Precoded Time-Frequency-Packed Multicarrier Faster-than-Nyquist Transmission

Mrinmoy Jana
ECE Dept., University of
British Columbia, BC, Canada
Email: mjana@ece.ubc.ca

Lutz Lampe
ECE Dept., University of
British Columbia, BC, Canada
Email: lampe@ece.ubc.ca

Jeebak Mitra
Huawei Technologies
Ottawa, ON, Canada
Email: jeebak.mitra@huawei.com

***Abstract*—We consider time and frequency-packed (TFP) multicarrier faster-than-Nyquist (MFTN) transmission that realizes non-orthogonal linear modulation to achieve higher spectral efficiency (SE) compared to the well-known orthogonal transmission at Nyquist rate. Such SE benefits come at the price of inter-symbol interference (ISI) and inter-carrier interference (ICI), which usually are equalized through computationally demanding receiver-side processing. In this paper, we investigate an alternative approach of pre-equalizing the interference at the transmitter, for the first time in an MFTN system where symbols are packed in both time and frequency dimensions. First, we present a two-dimensional linear precoding technique to jointly mitigate the ISI and ICI at the transmitter. Second, we propose a partial precoding strategy to pre-equalize the ISI at the transmitter, while ICI being compensated at the receiver through an iterative interference cancellation method. We validate the flexibility of our designs by simulating a precoded coherent optical fiber communication system, as a practical application example for MFTN. Our numerical results confirm significant performance and complexity advantages of the proposed precoding schemes over existing interference mitigation approaches.**

***Index Terms*—Multicarrier faster-than-Nyquist (MFTN) transmission, time-frequency-packing (TFP), precoding, pre-equalization.**

I. Introduction

Data traffic for the next generation communication systems is increasing at a staggering pace. Such dramatic growth of data rates warrants improved bandwidth-efficient transmission schemes. One way to accomplish this is to employ multicarrier faster-than Nyquist (MFTN) transmission, also known as time-frequency packing (TFP) [1]–[7], which relinquishes the orthogonality condition imposed by the Nyquist criterion, by deliberately reducing the time and frequency spacing of the adjacent symbols. Transmitting at a faster-than-Nyquist (FTN) rate theoretically provides a higher achievable rate [1]. From a practical implementation perspective, FTN signaling is advantageous for transmission systems such as the point-to-point microwave links [8] and longhaul optical fiber communication [2], [6], [7], [9], where the application of very high modulation formats to increase the spectral efficiency (SE) is challenging due to phase-noise and channel nonlinearities. Denser TFP schemes are also being considered in the context of new modulation formats for the cellular networks [5].

However, MFTN transmission introduces inter-symbol interference (ISI) and inter-carrier interference (ICI). Therefore, enjoying the SE benefits of the TFP systems entails successful mitigation of such interference. For this, a significant volume of work considers Bahl-Cocke-Jelinek-Raviv (BCJR) based maximum a-posteriori probability (MAP) equalization [1], [2], [6] or frequency domain equalization (FDE) techniques [10] for ISI mitigation, and/or other sub-optimal iterative ICI mitigation methods [1], [3], [7], [9], [11]. However, the complexity of such turbo-equalization schemes is substantial. Therefore, exploiting the fact that the MFTN interference is known at the transmitter, in this paper, we turn our attention to pre-equalization techniques that can significantly diminish or completely eliminate the computational burden from equalization at the receiver. Since the well-known Tomlinson-Harashima precoding (THP) manifests significant "precoding loss" and "modulo-loss" [12], particularly in an FTN transmission (see e.g. [13]), we consider linear precoding [8], [13] in this work. For our *first* contribution, we present a new two-dimensional (2-D) linear pre-equalization (LPE) technique, as an extension of the one-dimensional (1-D) LPE proposed in [13]. By orthogonalizing the FTN transmission through joint filtering of the constituent sub-channels (SCs) of an MFTN system, 2-D LPE yields optimal error-rate performance, which makes it competitive to computationally prohibitive and buffer-space constrained BCJR based equalization algorithms. However, such a precoding method is restrictive in terms of the time and frequency compression achievable in an MFTN transmission. To address this problem, as a *second* contribution, we propose a sub-optimal partial precoding (PP) strategy, which facilitates transmitter-side 1-D LPE precoding for the individual SCs of the TFP system, followed by a receiver-side turbo ICI cancellation. We validate the advantages of our proposed precoding schemes over the existing TFP interference mitigation methods, through numerical simulations of a coherent TFP optical wavelength division multiplexed (WDM) superchannel transmission, which has attracted significant attention more recently [2], [6], [7], [9].

