathbb{H}(B_i|Y) \right]^+ \quad (7)$$

where B_1, B_2, \dots, B_m is the binary Gray label in each real dimension, $R_{\text{MPDM}}(n)$ is the MPDM rate for DM output length n , the addition of 1 accounts for the un-shaped sign bit, $\sum_{i=1}^m \mathbb{H}(B_i|Y)$ is the uncertainty per real dimension of the BICM demapper, Y is the channel output, and $[\cdot]^+ = \max(0, \cdot)$. While systems designed with hardware implementation in mind may have a constraint on codeword length, in practice, this length may easily exceed 10,000 bits [21].

Fig. 6 shows the performance of the ideal shaped and unshaped 16QAM, 64QAM and 256QAM over varying SNR. We note that for an ideal distribution matcher, all modulation formats considered approach the AWGN capacity for low SNR.

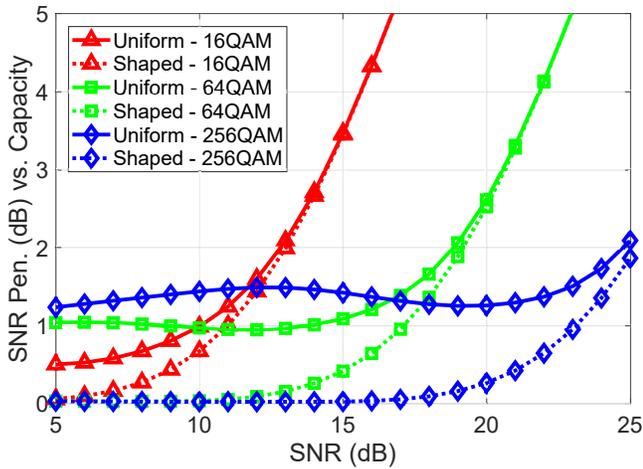


Fig. 7. Comparison of the SNR penalty compared with the AWGN capacity (i.e. Shannon bound) for 16QAM, 64QAM and 256QAM with and without probabilistic shaping. The simulations of shaped QAM assume the use of an ideal distribution matcher, with no rate loss and an optimized Maxwell-Boltzmann distribution.

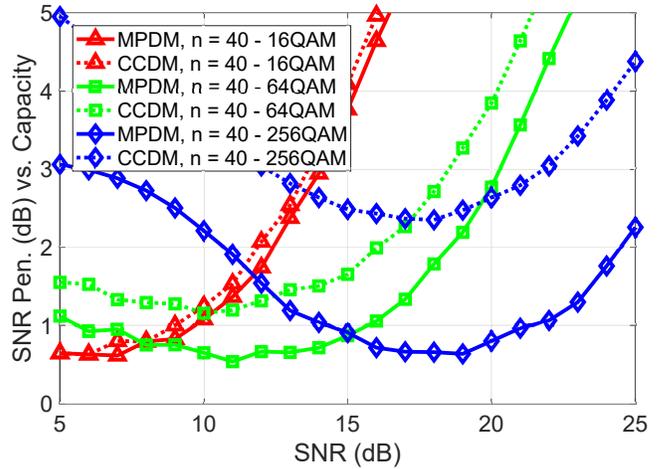


Fig. 9. Comparison of the SNR penalty compared with the AWGN capacity (i.e. Shannon bound) for 16QAM, 64QAM and 256QAM with distribution matchers having sequence length $n = 40$. CCDDM and MPDM are compared for an optimized Maxwell-Boltzmann distribution.

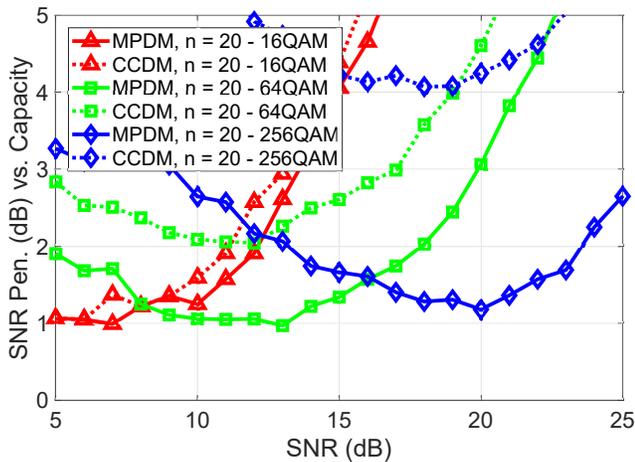


Fig. 8. Comparison of the SNR penalty compared with the AWGN capacity (i.e. Shannon bound) for 16QAM, 64QAM and 256QAM with distribution matchers having sequence length $n = 20$. CCDDM and MPDM are compared for an optimized Maxwell-Boltzmann distribution.

