Joint Channel Estimation and MAP Detection of Probabilistically Shaped QAM

Yilun DU
Laboratoire PRISME/IRAuS
Université d'Orléans
Orléans, France
yilun.du@etu.univ-orleans.fr

Karim Abed-Meraim, *Fellow, IEEE*
Laboratoire PRISME/IRAuS
Université d'Orléans
Orléans, France
karim.abed-meraim@univ-orleans.fr

Ouahbi Rekik
L2TI, UR 3043
Université Sorbonne Paris Nord
Paris, France
ouahbi.rekik@univ-paris13.fr

***Abstract*—This paper introduces a novel blind solution for joint channel estimation and data detection of convolutive multichannel communications systems where Probabilistic Shaping (PS) is used. The convolutive channel estimation is based on the Expectation-Maximization (EM) algorithm, and data detection is achieved by re-utilizing the probabilities obtained within the EM framework, based on a maximum *a posteriori* (MAP) estimation. This work shows that, even if the source kurtosis is close to the Gaussian one, the blind estimation is still possible via the proposed method contrary to other existing HOS (Higher Order Statistics) based methods. Simulation results show that our algorithm provides an interesting channel estimation accuracy and a much better performance in terms of symbol error rate (SER) as compared to Hyperbolic-Givens multi-modulus algorithm (HG-MMA) which is based on multi-modulus criterion.**

Index Terms*—Probabilistic Shaping, QAM modulation, EM algorithm, Maximum *a Posteriori

I. Introduction

In the context of wireless and coherent optical communications, PS has become a promising technique to enable some recent record-setting transmission experiments [1], [2]. Generally, PS is able to provide fine-grained rate adaptability (flexibility) and energy efficiency (sensitivity) gains [3]. PS technique also allows high- (low-) energy symbols to occur with relatively lower (higher) probability, which reduces the average transmission energy per bit and consequently improves the wireless reach in mobile communication systems. In optical fiber transmission, the implementation of PS can also effectively reduce the average power of laser signal, which allows the optical link to have a higher tolerance to fiber nonlinear effects. What's more, PS technique realized through probabilistic amplitude shaping (PAS) architecture provides an elegant solution to the combination of shaping and coding [4], [5], which was a long-standing problem of PS. PAS architecture decouples coding and shaping so that each can be done in a parallel way, which allows the implementation of off-the-shelf modern soft-decision forward-error correction (SD-FEC) in the PS-based communication systems.

However, PS also brings significant challenges for hardware and digital signal processing (DSP) algorithms in transceiver. An unavoidable problem in DSP is that the conventional channel estimation and equalization algorithms, which have been designed for quadrature amplitude modulation (QAM) with uniform symbol probability distribution and relying on high order statistics (HOS) of data, have poor performance when PS is implemented. This is because PS gives a Gaussian-like probability distribution, which affects significantly the identification methods based on the HOS. Detailed analysis of convergence failure of constant-modulus algorithm (CMA) and MMA due to the closeness to Gaussianity of sources has been presented in [6].

For this reason, in most experiment demonstrations, data-aided techniques are often utilized to solve such a problem [7], [8]. In this paper, we show that this shortcoming, in the blind context, is not inherent to the problem at hand and affects only methods based on fourth order cumulants, e.g. CMA- and MMA-like methods. For that, we propose to exploit the full HOS information via a maximum likelihood approach to overcome this limitation. Hence, a blind EM-based algorithm for PS, which realizes blind channel estimation and data detection in MIMO communications systems; when considering convolutive (multi-tap) channels is proposed. This is an iterative method, whose objective is to find the maximum likelihood estimates of channel's parameters in statistical models that are related to unobserved variables and associated to a Markov process. Moreover, as is mentioned above, the proposed method can work compatibly with modern SD-FEC thanks to PAS architecture of PS technique. Simulation results show that the proposed solution exhibits very promising channel estimation accuracy and much better performance of data detection as compared to standard blind methods such as the recently proposed HG-MMA algorithm [9].

