Multi-Layer Modulation for Intensity Modulated Direct Detection Optical OFDM

Rong Zhang and Lajos Hanzo
CSPC Group, School of ECS, University of Southampton, SO17 1BJ, UK

***Abstract*—A Multi-Layer Modulation (MLM) aided Intensity-Modulated Direct-Detection (IM/DD) DC-Biased Optical OFDM (DCO-OFDM) and Asymmetrically Clipped Optical OFDM (ACO-OFDM) are considered. More explicitly, we propose a Double Turbo Receiver (DTR) for jointly detecting the MLM and for compensating the clipping distortion. Additionally, a Genetic Algorithm (GA) aided weight optimisation is pursued for seeking an increased MLM Bits Per Symbol (BPS) throughput. Our numerical results demonstrate that for ACO-OFDM, at the throughput of $\eta = 3$ BPS, a 3dB gain was attained for $Q = 15$ DTR iterations without requiring weight optimisation. For DCO-OFDM, an even more significant gain of 12dB and 8dB was observed at the throughput of $\eta = 3$ BPS and $\eta = 4$ BPS, respectively without any clipping distortion compensation.**

I. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) has been invoked in diverse application areas, such as wired Asymmetric Digital Subscriber Lines (ADSL) [1], in wireless communications [2] as well as in optical communications [3]. The rationale of widely exploiting the OFDM technique is multi-fold and we refer to the fundamentals in [1], [3], [4]. Compared to the relatively mature wireless applications, OFDM has only recently been applied to optical communications [4]. The concept of Optical OFDM (OOFDM) brings about several benefits, since it is resilient to dispersion-induced impairments, hence potentially dispensing with the traditional optical pre- and post- compensation techniques [5]. More particularly, its high tolerance to both chromatic dispersion and to polarization mode dispersion extends the attainable distance to thousands of kilometers [6]. Additionally, OFDM is also applied in indoor optical wireless systems, in plastic optical fibres and in passive optical networks etc [4].

To elaborate a little further, in wireless OFDM systems the information is carried in the electronic domain using bipolar signalling, where coherent receivers may be used. By contrast, in Intensity-Modulated Direct-Detection (IM/DD) aided optical systems, the information is carried in the optical domain in terms of light intensity and it is unipolar, which constraints the modulating signal to positive real values [3][1]. Since the baseband OFDM signals are generally complex-valued and bipolar, some modifications are inevitable, when migrating OFDM to the unipolar optical domain. We focus on two classic types of IM/DD aided OOFDM systems, namely on DC-Biased Optical OFDM (DCO-OFDM) [8] and on Asymmetrically Clipped Optical OFDM (ACO-OFDM) [9]–[11]. In [12], ACO-OFDM and DCO-OFDM were compared to other classic optical modulation schemes, such as On-Off Keying (OOK) and Pulse-Position Modulation (PPM). When comparing these two schemes, it was shown in [13] that ACO-OFDM typically requires a lower average optical power for a given Bit Error Ratio (BER) and data rate than DCO-OFDM. When higher-order Phase Shift Keying (PSK) / Quadrature Amplitude Modulation (QAM) are considered, in [14], adaptive modulation was applied. In addition to conventional Fast Fourier Transform (FFT) and Inverse Fast Fourier Transform (IFFT)-based OOFDM, Discrete Hartley Transform (DHT) was investigated in [15].

In contrast to conventional modulation based OOFDM, in this paper, we conceived the Multi-Layer Modulation (MLM)-aided OOFDM. Inspired by the classic information theoretic concepts, in MLM, multiple modulated signals are first specifically weighted and then superimposed. As a result, in contrast to the lattice-structured conventional modulation constellation, the MLM constellation exhibits a Gaussian distribution. Hence, conventional threshold-based demodulation is no longer applicable and more complex probability-based iterative demodulation is pursued, Multi-Level-Coding (MLC) [16] constitutes a close relative of MLM, where the different overlaid layers are protected by different channel coding rates. MLM is also reminiscent of the Super-Position Coding (SPC) [17], where the different superimposed layers are transmitted at different power levels by assuming idealised capacity-achieving perfect channel codes. When viewing MLM as a multiplexing scheme, Code Division Multiplexing (CDM) [18] also bears some conceptual similarities. However, there are also a few differences. Firstly, multi-code transmissions require orthogonal spreading codes. Secondly, they do not have layer-specific weights as in MLM. In a nutshell, the

[1]We consider IM/DD aided OOFDM in this paper, while leaving its coherent OOFDM [7] counterpart for future work. Henceforth, we simply refer to IM/DD aided OOFDM as OOFDM.

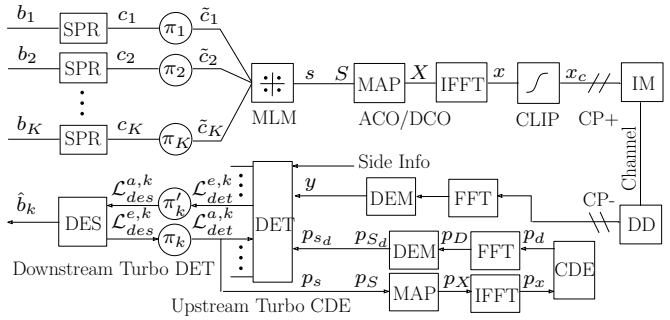

Fig. 1. Transceiver architecture with many-to-one type MLM constellation. Moreover, the acronyms used are SPR: spreading; π/π': interleaving / de-interleaving; MLM: multi-layer modulation; MAP: mapping; IFFT/FFT: inverse fast Fourier transform / fast Fourier transform; CP+/-: add / remove cyclic prefix; IM: intensity modulation; DD: direct detection; DEM: de-mapping; DET: detector; DES: de-spreading; CDE: clipping distortion estimation; ACO: asymmetrically clipped optical; DCO: DC-biased optical

advantages of MLM in optical systems are:

- theoretically, it can offer layer-dependent unequal protection as in MLC [19], [20] and its non-uniform constellation satisfies the near IM/DD channel capacity properties [21].
- practically, it is capable of offering a fine throughput versus robustness granularity, as in multi-code transmissions.

Against the above background, *we propose a new hybrid optical scheme based on MLM aided OOFDM. In addition to intrinsically amalgamating these two techniques, we conceive a powerful Double Turbo Receiver (DTR) architecture in the electronic domain, which jointly treats the multi-layer detection problem of MLM and the clipping distortion compensation problem of OOFDM. Importantly, we find the layer-specific weights that the MLM scheme should obey in the context of OOFDM with the aid of a Genetic Algorithm (GA) by investigating our DTR architecture in a semi-analytic performance evaluation approach.*

Our paper is organised as follows. In Section II, we provide a bird's eye view of the proposed hybrid transceiver. We then provide mathematical insights into the operation of our MLM-aided OOFDM in Section III. These discussions are followed by simulations in Section IV and conclusions in Section V.

II. Transceiver Architecture

A. General Description

As seen in the top left of the transceiver block diagram of Fig 1, the information bit-streams consist of K superimposed layers, where each layer is individually random spread. Without loss of generality, we consider the kth layer's information bit-stream of length L_b bits denoted as $\boldsymbol{b}_k = \{b_k(l), l = 1, \ldots, L_b\}$. After rate-$R_c$ spreading, the resultant chips of length $L_c = L_b/R_c$ in the kth layer are denoted as $\boldsymbol{c}_k = \{c_k(l), l = 1, \ldots, L_c\}$. The chip-stream is then interleaved by a layer-specific interleaver π_k, yielding $\tilde{\boldsymbol{c}}_k$. The resultant chip-streams of the K layers are superimposed according to the specifically designed MLM weighting pattern to be introduced in Section II-B, yielding the composite symbol-stream of length L_s as $\boldsymbol{s} = \{s(l), l = 1, \ldots, L_s\}$.

The MLM symbol-stream is then subjected to the specific mapping rule of the ACO-OFDM/DCO-OFDM schemes to be introduced in Section II-C and are then subsequently entered into the OOFDM block constituted by the Inverse Fast Fourier Transform (IFFT), Peak-Clipping (CLIP) and Cyclic Prefix insertion (CP+) operation of Fig 1. The resultant Time Domain (TD) OOFDM signal in electrical domain is then converted into optical domain (E/O) by IM and transmitted over the optical channel. Following the DD operation carried out at the receiver of Fig 1 for converting from optical domain back to electrical domain (O/E), the reverse operations are performed in the OOFDM block constituted by the CP removal (CP-) and Fast Fourier Transform (FFT). Finally, in order to recover the original K-layers' information bit-streams, the DTR architecture to be introduced in Section II-D is employed.

