Temporal probabilistic shaping for mitigation of nonlinearities in optical fiber systems

Yankov, Metodi Plamenov; Larsen, Knud J.; Forchhammer, Søren

Published in:
Journal of Lightwave Technology

Link to article, DOI:
10.1109/JLT.2017.2671452

Publication date:
2017

Document Version
Peer reviewed version

Temporal probabilistic shaping for mitigation of nonlinearities in optical fiber systems

Metodi P. Yankov, *Member, IEEE*, Knud J. Larsen and Søren Forchhammer, *Member, IEEE*

***Abstract*—In this paper, finite state machine sources (FSMSs) are used to shape quadrature amplitude modulation (QAM) for nonlinear transmission in optical fiber communication systems. The previous optimization algorithm for FSMSs is extended to cover an average power constraint, thus enabling temporal optimization with multi-amplitude constellations output, such as QAM. The optimized source results in increased received SNR and thereby increased achievable information rates (AIR)s under memoryless assumption. The AIR is increased even further when taking the channel and transmitter memory into account via trellis processing at the receiver. Significant gains are reported in the highly nonlinear region of transmission for an FSMS of up to second order and 16QAM and particularly for unrepeated transmission. At the optimal launch power of WDM transmission, the FSMS order needs to be increased further in order to notably outperform previous probabilistic shaping schemes.**

***Index Terms*—probabilistic shaping, finite state machines, nonlinear distortion.**

I. Introduction

Constellation shaping is currently among the hot topics for investigation in the fiber optic communications community. Shaping is generally achieved via digital processing, which allows for improving the performance of legacy fiber networks in terms of e.g. the achievable rates and maximum link distance without making changes to the links themselves. In particular, probabilistic shaping of quadrature amplitude modulation (QAM) constellations is gathering popularity due to its simple integration with standard digital signal processing techniques. As recently demonstrated [1]–[4], probabilistic shaping can be integrated into coded modulation schemes with powerful forward error correcting codes, thus allowing for increased link distance while also approaching the high data rate demand that the fiber systems face. Most such systems operate based on a memoryless channel assumption due to the reduced complexity of the receiver. However, due to the interactions of chromatic dispersion (CD) and nonlinear effects (self-phase and cross-phase modulation, SPM and XPM, respectively), this assumption is often too strong. The achievable rates of such receivers therefore diminish at high optical signal to noise ratio (SNR) (high launch power, respectively) [5].

Memoryless probabilistic shaping with memoryless receiver processing provides increased data rates, however, it generally does not allow for increased launch power and the gains are therefore limited [1].

Temporal shaping for the fiber channel was analyzed in [6], where it was suggested that superior gains to those of linear channels can be achieved. Input constellations constrained to a multi-dimensional (MD) ball are studied in [6], and are shown to significantly outperform uniformly distributed MD cube, i.e. uniformly, independent, identically distributed (i.i.d.) QAM symbols with alphabet cardinality tending to infinity. Such ball-shaped constellations are difficult to realize in practice due to 1) large (infinite) cardinality in each dimension; 2) dependence of dimensions on both past and future symbols. In practice, the cardinality is limited (e.g. 16 in 16QAM). The dependence between symbols in time can be realized by e.g. finite state machine sources (FSMSs) (equivalent to Markov sources in the context of this paper), which were treated in [7] for linear channels with memory. A procedure was derived for optimization of the source in terms of its state transition probabilities and the constellation constrained capacity was found with BPSK input. This method was modified for application to the nonlinear optical fiber in [8], where it was shown that shaping gains are possible even with QPSK input constellations.

In this paper, FSMSs are studied as input to the fiber channel. The algorithm from [7] is extended to cover nonlinear channels and FSMSs with multi-amplitude constellation output, allowing for high spectral efficiency. In Section IV, results are shown in terms of mutual information (MI) between channel input and output, where the channel memory was (partly) taken into account at the receiver.

A. Notation

The following notation will be used. Capital letters, e.g. X denote random variables, which come from an alphabet, denoted with caligraphic letters, e.g. $\mathcal{X}$, of size $|\mathcal{X}|$. Lower case letters denote realization of the random variable, e.g. x_k denotes the value of X at time k, and x^j denotes the $j-$th symbol from $\mathcal{X}$. The sequence in time from t to k is denoted x_t^k. The probability mass function (PMF) of X is defined as p_X, and the probability that X takes on value x is $p_X(X = x) = p(x)$ for short notation.

The main performance metric used in this paper is the mutual information (MI) between the channel input and output $\mathcal{I}(X_1^K; Y_1^K)$, which represents the achievable information rate (AIR) on the fiber with the corresponding receiver processing. Bit error rates (BER) are also studied.

The authors are with the Department of Photonics Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark: e-mail: meya@fotonik.dtu.dk

Part of this work was presented at the European Conference on Optical Communications, ECOC (2016).

II. Finite State Machine Sources and Channels

A FSMS is defined by its states, state transitions and output, associated with each transition. In this paper (with the exception of Section III-D), the state S is defined as the previous N symbols, emitted by the source x_{k-N}^{k-1} (also denoted by the column vector $[x_{k-N}, x_{k-N+1}, \dots x_{k-1}]$), where N is the *order* of the source. The symbols X, which are output of the FSMS and input to the channel in this paper are instances of a QAM constellation. A branch (or transition) of the source is defined as $b \in \mathcal{B} \triangleq \{i, j\}$, denoting exiting state i with the symbol x^j. We then have $|\mathcal{S}| = |\mathcal{X}|^N$ and $|\mathcal{B}| = |\mathcal{X}|^{N+1}$. Optimization of the transition PMF $p_{X|S}$ allows for *temporal probabilistic shaping*, generalizing memoryless probabilistic shaping, for which $p_{X|S} = p_X$ and p_X is optimized. When the FSMS output sequence x_1^K completely defines the state sequence s_1^K, the MI $\mathcal{I}(X_1^K; Y_1^K) = \mathcal{I}(B_1^K; Y_1^K)$[1].

In this paper, the finite state machine channel (FSMC) model from Fig. 1 is adopted at the receiver as an auxiliary channel (see Section III for more detail). While it is not an entirely accurate representation of the fiber, it allows for some of the channel memory to be captured. The main assumption of the model is that the memory is finite, i.e. the current channel output sample depends on the previous M input symbols, while disregarding the future. This extends the memoryless assumption, for which $p(y_1^K|x_1^K) = \prod_{k=1}^K p(y_k|x_k)$. Another assumption in this model is stationarity, i.e. the PMFs $p_{X|S}$ and $p_{Y|X,S}$ are time-independent.

