Probabilistic Shaping in Time-Frequency-Packed Terabit Superchannel Transmission

Mrinmoy Jana, Lutz Lampe, Jeebak Mitra, Wang Jin, and Kevin Law

***Abstract*—To combine the individual benefits of probabilistic-shaping (PS) and time-frequency-packing (TFP), we consider for the first time PS-TFP wavelength-division multiplexing (WDM) superchannels. However, TFP introduces inter-symbol interference (ISI) and/or inter-carrier interference (ICI). Moreover, the presence of reconfigurable optical add-drop multiplexers in the fiber links may further degrade the system performance. In this letter, we efficiently handle such challenges to present PS-TFP superchannels enabling Terabit-per-second data rates. For this, we perform a joint ISI and ICI channel estimation in tandem with turbo equalization to mitigate the interference. We investigate optimizing the parameters in the shaping and packing dimensions to achieve a desired target spectral efficiency. We show through our numerical results that such an optimized PS-TFP transmission leads up to 1.2 dB performance improvement over an unshaped Nyquist WDM system under similar conditions.**

***Index Terms*—Faster-than-Nyquist (FTN), Time-frequency packing (TFP), Super-Nyquist, Probabilistic shaping (PS), Superchannel.**

I. Introduction

TERABIT-per-second (Tbps) data rates per carrier are being targeted in the next generation optical fiber networks [1], [2]. To accomplish such a target, transmitting very high baud rates can be challenging due to the bandwidth limitations of the opto-electronics. Therefore, increasing the spectral efficiency (SE) is a key requirement for these networks. As an alternative to applying higher-order modulation formats, which are sensitive to fiber nonlinearity and laser linewidth, an attractive choice to increase the SE is to enable time-frequency packing (TFP) wavelength-division-multiplexing (WDM) superchannel transmission [2]. TFP allows controlled overlap of the symbols in time or frequency or both. However, such an SE improvement comes at the cost of introducing inter-symbol interference (ISI) and inter-carrier interference (ICI). Probabilistic shaping (PS) [3] is another enabling technology that has received renewed interest in the optical fiber communication literature more recently. By assigning non-uniform probability mass function (PMF) to the transmitted constellation, theoretically, PS can provide an ultimate shaping gain of 1.53 dB [3].

We investigate combining PS with TFP for the first time, to reap the aggregate benefits of the two technologies. We consider WDM superchannels achieving Tbps data rates, in the presence of narrow optical filtering due to the reconfigurable optical add-drop multiplexers (ROADMs) implemented through wavelength selective switches (WSSs). Such a severe filtering may lead to strong ISI for the edge subchannels (SCs) of a superchannel. By packing the symbols in time and/or frequency, and thereby reducing the aggregate signal bandwidth (BW), TFP can moderate the impact of such filtering, at the expense of introducing ISI/ICI of its own. With PS providing an additional degree of freedom, optimizing the transmission parameters in the time, frequency, and shaping dimensions can lead to significant performance improvements, and therefore, such an investigation is crucial for a practical transceiver design.

We apply powerful digital signal processing (DSP) techniques for the proposed PS-TFP systems to efficiently handle interference. For this, we employ a joint ISI and ICI channel estimation algorithm in conjunction with a turbo ISI cancellation (ISIC) and ICI cancellation (ICIC) structure. To restrict the receiver-side DSP complexity, at the beginning, we investigate frequency-only-packed (FP) transmissions. Our numerical results show that such a scheme achieves a similar performance as that of a time-only-packed (TP) system [1], with a significantly lower computational cost. Next, motivated by the fact that TFP can theoretically provide higher achievable rates than TP and FP [2], we consider packing of symbols in both time and frequency dimensions to obtain improved performance, at the expense of additional computational complexity. We show that by optimizing the shaping and packing parameters, substantial performance gains are achieved over unshaped Nyquist WDM transmission for the same data rate.