II. System Model

We consider the baseband system model for precoded MFTN transmission under an additive white Gaussian noise (AWGN) channel shown in Fig. 1. For each k^{th} SC, $k \in 1, 2, \cdots, N$, with N being the total number of SCs, a low-density parity-check (LDPC) coded and modulated data stream a_k is either jointly or separately precoded by a linear feedback filter (FBF). The precoded signal d_k is then frequency shifted to produce x_k, which is converted to an analog signal and shaped by a root-raised cosine (RRC) pulse h with a roll-off

978-1-5386-6528-2/19/$31.00 ©2019 IEEE

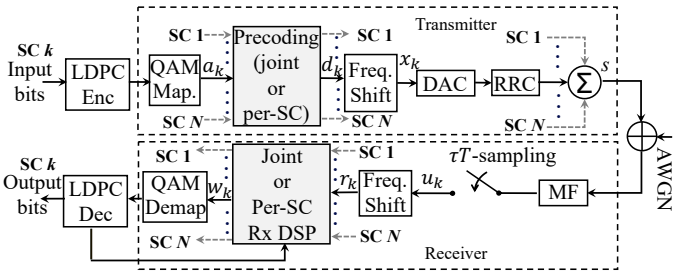
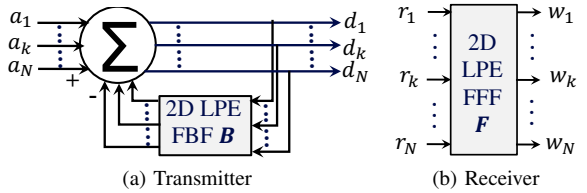


Fig. 1. Precoded-MFTN AWGN system model.



(a) Transmitter (b) Receiver

Fig. 2. 2-D LPE, where the shaded blocks represent additional signal processing compared to unprecoded MFTN systems.

factor β. The baseband equivalent aggregate signal s at the MFTN transmitter can be expressed as [9]

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

where $\Delta f = \xi \frac{1+\beta}{T}$ is the frequency-spacing between the adjacent SCs, with $0 < \tau \leq 1$ and $\xi > 0$ denoting the time and frequency compression ratios, respectively, such that $\tau = \xi = 1$ corresponds to Nyquist signaling, $\frac{1}{T}$ is the baud rate per SC, and l is the symbol index. At the receiver, the RRC matched-filtered and τT-sampled digital samples u_k of the k^{th} SC are frequency-shifted to produce the signal r_k that is jointly or separately processed by a feedforward filter (FFF). Thereafter, the digital samples are sent as inputs to the demapper and the LDPC decoder. For ease of characterization of the precoded TFP systems, we state the following proposition.

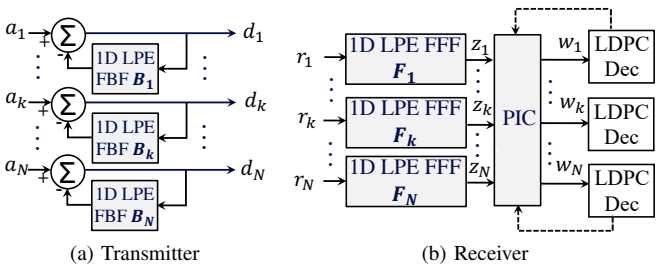
Proposition 1. *For the k^{th} SC, $k = 1, 2, \ldots, N$, shown in Fig. 1, frequency shifts of the precoded signal d_k at the transmitter and the τT-sampled signal u_k at the receiver by an amount $-\omega_0 \left(k - \frac{N+1}{2}\right)$ and $\omega_0 \left(k - \frac{N+1}{2}\right)$, respectively, where $\omega_0 = 2\pi\Delta f \tau T$, translates the overall TFP channel into a linear time-invariant (LTI) system, and the z-transform $\boldsymbol{H}(z)$ of the corresponding two-dimensional (2D) channel response is a Hermitian matrix polynomial.*

Proof: See Appendix A. ∎

III. Joint Precoding: 2-D LPE

Schematics of the 2-D LPE are shown in Fig. 2(a)-2(b) corresponding to the transmitter and receiver filtering operations, respectively. In Fig. 2(a), the modulated data symbols a_k, $k = 1, 2, \ldots, N$ from all SCs are jointly processed by a linear 2-D FBF $\boldsymbol{B}$ to produce the precoded symbols d_k, translating the overall precoding operation into an effective infinite impulse response (IIR) filter. The minimum-phase property of the FBF guarantees the stability of the IIR operation [12][1]. At

[1]We remark that different from a conventional ISI channel, such linear IIR filtering does not induce a precoding loss in an FTN system (see [13]), since the MFTN "channel" is part of the transmitter.



(a) Transmitter (b) Receiver

Fig. 3. Partial precoding, where the shaded blocks represent additional signal processing compared to unprecoded MFTN systems.