II. System Model

This section introduces the channel model and notations adopted in this paper. A MIMO communications system with N_t independent and identically distributed (i.i.d.) source signals is considered, where we denote by $u_t(k)$ the transmitted symbol sequence by t-th source. The transmitted signal is generated according to *Maxwell-Boltzmann* (MB) distribution [3] given by:

$$P_X(x) = \frac{\exp\left(-\lambda x^2\right)}{\sum_{t_0 \in \mathcal{J}} \exp(-\lambda t_0^2)}, x \in \mathcal{J} \tag{1}$$

where λ is a parameter that controls the entropy of the source X, denoted by $H(X)$ and expressed as $H(X) = -\sum_{X\in\mathcal{J}} P_X(x)\log_2 P_X(x)$, with unit bits/symbol. $\mathcal{J} = \{\pm1, \pm3, ..., \pm(D-1)\}$ represents the alphabet set of PS-D-PAM with its probabilistic vector being given by $\boldsymbol{p} = [P_X(1-D), ..., P_X(-1), P_X(1), \ \ ..., P_X(D-1)]^{\mathrm{T}}$. For PS-$D^2$-QAM, whose entropy is given by $2H(X)$ in this context, its distribution is given by the probabilistic matrix $\mathbf{Q} = \boldsymbol{p}\boldsymbol{p}^{\mathrm{T}}$ so that the probability of each complex symbol is given by $P(x + \mathrm{j}y) = P_X(x)P_X(y)$, where $\mathrm{j} = \sqrt{-1}$ and $x, y \in \mathcal{J}$.

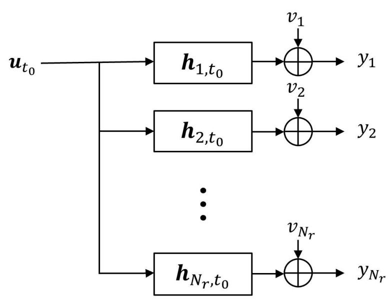

Fig. 1: Channel between t_0-th source and all the receivers

The architecture of the considered MIMO convolutive channel is illustrated in Fig. 1, in which, for simplification, only the channels between t_0-th source and all the receivers is shown, while for the other sources, the channel structure will be the same. At the receiver side, N_r receive antennas are considered, so that the r-th received signal, denoted by $y_r(k)$ with $1 \leqslant r \leqslant N_r$, is given by:

$$y_r(k) = \sum_{t=1}^{N_t}\sum_{n=0}^{M} h_{r,t}(n)u_t(k-n) + v_r(k), \tag{2}$$

where $k = 1, 2, ..., N_s$. N_s is the sample size and $h_{r,t}(n)$ refers to the finite impulse response coefficients of the channel, with a length of $(M+1)$, associated with the t-th transmitter and the r-th receiver. $v_r(k)$ denotes an additive white circular Gaussian noise with variance σ_v^2, which is independent among the receivers.

In order to express the system model in a more compact form, let's define the following vectors and matrices: $\boldsymbol{h}_r = [h_{r,1}(0), ..., h_{r,1}(M), ..., h_{r,N_t}(0), ..., h_{r,N_t}(M)]^{\mathrm{T}}$, $\boldsymbol{u}(k) = [u_1(k), ..., u_1(k-M), ..., u_{N_t}(k), ..., u_{N_t}(k-M)]^{\mathrm{T}}$, $\boldsymbol{y}(k) = [y_1(k), ..., y_{N_r}(k)]^{\mathrm{T}}$, $\mathbf{H} = [\boldsymbol{h}_1, ..., \boldsymbol{h}_{N_r}]^{\mathrm{T}}$, $\boldsymbol{v}(k) = [v_1(k), ..., v_{N_r}(k)]^{\mathrm{T}}$. Accordingly, (2) can be re-written as:

$$\boldsymbol{y}(k) = \mathbf{H}\boldsymbol{u}(k) + \boldsymbol{v}(k). \tag{3}$$

The system model presented by (3) can be considered as a Markov process whose state vector is defined as $\boldsymbol{s}(k) = [u_1(k-1), ..., u_1(k-M), ..., u_{N_t}(k-1), ...u_{N_t}(k-M)]^{\mathrm{T}}$ that consists of N_tM successive symbols, belonging to a set of D^{2N_tM} states, denoted by $\mathcal{S} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, ..., \boldsymbol{b}_N\}$, $N = D^{2N_tM}$. The transmitted symbols are assumed to be i.i.d. so that $P(\boldsymbol{s}(k) = \boldsymbol{b}_i) = P_i = \prod_{t=1}^{N_t}\prod_{n=1}^{M} P(u_t(k-n))$. The transition vector, associated with the transition between two successive state vectors $\boldsymbol{b}_i$ and $\boldsymbol{b}_j$, is expressed by $\boldsymbol{t}_{ij} = \boldsymbol{t}(k) = [u_1(k), ..., u_1(k-M), ..., u_{N_t}(k), ...u_{N_t}(k-M)]^{\mathrm{T}}$, containing $N_t(M+1)$ symbols. Similarly, $P(\boldsymbol{t}_{ij})$ can be expressed as $P_{ij} = P(\boldsymbol{s}(k) = \boldsymbol{b}_i, \boldsymbol{s}(k+1) = \boldsymbol{b}_j) = \prod_{t=1}^{N_t}\prod_{n=0}^{M} P(u_t(k-n))$. The set of all possible transition vectors, denoted by $\mathcal{T}$, contains $D^{2N_t(M+1)}$ elements.

Our very objective is to estimate jointly the channel and the data in a blind (eventually semi-blind) context. In the known channel case, it is established that when parameter λ in (1) corresponds to the maximum kurtosis value, the system performance (in term of symbol error rate) is significantly improved [3]. Unfortunately, in such a case the PS signal's kurtosis is close to 3, the one of the Gaussian distribution, which makes the blind channel estimation ineffective for methods relying on 4-th order information, e.g. CMA- or MMA-like methods [6]. According to some numerical analysis, the maxima of kurtosis of PS-16-QAM and PS-64-QAM are 2.778 and 2.999, respectively.

In this work, we highlight the fact that even if the kurtosis is close to 3 (i.e., the 4th order cumulant is close to zero), the HOS information of our probabilistically shaped constellation is still available in cumulants of higher order. Table I illustrates this claim by comparing the 6th order and 4th order cumulant values of PS-16-QAM and PS-64-QAM when their kurtosis are maximal (i.e. closest to 3).

TABLE I: Value of n-th order cumulant of PS-QAM at maximum of kurtosis

Order of cumulant	2	4	6
Cumulant value (PS-16-QAM)	3.608	-1.446	11.35
Cumulant value (PS-64-QAM)	5.988	-0.02467	0.3955

To exploit such an HOS information, we propose next to use a maximum likelihood approach which optimization is performed via the EM algorithm.

III. EM-based Estimation

This section details the proposed solution for blind estimation of MIMO channels. A maximum likelihood (ML)-based approach is proposed, which is minimized by using the EM algorithm. This algorithm is an iterative procedure aiming at finding maximum likelihood or maximum a posteriori estimates of parameters in statistical models, depending on unobserved latent variables.

In what follows, the parameters to be estimated are the channel impulse response coefficients and the noise variance, which are grouped in parameter vector $\boldsymbol{\rho} = \left[\mathrm{vec}^{\mathrm{T}}(\mathbf{H}), \sigma_v^2\right]^{\mathrm{T}}$, where $\mathrm{vec}(\cdot)$ denotes the vectorization of a matrix. Also, for convenience, a series of successive observations $(\boldsymbol{y}(t_1), \boldsymbol{y}(t_2), ..., \boldsymbol{y}(t_n))$ and state vectors $(\boldsymbol{s}(t_1), \boldsymbol{s}(t_2), ..., \boldsymbol{s}(t_n))$ are denoted by matrices $\mathbf{Y}_{[t_1:t_n]}$ and $\mathbf{S}_{[t_1:t_n]}$, respectively. Therefore, the observations $\mathbf{Y}_{[t_1:t_n]}$ represent the incomplete data, the states $\mathbf{S}_{[t_1:t_n]}$ stand for the missing data, while the complete data are given by $(\mathbf{Y}_{[t_1:t_n]}, \mathbf{S}_{[t_1:t_n]})$.

The EM algorithm is a well known technique that optimizes the ML function via the iterative maximization of an auxiliary function (given below in eq.(4)). For that, one alternates between two steps: E-step (for the evaluation of the auxiliary function) and M-step (for its maximization). Next, we propose to adapt it to our specific context to achieve the desired channel estimation and symbol detection objective.

