Entropy Allocation Optimization for PS-OFDM With Constellation Partitioning Based Modeling

Rui Zhang, You-Wei Chen, Wenlong Mou, Shuang Yao, and Gee-Kung Chang, *Fellow, IEEE, Fellow, OSA*

***Abstract*—We propose and experimentally demonstrate an entropy (bits/symbol) allocation scheme for orthogonal frequency division multiplexing (OFDM) signal in a millimeter wave (mmW) over optical fiber fronthaul link. Under the condition of additive white Gaussian noise (AWGN) channel, a closed-form expression between the symbol error rate (SER) of probabilistically shaped quadrature amplitude modulation (QAM) signal and signal-to-noise ratio (SNR) is derived. With a fixed averaged target entropy of the overall signal frame, we use a projected mirror descent algorithm to maximize the averaged SER of the channels based on the pre-measured SNR information. The gradient information is estimated by finite difference using the modeling result. Experimental results in a fiber-mmW converged system validate the proposed method, which allocates more entropy for higher SNR regions compared to a simple scheme that allocates the entropies proportionally based on Shannon's formula. It yields a smoother NGMI and can meet the 0.86 NGMI threshold over the operation frequencies, while uniform QAM signal and Shannon formula based algorithm fails the threshold requirement. Up to 1.8-dB and 0.6-dB received optical power (ROP) gain are achieved as compared with uniform QAM signal and signal with Shannon formula-based allocation, respectively. Compared with the uniform signal, up to 0.7-bits/symbol GMI improvement is also demonstrated.**

***Index Terms*—Entropy allocation optimization, fiber-millimeter transmission, probabilistic shaping.**

I. Introduction

As THE data requirement of optical network grows due to the emerging applications of 8K video, Internet of things (IoT), and vehicle-to-everything communication, fiber-wireless integrated fronthaul network are expected to fully exploit the spectral resources and maximize the capacity of the physical layer infrastructure [1], [2]. The adoption of OFDM modulation format provides high spectral efficiency. Millimeter-wave (mmW) transmission can accommodate more available bandwidth [3], [4] in the last-mile wireless delivery. However, mmW has a high attenuation factor and is sensitive to uneven frequency responses induced by frequency selective fading or local oscillator leakage [4], [5]. Therefore, increasing the link reliability and dealing with the uneven channel frequency response are critical for fiber-mmW converged links.

On the other hand, probabilistic shaping (PS) has attracted lots of research attentions in fronthaul applications. PS signal with the Maxwell-Boltzmann distribution shows enhanced tolerance to ASE noise with increased Euclidean distance in an averaged power limited system [6]. Therefore, it has been applied in the single carrier mmW link to enhance the channel capacity or the receiver sensitivity [4], [7]. Some reported works apply a uniform entropy (bits/symbol) distribution to the whole OFDM frame with a performance improvement [8]–[10]. However, part of the subcarriers in an OFDM signal may be degraded or distorted significantly due to an uneven frequency response channel. To solve the issue, PS provides the flexibility of tuning entropy value arbitrarily by adjusting the probabilities of the transmitted symbols [11]–[13]. Per-subcarrier based PS entropy allocation has been employed via the pre-measured signal to noise ratio (SNR) of each sub-block to maximize the channel capacity or enhance the received performance in various links (e.g., visible light communication, optical fiber, mmW) [13], [14]. Compared with the conventional bit-loading scheme tunes the entropy values by integers, PS outperforms by enabling a continuous entropy tuning. However, per-subcarrier tuning scheme shows increased complexity and every subcarrier requires sufficient long frame length to ensure low rate loss. Therefore, a simple Shannon-formula-based PS-unit allocation scheme, which allocates the entropies proportionally based on Shannon's formula with a fixed transmission data rate, has been reported [15], [16]. By grouping several subcarriers into one PS unit, the processing complexity and the required frame duration can both be reduced. However, the Shannon-Hartley theorem cannot describe the practical system capacity perfectly since forward error correction (FEC) is not ideal in practice [17]. In other words, the entropy allocation using the Shannon formula is not optimized when we consider the non-ideal FEC. Moreover, when the target information rate is not equal to the channel capacity, Shannon formula does not give guidance on the optimal allocation scheme to minimize BER with a given information rate. In particular, simply apportioning the entropy proportionally based on Shannon capacity is not necessarily optimal.

In this paper, we propose the entropy per PS unit allocation based on the optimization of pre-FEC metrics. In the analysis,

Manuscript received March 17, 2020; revised May 16, 2020 and June 28, 2020; accepted July 9, 2020. Date of publication July 13, 2020; date of current version November 1, 2020. This work was supported by an NSF Grant Award 1821819 to Industry/University Cooperative Resrearch Center for Fiber Wireless Integration and Networking at Georgia Institute of Technology. *(Corresponding author: Rui Zhang.)*

Rui Zhang, You-Wei Chen, Shuang Yao, and Gee-Kung Chang are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30308 USA (e-mail: ruizhangece@gatech.edu; yuwei.chen@ece.gatech.edu; syao65@gatech.edu; gkchang@ece.gatech.edu).