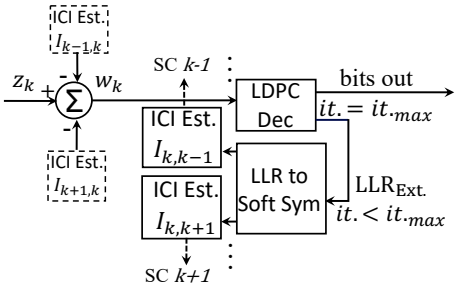


Fig. 4. ICI mitigation through PIC.

the receiver shown in Fig. 2(b), the frequency-shifted symbols r_k, $k = 1, 2, \ldots, N$ from all SCs are further jointly processed by a linear 2-D FFF $\boldsymbol{F}$, which whitens the colored noise caused by FTN sampling. Thereafter, the filtered samples are sent as inputs to a symbol-by-symbol demapper. Inspired by [14], the FFF and the FBF matrix computation from the 2-D channel matrix $\boldsymbol{H}$ defined in Proposition 1 is summarized below.

Proposition 2. *With the Cholesky decomposition of the Hermitian matrix polynomial $\boldsymbol{H}(z)$ performed as*

$$\boldsymbol{H}(z) = \boldsymbol{V}(z)\boldsymbol{V}^{\mathrm{H}}\left(z^{-*}\right) , \quad (2)$$

where $\boldsymbol{V}(z) = \sum_{k \geq 0} \boldsymbol{V}_k z^{-k}$ is causal and minimum-phase, i.e., $\boldsymbol{V}_k = \boldsymbol{0}$ for $k < 0$, $\boldsymbol{V}(z)$ is nonsingular for $|z| \geq 1$ and $\boldsymbol{V}_0$ is lower triangular, the 2-D LPE FFF and FBF are given by

$$\boldsymbol{F}(z) = \boldsymbol{D}^{-1}\boldsymbol{J}\boldsymbol{V}^{-1}\left(z^{-*}\right)\boldsymbol{J} , \quad (3)$$

$$\boldsymbol{B}(z) = \boldsymbol{D}^{-1}\boldsymbol{J}\boldsymbol{V}^{\mathrm{H}}\left(z^{*}\right)\boldsymbol{J} , \quad (4)$$

respectively, with $\boldsymbol{J}$ being the $K \times K$ anti-diagonal identity matrix, K is the number of TFP-channel taps, $(\cdot)^*$ represents the complex conjugate of a complex scalar, $(\cdot)^{-*} = \frac{1}{(\cdot)^*}$, $[\cdot]^{\mathrm{H}}$ and $[\cdot]^{-1}$ denote the matrix Hermitian and matrix inverse, respectively, $\boldsymbol{D} = \boldsymbol{diag}\left(v^*_{0,K,K}, \cdots, v^*_{0,1,1}\right)$, $v_{0,i,j}$ are the i^{th} and j^{th} entries of $\boldsymbol{V}_0$, and $\boldsymbol{diag}(\cdots)$ denotes the diagonal matrix constructed with the specified elements.

Proof: See Appendix B. ∎

By converting the MFTN transmission into an orthogonal system similar to [13], 2-D LPE yields optimal error-rate performance without performing any iterations between the equalizer and the LDPC decoder. However, such precoding suffers from the following two drawbacks: (a) for the joint filtering across SCs, it requires access to the digital samples of all SCs at both the transmitter and the receiver, which precludes the realization of independent SC processing, and (b) 2-D LPE is feasible only for a restricted range of τ, ξ pairs

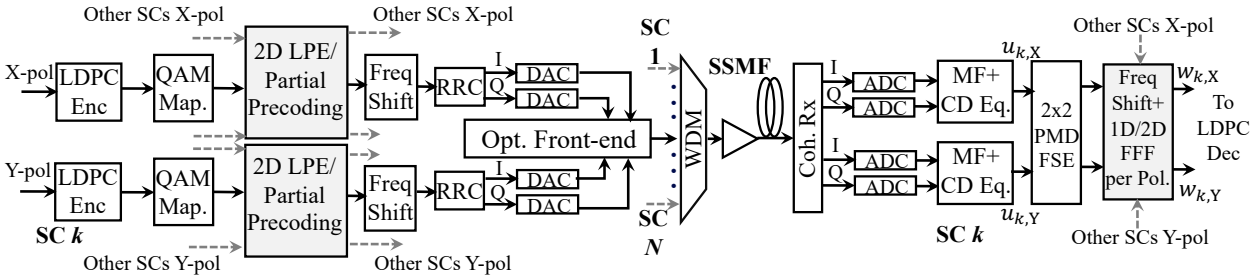


Fig. 5. Simulated MFTN system model: precoded DP TFP WDM optical superchannel transmission.

for a given β, since orthogonalizing the MFTN transmission requires that the net symbol rate is lower than the aggregate TFP bandwidth, and therefore,

$$\frac{N}{(1+\beta)\tau[(N-1)\xi+1]} \leq 1 \,. \tag{5}$$

Moreover, we also note that the computation of $\boldsymbol{B}$ and $\boldsymbol{F}$ requires the factorization (2) that is possible if the following Paley-Wiener condition (see e.g. [12], [15]) is satisfied

$$\tau T \int_{-\frac{1}{2\tau T}}^{\frac{1}{2\tau T}} \left| \log\det\left(\boldsymbol{H}\left(\mathrm{e}^{\mathrm{j}2\pi f \tau T}\right)\right)\right| df < \infty \,. \tag{6}$$

To address such limitations of the 2-D LPE, we propose another precoding strategy as follows.