A. E-step

The aim of this step is to calculate the auxiliary function, denoted by $Q(\boldsymbol{\rho};\boldsymbol{\rho}^{(l)})$, which is given by the conditional expectation of the complete-data log-likelihood, with respect to the conditional distribution of the missing data, given the observations and the current estimated parameter value at the l-th iteration, i.e. $\boldsymbol{\rho}^{(l)} = \left[\mathrm{vec}^{\mathrm{T}}(\mathbf{H}^{(l)}), (\sigma_v^2)^{(l)}\right]^{\mathrm{T}}$. Accordingly, such an auxiliary function is expressed as follows:

$$Q(\boldsymbol{\rho};\boldsymbol{\rho}^{(l)}) = \mathbb{E}\left(\log f_{\boldsymbol{\rho}}\left[\left(\mathbf{Y}_{[1:N_s]}, \mathbf{S}_{[1:N_s]}\right) \middle| \mathbf{Y}_{[1:N_s]}\right]; \boldsymbol{\rho}^{(l)}\right). \tag{4}$$

where $\mathbb{E}(\cdot)$ refers to the expectation w.r.t. the distribution of the missing data.

After some straightforward simplifications; and by ignoring terms that are irrelevant to $\boldsymbol{\rho}$, one obtains:

$$Q(\boldsymbol{\rho};\boldsymbol{\rho}^{(l)}) \propto \sum_{\boldsymbol{t}_{ij}\in\mathcal{T}} \sum_{k=1}^{N_s} \left(- N_r \log(\sigma_v^2) - \frac{\|\boldsymbol{y}(k) - \mathbf{H}\boldsymbol{t}_{ij}\|^2}{\sigma_v^2} \right) \eta_{\boldsymbol{\rho}^{(l)}}(k;i,j) \tag{5}$$

$$\eta_{\boldsymbol{\rho}^{(l)}}(k;i,j) = P\left(\boldsymbol{s}(k)=\boldsymbol{b}_i, \boldsymbol{s}(k+1)=\boldsymbol{b}_j \middle| \mathbf{Y}_{[1:N_s]}; \boldsymbol{\rho}^{(l)}\right) \tag{6}$$

where $\eta_{\boldsymbol{\rho}^{(l)}}(k;i,j)$ stands for the posterior probability of the two successive state vectors ($\boldsymbol{s}(k) = \boldsymbol{b}_i, \boldsymbol{s}(k+1) = \boldsymbol{b}_j$), given the observations $\mathbf{Y}_{[1:N_s]}$ with the estimated channel parameter $\boldsymbol{\rho}^{(l)}$. $\eta_{\boldsymbol{\rho}^{(l)}}$ can be calculated iteratively by using forward-backward variables denoted by $\alpha_{\boldsymbol{\rho}^{(l)}}(k;i)$ and $\beta_{\boldsymbol{\rho}^{(l)}}(k;j)$ [10], [11]. Neglecting factors independent from k, i and j, one can write:

$$\eta_{\boldsymbol{\rho}^{(l)}}(k;i,j) \propto \alpha_{\boldsymbol{\rho}^{(l)}}(k;i)\beta_{\boldsymbol{\rho}^{(l)}}(k+1;j)\phi_{\boldsymbol{\rho}^{(l)}}(k;i,j)P_{ij}, \tag{7}$$

$$\alpha_{\boldsymbol{\rho}^{(l)}}(k;i) = P\left(\mathbf{Y}_{[1:k-1]} \middle| \boldsymbol{s}(k)=\boldsymbol{b}_i; \boldsymbol{\rho}^{(l)}\right), \tag{7a}$$

$$\beta_{\boldsymbol{\rho}^{(l)}}(k+1;j) = P\left(\mathbf{Y}_{[k+1:N_s]} \middle| \boldsymbol{s}(k+1)=\boldsymbol{b}_j; \boldsymbol{\rho}^{(l)}\right), \tag{7b}$$

$$\phi_{\boldsymbol{\rho}^{(l)}}(k,i,j) = P\left(\mathbf{Y}_{[k]} \middle| \boldsymbol{s}(k)=\boldsymbol{b}_i, \boldsymbol{s}(k+1)=\boldsymbol{b}_j; \boldsymbol{\rho}^{(l)}\right). \tag{7c}$$