Using this model, the MI $\mathcal{I}(X_1^K; Y_1^K)$ can be computed via the BCJR algorithm [9], which was also suggested for the fiber channel with BPSK input [10]. On the other hand, the representation $\mathcal{I}(B_1^K; Y_1^K)$ allows for optimization of the state transition probabilities as $p(b) = \arg\max_{p(b)} \mathcal{I}(B_1^K; Y_1^K)$(under some constraints) [7]. In either case [7], [9], a trellis is built at the receiver side which allows for processing of sequences of symbols, thus capturing up to M symbols of the combined channel and transmitter memory via the likelihoods $p(y_k|x_{k-M}^k)$. In this paper (with the exception of Section III-D), the states in the receiver's trellis are represented by the previous M input symbols (as in Fig. 1). Note that while optimization of the FSMS transition probabilities requires BCJR processing with $M = N$, the MI can be computed with the memoryless assumption also when the input is an FSMS, and vice-versa. In general, $M = N$ is not a requirement for the system and different performance/complexity tradeoffs can be achieved with different choices of $\{M, N\}$.

III. FSMS Optimization for Multi-Amplitude and Multi-Phase Constellation

The general Blahut-Arimoto (BA) algorithm for optimizing an FSMS on a linear FSMC is given in [7, Lemma 44, Algorithm 45]. Due to it's lengthy derivation and notation, it is omitted here, however, some of the definitions and results therein will be reused with proper referencing in this paper. The limitations of the algorithm for direct application to the optical fiber are two-fold: 1) the likelihoods $p(y_k|x_{k-N}^k)$ are not known in closed form; 2) it does not support an average power constraint in its general form. An average power constraint is crucial for the fiber, since the channel, and thereby the likelihoods are power dependent. The algorithm is in its nature iterative, i.e., the PMF of the transitions is iteratively optimized until convergence. While a given average power constraint is ensured with constant amplitude constellations (such as QPSK in [7], [8]), FSMSs with multi-amplitude constellations output exhibit an average power, which is dependent on the PMF of the state transitions. We observed that without a power constraint, especially in the highly-nonlinear region of transmission, the algorithm from [7] results in convergence to a local optimum of higher MI than the initial value, but much lower than what is otherwise possible.

[1]since $\mathcal{H}(S_1^K|X_1^K) = 0 \Rightarrow \mathcal{I}(B_1^K; Y_1^K) = \mathcal{I}(S_1^K, X_1^K; Y_1^K) = \mathcal{I}(X_1^K; Y_1^K) + \mathcal{I}(S_1^K; Y_1^K|X_1^K) = \mathcal{I}(X_1^K; Y_1^K)$

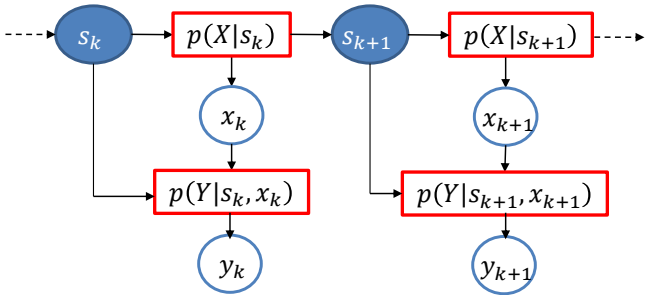


Fig. 1. FSMC with FSMS input model assumed in this work. For illustration purposes, the state of the channel is identical to the state of the source, which is generally not a requirement. The current symbol depends on the state and the PMF $p_{X|S}$, and the channel output depends on the vector $[s_k, x_k]$.

The average output power of an FSMS (input power to the channel) can be found as

$$P_{av} = \sum_{b \in \{i,j\}} p(b) \cdot |\alpha x^j|^2, \tag{1}$$

where α is a scaling factor, allowing for basic geometric shaping in terms of linear scaling of the constellation $\mathcal{X}$. In order to ensure a constant P_{av}, an additional Lagrange multiplier λ_P is introduced in the optimization process, making the noisy adjacency matrix from [7, Definition 43] $A_{i,j} = e^{T_{i,j} + \lambda_P \alpha |x^j|^2}$.

In order to circumvent the lack of a closed form expression for the channel likelihoods, the method of *mismatched decoding and auxiliary channel* is adopted (see e.g. [1], [2], [9]). The auxiliary channel in this case is a 2D Gaussian, i.e., the above mentioned likelihoods are modeled as Gaussian functions at y_k

$$p(y_k|B = b_{i,j}) = \mathcal{N}(\Sigma_{i,j}, \mu_{i,j}; [\mathrm{Re}\,[y_k], \mathrm{Im}\,[y_k]]^T), \tag{2}$$

where $\Sigma_{i,j}$ is a 2x2 covariance matrix and $\mu_{i,j}$ is the 2D mean. Such auxiliary function allows for capturing the non-circularity of the noise, which is particularly important when operating in the nonlinear region of transmission [11]. We note that it is possible to increase the dimensionality of the likelihoods to e.g. $p(y_{k-D}^{k+D}|B = b_{i,j})$, however, estimating of the parameters $\Sigma_{i,j}$ and $\mu_{i,j}$ in this case requires an exponentially increased number of samples [12, Ch. 1.4]. Furthermore, we have verified that increasing the dimensionality to $2 \cdot D + 1 = 3$ does not result in significant performance improvement. For the rest of the paper, $D = 0$.

Since the likelihoods are only approximations to the true channel, the result is a lower bound on the MI [9]. The overall precision of the approximation governs the tightness of the lower bound. The finite memory FSMC model adopted at the receiver in this work represents a further approximation of the true likelihood, and thus has an additional effect on the tightness of the MI bound.

The updated BA algorithm is given in Algorithm 1. We have slightly abused the notation to add the PMF and the scaling α to the arguments of the MI in Steps 21 and 22.

Algorithm 1 Algorithm for optimizing an FSMS on a nonlinear optical fiber channel

