Abstract—Product distribution matching (PDM) is proposed to generate target distributions over large alphabets by combining the output of several parallel distribution matchers (DMs) with smaller output alphabets. The parallel architecture of PDM enables low-complexity and high-throughput implementation. PDM is used as a shaping device for probabilistic amplitude shaping (PAS). For 64-ASK and a spectral efficiency of 4.5 bits per channel use (bpcu), PDM is as power efficient as a single full-fledged DM. It is shown how PDM enables PAS for parallel channels present in multi-carrier systems like digital subscriber line (DSL) and orthogonal frequency-division multiplexing (OFDM). The key feature is that PDM shares the DMs for lower bit-levels among different sub-carriers, which improves the power efficiency significantly. A representative parallel channel example shows that PAS with PDM is 0.93 dB more power efficient than conventional uniform signaling and PDM is 0.35 dB more power efficient than individual per channel DMs.

Index Terms—Probabilistic amplitude shaping, Distribution matcher, Rate adaptation, Parallel channels, Bit Loading, DSL, OFDM, Coded modulation

I. INTRODUCTION

Higher-order modulation is indispensable in mobile, satellite, cable, and fiber-optic communication to achieve the high spectral efficiency (SE) required for data applications. Transceivers must be flexible, i.e., they should support different SEs so they can adapt to the link quality at hand and deliver the best possible connectivity. Conventional coded modulation uses uniform distributions on the constellation points. This has two disadvantages. First, uniform distributions suffer a power inefficiency of up to 1.53 dB. Second, flexibility can be achieved only by supporting a large number of modcods, i.e., combinations of modulation formats and channel codes. For example DVB-S2X requires supporting 116 modcods [1].

One approach that has been proposed is geometric shaping (GS) [2], [3] which uses constellations with non-equidistant signal points. While improved power efficiency was observed, the problem of flexibility remains. A second approach is probabilistic shaping (PS) that uses equidistant signal points with a non-uniform distribution. For an overview of PS schemes, see [4, Sec. II] and references therein. Recently, we proposed probabilistic amplitude shaping (PAS) [4], a PS architecture that concatenates a distribution matcher (DM) [5], [6] as a shaping device with forward error correction (FEC), see Fig. 1. PAS achieves the optimal power efficiency and enables flexible SE with only one FEC code [4, Sec. VIII]. PAS has been successfully integrated with low-density parity-check (LDPC) codes [4], turbo codes [7], SC-LDPC codes [8], polar codes [9], and nonbinary codes [10]. In comparison [11], PAS is over 0.3 dB more power efficient than Non-Uniform Constellations (NUC) [3], a GS implementation advocated by the Advanced Television Systems Committee (ATSC) 3.0 standard. PAS is being considered for inclusion in the 5G standard [12]. The benefits of PAS for fiber-optic communication were recently showcased in a field trial [13] and future optical modems will implement PS [14, Sec. V-A].

The enabling technology for PAS is the DM, which transforms a binary data sequence into a sequence of symbols with a desired distribution. For an overview of existing DM algorithms, see [6, Sec. I] and references therein. For implementation, fixed-to-fixed length DMs are desirable. For high-throughput applications, efficient DM encoding is required. Furthermore, fixed-to-fixed length DMs require a large block length to work well [15].

In many practical settings, the data link is well modelled by a set of non-interacting parallel channels. Examples
include multi-carrier transmission such as orthogonal frequency division multiplexing (OFDM), discrete multitone (DMT), and multi-antenna transceivers when the singular value decomposition (SVD) of the channel matrix is used to orthogonalize the system. Employing current DM algorithms in such scenarios is challenging, as techniques like bit-loading partition the transmitted sequence in several short segments, each with an individual constellation size and distribution, which potentially causes a significant rate loss.

In this work, we propose a novel DM architecture called product distribution matching (PDM), which internally uses a collection of parallel DMs with smaller output alphabets to synthesize the desired distribution as a product distribution. A preferable implementation uses binary output alphabets for the individual DMs. This approach both facilitates high-throughput applications by parallelization and reduces the rate loss for short output lengths, which makes the PDM particularly amenable for large constellations and high-throughput.