IV. Partial Precoding (PP)

PP encapsulates a conceptual combination of transmitter-side pre-equalization and receiver-side equalization of TFP interference, as shown in Fig. 3(a)-3(b). In order to pre-mitigate the MFTN-ISI, PP employs separate 1-D LPE FBFs and FFFs for each SC, at the transmitter and receiver, respectively. For this, the FFFs and FBFs are computed based on the spectral factorization of the diagonal entries of the channel response $\boldsymbol{H}(z)$ (see e.g. [8], [13] for the computational details). Thereafter, the MFTN-ICI is mitigated at the receiver through an iterative parallel interference cancellation (PIC) approach in a turbo fashion as detailed in Fig. 4, similar to [1], [3], [9], [11].

As shown in Fig. 4, PIC enables the extrinsic log-likelihood-ratios (LLRs) fed back from the LDPC decoders to estimate and cancel the soft-estimates of the ICI stemming from the adjacent SCs, iteratively. For example, each LDPC iteration uses the extrinsic LLRs from the $(k-1)^{\text{th}}$ and $(k+1)^{\text{th}}$ SCs to compute the soft estimates of the data symbols corresponding to the neighboring SCs [9], [11]. Next, the soft-estimates $I_{k-1,k}$ and $I_{k+1,k}$ are computed and subtracted from the 1-D LPE FFF output symbols z_k of the k^{th} SC, where $I_{i,j}$ denotes the ICI from the i^{th} SC to the j^{th} SC, $i,j \in 1,2,\cdots,N$.

While PIC is well investigated in the FTN literature as a means to counter ICI, in this work, we apply it for the first time in tandem with precoding. With this design, PP based TFP systems completely eliminate the MFTN-ISI without performing computationally challenging BCJR iterations, and can also offer significant performance advantage over unprecoded PIC-only ISI and ICI equalization approach, such as [3]. We validate this claim through numerical simulations in Section V. We remark that for a given roll-off β, implementation of PP is feasible for the restricted range $\tau \geq \frac{1}{1+\beta}$ [13]. However, any amount of ξ can be accommodated through the PP implementation, which allows more flexible precoded TFP design compared to 2-D LPE. However, such benefits come at the expense of sub-optimal performance and iterative detection at the receiver that entails higher complexity and buffering.