B. Multi-Layer Modulation

In a simple guise, MLM may be interpreted as a superposition of multiple specifically weighted spread layers. Mathematically, MLM may be written as

$$\boldsymbol{s} = \sum_{k=1}^{K} \rho_k f_m[\tilde{\boldsymbol{c}}_k], \tag{1}$$

where f_m denotes a basic modulation format, e.g. BPSK or QPSK. Moreover, ρ_k denotes the layer-specific weight detailed in [22], which will be optimised in Section III-B and plays a key role in the MLM design. In this paper, we investigate QPSK-based MLM in Section IV, but for the sake of conceptual simplicity, we only detail real-valued dimension of QPSK, namely BPSK, when we discuss our receiver algorithm in Section III. Finally, we define the resultant Bits Per Symbol (BPS) throughput of our MLM scheme as $\eta = 2KR_c$, where the multiplier 2 arises from the fact that mapping two BPSK-based streams to two orthogonal dimensions (real and imaginary) doubles the throughput without degrading the BER performance.

C. ACO-OFDM and DCO-OFDM

In IM/DD optics, real and positive signalling is required. In order to maintain real signalling, Hermitian symmetry is exploited by both the ACO and DCO. As for tackling the challenge of having no negative light intensity, ACO and DCO resort to different solutions.

1) ACO-OFDM: For ACO-OFDM, the MLM symbol-stream $\boldsymbol{s}$ of Eq (1) is serial-to-parallel converted and

mapped to the symbol matrix $\boldsymbol{S} \in \mathbb{C}^{N/4\times L_o}$, which is denoted as

$$\boldsymbol{S} = \{S(n,l), n = 1,\ldots,N/4, l = 1,\ldots,L_o\}, \quad (2)$$

where N represents the number of OOFDM sub-channels and L_o represents the number of OOFDM symbols. After ACO-OFDM mapping, the ACO-OFDM Frequency Domain (FD) symbol matrix $\boldsymbol{X} \in \mathbb{C}^{N\times L_o}$ is constituted by the entries of

$$X(n,l) = \begin{cases} S(n/2,l) & \text{if } n \le N/2 \text{ and is even} \\ S^*(\frac{N-n+2}{2},l) & \text{if } n > N/2 \text{ and is even} \\ 0 & \text{if } n \text{ is odd.} \end{cases} \quad (3)$$

It becomes plausible that the ACO-OFDM mapping obeys the Hermitian symmetry property, which allows us to create real-valued TD signal samples $\boldsymbol{x} \in \mathbb{R}^{NL_o\times 1}$ denoted as

$$\boldsymbol{x} = \{x[(l-1)N+m], m = 1,\ldots,N, l = 1,\ldots,L_o\}, \quad (4)$$

after classic IFFT operation. More explicitly, the odd-indexed FD sub-channels are set to zero such that the first-half of the TD signal samples are copied in the second-half of the TD signal samples, albeit with their signs flipped. As a result, the TD signal samples can be losslessly conveyed with all the negative parts clipped at zero. When the positive peaks are additionally clipped at the upper limit of $\boldsymbol{x}_{ul}$, the resultant clipped unipolar TD signal samples $\boldsymbol{x}_c \in \mathbb{R}^{NL_o\times 1}$ may be written as

$$\phi(\boldsymbol{x}): \quad \boldsymbol{x}_c = \begin{cases} \boldsymbol{x}_{ul} & \text{if } \boldsymbol{x} \ge \boldsymbol{x}_{ul} \\ \boldsymbol{x} & \text{if } 0 < \boldsymbol{x} < \boldsymbol{x}_{ul} \\ \boldsymbol{0} & \text{if } \boldsymbol{x} \le 0. \end{cases} \quad (5)$$

2) DCO-OFDM: For DCO-OFDM, the MLM symbol matrix $\boldsymbol{S} \in \mathbb{C}^{N/2\times L_o}$ is denoted by

$$\boldsymbol{S} = \{S(n,l), n = 1,\ldots,N/2-1, l = 1,\ldots,L_o\}, \quad (6)$$

and the resultant DCO-OFDM FD symbol matrix $\boldsymbol{X} \in \mathbb{C}^{N\times L_o}$ is constituted by the entries of

$$X(n,l) = \begin{cases} S(n-1,l) & \text{if } n \le N/2, n \ne 1 \\ S^*(N-n+1,l) & \text{if } n > N/2, n \ne N/2+1 \\ 0 & \text{if } n = 1, n = N/2+1, \end{cases} \quad (7)$$

This is then followed by the IFFT-related modulation to generate the TD signal samples. The transmitted TD signal samples $\boldsymbol{x} \in \mathbb{R}^{NL_o\times 1}$ are subsequently subject to a DC-bias of ρ_0, which ensures that the negative values can be avoided with a high probability. Following the convention of OOFDM, we quote the level of DC-bias as the power of ρ_0 relative to the power of TD signal samples $\boldsymbol{x}$ and express it in dB as $\rho_{0_{dB}} = 10\log_{10}(\rho_0^2/\mathbb{E}[x^2]+1)$. Since the DCO-OFDM TD signal samples obeying the Gaussian distribution exhibit a high Peak to Average Power Ratio (PAPR), it remains inevitable to have a low portion of negative TD components, which may be eliminated by clipping at zero. When combined with positive peak clipping at the upper limit of $\boldsymbol{x}_{ul}$, the resultant clipped unipolar TD signal samples $\boldsymbol{x}_c \in \mathbb{R}^{NL_o\times 1}$ may be written as

$$\phi(\boldsymbol{x}): \quad \boldsymbol{x}_c = \begin{cases} \boldsymbol{x}_{ul} & \text{if } \boldsymbol{x}+\rho_0 \ge \boldsymbol{x}_{ul} \\ \boldsymbol{x}+\rho_0 & \text{if } 0 < \boldsymbol{x}+\rho_0 < \boldsymbol{x}_{ul} \\ \boldsymbol{0} & \text{if } \boldsymbol{x}+\rho_0 \le 0. \end{cases} \quad (8)$$

3) Comparisons: When comparing between ACO-OFDM and DCO-OFDM, the net throughput per OOFDM symbol for DCO-OFDM is approximately twice of that in ACO-OFDM, when the number of sub-channels N is high. However, this doubled throughput is achieved by degrading the power efficiency of DCO-OFDM owing to imposing a DC-bias of ρ_0. Another difference is the half-wave symmetry of the ACO-OFDM TD signal ensures that the clipping-distortion imposed by the removal of the negative amplitudes only occurs at the odd-indexed FD sub-channels carrying no data and hence can be ignored, despite the amplitude of the ACO-OFDM TD signal samples is reduced by a factor of $\alpha = 1/2$ [2]. On the other hand, the clipping-distortion of the DCO-OFDM TD signal samples cannot be ignored and hence imposes an error floor, if no specific counter-measure is employed.

