

# Novel Distribution Matcher Design for Short Length Frames Based on Non-Binary Convolutional Codes

Rami Klaimi

*MIMOPT Technology*

149 Avenue du Maine - 75014 Paris - France  
rami.klaimi@mimopt.com

Akram Abouseif

*MIMOPT Technology*

149 Avenue du Maine - 75014 Paris - France  
akram.abouseif@mimopt.com

Ghaya Rekaya-Ben Othman

*Telecom Paris, Institut Polytechnique de Paris*

19 Place Marguerite Perey 91120 Palaiseau, France  
ghaya.rekaya@telecom-paris.fr

Yves Jaouën

*Telecom Paris, Institut Polytechnique de Paris*

19 Place Marguerite Perey 91120 Palaiseau, France  
yves.jaouen@telecom-paris.fr

**Abstract**—We propose a distribution matcher that enables probabilistic constellation shaping while ensuring low-complexity dematching techniques. The proposal is based on non-binary convolutional codes, designed to respect a given optimal symbol distribution. In addition to lowering the dematching complexity, the proposed structure is shown to reduce the latency, to respect the target distribution with a low overhead and to outperform existing solutions with more than 0.3dB. It is also shown that, while being able to respect the target distribution for short frame lengths, the proposed technique helps enhancing the resilience of the optical system in question to the non-linearity effects.

**Index Terms**—Probabilistic shaping, distribution matcher, non-binary convolutional code, optical channel

## I. INTRODUCTION

In order to better approach the channel capacity limits for different communication systems, two main tracks are followed: designing a very good error correcting code, and insuring the optimal distribution at the input of the communication channel. The first track has been widely targeted throughout the years with the introduction of several forward error correction (FEC) codes. However, achieving the optimal channel input symbol distribution has been an open question for a while. In [1], [2], probabilistic constellation shaping (PCS) was developed and coupled with the FEC codes in order to insure good error correction and to optimize the channel input. The input distribution was proven to be near-optimum for optical systems when emulating a Gaussian distribution [3].

The design of an optimal and low-complex PCS scheme remains a challenge for nowadays researches. The shaper, chosen to be placed as an outer code at the transmitter end (before the FEC encoder) in order to avoid severe error propagation as explained in [4], should approach the optimal distribution while maintaining low-complex unshaping procedure at the receiver end. In [5], the constant composition distribution matcher (CCDM) was proposed, and offered a solution to pave the way for other algorithms such as the ones developed in [6]. The proposed PCS schemes are shown to achieve optimal distribution for long frames [5]. Recently, it

was shown that applying the shaping on short frame sizes is of high importance to tolerate the nonlinear effects [7]–[9]. In addition, short-length shaping would also be beneficial to lower the receiver complexity, and consequently limiting the added latency. However, the existing shaping algorithms fail to approach the desired distribution when applied to short frame sizes [7].

In this work, we aim at designing a new DM, targeting the optimization of the input distribution, and insuring the flexibility of the system, while offering the possibility of applying short length frames. The proposed distribution matcher (DM) is based on non-binary convolutional codes as defined in [10]. This choice is motivated by the offered flexibility of these codes, their ease of decoding and their performances when applied to short frames.

The rest of this paper is organized as follows: Section II is dedicated to introduce the adopted non-binary convolutional code structure. In Section III we explain how convolutional codes are used to construct distribution matchers for high-order QAM constellations, before giving an application example to the 16-QAM in Section IV. The proposal is assessed in Section V, where simulation results are shown in comparison with state-of-art distribution matchers and the resilience of the proposal to the non-linearity effects is studied in function of the frame length. And finally, Section VI concludes the paper.