In the final part of this work, we propose extended PDM for parallel channels, which shares the component DMs for lower bit-levels among different sub-carriers. Extended PDM can be applied, e.g., in OFDM and DMT. We provide a representative example where extended PDM is 0.93 dB and 0.35 dB more power efficient than uniform signaling and individual per sub-carrier DMs, respectively, and operates close to the waterfilling limit. All simulation results were obtained using the DM implementations by [16].

This work is structured as follows. Sec. II reviews DMs and PAS and states achievable rate expressions for system design. In Sec. III, we introduce the PDM architecture and present finite length simulation results for 64-QAM. Sec. IV shows how extended PDM can be utilized to operate PAS close to the waterfilling limit of parallel channels. We conclude in Sec. V.

II. PRELIMINARIES

A. Distribution Matching (DM)

DMs [5], [6] transform a sequence of uniformly distributed input bits into an output sequence of symbols from an alphabet \( \mathcal{A} \) with a desired distribution. A fixed-to-fixed length DM maps \( k \) input bits \( d^k \) to \( n \) output symbols \( a^n = \text{dm}(d^k) \). The mapping \( \text{dm} \) is invertible, i.e., \( d^k \) can be recovered from \( a^n \) by applying the inverse mapping \( \text{dm}^{-1} \). Fixed-to-fixed length DMs can be implemented by the constant composition distribution matcher (CCDM) [6], for binary output alphabets see also [17]. A DM is characterized by the following parameters.

- The rate is
  \[
  R_{\text{dm}} = \frac{k}{n} \left\lceil \frac{\text{bits}}{\text{output symbol}} \right\rceil. \tag{1}
  \]

- The output distribution is
  \[
  P_{A}(a) = \sum_{d^k \in (0,1)^k} P_{\text{dm}}(d^k) P_{d^k}(a), \quad a \in \mathcal{A} \tag{2}
  \]
  where \( P_{d^k} \) is the empirical distribution of the sequence \( a^n \), i.e.,
  \[
  P_{a^n}(a) = \frac{|\{i : a_i = a\}|}{n}, \quad a \in \mathcal{A}. \tag{3}
  \]

B. Amplitude Shift Keying Modulation

We consider \( 2^m \)-amplitude shift keying (ASK) constellations

\[
\mathcal{X} = \{ \pm1, \pm3, \ldots, \pm(2^m - 1) \} \tag{5}
\]

with amplitude alphabet

\[
\mathcal{A} = \{1,3,\ldots,2^m - 1\}. \tag{6}
\]

We use label functions \( \beta : \mathcal{X} \rightarrow \{0,1\}^m \) and corresponding bit mappers \( \chi : \{0,1\}^m \rightarrow \mathcal{X} \). For all labels \( B = B_1 \ldots B_m \) in this work, the first bit \( B_1 \) labels the sign \( S \) according to

\[
B_1 = \begin{cases} 0, & S = 1 \\ 1, & S = -1. \end{cases} \tag{7}
\]

Consequently, \( B_2 \ldots B_m \) label the amplitudes and each label function \( \beta \) implies an amplitude label function \( \beta_A \) and each bit-mapper \( \chi \) implies an amplitude bit-mapper \( \chi_A \). Two labels are of special interest, namely the binary reflected Gray code (BRGC) [18] and the natural based binary code (NBBC) [4, Sec. VI.C] where the amplitude label is a natural binary code (NBC). The two labels are illustrated for 8-ASK in Table I.

| BRGC  | 000 | 001 | 011 | 010 | 110 | 111 | 101 | 100 |
|------|-----|-----|-----|-----|-----|-----|-----|-----|
| NBBC | 000 | 001 | 010 | 011 | 111 | 110 | 101 | 100 |

- The rate loss is the difference of the DM rate and the entropy rate of a discrete memoryless source (DMS) \( P_A \), i.e.,
  \[
  R_{\text{loss}} = \mathbb{H}(A) - \frac{k}{n}. \tag{4}
  \]

By [6, Sec. III.B], the rate loss of CCDM vanishes for large output lengths \( n \). In this work, we are interested in DMs with relatively short output lengths and we therefore need to account for the rate loss in our system design.

