Hierarchical distribution matching with massively parallel interfaces for fiber-optic communications [invited]

Hierarchical Distribution Matching with Massively Parallel Interfaces for Fiber-Optic Communications

Tsuyoshi Yoshida, Erik Agrell, and Magnus Karlsson

(Invited Paper)

***Abstract*—The design of the distribution matching (DM) encoder and decoder is essential in the implementation of probabilistic shaping. Recently, techniques for low-complexity implementation of DM have been studied. This work consists of three contributions on this topic. Firstly, the mismatch between required throughput and clock frequency in the electric circuitry of fiber-optic transceivers is explained. The throughput of one DM module determines the number of parallel DM modules needed, which in turn determines the circuit size and power consumption. Our previously proposed hierarchical DM (HiDM) has massively parallel input/output interfaces and thus around 100 times fewer instances are required compared with run-length-coding-based DM. Secondly, the HiDM construction is exemplified and described in detail for a DM word length of more than 100 symbols. Thirdly, the capability of HiDM to shape probability mass functions suitable for nonlinear fiber-optic channels is demonstrated, considering higher-order moments.**

***Index Terms*—Coding, hierarchical distribution matching, implementation, modulation, optical fiber communication, probabilistic shaping, reverse concatenation, throughput.**

I. Introduction

Constellation shaping has been deeply investigated over several decades to approach the Shannon capacity over the additive white Gaussian noise (AWGN) channel. The two main types of shaping schemes are *geometric shaping* [1] and *probabilistic shaping* (PS) [2], [3]. Fiber-optic communication channels with optical amplifiers are suitable target applications for PS. The first reason is the existence of the linear optical amplifier. When the optical signal is shaped, the average optical power inside an optical modulator is reduced, but the power will soon be recovered by optical amplifiers, which gives an almost linear gain without waveform degradation. The second reason is the channel stability because of the confined waveguide (fiber) transmission.

Probabilistic amplitude shaping (PAS) [4] provides an attractive method to implement PS by using reverse concatenation, which means forward error correction (FEC) inside the shaping. The PAS scheme was early examined in optical fiber communications [5] and had a significant impact on the community. The shaping encoding and decoding functions for PAS are called *distribution matching* (DM) and *distribution dematching* (invDM), resp. DMs can be classified in terms of symbolwise or bitwise operation, and further into computation-based or LUT-based. The LUT-based DMs can be further classified into fixed- or variable-length LUTs. A symbolwise DM directly controls the probability mass function (PMF) of the output symbols [6]–[16], whereas a bitwise DM [17]–[19] controls the probability of the output bits in a bit tributary, and the PMF of the symbol is controlled by multiple binary bit tributaries, according to an architecture called bit-level DM or product DM [20], [21]. LUT-based DMs, on the other hand, provide lower computational complexity at the expense of memory. LUT-based DMs with fixed-length interfaces include [7], [10], [12], [17], while others use (virtually) variable-length interfaces [8], [14], [16], [18].

In the original PAS scheme, *constant-composition DM* (CCDM) [6] was employed, which is a symbolwise, computation-based DM, similar to the arithmetic coding scheme proposed in [22]. Our previous works include DM based on *run-length coding* (RLC) [18], which is a bitwise, LUT-based, variable-length DM, and *hierarchical DM (HiDM)* [12], [23], which is a symbolwise, LUT-based, fixed-length DM. HiDM, having a unique tree structure of LUTs, shows good performance, reasonable implementation complexity leading to low power consumption, high throughput, and small error rate increase in the invDM processing.

A main issue in the design of logic circuitry for optical fiber communications is the mismatch between the required throughput (several 100 Gb/s to a few Tb/s) and the clock frequency of the electrical circuitry (several 100 MHz). Most PS coding schemes operate in a highly sequential manner, so that their numbers of input/output bits per clock cycle (throughput) and their numbers of physical wirings (bus widths) would be one or a few bits. To realize transmission at 1 Tb/s using a 500 MHz clock, 2000 parallel *instances* are required if the number of input/output bits is only one bit per clock cycle. Even if one DM module operating at 500 MHz for 500 Mb/s is very small, the total required circuit area would be 2000 times larger. On the other hand, HiDM can input and output several 100 bits or even 1000 bits per clock cycle because of its massively parallel interfaces. To realize 1 Tb/s from a 500 MHz clock, we need just a few instances. This is an important advantage in high-speed fiber-optic communications.