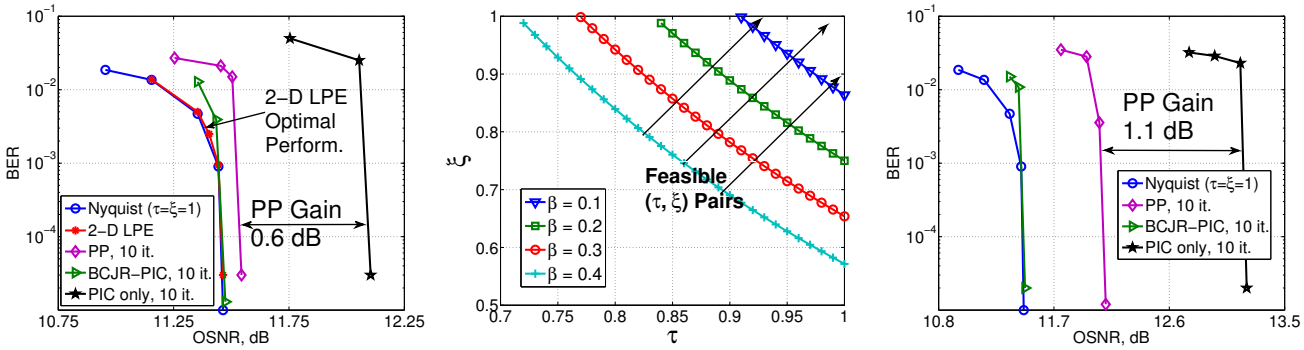
V. Simulation results

In this section, we validate the effectiveness of the proposed precoded MFTN designs by way of numerical simulations using parameter settings relevant for practical coherent optical WDM superchannel transmission, which is a prime candidate for the introduction of FTN. For the simulations, we consider a dual-polarized (DP) quarternary phase-shift keying (QPSK) 3-SC TFP WDM superchannel having per SC baud rate 40 Gbaud[2]. In the simulation setup shown in Fig. 5, the transmitter and receiver blocks for the discrete-time baseband modules are same as those in Fig. 1 except that the data processing for each of the two polarizations is performed separately for each SC. The baseband analog data after the digital-to-analog converter (DAC) is processed by the opto-electronic front-end and transmitted as an optical signal through a 1000 km standard single-mode fiber (SSMF) with chromatic dispersion (CD) parameter value -18 ps^2/km, polarization mode dispersion (PMD) 0.5 ps$/\sqrt{\text{km}}$, and then is received by the coherent optical receiver. LDPC codes from the DVB-S2 standard with rate 0.9 and codeword-length 64800 bits, $\beta = 0.3$, 8-tap TFP-ISI for BCJR, 20-tap LPE-FBF, 200-tap LPE-FFF, and maximum iteration count of 10 between the PIC and the LDPC decoder are considered for the simulations. Perfect frequency synchronization and phase-lock are assumed. Moreover, matched filtering at the receiver is combined with the time-invariant frequency domain CD compensator using overlap-and-add method, and for PMD compensation, we use a 19-tap 2×2 butterfly-type fractionally-spaced adaptive least-mean square (LMS) equalizer[3].

We first show the advantages of the 2-D LPE and PP in Fig. 6(a) by plotting the coded bit-error rate (BER) performance averaged over both polarizations and all SCs, as a function of the optical signal-to-noise ratio (OSNR) [6]. For reference, we also add the error rate curves in Fig 6(a)

[2]For additional information on fiber-optical channel and noise characterization, DP and WDM systems, interested readers are referred to [6], [7].

[3]For the PP TFP systems, independent SC processing enables us to employ the PMD equalizer after the 1-D LPE FFFs, similar to [13], where the LMS filter is trained with the known MFTN-interference induced pilots. However, the matrix filtering operations in the 2-D LPE requires the placement of the FFF before the PMD equalizer, for which the LMS update equations need to be modified compared to a conventional Nyquist WDM transmission [6], [7]. Such modification is detailed in Appendix C.



(a) BER vs. OSNR, $\beta=0.3, \tau=0.85, \xi=0.88$. (b) Feasible range of τ, ξ for 2-D LPE. (c) BER vs. OSNR, $\beta=0.3, \tau=0.8, \xi=0.9$.

Fig. 6. Numerical results. DP-QPSK, 3-SC WDM superchannels, showing the benefits and limitations of 2-D LPE and PP.

TABLE I
COMPLEXITY, MEMORY AND LATENCY PER CODEWORD

Method	Complexity	Memory	Latency
2-D LPE	$\mathcal{O}\left(N^2 N_\mathrm{f} L_\mathrm{w}\right)$	$\mathcal{O}(N L_\mathrm{b})$	$\mathcal{O}(L_\mathrm{b})$
PP	$\mathcal{O}\left((M+L_\mathrm{c}+N_\mathrm{f}) N I_\mathrm{m} L_\mathrm{w}\right)$	$\mathcal{O}(N L_\mathrm{b} I_\mathrm{m})$	$\mathcal{O}(L_\mathrm{b} I_\mathrm{m})$
PIC-Only	$\mathcal{O}\left((M+L_\mathrm{c}+L_\mathrm{s}) N I_\mathrm{m} L_\mathrm{w}\right)$	$\mathcal{O}(N L_\mathrm{b} I_\mathrm{m})$	$\mathcal{O}(L_\mathrm{b} I_\mathrm{m})$
BCJR-PIC	$\mathcal{O}\left((M^{\frac{L_\mathrm{s}}{2}}+L_\mathrm{c}) N I_\mathrm{m} L_\mathrm{w}\right)$	$\mathcal{O}(N L_\mathrm{b} I_\mathrm{m})$	$\mathcal{O}(L_\mathrm{b} I_\mathrm{m})$

corresponding to the following three scenarios: (a) Nyquist WDM transmission having the same baud rate and therefore, larger bandwidth, (b) BCJR based ISI equalization in conjunction with PIC for ICI mitigation as in [1], denoted by the legend "BCJR-PIC, 10 it." and (c) PIC based ISI and ICI cancellation as in [3], indicated by the label "PIC only, 10 it". As shown in the figure, 2-D LPE achieves similar performance as that of a Nyquist WDM system and an MFTN system employing BCJR-PIC, by successfully pre-equalizing the ISI and ICI completely. For this, 2-D LPE relies on simple non-iterative filtering operations, as opposed to the substantially complex and buffer-space constrained BCJR algorithm that is impractical especially for larger constellations. Fig. 6(a) also suggests that PP yields 0.6 dB performance improvement over the PIC-only receiver structure. Moreover, it produces sub-optimal performance compared to the 2-D LPE for this particular combination of τ and ξ that is well within the range specified by the inequality (5). However, as shown through the subsequent results, PP is more effective for stricter values of τ and ξ pairs, for which 2-D LPE precoding is infeasible.