Wenlong Mou is with the department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: wmou@eecs.berkeley.edu).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2020.3008772

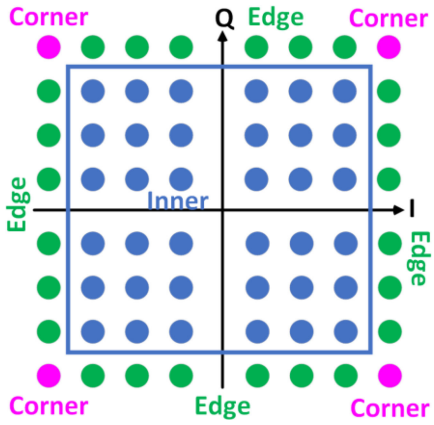

Fig. 1. The 64QAM constellation diagram of the inner, corner and edge symbols in modeling.

the Maxwell-Boltzmann distribution, which is the optimal distribution under AWGN [18], is assumed. The model gives a closed-form expression of the SER versus SNR, which guides the allocation scheme. The modeling procedure applies to different detection schemes such as maximum likelihood (ML) detector and maximum a posteriori (MAP) detector. We reformulate the PS-unit based entropy allocation scheme to an optimization problem with pre-measured SNR, while the constraint that the averaged entropy and spectral efficiency of the whole data frame remain unchanged. A projected mirror descent algorithm is utilized to maximizes the averaged SER (thus Q factor) among PS units. We also conduct a proof-of-concept experiment in a fiber-mmW integrated fronthaul link to validate the proposed model and algorithm.

This paper is organized as follows. Section II describes the principle of operation, including the channel modeling of the PS signal and the optimization scheme based on the projected mirror descent algorithm. Section III presents the experimental setup of the fiber-mmW integrated fronthaul while Section IV compares the experimental results with different allocation algorithms. Finally, Section V gives concluding remarks.

II. Principle of Operation

This section presents the channel modeling of the PS QAM signal and the entropy (bits/symbol) optimization scheme. Note that in our analysis, an ML detector, which makes the decision based on the minimum distance, is assumed. The modeling can be further extended to a MAP detector, which gives better performance but a higher complexity [19].

A. Channel Modeling Based on Constellation Partitioning

Firstly, an M-QAM signal is divided into three parts: inner symbols, edge symbols and corner symbols. Fig. 1 presents an example of the division scheme for a 64-QAM signal. Assume the amplitude of the M-QAM signal is a_I and a_Q ($\sqrt{\mathrm{M}}$ PAM alphabet) for I and Q part respectively, we can write the signal with AWGN as [17]:

$$\mathrm{r} = \sqrt{E_g} \cdot \begin{bmatrix} a_I \\ a_Q \end{bmatrix} + \begin{bmatrix} n_I \\ n_Q \end{bmatrix}, \tag{1}$$

where $n_I$and n_Q are i.i.d. $N(0, \mathrm{N}_0/2)$, and E_g is the energy of the received pulse. The statistics of the noise complex envelope shows white Gaussian noise with PSD $\mathrm{N}_0/2$. Then we can calculate symbol error rate (SER) as a function of E_g and N_0.

Assume the transmitted signal obeys Maxwell-Boltzmann distribution in the PS scenario, the SER can be represented as the weighted sum of error probabilities of different parts:

$$\begin{aligned}
\mathrm{P}_e &= 4 \cdot \frac{\exp\left(-2\lambda\left(\sqrt{M}-1\right)^2\right)}{Z} \cdot P\left(wrong|corner\right) \\
&+ \left(8 \cdot \sum_{i=1}^{\sqrt{M}/2-1} \left(\frac{\exp\left(-\lambda\left(\left(\sqrt{M}-1\right)^2+(2i-1)^2\right)\right)}{Z}\right)\right) \\
&\cdot P\left(wrong|edge\right) \\
&+ \left(4 \cdot \sum_{j=1}^{\sqrt{M}/2-1} \sum_{i=1}^{\sqrt{M}/2-1} \left(\frac{\exp\left(-\lambda\left((2j-1)^2+(2i-1)^2\right)\right)}{Z}\right)\right) \\
&\cdot P\left(wrong|inner\right) \\
&= \alpha_1 \cdot P\left(wrong/corner\right) + \alpha_2 \cdot P\left(wrong/edge\right) \\
&\quad + \alpha_3 \cdot P\left(wrong|inner\right),
\end{aligned} \tag{2}$$

where the weight coefficients are functions of λ and correspond to the probabilities of each part. $P(wrong|corner)$, $P(wrong|edge)$ and $P(wrong|inner)$ represent the error probabilities of corner symbols, edge symbols and inner symbols, which can be obtained as a function of $Q(\sqrt{2E_g/N_0})$ as shown in the appendix. Z is the normalization term: $4 \cdot \sum_{i=1}^{\sqrt{M}/2} \sum_{j=1}^{\sqrt{M}/2} (\exp(-\lambda((2j-1)^2+(2i-1)^2)))$.