## II. PRINCIPLE OF NON-BINARY CONVOLUTIONAL CODES

In this work, we use non-binary convolutional codes (NB-CC), with a modified structure to the one proposed in [10]. It was first proposed to construct non-binary turbo codes for wireless communication. In Fig. 1, the adopted NB-CC structure is illustrated with two memory elements. When considered with a given Galois field (GF) of the order  $q$  (i.e.  $q$  elements), the input  $In$  and outputs  $(s, p)$  are in the same GF, while the trellis of the code shows  $q^2$  states. The recursion is assured by  $\{r_1, r_2\} \in \text{GF}(q)^2$ , and the parity  $p$  is calculated using  $\{a_1, a_2, a_3, a_4\} \in \text{GF}(q)^4$ . For a frame  $In$  of  $K$  GF( $q$ ) symbols,  $K$  parity symbols are generated. Puncturing

techniques on the information or parity symbols can be applied in order to achieve the desired rate.

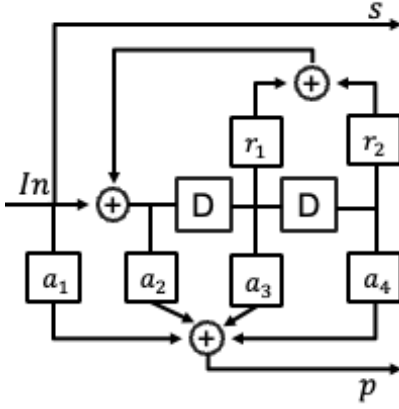


Fig. 1. Structure of 2-memory non-binary convolutional code

### III. DISTRIBUTION MATCHER BASED ON TIME-VARIANT CONVOLUTIONAL CODES

Constructed from two pulse amplitude modulations, one on the *in-phase* ( $I$ ) axis and the other one on the *quadrature* ( $Q$ ) axis, a quadrature amplitude modulation (QAM) of order  $2^m$  enjoys a maximum number of  $\sqrt{2^m}$  different amplitude levels. When uniform input is required, the  $m$  bits in a  $2^m$ -QAM symbol are generated from a uniform distribution, and the different amplitude levels are selected with the same probability. However, when Gaussian input is required, the probabilities of selecting the different amplitude levels are assigned accordingly. In that case, the two sign bits of the  $2^m$ -QAM symbol are the only ones that respect a uniform distribution, whilst the other  $m - 2$  bits have to respect the target distribution.

In this work we propose the use of NB-CCs over  $GF(4)$  (Fig. 1) to emulate the Gaussian distribution by generating the  $m - 2$  non-uniform QAM bits. This structure generates parities over  $GF(4)$  at each time instance. To obtain a  $2^m$ -QAM constellation symbol,  $m/2$  encoding processes are required since each process generates a symbol of two bits. Since the sign bits are uniformly distributed, one of these  $m/2$  encoding processes should respect a uniform distribution, while the other ones will generate non-uniform outputs.

In Fig. 2, a time variant trellis is proposed, where a constellation symbol is generated through  $m/2$  time instances and from  $m/2$  different NB-CCs structures. The first trellis, denoted as "uniform", results from a uniform NB-CC designed respecting the constraints of [10]. This trellis generates uniformly distributed symbols of two bits that are mapped to the sign bits of the  $2^m$ -QAM symbols. On the other hand, the non-uniform trellises are generated from NB-CCs that respect a non-uniform distribution at their parity output. In that case, the parity  $p$  is not generated from the parity polynomial as explained in Fig. 1, but from a look up table designed to respect the target Gaussian distribution. The same non-uniform

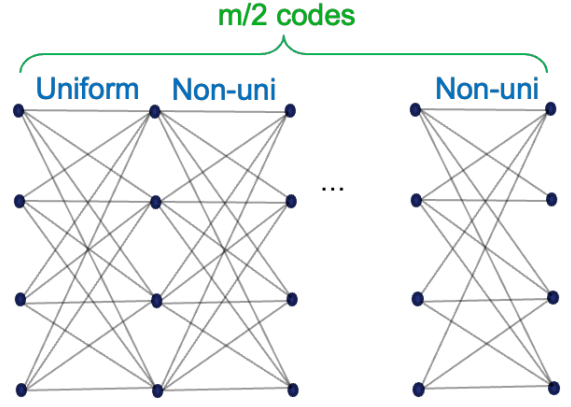


Fig. 2. Time-variant trellis of the CC-based DM

process is repeated  $m/2 - 1$  times, in order to generate the  $m - 2$  non-uniform bits of a  $2^m$ -QAM constellation symbol. In Table I, the required number of encoding processes are shown for different QAM constellation orders.