In this work, we firstly raise important issues of throughput and bus width by exemplifying two of our previously proposed DMs [12], [18]. Next, we explain recommended design principles of HiDM in detail to make our previous work

T. Yoshida is with Information Technology R&D Center, Mitsubishi Electric Corporation, Kamakura, 247-8501, Japan. He also belongs to Graduate School of Engineering, Osaka University, Suita, 505-0871, Japan (e-mail: Yoshida.Tsuyoshi@ah.MitsubishiElectric.co.jp).

M. Karlsson is with the Dept. of Microtechnology and Nanoscience and E. Agrell is with the Dept. of Electrical Engineering, both at Chalmers University of Technology, SE-41296 Gothenburg, Sweden.

This work was partly supported by "Massively Parallel and Sliced Optical Network (MAPLE)," the Commissioned Research of National Institute of Information and Communications Technology (NICT), Japan (project no. 20401) and by the Swedish Research Council (project no. 2017-03702).

TABLE I
PARAMETERS IN REVERSE CONCATENATION PS SYSTEMS.

Notation	Description
m	number of bits per QAM symbol
m_{sb}	number of shaped bit tributaries per QAM symbol
N_{s}	number of PAM symbols per DM word
N_{in}	number of information bits per DM word

[12] more reproducible. Finally, we show that the DM word choice in HiDM can be flexibly adapted to different PMFs of the transmitted symbols by considering higher-order moments. This was partly addressed in [24] to improve the performance over nonlinear fiber links.

II. THROUGHPUT AND BUS WIDTH AT THE INTERFACES

In this section, we compare two previously proposed LUT-based DMs in terms of throughput and bus width. Some key parameters of reverse concatenation PS are defined in Tab. I. For simplicity, nonshaped bits are excluded from the explanation.

A. Run-length-coding-based bitwise DM

As a submodule for bit-level DM, we proposed a binary-output DM with RLC (variable-length coding) [18]. The encoder schematic is shown in Fig. 1. The incoming N_{in} bits are first demultiplexed into K lanes. In each lane j, the input bits are mapped into a binary word s_j in which exactly half of the bits are 1 using a uniformalizer (UFL), which employs bit flipping and adjustment sequence insertion. Blocks of K bits are converted into RLC words, having a length from 1 to 2^K bits, using a variable-length LUT as in Tab. II. Finally, the RLC words are stored in a first-input/first-output (FIFO) buffer, where they are concatenated into a DM word and output with some latency to account for the DM conversion speed. As the RLC codebook is prefix-free, the codewords are uniquely invertible at the receiver by reading the bits from the beginning. The RLC word corresponding to input 111 in Tab. II is chosen to be 00000000 instead of the more natural 0000000, since this makes the DM word length fixed at $(2^K+1)N_{\mathrm{in}}/(2K)$ for all inputs, at the expense of a slightly higher rate loss. More details can be found in [18].

The key element in a hardware implementation of this bitwise DM with RLC is the variable-length LUT. The number of entries (addresses) is significantly smaller than that with a fixed-length LUT to realize the same rate loss. On the other hand, it is known that a variable-length LUT is not straightforward [25], [26]. An available LUT element usually has a fixed bus width at input/output interfaces. Thus how to realize a virtual variable-length LUT with a fixed-length LUT or other available elements is a critical issue for the implementation. According to the exemplified RLC in Tab. II, the bus width at the output interface can be 8 bits. For example, in case that the output length is shorter than 8, arbitrary bits should be padded. The address for writing into the FIFO buffer is updated after writing the current RLC word based on its effective (unpadded) length.

The throughputs for the bitwise DM with RLC in Fig. 1 are K bits at the input interface and $(2^K+1)/2$ bits at the output.

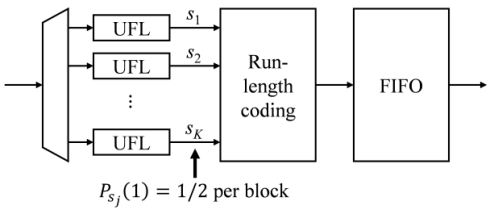

Fig. 1. Schematic for bitwise DM encoding with RLC and periodical uniformalization.