Moreover, $Q(\sqrt{2\mathrm{E}_g/\mathrm{N}_0})$ can be reprenseted with signal SNR and the mean value of signal power. The energy per bit can be written as [17]:

$$E_b = \frac{\text{Energy per pulse}}{\text{entropy}} = \frac{\mathbb{E}\left(a_I^2 + a_Q^2\right)}{H} E_g, \tag{3}$$

where *H* is the signal entropy and mean value is

$$\begin{aligned}
&\mathbb{E}\left(a_I^2 + a_Q^2\right) \\
&= \sum_{i=1}^{\sqrt{M}/2} \sum_{j=1}^{\sqrt{M}/2} \left(\frac{\exp\left(-\lambda\left((2j-1)^2+(2i-1)^2\right)\right)}{Z} \right. \\
&\qquad \left. \cdot \left((2j-1)^2+(2i-1)^2\right)\right).
\end{aligned} \tag{4}$$

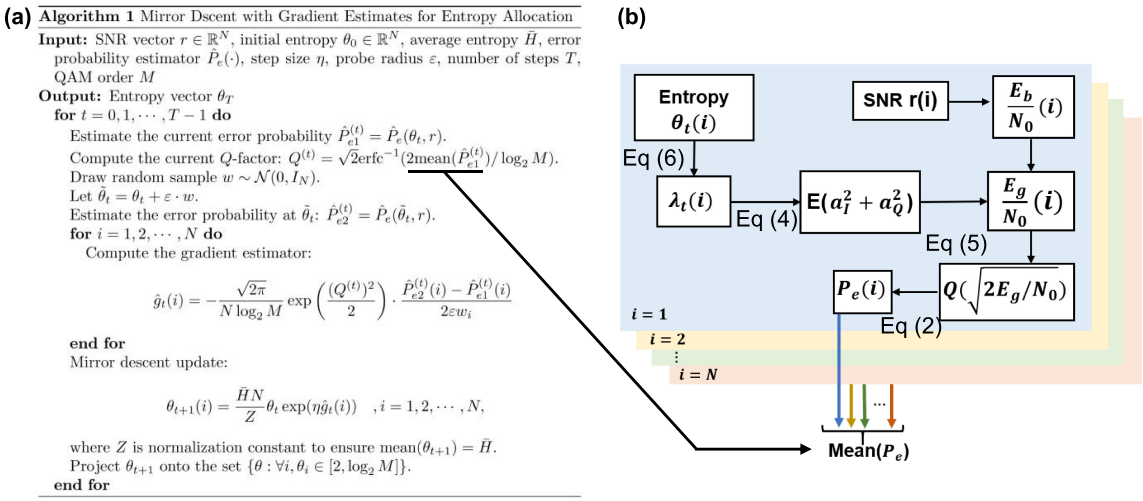

Fig. 2. Algorithm block diagram. (a) Modeling based algorithm using mirror descent with gradient estimates for entropy allocation. (b) Chart flow for calculating P_e given SNR and entropy.

With $E_b/N_0 = SNR/H$, the Q function can be written as:

$$Q\left(\sqrt{2E_g/N_0}\right) = Q\left(\sqrt{2E_b \cdot H \Big/ \left(N_0 \cdot \mathbb{E}\left(a_I^2 + a_Q^2\right)\right)}\right)$$
$$= Q\left(\sqrt{\frac{2E_b}{N_0} \cdot \frac{H}{\mathbb{E}\left(a_I^2 + a_Q^2\right)}}\right) = Q\left(\sqrt{\frac{2SNR}{\mathbb{E}\left(a_I^2 + a_Q^2\right)}}\right). \quad (5)$$

Combining Eq. (2) with Eq. (5), we obtain P_e as a function of SNR, λ, and modulation order *M*. The bit error rate (BER) can be approximated as $P_e/\log_2 M$ with gray mapping [17]. Note that λ can also be represented with signal entropy:

$$\mathrm{H} = 4 \cdot \sum_{i=1}^{\sqrt{M}/2} \sum_{j=1}^{\sqrt{M}/2} \frac{\exp\left(-\lambda\left((2j-1)^2 + (2i-1)^2\right)\right)}{Z}. \quad (6)$$

Finally, we can get Q factor as a function of SNR, signal entropy and QAM order M with $Q^2(dB) = 20log_{10}(\sqrt{2}erfc^{-1}(2BER))$ [20].

B. Entropy Optimization Design

Fig. 2(a) illustrates the entropy optimization algorithm based on the model that we built up in the previous section. Most of the commercial wireless systems operate with multicarrier waveform, i.e., OFDM, in which each subcarrier occupies small bandwidth. Therefore, each subcarrier can be approximated by AWGN channel due to smaller bandwidth even if the received signal suffers from non-uniform frequency responses. On the other hand, the operating ROP point of practical system (nearby the pre-FEC thresholds) is dominated by ASE noise. This implies that the model built with AWGN assumption can still work. Assume that there are *N* PS units (each PS unit is comprised of few subcarriers), we can obtain the pre-measured channel SNR vector$r \in R^N$ for every PS unit. One constraint of the optimization algorithm is that average entropy of the whole signal frame equals target entropy $\bar{H}$. Given *r*, $\bar{H}$, the initial entropy vector θ_0, step size η, probe radius ε and number of steps *T*, we can compute the allocated entropy step by step, as shown in Fig. 2.

We use a projected mirror descent algorithm for maximizing the Q factor (thus the averaged SER) among the PS units, where the gradient information is estimated by taking finite difference for function value information using the modeling result. Note that we need to estimate the gradient of the Q-factor with respect to entropy of each PS unit. This can be implemented via chain rule: we can compute $\partial Q/\partial P_e(i)$ analytically, and compute $\partial P_e(i)/\partial\theta_i$ numerically by finite difference. Multiplying them together, we obtain the gradient estimator $\hat{g}$ in the Algorithm described by Fig. 2(a). Mirror descent is a variant of gradient descent which adapts to the geometry of the problem and enforces the constraint conveniently, with possibly faster convergence rate. (See, [21], Chapter 3). In general, for any $x, y \in R^n$, the algorithm uses a convex mirror map function Φ to define the Bregman divergence $D_\Phi(y, x) = \Phi(y) - \Phi(\mathrm{x}) - \langle y - x, \nabla\Phi(x)\rangle$. Suppose we are given a step size η and a gradient estimator g. Let x^t be the current iterate and x^{t+1} be the next iterate, the algorithm takes the update:

$$x^{t+1} = \arg\min_y \eta\langle g(x^t), y - x^t\rangle + D_\Phi(y, x^t), \quad (7)$$

where $\langle g(x^t), y - x^t\rangle$ denotes the inner product between the vectors $g(x^t)$ and $y - x^t$.