Initialize: P_{thr} - threshold for achieved average power
Initialize: $\lambda_{min}, \lambda_{step}$ - conditions for numerical estimation of λ.
1: **for** $\alpha \in [1/\max(|X|); 1/\min(|X|)]$ [13] **do**
Initialize: p_B, such that the power constraint $\sum_{b\in i,j} p(b)|\alpha x^j|^2 = P_{in}$ is satisfied
2: **while** Not converged $p(b)$ **do**
3: Generate the sequences s_1^K, x_1^K and b_1^K according to $s_k = x_{k-M}^{l-1}$, $x_k \sim p_{X|S}$, $b_k = [s_k, x_k]$
4: Generate y_1^K by solving the NLSE via the split-step Fourier method (SSFM)
5: **for** each $b \in \{i, j\}$ **do**
6: Find the set of indices $k \in \mathcal{K}$ for which $s_k = s^i$ and $x_k = x^j$.
7: Estimate $\Sigma_{i,j} = \mathrm{Cov}\left[\mathrm{Re}\left[Y_\mathcal{K}\right], \mathrm{Im}\left[Y_\mathcal{K}\right]\right]^T$
8: Estimate $\mu_{i,j} = \mathbb{E}_k\left[\mathrm{Re}\left[Y_\mathcal{K}\right], \mathrm{Im}\left[Y_\mathcal{K}\right]\right]^T$
9: Set $p(y|B = b_{i,j}) = \mathcal{N}(\Sigma_{i,j}, \mu_{i,j}; [\mathrm{Re}[y], \mathrm{Im}[y]]^T)$
10: **end for**
11: Calculate $T_{i,j}$ according to [7, Lemma 70].
12: Set $\lambda = \lambda_{min}$
13: **do**
14: Optimize p_B and p_S according to [7, Lemma 44], with $A_{i,j} = e^{T_{i,j}+\lambda_P \alpha |x^j|^2}$.
15: Calculate $P_\lambda = \sum_{b\in i,j} p(b)|\alpha x^j|^2$
16: $\lambda = \lambda + \lambda_{step}$
17: **while** $|P_\lambda - P_{in}| > P_{thr}$
18: **end while**
19: Calculate $\mathcal{I}_\alpha(X_1^K; Y_1^K) = \mathcal{I}_\alpha(B_1^K; Y_1^K)$ [7, Eq. (42)]
20: **end for**
21: $\hat{\alpha} = \arg\max_\alpha \mathcal{I}_\alpha(X_1^K; Y_1^K)$
22: $p_B = \arg \mathcal{I}_{\hat{\alpha}}(X_1^K; Y_1^K, p_B)$

The newly introduced λ is found numerically by a search with a step of size λ_{step}. As long as λ_{step} is kept small, the average power constraint can be achieved with reasonable accuracy P_{thr}. In our simulations, $P_{thr} = 0.005$ dBm and $\lambda_{step} = 0.001$, which resulted in observed convergence in all investigated cases.

A block diagram of the optimization process is given in Fig. 2. As mentioned, MI estimation, while requiring the same BCJR processing, is generally independent of the FSMS optimization and can be performed with different parameters (e.g. different M, D, K and different auxiliary channel $p(y_k|x_k, s_k)$). Algorithm 1 can be considered a combination

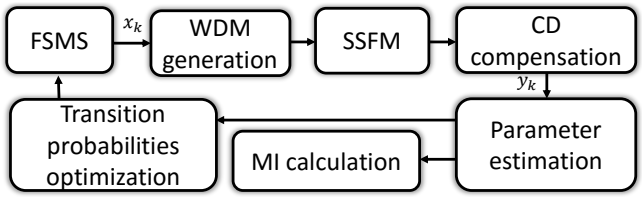


Fig. 2. A block diagram of Algorithm 1.

of the algorithms from [13] for AWGN channel, which was later modified to fit the optical fiber channel under memoryless assumption, for which the channel likelihoods are input-dependent [1], and then further modified to include the generalizations of FSMC with FSMS input [7].

A. Demodulation with a BCJR receiver

Practical receivers can be built with similar techniques as used in Algorithm 1. The goal of the demodulator is to estimate log-likelihood ratios (LLR)s of the bits, which in turn requires the symbol posterior probabilities $p(x_k|y_1^K)$. In Algorithm 1, the posterior probabilities $p(b_k|y_1^K)$ are implicitly calculated via forward *and backward* recursions during the calculation of the $T_{i,j}$ values. The posteriors can then be obtained by marginalization as

$$p(X_k = x^j|y_1^K) = \sum_i p(S_k = s^i, X_k = x^j|y_1^K) = \sum_i p(B_k = \{i, j\}|y_1^K). \quad (3)$$

The LLRs are then computed by standard techniques [14]. The MI in this case is estimated as $\mathcal{I}(X_1^K; Y_1^K) = \mathcal{H}(X_1^K) - \mathcal{H}(X_1^K|Y_1^K)$ and provides an AIR.

The MI can also be estimated as $\mathcal{I}(X_1^K; Y_1^K) = \mathcal{H}(Y_1^K) - \mathcal{H}(Y_1^K|X_1^K)$ as in e.g. [9]. In that case, the same trellis is used, but $p(y_1^K)$ is calculated instead of $p(b_1^K|y_1^K)$. This calculation requires only a forward recursion and thus less complexity. The MI estimates of both approaches are equivalent, however, the latter approach does not allow for demodulation as in Eq. (3).

B. Complexity issue

In order to accurately estimate the likelihoods $p(Y|B = \{i, j\})$, the set of indices $k \in \mathcal{K}_{i,j}$ for which $[s_k, x_k] = [s^i, x^j]$ needs to be large. However, the maximum length K of the input/output sequence is limited by the maximum memory of the simulator. As M, N and thereby $|\mathcal{B}|$ grow and K is fixed, $|\mathcal{K}_{i,j}|$ is reduced and the likelihoods become more inaccurate. Furthermore, the standard method for calculating the MI $\mathcal{I}(X_1^K; Y_1^K)$ is to generate long sequences x_1^K and y_1^K, and use the fact, that [9]

$$\mathcal{H}(X_1^K|Y_1^K) = \lim_{K\to\infty} \frac{1}{K} \log_2 p(x_1^K|y_1^K), \quad (4)$$

$$\mathcal{H}(Y_1^K) = \lim_{K\to\infty} \frac{1}{K} \log_2 p(y_1^K). \quad (5)$$

The values $T_{i,j}$ in Step 11 in Algorithm 1 are estimated in a similar manner from the sub-sequences of symbols $\mathcal{K}_{i,j}$. As

the size $|\mathcal{K}_{i,j}|$ is reduced, the approximation for $T_{i,j}$ in [7, Lemma 70] becomes poorer and results in inaccurate MI estimation and optimization. This issue is even more pronounced when the input PMF is non-uniform, since some sequences appear seldom, and their $T_{i,j}$ values and contributions to the MI are improperly calculated. A couple of suggestions for circumventing this problem are given below.

C. Maxwell-Boltzmann distribution

While it was shown in [1] that with memoryless receiver processing the modified BA algorithm provides only marginal improvement over Maxwell-Boltzmann (MB) distributions, such analogy is not possible here, since the MB distribution is inherently memoryless. To that end, the multi-dimensional MB (MDMB) PMF of the branches in an FSMS is defined from the probabilities

$$p(B=\{i,j\},\lambda_{MB}) \propto \cdot \exp(-\lambda_{MB}\left[s^i,x^j\right]^H\left[s^i,x^j\right]), \quad (6)$$

where $[\cdot]^H$ denotes a conjugate transpose of the column vector. An average power constraint P_{av} is ensured with MDMB by scaling the constellation $\mathcal{X}$ to $\hat{\mathcal{X}} = \mathcal{X}/\sqrt{\sum_{b\in\{i,j\}} p(b,\lambda_{MB})|x^j|^2}$. The MDMB can be optimized by brute-force searching through a suitable range of the parameter λ_{MB}. The dimensionality of the optimization problem is thus reduced to 1, however, the dimensionality of the MI estimation problem is still governed by the cardinality $|\mathcal{B}| = |\mathcal{X}|^{M+1}$ of the branches in the receiver.