C. PAS Transmitter

The PAS architecture implements probabilistically shaped ASK modulation. The PAS transmitter is displayed in Fig. 3 and works as follows (for a more detailed description, see [4, Sec. IV]). A DM maps \( k \) data bits to \( n \) amplitudes \( A^n \), which are represented by \( n(m-1) \) amplitude bits. The amplitude bits and \( \gamma n \) additional data bits are multiplied with the parity generating part \( P \) of a systematic generator matrix \( [I|P] \) to generate \( (1-\gamma)n \) redundancy bits. The redundancy bits and the additional data bits are mapped to \( n \) signs \( S^n \), which are multiplied symbolwise with the amplitudes \( A^n \). The FEC code instantiated by \( P \) has rate

\[
c = \frac{n(m-1) + \gamma n}{mn} = \frac{m-1 + \gamma}{m} \tag{8}
\]
and the fraction of signs used for data bits is
\[ \gamma = 1 - (1 - c)m. \] (9)

PAS requires \( 0 \leq \gamma \leq 1 \). The transmission rate of PAS is the number of data bits per ASK symbol given by
\[ R_t = \frac{k}{n} + \gamma. \] (10)

**D. Channel Model**

The generated signal points \( A_i \cdot S_i \) are multiplied by the constellation scaling \( \Delta \) and transmitted over an additive white Gaussian noise (AWGN) channel. We define \( X_i = \Delta \cdot A_i \cdot S_i \).

At a generic time instance, the channel model is
\[ Y = X + Z \] (11)

where \( Z \) is zero mean Gaussian noise with unit variance. The signal-to-noise ratio (SNR) is
\[ \text{SNR} = \mathbb{E}[X^2]. \] (12)

**E. PAS Achievable Rate**

We consider a PAS receiver with a bit-metric decoder. The PAS transmitter defines the label \( B^{\text{fec}}_1 = \beta^{\text{fec}}(X) \) where \( B^{\text{fec}}_1 = B_1 \) is the sign label and where \( B^{\text{fec}}_2 \cdots B^{\text{fec}}_m \) is the amplitude label. We assume a uniform sign distribution, i.e.,
\[ P_{B_1}(0) = P_{B_1}(1) = P_S(-1) = P_S(1) = \frac{1}{2}. \] (13)

We refer to [4, Sec. IV.A] for a justification of this assumption. A binary demapper calculates the soft-informations
\[ L_j = \log \frac{P_{Y|B^{\text{fec}}_j}(y|0)}{P_{Y|B^{\text{fec}}_j}(y|1)}, \quad j = 1, 2, \ldots, m \] (14)

which are passed to a binary decoder. By [19], an achievable rate for a bit-metric decoder is
\[ R_{\text{bmd}} = \left[ \mathbb{H}(X) - \sum_{j=1}^{m} \mathbb{H}(B^{\text{fec}}_j|Y) \right]^+ \] (15)

where \([\cdot]^+ = \max(0, \cdot). \) For PAS, \( R_{\text{bmd}} \) must be evaluated using \( P_X = P_A P_S \). The rate \( R_{\text{bmd}} \) is an achievable rate for the PAS receiver with BMD if
\[ \mathbb{H}(A) + \gamma = R_{\text{bmd}}. \] (16)
Note that the label $B^{dm}$ is not required to be the same as the label $B^{fc}$ that is used by the FEC encoder and decoder. We choose the NBC for the amplitude label $B_2^{dm}, \ldots, B_m^{dm}$, the BRGC for the FEC label $B^{fc}$ and we optimize (20) over the binary distributions $P_{B^{dm}_j}$, $j = 2, \ldots, m$ (recall that the sign distribution $P_{B_1}$ is uniform) and the constellation scaling $\Delta$. In Fig. 4, we display the resulting achievable rate for 8-ASK. We observe that the product constraint (19) leads to virtually no performance loss.

