Optical Time-Frequency Packing: Principles, Design, Implementation, and Experimental Demonstration

Marco Secondini, Tommaso Foggì, Francesco Fresì, Gianluca Meloni, Fabio Cavaliere Member, IEEE, Giulio Colavolpe Senior Member, IEEE, Enrico Forestieri Member, IEEE, Luca Potì Member, IEEE, Roberto Sabella Senior Member, IEEE, and Giancarlo Prati Fellow, IEEE

Abstract—Time-frequency packing (TFP) transmission provides the highest achievable spectral efficiency with a constrained modulation format and detector complexity. In this work, the application of the TFP technique to fiber-optic systems is investigated and experimentally demonstrated. The main theoretical aspects, design guidelines, and implementation issues are discussed, focusing on those aspects which are peculiar to TFP systems. In particular, adaptive compensation of propagation impairments, matched filtering, and maximum a posteriori probability detection are obtained by a combination of a butterfly equalizer and four low-complexity parallel Bahl-Cocke-Jelinek-Raviv (BCJR) detectors. A novel algorithm that ensures adaptive equalization, channel estimation, and a proper distribution of tasks between the equalizer and BCJR detectors is proposed. A set of irregular low-density parity-check codes with different rates is designed to operate at low error rates and approach the spectral efficiency limit achievable by TFP at different signal-to-noise ratios. An experimental demonstration of the designed system is finally provided with five dual-polarization QPSK-modulated optical carriers, densely packed in a 100 GHz bandwidth, employing a recirculating loop to test the performance of the system at different transmission distances.

Index Terms—Time-frequency packing, faster-than-Nyquist signaling, information theory, optical fiber communication, coherent optical systems.

I. INTRODUCTION

Next generation optical systems will use coherent detection and advanced signal processing for enabling the transmission of extremely high bit rates. Currently deployed 100 Gb/s single-carrier systems typically operates on a 50 GHz grid spacing, employing quadrature phase-shift keying (QPSK) modulation with polarization multiplexing to meet the required 2 bit/s/Hz spectral efficiency (SE), with a potential reach of thousands of kilometers. This, considering the actual power and bandwidth limitations of typical fiber-optic links, still leaves a significant margin for improvement with respect to channel capacity. On the other hand, enabling multi-Tb/s-per-channel transmission on the same links will require to operate as close as possible to the Shannon limit, achieving a much higher SE and possibly adapting it to the available signal-to-noise ratio (SNR). While the use of coherent detection, digital signal processing (DSP), and soft-decoding forward error correction is not in question in long-haul systems, a few different options are being considered for the optical transport format. Besides a high SE, the selected format should also offer best performances in terms of energy efficiency, cost, and reliability, the complexity of the required DSP being one of the driving factors for all those issues. From a system point of view, the whole problem can be summarized as finding the best combination of modulation and coding that maximizes SE for a given SNR and constrained complexity. In optical communications, orthogonal signaling is typically adopted to ensure the absence of inter-symbol interference (ISI) and inter-carrier interference (ICI). For instance, both Nyquist wavelength-division multiplexing (WDM) and orthogonal frequency-division multiplexing (OFDM) solutions, whose performance and complexity are basically equivalent on the fiber-optic channel, employ orthogonal signaling. In both cases, the orthogonality condition sets a lower limit to time- and frequency-spacing (the Nyquist criterion), such that the achievable SE is limited by the number of levels of the underlying modulation format. In fact, higher SE requires higher-level modulation (e.g., 16-ary quadrature amplitude modulation (QAM)), with higher complexity and lower resilience to nonlinear effects.

Recently, a different approach has been proposed which, giving up the orthogonality condition, allows to overcome the Nyquist limit and achieve a higher SE with low-level modulations [4]–[8]. This time- and frequency-packing (TFP) approach is an extension of well known faster-than-Nyquist (FTN) signaling [9]. In FTN signaling, pulses can be packed closer than the Nyquist limit without performance degradation, provided that the minimum Euclidean distance of the system is not reduced and the optimum detector is employed (Mazo limit) [9]. Analogously, a closer packing can be achieved also in frequency domain without performance degradation (two dimensional Mazo limit [10]. In other words, by increasing signaling rate for a fixed pulse bandwidth (or, equivalently,
by reducing pulse bandwidth for a fixed signaling rate), some bandwidth resources are saved at the expense of introducing ISI. A similar approach has been experimentally demonstrated also in [11]. FTN, however, does not provide the best performance in terms of SE, and has a limit in the complexity of the required detector (which can be very high). On the other hand, TFP overcomes this limit and seeks the best solution by dividing the problem in three parts: i) set the desired input constellation (e.g., QPSK) and detector complexity; ii) find the optimum time- and frequency-spacing which provide the maximum achievable SE for the given input constellation and detector complexity; iii) select a proper code to approach as close as desired the achievable SE (information theory guarantees that such a code exists).

In this work, after introducing the theoretical aspects of the TFP approach, we discuss the design procedure and implementation of a TFP fiber-optic system and experimentally investigate its performance. Section II introduces the TFP approach and the basic concept of achievable SE for a mismatched decoder, whose maximization is the key aspect of TFP. In Section III we explain how to design a TFP system and find the optimum modulation parameters (time and frequency spacing) that maximize the achievable SE for a given transmitter and receiver complexity and a set of irregular low-density parity-check (LDPC) codes to practically approach the achievable SE. The practical implementation of a TFP fiber-optic system is discussed in Section VI focusing on the DSP part, which is the only one to require some modifications with respect to a standard WDM system employing coherent detection. The experimental demonstration of the designed TFP system is addressed in Section VII five closely-packed 40 GBd dual-polarization (DP) quaternary phase-shift keying (QPSK) channels are transmitted through a recirculating loop, keeping the net SE beyond the theoretical limit of Nyquist-WDM (4 bit/s/Hz) up to 6000 km; higher SEs are achieved at shorter distances by adapting the TFP configuration and code rate to the available OSNR, achieving a net SE of more than 7 bit/s/Hz (for DP-QPSK transmission) at a distance of 400 km. A discussion of the results is provided in Section VII and conclusions are finally drawn in Section VII.