D. Double Turbo Receiver

Let us now continue by introducing a unified framework for the receiver of MLM-aided OOFDM in electronic domain. Theoretically, following the Bussgangs theorem [23], the OOFDM clipping process may be described as

$$\phi(\boldsymbol{x}) = \alpha\boldsymbol{x} + \boldsymbol{d}, \quad (9)$$

where $\boldsymbol{d}$ represents the distortion component that is uncorrelated with $\boldsymbol{x}$, while α represents the scalar penalty imposed by clipping, which is given by $\alpha = \mathbb{E}\left[\boldsymbol{x}^\dagger\phi(\boldsymbol{x})\right]/\mathbb{E}\left[||\boldsymbol{x}||^2\right]$. OOFDM is capable of combating Inter-Symbol-Interference (ISI), provided that we append a sufficiently long CP before optical modulation, which is removed at the receiver after applying DD and before applying FFT. As a result, by letting the E/O and O/E related factor be unity, the discrete-time model of the lth received OOFDM symbol after FFT may be written as

$$\boldsymbol{r}(l) = \boldsymbol{F}\boldsymbol{H}_t\boldsymbol{x}_c(l) + \boldsymbol{W}(l), \quad (10)$$

$$= \boldsymbol{F}\boldsymbol{H}_t[\alpha\boldsymbol{F}^\dagger\boldsymbol{X}(l) + \boldsymbol{d}(l)] + \boldsymbol{W}(l), \quad (11)$$

$$= \alpha\boldsymbol{H}_f\boldsymbol{X}(l) + \boldsymbol{H}_f\boldsymbol{D}(l) + \boldsymbol{W}(l), \quad (12)$$

where $\boldsymbol{x}_c(l) = [x_c[(l-1)N+1],\ldots,x_c(lN)]^{\mathrm{T}}$ and $\boldsymbol{d}(l) = [d[(l-1)N+1],\ldots,d(lN)]^{\mathrm{T}}$ represent the lth segment of TD signal samples and distortion samples, respectively, while $\boldsymbol{X}(l)$ and $\boldsymbol{D}(l)$ represent their corresponding FD counterparts. Furthermore, $\boldsymbol{H}_t$ is a circulant

[2]However, when a positive peak-clipping is applied to the signal capped at x_{ul}, clipping distortion will occur also in ACO-OFDM.

matrix denoting the TD channel response, while $\boldsymbol{H}_f$ is a diagonal matrix denoting the FD channel response constituted by the eigenvalues of $\boldsymbol{H}_t$ transformed by FFT matrix of $\boldsymbol{F}$ and IFFT matrix of $\boldsymbol{F}^\dagger$. Finally, $\boldsymbol{W}(l)$ represents the Additive White Gaussian Noise (AWGN). After DE-Mapping (DEM) of $\boldsymbol{r}(l)$, the resultant data-bearing received signal samples become

$$\boldsymbol{y}(l) = \alpha \boldsymbol{H}\boldsymbol{S}(l) + \boldsymbol{H}\boldsymbol{S}_d(l) + \boldsymbol{N}(l), \tag{13}$$

where DEM may be seen as the puncturing of $\boldsymbol{r}(l)$ at the indices of $\{2, 4, \ldots, N/2\}$ for ACO-OFDM and at the indices of $\{2, 3, \ldots, N/2\}$ for DCO-OFDM, yielding the pairs $\boldsymbol{H}_f \mapsto \boldsymbol{H}$, $\boldsymbol{X}(l) \mapsto \boldsymbol{S}(l)$, $\boldsymbol{D}(l) \mapsto \boldsymbol{S}_d(l)$, $\boldsymbol{W}(l) \mapsto \boldsymbol{N}(l)$. The samples $\{\boldsymbol{y}(l), l = 1, \ldots, L_o\}$ are then parallel-to-serial converted and for simplicity, we consider a particular entry of the resultant serial signal samples, which is expressed as

$$y = \alpha h s + h s_d + n. \tag{14}$$

Based on Eq (14), the DTR architecture seen in Fig 1 is employed, which includes both the downstream turbo iteration and the upstream turbo iteration, where we define downstream / upstream as the signal processing that occurs after / before DET in Fig 1. More explicitly,

- in downstream turbo iteration, the kth DETector (DET) generates the *extrinsic* information $\mathcal{L}_{det}^{e,k}$ for the interleaved chip-stream $\tilde{\boldsymbol{c}}_k$. After deinterleaving, this soft information is then used as *a priori* information of $\mathcal{L}_{des}^{a,k}$ for feeding the DE-Spreading (DES) block of Fig 1 to generate the updated *a priori* information $\mathcal{L}_{det}^{a,k}$ for further enhancing the *extrinsic* information of the interleaved chip-stream $\tilde{\boldsymbol{c}}_k$. These iterations are repeated a number of times in order to deliver the decisions based on the *a posteriori* information of the bit-stream $\boldsymbol{b}_k$.
- additionally, we also employ upstream iterations to suppress the clipping-distortion as detailed in [24]–[27]. During each downstream iteration, the updated *a priori* information $\mathcal{L}_{det}^{a,k}$ is entered into the Clipping Distortion Estimation (CDE) block of Fig 1 by producing both the estimates and the variance of the clipping distortion imposed on the transmitter. Then the reconstructed clipping distortion estimates are cancelled in the DET of Fig 1 before delivering the updated *extrinsic* information $\mathcal{L}_{det}^{e,k}$ of the chip-stream $\tilde{\boldsymbol{c}}_k$. Moreover, as seen in Fig 1, the statistics involved in each stage of the upstream iterations are denoted by $p_{\boldsymbol{s}_d}, p_{\boldsymbol{s}}, p_{\boldsymbol{S}_d}, p_{\boldsymbol{S}}, p_{\boldsymbol{d}}, p_{\boldsymbol{x}}, p_{\boldsymbol{D}}$ and $p_{\boldsymbol{X}}$.

The upstream turbo CDE assists the downstream turbo DET of Fig 1 to converge, hence reducing/removing the potential error floor.

III. Design and Optimisation

A. Mathematical Representation

The above conceptual discussion of the DTR block of Fig 1 in Section II-D may be neatly represented in mathematical form. Considering the model of Eq (14), the DET evaluates the marginal *a posteriori* probability of $\tilde{c}_k$ as

$$P(\tilde{c}_k|y) = \overbrace{p(y, \{\tilde{c}_i\}_{i\neq k}, s_d|\tilde{c}_k)}^{\text{likelihood}} \overbrace{P(\tilde{c}_k)}^{\text{probability}}, \tag{15}$$

$$= \underbrace{p(y|\{\tilde{c}_i\}_{i\neq k}, s_d, \tilde{c}_k) \prod_{i\neq k} P(\tilde{c}_i) P(s_d)}_{\text{extrinsic}} \overbrace{P(\tilde{c}_k)}^{\text{a priori}}, \tag{16}$$

where the *a posteriori* probability at the left hand side of Eq (15) may be decomposed based on Bayes' Theorem into the likelihood function given by the first multiplicative term and the probability term seen at the right hand side of Eq (15), where the likelihood function may be further decomposed according to Eq (16). Since the probabilities of $P(\{\tilde{c}_i\}_{i\neq k})$ and $P(s_d)$ are mutually independent, they may be represented in a product form, namely as $P(\{\tilde{c}_i\}_{i\neq k}, s_d) = \prod_{i\neq k} P(\tilde{c}_i) P(s_d)$. As a result, our downstream turbo DET of Fig 1 estimates the *extrinsic* information part of Eq (16), while our upstream turbo CNE estimates $P(s_d)$, given $P(\tilde{c}_k)$ and $P(\{\tilde{c}_i\}_{i\neq k})$.

An exhaustive evaluation of the *extrinsic* information imposes an excessive computational complexity, which requires the summing of all possible combinations of $P(\{\tilde{c}_i\}_{i\neq k})$ and $P(s_d)$. Hence we resort to an algorithm exhibiting a linear computational complexity, where we rewrite Eq (14) as

$$y = \alpha h_k \tilde{c}_k + \sum_{i\neq k} \alpha h_i \tilde{c}_i + h s_d + n, \tag{17}$$

where $h_k = h\rho_k$ denotes the effective channel, while the second and third term represent the multi-layer interference and clipping noise, respectively. The idea of our low complexity receiver is to ensure that after interference (I) cancellation, we arrive at the decontaminated received signal of

$$y - I = \alpha h_k \tilde{c}_k + \sum_{i\neq k} \alpha h_i (\tilde{c}_i - M_{\tilde{c}_i}) + h(s_d - M_{s_d}) + n, \tag{18}$$

where the multi-layer interference estimates of $M_{\tilde{c}_k}$ and the clipping distortion estimates of M_{s_d} are derived in Section III-A1 and in Section III-A2, respectively.