In order to enforce the constraint that the average entropy $\bar{H}$ over channels is fixed, we choose the mirror map function to be $\Phi(x) = \sum_{i=1}^n \frac{x_i}{N\bar{H}} \log \frac{x_i}{N\bar{H}}$. In this case, the function $D_\Phi(x, y)$ is the Kullback-Leibler (KL) divergence between $x/N\bar{H}$ and $y/N\bar{H}$, the normalized entropy vectors seen as probability distribution over N coordinates. The one-step update defined above can be written in closed form as exponential weight update, which is shown in Algorithm 1.

Define the norm $\|x\|_1 = \sum_i |x_i|$. It is known (See [21]) that within a domain of diameter R under the $\|\cdot\|_1$ metric, for a convex functions f such that: $\max_i |\partial f/\partial x_i(x)| \le L$, the minimization of f converges to the optimal solution with

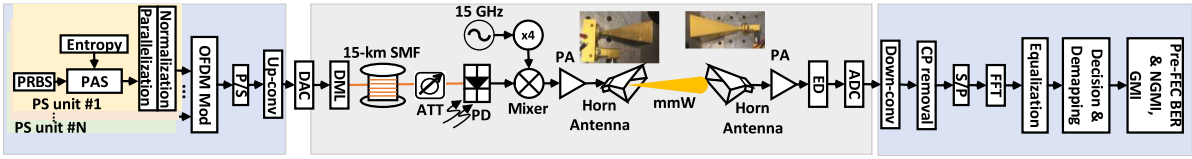

Fig. 3. Experimental setup and DSP diagram. P/S: Parallel to Serial converter; DAC: digital to analog converter, DML: Directly Modulated laser, PA: power amplifier, ATT: attenuator, PD: photodetector, ED: envelope detector, ADC: analog to digital converter, S/P: Serial to Parallel converter.

rate $RL/\sqrt{T}$. Though the objective function in the entropy allocation problem is generally non-convex, it is locally convex in a region around a local minimum. Thus, an initialization point acquired by the heuristics-based allocation method satisfies the convergence rate in the local region. For a system with *N* PS units, the diameter R under the $\|\cdot\|_1$ metric scales linearly with N, as entropies are bounded. Note that in our model, the one-dimensional mapping from SER to the Q factor and the mapping from entropy to SER both have bounded derivatives in a bounded region, the Lipschitz constant L scales with $1/N$, due to the averaging of SER per PS unit. Overeall, the algorithm has a rate of convergence independent of number of PS units. This is made possible by the mirror descent algorithm which exploits the structure of the objective function and the constraint set. Finally, the mirror descent algorithm directly enforces the average entropy constraint that the averaged entropy of *N* PS units equals to $\bar{H}$. In each iteration of Algorithm 1, we perform a mirror descent update. To estimate the gradient of Q, we sample an entropy vector near the current vector and calculate its SERs per channel and final Q factor. We can take the difference between the old and new SERs and Q factors, from which the gradient of Q can be computed. Note that given an entropy vector, we can compute the signal SER based on the block diagram in Fig. 2(b) (principle in Section II.A). The current signal error probability can be obtained from taking average of the SER of the *N* PS units.

III. Experimental Setup

Fig. 3 illustrates the experimental setup and DSP diagram. In the transmitter DSP, subcarriers in the same PS unit share the same entropy and they are combined together to generate the target entropy with a smaller rate loss. For every PS unit, a pseudorandom binary sequence (PRBS) is loaded to a probabilistic amplitude shaping (PAS) block with the assigned entropy from the optimization algorithm. In the PAS block, the constant composition distribution matching (CCDM) generates the desired probability distribution for the amplitude part of the I and Q signal, respectively. We choose the target distribution as the Maxwell-Boltzmann distribution. Then the amplitude part combines with the sign part, which gives the QAM signal. Details of the PAS can be found in [16]. The symbol sequence in each PS unit is then parallelized into several subcarriers and normalized. In this demonstration, the PS-64QAM OFDM signal consists of 800 subcarriers and 54 symbols with a subcarrier spacing of 750 kHz, while PS-256QAM OFDM has a similar setting but with a subcarrier spacing of 500 kHz. The IFFT size is 2048 for both schemes and 5.12% CP is inserted. After OFDM modulation (IFFT operation and CP insertion) and P/S conversion, the complex OFDM signal is upconverted to an intermediate frequency (IF) signal with a carrier frequency of 500 MHz. The IF signal is loaded into a digital to analog converter (DAC) with a sampling rate of 16 GS/s and 8-bits ENOB. Then it is modulated by a 10-GHz directly modulated laser (DML). The optical signal is delivered through a 15-km single-mode fiber (SMF) and detected by a 10 GHz photodetector (PD). An attenuator before the PD controls the ROP in the system. The electrical IF signal is then converted to a 60-GHz mmW signal by a 18-GHz electrical mixer (input third order intercept equals to 13 dBm), a 15-GHz RF source and a frequency quadrupler. A wideband power amplifiers (PA) are employed to boost the signal on both sides of antennae. A horn antenna (50 GHz to 75 GHz) with 25-dBi gain radiates the 60-GHz radio signal over the air while a symmetric horn antenna receives the signal after 1.5-m wireless transmission. A 1-GHz envelope detector (ED) down-converts the radio signal to the IF signal again. Then a 5-GS/s ADC recovers the digital waveform. In the receiver DSP, we first down-converts the IF signal to baseband signal. After CP removal, S/P conversion, FFT and equalization with a one-tap equalizer, we perform the decision and de-mapping, which can be utilized to calculate the pre-FEC metrics such as BER, generalized mutual information (GMI), and normalized GMI (NGMI).