In Fig. 6(b), we plot the range of τ and ξ where the spectral factorization (2) and thereby, 2-D LPE precoding is infeasible, for varying β. To numerically evaluate such range of values that does not satisfy (6), we observe the presence of spectral zeros in the overall TFP channel $H(z)$. Fig. 6(b) indicates that higher values of the RRC roll-off translates to a larger range of feasible τ and ξ values for the 2-D LPE. Furthermore, we note that the plots in the figure also correspond to the inequality (5). This means that the dimensionality and factorization constraints are equivalent for the considered precoded TFP systems.

Finally, to show the usefulness of PP in more detail, we deliberately choose a pair of time and frequency compression ratios in Fig. 6(c) such that (5) is violated for $\beta = 0.3$, and therefore, 2-D LPE can not be employed. Fig. 6(c) shows that the BCJR-PIC outperforms PP by 0.65 dB at the price of significantly higher complexity. However, under such transmission scenarios, PP offers 1.1 dB performance gains over PIC-only equalization scheme having similar computational cost. The details of the receiver complexity, latency and memory requirements for the different interference mitigation schemes are furnished in Table I, where M, L_s, L_c, N_f, L_b, I_m denote the modulation order, truncated ISI and ICI-taps length, 1D/2D LPE FFF taps length, LDPC codeword length in bits and the maximum turbo iteration count, respectively, and $L_\mathrm{w} = \frac{L_\mathrm{b}}{\log_2 M}$ is the number of modulated symbols corresponding to each codeword. Values of the above parameters considered for our simulations are mentioned at the beginning of this section. Benefits of the proposed precoded systems can be seen in the performance-complexity trade-off, through suitable precoding technique selection depending on the MFTN parameters.

VI. CONCLUSIONS

We presented two precoding approaches for the first time in MFTN systems that enable packing of symbols in *both* time and frequency dimensions. First, a matrix linear filtering based 2-D LPE precoding is proposed that performs joint processing of the SCs to completely eliminate TFP ISI and ICI, and thereby, accomplishes optimal error rate performance. However, functionality of such precoding is limited to a restricted range of time and frequency compression. Second, we presented PP that facilitates independent processing of SCs at the transmitter for mitigating ISI, but operates in an iterative fashion with the LDPC decoder, to eliminate ICI at the receiver. Simulation results for a DP QPSK TFP WDM optical superchannel suggests up to 1.1 dB performance gains by the proposed precoding techniques over existing interference mitigation methods having similar or significantly higher computational cost, buffer space and latency requirement.

ACKNOWLEDGEMENT

This work was supported by Huawei Canada and the Natural Sciences and Engineering Research Council of Canada (NSERC).

APPENDIX A
PROOF OF PROPOSITION 1

Projecting the received signal component onto the basis functions $h(t-l\tau T)\mathrm{e}^{\mathrm{j}2\pi\left(k-\frac{N+1}{2}\right)\Delta f t}$ as per [1], the resulting

matched filtered analog signal for the k^{th} SC, $k=1,2,\ldots,N$, can be written as

$$\hat{u}_k(t)=s(t)\mathrm{e}^{-\mathrm{j}2\pi(k-\frac{N+1}{2})\Delta f t}\star h(t)\,, \tag{7}$$

where $\star$ denotes linear convolution. Writing $u_k[n]$ as the τT samples of $\hat{u}_k(t)$ in (7), we get

$$u_k[n]=\sum_{m=1}^{N}\left(x_m[n]\mathrm{e}^{j\omega_0(m-k)n}\star g_{0,m-k}[n]\right), \tag{8}$$

where $g_{u,v}$ denotes τT samples of $f_u(t)\star f_v(t)$ with $f_u(t)=h(t)\mathrm{e}^{j2\pi u\Delta f t}$. Multiplying both sides of (8) by $\mathrm{e}^{j\omega_0(k-\frac{N+1}{2})n}$ for all $k=1,2,\ldots,N$, we obtain

$$r_k[n]=\sum_{m=1}^{N}\left(d_m[n]\star g_{k-\frac{N+1}{2},m-\frac{N+1}{2}}[n]\right)\,, \tag{9}$$

which shows that the frequency-shift operations convert the TFP transmission into an LTI system, with impulse responses given by $g_{k-\frac{N+1}{2},m-\frac{N+1}{2}}$, $k,m=1,2,\ldots,N$. Denoting by $\boldsymbol{H}(z)$ the z-transform of the 2-D channel, conjugate symmetry of the sequences $g_{k-\frac{N+1}{2},m-\frac{N+1}{2}}[n]$ implies $\boldsymbol{H}(z)=\boldsymbol{H}^{\mathrm{H}}(z^{-*})$, which completes the proof.