II. System Model

The block diagram of a dual-polarized (DP) PS-TFP WDM superchannel transmission is shown in Fig. 1 for the X polarization (pol.) of the m^{th} SC, $m = 1, 2, \ldots, N$, where N is the total number of SCs in the superchannel. Uniformly distributed input bits for each pol. are processed by a constant-composition distribution matcher (CCDM) to feed a low-density parity-check (LDPC) encoder and a bit mapper, that produces a sequence of shaped M-ary quadrature amplitude modulation (QAM) symbols a_m having a desired PMF [3]. Thereafter, the digital samples are pulse-shaped by a root-raised cosine (RRC) filter p with a roll-off factor β, and transmitted as a WDM signal using electrical digital-to-analog converters (DACs) and an optical frontend comprising of a driver and a Mach-Zehnder (MZ) modulator. The baseband transmitted faster-than-Nyquist (FTN) signal can be written as

$$s(t) = \sum_{l} \sum_{m} a_m[l] p(t - l\tau T) \mathrm{e}^{\mathrm{j}2\pi\left(m - \frac{N+1}{2}\right)\Delta f t}, \quad (1)$$

where $0 < \tau \leq 1$ is the FTN acceleration factor for each SC, $\frac{1}{\tau T}$ is the per-SC baud rate, $\Delta f = \xi \frac{1+\beta}{T}$ denotes the frequency separation between the SCs, with $\xi > 0$ being the frequency compression ratio, such that $\tau = \xi = 1$ corresponds to Nyquist signaling. While propagating through the optical fiber, the

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Huawei Tech., Canada. M. Jana and L. Lampe are with the ECE Dept., Univ. of Brit. Col., BC, Canada. Email: {mjana, lampe}@ece.ubc.ca. J. Mitra and W. Jin are with Huawei Tech., Ottawa, Canada. Email: {jeebak.mitra, wangjin2}@huawei.com. K. Law is with Huawei Tech., Shenzen, China. Email: k.law@huawei.com.

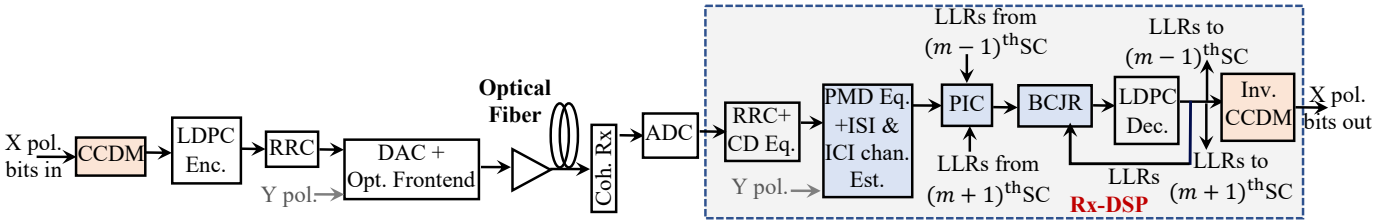


Fig. 1. PS-TFP simulation block diagram, shown for the X-pol. of the m^{th} SC. The gray shaded block marked with the dashed border represents the Rx-DSP, the blue and the red shaded blocks represent DSP modifications specific to TFP and PS, respectively.

transmitted signal is affected by fiber impairments and severe optical filtering due to WSSs. At the coherent receiver (Rx), the digital samples are fed as inputs to the Rx-DSP module for further processing. The details of the DSP operations are provided in the following.

III. PS-TFP Transmission

In the receiver data-path, the RRC matched-filtered samples are first equalized by the chromatic dispersion (CD) equalizer. Next, the equalized samples are processed by an adaptive data-aided least-mean-square (LMS) based estimation algorithm that jointly estimates the 2×2 polarization mode dispersion (PMD) equalizer coefficients and the ISI and ICI channel impulse responses together. Thereafter, the received signal for each pol. branch is processed by the turbo ICIC and ISIC modules. Finally, the output bits are generated following the LDPC decoding and the inverse-CCDM operation. The adaptive channel estimation and interference mitigation algorithms are described as follows.