IV. Experimental Results and Analysis

A. Spectral Efficiency of the Signals

We utilize two pre-FEC thresholds in the analysis: 1) pre-FEC BER equals to 3.8×10^{-3} (HD-FEC) with 7% FEC OH (code rate = 0.935) [22]; 2) NGMI threshold of 0.86 (ideal FEC) with 16.3% FEC OH (code rate = 0.86) [13], [23]. Considering the rate loss is negligible, the spectral efficiency (SE) is expressed as [6], [16]:

$$SE = H - (1 - R) \cdot \log_2 M, \tag{8}$$

where *H* is the target entropy, *R* is FEC code rate and *M* is QAM order. We can observe that for higher FEC OH (lower FEC code rate), the PS signal needs to increase *H* more significantly. In the experiment, we choose PS 64-QAM signal (4.28 bits/symbol) versus Uniform 16QAM, and PS 256 QAM (6.28 bits/symbol) versus Uniform 64QAM. The SEs with the two FEC OH are summarized in Table I. Note that with 3.8×10^{-3} pre-FEC threshold (OH 7%), the SE of the PS signal is slightly higher than the Uniform signal. In the experiment, the DM rate losses of the PS 64QAM and PS 256QAM are 0.038 bits/symbol and 0.098 bits/symbol, respectively. However, in our following analysis, the PS signal with 7% OH (with higher SE) still shows sensitivity gain compared with uniform signal.

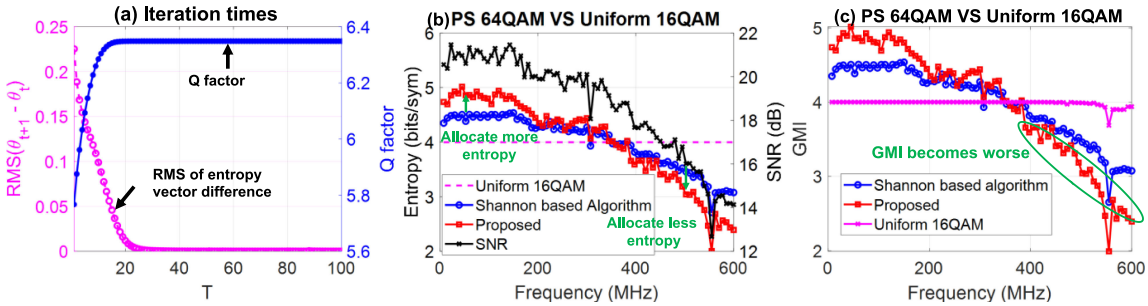

Fig. 4. (a) Convergence speed of the optimization algorithm (PS 64QAM, ROP $= -1$ dBm). PS 64QAM versus Uniform 16QAM. (b) SNR/Assigned entropy over frequency with different algorithms at ROP of 0 dBm. (c) GMI over frequency with different algorithms at ROP of 0 dBm.

TABLE I
SE OF DIFFERENT SIGNALS WITH DIFFERENT FEC OH

	7% FEC OH	16.3% FEC OH
SE of Uniform 16QAM	3.74 bits/symbol	3.44 bits/symbol
SE of PS 4.28 64QAM	3.89 bits/symbol	3.44 bits/symbol
SE of Uniform 64QAM	5.61 bits/symbol	5.16 bits/symbol
SE of PS 6.28 256QAM	5.76 bits/symbol	5.16 bits/symbol

TABLE II
COMPUTATIONAL COMPLEXITY SUMMARY

	PER ITERATION	ITERATION TIMES
Shannon's formula based method	O(MN)	1
Proposed	O(MN)	30

B. Compututional Complexity and Parameter Setting

In the following, we analyze the s complexity of the proposed algorithm and it is summarized in Table II. Firstly, the number of iterations is independent of the number of PS units, according to the analysis in Section II.B. Therefore, we only need to study the per-iteration cost for the algorithm. In each iteration, we compute the error probability for each PS unit independently and calculate the averaged Q factor twice, by taking the difference of which we estimate the gradient. The computational complexity scales linearly with N, the number of PS units in the system. For each PS unit, we need to carry on the computation in Eq. (2–5) for a given entropy level, which takes the computational complexity scaling linearly with the QAM order M. Overall, the computational complexity for the algorithm is $O(MN)$. The computational complexity thus remains the same order as the approach of allocating proportionally by Shannon's formula. It is worth noting that the computations in each channel and the computation related to each (i, j) pair (Eq. (2–5)) can be implemented in parallel, making it potentially fast with a hadrware implementation.