II. TIME-FREQUENCY PACKING

In order to summarize the general ideas behind TFP, we refer here to an ideal dual-polarization AWGN channel. Rather than as a specific modulation format, TFP should be regarded as a design procedure for the optimization of a class of modulation formats—namely, multicarrier linear modulations, to which both Nyquist WDM and OFDM belong. Many communication systems employ this kind of modulation to encode information onto waveforms which can be practically generated and reliably transmitted through a given communication channel. The low-pass equivalent model of a generic linearly-modulated multicarrier system is schematically depicted in Fig. 1. All the equally-spaced carriers are linearly modulated with the same modulation format and shaping pulse

\[ p(t) = \mathcal{F}^{-1}\{P(f)\} \]

The complex envelope of the transmitted signal is

\[ x(t) = \sum_{k} x_k^{(t)} p(t - kT)e^{j2\pi Ft} \]  

(1)

where \( x_k^{(t)} \) is the transmitted symbol (a two-component vector, one per each polarization) on the \( \ell \)-th carrier at time \( kT \), \( T \) is the symbol time (or time spacing between adjacent symbols), \( F \) the frequency spacing between adjacent carriers, and, for simplicity, a perfect time and phase synchronization among the carriers is assumed. Signal (1) is corrupted by additive white Gaussian noise (AWGN) \( n(t) \) and demodulated by a bank of matched filters and symbol-time samplers. Denoting by \( x = \{x_k^{(t)}\} \) the set of transmitted symbols and \( y = \{y_k^{(t)}\} \) the set of channel outputs, the SE of the system (bit/s/Hz) is

\[ \eta = \frac{I(X;Y)}{FT} \]  

(2)

where \( I(X;Y) \) is the average mutual information rate (bit/symbol) between input and output [13], and \( 1/FT \) is the inverse of the time-frequency spacing product, which equals the number of symbols transmitted per second per Hertz. Typically, these modulation formats are designed to avoid both ISI and ICI by imposing proper orthogonality constraints on the employed waveforms (e.g., Nyquist-WDM or OFDM). This, in turn, poses a constraint on the pulse shape and sets a limit to the minimum time and frequency spacing between pulses (Nyquist limit). When this orthogonal signaling approach is employed, \( I(X;Y) = I(X_k^{(t)};Y_k^{(t)}) \), \( \forall k, \ell \) is achievable by a symbol-by-symbol detector and depends only on the modulation format and signal-to-noise ratio. On the other hand, the minimum value of the time-frequency spacing

1When employing single-user detectors, which is the case considered in this work, the actual phase and time shift between carriers is typically irrelevant and has a negligible impact on the achievable information rate, both in terms of linear cross-talk among carriers (which, as shown later, are only slightly overlapped), and of inter-channel nonlinearity [12]. In fact, the experimental demonstration of Section VII does not employ any phase locking or time synchronization among carriers.

2Here, \( \eta \) denotes the maximum rate per unit bandwidth at which information can be reliably transmitted through the channel, where maximization is performed over all possible encoding of information bits on transmitted symbols. In the following, we will also consider other two slightly different definitions of SE, where specific constraints on the detector or coding strategy are imposed. Moreover, upper-case letters denote random variables, while lower-case letters denote their realizations.

Figure 1. Low-pass equivalent model employed to design the TFP system.
product at denominator of (2) is set by the Nyquist limit and is \( FT = 1 \)—achievable, for instance, by using pulses with a rectangular spectrum \( F(f) = \sqrt{1/B} \text{rect}(f/B) \) of low-pass bandwidth \( B \) and setting \( F = 2B \) and \( T = 1/2B \), as in Nyquist-WDM. Thus, the only way to increase \( \eta \) is through the numerator of (2), by increasing the cardinality of the modulation alphabet. The main drawback of this approach is that a high SE is obtained at the expense of strict requirements on spectral shaping and complicated modulation formats. For DP-QPSK modulation, the upper limit is \( \eta = 4 \text{ bit/s/Hz} \).

On the contrary, if we give up the orthogonality condition, we have no constraints on the choice of \( p(t) \), \( F \), and \( T \). Thus, we can select a shaping pulse \( p(t) \) that is compatible with the available hardware components, and try to increase \( \eta \) by reducing the denominator \( FT \) below the Nyquist limit, without changing the modulation format. This way, however, we also introduce ICI and ISI and, therefore, reduce the numerator of (2), by increasing the cardinality of the modulation alphabet. The result obtained through (2) depends on the shaping pulse considered in (1), but is independent of its bandwidth \( B \). In practice, once the optimum normalized time and frequency spacing \( BT \) and \( F/B \) are found through (2), one transmission parameter (e.g., the bandwidth \( B \)) can be set arbitrarily, while the other two (e.g., \( F \) and \( T \)) are scaled accordingly to preserve the optimized \( BT \) and \( F/T \) values.

The last step of the TFP method is common to almost any digital communication system and consists in finding a coding strategy that, by properly encoding information bits on transmitted symbols \( \{ x_k^{(0)} \} \), operates as close as possible to (2)—information theory guarantees that such a code does exist. Though similar coding strategies can be adopted in TFP and orthogonal signaling, this step has some peculiarities related to the presence of ISI and ICI which will be discussed in the next section.

### III. System Design

In this section, we show how to design a multicarrier fiber-optic system by employing the TFP approach described in the previous section. We refer, again, to the ideal low-pass equivalent scheme reported in Fig. 1, which, under some assumptions, is a reasonable representation of the fiber-optic channel and consider a DP-QPSK modulation alphabet. A sequence of i.i.d. symbols \( \{ x_k^{(0)} \} \), drawn from a DP-QPSK alphabet, modulates the selected carrier \( (k = 0, \text{ in the scheme}) \) at rate \( 1/T \) with a real shaping pulse \( p(t) \). All the modulated carriers are then combined and transmitted through an AWGN channel; it is a lower bound to the mutual information rate \( I(X; Y) \) on the real channel; it is achievable by the maximum \( a \ posteriori \) probability (MAP) detector designed for the selected auxiliary channel; and it can be simply evaluated through simulations (15). The auxiliary channel, though arbitrary, is conveniently chosen as the one providing the best trade-off between performance and complexity: the closer the auxiliary channel to the real channel, the higher is the AIR \( \hat{I}(X^{(0)}; Y^{(0)}) \) (and closer to \( I(X; Y) \)); the simpler the auxiliary channel, the simpler is the MAP detector required to achieve \( \hat{I}(X^{(0)}; Y^{(0)}) \). Finally, time and frequency spacing are optimized by maximizing the achievable SE with the selected detector

\[
\hat{\eta}_{\text{max}} = \max_{F,T>0} \frac{\hat{I}(X^{(0)}; Y^{(0)})}{FT} \leq \eta \tag{4}
\]

Optimization (4) is the very essence of TFP, which distinguishes it from FTN or other non-orthogonal signaling techniques. The result obtained through (4) depends on the shaping pulse considered in (1), but is independent of its bandwidth \( B \). In practice, once the optimum normalized time and frequency spacing \( BT \) and \( F/B \) are found through (4), one transmission parameter (e.g., the bandwidth \( B \)) can be set arbitrarily, while the other two (e.g., \( F \) and \( T \)) are scaled accordingly to preserve the optimized \( BT \) and \( F/T \) values.