TABLE II
AN EXAMPLE OF RLC ($K = 3$).

Input bits $s_1 s_2 s_3$	Input symbol	RLC word	Effective RLC word length
000	0	1	1
001	1	01	2
010	2	001	3
011	3	0001	4
100	4	00001	5
101	5	000001	6
110	6	0000001	7
111	7	00000000	8

Relevant values of K are from 3 to 6, so the throughputs range from 3 to 6 bits at the input and from 4.5 to 32.5 bits at the output. In these cases, the bus widths should be 6 and 33 bits at input and output interfaces, resp. Since this is a bit-level DM, m_{sb} instances are needed control the PMF of one quadrature amplitude modulation (QAM) symbol. At a clock frequency of 500 MHz, the minimum throughput at the encoding output is $4.5 \cdot 500 = 2.25$ Gsymbol/s. To achieve a symbol rate of 100 Gsymbol/s, $\lceil 2 \cdot 100/2.25 \rceil m_{\mathrm{sb}} = 89 m_{\mathrm{sb}}$ instances are required for polarization-multiplexed PS-QAM signal generation. At the same condition, the throughput of CCDM [6] is one pulse amplitude modulation (PAM) symbol per clock cycle, i.e., 500 Msymbol/s, and $\lceil 2 \cdot 100/0.5 \rceil = 400$ instances are required. This would be the same for other computation-based symbolwise DMs. If m_{sb} is 4 (i.e., 64-QAM), the required number of instances is comparable between CCDM and bit-level DM with RLC, but CCDM needs complex arithmetic coding [6], [12]. A drawback with bitwise DMs such as this RLC-based scheme is that the obtained PMFs are usually constrained to products of bit probabilities[1], which is not the case for symbolwise DMs.

B. Hierarchical DM

Fig. 2 shows the schematic of HiDM, which is a LUT-based fixed-length-to-fixed-length conversion scheme [12]. The parameters in HiDM are defined in Tab. III. The N_{in} input bits (excluding sign bits) are partitioned and input to LUTs, hierarchically placed in L layers. In the top layer, an LUT receives s_L bits and outputs $u_L = t_{L-1} r_{L-1}$ bits. In layer $\ell = L-1, L-2, \ldots, 2$, each LUT receives r_ℓ bits from layer $\ell+1$ and s_ℓ bits from the input of the DM as information bits. Totally $v_\ell = r_\ell + s_\ell$ bits are converted into $u_\ell = t_{\ell-1} r_{\ell-1}$ bits. These bits are fed into $t_{\ell-1}$ LUTs in layer $\ell - 1$, which each receive $r_{\ell-1}$ bits. In layer 1, each LUT receives r_1 bits from layer 2 and s_1 bits from the input of the DM. Totally

[1]If a bitwise DM generates binary DM words with a fixed number of ones, a parallel amplitude architecture [27] can approximate an arbitrary PMF.

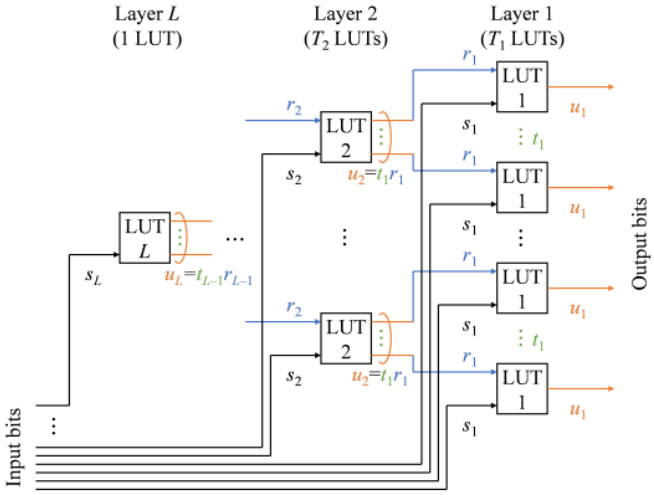

Fig. 2. Schematic of HiDM encoding.

TABLE III
KEY PARAMETERS IN HIDM.