Fig. 7 shows the SNR penalty that each of the above formats incurs, relative to the Shannon bound. Here, we note that while all formats may approach the bound with an ideal distribution matcher, the penalty for uniform QAM modulation is smaller for lower SNRs. Between 5 and 10 dB, 16QAM has a penalty of 0.5–1 dB, while 64QAM has a penalty of around 1 dB from 10–15 dB SNR.

We note from Fig. 8 that the SNR penalty relative to the Shannon bound is significant for both MPDM and CCDDM with length $n = 20$. In particular, we note that the performance of both MPDM and CCDDM is significantly worse than uniform 16QAM for SNRs of 5–10 dB, while MPDM 64QAM can achieve performance comparable to that of uniform 64QAM between 8–14 dB SNR. While the performance improvement of MPDM increases with modulation cardinality, we note that

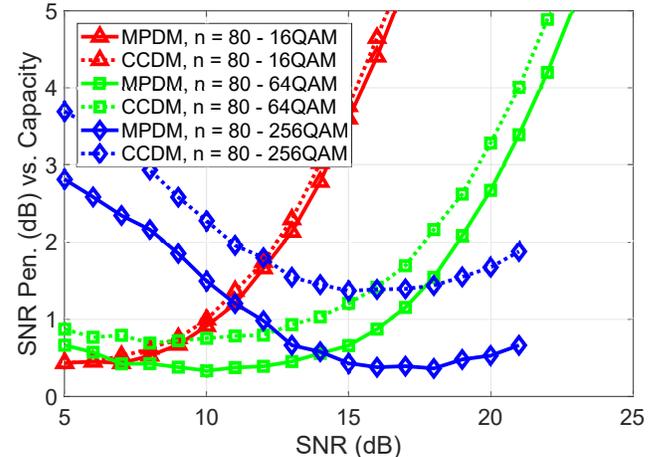


Fig. 10. Comparison of the SNR penalty with the AWGN capacity (i.e. Shannon bound) for 16QAM, 64QAM and 256QAM with distribution matchers having sequence length $n = 80$. CCDDM and MPDM are compared for an optimized Maxwell-Boltzmann distribution.

the penalty compared with the Shannon bound is at best around 1 dB, and that higher cardinality formats with low SNR suffer extremely large penalties.

We observe from Fig. 9 that the SNR penalty relative to the Shannon bound is reduced in all cases with DM length increased to $n = 40$. While MPDM and CCDDM 16QAM both offer performance similar to uniform 16QAM for 5–10 dB SNR, MPDM 64QAM and 256QAM exhibit less than 1 dB penalty for 6–15 dB SNR and 15–21 dB SNR respectively. Although there are still significant penalties for higher cardinality formats with low SNR, we note that significantly better than uniform performance may be observed over a wide range of SNR for MPDM 64QAM and 256QAM, while CCDDM suffers from significant penalty.

In Fig. 10, we see the penalties further reduced as DM length is increased to $n = 80$. While the penalty of MPDM

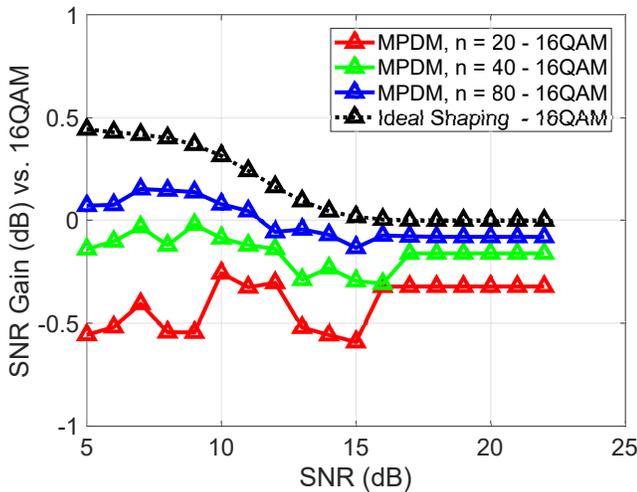


Fig. 11. Comparison of the SNR gain over uniform 16QAM achieved by an MPDM distribution matcher with sequence length $n = 20, 40, 80$. An ideal distribution matcher with no rate loss is also shown.