The maximum achievable SE depends also on the given SNR (it increases as the SNR increases). However, the optimum \( F \) and \( T \) depend only slightly on it, such that a single optimization can be adopted for a wide range of SNRs (i.e., of link distances).

The last step of the TFP method is common to almost any digital communication system and consists in finding a coding strategy that, by properly encoding information bits on transmitted symbols \( \{ x_k^{(0)} \} \), operates as close as possible to (2)—information theory guarantees that such a code does exist. Though similar coding strategies can be adopted in TFP and orthogonal signaling, this step has some peculiarities related to the presence of ISI and ICI which will be discussed in the next section.

### III. System Design

In this section, we show how to design a multicarrier fiber-optic system by employing the TFP approach described in the previous section. We refer, again, to the ideal low-pass equivalent scheme reported in Fig. 1, which, under some assumptions, is a reasonable representation of the fiber-optic channel and consider a DP-QPSK modulation alphabet. A sequence of i.i.d. symbols \( \{ x_k^{(0)} \} \), drawn from a DP-QPSK alphabet, modulates the selected carrier \( (k = 0, \text{ in the scheme}) \) at rate \( 1/T \) with a real shaping pulse \( p(t) \). All the modulated carriers are then combined and transmitted through an AWGN channel; it is a lower bound to the mutual information rate \( I(X; Y) \) on the real channel; it is achievable by the maximum \( a \ posteriori \) probability (MAP) detector designed for the selected auxiliary channel; and it can be simply evaluated through simulations (15). The auxiliary channel, though arbitrary, is conveniently chosen as the one providing the best trade-off between performance and complexity: the closer the auxiliary channel to the real channel, the higher is the AIR \( \hat{I}(X^{(0)}; Y^{(0)}) \) (and closer to \( I(X; Y) \)); the simpler the auxiliary channel, the simpler is the MAP detector required to achieve \( \hat{I}(X^{(0)}; Y^{(0)}) \). Finally, time and frequency spacing are optimized by maximizing the achievable SE with the selected detector

\[
\hat{\eta}_{\text{max}} = \max_{F,T>0} \frac{\hat{I}(X^{(0)}; Y^{(0)})}{FT} \leq \eta \tag{4}
\]

Optimization (4) is the very essence of TFP, which distinguishes it from FTN or other non-orthogonal signaling techniques. The result obtained through (4) depends on the shaping pulse considered in (1), but is independent of its bandwidth \( B \). In practice, once the optimum normalized time and frequency spacing \( BT \) and \( F/B \) are found through (4), one transmission parameter (e.g., the bandwidth \( B \)) can be set arbitrarily, while the other two (e.g., \( F \) and \( T \)) are scaled accordingly to preserve the optimized \( BT \) and \( F/T \) values.

The maximum achievable SE depends also on the given SNR (it increases as the SNR increases). However, the optimum \( F \) and \( T \) depend only slightly on it, such that a single optimization can be adopted for a wide range of SNRs (i.e., of link distances).

The last step of the TFP method is common to almost any digital communication system and consists in finding a coding strategy that, by properly encoding information bits on transmitted symbols \( \{ x_k^{(0)} \} \), operates as close as possible to (2)—information theory guarantees that such a code does exist. Though similar coding strategies can be adopted in TFP and orthogonal signaling, this step has some peculiarities related to the presence of ISI and ICI which will be discussed in the next section.
channel with noise $\mathbf{u}(t)$. At receiver side, the selected carrier is demodulated by a matched filter and a symbol-time sampler. Received samples $\{y_k\}$ are finally sent to a MAP symbol detector that operates on the output of the matched filter \cite{BCJR} and is implemented through the algorithm by Bahl, Cocke, Jelinek, and Raviv (BCJR) \cite{BCJR}. The MAP symbol detector is matched to an auxiliary channel, whose selection determines the distributions $q(y^{(0)}|x^{(0)})$ and $q(y^{(0)})$ to be used in \cite{BCJR}. In particular, as an auxiliary channel we take an approximation of the real channel, obtained from the latter by neglecting ICI, truncating ISI to the first $L_T \leq L$ pre- and post-cursor symbols (after the matched filter) —where $2L+1$ is the actual memory of the channel and $L_T$ is a design parameter strictly related to detector complexity —and increasing the noise variance $\sigma_n^2 > \sigma_n^2$ (up to a numerically optimized value) to account for the neglected ICI and ISI. This choice provides a reasonable trade-off between performance (how tight is the bound in \cite{BCJR}) and complexity (of the matched BCJR detector). Moreover, as the DP-QPSK modulation can be seen as the combination of four orthogonal BPSK modulations (one per each quadrature component of each state of polarization of the signal), four independent and identical BCJR detectors with $2L_T$ states are used to separately detect the four BPSK components.

Although also the pulse shape $p(t)$ can be optimized to maximize the SE \cite{19}, in this work we consider only the pulse shape obtained by employing a ninth-order type I Chebyshev filter (selected among the electrical low-pass filters available in our laboratory) with 3 dB bandwidth $B$. For the given shape, \cite{3} is evaluated through numerical simulations as explained in \cite{15} on a grid of values of the normalized time and frequency spacing $TB$ and $F/B$, seeking the maximum SE \cite{4} and the corresponding optimum spacings. To account for unsynchronized channels and unlocked lasers, each modulated carrier is also subject to a random phase and time shift and polarization rotation. The optimization is performed considering a truncated channel memory $L_T = 3$ (8-state MAP symbol detectors) and two different values of the SNR per bit (defined as the ratio $E_b/N_0$ between the mean energy per bit and the noise power spectral density and related to the OSNR through $\text{OSNR} = R_bE_b/(2N_0B_{\text{ref}})$, where $R_b$ is the total net bit rate and $B_{\text{ref}} \simeq 12.5$ GHz the conventional reference bandwidth) of 7.5 and 22.5 dB. The corresponding contour plots of the achievable SE are reported in Fig. 2(a) and (b).