Remark 1. The information-theoretic work [20] considered only the case when $B^{dm} = B^{fc}$, in which case (20) becomes

$$R_{bicm} = \sum_{j=1}^{m} \mathbb{E}(B_j^{fc}, Y)$$

which is the so-called BICM capacity. As shown in Fig. 4, $R_{bicm}$ is less power efficient than $R_{bicm}^{II}$ although the difference is small.

B. PDM

PDM can efficiently generate the product distributions introduced in the previous subsection. The PDM is displayed in Fig. 2. $k$ binary data bits are demultiplexed into $m - 1$ parallel blocks of lengths $k_2, \ldots, k_m$. The $m - 1$ parallel binary DMs output $m - 1$ shaped binary sequences of length $n$. A bit mapper $\chi_{dm}$ recombiners the $m - 1$ sequences and outputs one shaped amplitude sequence of length $n$.

C. PDM Rate Loss

The rate and the output distribution of the $j$th DM is $k_j/n$ and $P_{B_j^{dm}}$, respectively. The total rate of the PDM is

$$\frac{k}{n} = \frac{k_2 + \cdots + k_m}{n}$$

and the total rate loss of the PDM is the sum of the individual rate losses, i.e.,

$$R_{loss} = \sum_{j=2}^{m} \left[ \mathbb{H}(B_j^{dm}) - \frac{k_j}{n} \right].$$

D. PDM for the AWGN Channel

For the AWGN channel, we use the NBC for the bit-mapper $\chi_{dm}$ and we choose binary DM distributions that minimize the overall power. Ignoring the rate loss for now, the optimization problem is

$$\begin{align*}
\text{minimize} & \quad \mathbb{E}[X^2] \\
\text{subject to} & \quad \sum_{j=2}^{m} \mathbb{E}(B_j) = R_{dm} \\
& \quad X = \chi^{nbhc}(B).
\end{align*}$$

To account for the rate loss, we replace the sum-entropy constraint in (24) by a sum-rate constraint, where the $j$th rate $k_j/n$ is the rate required to implement the DM output distribution $P_{B_j}$. Altogether, we choose the component DMs via

$$\begin{align*}
\text{minimize} & \quad \mathbb{E}[X^2] \\
\text{subject to} & \quad \sum_{j=2}^{m} k_j/n = R_{dm} \\
& \quad X = \chi^{nbhc}(B).
\end{align*}$$

E. Simulation Results

We numerically compare different DM implementations by using 64-ASK and a target SE of $R_t = 4.5$ bpcu. We employ a 32-ary DM as a reference as suggested in [4, Sec. V]. The performance of this system is compared to a PDM setup with 1 ($B_2^{dm}$), 2 ($B_2^{dm}, B_3^{dm}$), 3 ($B_2^{dm}, B_3^{dm}, B_4^{dm}$), 4 ($B_2^{dm}, B_3^{dm}, B_4^{dm}, B_5^{dm}$) and 5 ($B_2^{dm}, B_3^{dm}, B_4^{dm}, B_5^{dm}, B_6^{dm}$) individually shaped bit-levels and corresponding binary DMs. The product distribution has been obtained by following the approach of Sec. III-D, while imposing a uniform distribution on the unshaped bit-levels.