1) Downstream Turbo DET: The soft information exchanged in the downstream turbo DET is constituted by the Log-Likelihood Ratio (LLR) of $\tilde{c}_k$, where the *extrinsic* LLR delivered by the DET of chip $\tilde{c}_k$ is given by

$$\mathcal{L}_{det}^{e,k}(\tilde{c}_k) = \ln \frac{p(y - I|\tilde{c}_k = +1)}{p(y - I|\tilde{c}_k = -1)} \tag{19}$$

$$= 2\alpha h_k (y - I)/V, \tag{20}$$

where we implicitly approximate the interference components I as Gaussian in the derivation. In Eq (20),

the interference estimates I and the corresponding interference plus noise variance V may be constructed as

$$I = \sum_{i \neq k} \alpha h_i M_{\tilde{c}_i} + h M_{s_d}, \tag{21}$$

$$V = \sum_{i \neq k} |\alpha h_i|^2 V_{\tilde{c}_i} + |h|^2 V_{s_d} + \sigma^2. \tag{22}$$

Moreover, the soft chip estimate $M_{\tilde{c}_k}$ and its variance $V_{\tilde{c}_k}$ are determined by the *a priori* LLR provided by the DES and may be written as $M_{\tilde{c}_k} = \tanh[\mathcal{L}^{a,k}_{det}(\tilde{c}_k)]$ and $V_{\tilde{c}_k} = 1 - M^2_{\tilde{c}_k}$, respectively. Let us now discuss the expression of the clipping noise estimates M_{s_d} and its variance V_{s_d}.

2) Upstream Turbo CNE: Having collected all the length L_c soft chip estimates $M_{\tilde{\boldsymbol{c}}_k}$ and their variance $V_{\tilde{\boldsymbol{c}}_k}$, it becomes straightforward to obtain the soft estimates of the MLM symbols $M_{\boldsymbol{s}} = \sum_{k=1}^{K} \rho_k M_{\tilde{\boldsymbol{c}}_k}$ and their variance as $V_{\boldsymbol{s}} = \sum_{k=1}^{K} |\rho_k|^2 V_{\tilde{\boldsymbol{c}}_k}$. Collecting all length L_s soft estimates of the MLM symbols and mirroring the operations carried out at the transmitter, we can obtain the soft estimates (variance) of the MLM symbol matrix $M_{\boldsymbol{S}}$ ($V_{\boldsymbol{S}}$) and consequently the estimates of the OOFDM FD symbol matrix $M_{\boldsymbol{X}}$ ($V_{\boldsymbol{X}}$). Hence, we can generate the soft estimates of the OOFDM TD signal samples as $M_{\boldsymbol{x}} = \boldsymbol{F}^{\dagger} M_{\boldsymbol{X}}$ and their corresponding variance as $V_{\boldsymbol{x}} = \boldsymbol{I}_N V_{\boldsymbol{X}}/N$, where the availability of $M_{\boldsymbol{x}}$ and $V_{\boldsymbol{x}}$ enables us to generate the next estimates of $M_{\boldsymbol{d}}$ and $V_{\boldsymbol{d}}$.

Recall from Eq (9) that the Bussgangs theorem [23] ensures having uncorrelated variables $\boldsymbol{d}$ and $\boldsymbol{x}$, hence we consider a particular TD distortion sample, where the corresponding soft estimate and TD residual clipping noise variance are

$$M_d = \mathbb{E}[x_c - \alpha x] = \int [\phi(x) - \alpha x] p(x) dx, \tag{23}$$

$$V_d = \mathbb{E}[|d - M_d|^2] = \int |\phi(x) - \alpha x - M_d|^2 p(x) dx, \tag{24}$$

with $p(x)$ obeying the Gaussian distribution having a mean and variance of M_x and V_x, respectively. Hence we have $p(x) \sim \mathcal{N}(M_x, V_x)$. After collecting all TD distortion sample statistics and arranging them into $M_{\boldsymbol{d}}$ and $V_{\boldsymbol{d}}$ of dimension $(N \times L_o)$, we transform them back to the FD by applying $\boldsymbol{F}$. As a result, the soft estimates of the FD distortion $M_{\boldsymbol{D}}$ and of the FD residual clipping noise variance $V_{\boldsymbol{D}}$ are given by $M_{\boldsymbol{D}} = \boldsymbol{F} M_{\boldsymbol{d}}$ and $V_{\boldsymbol{D}} = \boldsymbol{I}_N V_{\boldsymbol{d}}/N$. This is then followed by the DEM process to get $M_{\boldsymbol{S}_d}$ and $V_{\boldsymbol{S}_d}$ for the data-bearing FD positions. Finally, they are parallel-to-serial converted to yield $M_{\boldsymbol{s}_d}$ and $V_{\boldsymbol{s}_d}$, as required by Eq (21) and Eq (22).

Remarks: This complexity of our downstream Turbo DET increases linearly with the number of layers stacked, as well as with the number of iterations. However, this is the lowest possible complexity one can claim, when detecting a Gaussian distributed signal, which is in contrast to the classic QAM/PSK modulation that has a structured constellation. Furthermore, the processing complexity is imposed in the electronic domain, where powerful baseband signal processing techniques may be invoked, instead of relying on optical-domain processing. Finally, the calculation of M_{s_d} and V_{s_d} is only needed, when clipping distortion compensation is used, while it can be excluded, if one does not want to take the effects of clipping noise into account. Hence, it is not an inherent complexity of our multi-layer interference cancellation algorithm.

Since the calculation of M_{s_d} and V_{s_d} are optional, we thus characterize the computational complexity in terms of the number of multiplications required by the receiver's downstream turbo DET. As a result, now we have

$$I = \alpha \sum_{i \neq k} h_i M_{\tilde{c}_i}, \tag{25}$$

$$V = \alpha^2 \sum_{i \neq k} |h_i|^2 V_{\tilde{c}_i} + \sigma^2. \tag{26}$$

Firstly, we may require K multiplications for evaluating $A = \{h_i M_{\tilde{c}_i}, i \in [1, K]\}$ and $3K$ multiplications for determining $B = \{(\mathcal{R}(h_i)^2 + \mathcal{I}(h_i)^2) V_{\tilde{c}_i}, i \in [1, K]\}$, where these records are shared across all K calculations of Eq (15) for all K layers. Hence, to evaluate Eq (25), we require $(K-1)$ multiplications within the summation, thus the effective number of multiplications is $(K-1)/K$ plus one additional multiplication with α. Similarly, to calculate Eq (26), the effective number of multiplications is $3(K-1)/K$ plus two additional multiplication associated with α^2. Regarding Eq (15), it needs 4 multiplications. In summary, the number of multiplications required per layer for the calculation of Eq (15) is $\zeta = 4 + (K-1)/K + 1 + 3(K-1)/K + 2 = 7 + 4(K-1)/K$. For $K >> 1$, we may quantify the computational complexity associated with the receiver's downstream turbo DET as $\zeta = 11$ per layer per iteration.

Algorithm 1 Pseudo-code of Double Turbo Receiver

Set: $\{M_{\boldsymbol{s}_d} = 0, V_{\boldsymbol{s}_d} = 1\}$, $\{M_{\tilde{\boldsymbol{c}}_k} = 0, V_{\tilde{\boldsymbol{c}}_k} = 1, k = 1, \ldots, K\}$
for $q = 1$ to Q **do**
  for $k = 1$ to K **do**
    Get: $\mathcal{L}^{e,k}_{det}(\tilde{\boldsymbol{c}}_k)$ using Eq (20), Eq (21), Eq (22)
    Update: $\mathcal{L}^{a,k}_{det}(\tilde{\boldsymbol{c}}_k)$ using
      $\mathcal{L}^{a,k}_{des}(\boldsymbol{c}_k) = \pi'_k\{\mathcal{L}^{e,k}_{det}(\tilde{\boldsymbol{c}}_k)\}$
      $\mathcal{L}^{e,k}_{des}(\boldsymbol{c}_k) = f\{\mathcal{L}^{a,k}_{des}(\boldsymbol{c}_k)\}$
      $\mathcal{L}^{a,k}_{det}(\tilde{\boldsymbol{c}}_k) = \pi_k\{\mathcal{L}^{e,k}_{des}(\boldsymbol{c}_k)\}$
  end for
  Get: $\{M_{\tilde{\boldsymbol{c}}_k}, V_{\tilde{\boldsymbol{c}}_k}\ k = 1, \ldots, K\} \rightarrow$
    $\{M_{\boldsymbol{s}}, V_{\boldsymbol{s}}\} \rightarrow \{M_{\boldsymbol{S}}, V_{\boldsymbol{S}}, M_{\boldsymbol{X}}, V_{\boldsymbol{X}}\} \rightarrow \{M_{\boldsymbol{x}}, V_{\boldsymbol{x}}\}$
  Get: $\{M_{\boldsymbol{d}}, V_{\boldsymbol{d}}\}$ using Eq (23), Eq (24) $\rightarrow$
    $\{M_{\boldsymbol{D}}, V_{\boldsymbol{D}}\} \rightarrow \{M_{\boldsymbol{S}_d}, V_{\boldsymbol{S}_d}\} \rightarrow \{M_{\boldsymbol{s}_d}, V_{\boldsymbol{s}_d}\}$
end for

$*$ π'_k and f denote de-interleaving and soft de-spreading operations.