Observe that this type of constraint contains the ball-shaped input from [15], [16] as a special case, since for large $|\lambda_{MB}|$, most of the mass of the constellation is concentrated on a single amplitude in the multi-dimensional space (the points with largest amplitude for $\lambda < 0$ and the points with lowest amplitude for $\lambda > 0$) . When $|\mathcal{X}| \to \infty$ and $|\lambda_{MB}|$ is large, the constellation is a multi-dimensional sphere.

D. Merging of states

As seen from Eq. (6), many branches exhibit equivalent probabilities by definition. It would therefore be interesting to merge such branches, thereby reducing the dimensionality of the optimization problem and trellis processing at the receiver, which in turn allows for longer memory to be captured. Care must be taken during such merging in order to ensure that the state sequence is uniquely defined and can be traced at the receiver side with known initial state.

For example, merging based solely on the value $\left[s^i,x^j\right]^H\left[s^i,x^j\right]$ constitutes a violation of this rule since a certain state sequence s_1^K can be generated by different symbol sequences, making reliable communication impossible with trellis processing at the receiver[2].

Instead, we propose to merge states with equal value of the multi-dimensional vector $\left[|s^i(1)|,|s^i(2)|,\ldots|s^i(N)|\right]$, where $s^i(n)$ is the $n-$th element of the vector s^i. This results in parallel branches (transitions) between states, corresponding to symbols from $\mathcal{X}$ with the same amplitude. Such cases are treated in [7, Section IV-C]. The dimensionality of the MI estimation and optimization problems is thus reduced from $|\mathcal{X}|^{N+1}$ to $|\mathcal{X}| \cdot |\mathcal{A}_{\mathcal{X}}|^N$, where $\mathcal{A}_{\mathcal{X}}$ is the set of unique amplitudes of the elements in $\mathcal{X}$ (e.g. for 16QAM $|\mathcal{A}_{\mathcal{X}}| = 3$).

[2]since $\mathcal{H}(S_1^K|X_1^K) \neq 0 \Rightarrow \mathcal{I}(B_1^K;Y_1^K) = \mathcal{I}(S_1^K,X_1^K;Y_1^K) = \mathcal{I}(X_1^K;Y_1^K) + \mathcal{I}(S_1^K;Y_1^K|X_1^K) \neq \mathcal{I}(X_1^K;Y_1^K)$

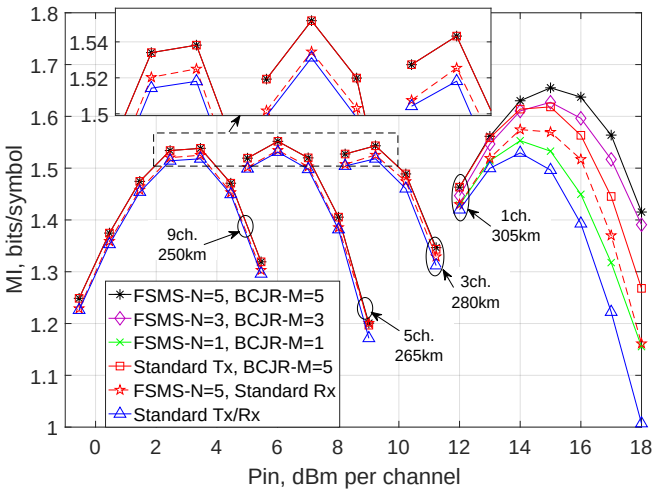


Fig. 3. Unrepeated transmission of FSMS with QPSK. The optimized FSMS enables increased launch power and gains of around 0.15 bits/symbol for single channel. The gains are decreased to ≈ 0.02 bits/symbol in a WDM setup. which is mostly attributed to the BCJR processing at the receiver. Memoryless processing gains are < 0.01 bits/symbol.

IV. Results

The MI in this work is measured in bits/symbol as in Step 19 of Algorithm 1. As mentioned before, it represents an AIR with the particular receiver processing. A single-polarization link is considered. A "Standard Rx" processing corresponds to a BCJR without memory, i.e. $M = 0$, or a simple 2D Gaussian receiver [1], [11]. A "Standard Tx" corresponds to uniform i.i.d. input symbols. At the receiver side, a matched filter is applied and chromatic dispersion compensation is performed in the frequency domain.

A. Unrepeated transmission

We study the gains in maximum AIR for an unrepeated link at the optimal launch power. To that end, the split-step Fourier method (SSFM) is used for a link of a certain length, and an Erbium doped fiber amplifier (EDFA) with a noise figure of 5 dB is included at the end of the link, which sets the received power to 0 dBm. The rest of the fiber and transmitter parameters are given in Table I. In Fig. 3, the MI is given for QPSK input. We see that for single channel, temporal shaping allows for increased optimal launch power and results in gradually increasing AIR improvement with the FSMS order to ≈ 0.15 bits/symbol at $N = 5$. We note that around half of the gain is also obtained with memoryless processing. This is a result of the improved effective SNR at the receiver, in turn achieved by suppressing unwanted sequences in the FSMS by assigning them a lower probability of occurrence. This gain is even greater than the gain of a BCJR processing for uniform i.i.d. input symbols, which means that through shaping, complexity is effectively transferred from the receiver

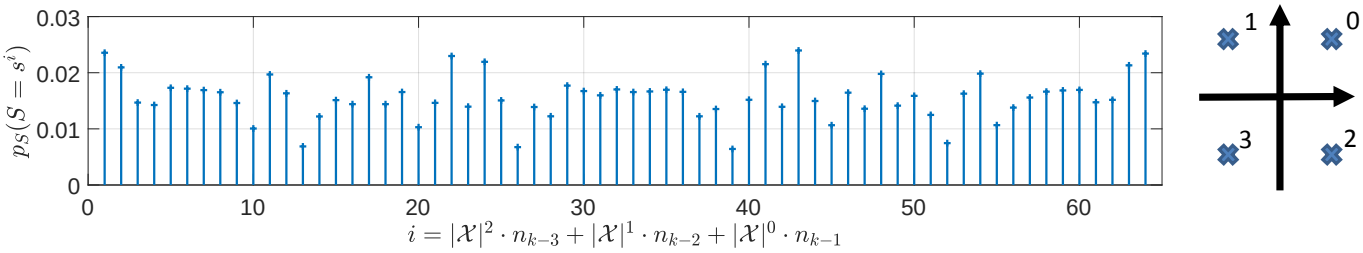


Fig. 4. Optimized PMF of the states, $N = 3$, $|\mathcal{X}| = 4$. The state numbering is given as label, where n_k is the index from the QPSK alphabet, which appears at time k (indexing of the QPSK symbols is given on the right).