A. PMD Equalizer, ISI and ICI Channel Estimation

While the TFP-induced interference is a-priori known, an additional channel estimation leads to performance improvements by minimizing the residual interference [4]. Moreover, such an estimation can also account for the additional electrical/optical filtering, such as the WSSs, in the optical links. To accomplish this, we extend the concept presented in [4] for a TP system to our PS-TFP transmission, to jointly estimate the PMD equalizer, and the ISI and ICI channel taps. We remark that using a similar strategy, a more generalized algorithm can also be developed to estimate additional transmitter, receiver and channel impairments (see e.g. [5] for further details.) For the TFP transmission considered here, the adaptive estimation algorithm for the PMD equalizer, ISI and ICI channel is summarized in the following lemma. For this, we denote the shaped constellation symbols and the input received samples to the 2×2 PMD equalizer by the column vectors $\boldsymbol{a}_k^{(m)}$ and $\boldsymbol{r}_k^{(m)}$, respectively, which are formulated by stacking the X and Y pol. signals of the m^{th} SC at the k^{th} sample time.

Lemma 1. *For the m^{th} SC, the LMS update equations for the L_{p}-tap 2×2 PMD equalizer, $(2L_{\text{s}}+1)$-tap ISI channel, and the $(2L_{\text{c}}+1)$-tap ICI channel can be written as*

$$\boldsymbol{C}_{i,k+1}^{(m)} = \boldsymbol{C}_{i,k}^{(m)} - \mu_{\text{c}} \odot \boldsymbol{e}_k^{(m)} \left(\boldsymbol{r}_{k-i}^{(m)}\right)^{\text{H}}, \tag{2}$$

$$\boldsymbol{h}_{j,k+1}^{(m)} = \boldsymbol{h}_{j,k}^{(m)} + \mu_{\text{h}} \text{Re}\left[\left(\boldsymbol{e}_k^{(m)} \odot \left(\boldsymbol{a}_{k-j}^{(m)} + \boldsymbol{a}_{k+j}^{(m)}\right)^*\right)\right], \tag{3}$$

$$\boldsymbol{g}_{w,k+1}^{(n,m)} = \boldsymbol{g}_{w,k}^{(n,m)} + \mu_{\text{g}}\left(\boldsymbol{e}_k^{(n)} \odot \left(\hat{\boldsymbol{a}}_{k-w}^{(m)}\right)^* + \left(\boldsymbol{e}_k^{(n)}\right)^* \odot \hat{\boldsymbol{a}}_{k+w}^{(m)}\right), \tag{4}$$

where $\boldsymbol{C}_{i,k}^{(m)}$ is the i^{th} matrix-tap of the $\frac{T}{2}$-spaced PMD equalizer such that $0 \le i \le L_{\text{p}}-1$, $\boldsymbol{h}_{j,k}^{(m)}$ and $\boldsymbol{g}_{w,k}^{(n,m)}$ are the j^{th} and the w^{th} symbol-spaced ISI and ICI tap vector, respectively, such that $0 \le j \le L_{\text{s}}$ and $0 \le w \le L_{\text{c}}$, $\mu_{\text{c}} > 0$, $\mu_{\text{h}} > 0$ and $\mu_{\text{g}} > 0$ are the step size parameters, $\text{Re}[\cdot]$, $(\cdot)^*$, $(\cdot)^{\text{H}}$ and $\odot$ denote the real-part, complex conjugation, matrix hermitian and elementwise vector product operations, respectively, $\boldsymbol{e}_k^{(m)}$ is the error signal vector for the LMS adaptation, and $\hat{\boldsymbol{a}}_k^{(n)} = \boldsymbol{a}_k^{(n)} \text{e}^{\pm \text{j}2\pi\Delta f k}$ denotes the rotated constellation symbol for the n^{th} SC, with $\pm$ sign determined from (1) depending on the relative positions of the SCs.

Proof: See Appendix. ∎

For the initial LMS convergence, a continuous block of pilot symbols are transmitted at the beginning of data transmission, followed by periodic pilots insertion for slow adjustments of the estimates. The estimated ISI and ICI taps are provided as inputs to the interference mitigation module as detailed next.