B. Constellation Optimisation

Before embarking on finding the best possible weights with the aid of GA, we first discuss our performance evaluation method conceived for the DTR introduced in Section III-A, which will then be used in our optimisation efforts.

1) Performance Evaluation: Semi-analytic techniques dispense with bit-by-bit Monte Carlo simulations. Instead, they rely either on Signal-to-Noise-Ratio (SNR) tracking [28] or mutual information evolution [29]. In our DTR discussed above, we employ both techniques, since the DES of Fig 1 has an explicit input/output mutual information transfer rule. More explicitly, let us consider Eq (17), where the average SNR γ_k^q at the qth iteration of the kth layer may be written as

$$\gamma_k^q = \frac{|\alpha\rho_k|^2}{\sum_{i\neq k}|\alpha\rho_i|^2\mathbb{E}[V_{\tilde{c}_i}] + \mathbb{E}[V_{s_d}] + \sigma_e^2}, \tag{27}$$

where $\mathbb{E}[V_{\tilde{c}_k}]$ and $\mathbb{E}[V_{s_d}]$ represent the average residual interference variance $V_{\tilde{c}_k}$ of the kth layer and clipping noise variance V_{s_d}, respectively and $\sigma_e^2 = \sigma^2/\mathbb{E}[|h|^2]$.

The average SNR is the performance metric of the DET component, while we employ the mutual information as the metric of the DES component of Fig 1. The average SNR may be related to the *extrinsic* mutual information of the DET block of Fig 1 by $I_{det,k}^{e,q} = J(2\sqrt{\gamma_k^q})$, where $J(\cdot)$ is the Jacobian function detailed in [30]. For the DES component, there is a monotonic relationship between the *a priori* mutual information I_{des}^a and the *extrinsic* mutual information I_{des}^e. For the kth DES, it may be explicitly written as

$$I_{des,k}^{e,q} = J\left(\sqrt{(R_c^{-1}-1)J^{-1}[(I_{des,k}^{a,q})^2]}\right). \tag{28}$$

The *extrinsic* mutual information $I_{des,k}^{e,q}$ is then fed back as the *a priori* mutual information $I_{det,k}^{a,q}$ to the DET of Fig 1 during the $(q+1)$st iteration.

For quantifying the SNR γ_k^{q+1} experienced at the $(q+1)$th iteration, two transformation functions are defined for mapping the mutual information metric to the variance metric, namely

$$\mathbb{E}[V_{\tilde{c}_k}] = T_1(I_{det,k}^{a,q}); \tag{29}$$

$$\mathbb{E}[V_{s_d}] = T_2(I_{det,k}^{a,q}, \forall k) \tag{30}$$

Given a particular value of the DET's *a priori* mutual information $I_{det,k}^{a,q}$, we can generate the corresponding LLR samples of $\mathcal{L}_{det}^{a,k}(\tilde{c}_k)$ and consequently generate improved soft estimates of $M_{\tilde{\mathbf{c}}_k}$, which hence results in a reduced residual interference plus noise variance of $V_{\tilde{\mathbf{c}}_k}$. As a result, after generating the sample average of the residual interference plus noise variance, we arrive at the relationship of T_1. On the other hand, the relationship of T_2 is more involved, since it is a multi-parameter function, which depends both on the clipping function $\phi(\boldsymbol{x})$ as well as on the number of OOFDM sub-channels N. Exploiting the LLR samples of $\mathcal{L}_{det}^{a,k}(\tilde{\mathbf{c}}_k), k = 1,\ldots,K$ generated and mirroring the operations introduced in Section III-A2, we can establish the relationship of T_2.

Hence the SNR γ_k^{q+1} of the $(q+1)$th iteration becomes

$$\gamma_k^{q+1} = \frac{|\alpha\rho_k|^2}{\sum_{i\neq k}|\alpha\rho_i|^2 T_1(I_{det,i}^{a,q}) + T_2(I_{det,k}^{a,q}, \forall k) + \sigma_e^2}. \tag{31}$$

After Q iterations, the output *a posteriori* mutual information of the DES of Fig 1 is given by

$$I_{des,k}^{p,Q} = J\left(\sqrt{J^{-1}[(I_{des,k}^{a,Q})^2]/R_c}\right) \tag{32}$$

with the corresponding BER of $P_{b,k} = \mathcal{Q}[J^{-1}(I_{des,k}^{p,Q})/2]$.

2) Optimisation Procedure: Let us now search for the optimised weights so that given a particular design of $\{\rho_k, k = 1,\ldots,K\}$, the target BER P_b^* may be achieved at the output of the DES by all K layers after a predefined number of iterations Q. Since the off-line modelling of the multi-parameter function T_2 is rather complex, we exclude it from the optimisation procedure, which is achieved by treating the clipping noise as irreducible. However, we will execute upstream turbo CNE in our Monte-Carlo simulations in Section IV where needed [3]. Furthermore, we normalise the clipping noise plus AWGN components to unity during the optimisation and opt for using the objective function constituted by the minimisation of the total weights ρ_k, which is then formulated as

$$\min \sum_{k=1}^{K} |\rho_k|^2 \tag{33}$$

$$s.t. \quad P_{b,k} <= P_b^* \quad k = 1,\ldots,K \tag{34}$$

$$\rho_k > 0 \quad k = 1,\ldots,K. \tag{35}$$

Observe that the above optimisation formulation constitutes a lower-bounded minimisation problem subject to a non-linear constraint, which states that the BER of all K layers after Q iterations should be no worse than the target P_b^*. Since the convexity for the above problem is not guaranteed, using gradient-based algorithms may not be appropriate. Hence, we invoke a powerful randomly-guided bio-inspired GA [31].

The evolution process of the GA usually commences from a population of randomly generated legitimate solutions termed as *individuals*, which are subjected to genetic operations. In each consecutive GA generation, the fitness of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population based on their

[3] Strictly speaking, optimising without considering the offline modelling of T_2 results into optimising for a higher noise level, which leads to sub-optimum weights at the typical noise level. Although they are not optimum in the strict sense, they guarantee meeting the BER targets we set for the optimisation. Hence, when the actual noise level is no higher than the noise level used during the optimisation, our resultant weights must also guarantee meeting the BER targets, regardless, whether CNE is considered. In simple terms, the resultant weights meet the BER target even in the worst-case scenario.

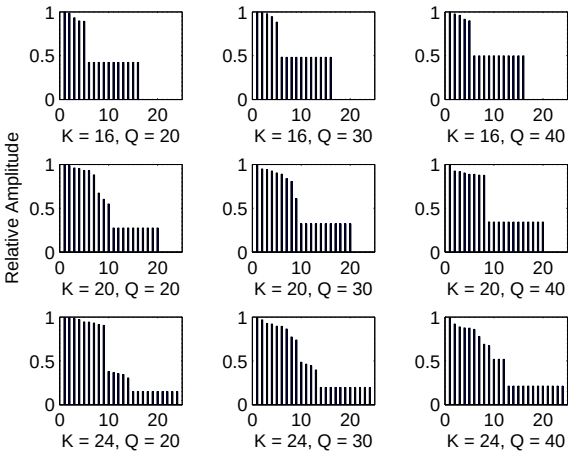

Fig. 2. Optimised amplitude of MLM for OOFDM for $K = 16, 20, 24$ layers associated with $Q = 20, 30, 40$ iterations and the spreading code rate of $R_c = 1/8$ at the target BER of $P_b^* = 10^{-4}$ in the context of relative amplitude pattern, where the amplitude of each layer was normalised to the maximum amplitude.