In the experiment, we set η as 0.1 and ε as 10^{-9}. The initial entropy is chosen to be proportional to Shannon's formula, in order to yield a faster convergence. Firstly, we investigate the number of steps (iteration) T based on the pre-measured SNR in the setup. Fig. 4(a) demonstrates root mean square (RMS) of the difference between the current allocated entropy vector θ_{t+1} and the allocated entropy vector θ_t in previous iteration with the increasing of the iteration number T. The predicted Q factor in every iteration is also presented, which aligns with the trend of the RMS of entropy vector difference. One can note that the variance of the Q factor is relatively small after 15 times of iterations while the allocated entropy RMS converges when T is around 20. In the experiment, T is fixed as 30 to ensure that the convergence results are always achieved.

Moreover, by sweeping the PS unit size from 10 to 800 using the Shannon formula based algorithms, 10 subcarriers per PS unit yields the best performance among the sweeping regions. Note that one subcarrier loading might give better entropy loading resolution but yield significant rate loss in our experiment setup. Thus, to make a trade-off between entropy loading resolution and rate loss optimization, PS unit size is fixed as 10 subcarriers per unit in the following algorithm comparisons. Pre-FEC BER, GMI, and normalized GMI (NGMI) are utilized to characterize system performance. The GMI denotes the maximum number of information bits per transmit symbol in the bit-interleaved coded modulation (BICM) AWGN auxiliary channel, with ideal binary FEC decoding. The NGMI implies the binary FEC code rate.

C. Performance Analysis

Firstly, the channel SNR distribution over frequency is pre-measured through the EVM of training signal via [24]:

$$SNR = 1/EVM^2. \tag{9}$$

The training symbols are 20 QPSK symbols. Based on the pre-measured SNR, we can compute the assigned entropy using different algorithms. The Shannon formula based allocation scheme allocates the entropies proportionally based on Shannon's formula with a fixed total target entropy as shown below [15], [16]:

$$H_j = \frac{S \cdot H \cdot \log_2(1 + SNR_j)}{\sum_i \log_2(1 + SNR_i)}, \tag{10}$$

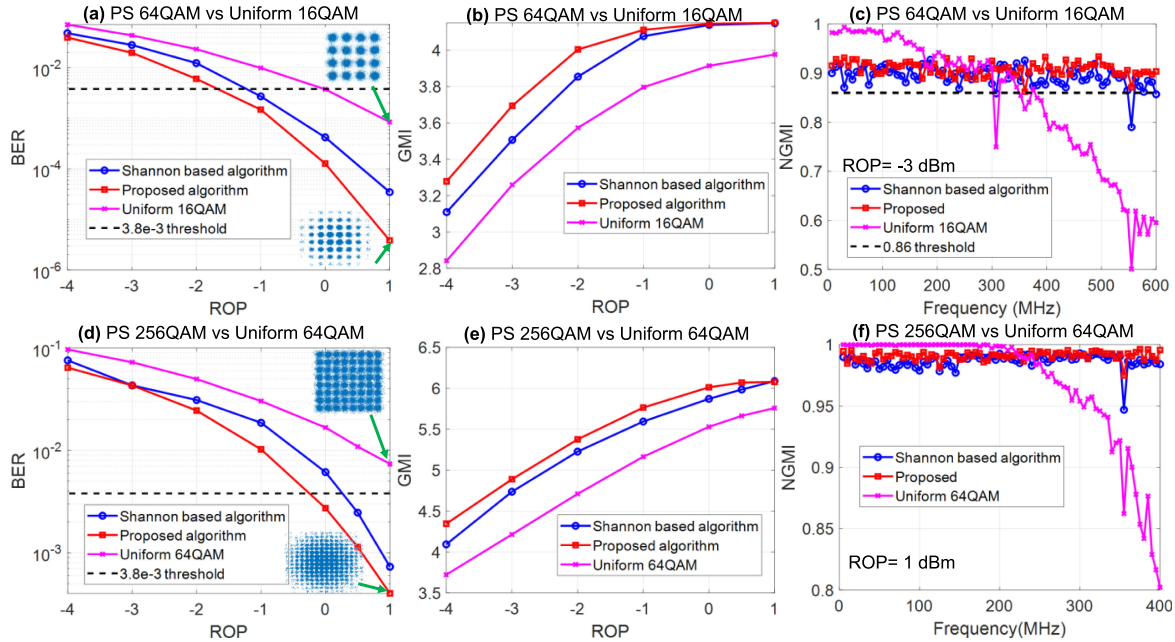

Fig. 5. Experimental results of different algorithms. PS 64QAM vs Uniform 16QAM regarding (a) Pre-FEC BER versus ROP; (b) GMI versus ROP; (c) NGMI over frequency. PS 256 QAM vs Uniform 64 QAM regarding (d) Pre-FEC BER versus ROP; (e) GMI versus ROP; (f) NGMI over frequency.

where H_j and SNR_j denote allocated entropy (bits/symbol) and pre-measured SNR of the j_{th} PS unit, respectively. H is the target averaged entropy of the whole data frame as indicated in part A.The proposed algorithm is presented in Fig. 2(a) and Section II.B.

Fig. 4(b) shows the allocated entropy result. The black curve is the pre-estimated SNR curve of the channel. Both the proposed optimization algorithm and the algorithm based on the Shannon-Hartley theorem can track the SNR fluctuations over the frequency and show similar shape to the SNR curve. Compared with the algorithm based on the Shannon formula, the proposed algorithm allocates more entropy for the data with higher SNR while assigns less entropy for the data with lower SNR. This is due to the modeling result (Section II. A) changes the maximum supported capacity given a SNR.