B. Turbo Interference Cancellation

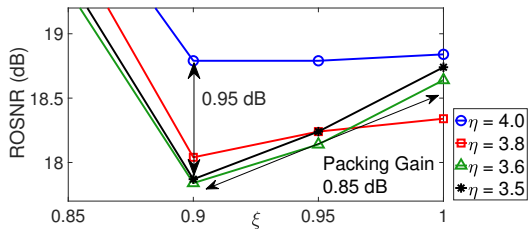
To mitigate the interference, we employ a turbo parallel interference cancellation (PIC) based ICIC method in tandem with a Bahl-Cocke-Jelinek-Raviv (BCJR) based ISIC structure, similar to [2]. The ICIC scheme uses the estimated ICI impulse response and the a-posteriori log-likelihood ratios (LLRs) fed back from the LDPC decoders of the neighboring SCs to subtract the soft estimates of the ICI from the received samples of each SC, iteratively. Thereafter, the BCJR-ISIC mitigates the ISI using the estimated ISI channel, and provides LLRs back to the LDPC decoder. To account for PS, prior to the first LDPC iteration, the PMF of the shaped constellation is used by the BCJR-ISIC for the a-priori probability computation. At the end of the final iteration, bits are generated from the LDPC decoders and fed as inputs to the inverse-CCDM to produce the output bits.

TABLE I
Simulation Settings

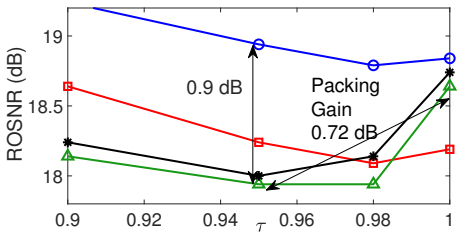
Baud rate per SC (Gbaud)	Tot. Nyquist BW (GHz)	WSS 3-dB BW (GHz)	λ in MB distr.	$\eta=$ Raw BPS per pol.
68.50	226.05	225	0	4.0
72.11	237.96		0.0855	3.8
76.11	251.16		0.1249	3.6
78.29	258.36		0.1419	3.5
90.00	297.00	300	0.2207	3.0
96.43	318.22		0.2528	2.8
103.85	342.71		0.2875	2.6

IV. Numerical Results

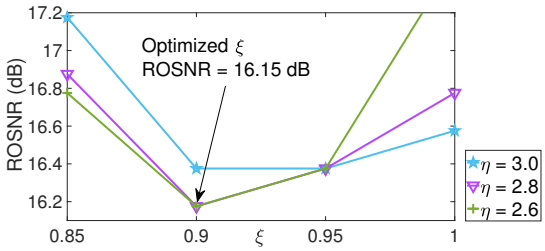
To exhibit the superior performance of the proposed design, we consider DP 16-QAM PS-TFP superchannels having 3 SCs, where each SC has a data rate of 400 Gbps, such that the 1.2 Tbps superchannels are packed within an aggregate BW of 225 GHz or 300 GHz, corresponding to two different realistic SE targets [1]. We consider $\beta = 0.1$,



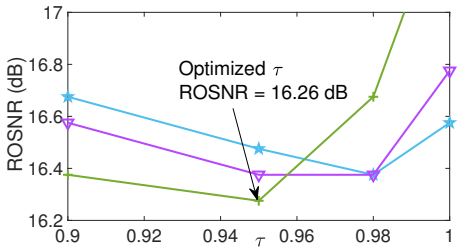
(a) 225 GHz system, FP only ($\tau = 1$).



(b) 225 GHz system, TP only ($\xi = 1$).



(c) 300 GHz system, FP only ($\tau = 1$).



(d) 300 GHz system, TP only ($\xi = 1$).