In principle, the TFP optimization procedure described here can be applied also to a realistic fiber-optic channel, as the AIR definition \cite{3}, its properties, and the simulation-based method for its computation \cite{15} are valid for any channel. The only requirement is that of computing the output sequence $\{y_k\}$ for the desired real channel (e.g., through the split-step Fourier method). This, however, significantly increases the computation time required to estimate a single AIR value and makes the optimization procedure cumbersome. For this reason, we decided to optimize the system in the absence of nonlinear effects, and then tested the obtained suboptimum configuration (the one in Fig. 2) over a realistic link. A numerical estimate of the achievable SE with the suboptimum configuration over the realistic link (including nonlinear effects) is reported in Section VI (Fig. 12) and compared to experimental results.

The achievable SE obtained with this design procedure can be practically approached by employing properly designed codes. When the TFP technique is adopted, and thus ISI is intentionally introduced, codes designed for the AWGN channel no longer perform satisfactorily. So a redesign is required. We designed proper LDPC codes specifically tailored for the ISI channels resulting from the adoption of the TFP technique. The adopted procedure is based on two steps. The heuristic technique for the optimization of the degree distributions of the LDPC variable and check nodes proposed in \cite{20} is first adopted. This technique consists of a curve fitting on extrinsic information transfer (EXIT) charts, is based on a Gaussian assumption on all messages involved in the iterative process, and is much simpler than other optimization techniques, such as density evolution, which require intensive computational efforts. The parameters of the designed codes are reported in Table I where $r$ denotes the rate of the code and the degree distributions of variable and check nodes are provided by using the notation in \cite{21}. In any case, the codeword length is $N = 64800$.

Once the degree distributions of the LDPC variable and check nodes have been designed, the parity check matrix of an LDPC code with those degree distributions is built through the very effective PEG algorithm \cite{22}, \cite{23}, which allow to design an LDPC code whose underlying Tanner graph has a large girth. The BER curves for uncoded QPSK transmission and for the designed LDPC codes (independent encoding of the in-phase and quadrature components of each polarization) obtained through numerical simulations for the back-to-back system with the TFP configuration adopted in the experimental setup (constrained optimum at 40 Gb/s) are reported in Fig. 3. With this TFP configuration, the 8/9 LDPC code requires $E_b/N_0 \simeq 9.3$ dB and provides an SE of about 7.1 bit/s/Hz. Thus, we can compare it with Fig. 2(a), which shows that approximately the same SE (7.2 bit/s/Hz) is theoretically achievable at $E_b/N_0 = 7.5$ dB. This means that the designed code over the (back-to-back) TFP channel has a penalty of less than 2 dB with respect to the theoretical limit provided by the AIR in \cite{1}. The gap between the actual rates achieved by the designed codes over the fiber-optic channel (including nonlinear effects) and the AIR over the same channel is numerically investigated in Section VI (Fig. 12) and is between 2 and 3 dB for all the codes. Detection of this kind of codes is typically characterized by the presence of error floors at high SNRs. In our simulations, we transmitted up to 10000 codewords without observing any floor, meaning that error floors, if present, are probably located at a BER lower than $10^{-8}$. In any case, outer hard-decision external codes with very low overhead can be employed to correct the residual errors and remove the floor. For instance, in the DVB-S2 standard, where LDPC codes with same length and rates as in Tab. I are adopted, outer BCH codes with less than 0.4% overhead are used to correct from 8 to 12 residual errors (depending on the rate) per codeword \cite{24}. 


Figure 2. Contour plots of the achievable SE (obtained by numerical simulations and shown with increments of 0.2) as a function of the normalized time and frequency spacing for DP-QPSK modulation on the AWGN channel for: (a) $E_b/N_0 = 7.5\, \text{dB}$; (b) $E_b/N_0 = 22.5\, \text{dB}$. The maximum value (+) and the value obtained with one of the configurations adopted in the experimental setup (⊙) are also reported at the corresponding coordinates.

| $r$ | variable node distribution | check node distribution |
|-----|----------------------------|-------------------------|
| $r = 8/9$ | $0.333318x + 0.6x^2 + 0.0666621x^3$ | $0.000277778x^2 + 0.9989335x^2 + 0.000787037x^4$ |
| $r = 5/6$ | $0.249985x + 0.666662x^2 + 0.083333x^4$ | $0.000679012x^4 + 0.99858x^4 + 0.000740741x^4$ |
| $r = 4/5$ | $0.2x + 0.699985x^2 + 0.100015x^4$ | $0.999383x^4 + 0.000617284x^4$ |
| $r = 3/4$ | $1.54321 \cdot 10^{-5} + 0.166661x + 0.75x^2 + 0.083339x^4$ | $9.25926 \cdot 10^{-5} + 0.999907x^4$ |
| $r = 2/3$ | $0.111696x + 0.777779x^2 + 0.111111x^3$ | $0.999861x^6 + 0.000138889x^6$ |

Table I

**CODE RATES AND DEGREE DISTRIBUTIONS OF THE DESIGNED LDPC CODES.**

**IV. SYSTEM IMPLEMENTATION**

A fiber-optic system based on the TFP approach can be implemented by using the same hardware configuration typically used for WDM systems based on coherent detection [25]. A significant difference is in the DSP algorithms actually required at the receiver. Moreover, some care should be taken to ensure that the transmitted signal is linearly modulated as in [1]. In this section, we will refer to the transmitter and receiver implementation schemes shown in Fig. 4, focusing on those elements which are peculiar to the TFP implementation. Practical details about the experimental setup actually employed in the experimental demonstration will be given in Section [1].