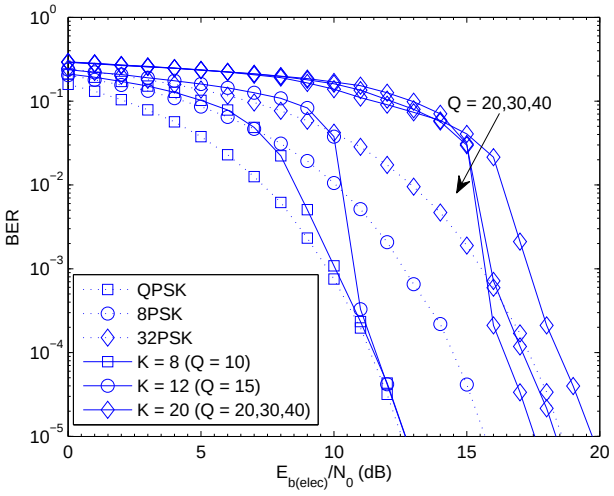

Fig. 3. BER performance comparison of MLM-aided ACO-OFDM to conventional PSK-aided ACO-OFDM for comparable BPS throughput values, namely $\eta = \{2, 3, 4, 5, 6\}$.

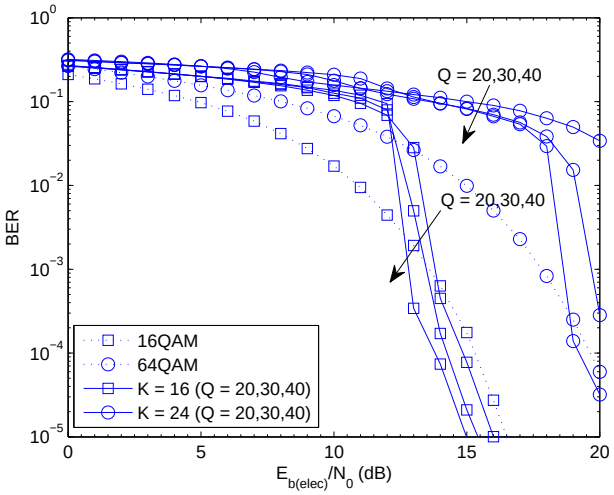

Fig. 4. BER performance comparison of MLM-aided ACO-OFDM to conventional QAM-aided ACO-OFDM for comparable BPS throughput values, namely $\eta = \{2, 3, 4, 5, 6\}$.

'quality' termed as *fitness*, and again modified using the *crossover*, *mutation* etc GA operations to form a new population. The new population is then used in the next iteration of the algorithm. Commonly, the algorithm terminates when either a maximum number of GA generations has been produced, or a satisfactory fitness level has been reached for the population. Despite GA's powerful capability, the potential drawback of GAs is that they may not find the global optimum. To provide further insights, in Table I, we include the optimised weights for $K = 16, 20, 24$ superimposed layers associated with $Q = 20, 30, 40$ iterations and for the spreading code rate of $R_c = 1/8$ at the target BER of $P_b^* = 10^{-4}$. Also in Fig 2, these are plotted in the context of relative weighting pattern as y-axis, where the weight of each layer was normalised to the maximum weight.

IV. Performance Evaluation

We now present numerical results for charactering our MLM-aided OOFDM versus both the electrical $E_{b(elec)}/N_0$ as well as versus the optical $E_{b(opt)}/N_0$ having $N = 128$ FD sub-channels. We investigate the classic AWGN channel so as to focus our attention on the modulation-related gain of our MLM-aided OOFDM scheme. Hence our results become free from the restrictions of a specific optical transmission medium and application, which was also the case considered in [12]. In our simulations, we set the spreading code rate to $R_c = 1/8$ and the number of superimposed layers to $K = \{8, 12, 16, 20, 24\}$. The associated BPS throughput then becomes $\eta = \{2, 3, 4, 5, 6\}$, corresponding to the throughput of QPSK, 8PSK, 16QAM, 32PSK and 64QAM in the context of conventional OOFDM, respectively. Additionally, when we have a BPS throughput of $\eta \geq 4$, weight optimisation is required. Finally, we impose no positive peak-clipping for ACO-OFDM and DCO-OFDM so as to compare our system to the benchmarkers characterized in [12], [15], [32]. Note that the clipping noise imposed by positive peak-clipping should be considered, when designing a practical system. In our simulations, we used random interleavers and sign-alternating spreading codes, such as $\{+1, -1, +1, -1, \dots\}$.

A. BER Performance versus $E_{b(elec)}/N_0$

1) ACO-OFDM: Fig 3 and Fig 4 compare the BER performance of MLM-aided ACO-OFDM to both PSK-aided ACO-OFDM and to QAM-aided ACO-OFDM for comparable BPS throughput values. Since no positive peak-clipping is imposed, no clipping noise affects the

TABLE I
GA OPTIMISED AMPLITUDE PATTERN FOR $K = 16, 20, 24$ WITH $Q = 20, 30, 40$ ITERATIONS AND SPREADING CODE RATE OF $R_c = 1/8$ AT TARGET BER OF $P_b^* = 10^{-4}$.

$K = 16$				$K = 20$				$K = 24$			
Q	20	30	40	Q	20	30	40	Q	20	30	40
1	0.3952	0.3735	0.3713	1	0.3496	0.3512	0.3526	1	0.3325	0.3457	0.3582
2	0.3883	0.3701	0.3620	2	0.3445	0.3335	0.3264	2	0.3302	0.3343	0.3297
3	0.3689	0.3655	0.3565	3	0.3355	0.3297	0.3234	3	0.3270	0.3204	0.3182
4	0.3545	0.3534	0.3401	4	0.3332	0.3248	0.3185	4	0.3230	0.3183	0.3139
5	0.3528	0.3294	0.3334	5	0.3256	0.3169	0.3130	5	0.3140	0.3094	0.3127
6	0.1672	0.1802	0.1853	6	0.3247	0.3121	0.3130	6	0.3133	0.3090	0.3078
7	0.1671	0.1799	0.1851	7	0.3080	0.2937	0.3094	7	0.3099	0.2987	0.2787
8	0.1671	0.1799	0.1851	8	0.2352	0.2830	0.3088	8	0.3043	0.2670	0.2472
9	0.1669	0.1799	0.1851	9	0.2105	0.2138	0.1211	9	0.3007	0.2553	0.2410
10 ... 16	0.1668	0.1799	0.1850	10	0.1919	0.1132	0.1210	10	0.1261	0.1677	0.1856
				11 ... 20	0.0961	0.1132	0.1210	11	0.1223	0.1600	0.1855
								12	0.1178	0.1535	0.1855
								13	0.1146	0.1373	0.0764
								14	0.1018	0.0683	0.0764
								15 ... 24	0.0499	0.0683	0.0764

data-bearing sub-channels, consequently no upstream turbo CNE is executed.

Fig 3 shows that the performance of MLM associated with $K = 8$ layers converges to that of QPSK after $Q = 10$ iterations, hence no performance improvement/degradation was experienced. However, when considering a BPS throughput of $\eta = 3$, the performance of MLM associated with $K = 12$ layers also converges to that of QPSK after $Q = 15$ iterations. Hence, this results in about a gain of 3dB, when compared to 8PSK at a BER of 10^{-5}. When a higher BPS throughput of $\eta = 5$ is considered, the MLM associated with $K = 20$ layers has to invoke $Q = 40$ iterations to achieve a gain of 1dB over 32PSK at a BER of 10^{-5}. Fig 4, when the BPS throughput is $\eta = 4$, the performance of MLM associated with $K = 16$ layers becomes superior to that of 16QAM for $Q \geq 20$ iterations. Finally, when the BPS throughput of $\eta = 6$ is considered, the MLM becomes only capable of marginally outperforming 64QAM for $Q = 40$ iterations.

These investigations imply that for ACO-OFDM, the performance gain of MLM designed for a BPS throughput of $\eta \geq 5$ is associated with a high complexity, while it may be deemed most attractive for $\eta = 3$, where a performance gain of 3dB was achieved at a complexity associated with $Q = 15$ without using any weight optimisation. Hence, as long as the baseband signal processing is capable of handling this increased complexity, a performance gain of 3dB becomes affordable.

2) DCO-OFDM: Fig 5 and Fig 6 compare the BER performance of MLM-aided DCO-OFDM to both conventional PSK-aided DCO-OFDM and to QAM-aided DCO-OFDM by arranging for a comparable BPS throughput and DC-bias of $\rho_{0_{dB}} = 7$dB. Although no positive peak-clipping is imposed, it still requires upstream turbo CNE when appropriate, which was experimentally found when $\eta \geq 5$.