Fig. 2. TP vs. FP Tbps superchannels.

varying τ and ξ, LDPC codes with rate 0.8 and codeword-length 64800 bits, and varying per-SC baud rates, as shown in Table I. PS configurations with the Maxwell-Boltzman (MB) distribution having different values of the parameter λ [3] are employed for our numerical evaluation. In Table I, we compute the parameter "Total Nyquist BW" as the product $3(1+\beta) \times$ baud rate per SC, and η denotes the uncoded bits per symbol (BPS) per pol. Perfect frequency synchronization and phase-lock are assumed. We transmit 30,000 symbols at the beginning of data transmission for the initial LMS convergence, followed by 4% periodic pilot symbols insertion uniformly across the data frame. We consider 1 WSS in the optical link, where each WSS is modeled as a 6^{th} order Gaussian filter [2]. Moreover, $L_{\text{s}} = 3$ corresponding to 64-state BCJRs, $L_{\text{c}} = 9$, and a maximum iteration count of 10 between the ISIC/ICIC and the LDPC decoders are considered.

A. Optimized TP & FP Transmission: Performance vs. Complexity

As shown in Table I, to achieve the same data rate with more shaping, corresponding to a lower value of η, a higher baud rate is required. However, a higher baud rate occupies a larger BW, which suffers more from the ISI due to optical filtering from WSS. Therefore, to obtain the maximum shaping gain, an optimal η should be applied for a given WSS 3-dB BW and a target data rate. Furthermore, to reduce the impact of ISI due to WSS on the edge SCs, TP and/or FP can be enabled that reduces the aggregate superchannel BW [2]. We remark that a shaped TP or FP transmission is a special case of the PS-TFP system described in Section III, with $\xi = 1$ or $\tau = 1$, respectively. By optimizing the amount of temporal and spectral overlap, the maximum packing gain can be achieved for such systems [2]. In Fig. 2, we show such optimality for the 225 GHz and the 300 GHz TP and FP systems, by plotting the required optical signal-to-noise ratio (ROSNR) to obtain an error-free transmission, as a function of τ or ξ. For example, with the 225 GHz FP system shown in Fig. 2(a), we highlight a packing gain of 0.85 dB, corresponding to an optimal $\eta = 3.6$ bits/symbol/pol and an optimal $\xi = 0.9$. For this ξ value,

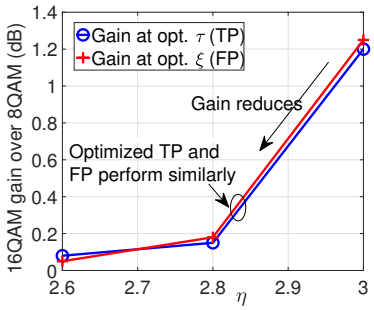


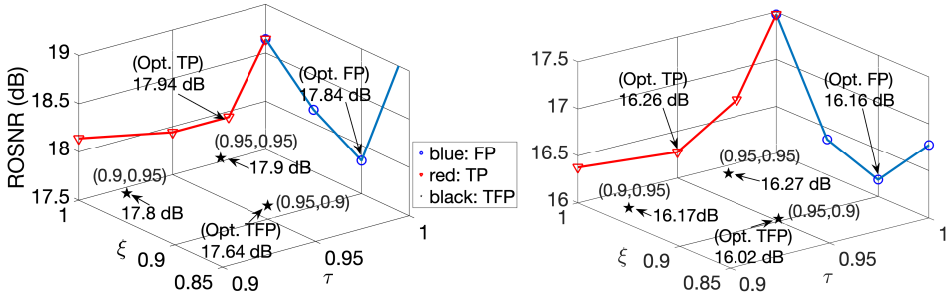
Fig. 3. ROSNR gain of PS 16-QAM compared to PS 8-QAM, for the optimized TP and FP configurations. 1.2 Tbps superchannels in 300 GHz.

we also observe an ROSNR gain of 0.95 dB compared to $\eta = 4$ bits/symbol/pol.