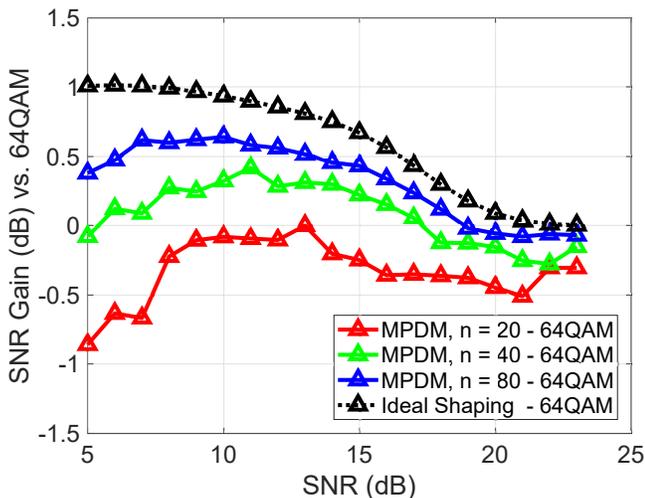


Fig. 12. Comparison of the SNR gain over uniform 64QAM achieved by an MPDM distribution matcher with sequence length $n = 20, 40, 80$. An ideal distribution matcher with no rate loss is also shown.

16QAM is not reduced significantly compared with CCDM 16QAM, we note that significant reductions in penalty are observed for MPDM 64QAM and 256QAM, and penalty of approximately 0.5 dB or better is achievable from 5–20 dB.

Fig. 11 shows the SNR gain relative to uniform 16QAM for MPDM 16QAM with DM lengths of $n = 20, 40, 80$. We note that only the longest DM achieves a small gain for SNRs from 5–12 dB, while shorter DMs incur some penalty.

Fig. 12 shows the SNR gain relative to uniform 64QAM for MPDM 64QAM with DM lengths of $n = 20, 40, 80$. We note that better than uniform performance can be achieved over a wide range of SNRs with $n = 40$. Additionally, most of the available shaping gain can be achieved over a wide range of SNRs for $n = 80$.

Fig. 13 shows the SNR gain relative to uniform 256QAM for MPDM 256QAM with DM lengths of $n = 20, 40, 80$.

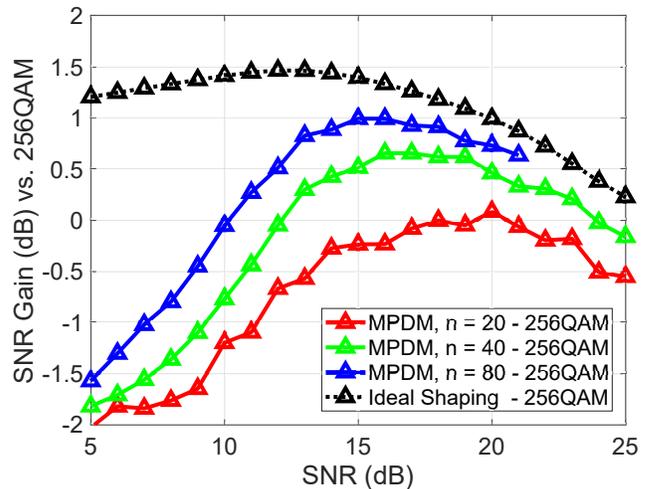


Fig. 13. Comparison of the SNR gain over uniform 256QAM achieved by an MPDM distribution matcher with sequence length $n = 20, 40, 80$. An ideal distribution matcher with no rate loss is also shown.

We again note that $n = 40$ can achieve again compared with uniform 256QAM over a wide range of SNRs, and that $n = 80$ can achieve most of the available shaping gain from approximately 13–22 dB. For all lengths considered here, we note that there are significant penalties for operation in the low SNR regime.

V. CONCLUSIONS

We have examined the performance of multiset partition distribution matching (MPDM) over a wide range of SNRs with 16QAM, 64QAM and 256QAM, for short sequence lengths of 20, 40 and 80 symbols. While the available gains are limited in the region considered for 16QAM, significantly better than uniform performance may be achieved for 64QAM and 256QAM using an MPDM with length $n = 40$, and most of the available shaping gain may be achieved over a wide range of SNRs for 64QAM and 256QAM when considering an MPDM with length $n = 80$.