TABLE I  
ENCODING PROCESSES TO SHAPE  $2^m$ -QAM CONSTELLATIONS

$q$	Uniform trellis	Non-Uniform trellis
16	1	1
64	1	2
256	1	3

### IV. NB-CC DM CONSTRUCTION FOR 16-QAM

As previously stated, to design a NB-CC based DM for a 16-QAM having symbols of  $m = 4$  bits, one uniform NB-CC convolutional code is used alongside with one non-uniform NB-CC. These codes are 2-memory elements structure and designed over  $GF(4)$ , resulting in code trellises of 16 states. In Fig. 3, a small part of the resulting trellises is given for the sake of simplicity, where only four states are considered from the 16 states codes. As shown in this figure, the first code, whose parity symbols are generated respecting Fig. 1 and whose responsible of generating the sign bits, shows uniformly distributed parities on its generated transitions. In this case, each state generates a transition for each parity value from the ensemble  $\{00, 01, 10, 11\}$ . On the other hand, the non-uniform NB-CC, responsible of generating the two amplitude bits, does not respect the parity generation of Fig. 1. In this case, the parities are filled from a look up table to respect the given distribution. This is also shown in Fig. 3, where the parity  $p = 11$  is shown to be generated with the highest probability in this example.

### V. SIMULATION RESULTS

1) *System setup*: In this section, the proposed PCS will be evaluated under different simulation conditions. As shown in Fig. 4, we consider the implementation of transmission system of dual polarized (DP) single mode G-652 type fiber

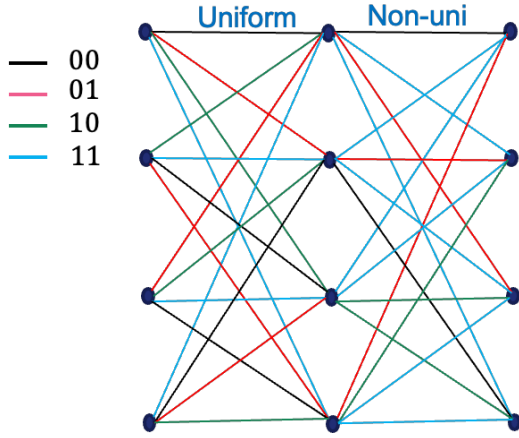


Fig. 3. Encoding trellises to shape a 16-QAM symbol

(SMF) with baud rate of 96 GBaud. Twelve spans construct the considered SMF, where the span length is 100 Km and each span followed by an EDFA amplifier. The used numerical parameters are presented in Fig.4. At the receiver side, chromatic dispersion compensation ( $CD^{-1}$ ) and the MIMO equalization based on training sequence are done in the frequency domain.