Since the system employs single-user detectors, an independent transmitter and receiver pair is used per each optical carrier. Each optical carrier is thus generated at the desired wavelength (e.g., by an external-cavity laser (ECL)), modulated, optically multiplexed with the other modulated carriers, transmitted through the optical link, extracted by an optical demultiplexer, and independently detected. In each transmitter, the in-phase and quadrature components of two orthogonal states of polarization are independently and linearly modulated by a pair of nested Mach-Zender modulators (MZMs). In principle, the desired pulse shape $p(t)$ can be obtained either operating on the electrical signals that drive the modulator (through a low-pass filter (LPF), as actually shown in the scheme of Fig. 4) or on the optical signal after the modulator (through an optical band-pass filter), provided that the overall equivalent low-pass impulse response of the transmitter (driver, modulator, electrical filter and/or optical filter) is $p(t)$ and that linearity of the modulator is preserved by employing a driving voltage significantly lower than the modulator half-wave voltage $V_{ph}$. A comparison of the performance obtained by optical or electrical filtering when employing different driving voltages is presented in [26]. An alternative modulation scheme, where linear modulation [1] is obtained by operating the MZM at its maximum driving voltage (to reduce its insertion loss), may also be devised. For instance, by using an additional MZM as a pulse carver and employing...
optical filtering to obtain the desired pulse shape $p(t)$ and $BT$ product $[5, 8]$, the nonlinearity of the MZM affects only (and slightly) the overall pulse shape $p(t)$, but does not introduce nonlinear ISI. In this scheme, however, the insertion loss saved by increasing the MZM driving voltage is replaced by the additional loss introduced by the pulse carver and optical filter. Finally, a possible implementation based on an arrayed waveguide grating device that filters and multiplexes all the frequency subchannels is in the optical domain has been proposed in $[27]$. In this work, we consider a modulation scheme based on a single MZM (driven at low voltage) and analogue electrical filtering which, at the present, seems to be the most practical choice in terms of cost and complexity. Moreover, as discussed in Section 3, the choice of $p(t)$ is not critical and a reasonably good performance can be obtained by employing available analogue low-pass filters, as shown in Fig. 2.

At the receiver side, each optical carrier is demodulated by employing a phase- and polarization-diversity coherent detection scheme. After optical demultiplexing, each carrier is split into two orthogonal states of polarizations, which are then separately combined with the optical field of a local oscillator (LO) laser in a $2 \times 4 90^\circ$ optical hybrid and detected with two pairs of balanced photodetectors. The four resulting electrical signals (the in-phase and quadrature components of each state of polarization) are then sampled by an analog-to-digital converter (ADC) with a bandwidth $B$ at least equal to the (low-pass) bandwidth of the shaping pulse $p(t)$ and a sampling rate of $2B$. The remaining part of receiver processing is digitally implemented according to the scheme depicted in Fig. 3, assuming a sampling rate of $1/T$.

Note that, since TFP is employed, the required bandwidth and sampling rate are typically lower than $1/(2T)$ and $1/T$, respectively, and digital upsampling can be employed to achieve the $1/T$ rate required for symbol-time processing, without any performance degradation. The $k$-th received column vector of samples $r_k = (r_{1,k}, r_{2,k})^T$ (one complex sample per state of polarization) is first processed to compensate for the presence of any large and slowly varying frequency offset $f_o$ between the transmit and receive lasers. The estimate $f_o$ is obtained during the training phase (on a known training sequence) by employing the frequency estimation algorithm described in $[28]$, and then slowly updated based on decisions. Compensated samples $s_k = r_k e^{-j2\pi f_o k T}$ are then processed by an adaptive 2-D $N_c$-tap synchronous feed-forward equalizer (FFE) that compensates for linear propagation impairments, such as group-velocity dispersion (GVD), polarization rotations, and polarization-mode dispersion (PMD), and completes (as explained later) the implementation of the matched filter. At the output of the equalizer, the components $z_{1,k}$ and $z_{2,k}$ of the equalized samples $z_k$ are then separated and independently processed. For each component, decision-directed carrier phase estimation (CPE) based on the Tikhonov parametrization algorithm $[29]$ and taking into account ISI is employed to cope with the laser phase noise. Finally, the in-phase and quadrature components of the compensated samples $y_{1,k} = z_{1,k} e^{-j\hat{\theta}_{1,k}}$ and $y_{2,k} = z_{2,k} e^{-j\hat{\theta}_{2,k}}$ are separated and sent to four parallel $2^{L_{r}}$-state BCJR detectors $[18]$, followed by four LDPC decoders. The BCJR detectors and LDPC decoders iteratively exchange information to achieve MAP detection according to the turbo principle $[30]$. At each iteration, as new (more accurate) preliminary decisions are available from the decoders, the CPEs update the phase estimates $\hat{\theta}_{1,k}$ and $\hat{\theta}_{2,k}$ and a new set of compensated samples $y_{1,k}$ and $y_{2,k}$ is fed to the BCJR detectors. At the first iteration, as preliminary decisions are not available, the CPEs exploit pilot symbols (evenly inserted in the transmitted sequence at rate $r_p$) to provide a rough initial estimate of the phase and make the iterative process bootstrap.

The equalizer should be configured to make the low-pass equivalent model of the system as close as possible to the ideal scheme considered in Fig. 1. Considering that the amplified-spontaneous-emission (ASE) noise accumulated during propagation can be modeled as independent AWGN on each polarization at the input (or, equivalently, at the output) of the fiber, and that the transfer matrix of the fiber $H_f(f)$ (in the linear regime) is unitary, i.e., $H_f(f)^{-1} = H_f(f)^\dagger$, the required transfer matrix of the 2D-FFE equalizer should be

$$H_{\text{eq}}(f) = H_f(f)^\dagger P(f)^* / H_{te}(f)$$  \hspace{1cm} (5)
where $H_{fe}(f)$ is the low-pass equivalent transfer function of the optoelectronic front-end (optical filter, photodetector, and ADC). In this case, the corresponding overall channel transfer matrix would be $H(f) = |P(f)|^2 I$, with $I$ the 2 × 2 identity matrix, independently of the actual transfer matrix of the fiber. The evaluation of (5) for the system at hand requires an accurate characterization of transmitter and receiver front-end, and an adaptive estimate of the fiber transfer matrix $H_f(f)$.

Here, instead, taking inspiration from [16], an algorithm has been devised that configures the equalizer according to (5), without requiring a separate knowledge of $H_f(f), P(f)$, and $H_{fe}(f)$. Denoting by $C_i$ the 2 × 2 matrix of coefficients of the $i$-th tap of the equalizer, the equalized samples are

$$
z_k = \sum_{i=0}^{N_e-1} C_i s_{k-i}$$

(6)