By ignoring the terms that are independent from $\boldsymbol{\rho}$, $\phi_{\boldsymbol{\rho}^{(l)}}$ can be computed by:

$$\phi_{\boldsymbol{\rho}^{(l)}} \propto \left((\sigma_v^2)^{(l)}\right)^{-N_r} \exp\left(-\frac{\|\boldsymbol{y}(k) - \mathbf{H}^{(l)}\boldsymbol{t}_{ij}\|^2}{(\sigma_v^2)^{(l)}}\right). \tag{8}$$

The forward and backward variables can be computed recursively by:

$$\alpha_{\boldsymbol{\rho}^{(l)}}(k+1;i) = \sum_{c\in\mathcal{F}(i)} \alpha_{\boldsymbol{\rho}^{(l)}}(k;c)\phi_{\boldsymbol{\rho}^{(l)}}(k,c,i)P_c, \tag{9}$$

$$\beta_{\boldsymbol{\rho}^{(l)}}(k;j) = \sum_{c\in\mathcal{B}(j)} \beta_{\boldsymbol{\rho}^{(l)}}(k+1;c)\phi_{\boldsymbol{\rho}^{(l)}}(k,j,c)P_c, \tag{10}$$

where P_c is the probability of the state vector c, $\mathcal{F}(i)$ (resp. $\mathcal{B}(j)$) denotes the set that contains all the state vectors in forward (resp. backward) connected to $\boldsymbol{b}_i$ (resp. $\boldsymbol{b}_j$), which is referred to as predecessors (resp. successors) hereinafter. The number of predecessors (successors) of a state vector is D^{2N_t}. Here the initialization for forward and backward iterations is given by $\alpha_{\boldsymbol{\rho}^{(l)}}(1;i) = \beta_{\boldsymbol{\rho}^{(l)}}(N_s;j) = \frac{1}{D^{2MN_t}}$, for all possible i and j.

B. M-step

The aim of M-step is to find the vector $\boldsymbol{\rho}^{(l+1)}$ that satisfies:

$$\boldsymbol{\rho}^{(l+1)} = \arg\max_{\boldsymbol{\rho}} Q\left(\boldsymbol{\rho};\boldsymbol{\rho}^{(l)}\right). \tag{11}$$

According to (5), $Q\left(\boldsymbol{\rho};\boldsymbol{\rho}^{(l)}\right)$ has a quadratic form, thus $\frac{\partial Q(\boldsymbol{\rho};\boldsymbol{\rho}')}{\partial\boldsymbol{\rho}} = 0$ leads to an unique solution, which can be expressed by:

$$\mathbf{H}^{(l+1)} = \mathbf{R}_{yt}\mathbf{R}_{tt}^{-1}, \tag{12}$$

$$\left(\sigma_v^2\right)^{(l+1)} = \frac{1}{N_s N_r} \mathrm{tr}\left(\mathbf{R}_{yy} - \mathbf{H}^{(l+1)}\mathbf{R}_{ty}\right), \tag{13}$$

where auto-correlation matrices $\mathbf{R}_{yy}$, $\mathbf{R}_{tt}$ and cross-correlation matrix $\mathbf{R}_{yt}$ are given by:

$$\mathbf{R}_{yy} = \sum_{k=1}^{N_s} \boldsymbol{y}(k)\boldsymbol{y}^{\mathrm{H}}(k), \tag{14}$$

$$\begin{aligned}\mathbf{R}_{ty} = \mathbf{R}_{yt}^{\mathrm{H}} &= \sum_{k=1}^{N_s} \sum_{\boldsymbol{t}_{ij}\in\mathcal{T}} \boldsymbol{t}_{ij}\boldsymbol{y}^{\mathrm{H}}(k)\eta_{\boldsymbol{\rho}^{(l)}}(k;i,j) \\ &= \sum_{k=1}^{N_s} \mathbb{E}\left(\boldsymbol{t}(k) \middle| \mathbf{Y}_{[1:N_s]}; \boldsymbol{\rho}^{(l)}\right) \boldsymbol{y}^{\mathrm{H}}(k),\end{aligned} \tag{15}$$