We performed a similar investigation with DP star-8 QAM 1.2 Tbps superchannels packed in a 300 GHz aggregate BW, corresponding to the last three entries of baud rates and η values in Table I. A comparison between the 16-QAM and 8-QAM systems is shown in Fig. 3 at the optimal TP and FP configurations, where the ROSNR gain produced by the 16-QAM system over the 8-QAM transmission is plotted against varying η. The plots suggest that as η decreases, the OSNR gains by 16-QAM reduce. The figure also indicates that the optimized TP and FP behave almost similarly. Such a trend can also be observed by comparing Fig. 2(a) with Fig. 2(b), and Fig. 2(c) with Fig. 2(d), which shows that the lowest achieved ROSNRs for the FP and TP systems are similar. However, to achieve a similar performance, the FP systems employ turbo-PIC ICIC, whereas the shaped TP systems, similar to the ones presented in [1], apply BCJR-ISIC, the DSP complexity of which is significantly higher than the PIC [2]. Computational complexity for both schemes is shown in Table II. For the simulation parameters used in this work, BCJR-ISIC requires 30-times additional computations compared to the turbo ICIC.

B. Shaped TFP Performance

To show the benefits of TFP, we evaluate the performance of the PS-TFP WDM transmission, as described in Section III, under the same data rate and BW constraints. In Fig. 4(a)-4(b), we show the ROSNR as a function of both τ and ξ,



(a) 225 GHz system, $\eta = 3.6$ bits/symbol/pol.

(b) 300 GHz system, $\eta = 2.6$ bits/symbol/pol.

Fig. 4. Optimization of τ and ξ for the PS-TFP superchannels.

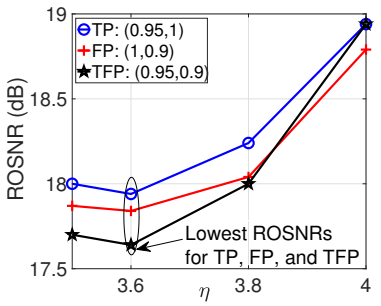


Fig. 5. ROSNR vs. η for the 225 GHz system, with three different (τ, ξ) combinations, corresponding to a TP, FP, and TFP transmission.

by means of three-dimensional plots. As shown in the figures, the lowest achievable ROSNRs for the PS-TFP systems are lower than those of the the optimized TP and FP transmission by $0.2 - 0.3$ dB. Since finding an analytical expression for the ROSNR as a function of TFP parameters seems to be intractable, the optimal τ and ξ for the TFP systems need to be numerically evaluated for each combination of the transmission and shaping parameters, and the employed ISIC and ICIC method. Comparing Fig. 2(a) with Fig. 4(a) indicates that for the 225 GHz system, the lowest ROSNR achieved by the PS-TFP transmission, with $\eta = 3.6$ bits/symbol/pol, $\tau = 0.95$ and $\xi = 0.9$, corresponds to a gain of 1.2 dB over an unshaped Nyquist WDM transmission, for which $\eta = 4$ bits/symbol/pol, $\tau = 1$ and $\xi = 1$. Finally, we plot the ROSNR as a function of η in Fig. 5 for the 225 GHz system, with three different τ and ξ combinations. As shown in the figure, for these specific TP, FP and TFP configurations, lowest ROSNRs are achieved at $\eta = 3.6$ bits/symbol/pol., through an optimal trade-off between the WSS induced distortion and shaping gain. However, such temporal and spectral compression values are not optimal for all shaping configurations, which requires further optimization in the (τ, ξ) space for each value of η.

V. Conclusion

To reap the combined benefits of PS and TFP, we optimize the use of both the concepts for the first time to transmit Tbps superchannels in the presence of severe bandwidth limitation. We presented powerful interference channel estimation and mitigation techniques to achieve this. By optimizing the parameters in the shaping and the packing dimensions, we showed that FP systems employing turbo-PIC ICIC achieved similar performance as TP systems applying BCJR-ISIC, with significantly lower computational complexity. However, PS-TFP superchannels outperformed both TP and FP systems to achieve up to 1.2 dB ROSNR gain over an unshaped Nyquist WDM transmission.