On the other hand, we also compare the GMI over frequency for different algorithms as presented in Fig. 4(c). The GMI trend is similar to what we observed in Fig. 4(b). For the region with higher SNR (e.g., 0 to 350 MHz), the proposed algorithm has better GMI due to more allocated entropy.

Then we sweep the ROP and compare the sensitivity performance of the algorithms. Fig. 5(a) to (c) present the performance of PS 4.28 64QAM versus uniform 16QAM while Fig. 5(d) to (f) show the performance of the PS 6.28 256QAM versus uniform 64QAM. In the experiment, the pre-FEC BER is calculated using ML hard decision. NGMI utilizes the bit-wise log-likelihood ratio and assumes an ideal binary FEC. In practice, the pre-FEC characterization parameter depends on the specific FEC scheme implemented in the system. For instance, pre-FEC BER is utilized for HD-FEC while GMI is more suitable for SD-FEC.

Fig. 5(a) shows that, compared with uniform 16QAM, the Shannon formula based algorithm has a 1.2-dB ROP gain with the BER threshold of 3.8×10^{-3} while the proposed algorithm shows a 1.8-dB gain. Note that the SE or information rate of the PS signal is higher than that of the Uniform signal in this case. The sensitivity gain could be higher if keeping the SE. Fig. 5(b) illustrates the GMI versus ROP. In comparison to the Shannon formula based algorithm, our proposed algorithm shows an additional 0.18-bits/symbol improvement of GMI value when the ROP ranges from -4 dBm to -2 dBm. When the ROP is higher, the two algorithms gradually converge due to saturation to its maximum supported entropy. The uniform 16QAM is always worse than the curves with entropy allocation. Fig. 5(c) compares the NGMI performances over frequency for three algorithms. Reference line of NGMI $= 0.86$ is provided. PS-OFDM 64-QAM signal with the proposed algorithm has the smoothest NGMI curve. The schemes with uniform 16QAM and Shannon formula based algorithm cannot pass the 0.86 threshold in some frequency regions. For the sharp "dip" of the Uniform 16QAM signal (where the signal has a very significant SNR degradation such as $\sim$300 MHz or 550 MHz), the Shannon formula based algorithm still shows a small "dip" while the proposed algorithm shows a flat NGMI.

As for PS 256QAM versus uniform 64QAM, in comparison to the Shannon formula based algorithm, the proposed algorithm has a 0.5-dB sensitivity gain with respect to 3.8×10^{-3} pre-FEC BER threshold (Fig. 5(d)). As for GMI in Fig. 5(e), the proposed algorithm also has better performance (0.7 bits/symbol GMI improvement) while the Shannon based equation only shows 0.4 bits/symbol GMI enhancement. As illustrated in Fig. 5(f), the channel response is smoother so the difference between the Shannon formula based algorithm and the proposed algorithm becomes smaller. Overall, the NGMI of the proposed algorithm still lies above the Shannon formula based algorithm (blue curve).

V. Conclusion

In this paper, we have proposed a novel algorithm for entropy allocation in the PS-OFDM mobile fronthaul link. The closed-form expression of the pre-FEC SER versus SNR under AWGN channel condition is derived. Based on the modeling result, the optimization algorithm utilizes the pre-measured SNR information and allocates the entropy intelligently using a projected mirror descent algorithm with the pre-measured SNR information. Experimental demonstration indicates a convergence speed of 20 to 30 iteration times. The proposed algorithm allocates more entropy for regions with higher SNR and less entropy for regions with lower SNR. Furthermore, compared with a simple Shannon formula based algorithm, the result also shows a smoother NGMI performance, up to 0.25 bits/symbol improvement of GMI value, and up to 0.6-dB sensitivity improvement with uneven frequency response. MAP detection can further optimize the scheme with better performance. In summary, the proposed algorithm is promising to extend the link power budget or improve the signal quality in a fiber-mmW converged fronthaul, especially in an uneven channel condition.

Appendix

The error probability of the corner symbols can be written as [17]:

$$\begin{aligned}\mathrm{P}\left(wrong|corner\right) &= 1 - P\left(correct|corner\right)\\ &= 1 - P\left(n_I < \sqrt{E_g} \& n_Q < \sqrt{E_g}\right)\\ &= 1 - P\left(n_I < \sqrt{E_g}\right) \cdot P\left(n_Q < \sqrt{E_g}\right)\\ &= 1 - \left(1 - Q\left(\sqrt{2E_g/N_0}\right)\right)^2\\ &= 2Q\left(\sqrt{2E_g/N_0}\right) - Q^2\left(\sqrt{2E_g/N_0}\right),\end{aligned} \tag{11}$$

where Q is the Q function. Similarly, we can also obtain the error probability of the symbol for edge case and inner case, respectively [17].

$$\begin{aligned}\mathrm{P}\left(wrong|\mathrm{edge}\right) &= 1 - P\left(|n_I| < \sqrt{E_g} \& n_Q < \sqrt{E_g}\right)\\ &= 3Q\left(\sqrt{2E_g/N_0}\right) - 2Q^2\left(\sqrt{2E_g/N_0}\right).\end{aligned} \tag{12}$$

$$\begin{aligned}\mathrm{P}\left(wrong|\mathrm{inner}\right) &= 1 - P\left(|n_I| < \sqrt{E_g} \& |n_Q| < \sqrt{E_g}\right)\\ &= 4Q\left(\sqrt{2E_g/N_0}\right) - 4Q^2\left(\sqrt{2E_g/N_0}\right).\end{aligned} \tag{13}$$

References

[1] G. Chang and Y. Chen, "Key fiber wireless integrated radio access technologies for 5G and beyond," in *Proc. 24th OptoElectron. Commun. Conf. Int. Conf. Photon. Switching Comput.*, 2019.