$$\begin{aligned}\mathbf{R}_{tt} &= \sum_{k=1}^{N_s} \sum_{\boldsymbol{t}_{ij}\in\mathcal{T}} \boldsymbol{t}_{ij}\boldsymbol{t}_{ij}^{\mathrm{H}}\eta_{\boldsymbol{\rho}^{(l)}}(k;i,j) \\ &= \sum_{k=1}^{N_s} \mathbb{E}\left(\boldsymbol{t}(k)\boldsymbol{t}^{\mathrm{H}}(k) \middle| \mathbf{Y}_{[1:N_s]}; \boldsymbol{\rho}^{(l)}\right).\end{aligned} \tag{16}$$

To limit the number of iterations one can set a conditions:

$$\frac{\|\boldsymbol{\rho}^{(l+1)} - \boldsymbol{\rho}^{(l)}\|}{\|\boldsymbol{\rho}^{(l)}\|} < \epsilon, \tag{17}$$

where ϵ is a small positive number chosen as threshold.

C. Data detection

Given the observation sequence and the estimated channel parameters at the end of the iterative process, denoted by $\mathbf{Y}_{[1:N_s]}$ and $\boldsymbol{\rho}^{(\mathrm{end})}$ respectively, an optimal criterion referred to as minimum symbol-error probability [12] can be easily integrated in the EM structure. The way to minimize symbol-error probability is to find the symbol that maximizes the posterior probability of transmitted symbol at each instant given the full observations, which can be formulated as:

$$\hat{u}_t(k) = \arg\max_{a\in\mathcal{A}} P\left(u_t(k) = a \middle| \mathbf{Y}_{[1:N_s]}; \boldsymbol{\rho}^{(\mathrm{end})}\right), \tag{18}$$

where $\mathcal{A}$ denotes the alphabet set of the considered PS-D^2-QAM, whereas $\hat{u}_t(k)$ stands for the estimated symbol from t-th source at instant k.

According to the posterior probability $\eta_{\boldsymbol{\rho}^{(\text{end})}}(k;i,j)$ of two successive state vectors ($\boldsymbol{s}(k) = \boldsymbol{b}_i, \boldsymbol{s}(k+1) = \boldsymbol{b}_j$) given the observations $\mathbf{Y}_{[1:N_s]}$, the probability term in (18) can be computed by:

$$P\left(u_t(k)=a_m \middle| \mathbf{Y}_{[1:N_s]}; \boldsymbol{\rho}^{(\text{end})}\right) = \sum_{i,j\in\mathcal{E}(m)} \eta_{\boldsymbol{\rho}^{(\text{end})}}(k;i,j), \quad (19)$$

where $\mathcal{E}(m)$ is the set of all transition vectors that satisfies $\boldsymbol{t}_{ij} = [u_1(k), ..., u_1(k-M), ..., u_t(k) = a_m, ..., u_t(k-M), ..., u_{N_t}(k), ..., u_{N_t}(k-M)]^{\mathrm{T}}$. Consequently, the EM-based algorithm, presented in this work, allows to achieve a joint blind channel estimation and data detection.

IV. Simulation results

In this section, we analyse the performance of the proposed blind EM-based algorithm in the case of a 2×2 convolutive MIMO system. The analysis consists of two parts.

The first part is aimed to demonstrate the channel estimation performance. To do so, a benchmark is set by using a fully-pilot-based channel estimation (i.e. all transmitted symbols are assumed to be known), as showed in [13]. The performance is evaluated through the Normalized Mean-Square Error (NMSE), expressed as:

$$\text{NMSE} = \frac{1}{N_0} \sum_{n=1}^{N_0} \|\hat{\boldsymbol{h}}_n - \boldsymbol{h}_{\text{real}}\|^2 / \|\boldsymbol{h}_{\text{real}}\|^2, \quad (20)$$

where $\hat{\boldsymbol{h}}_n$ denotes the estimated channel coefficients at the n-th run while $\boldsymbol{h}_{\text{real}}$ denotes the real (exact) channel coefficient vector. N_0 stands for the number of Monte-Carlo runs. The second part focuses on demonstrating the performance of data detection or source estimation. To highlight the performance gain of the proposed EM-based algorithm, a comparison with the HG-MMA [9] solution is conducted, in the case of both uniformly distributed QAM and PS-QAM. For channel coefficients, we take a channel model considered in optical communications context [14], which is given by:

$$\begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} = \sum_{n=0}^{M} \mu_n \begin{bmatrix} \cos\theta_n & e^{-\mathrm{j}\phi_n}\sin\theta_n \\ -e^{\mathrm{j}\phi_n}\sin\theta_n & \cos\theta_n \end{bmatrix} \begin{bmatrix} u_1(k-n) \\ u_2(k-n) \end{bmatrix} \quad (21)$$

where M is referred to as channel degree, μ_n are constants that control the intensity of inter-symbol interference (ISI), while θ_n and ϕ_n represent the azimuth and elevation rotation angles respectively, corresponding to the n-th channel's tap.

Simulation parameters are listed in Table II, which are applied to all simulations unless it is otherwise indicated. To avoid ill convergence of the considered algorithms (i.e. convergence to local extrema instead of the global one), we initialize the iterative process by a least squares (LS) channel estimate using N_p pilots.

Fig. 2 depicts the channel estimation performance of the proposed algorithm in terms of NMSE vs. SNR. The result

TABLE II: Common simulation parameters

Parameter	Notation & Value
Constellation type	PS-16-QAM
Entropy of source	$H = 3$, 3.5 or 4
Number of data symbols	$N_s = 500$
Number of pilot symbols for initialization	$N_p = 20$
Number of transmitters	$N_t = 2$
Number of receivers	$N_r = 2$
Order of channel	$M = 1$
Number of Monte-Carlo runs	$N_0 = 100$
Threshold to limit number of iterations	$\epsilon = 10^{-3}$

presented by h_{init} is obtained from a LS-based channel estimation by using N_p pilots only. The fully-pilot-based channel estimation $\mathrm{h}_{\text{full-pilot}}$ is obtained in the same way as h_{init}, but with a total number of pilots of $(N_s + N_p)$, which is considered as an upper limit of $\mathrm{h}_{\text{EM-B}}$. It can be observed that the proposed algorithm increases the estimation accuracy by minimising the NMSE, compared with that obtained after initialization (i.e. h_{init}). Moreover, the performance of the proposed blind EM-based solution reaches the upper limit in this case when SNR $\geqslant 15\,\text{dB}$.

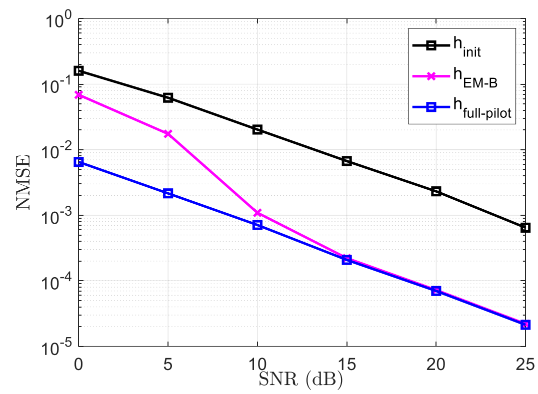

Fig. 2: NMSE vs. SNR for pilot initialization (h_{init}), blind EM-based algorithm ($\mathrm{h}_{\text{EM-B}}$) and fully-pilot-based estimation ($\mathrm{h}_{\text{full-pilot}}$), in the case of PS-16-QAM with $H = 3$

Fig. 3 shows the performance comparison. One can observe that, as expected, both algorithms converges well with similar performance when uniformly distributed 16-QAM is implement, but HG-MMA fails to converge with PS-16-QAM for entropy H=3.5 and H=3 (almost Gaussian case). Note that, for $H = 3$, the source kurtosis is approximately 2.9, very close to the Gaussian kurtosis 3. When entropy decreases, EM not only succeeds to estimate properly the channel but shows a better performance. The improvement is due to the change of constellation because the source is closer to QPSK.