We first consider the results of Fig. 5 which illustrates the finite length loss of all considered configurations. The DM rate loss (4) and the PDM rate loss (23) is converted to an “SNR loss” by

$$\text{SNR}_{loss} = 10 \log_{10} \left( \frac{P_{bicm}^{-1}(P_X, R_t + R_{loss})}{R_{bicm}^{-1}(P_X, R_t)} \right).$$

As a rule of thumb, the following expression may be useful as a rough estimate:

$$\text{SNR}_{loss,awgn} = 10 \log_{10} \left( \frac{\sigma^2(R_t + R_{loss}) - 1}{2 \sigma R_t - 1} \right) \approx R_{loss} \cdot 20 \log_{10} 2 \approx R_{loss} \cdot 6 \text{ dB}.$$

We observe that the PDMs have an aggregated rate loss that is significantly lower than the rate loss of the 32-ary DM. The resulting performance is comparable only for output lengths of more than $10^4$ symbols.

To further illustrate the flexibility of the transmitter design, we consider a coded scenario with a rate 9/10 LDPC block.
code from the DVB-S2 standard [21] of block length 64 800 bits and a corresponding DM output length of 10 800 symbols. This choice allows for a fair comparison, as both the parallel binary DMs and the 32-ary DM have a similar performance. Fifty iterations are used for the belief propagation (BP) decoding.

As shown in Fig. 6, a PDM with 3 shaped bit-levels achieves a similar performance as the 32-ary DM. If only 2 bit-levels are shaped, the loss in energy efficiency is 0.4 dB at a target frame error rate (FER) of $10^{-5}$. Table II illustrates that these observations are reflected by the asymptotic achievable rates of Sec. II-E, which were evaluated for the corresponding optimized distributions. While the required SNRs to achieve an SE of 4.5 bpcu are close for 3, 4 and 5 shaped bit-levels, larger gaps can be observed for 1 or 2 shaped bit-levels.

![Fig. 6. Performance comparison of the proposed PDM for 64-ASK and a target SE of 4.5 bpcu and different number of shaped bits.](image)

| DM configuration      | Required SNR [dB] |
|-----------------------|-------------------|
| 32-ary DM             | 27.13             |
| PDM 1 Bit shaped      | 28.29             |
| PDM 2 Bits shaped     | 27.48             |
| PDM 3 Bits shaped     | 27.35             |
| PDM 4 Bits shaped     | 27.32             |
| PDM 5 Bits shaped     | 27.31             |

IV. PROBABILISTIC SHAPING FOR PARALLEL CHANNELS

A. System Model

We consider $L$ parallel channels with the I/O relation

$$Y_\ell = h_\ell X_\ell + Z_\ell, \quad \ell = 1, 2, \ldots, L.$$  (28)

The noise terms $Z_\ell$ are zero mean Gaussian with unit variance. The $h_\ell$ model the channel gains and we assume that both the receiver and transmitter have full channel state information, i.e., they both know the channel gains $h_\ell$ and the noise variance. We consider coding over $n$ channel uses of each channel, which results in total in $L \cdot n$ channel uses. This choice is for clarity of exposition; the scheme can easily be generalized.

B. Waterfilling [22, Sec. 5.4.6]

The transmitter has an average power budget $P$, i.e., the inputs are subject to the sum-power constraint

$$\frac{1}{L} \sum_{\ell=1}^{L} \mathbb{E}[X_\ell^2] \leq P.$$  (29)

The average SE

$$\frac{1}{L} \sum_{\ell=1}^{L} \frac{1}{2} \log_2 (1 + h_\ell^2 P_\ell)$$  (30)

is achievable with the channel inputs $X_\ell$ being independent zero mean Gaussian with variance $P_\ell$. The average SE is maximized by waterfilling, i.e.,

$$P_\ell^* = \left[ \frac{1}{\lambda} - \frac{1}{h_\ell^2} \right]^+, \quad \lambda: \frac{1}{L} \sum_{\ell=1}^{L} P_\ell^* = P.$$  (31)

Suppose that $P_\ell^*$ is positive. The SE allocated to channel $\ell$ is then

$$C_\ell = \frac{1}{2} \log_2 \frac{h_\ell^2}{\lambda}.$$  (32)

Based on $C_\ell$, we choose the constellation size $2^{m_\ell}$ so that

$$m_\ell \approx C_\ell + 1.$$  (33)

to avoid reduced SE because of too small constellation sizes. Let $m = \max_\ell m_\ell$ denote the maximum constellation size.