Notation	Description
ℓ	layer index
L	number of layers
t_ℓ	number of LUTs in layer ℓ connected to a single LUT in layer $\ell+1$
T_ℓ	number of LUTs in layer ℓ
r_ℓ	number of input bits in an LUT in layer ℓ from layer $\ell+1$
s_ℓ	number of input information bits in an LUT in layer ℓ
v_ℓ	number of input bits in an LUT in layer ℓ ($r_\ell+s_\ell$)
u_ℓ	number of output bits in an LUT in layer ℓ ($mN_\mathrm{s}/2$ if $\ell=1$ or $t_{\ell-1}r_{\ell-1}$ else)

$v_1 = r_1 + s_1$ bits are converted into u_1 bits, which corresponds to u_1/m_sb QAM symbols. The number of DM output bits and QAM symbols are $m_\mathrm{sb}N_\mathrm{s}/2 = T_1u_1 = (\Pi_{\ell=1}^{L-1}t_\ell)u_1$ and $N_\mathrm{s}/2$, resp.

HiDM has massively parallel input/output interfaces, which are well suited for hardware implementation. The bus widths are $\sum_{\ell=1}^{L} T_\ell s_\ell$ and T_1u_1 at the input and the output of the DM encoding, resp. If for example $L=7$, $T_\ell = 2^{7-\ell}$, $s_1 = \cdots = s_6 = 5$, and $s_7 = 10$, then the bus widths are 640 bits at both the input and the output interfaces. At a clock frequency of 500 MHz, the output throughput is 320 Gb/s or $320/m_\mathrm{sb}$ Gsymbol/s. Under a symbol rate of 100 Gsymbol/s, $\lceil 2\cdot 100/(320/m_\mathrm{sb})\rceil = \lceil (5/8)m_\mathrm{sb}\rceil$ instances are required. Thus, the larger bus width supports around 100 times larger throughput, which requires 100 times fewer instances compared with the RLC-based DM in Sec. II-A.

III. DESIGN AND EVALUATION OF HIDM

A fiber-optic communication channel can be approximated by the AWGN channel with an average power constraint if the dominant impairment is amplified spontaneous emission noise from optical amplifiers. The target PMF for such channel is the discrete Gaussian, or Maxwell–Boltzmann (MB) distribution. For simplicity, the shaped QAM symbols are obtained by combining two shaped PAM symbols. In an example for PS-256-QAM generation [12], the total number of bits per PAM symbol $m/2$ is 4, and both the sign bit (the most significant bit) and the least significant bit are not shaped. Only the second and third significant bits are shaped in each dimension, so that $m_\mathrm{sb}=4$. Tab. IV exemplifies the parameters used. The number of DM input bits per DM word $\sum_{\ell=1}^{L} T_\ell s_\ell$ is 507, and the number of DM output bits per DM word $m_\mathrm{sb}N_\mathrm{s}/2 = T_1u_1$ is 640. Thus the maximum spectral efficiency per 2D symbol is $\beta = 2(1+mN_\mathrm{in}/N_\mathrm{s}) = 2(2+507/320) = 7.169$ bit per channel use (bpcu). The entropy of a 2D symbol $2H(X)$ will be larger than β, where X denotes a PAM symbol.

TABLE IV
CHOSEN PARAMETERS USED IN [12, TAB. IV, FIG. 4].

ℓ	t_ℓ	T_ℓ	r_ℓ	s_ℓ	v_ℓ	u_ℓ
7		1		5	5	12
6	2	2	6	5	11	12
5	2	4	6	5	11	12
4	2	8	6	5	11	12
3	2	16	6	5	11	12
2	2	32	6	5	11	12
1	2	64	6	3	9	10

The values of T_ℓ, v_ℓ, and u_ℓ determine the accumulated size of the LUTs, i.e., $\sum_{\ell=1}^{L} T_\ell 2^{v_\ell} u_\ell$ for DM. If a simple mirror structure is employed for the invDM, its size will be[2] $\sum_{\ell=1}^{L} T_\ell 2^{u_\ell} v_\ell$. Thus, there would be practical constraints on the values of v_ℓ and u_ℓ, which depend on the acceptable hardware resource usage. Under such constraints, a binary tree structure ($t_\ell = 2$, $\forall \ell$) gives the best shaping performance.