Fig 5 shows that the performance of MLM associated with $K = 8$ and $K = 12$ layers converges to that of QPSK after $Q = 10$ and $Q = 15$ iterations, respectively, while no upstream turbo CNE is needed. On the other hand, 8PSK-aided DCO-OFDM results in an irreducible error floor at the BER level of 10^{-5}. Hence, when we have a BPS throughput of $\eta = 3$, the MLM-aided design achieves approximately 12dB gain at a BER of 10^{-5}. When a higher BPS throughput of $\eta = 5$ is required, the MLM associated with $K = 20$ layers has to invoke upstream turbo CNE. As a result, MLM is capable of significantly reducing the error floor compared to that of the identical-BPS 32PSK for $Q = 20, 30, 40$ iterations. Fig 6, when we have a BPS throughput of $\eta = 4$, the performance of MLM associated with $K = 16$ layers without invoking upstream turbo CNE results in no error floor, while the $\eta = 4$ BPS 16QAM exhibited an error floor. Furthermore, MLM with $K = 16$ layers achieved a 8dB and 9dB gain at a BER of 10^{-5} after $Q = 20$ and

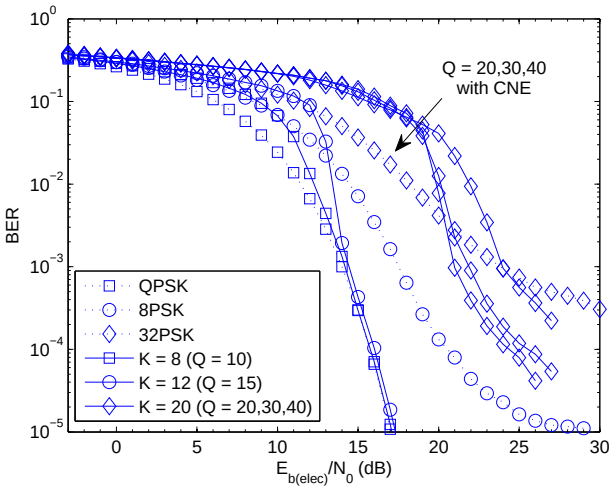

Fig. 5. BER performance comparison of MLM-aided DCO-OFDM to conventional PSK-aided DCO-OFDM by arranging for the BPS throughput to be $\eta = \{2, 3, 4, 5, 6\}$ and for a DC-bias of $\rho_{0_{dB}} = 7$dB.

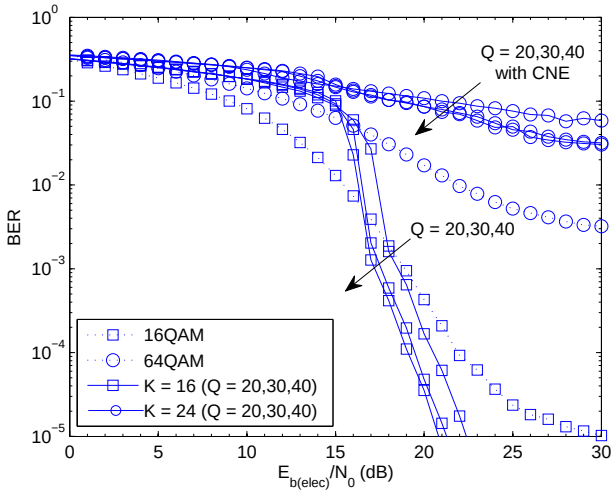

Fig. 6. BER performance comparison of MLM-aided DCO-OFDM to conventional QAM-aided DCO-OFDM by arranging for the BPS throughput to be $\eta = \{2, 3, 4, 5, 6\}$ and for a DC-bias of $\rho_{0_{dB}} = 7$dB.

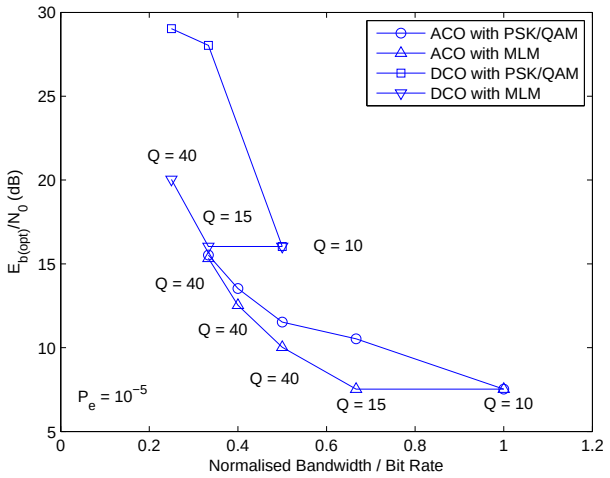

Fig. 7. Comparison of the required $E_{b(opt)}/N_0$ of MLM and conventional modulation for achieving BER of 10^{-5} for both ACO-OFDM and DCO-OFDM with DC-bias $\rho_{0_{dB}} = 7$dB.

$Q = 40$ iterations, respectively. Finally, when the BPS throughput of $\eta = 6$ is considered, the performance of MLM becomes worse than that of 64QAM even for $Q = 40$ iterations, despite using upstream turbo CNE.

These investigations imply that for DCO-OFDM and for a BPS throughput of $\eta \leq 5$, the MLM-aided design is preferred. Importantly, our MLM receiver is innately capable of incorporating upstream turbo CNE and for a BPS throughput of $\eta \leq 4$, a significant performance gain is achieved even without turbo CNE, which renders our design quite attractive.

B. Required $E_{b(opt)}/N_0$ versus Normalised Bandwidth

Let us now investigate our MLM-aided OOFDM system's required $E_{b(opt)}/N_0$ as a function of the normalised bandwidth per bit rate. Following [12], the normalized bandwidth per bit rate of DCO-OFDM is $(1 + 2/N)/\eta$, while that of ACO-OFDM is $2(1 + 2/N)/\eta$. On the other hand, the conversion between $E_{b(opt)}/N_0$ and $E_{b(elec)}/N_0$ depends on the distribution of the TD signal. For ACO-OFDM, we have $E_{b(opt)}/N_0 = E_{b(elec)}/\pi N_0$, while for DCO-OFDM, we have $E_{b(opt)}/N_0 = \rho^2 E_{b(elec)}/(\mathbb{E}[x^2] + \rho^2)N_0$.

Fig 7 compares the required $E_{b(opt)}/N_0$ of MLM and that of conventional modulation required for achieving a BER of 10^{-5} for both ACO-OFDM and DCO-OFDM with $\rho_{0_{dB}} = 7$dB. It shows that for ACO-OFDM, MLM generally requires a lower $E_{b(opt)}/N_0$ than conventional modulation for both target BERs considered and attains its highest gain at the normalised bandwidth per bit rate of $2/3$ with $Q = 15$ iterations, despite dispensing with weight optimisation. When DCO-OFDM is considered, the improvements of MLM over conventional modulation become more prominent, where MLM-aided DCO-OFDM is less prone to clipping distortion.

V. Conclusions

This paper introduced the MLM designed for IM/DD OOFDM, where both the concept of MLM and its related DTR algorithms were detailed. More importantly, a GA-aided optimum weighting pattern design was conceived with the aid of a semi-analytic tracking technique. Significant gains were demonstrated by comparing our MLM-aided design to both the conventional ACO-OFDM and DCO-OFDM designs both as a function of $E_{b(elec)}/N_0$ and that of $E_{b(opt)}/N_0$, where these have been achieved at the expense of an increased transceiver complexity. Finally, we will investigate more practical optical channels in our future work.

Appendix A

We now provide the pseudo-code of our GA-aided optimisation algorithm along with some remarks. Firstly, the GA-related parameters have to be carefully selected, since different configurations lead to slightly different results. Since the focus of our paper is not on global optimisation but on its results, we do not elaborate here further. Secondly, the loops controlled by N_{test} and N_{seed} are not compulsory, but they were experimentally found beneficial, where N_{test} repeats the GA-aided search several times, since a single run may get trapped in local optima, while N_{seed} assists in finding different paths in parallel so as to improve the optimisation efficiency. In GA parlance, N_{test} represents the *exploitation*, while N_{seed} the *exploration* operations.