[2] R. Zhang *et al.*, "An ultra-reliable MMW/FSO A-RoF system based on coordinated mapping and combining technique for 5G and beyond mobile fronthaul," *J. Lightw. Technol.*, vol. 36, no. 20, pp. 4952–4959, Oct. 2018.

[3] H. Wang, C. Cheng, C. Tsai, Y. Chi, and G. Lin, "28-GHz wireless carrier heterodyned from orthogonally polarized tri-color laser diode for fading-free long-reach MMWoF," *J. Lightw. Technol.*, vol. 37, no. 13, pp. 3388–3400, Jul. 2019.

[4] X. Li, J. Yu, and G.-K. Chang, "Photonics-aided millimeter-wave technologies for extreme mobile broadband communications in 5G," *J. Lightw. Technol.*, vol. 38, no. 2, pp. 366–378, Jan. 2020.

[5] Z. Cao, L. Zou, L. Chen, and J. Yu, "Impairment mitigation for a 60 GHz OFDM radio-over-fiber system through an adaptive modulation technique," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 3, no. 9, pp. 758–766, Sep. 2011.

[6] G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic shaping and forward error correction for fiber-optic communication systems," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 230–244, Jan. 2019.

[7] M. Kong and J. Yu, "Performance improvement on a mimo radio-over-fiber system by probabilistic shaping," *Opt. Commun.*, vol. 407, pp. 87–91, 2018.

[8] Y. Ha, W. Niu, and N. Chi, "Post equalization scheme based on deep neural network for a probabilistic shaping 128 QAM DFT-S OFDM signal in underwater visible light communication system," in *Proc. 18th Int. Conf. Opt. Commun. Netw.*, 2019.

[9] J. Shi *et al.*, "Improved performance of high-order qam ofdm based on probabilistically shaping in the datacom," in *Proc. Opt. Fiber Commun. Conf. Expo.*, 2018.

[10] J. Shi *et al.*, "Probabilistically shaped 1024-QAM OFDM transmission in an im-dd system," in *Proc. Opt. Fiber Commun. Conf.*, 2018, Paper W2A.44.

[11] F. Tian *et al.*, "Probabilistic shaped trellis coded modulation with generalized frequency division multiplexing for data center optical networks," *Opt. Express*, vol. 27, pp. 33159–33169, 2019.

[12] F. Steiner, P. Schulte, and G. Bocherer, "Approaching waterfilling capacity of parallel channels by higher order modulation and probabilistic amplitude shaping," in *Proc. 52nd Annu. Conf. Inf. Sci. Syst.*, 2018.

[13] D. Che and W. Shieh, "Achievable rate comparison between entropy and bit loading in a 100-Gb/s DM-DD DMT system," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2019.

[14] C. Xie *et al.*, "Achievable information rate enhancement of visible light communication using probabilistically shaped OFDM modulation," *Opt. Express*, vol. 26, pp. 367–375, 2018.

[15] R. Zhang *et al.*, "Joint optimization of processing complexity and rate allocation through entropy tunability for 64-/256-QAM based radio fronthauling with LDPC and PAS-OFDM," in *Proc. Opt. Fiber Commun. Conf.*, 2020, Paper M2F.2.

[16] R. Zhang, Y. Chen, W. Mou, and G. Chang, "Rate redundancy and entropy allocation for PAS-OFDM based mobile fronthaul," *J. Lightw. Technol.*, to be published.

[17] J. R. Barry, D. G. Messerschmitt, and E. A. Lee, *Digital Communication*, 3rd ed., Norwell, MA, USA: Kluwer Academic Publishers, 2003.

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

[19] S. Hu *et al.*, "Map detection of probabilistically shaped constellations in optical fiber transmissions," in *Proc. Opt. Fiber Commun. Conf.*, 2019, Paper W1D.3.

[20] T. Mizuochi, "Recent progress in forward error correction and its interplay with transmission impairments," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, no. 4, pp. 544–554, Jul./Aug. 2006.

[21] A. S. Nemirovskiĭ and D. B. IUdin, *Problem Complexity and Method Efficiency in Optimization*. Hoboken, NJ, USA: Wiley, 1983.

[22] J. Zhang *et al.*, "280 Gb/s IM/DD PS-PAM-8 transmission over 10 km ssmf at O-band for optical interconnects," in *Proc. Opt. Fiber Commun. Conf.*, 2020, Paper M4F.1.

[23] J. Cho, L. Schmalen, and P. J. Winzer, "Normalized generalized mutual information as a forward error correction threshold for probabilistically shaped QAM," in *Proc. Eur. Conf. Opt. Commun.*, 2017.

[24] R. A. Shafik, M. S. Rahman, and A. R. Islam, "On the extended relationships among EVM, BER and SNR as performance metrics," in *Proc. Int. Conf. Elect. Comput. Eng.*, 2006, pp. 408–411.