The LUT contents are determined from layer 1 and up. There are 2^{u_1} output-word candidates for LUT1, of which 2^{v_1} should be selected based on some criterion, e.g., minimum average symbol energy $E = \mathbb{E}[X^2]$. Thus, the output-word candidates are sorted by increasing E, assuming a Gray-mapped PAM constellation. The top 2^{v_1} candidates are selected, and assigned input symbols in natural order (i.e., $0\cdots00$, $0\cdots01$, …, $1\cdots11$), with $0\cdots00$ assigned to the word with the smallest E. The process then continues with layers $\ell = 2,\ldots,L$. There are 2^{u_ℓ} output-word candidates for the level-ℓ LUT. For each candidate, E is computed based on the selected contents for LUTs $1,\ldots,\ell-1$. The output words are again sorted by increasing E, and the top 2^{v_2} are selected.

We generated PS-256-QAM signals having a DM word length of 320 16-PAM symbols by employing CCDM [6] and HiDM [12]. The target PMF for CCDM was set to the MB distribution with $\beta = 7.169$ bpcu. For HiDM, the scheme exemplified above and in Tab. IV was used, which also has $\beta = 7.169$ bpcu. In Tab. V [12], some key statistics of the shaped PAM symbols X and QAM symbols X_c are summarized, viz. the PMF P_X, average QAM symbol energy $E_\mathrm{c} = \mathbb{E}[|X_\mathrm{c}|^2]$, QAM symbol entropy $H(X_\mathrm{c}) = 2H(X)$, maximum spectral efficiency β, rate loss $R_\mathrm{loss} = H(X_\mathrm{c}) - \beta$, and constellation gain $G = (2^\beta - 1)d_\mathrm{min}^2/(6E_\mathrm{c})$, where d_min denotes the minimum Euclidean distance. The rate loss of a QAM symbol was 0.07 and 0.08 bpcu for CCDM and HiDM, resp. In each case, the constellation gain G was more than 1 dB, while G is 0 dB for uniform square QAM. The gap in G from the ideal Maxwell–Boltzmann (MB) distribution was within 0.4 dB even though we did not shape the least

[2]There may exist techniques to reduce the LUT size without sacrificing performance.

TABLE V
STATISTICS OF THE SHAPED SYMBOLS [12].

	CCDM	HiDM	MB
N_s (PAM symbols)	320	320	–
$P_{\|X\|}(1)$	0.2453	0.2376	0.2628
$P_{\|X\|}(3)$	0.2453	0.2376	0.2355
$P_{\|X\|}(5)$	0.1625	0.1684	0.1891
$P_{\|X\|}(7)$	0.1625	0.1684	0.1360
$P_{\|X\|}(9)$	0.0719	0.0757	0.0877
$P_{\|X\|}(11)$	0.0719	0.0757	0.0506
$P_{\|X\|}(13)$	0.0203	0.0183	0.0262
$P_{\|X\|}(15)$	0.0203	0.0183	0.0121
E_c	74.00	74.70	68.31
$H(X_c)$ (bpcu)	7.242	7.252	7.169
β (bpcu)	7.169	7.169	7.169
R_{loss} (bpcu)	0.073	0.083	0
G (dB)	1.097	1.056	1.444

significant bit.[3]

IV. FLEXIBLE TUNING OF TWO-DIMENSIONAL PMFS

In fiber-optic links where AWGN is not the dominant impairment, different PMFs than MB can give better performance. The received signal-to-noise ratio (SNR) after propagation through a nonlinear fiber-optic channel depends on the transmitted PMF, especially for short links with negligible chromatic and polarization-mode dispersion, where the transmitted waveform shape is maintained. The nonlinear self-channel interference increases with the excess kurtosis $\Phi = \mathbb{E}[|X_c|^4]/\mathbb{E}^2[|X_c|^2] - 2$ [28], [29] of the QAM symbols X_c, which equals 0 for a Gaussian distribution.

HiDM can shape PMFs in an arbitrary number of dimensions as long as the complexity is acceptable. Here, we improve the tolerance to fiber nonlinearity by two-dimensional shaping using HiDM. The LUTs are designed as in Sec. III, except that the output-word candidates are sorted by increasing $\mathbb{E}\left[|X_c|^{F/2}\right]$, for some $F = 1, 2, \ldots, 8$, to reduce Φ.