C. PAS for Parallel Channels

PAS can easily be combined with parallel channels. This is illustrated in Fig. 7. A DM device transforms data bits into a sequence of amplitudes for each channel, which are then combined with sign bits originating from a common encoding device. In its simplest form, this DM device consists of individual DMs, each with its output alphabet size matched to the corresponding constellation size, see Fig. 8.

D. PDM for Parallel Channels

The PDM suggests an alternative way to generate $L$ amplitude sequences for distinct constellation sizes. For example, suppose we have $L = 2$ different channels and need a length $n$ amplitude sequence for 4-ASK and a length $n$ sequence for 8-ASK. The PDM needs one binary DM for 4-ASK and two binary DMs for 8-ASK. As illustrated in Fig. 10, the idea is now to use for the first amplitude bit-level $B_2$ of 4-ASK and 8-ASK a single binary DM with output length $n_2 = 2 \cdot n$ and to generate the second amplitude bit-level $B_3$ for 8-ASK by a second binary DM with output length $n_3 = n$. The potential benefit of this approach is twofold: first, using PDM should reduce the rate loss, and second, replacing two DMs of lengths $n$ by one single DM of length $2 \cdot n$ should reduce the rate loss even further. Fig. 9 shows this extended PDM scheme. It provides the same interface to PAS as the naive approach that uses $L$ individual DMs.
Simultaneously using one DM on more than one constellation size imposes restrictions on the distribution families that can be generated by extended PDM. We next argue how extended PDM can be used to generate families of Gaussian-like distributions. The maximum constellation size is $2^m$ and we choose the $m−1$ DM output distributions so that an NBBC mapper generates a Gaussian-like distribution. By grouping $2^j$ neighbouring signal points together, the distribution of these signal point groups is still Gaussian-like, and it is given by the product distribution generated by the first $m−1−j$ DMs. This suggests that by using only $m−1$ DMs, we can simultaneously generate Gaussian-like distributions on $4, 8, \ldots, 2^m$-ASK constellations. An example is shown in Fig. 10.

E. Parametrization

We next state the parameters of the FEC code and the PDM so that the parallel PAS operates at a specific SE. For the considered case where we use each of the $L$ channels $n$ times, the block length of the binary FEC code is

$$n_{\text{code}} = \sum_{\ell=1}^{L} m_{\ell} \cdot n \quad (34)$$
and formulas (8) and (9) generalize to

$$c = \frac{\sum_{\ell=1}^{L} (m_\ell - 1 + \gamma)}{\sum_{\ell=1}^{L} m_\ell}$$  \hspace{1cm} (35)$$

$$\gamma = 1 - (1 - c) \frac{1}{L} \sum_{\ell=1}^{L} m_\ell.$$  \hspace{1cm} (36)

The DM output lengths are given by

$$n_j = \sum_{\ell=1}^{L} \mathbb{1}(m_\ell \geq j) \cdot n, \quad j = 2, 3, \ldots, m$$  \hspace{1cm} (37)

and the corresponding DM input lengths are $k_2, k_3, \ldots, k_m$. The average SE of the overall system is now

$$R_t = \frac{\sum_{j=2}^{m} k_j}{L \cdot n} + \gamma$$  \hspace{1cm} (38)

$$= \frac{1}{L \cdot n} \left[ \sum_{j=2}^{m} \mathbb{H}(D_{j}^{dm})n_j \right] + \gamma - R_{\text{loss}}.$$  \hspace{1cm} (39)

### F. Waterfilling for PAS

For the $L$ parallel channels, suppose we have chosen the constellation sizes $2^{m_\ell}$, $\ell = 1, \ldots, L$ and suppose further we have chosen the code rate $c$ and thereby the fraction $\gamma$ of signs used for data bits. To achieve the target rate $R_t$, the rate assigned to the amplitudes is thus $R_{dm} = R_t - \gamma$, which results in the following constraint for the amplitude distributions (ignoring the rate loss):

$$1 \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{H}(A_\ell) = R_{dm}.$$  \hspace{1cm} (40)