TABLE I
SYSTEM AND FIBER PARAMETERS

Fiber loss	$\alpha = 0.2$ dB/km
Non-linear coefficient	$\gamma = 1.3$ (W·km)$^{-1}$
Dispersion	$D = 17$ ps/(nm·km)
Central wavelenght	$\lambda_0 = 1.55$ μm
SSFM step	1 km
Symbol rate	28 GBd
Channel spacing	30 GHz
No. of WDM channels	1, 3, 5, 9
Pulse shape	Square root raised cosine
Roll-off factor	0.01

to the transmitter. In WDM scenarios, the gains are decreased and are mainly attributed to the BCJR receiver processing. Only marginal gains are achieved with an optimized FSMS and memoryless processing at the receiver (discussion about this result is provided below).

In Fig. 4, the optimized state PMF p_S is given at the optimal launch power of $P_{in} = 15$ dBm for single channel transmission. The state indexing is given in the label of the axes. We see that some symmetry and structure exist in the optimized state probabilities. However, we note that this solution is not unique due to the dependence of the channel on the input, which does not guarantee convergence to a unique optimum of Algorithm 1. Due to the symmetry of the constellation, a different numbering of the QPSK symbols (e.g. by taking the conjugate constellation and keep the order in Fig. 4) results in exactly the same distribution of the sequence x_1^k and should therefore result in the same AIR. We have not yet discovered a way to exploit this fact in the optimization process.

In Fig. 5, the MI is given for a 16QAM FSMS for different number of channels. The fiber length in each case is chosen such that the MI is between 2.5 and 3 bits/symbol, which is where the maximum shaping gain is expected. The maximum length of the sequence we were able to simulate is $5 \cdot 10^5$ symbols, which limits the length of the memory for BCJR processing and the FSMS order to $M = N = 2$ (see Section III-B). We see a gain of ≈ 0.2 bits/symbol with an FSMS of order $N = 2$ over a single channel. In the linear region of transmission, the gains are achieved by memoryless shaping with the algorithm from [1] and by Maxwell-Boltzmann optimization. This is expected, as the channel can be considered memoryless AWGN in these cases. When increasing the launch power, the memoryless shaping gain disappears. The MDMB with $N = M = 2$ achieves ≈ 0.15 bits/symbol at the optimal launch power, which is comparable to the gain of a standard transmitter with BCJR processing and $M = 2$. We also study the reduced dimensionality FSMS and BCJR processing from Section III-D. We see that even though the FSMS order and the BCJR memory are increased to $N = M = 5$, the shaping gains in the nonlinear region almost completely disappear.

Since both of the above mentioned complexity reduction techniques rely on abusing similarities in the amplitude of the constellation points and states thereof, we conclude that amplitude processing is insufficient for obtaining significant shaping gains in the highly nonlinear region of the fiber when limited memory is exploited and when the constellation is of low cardinality (16QAM considered here vs. 256QAM in e.g. [16])[3]. Different merging criteria are therefore needed in order to increase the order of the FSMS while retaining the complexity and preserving the shaping gains from Fig. 5. This is further supported by Fig. 6, where the optimized PMF of the branches is given as a function of the MD amplitude of the branch $A(b) = \left[s^i, x^j\right]^H \left[s^i, x^j\right]$ for an FSMS with $N = 2$ optimized with Algorithm 1 and an MDMB PMF. We see that branches with equal amplitude can exhibit substantially different probability of occurrence, which cannot be attributed to computational inaccuracy. An MDMB fit is also plotted to the expected value of the probability of occurrence of each amplitude, which we see provides a good approximation. However, the lack of further diversity of the MDMB PMF restricts the shaping gains.

When the number of channels is increased, the gains are reduced similarly to the QPSK case. However, for 3 channels, the 16QAM FSMS, including with memoryless receiver processing, allows for slightly improved AIR w.r.t. standard probabilistic shaping methods (FSMS with $N = 0$). When the number of channels is increased, the intra-channel memory becomes more inconsequential, and taking it into account at the receiver and/or transmitter provides only marginal gains. For 9 WDM channels, ≈ 0.02 bits/symbol are achieved with $N = 2$ w.r.t. $N = 0$. MDMB achieves approximately the same gain. We note that an FSMS with $N = 2$ and memoryless

[3]We note that the MDMB optimization is one-dimensional, which allows to optimize the input and generate sequences with large N. We did not observe notable improvements with $N = 20$ and memoryless processing at the receiver.