As in [24], the base constellation is 64-QAM, 32-QAM, or a 1:1 hybrid of 16-QAM and 32-QAM. The target number of coded bits per QAM symbol is $m = 4.25$ bpcu. The assumed FEC code rate is $5/6$, so the target information rate is 3.542 bpcu. When we employ 64-QAM, 32-QAM, or hybrid 32/16-QAM, the FEC throughput increases by $6/4.25 - 1 = 41.1\%$, $5/4.25 - 1 = 17.6\%$, or $44.5/4.25 - 1 = 5.9\%$, resp., compared with uniform signaling. Fig. 3 shows the PMFs generated by HiDM for different F values. The PMFs for the linear AWGN channel have a relatively high peak at small amplitudes, wheras the the PMFs for nonlinear channels are more uniform.

Fig. 4 illustrates the tradeoff between linear and nonlinear performance. The horizontal axis shows the normalized generalized mutual information (NGMI) [30], [31] or asymmetric information (ASI) [32], [33] with matched bit-metric decoding [4] over the AWGN channel with an SNR of 12 dB. The vertical axis shows the excess kurtosis Φ, which approximately quantifies the nonlinear interference. PS-64-QAM and PS-32-QAM show comparable linear performance. They are almost

[3]If the least significant bit is shaped, the energy gap will be reduced to less than 0.3 dB.

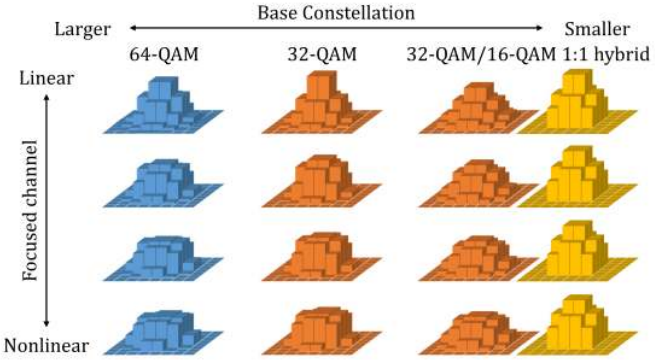

Fig. 3. Two-dimensional PMFs obtained by HiDM for (from top to bottom) $F = 2, 4, 6, 8$.

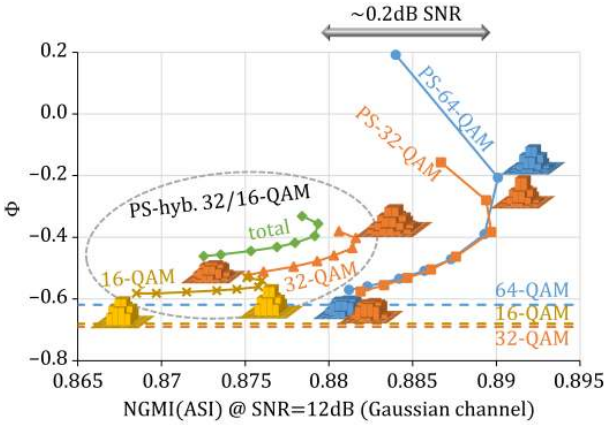

Fig. 4. Tradeoff between linear and nonlinear performance. The top marker on each curve corresponds to $F = 1$ and the bottom one to $F = 8$.

0.2 dB better in terms of required SNR than hybrid PS-32/16-QAM. When the DM word lists in the LUTs are adapted to nonlinear fiber-optic channels by increasing F, the linear performance degrades, but the nonlinear performance improves. For PS-64-QAM and PS-32-QAM with high F, Φ becomes comparable with uniform 64-QAM, at the expense of a linear performance reduction of about 0.2 dB.

The best linear performance was observed at $F = 2$, with minimum E_c, for PS-64-QAM and PS-hybrid-32/16-QAM, and $F = 3$ for PS-32-QAM. To investigate why the best F is 3 for PS-32-QAM, we computed the rate loss as shown in Fig. 5. The rate loss decreases as F increases, i.e., as Φ decreases, for PS-64-QAM and PS-32-QAM. The balance between E_c and rate loss causes the peculiar performance of PS-32-QAM in Fig. 4.