Recall that the inputs $X_\ell$ are given by $\Delta_\ell A_\ell S_\ell$ where

$$A_\ell S_\ell \in \{ \pm 1, \pm 3, \ldots, \pm (2^{m_\ell} - 1) \}.$$  \hspace{1cm} (41)

The average power on the $\ell$th channel is $\mathbb{E}[|\Delta_\ell A_\ell|^2]$ and depends on the distribution $P_{A_\ell}$ and the constellation scaling $\Delta_\ell$. We use the following strategy: to ensure a similar detection reliability on each channel, independent of the chosen amplitude distributions, we choose

$$\Delta_\ell = \frac{\Delta}{h_\ell}.$$  \hspace{1cm} (42)

In this way, two neighbouring constellation points have the distance $2\Delta$ on all channels. The average power on each channel is $\frac{\Delta^2}{h_\ell^2} \mathbb{E}[A_\ell^2] \propto \frac{1}{h_\ell^2} \mathbb{E}[A_\ell^2]$. Next, we calculate the amplitude distributions by

$$\text{minimize} \quad P_{A_1, \ldots, A_L} \sum_{\ell=1}^{L} \frac{1}{h_\ell^2} \mathbb{E}[A_\ell^2]$$

subject to

$$\frac{1}{L} \sum_{\ell=1}^{L} \mathbb{H}(A_\ell) = R_{dm}.$$  \hspace{1cm} (43)

To account for rate loss, the sum-entropy constraint is replaced by a DM sum-rate constraint. For extended PDM, the sum-entropy and sum-rate expressions from (39) and (38) are used, respectively.

### G. Simulation Results

To evaluate the performance of parallel PAS and extended PDM, we employ the following example of 3 parallel channels,
Observing in Fig. 11 that the PDM setup improves over the reference strategy at a FER of $10^{-2}$ by 0.35 dB. This is mainly because of the decreased rate loss as shown in the asymptotic achievability plot of Fig. 12. We plot the average achievable rate over all parallel channels vs. the average sum power for both schemes and their specific input distributions. The power assignment is optimized via mercury/waterfilling. We also plot three horizontal lines at 3.09 bpcu, 3.099 bpcu and 3.16 bpcu, which denote $R_t$, $R_t + R_{\text{loss,pdm}}$ and $R_t + R_{\text{loss,ref}}$, respectively. The crossing of the last two horizontal lines with their respective achievability curves are labeled as OP$_{\text{pdm,virt}}$ and OP$_{\text{ref,virt}}$. They indicate virtual operating points that would be achievable with the currently used input distributions. Because of the rate loss, the actual operating points are given by the orthogonal projections of these points on the actual SE curve, however. Their difference in SNR of 0.4 dB accurately predicts the gap of 0.35 dB that we observe in the coded result in Fig. 11. Compared to uniform distributions, the asymptotic gain (accounting for the rate loss) is 0.93 dB. The gap to the waterfilling solution is 0.32 dB.

V. CONCLUSION

We proposed product distribution matching (PDM), an architecture that uses binary DMs in parallel. This parallelization enables high-throughput implementations of DMs. The binary component DMs of PDM reduce complexity. We have shown that PDM performs as well as higher-order DMs for long block lengths and that PDM can perform much better than higher-order DMs for short block lengths. We have proposed extended PDM, which enables PAS to operate close to the waterfilling limit of multi-carrier transmission schemes such as OFDM.

ACKNOWLEDGMENT

The authors would like to thank Gerhard Kramer for fruitful discussions and comments on drafts of this manuscript.

REFERENCES

[1] Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications; Part 2: DVB-S2 Extensions (DVB-S2X), European Telecommunication Standards Inst. (ETSI) Std. EN 302 307-2, Rev. 1.1.1, 2014.