Denoting by $x_k$ the $k$-th column vector of transmitted symbols, by $h_k$ the column vector of the two $i$-th coefficients of the desired (but unknown) overall impulse responses at the output of the matched filter (one per polarization), and by $g_k = (e^{j\theta_1 k}, e^{j\theta_2 k})^T$ the column vector of the phase estimates for the $k$-th samples on the two polarizations, the error with respect to the desired channel response is

$$e_k = g_k^* \circ z_k - \sum_{i=1}^{L} h_i \circ x_{k-i}$$

(7)

where $\circ$ denotes the Hadamard (entrywise) product. As shown in [16], the variance of each element of (7) is minimum when the matched filter condition is met, i.e., when $H_{fe}(f) H_f(f) H_{eq}(f) = P_s(f) I$. Given the unitarity of $H_f(f)$, this is equivalent to (5) and provides the desired overall response $H(f) = |P(f)|^2 I$. Thus, both the required equalizer coefficients and the desired channel coefficients of the Ungerboeck observation model can be simultaneously estimated by an iterative data-aided stochastic-gradient algorithm that minimizes the variance of (7). By holding $h_0$ constant (to an arbitrary value) and forcing the symmetry condition $h_{-i} = h_i^*$, the update law for the equalizer coefficients and the estimated channel coefficients are, respectively

$$C_i^{(k+1)} = C_i^{(k)} - \alpha_c (g_k \circ e_k) s_{k-i}^*, \quad 0 \leq i \leq N_e - 1$$

(8)

$$h_i^{(k+1)} = h_i^{(k)} + \alpha_h (e_k \circ x_{k-i}^* + e_k^* \circ x_{k+i}), \quad 1 \leq i \leq L_T$$

(9)

where $\alpha_c$ and $\alpha_h$ are the step-size gains. Updates (8) and (9) require knowledge of the transmitted symbols. While the equalizer coefficients need to be continuously updated to track variations of the fiber-optic channel, coefficients $\{h_i\}$ of the overall channel response do not change with time and can be estimated only once when setting up the link. The initial convergence of the algorithm can be guaranteed by the use of a known training sequence, while a slow tracking of the fiber channel can be achieved by updating only the equalizer coefficients according to (8), possibly at a much lower rate than $1/T$ and with a significant delay. This allows to use pilot symbols and/or to replace transmitted symbols with final decisions (after successful decoding of the whole codeword), with a negligible impact on information rate and performance.

The computation of the channel metric for the BCJR algorithm requires knowledge of the channel coefficients $\{h_k\}$ and of the noise variance. Thus, once estimated by (9), channel coefficients are passed to the BCJR processing blocks together with an estimate of the variance of (7).

V. EXPERIMENTAL DEMONSTRATION

Fig. 3 shows the experimental setup employed for the practical implementation of the TFP system and for the transmission experiments. Five external-cavity lasers (ECL) are grouped into two sets (odd and even channels), which are separately modulated by means of two integrated nested Mach-Zehnder modulators (IQ-MZM). Bandwidth, rate, and spacing of the five TFP channels are optimized under some constraints posed by the available hardware according to the design procedure described in Section III to maximize the achievable SE with the desired detector complexity. In particular, the optical carrier spacing is set to $F = 20 \text{GHz}$ and the binary electrical signals that drive the in-phase (I) and quadrature (Q) port of each IQ-MZM are modulated at a rate $R = 1/T = 40 \text{GBd}$ and filtered by a ninth-order Chebyshev low-pass filter (LPF) with a bandwidth $B = 10 \text{GHz}$ (cut-off frequency). The peak-to-peak modulation voltage is set to $V_{pp} = 1.5 \text{V}$, while the half-wave voltage of each MZM is $V_{\pi} = 2.8 \text{V}$. For the same fixed bandwidth and spacing, lower transmission rates of 35 and 30 GBd are also considered. Polarization multiplexing is emulated by means of a 50/50 beam splitter, an optical delay line, and a polarization beam combiner (PBC). Each I and Q component is modulated by a sequence of random information bits, which are independently encoded according to one of the LDPC codes reported in Tab. 1. Odd and even channels are then combined by means of a $2 \times 1$ optical coupler (OC). The optical spectrum of the transmitted TFP superchannel (at the input of the recirculating loop) is depicted in Fig. 4.

At the receiver side, one of the five TFP channels is detected by employing coherent phase- and polarization-diversity detection and setting the local oscillator (LO) at the nominal wavelength of the selected channel. The received optical signal is mixed with the LO through a polarization-diversity $90^\circ$ hybrid optical coupler, whose outputs are sent to four couples of balanced photodiodes. The four photodetected signals are sampled and digitized through a 20 GHz 50 GSa/s real-time oscilloscope in separate blocks of one million samples at a time, corresponding to about 12 codewords (at 40 GBd) per each quadrature component. After digital resampling at rate $1/T$ (one sample per symbol), each block is processed off-line according to the scheme of Fig. 5 with $N_e = 23$ equalizer

---

4As variations of the fiber-optic channel typically take place on a time scale of milliseconds, the channel remains approximately constant over many consecutive codewords.

5This configuration is twice more “packed” than allowed by the Nyquist limit: its time-frequency spacing product is $FT = 0.5$, meaning that two QPSK symbols are transmitted per second per Hertz per polarization.
taps and $L_T = 3$ truncated channel memory (8-state BCJR detector). The first received codeword of each block (on each quadrature component) is used as a training sequence for the convergence of the DSP algorithms (initial estimate of the frequency offset $\hat{f}_0$, equalizer coefficients $C_i$, and channel coefficients $h_i$), while the others are effectively employed to measure system performance. This is not considered in the computation of the SE as, in a real system, the training sequence would be transmitted only once. After decoding of each codeword, the equalizer coefficients are then slowly updated (one update each 500 decoded symbols) according to (8) by employing decisions. Pilot symbols at rate $r_p = 1/400$ are finally employed (and accounted for in the SE computation) to make the iterative decoding process (CPE, BCJR detection, and LDPC decoding) bootstrap. A maximum of 20 turbo iterations is considered.

Bit-error rate (BER) measurements are performed off-line by averaging over a total of about 500 randomly selected codewords (of length 64800). This sets a limit for the minimum measurable BER at about $10^{-7}$, and for reliable BER measurements at about $10^{-6}$. Transmission is considered to be error-free when all the information bits are correctly decoded at the receiver, which means, in fact, $BER < 10^{-6}$ with high probability. As discussed in Section III, detection of the adopted LDPC codes is characterized by the presence of possible error floors at $BER < 10^{-8}$ (which where, therefore, never observed neither experimentally nor in simulation), which can be practically removed by concatenating outer hard-decision BCH codes with small additional overhead ($<0.4\%$), complexity, and latency [24]. We account for this fact by virtually including a BCH code with rate $r_{BCH} = 0.996$ (0.4% redundancy) in the computation of the experimentally achieved SE, which is therefore defined as the actual amount of information that is reliably transmitted (error-free within measurement accuracy) per unit time and bandwidth once the overload due to the LDPC code, outer BCH code (virtually present), and pilot symbols is removed

$$\hat{\eta} \triangleq \frac{4r_{LDPC}r_{BCH}(1 - r_p)}{FT} \leq \eta_{\text{max}} \quad (10)$$

Depending on the available $E_b/N_0$ and on the presence of uncompensated transmission impairments, different code rates $r_{LDPC}$ are required to obtain reliable transmission. The transmission system can thus be adapted to finely adjust the information rate to the channel conditions (accumulated noise and propagation penalties) by changing $r_{LDPC}$ (selected among the values available in Tab. I) while keeping the transmission rate $1/T$ and channel spacing $F$ constant. A wider tuning of the SE is finally obtained by changing also the transmission rate. In particular, rates of 40, 35, and 30 GBd are considered. The difference between the experimentally achieved SE (10) and the theoretically achievable SE (4) depends on the performance (and available rate granularity) of the designed LDPC codes and on the presence of any additional impairment unaccounted for in this design procedure (e.g., modulator imperfections, nonlinearity, etc.).