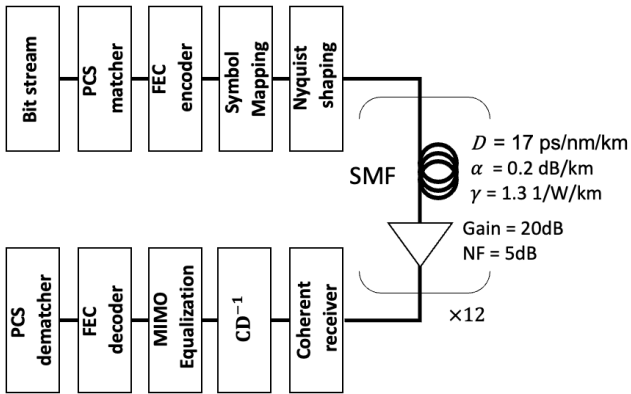


Fig. 4. Transmission chain

2) *Comparison with state-of-art codes*: The goal of our proposal is to design good DM for short-length frames. For that, it is compared with the CCDM algorithm from [5], simulated based on the web package at [11], for different code lengths. A classical (1944, 1458) protograph based low-density parity-check (LDPC) code is applied after both DMs, and the results are the post-FEC error correcting performances. In the first comparison, the input FEC frame is obtained from the output of one DM procedure, resulting in the use of long frame DM of  $K = 1152$  bits, with a PCS rate  $R = 3.2$  bits/2D symbol. In this case, the CCDM shows better performance than the NB-CC, outperforming it with 0.5 dB as shown in Fig. 5. When designing both DMs with shorter frames, while maintaining the same rate, the input FEC frame is composed from several shaping procedure and an equivalent performance is observed for  $K = 192$  bits. However, by more shortening the frame length, the CCDM fails to maintain its

good performance, and the NB-CC starts to outperform it with more than 0.3 dB. From this figure, we can clearly conclude that the NB-CC preserves a similar performance for different frame lengths, which is not the case for the CCDM.

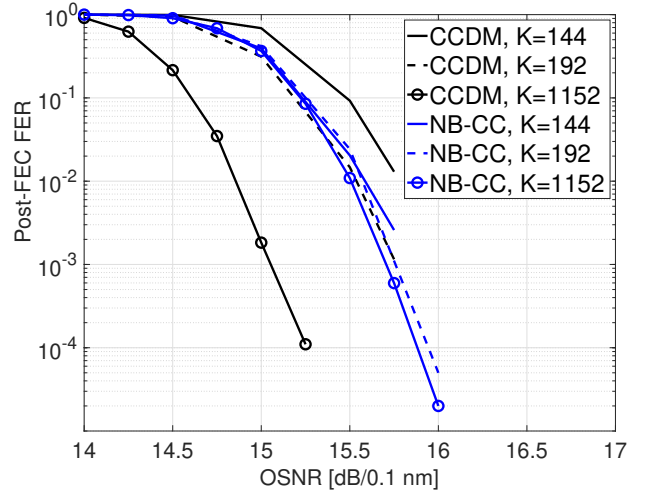


Fig. 5. Frame error rate (FER) comparison between the proposed DM (with  $R = 3.2$  bits/2D symbol) and [5], with the consideration of FEC code

3) *Resilience to the non-linear effect*: We will study in this part the tolerance of the proposal to the non-linear (NL) effects. Here we consider the previously used long ( $K = 1152$ ) and short frames ( $K = 144$ ) with the same rate ( $R = 3.2$  bits/2D symbol). In Fig. 6, the system tolerance to NL effects is assessed without the consideration of FEC code. It is shown again that the NB-CC is not highly affected by its frame size in the linear regime. The limited observed differences are mitigated in the presence of FEC code as shown in Fig. 5. On the other hand, the proposal is showing better NL tolerance with short length frames, which is in accordance with the propositions of [7]–[9], where it was verified that short length shaping can be advantageous to enhance the system resilience to the NL effects.

4) *Flexibility to achieve different rates*: Since convolutional codes are used in the proposed PCS, different code rates can be achieved by simple puncturing technique. In the case of a 16-QAM, the mother code rate of the proposed NB-CC is  $R = 2$  bits/2D symbol, since for each input 16-QAM symbol, 4 parity bits are generated from the two considered encoding processes. To achieve other rates, puncturing is performed on the information bits to conserve the generated parities. In Table II, three examples are given for puncturing masks to be applied on the information bits in order to achieve the required rate.