[2] M. F. Barsoum, C. Jones, and M. Fitz, “Constellation Design via Capacity Maximization,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun. 2007, pp. 1821–1825.

[3] N. S. Loghin, J. Zöllner, B. Mouhouche, D. Anzarregui, J. Kim, and S. L. Park, “Non-Uniform Constellations for ATSC 3.0,” IEEE Trans. Broadcast., vol. 62, no. 1, pp. 197–203, Mar. 2016.

[4] G. Böcherer, F. Steiner, and P. Schulte, “Bandwidth efficient and rate-matched low-density parity-check coded modulation,” IEEE Trans. Commun., vol. 63, no. 12, pp. 4651–4665, 2015.

[5] G. Böcherer and R. Mathar, “Matching dyadic distributions to channels,” in Proc. Data Compression Conf., Snowbird, UT, USA, 2011, pp. 23–32.

[6] P. Schulte and G. Böcherer, “Constant composition distribution matching,” IEEE Trans. Inf. Theory, vol. 62, no. 1, pp. 430–434, 2016.

[7] P. Yuan, “Rate-matched coded modulation for wireless transmission,” Master’s thesis, Technical University of Munich, Institute for Communications Engineering, 2015.

[8] 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. 8, Apr. 2016.
[9] T. Prinz, “Polar codes for higher-order modulation and probabilistic amplitude shaping,” Master’s thesis, Technical University of Munich, Institute for Communications Engineering, 2016.

[10] J. J. Boutros, F. Jardel, and C. Méasson, “Probabilistic shaping and non-binary codes,” *arXiv preprint arXiv:1701.07976*, 2017.

[11] F. Steiner and G. Böcherer, “Comparison of Geometric and Probabilistic Shaping with Application to ATSC 3.0,” in *Int. ITG Conf. Source Channel Coding*, Hamburg, Germany, Feb. 2017.

[12] “Signal shaping for QAM constellations,” Huawei, Athens, Greece, Tech. Rep., Feb. 2018, 3GPP TSG-RAN no. 88. [Online]. Available: http://www.3gpp.org/ftp/tnsran/WG1_RL1/TSGR1_88/Docs/R1-1701712.zip

[13] W. Idler, F. Buchali, L. Schmalen, E. Lach, R.-P. Braun, G. Böcherer, P. Schulte, and F. Steiner, “Field trial of a 1 Tbit/s super-channel network using probabilistically shaped constellations,” *J. Lightw. Technol.*, vol. 99, no. PP, pp. 1–1, 2017.

[14] K. Roberts, Q. Zhuge, I. Monga, S. Gareau, and C. Laperle, “Beyond 100 Gb/s: Capacity, Flexibility, and Network Optimization,” *Journal of Optical Communications and Networking*, vol. 9, no. 4, pp. C12–C24, Apr. 2017.

[15] P. Schulte and B. C. Geiger, “Divergence scaling of fixed-length, binary-output, one-to-one distribution matching,” *arXiv preprint arXiv:1701.07371*, 2017.

[16] http://www.shapecomm.de.

[17] T. Ramabadran, “A coding scheme for m-out-of-n codes,” *IEEE Trans. Commun.*, vol. 38, no. 8, pp. 1156–1163, Aug. 1990.

[18] F. Gray, “Pulse code communication,” U. S. Patent 2 632 058, 1953.

[19] G. Böcherer, “Achievable rates for shaped bit-metric decoding,” *arXiv preprint*, 2014. [Online]. Available: http://arxiv.org/abs/1410.8075

[20] A. G. i Fabregas and A. Martinez, “Bit-interleaved coded modulation with shaping,” in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2010.

[21] Digital Video Broadcasting (DVB); 2nd Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and Other Broadband Satellite Applications (DVB-S2), Std. EN 302 307, Rev. 1.2.1, 2009.

[22] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[23] Unified high-speed wire-line based home networking transceivers - System architecture and physical layer specification, ITU Std. G.9960, Oct. 2010.