Long-distance transmission is emulated by using a recirculating loop, composed of two 40 km long spans of standard single mode fiber, two optical amplifiers, a polarization scrambler (POL-S), and a gain equalizer filter (GEF). The total dispersion accumulated during propagation through the recirculating loop is compensated by a static frequency-domain equalizer, placed in front of the 2D-FFE equalizer and configured according to the selected link length.

A. Back to back measurements

The back-to-back performance of the system is investigated by bypassing the recirculating loop and measuring only the BER of the central channel. In order to experimentally verify the TFP optimization performed numerically in Section III, signals at different baud rates are generated and coded with
BER

Figure 8. Experimental back-to-back performance of the TFP system (only central channel): BER with the 40 Gbd DP-QPSK configuration and different code rates.

Figure 9. Experimental back-to-back performance of the TFP system (only central channel): achieved SE with the 40 Gbd, 35 Gbd, and 30 Gbd DP-QPSK configuration.

B. Transmission experiments

Transmission experiments are performed by properly setting the number of rounds that the signal travels through the recirculating loop in Fig. 6. The launch power is optimized to obtain the best trade-off between noise and nonlinear propagation effects. For the sake of simplicity, it is assumed that the optimal launch power is independent of the transmission distance and code rate. The optimization is performed by setting the same power for the five channels and measuring the performance of the third (central) one, which is the most affected by inter-channel nonlinearity. Fig. 10 shows the maximum achievable transmission distance as a function of the launch power, for either 30 or 40 Gbd transmission and a fixed SE $\bar{\eta} \approx 5.3$ bit/s/Hz (obtained with code rates of $8/9$ and $2/3$, respectively). A slightly different result is obtained for 30 Gbd and 40 Gbd transmission, the optimum launch power being -5 and -6 dBm per channel, respectively. However, in the following measures, the same launch power of -5 dBm per channel is used for any transmission rate. Once the launch power has been set, the maximum achieved SE $\bar{\eta}$—defined in (10) and obtained by selecting the highest code rate (among the available ones reported in Tab. I) guaranteeing error-free transmission—is measured as a function of the transmission distance for a transmission rate of 30, 35, and 40 Gbd. Fig. 11 shows the results for each of the five TFP channels (symbols) as well as for the whole super-channel (lines, corresponding to the worst-performing channel). In the same figure, the code rates employed to achieve the measured SEs are also indicated. Due to inter-channel nonlinearity (as $F = 2B$, linear crosstalk among channels is practically negligible), the central channel is typically the worst performing, while the outer channels are the best performing. This is more evident at higher transmission rates. At short distances, i.e. at high OSNRs, the achieved SE is much higher (about 7.1 bit/s/Hz at 400 km) than the theoretical limit of 4 bit/s/Hz achievable by Nyquist-WDM transmission with same DP-QPSK modulation format, and remains higher up to almost 6000 km. Moreover, the SE can be adapted to the propagation conditions by simply changing the code rate or, for significant OSNR variations, the amount of packing (i.e., the baud rate $1/T$ or, equivalently, the bandwidth $B$ and the frequency spacing $F$), without changing the modulation format and transceiver hardware.

VI. DISCUSSION

As a recent and not yet fully mature technique, TFP can be still improved in terms of performance and complexity. In this sense, the implementation proposed in this work is intended to demonstrate the technical feasibility and good performance of TFP and should not be considered as the ultimate TFP solution. In fact, there are several options to improve the performance (SE vs. distance) of the proposed
could provide up to 2 dB of improvement in the required SNR, as suggested by the comparison between Fig. [3] and [8]. A similar improvement could be achieved also by designing better LDPC codes (with a lower distance from the Shannon limit), as discussed in Section III. Such an improvement in the required SNR would then translate into an almost proportional improvement in terms of maximum transmission distance. This is shown in Fig. 12, which compares the achieved SE (10) measured in the experimental setup (the 40 GBd configuration), the achieved SE (10) estimated through numerical simulations (including nonlinear effects but without accounting for TX or RX imperfections), and the achievable SE (4) estimated through numerical simulations (i.e., for the best possible code). The difference between the first two curves (experiments and simulations) is due to the aforementioned imperfections of the experimental setup, which are not considered in the numerical simulations and could be possibly removed. Moreover, for high OSNR values (i.e., at short distances), electronic noise and quantization effects at the receiver (not included in the simulations) become relevant with respect to optical noise and causes an additional penalty with respect to numerical simulations. This explains why the difference between the two curves (in dBs along the x-axis) is not constant and increases at short distances. Finally, the difference between the last two curves (experiments and simulations) is due to the aforementioned imperfections of the experimental setup, which are not considered in the numerical simulations and could be possibly removed. Moreover, for high OSNR values (i.e., at short distances), electronic noise and quantization effects at the receiver (not included in the simulations) become relevant with respect to optical noise and causes an additional penalty with respect to numerical simulations. This explains why the difference between the two curves (experiments and simulations) is non-constant and increases at short distances. Finally, the difference between the last two curves (experiments and simulations) is due to the aforementioned imperfections of the experimental setup, which are not considered in the numerical simulations and could be possibly removed.