TABLE II  
INFORMATION BITS PUNCTURING MASKS

$R$ (bits/2D symbol)	3.2	3.55	3.7
Puncturing Period	4	8	12
Puncturing mask	1000	10...0	10...0

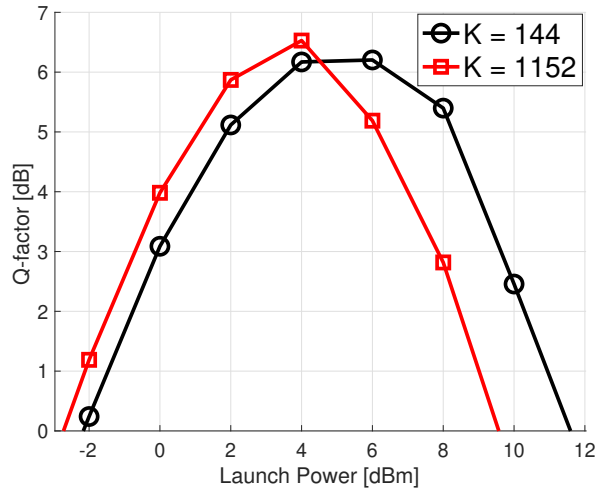


Fig. 6. Comparison of the effect of non-linearity between the cases of short length and long length PCS

## VI. CONCLUSIONS

In this work, a new DM is proposed, based on the use of time-varying NB-CCs. The implemented NB-CCs are easily decoded with well established state-of-art algorithms. The NB-CC DM ensures the flexibility to achieve any required rate of the overall system. When applied with short-length PCS, the proposal is shown to maintain its good performance and to outperform existing DMs with more than 0.3 dB. In addition, the proposal is more tolerant to NL effects and can guarantee lower latency and lower decoding complexity thanks to the use of short-length decoding.

## REFERENCES

- [1] G. Böcherer, “Probabilistic signal shaping for bit-metric decoding,” in *IEEE Int. Symp. on Inform. Theory*, IEEE, Honolulu, HI, USA, Jun. 2014, pp. 431–435.
- [2] G. Böcherer, F. Steiner, and P. Schulte, “Bandwidth efficient and rate-matched low-density parity-check coded modulation,” *IEEE Trans. on Commun.*, vol. 63, no. 12, pp. 4651–4665, 2015.
- [3] J. Cho and P. J. Winzer, “Probabilistic constellation shaping for optical fiber communications,” *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1590–1607, 2019.
- [4] F. Steiner, “Coding for higher-order modulation and probabilistic shaping,” Ph.D. dissertation, Technische Universität München, 2020.
- [5] P. Schulte and G. Böcherer, “Constant composition distribution matching,” *IEEE Trans. on Inf. Theory*, vol. 62, no. 1, pp. 430–434, 2015.
- [6] P. Schulte, “Algorithms for distribution matching,” PhD dissertation, Technical University of Munich, Germany, 2019.
- [7] J. Cho and R. Tkach, “On the kurtosis of modulation formats for characterizing the nonlinear fiber propagation,” *J. of Lightwave Technology*, 2022.
- [8] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, “Mapping strategies for short-length probabilistic shaping,” in *45th European Conf. on Optical Commun. (ECOC 2019)*, IET, Dublin, Ireland, Jun. 2019.
- [9] T. Fehenberger, H. Griesser, and J.-P. Elbers, “Fiber nonlinearity mitigation by short-length probabilistic constellation shaping for pilot-aided signaling,” in *21th ITG-Symposium Photonic Networks*, VDE, Dec. 2020.
- [10] R. Klaimi, C. Abdel Nour, C. Douillard, and J. Farah, “Design of Low-Complexity Convolutional Codes over GF(q),” in *IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, UAE, Dec. 2018.
- [11] P. Schulte, F. Steiner, and G. Böcherer, *shapcomm WebDM: Online constant composition distribution matcher*, <http://dm.shapcomm.de>, Jul. 2017.