TABLE II
Computational complexity shown for each code-symbol, SC, polarization, and LDPC iteration. $M = 16$, $L_\mathrm{s} = 3$, $L_\mathrm{c} = 9$.

Operations	BCJR	PIC
ADD/SUB	$8M^{\frac{L_\mathrm{s}}{2}} - 4$	$M + 2L_\mathrm{c} + 1$
MUL/DIV	$16M^{\frac{L_\mathrm{s}}{2}} + 2$	$M + 2L_\mathrm{c} + 1$
Nonlin.	$8M^{\frac{L_\mathrm{s}}{2}} + 2$	0
Total	$32M^{\frac{L_\mathrm{s}}{2}}$	$2(M + 2L_\mathrm{c} + 1)$

Appendix

Proof of Lemma 1

The error signal for the LMS adaptation is computed as the difference between the PMD filter output and the "desired signal" [4], [5], whereby the combined error vector with the X and Y polarization symbols is written as

$$\boldsymbol{e}_k^{(m)} = \underbrace{\boldsymbol{f}_k^{(m)}}_{\text{filtered output}} - \underbrace{\boldsymbol{d}_k^{(m)}}_{\text{desired signal}} , \tag{5}$$

where

$$\boldsymbol{f}_k^{(m)} = \sum_{i=0}^{L_\mathrm{p}-1} \boldsymbol{C}_{i,k}^{(m)} \boldsymbol{r}_{k-i}^{(m)} , \tag{6}$$

$$\boldsymbol{d}_k^{(m)} = \sum_{j=-L_\mathrm{s}}^{L_\mathrm{s}} \boldsymbol{h}_{j,k}^{(m)} \odot \boldsymbol{a}_{k-j}^{(m)} + \sum_{n \neq m} \sum_{w=-L_\mathrm{c}}^{L_\mathrm{c}} \boldsymbol{g}_{w,k}^{(n,m)} \odot \hat{\boldsymbol{a}}_{k-w}^{(n)} . \tag{7}$$

From (5), we define the mean squared error (MSE) as

$$\mathcal{M}_\mathrm{tot} = \mathbb{E}\left(\sum_{m=1}^{N} \| \boldsymbol{e}_k^{(m)} \|^2 \right) , \tag{8}$$

where $\mathbb{E}(\cdot)$ denotes the expectation operator. Finally, exploiting the symmetry of the real-valued TFP ISI channel and the conjugate symmetry of the ICI channel [2], [5], and computing the gradients $\frac{\partial \mathcal{M}_\mathrm{tot}}{\partial \boldsymbol{C}_{i,k}^{(m)}}$, $\frac{\partial \mathcal{M}_\mathrm{tot}}{\partial \boldsymbol{h}_{j,k}^{(m)}}$ and $\frac{\partial \mathcal{M}_\mathrm{tot}}{\partial \boldsymbol{g}_{w,k}^{(n,m)}}$, we obtain the LMS update equations based on the well-known gradient decent algorithm [4], [5].

References

[1] L. Li *et al.*, "Non-Orthogonal WDM Systems with Faster than Nyquist Technology," in *Opt. Fiber Commun. Conf.(OFC)*, 2019, pp. 1–3.

[2] M. Jana *et al.*, "Interference Cancellation for Time-Frequency Packed Super-Nyquist WDM Systems," *IEEE Photon. Technol. Lett.*, vol. 30, no. 24, pp. 2099–2102, 2018.

[3] G. Böcherer *et al.*, "Probabilistic Shaping and Forward Error Correction for Fiber-Optic Communication Systems," *J. Lightw. Technol*, vol. 37, no. 2, pp. 230–244, 2019.

[4] M. Secondini *et al.*, "Optical Time-Frequency Packing: Principles, Design, Implementation, and Experimental Demonstration," *J. Lightw. Technol*, vol. 33, no. 17, pp. 3558–3570, 2015.

[5] M. Jana *et al.*, "Design of Time-Frequency Packed WDM Superchannel Transmission Systems," Accepted in *J. Lightw. Technol.*, 2020.