Appendix B
Proof of Proposition 2

Following the notations in [14], we define: $\boldsymbol{S}(z)=\boldsymbol{D}^{-1}\boldsymbol{J}\boldsymbol{V}^{H}(z^{*})\boldsymbol{J}$, $\boldsymbol{\Sigma}=\boldsymbol{D}^{\mathrm{H}}\boldsymbol{D}$, and $\boldsymbol{M}(z)=\boldsymbol{J}\boldsymbol{H}^{\mathrm{H}}(z^{*})\boldsymbol{J}$. Considering the conjugate symmetry of the overall impulse response, we have $\boldsymbol{M}(z)=\boldsymbol{H}(z)$. Therefore, we can write:

$$\boldsymbol{S}^{\mathrm{H}}(z^{-*})\boldsymbol{\Sigma}\boldsymbol{S}(z)=\boldsymbol{J}\boldsymbol{V}(z^{-1})\boldsymbol{V}^{\mathrm{H}}(z^{*})\boldsymbol{J} \tag{10}$$

$$=\boldsymbol{J}\boldsymbol{H}(z^{-1})\boldsymbol{J}=\boldsymbol{H}^{\mathrm{H}}(z^{-*})=\boldsymbol{H}(z). \tag{11}$$

Based on the above factorization, we obtain, as in [14],

$$\boldsymbol{F}(z)=\boldsymbol{\Sigma}^{-1}\boldsymbol{S}^{-\mathrm{H}}\left(z^{-*}\right), \tag{12}$$

$$\boldsymbol{B}(z)=\boldsymbol{S}(z)\,. \tag{13}$$

Substituting $\boldsymbol{S}(z)$ and $\boldsymbol{\Sigma}$ defined above produces the result.

Appendix C
2-D LPE PMD Equalizer LMS Algorithm

We consider a half-symbol spaced LMS equalizer for PMD mitigation [6]. Let us denote the 2×2 PMD compensating filter for the i^{th} SC, $i=1,2,\ldots,N$, by $\begin{bmatrix}c_{\mathrm{xx},i}[\nu,k] & c_{\mathrm{xy},i}[\nu,k]\\ c_{\mathrm{yx},i}[\nu,k] & c_{\mathrm{yy},i}[\nu,k]\end{bmatrix}$, where each entry of the matrix corresponds to the ν^{th} fractionally-spaced tap, $\nu=0,1,\ldots,N_{\mathrm{c}}-1$, at the k^{th} time index. Considering N_{f} static symbol-spaced taps for each filter-entry $f_{ij}[\mu]$, $i,j\in\{1,2,\ldots,N\}$, $\mu=0,1,\ldots,N_{\mathrm{f}}-1$, of the 2-D LPE static FFF matrix, we can write the X-pol and Y-pol outputs, respectively, for the i^{th} SC, $i=1,2,\ldots,N$, as

$$w_{i,\mathrm{X/Y}}[k]=\sum_{j=1}^{N}\boldsymbol{\mathcal{C}}^{\mathrm{H}}_{\mathrm{X/Y},j}[k]\,\tilde{\boldsymbol{U}}_j[k]\boldsymbol{\Upsilon}_{ij}\,, \tag{14}$$

where the subscript X/Y means "X respectively Y", $w_{i,\mathrm{X/Y}}$ and $u_{i,\mathrm{X/Y}}$ are the PMD equalizer input and the frequency-shift output for the i^{th} SC, respectively, shown in Fig. 5. Moreover, in (14):

$$\boldsymbol{\mathcal{C}}_{\mathrm{X},j}[k]=\left[\left\{c^{*}_{\mathrm{xx},i}[m,k]\right\}_{m=0}^{N_{\mathrm{c}}-1},\left\{c^{*}_{\mathrm{xy},i}[n,k]\right\}_{n=0}^{N_{\mathrm{c}}-1}\right]^{\mathrm{T}},$$

$$\boldsymbol{\mathcal{C}}_{\mathrm{Y},j}[k]=\left[\left\{c^{*}_{\mathrm{yx},i}[m,k]\right\}_{m=0}^{N_{\mathrm{c}}-1},\left\{c^{*}_{\mathrm{yy},i}[n,k]\right\}_{n=0}^{N_{\mathrm{c}}-1}\right]^{\mathrm{T}},$$

$$\tilde{\boldsymbol{U}}_j[k]=\left[\left\{\vartheta_j^{(\kappa)}[k]\right\}_{\kappa=0}^{N_{\mathrm{f}}-1}\right]\boldsymbol{P}_j\,,$$

$$\vartheta_j^{(\kappa)}[k]=\left[\left\{u_{i,\mathrm{X}}[k-2\kappa-\gamma]\right\}_{\gamma=0}^{N_{\mathrm{c}}-1},\left\{u_{i,\mathrm{Y}}[k-2\kappa-\gamma]\right\}_{\gamma=0}^{N_{\mathrm{c}}-1}\right]^{\mathrm{T}},$$

$$\boldsymbol{P}_j=\mathbf{diag}\Big(\underbrace{\mathrm{e}^{-\mathrm{j}\hat{\omega}_0 r_j k},\ldots,\mathrm{e}^{-\mathrm{j}\hat{\omega}_0 r_j(k-2(N_{\mathrm{f}}-1))}}_{N_{\mathrm{f}}}\Big),$$