V. Conclusion

The proposed solution is able to make full use of the *a priori* probability distribution of the considered PS constellation. Also, data detection is realized by re-utilizing the probabilities obtained within the EM framework, and based on the MAP. Simulation results show that the proposed solution provides

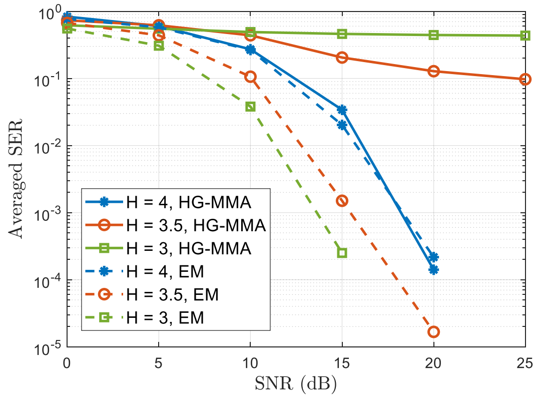

Fig. 3: Averaged SER vs. SNR for HG-MMA in the case of PS-16-QAM with varying entropy

an interesting channel estimation accuracy and a much better performance in terms of SER as compared to the HG-MMA solution ((HOS-based method)). Note that, in order to reduce the computational complexity, several simplified (approximated) alternatives are under study, which will be presented in future works.

References

[1] Han Sun, Mehdi Torbatian, Mehdi Karimi, Robert Maher, Sandy Thomson, Mohsen Tehrani, Yuliang Gao, Ales Kumpera, George Soliman, Aditya Kakkar, et al., "800G DSP ASIC design using probabilistic shaping and digital sub-carrier multiplexing," *Journal of Lightwave Technology*, vol. 38, no. 17, pp. 4744–4756, 2020.

[2] Shiwei Wang, Zijie Lu, Wei Li, Shi Jia, Lu Zhang, Mengyao Qiao, Xiaodan Pang, Nazar Idrees, Muhammad Saqlain, Xiang Gao, et al., "26.8-m THz wireless transmission of probabilistic shaping 16-QAM-OFDM signals," *APL Photonics*, vol. 5, no. 5, pp. 056105, 2020.

[3] Junho Cho and Peter J Winzer, "Probabilistic constellation shaping for optical fiber communications," *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1590–1607, 2019.

[4] Georg Böcherer, "Probabilistic signal shaping for bit-metric decoding," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 431–435.

[5] Georg Böcherer, Fabian Steiner, and Patrick Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Transactions on communications*, vol. 63, no. 12, pp. 4651–4665, 2015.

[6] André P des Rosiers and Paul H Siegel, "Effect of varying source kurtosis on the multimodulus algorithm," in *IEEE International Conference on Communications (Cat. No. 99CH36311)*. IEEE, 1999, vol. 2, pp. 1300–1304.

[7] Julian Renner, Tobias Fehenberger, Metodi P Yankov, Francesco Da Ros, Søren Forchhammer, Georg Böcherer, and Norbert Hanik, "Experimental comparison of probabilistic shaping methods for unrepeated fiber transmission," *Journal of Lightwave Technology*, vol. 35, no. 22, pp. 4871–4879, 2017.

[8] Fred Buchali, Fabian Steiner, Georg Böcherer, Laurent Schmalen, Patrick Schulte, and Wilfried Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," *Journal of Lightwave Technology*, vol. 34, no. 7, pp. 1599–1609, 2016.

[9] Syed Awais Wahab Shah, Karim Abed-Meraim, and Tareq Y Al-Naffouri, "Blind source separation algorithms using hyperbolic and givens rotations for high-order QAM constellations," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1802–1816, 2017.

[10] Ghassan Kawas Kaleh and Robert Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Transactions on Communications*, vol. 42, no. 7, pp. 2406–2413, 1994.

[11] Shun-Zheng Yu and Hisashi Kobayashi, "An efficient forward-backward algorithm for an explicit-duration hidden markov model," *IEEE signal processing letters*, vol. 10, no. 1, pp. 11–14, 2003.

[12] Edward A Lee and David G Messerschmitt, *Digital communication*, Springer Science & Business Media, 2012.

[13] Li Tao, Hui Tan, Chonghua Fang, and Nan Chi, "Volterra series based blind equalization for nonlinear distortions in short reach optical CAP system," *Optics Communications*, vol. 381, pp. 240–243, 2016.

[14] Seb J Savory, "Digital coherent optical receivers: Algorithms and subsystems," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 16, no. 5, pp. 1164–1179, 2010.