V. SUMMARY

Some aspects of low-complexity implementations of DM PS in fiber-optic communications were studied, in terms of throughput, bus width, and circuit area. HiDM has around 100 times larger throughput than a DM with RLC. A large-scale HiDM example was given in detail, realizing a DM word length of 160 256-QAM symbols using a 7-layer LUT hierarchy. The resulting energy gap from the ideal MB distribution is less than 0.4 dB, while keeping four bits per QAM symbol uniformly distributed (nonshaped). A simple method

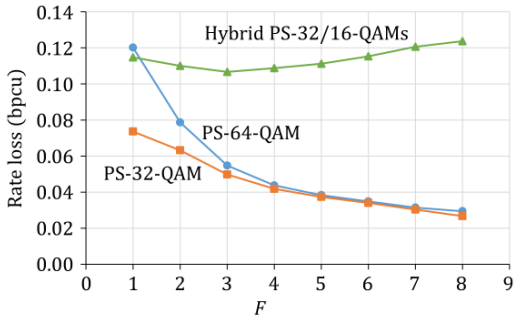

Fig. 5. Rate loss as a function of the sorting parameter F.

to flexibly adapt HiDM to nonlinear channels with granular base constellations was also described.

As shown in [34], this HiDM is useful also for a rudimentary form of joint source–channel coding. This realizes simultaneous data compression and probabilistic shaping, which can further reduce the required SNR or system power consumption in future optical networks. Thanks to the high throughput by massively parallel interfaces, both the encoding and decoding of HiDM, including joint source–channel coding, was implemented in a single field-programmable gate array chip [35].

Acknowledgment

We thank Koji Igarashi of Osaka University for fruitful discussions.

References

[1] G. D. Forney, Jr. and L.-F. Wei, "Multidimensional constellations—Part I: introduction, figure of merit, and generalized cross constellation," *IEEE J. Selected Areas Commun.*, vol. 7, no. 6, pp. 877–892, Aug. 1989.

[2] A. R. Calderbank and L. H. Ozarow, "Nonequiprobable signaling on the Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 726–740, July 1990.

[3] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 913–929, May 1993.

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

[7] J. Cho, S. Chandrasekhar, R. Dar, and P. J. Winzer, "Low-complexity shaping for enhanced nonlinearity tolerance," in *Proc. Eur. Conf. on Opt. Comm. (ECOC)*, Düsseldorf, Germany, Sep. 2016, p. W1C.2.

[8] G. Böcherer, F. Steiner, and P. Schulte, "Fast probabilistic shaping implementation for long-haul fiber-optic communication systems," in *Proc. Eur. Conf. on Opt. Comm. (ECOC)*, Göteborg, Sweden, Sep. 2017, p. Tu.2.D.3.

[9] F. Steiner, P. Schulte, and G. Böcherer. "Approaching waterfilling capacity of parallel channels by higher order modulation and probabilistic amplitude shaping," in *Proc. 52nd Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, USA, Mar. 2018.

[10] T. Yoshida, M. Karlsson, and E. Agrell, "Technologies toward implementation of probabilistic constellation shaping," in *Proc. Eur. Conf. on Opt. Comm. (ECOC)*, Roma, Italy, Sep. 2018, p. Th.1.H.1.

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

[12] T. Yoshida, M. Karlsson, and E. Agrell, "Hierarchical distribution matching for probabilistically shaped coded modulation," *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1579–1589, Mar. 2019.

[13] P. Schulte and F. Steiner, "Divergence-optimal fixed-to-fixed length distribution matching with shell mapping," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 620–623, Apr. 2019.

[14] J. Cho and P. J. Winzer, "Multi-rate prefix-free code distribution matching ," in *Proc. Opt. Fib. Commun. Conf. (OFC)*, San Diego, CA, USA, Mar. 2019, Paper M4B.7.

[15] Y. C. Gültekin, W. J. van Houtum, A. Koppelaar, and F. M. J. Willems, "Enumerative sphere shaping for wireless communications with short packets," *IEEE Trans. Wireless Commun.*, to appear, 2020.