$\hat{\omega}_0=\pi(1+\beta)\xi\tau$, $r_j=j-1-\frac{N-1}{2}$, $\boldsymbol{\Upsilon}_{ij}=\left[\left\{f_{i,j}[\mu]\right\}_{\mu=0}^{N_{\mathrm{f}}-1}\right]^{\mathrm{T}}$, where $[\cdot]^{\mathrm{T}}$ denotes the matrix transpose and the expression $\left\{x[j]\right\}_{j=N_1}^{N_2}$ denotes the row-vector $[x[N_1],\ldots,x[N_2]]$. Writing the error signals as

$$\varepsilon_{\mathrm{X/Y},i}[k]=u_{i,\mathrm{X/Y}}[k]-a_{i,\mathrm{X/Y}}[k-k_0]\,, \tag{15}$$

with k_0 being the decision delay, and $a_{i,\mathrm{X/Y}}$ being the modulated symbol for the i^{th} SC and the corresponding polarization, the LMS update equation for the i^{th} SC, $i=1,2,\ldots,N$, can be written as

$$\boldsymbol{\mathcal{C}}_{X/Y,i}[k+1]=\boldsymbol{\mathcal{C}}_{X/Y,i}[k]-\alpha\tilde{\boldsymbol{U}}_i[k]\sum_{l=1}^{N}\boldsymbol{\Upsilon}_{li}\varepsilon^{*}_{\mathrm{X/Y},l}[k]\,, \tag{16}$$

where $\alpha>0$ is the step-size parameter.

References

[1] F. Rusek, *Partial Response and Faster-than-Nyquist Signaling*, Ph.D. thesis, Lund University, 2007.

[2] G. Colavolpe, T. Foggi, A. Modenini, and A. Piemontese, "Faster-than-Nyquist and beyond: how to improve spectral efficiency by accepting interference," *Optics Express*, vol. 19, no. 27, pp. 26600–26609, December 2000.

[3] D. Dasalukunte, *Multicarrier Faster-than-Nyquist Signaling Transceivers*, Ph.D. thesis, Lund University, 2012.

[4] J. B. Anderson, F. Rusek, and V. Owall, "Faster-Than-Nyquist Signaling," *Proc. IEEE*, vol. 101, no. 8, pp. 1817–1830, 2013.

[5] P. Banelli, S. Buzzi, G. Colavolpe, A. Modenini, F. Rusek, and A. Ugolini, "Modulation formats and waveforms for 5G networks: Who will be the heir of OFDM?," *IEEE Commun. Mag.*, vol. 31, no. 6, pp. 80–93, Nov. 2014.

[6] 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.

[7] K.Shibahara, A. Masuda, S. Kawai, and M. Fukutoku, "Multi-stage Successive Interference Cancellation for Spectrally-Efficient Super-Nyquist Transmission," in *European Conf. on Opt. Commun. (ECOC)*, 2015, pp. 1–3.

[8] M. Jana, L. Lampe, and J. Mitra, "Dual-polarized Faster-than-Nyquist Transmission Using Higher-order Modulation Schemes," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5332–5345, 2018.

[9] M. Jana, L. Lampe, and J. Mitra, "Interference Cancellation for Time-Frequency Packed Super-Nyquist WDM Systems," *IEEE Photon. Technol. Lett.*, vol. 30, no. 24, pp. 2099–2102, 2018.

[10] S. Sugiura and L. Hanzo, "Frequency-Domain-Equalization-Aided Iterative Detection of Faster-than-Nyquist Signaling," *IEEE Trans. Veh. Technol.*, vol. 64, no. 5, pp. 2122–2128, May 2015.

[11] A. Barbieri, D. Fertonani, and G. Colavolpe, "Time-frequency Packing for Linear Modulations: Spectral Efficiency and Practical Detection Schemes," *IEEE Trans. on Commun.*, vol. 57, no. 10, pp. 2951–2959, 2009.

[12] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*, Wiley-IEEE Press, 2002.

[13] M. Jana, A. Medra, L. Lampe, and J. Mitra, "Pre-equalized Faster-than-Nyquist Transmission," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4406–4418, 2017.

[14] R.F.H. Fischer, "Sorted Spectral Factorization of Matrix Polynomials in MIMO Communications," *IEEE Trans. on Commun.*, vol. 53, no. 6, pp. 945–951, 2005.

[15] D. Youla and N. Kazanjian, "Bauer-Type Factorization of Positive Matrices and the Theory of Matrix Polynomials Orthogonal on the Unit Circle," *IEEE Trans. on Circuits and Sytems*, vol. 25, no. 2, pp. 57–69, 1978.