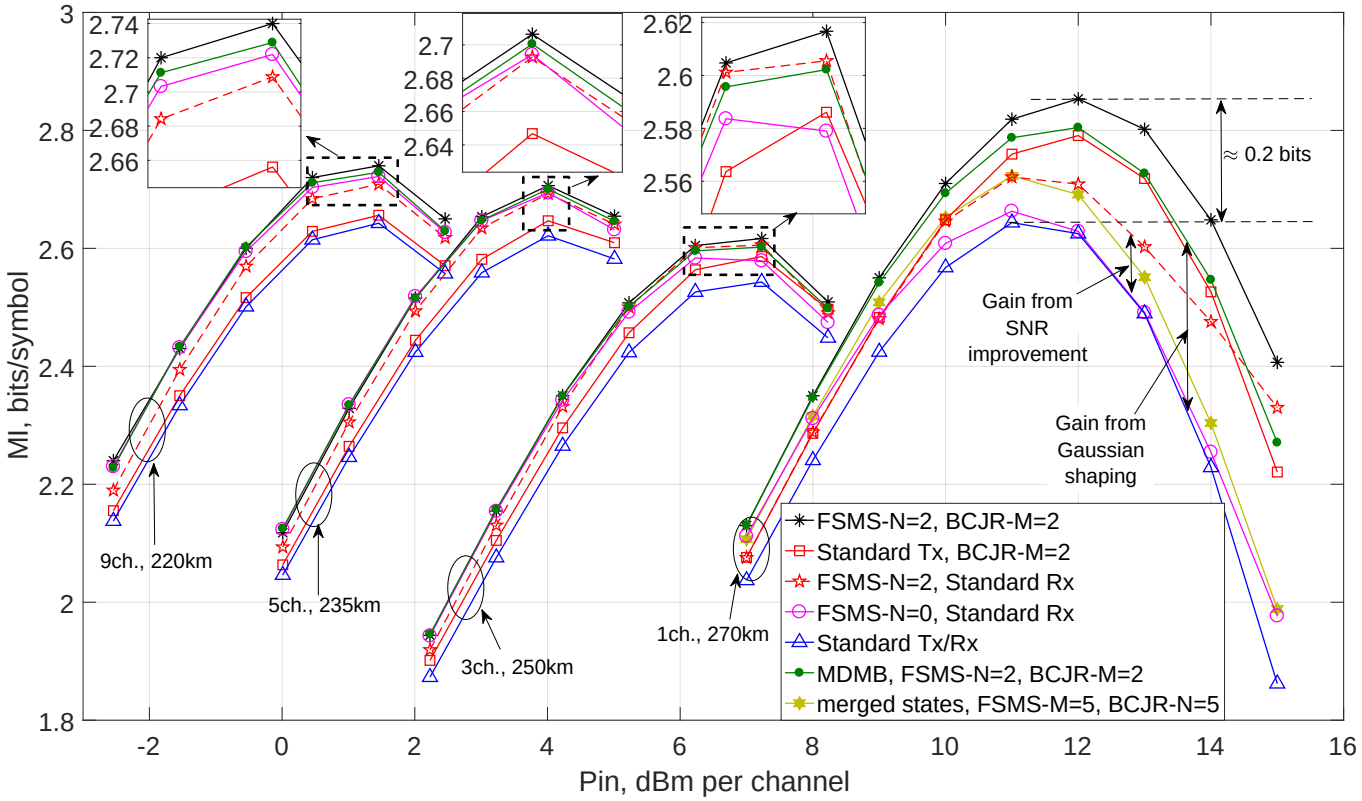


Fig. 5. Unrepeated transmission of FSMS with 16QAM output. The optimized FSMS enables increased launch power for single channel and gains of around 0.2 bits/symbol. In the nonlinear region, MDMB is outperformed by the optimization from Algorithm 1. The gains are decreased in WDM scenarios, and become marginally better than the gains of standard memoryless shaping for more than 3 WDM channels. In the latter case, FSMS of higher order is still beneficial, including with memoryless processing at the receiver.

processing is penalized in the linear region of transmission. This is due to the fact, that Algorithm 1 converges to a local optimum, which is different than that of a memoryless solution.

The resulting waveforms of the optimized FSMSs are further studied in terms of their peak-to-average power ratio (PAPR), and in particular, its complementary cumulative distribution function (CCDF) after pulse shaping. The CCDF shows how often a certain PAPR is exceeded, and is estimated numerically. An example of the single channel CCDF for an optimized FSMS of $N = 5$ and QPSK output is given in Fig. 7(a) for launch powers 12 dBm (linear region), 15 dBm (optimal) and 18 dBm (highly nonlinear region). We see that the FSMS allows for reduced PAPR of the pulse-shaped waveform by ≈ 0.35 dB and up to ≈ 0.8 dB at CCDF$= 10^{-4}$ and CCDF$= 10^{-6}$, respectively. This reduction allows for the shaping gains with QPSK alphabet, which otherwise has a constant envelope at the sampling points.

In Fig. 7(b), the PAPR which is exceeded with probability 10^{-4} is given for single channel and WDM as a function of the launch power for optimized FSMS of $N = 5$. Due to the central limit theorem, the PMF in each channel has less impact on the distribution of the multi-channel waveform and thereby its PAPR, when the number of channels is increased (the distribution will tend to Gaussian for infinite number of channels). For 9 channels, the PAPR cannot be reduced with the FSMS, which also results in diminished shaping gains.

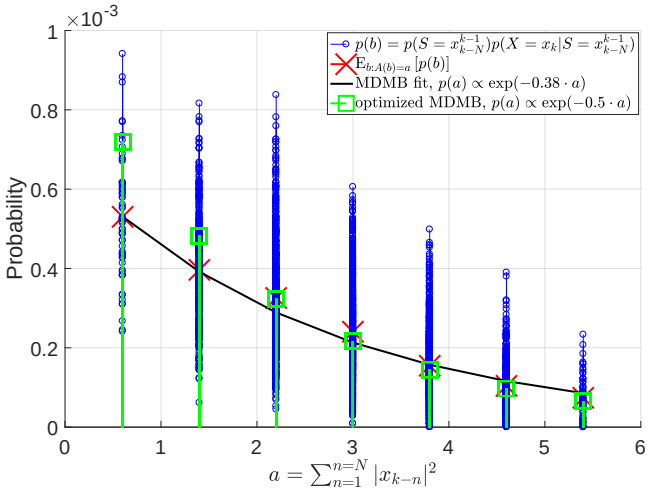
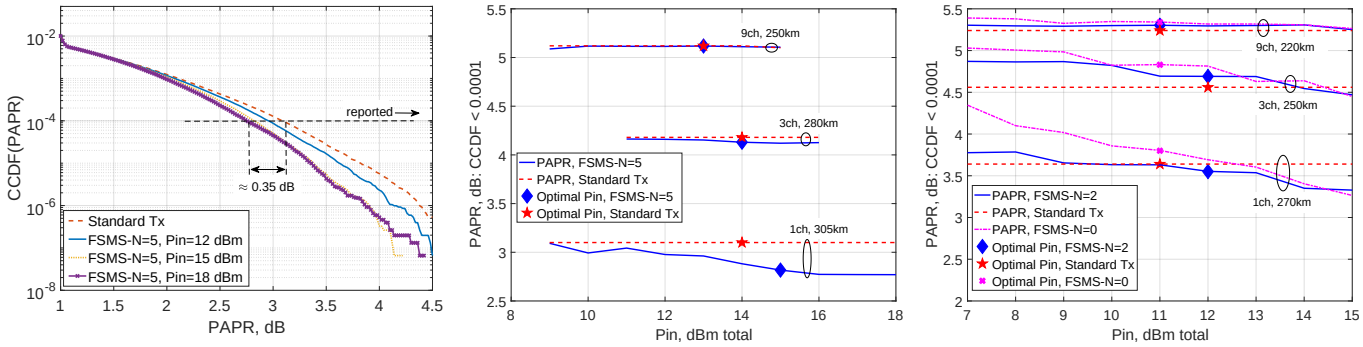


Fig. 6. Optimized PMF of the branches in an FSMS with 16QAM and $N = 2$ as a function of their MD amplitude for 270 km unrepeated transmission on a single channel. An MDMB fit to the expected value of the probability of occurrence of each amplitude is also shown, together with an independently optimized MDMB PMF as in Section III-C. The MDMB potentially allows for capturing the expected value of the probability of occurrence of each amplitude, but not the complete function.

In Fig. 7(c), the PAPR is given for 16QAM. In the linear region of transmission, high PAPR has no impact on the



(a) PAPR CCDF, single channel QPSK (b) QPSK PAPR at CCDF$< 10^{-4}$ (c) 16QAM PAPR at CCDF$< 10^{-4}$

Fig. 7. PAPR study for QPSK [**a**) and **b**)] and 16QAM [**c**)]. For a few WDM channels, the PAPR with single channel processing is reduced with an optimized FSMS, and shaping gain is thus achieved. For increased number of channels, due to the central limit theorem, the PMF in each channel has little impact on the PAPR of the multi-channel waveform.