[16] J. Cho, "Prefix-free code distribution matching for probabilistic constellation shaping," *IEEE Trans. Commun.*, to appear, 2020.

[17] T. Yoshida, M. Karlsson, and E. Agrell, "Short-block-length shaping by simple mark ratio controllers for granular and wide-range spectral efficiencies," in *Proc. Eur. Conf. on Opt. Comm. (ECOC)*, Göteborg, Sweden, Sep. 2017, p. Tu.2.D.2.

[18] T. Yoshida, M. Karlsson, and E. Agrell, "Low-complexity variable-length output distribution matching with periodical distribution uniformalization," in *Proc. Opt. Fib. Commun. Conf. (OFC)*, San Diego, CA, USA, Mar. 2018, p. M.4.E.2.

[19] Y. Koganei, K. Sugitani, H. Nakashima, and T. Hoshida, "Optimum bit-level distribution matching with at most $O(N^3)$ implementation complexity," in *Proc. Opt. Fib. Commun. Conf. (OFC)*, San Diego, CA, USA, Mar. 2019, Paper M4B.4.

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

[21] G. Böcherer, F. Steiner, and P. Schulte, "High throughput probabilistic shaping with product distribution matching," [Online]. Available: www.arxiv.org/abs/1702.07510

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

[23] S. Civelli and M. Secondini, "Hierarchical distribution matching: a versatile tool for probabilistic shaping," [Online]. Available: www.arxiv.org/abs/1911.08243

[24] T. Yoshida and N. Suzuki, "Flexible and low-power probabilistic shaping for fiber-optic communications," in *Proc. Signal Processing in Photonic Communications (SPPCom)*, Burlingame, CA, USA, July 2019, Paper SpT3E.2.

[25] Z. Aspar, Z. M. Yusof, and I. Suleiman, "Parallel Huffman decoder with an optimized look up table option on FPGA," in *2000 TENCON Proc. Intelligent Systems and Technologies for the New Millennium (Cat. No.00CH37119)*, vol. 1, pp. 73–76, 2000.

[26] H. J. Huang, C.-H. Fang, and C.-P. Fan, "Very large scale integration design of a low power and cost effective context based adaptive variable length coding decoder for H. 264/AVC portable applications," IET Image Processing, vol. 6, no. 2, pp. 104–114, 2012.

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

[28] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Opt. Express*, vol. 22, no. 13, pp. 16335–16362, June 2014.

[29] E. Sillekens, D. Semrau, G. Liga, N. A. Shevchenko, Z. Li, A. Alvarado, P. Bayvel, R. I. Killey, and D. Lavery, "A simple nonlinearity-tailored probabilistic shaping distribution for square QAM," in *Proc. Opt. Fib. Commun. Conf. (OFC)*, San Diego, CA, USA, Mar. 2018, Paper M3C.4.

[30] J. Cho, L. Schmalen, and P. Winzer, "Normalized generalized mutual information as a forward error correction threshold for probabilistically shaped QAM," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Göteborg, Sweden, Sep. 2017, Paper M.2.D.2.

[31] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications," *IEEE/OSA J. Lightw. Technol.*, vol. 37, no. 6, pp. 1590–1607, Mar. 2019.

[32] T. Yoshida, M. Karlsson, and E. Agrell, "Performance metrics for systems with soft-decision FEC and probabilistic shaping," *IEEE Photon. Technol. Lett.*, vol. 29, no. 23, pp. 2111–2114, Dec. 2017.

[33] T. Yoshida, A. Alvarado, M. Karlsson, and E. Agrell, "Post-FEC BER prediction for bit-interleaved coded modulation with probabilistic shaping," [Online]. Available: arxiv.org/abs/1911.01585

[34] T. Yoshida, M. Karlsson, and E. Agrell, "Joint source-channel coding via compressed distribution matching in fiber-optic communications," in *Proc. Opt. Fib. Commun. Conf. (OFC)*, San Diego, CA, USA, Mar. 2019, Paper M4B.6.

[35] T. Yoshida, M. Binkai, S. Koshikawa, S. Chikamori, K. Matsuda, N. Suzuki, M. Karlsson, and E. Agrell, "FPGA implementation of distribution matching and dematching," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Dublin, Ireland, Sep. 2019, Paper M.2.D.2.