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] 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.
- [3] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, “On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel,” *Journal of Lightwave Technology*, vol. 34, no. 22, pp. 5063–5073, Nov. 2016.
- [4] J. Cho, X. Chen, S. Chandrasekhar, G. Raybon, R. Dar, L. Schmalen, E. Burrows, A. Adamiecki, S. Corteselli, Y. Pan *et al.*, “Trans-atlantic field trial using probabilistically shaped 64-QAM at high spectral efficiencies and single-carrier real-time 250-Gb/s 16-QAM,” in *Proc. Optical Fiber Communication Conference (OFC)*. Los Angeles, CA, USA: Paper Th5B.3, Mar. 2017.
- [5] M. Yankov, K. Larsen, and S. Forchhammer, “Temporal probabilistic shaping for mitigation of nonlinearities in optical fiber systems,” *Journal of Lightwave Technology*, vol. PP, no. 99, pp. 1–9, 2017.

- [6] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.
- [7] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *IEEE Electronics Letters*, vol. 33, no. 6, pp. 457–458, Mar. 1997.
- [8] T. Yoshida, M. Karlsson, and E. Agrell, "Low-complexity variable-length output distribution matching with periodical distribution uniformization," in *Proc. Optical Fiber Communication Conference (OFC)*. San Diego, CA, USA: Paper M4E.2, Mar. 2018.
- [9] M. Dia, V. Aref, and L. Schmalen, "A compressed sensing approach for distribution matching," *CoRR*, vol. abs/1804.00602, 2018. [Online]. Available: <http://arxiv.org/abs/1804.00602>
- [10] P. Schulte and F. Steiner, "Shell mapping for distribution matching," *arXiv preprint arXiv:1803.03614*, Mar. 2018.
- [11] J. Cho, S. Chandrasekhar, R. Dar, and P. J. Winzer, "Low-complexity shaping for enhanced nonlinearity tolerance," in *Proc. European Conference on Optical Communications (ECOC)*. Düsseldorf, Germany: Paper W.1.C.2, Sep. 2016.
- [12] Y. C. Gültekin, W. van Houtum, S. Serbetli, and F. M. Willems, "Constellation shaping for IEEE 802.11," in *Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Montreal, QB, Canada, Oct. 2017.
- [13] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Multiset-Partition Distribution Matching," *arXiv preprint arXiv:1801.08445*, Jan. 2018.
- [14] E. P. da Silva, M. P. Yankov, F. Da Ros, S. Forchhammer, M. Galili, L. K. Oxenløwe, and D. Zibar, "Experimental comparison of gains in achievable information rates from probabilistic shaping and digital backpropagation for DP-256QAM/1024QAM WDM systems," in *Proc. European Conference on Optical Communications (ECOC)*, Düsseldorf, Germany, Sep. 2016.
- [15] G. Böcherer, P. Schulte, and F. Steiner, "High throughput probabilistic shaping with product distribution matching," *arXiv preprint arXiv:1702.07510*, Feb. 2017.
- [16] F. Steiner, P. Schulte, and G. Böcherer, "Approaching waterfilling capacity of parallel channels by higher order modulation and probabilistic amplitude shaping," in *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, March 2018, pp. 1–6.
- [17] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. New York, NY, USA: John Wiley & Sons, 2006.
- [18] W. Feller, *An introduction to probability theory and its applications. Volume 1*, 3rd ed. John Wiley & Sons, 1968.
- [19] G. Böcherer and R. Mathar, "Matching dyadic distributions to channels," in *Proceedings of the 2011 Data Compression Conference*, ser. DCC '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 23–32. [Online]. Available: <http://dx.doi.org/10.1109/DCC.2011.10>
- [20] G. Böcherer, "On joint design of probabilistic shaping and forward error correction for optical systems," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, March 2018, pp. 1–36.
- [21] K. Ishii, K. Dohi, K. Kubo, K. Sugihara, Y. Miyata, and T. Sugihara, "A study on power-scaling of triple-concatenated fec for optical transport networks," in *2015 European Conference on Optical Communication (ECOC)*, Sept 2015, pp. 1–3.