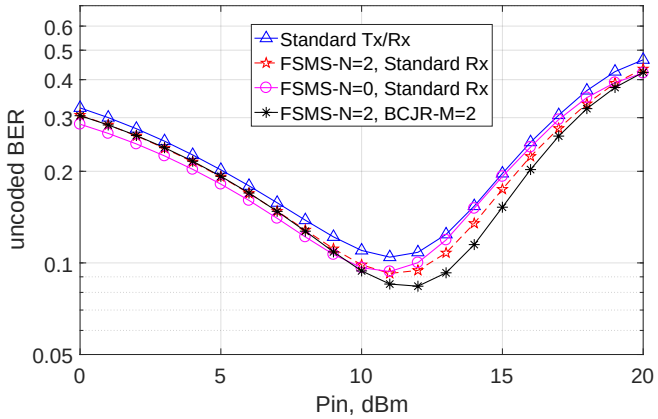


Fig. 8. BER of unrepeated transmission of FSMS with 16QAM output after 270 km with fixed PMFs, obtained at the respective optimal launch powers for FSMS with $N = 2$ and $N = 0$. The PMF with $N = 2$ becomes sub-optimal for linear transmission.

AIR, and the source converges to a high-PAPR solution, similar to the MB or MDMB. When the power is increased, a memoryless source does not allow for reduced PAPR, and thus provides very little gain at the optimal launch power. As mentioned, in the WDM cases, single-channel processing and optimization have less and less impact on the multi-channel waveform properties, and the gain is thus decreased (as seen in Fig. 5). The optimized source in such cases has a more Gaussian structure in order to mitigate the impact of the linear amplification noise.

We also study the robustness of fixed PMFs for $N = 2$ and $N = 0$, obtained at the respective optimal launch powers, while simultaneously demonstrating demodulation of the FSMS with BCJR processing. The 16QAM PMF for $N = 2$ is obtained at $P_{in} = 12$ dBm, and at $P_{in} = 11$ dBm for $N = 0$ (see Fig. 5). The uncoded BER with those PMFs is given in Fig. 8 for the entire launch power region at 270 km. The BER with $N = 2$ is decreased especially in the non-linear region. However, when the launch power is reduced, the PMF with $N = 2$ optimized for nonlinear transmission performs sub-optimally to the PMF with $N = 0$, which is optimized at a lower power and is mainly focused on mitigating the linear noise.

B. Amplified WDM system distance improvement

Finally, we study a WDM link with 9x28 GBaud channels and ideal distributed Raman amplification (IDRA), modeled as in [5]. The rest of the parameters are as in Table I. The results for 16QAM are given in Fig. 9. Similar to the unrepeated WDM transmission case, temporal shaping provides only marginal gains over memoryless shaping. Temporal shaping is even more challenging in this case due to the increased transmission length and the fact that the accumulated dispersion makes the channel memory longer. Short memory FSMSs ($N \leq 2$) therefore fail to produce significant improvements. Indeed, we observed that similar to the linear region of transmission, Algorithm 1 converges to a solution, for which the conditional probability $p(X = x^j|S)$ approximates the marginal distribution $p(X = x^j) = \sum_i p(B = \{i, j\})$, i.e., the current output is independent of the past. Such a solution is beneficial in mitigating the white part of the noise (linear and nonlinear), but barely provides SNR improvement.

In order to achieve temporal shaping gains (and in fact, increased memoryless shaping gains) in the case of WDM with significant XPM, cross-channel processing is required both at the transmitter and the receiver. This is further discussed in Section V.

V. Discussion and Future Work

One of the main limitations of Algorithm 1 is the requirements for very long sequences for accurate calculation of the values $T_{i,j}$ for high FSMS order N. Furthermore, generating the sequence y_1^K requires performing the SSFM, which is slow and cumbersome for long sequences. This limitation can be circumvented by employing some of the recent techniques for fiber channel modeling [17], [18]. However, in their present form, i.i.d. input symbols are assumed when generating the channel noise and the channel output samples. Generalizing these techniques to generating correlated noise samples would be of great interest for expanding the work in this paper to longer memory and higher FSMS order. This will allow for reduced running time, however, the complexity will still be dominated by the BCJR processing at the receiver. It is noted that Algorithm 1 is general, and can be applied directly also for multiple-input multiple-output type systems and/or

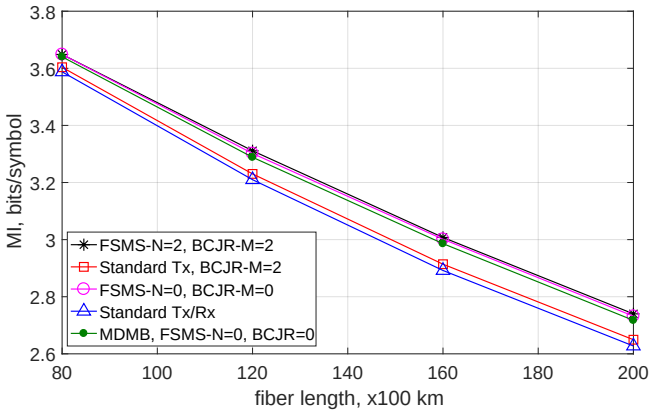


Fig. 9. Optimized sources for an IDRA 9x28 GBaud WDM transmission. Increasing the order to only $N = 2$ does not provide significant gains over standard probabilistic shaping. The gains from memory exploit and shaping appear to add-up.

multi-channel processing (e.g. super-channels). As seen from Fig. 3 and Fig. 5, the optimized FSMS allows for gains also with memoryless processing, thus operating at low complexity. Furthermore, the fact that gains are achieved in a 3-channel setup with only single channel processing indicate that further improvements in WDM scenarios are possible with cross-channel processing. Together with the fact, that the optimization is performed offline, this makes FSMSs a promising tool for achieving shaping gain in the above mentioned systems.

In this work, QAM constellations were chosen as outputs of the FSMS due to their popularity and practical transceiver benefits such as Gray coding, readily available digital signal processing, etc.. Other types of constellations can be directly adopted in this framework without loss of generality. In particular, geometrically optimized constellations, e.g. ring constellations [19], with probabilistically shaped temporal properties are of interest for future research.

As seen in Fig. 6, MDMB can provide a good fit to the *expected* value of the probability of occurrence of each amplitude. Introducing hyper-parameters, such as variance of the probability of occurrence may allow for describing a modified MDMB PMF which captures the required diversity of the probabilities of each amplitude for achieving the shaping gains from Fig. 5. Such closed form expressions of the PMF and reduced dimensionality optimization would in turn allow for increasing the FSMS order and potentially higher gains.

This work describes theoretical AIRs on the fiber, however, communicating binary data with this system is not trivial. In order to build a practical system, a method is required which maps the binary data to the branches of the FSMS with the specified PMF. This may be possible with the techniques from e.g. [20], [21]. Alternatively, non-binary coding and communication can be considered for easier shaping at the symbol level.