Acknowledgment

The financial support of the RC-UK under the auspices of the UK-India ATC in Wireless Communications is gratefully acknowledged.

Algorithm 2 GA-aided Amplitude Pattern Optimisation

// Define Problem
Set: K, Q, R_c, P_b^*
// Optimisation Environment
Set Starting Point: $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_K]$
Set Objectives: Eq (33)
Set Constraints:
    Linear constraint = []
    Non-linear constraint (see: nonlfun)
    Bounds = Eq (35)
Set GA-related Parameters
// Optimisation Procedure
for $i = 1$ to N_{test} **do**
  for $j = 1$ to N_{seed} **do**
    // Kernel Solver
    $\boldsymbol{\rho}_j = \text{GA}(\boldsymbol{\rho}, \text{Objectives, Constraints, Parameters})$
  end for
  $\boldsymbol{\rho} = \min_j \sum_{k=1}^{K} |\rho_{j,k}|^2$
end for
Normalise $\boldsymbol{\rho} = \boldsymbol{\rho} / \sqrt{\sum_{k=1}^{K} |\rho_k|^2}$

// **nonlfun**
Set: $T_1(I_{det,k}^{a,0}) = 1$, $k = 1, \ldots, K$
for $q = 1$ to Q **do**
  for $k = 1$ to K **do**
    $\gamma_k^q = |\alpha\rho_k|^2 / (\sum_{i \neq k} |\alpha\rho_i|^2 T_1(I_{det,i}^{a,q-1}) + 1)$
    $\gamma_k^q \xrightarrow{J(2\sqrt{\gamma_k^q})} I_{det,k}^{e,q} \xrightarrow{\text{Eq (28)}} I_{des,k}^{e,q} \xrightarrow{\text{interpolate}} T_1(I_{det,k}^{a,q})$
  end for
end for
Get $\forall k$: $I_{des,k}^{p,Q}$ using Eq (32) and $P_{b,k} = \mathcal{Q}[J^{-1}(I_{des,k}^{p,Q})/2]$
Set non-linear constraints: $P_{b,k} \leq P_b^*$, $k = 1, \ldots, K$

REFERENCES

[1] L. Hanzo, M. Munster, B. J. Choi, and T. Keller, *OFDM and MC-CDMA for Broadcasting Multi-User Communications, WLANs and Broadcasting*. Wiley-IEEE Press, 2003.

[2] R. Zhang and L. Hanzo, "Wireless cellular networks," *IEEE Vehicular Technology Magazine*, vol. 5, pp. 31–39, 2010.

[3] J. Armstrong, "OFDM for optical communications," *IEEE/OSA Journal of Lightwave Technology*, vol. 27, pp. 189–204, 2009.

[4] W. Shieh and I. Djordjevic, *OFDM for Optical Communications*. Academic Press, 2009.

[5] B. J. C. Schmidt, A. J. Lowery, and J. Armstrong, "Experimental demonstrations of electronic dispersion compensation for long-haul transmission using direct-detection optical OFDM," *IEEE/OSA Journal of Lightwave Technology*, vol. 26, pp. 196–203, 2008.

[6] A. J. Lowery, L. B. Du, and J. Armstrong, "Performance of optical OFDM in ultralong-haul WDM lightwave systems," *IEEE/OSA Journal of Lightwave Technology*, vol. 25, pp. 131–138, 2008.

[7] W. Shieh and C. Athaudage, "Coherent optical orthogonal frequency division multiplexing," *IEE Electronics Letters*, vol. 42, pp. 587–589, 2006.

[8] J. B. Carruthers and J. M. Kahn, "Multiple-subcarrier modulation for nondirected wireless infrared communication," *IEEE Journal of Selected Areas of Communications*, vol. 14, pp. 538–546, 1996.

[9] J. Armstrong and A. J. Lowery, "Power efficient optical OFDM," *IET Electronics Letters*, vol. 42, p. 370–371, 2006.

[10] B. K. C. Liang and J. Evans, "Diversity combining for asymmetrically clipped optical OFDM in IM/DD channels," in *Proc. of IEEE GLOBECOM OWC workshop*, Honolulu, USA, Dec.1-4, 2009.

[11] A. D. K. Asadzadeh and S. Hranilovic, "Receiver design for asymmetrically clipped optical OFDM," in *Proc. of IEEE GLOBECOM OWC workshop*, Houston, USA, Dec.5-9, 2011.

[12] J. Armstrong and B. J. C. Schmidt, "Comparison of asymmetrically clipped optical OFDM and DC-biased optical OFDM in AWGN," *IEEE Communications Letters*, vol. 12, pp. 343–345, 2008.

[13] R. Mesleh, H. Elgala, and H. Haas, "On the performance of different OFDM based optical wireless communication systems," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 3, pp. 620–628, 2011.

[14] J. Grubor and K.-D. Langer, "Efficient signal processing in OFDM-based indoor optical wireless links," *Journal of Networks*, vol. 5, pp. 197–211, 2010.

[15] M. Moreolo, R. Muoz, and G. Junyent, "Novel power efficient optical OFDM based on Hartley transform for intensity-modulated direct-detection systems," *IEEE/OSA Journal of Lightwave Technology*, vol. 28, pp. 798–805, 2010.

[16] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Transactions on Information Theory*, vol. 23, pp. 371–377, 1977.

[17] X. Ma and P. Li, "Coded modulation using superimposed binary codes," *IEEE Transactions on Information Theory*, vol. 50, pp. 3331–3343, 2004.

[18] K. Kitayama, "Code division multiplexing lightwave networks based upon optical code conversion," *IEEE Journal of Selected Areas of Communications*, vol. 16, pp. 1309–1319, 1998.

[19] I. B. Djordjevic and B. Vasic, "Multilevel coding in m-ary DPSK/differential QAM high-speed optical transmission with direct detection," *IEEE/OSA Journal of Lightwave Technology*, vol. 24, pp. 420–428, 2006.

[20] X. Zhou and J. Yu, "Multi-level, multi-dimensional coding for high-speed and high-spectral-efficiency optical transmission," *IEEE/OSA Journal of Lightwave Technology*, vol. 27, pp. 3641–3653, 2009.

[21] A. A. Farid and S. Hranilovic, "Capacity bounds for wireless optical intensity channels with gaussian noise," *IEEE Transactions on Information Theory*, vol. 56, pp. 6066–6077, 2010.

[22] R. Zhang and L. Hanzo, "A unified treatment of power-efficient superposition coding aided communications: Theory and practice," *IEEE Communications Surveys and Tutorials*, vol. 13, pp. 503–520, 2011.

[23] D. Dardari, V. Tralli, and A. Vaccari, "A theoretical characterization of nonlinear distortion effects in OFDM systems," *IEEE Transactions on Communications*, vol. 48, pp. 1755–1764, 2000.

[24] D. Kim and G. L. Stüber, "Clipping noise mitigation for OFDM by decision-aided reconstruction," *IEEE Communications Letters*, vol. 3, pp. 4–6, 1999.

[25] H. Chen and A. M. Haimovich, "Iterative estimation and cancellation of clipping noise for OFDM signals," *IEEE Communications Letters*, vol. 7, p. 305–307, 2003.

[26] H. Nikopour, A. K. Khandani, and S. H. Jamali, "Turbo-coded OFDM transmission over a nonlinear channel," *IEEE Transactions on Vehicular Technology*, vol. 54, pp. 1361–1371, 2005.

[27] F. Peng and W. E. Ryan, "MLSD bounds and receiver design for clipped OFDM channels," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 3568–3578, 2008.

[28] J. Boutros and G. Caire, "Iterative multiuser decoding: unified framework and asymptotic performance analysis," *IEEE Transactions on Information Theory*, vol. 48, pp. 1772–1793, 2002.

[29] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Transactions on Communications*, vol. 49, pp. 1727–1737, 2001.

[30] G. K. S. ten Brink and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Transactions on Communications*, vol. 52, pp. 670–678, 2004.

[31] J. C. MacKay, *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.

[32] J. M. Tang and K. A. Shore, "Maximizing the transmission performance of adaptively modulated optical OFDM signals in

multimode-fiber links by optimizing analog-to-digital converters," *IEEE/OSA Journal of Lightwave Technology*, vol. 25, pp. 787–798, 2006.