VI. Conclusion

In this paper, finite state machine sources (FSMS) were proposed for temporal probabilistic constellation shaping for nonlinear fiber optic systems. It was shown that shaping the probabilities of sequences of QAM symbols is beneficial and in some cases significantly outperforms memoryless shaping. Gains were demonstrated in links, dominated by self-phase modulation, where temporal shaping allows for increased launch power and thereby achievable information rates. In links, dominated by cross-phase modulation, such as wavelength division multiplexing systems, temporal shaping requires cross-channel processing and optimization and/or increased FSMS order in order to significantly outperform memoryless shaping.

VII. Acknowledgments

This work was supported by the Danish National Research Foundation (DNRF) Research Centre of Excellence, SPOC, ref. DNRF123. The authors would like to thank the reviewers for the suggestion of studying the WDM waveforms in terms of peak power properties.

References

[1] M. P. Yankov, F. D. Ros, E. P. da Silva, S. Forchhammer, K. J. Larsen, L. K. Oxenløwe, M. Galili, and D. Zibar, "Constellation shaping for WDM systems using 256QAM/1024QAM with probabilistic optimization," *IEEE Journal of Lightwave Technology*, vol. 34, no. 22, pp. 5146–5156, Nov. 2016.

[2] 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," *IEEE Journal of Lightwave Technology*, vol. 34, no. 7, pp. 1599–1609, Apr. 2016.

[3] T. Liu, C. Lin, and I. B. Djordjevic, "Advanced GF(3^2) nonbinary LDPC coded modulation with non-uniform 9-QAM outperforming star 8-QAM," *Optics Express*, vol. 24, no. 13, pp. 13 866–13 874, June 2016.

[4] C. Pan and F. R. Kschischang, "Probabilistic 16-QAM shaping in WDM systems," *IEEE Journal of Lightwave Technology*, vol. 34, no. 18, pp. 4285–4292, Sep. 2016.

[5] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *IEEE Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

[6] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On shaping gain in the nonlinear fiber-optic channel," in *Proc. of IEEE International Symposium on Information Theory*, July 2014, pp. 2794–2798.

[7] P. O. Vontobel, A. Kavčić, D. M. Arnold, and H.-A. Loeliger, "A generalization of the BlahutArimoto algorithm to finite-state channels," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1887–1918, May 2008.

[8] M. P. Yankov and S. Forchhammer, "Temporal probabilistic constellation shaping for WDM optical communication systems," in *Proc. of European Conference on Optical Communications (ECOC)*, Sep. 2016, p. W.1.C.5.

[9] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[10] I. B. Djordjevic, B. Vasic, M. Ivkovic, and I. Gabitov, "Capacity limits of optical fiber networks," *IEEE Journal of Lightwave Technology*, vol. 23, no. 10, pp. 3755–3763, Nov. 2005.

[11] T. A. Eriksson, T. Fehenberger, P. A. Andrekson, M. Karlsson, N. Hanik, and E. Agrell, "Impact of 4d channel distribution on the achievable rates in coherent optical communication experiments," vol. 34, no. 9, pp. 2256–2266, May 2016.

[12] C. M. Bishop, *Pattern recognition and machine learning, 2nd edition*. New York, NY: Springer, 2006.

[13] N. Varnica, X. Ma, and A. Kavčić, "Capacity of power constrained memoryless awgn channels with fixed input constellations," in *Proc. of GLOBECOM*, Nov. 2002, pp. 1339–1343.

[14] S. ten Brink, J. Speidel, and R.-H. Yan, "Iterative demapping and decoding for multilevel modulation," in *Proc. of IEEE Globecom*, Nov. 1998, pp. 579–584.

[15] R. Dar, M. Shtaif, and M. Feder, "New bounds on the capacity of the nonlinear fiber-optic channel," *Optics letters*, vol. 39, no. 2, pp. 398–401, Jan. 2014.

[16] O. Geller, R. Dar, M. Feder, and M. Shtaif, "A shaping algorithm for mitigating inter-channel nonlinear phase-noise in nonlinear fiber systems," *IEEE Journal of Lightwave Technology*, vol. 34, no. 16, pp. 3884–3889, Aug. 2016.
[17] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-channel nonlinear interference noise in WDM systems : Modeling and mitigation," *IEEE Journal of Lightwave Technology*, vol. 33, no. 5, pp. 1044–1053, Mar. 2015.
[18] A. Carena, G. Bosco, V. Curri, Y. Jiang, P. Poggiolini, and F. Forghieri, "EGN model of non-linear fiber propagation," *Optics Express*, vol. 22, no. 13, pp. 16 335–16 362, June 2014.
[19] T. Freckmann, R.-J. Essiambre, P. J. Winzer, G. J. Foschini, and G. Kramer, "Fiber capacity limits with optimized ring constellations," *IEEE Photonics Technology Letters*, vol. 21, no. 20, pp. 1496–1498, Oct. 2009.
[20] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4651–4665, Oct. 2015.
[21] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.

Metodi Yankov (S'13, M'16) received a B. Eng. degree from the Technical University of Sofia, Bulgaria in 2010 in the field of radio communications, and a M. Sc. degree from the Technical University of Denmark (DTU), Lyngby in 2012 in the area of signals and transmission technology for telecommunications. He obtained a PhD degree from DTU in March, 2016, and the title of his thesis was "Capacity estimation and near-capacity achieving techniques for digitally modulated communication systems". He has since been employed as a post-doc at the Coding and Visual Communications group at DTU, Department of Photonics Engineering. His interests include coded modulation and iterative receivers, as well as information theory of both wireless and optical channels.

Knud J. Larsen received his M.S.E.E. and Ph.D. degrees from the Technical University of Denmark, Kgs. Lyngby. He was at Ericsson Signal Systems as a senior engineer and project manager involved in design and implementation of communication systems and traffic control systems. Since 1986 he has been a professor at the Technical University of Denmark (previously the Institute of Telecommunication, now the Department of Photonics Engineering), teaching and doing research in communication theory and coding. He has worked extensively on studies of error correction for the European Space Agency and is now involved in projects related to optical transmission. Together with four colleagues he received the 1991 Information Theory Society Paper Award for a paper on algebraic geometry codes.

Søren Forchhammer (M'04) received the M.S. degree in engineering and the Ph.D. degree from the Technical University of Denmark, Lyngby, in 1984 and 1988, respectively. Currently, he is a Professor with DTU Fotonik, Technical University of Denmark, where he has been since 1988. He is Head of the Coding and Visual Communication Group at DTU Fotonik. His main interests include information theory, capacity calculations, two-dimensional information theory, signal processing for optical communication, source coding, image and video coding, distributed source coding, processing for image displays, and visual communications.