The TFP transmission technique beyond that achieved in our experiments. In the first place, we consider some improvements that, being related to the design, optimization, or implementation of the system, do not affect its complexity. As indicated in Fig. 2 due to some limitations in the available hardware, the adopted configuration is not exactly the optimum one. In fact, according to numerical simulations, the optimum configuration provides from 0.5 to 1 bit/s/Hz of improvement of the achievable SE (for low and high OSNRs, respectively) compared to the suboptimum configuration actually employed in the experimental setup. Moreover, as the optimization of Fig. 2 refers to an AWGN channel, a more accurate optimization could be sought by accounting also for nonlinear effects through approximate channel models, time-consuming simulations, or directly optimizing the experimental setup through extensive measurements (in increasing order of accuracy and required time). A non-negligible improvement should be achievable, still without increasing the system complexity, also by removing possible imperfections in the experimental setup (e.g., unbalance between I/Q components or polarizations at the modulator or at the opto-electronic front-end). This

In the second place, a significant improvement can be obtained at the expense of an additional DSP complexity. For instance, by modifying the detection strategy to account for a longer ISI (e.g., increasing the number of trellis states or considering channel shortening techniques [31], as described in [8] and/or also for ICI (multi-user detection), pulses can be more densely packed, achieving a higher SE. Finally, the DSP implemented in the experimental setup does not include any nonlinearity mitigation strategy, which could be adopted to improve the overall performance. For instance, as
confirmed by some preliminary results, the low-complexity digital backpropagation strategy proposed in [32] can be easily integrated in the DSP (replacing the static frequency-domain equalizer for dispersion compensation) to mitigate intra-channel nonlinearity and extend the reach. This subject is however outside the scope of the paper and is left for a future investigation.

Due to possible improvements and ongoing research, an accurate and comprehensive comparison with more conventional (and mature) techniques (e.g., Nyquist-WDM) is not yet available. A numerical comparison between TFP and Nyquist-WDM performance can be found in [8], where it is shown that TFP can achieve higher SE values than high-order modulation Nyquist-WDM over realistic long-haul systems. Moreover, TFP offers advantages also at a network level, as it provides high SE and flexibility (e.g., reach adaptation and filter configuration) without requiring transponders supporting multiple modulation formats [7]. In terms of complexity, TFP has the advantage of a simpler transmitter architecture (e.g., only two-level driving signals are needed to control the modulator; no DSP and digital-to-analog conversion are required to process the modulating signals; more relaxed constraints on the pulse shape can be considered) at the expense of a more complex DSP at the receiver (e.g., the 8-state BCJR detector employed in this work). On the other hand, the use of a TFP DP-QPSK format allows to employ only symbol-time processing at the receiver and greatly simplifies decision-directed algorithms compared to higher level modulation formats. This can partly compensate for the additional complexity of the BCJR detector.

VII. CONCLUSIONS

In this work, after reviewing the main theoretical aspects of the TFP approach, we have investigated its application to fiber-optic systems. The main challenges pertain to the peculiar nature of the channel (the optical fiber, impaired by linear and nonlinear propagation effects) and to the high data rates involved. We have thus discussed the implementation schemes, focusing on the main differences with respect to a conventional coherent WDM system. The only relevant difference is a modification of the DSP algorithms employed for detection. In the proposed scheme, a butterfly equalizer adaptively addresses propagation impairments and perform matched filtering, while intentional ISI due to TFP filtering is accounted for by a BCJR detector. To ensure a proper distribution of tasks between the equalizer and BCJR detector, an algorithm has been proposed that adaptively controls the equalizer and provide channel metrics to the BCJR detector. This makes the receiver fully adaptive, without requiring a priori knowledge of the adopted TFP configuration. Soft-decoding forward-error correction is finally employed. In particular, irregular LDPC codes with various code rates (in the range 2/3–8/9) and specifically optimized for the TFP channel have been designed to operate at low error rates. They approach (within about 3 dB) the SE limits achievable by the proposed techniques at different SNRs. The performance of the proposed system have been tested both experimentally and by simulations, demonstrating technical feasibility and good performance. Five closely-packed DP-QPSK channels were transmitted through a recirculating loop, keeping the net SE beyond the theoretical limit of Nyquist-WDM (4 bit/s/Hz) up to 6000 km. The channel bandwidth and spacing was held fixed to 20 GHz, while the transmission rate and code rate were adapted, depending on the transmission distance, to the available OSNR. At 400 km, a net SE of more than 7 bit/s/Hz was achieved by setting the transmission rate at 40 Gb/s (twice faster than the Nyquist limit) and the code rate to 8/9. The transmission distance was then gradually increased up to 6000 km, with a net SE which gradually decreased to about 4 bit/s/Hz (achieved with a 30 Gb/s transmission rate and a 2/3 code rate). Both the LDPC encoder and decoder were actually included in the experimental setup, as it is advisable in the presence of soft decoding, for which the use of a numerically evaluated “pre-FEC BER threshold” may be unreliable.

In conclusion, we have demonstrated that TFP with low-level modulation (e.g., DP-QPSK) can be considered as a practical and viable alternative to high-level modulations to achieve high SEs over long-haul fiber-optic links, providing good performance and high flexibility (reach adaptation and filter configuration) with simpler transponder architectures (single modulation format, relaxed constraints on pulse shape, no DSP and digital-to-analog conversion at the transmitter).

REFERENCES

[1] G. Bosco, V. Curri, A. Carena, P. Poggiolini, and F. Forghieri, “On the performance of Nyquist-WDM Terabit superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers,” J. Lightwave Technol., vol. 29, no. 1, pp. 53–61, 2011.
[2] W. Shieh, X. Yi, and Q. Yang, “Coherent optical OFDM: has its time come?,” vol. 7, pp. 234–255, Mar. 2008.
[3] A. Barbieri, G. Colavolpe, T. Foggi, E. Forestieri, and G. Prati, “OFDM vs. single-carrier transmission for 100 Gbps optical communication,” J. Lightwave Technol., vol. 28, pp. 2537–2551, September 1 2010.
[4] A. Barbieri, D. Fertonani, and G. Colavolpe, “Time-frequency packing for linear modulations: spectral efficiency and practical detection schemes,” IEEE Trans. Commun., vol. 57, pp. 2951–2959, Oct. 2009.
[5] G. Colavolpe, T. Foggi, A. Modernini, and A. Piemontese, “Faster-than-Nyquist and beyond: how to improve spectral efficiency by accepting interference,” Opt. Express, vol. 19, pp. 26600–26609, December 2011.
[6] L. Poti, G. Meloni, G. Berrettini, F. Fresi, M. Secondini, T. Foggi, G. Colavolpe, E. Forestieri, A. D’Errico, F. Cavaliere, R. Sabella, and G. Prati, “Casting 1 Tb/s DP-QPSK communication into 200 GHz bandwidth,” in Proc. European Conf. on Optical Commun. (ECOC’12), September 19-23, 2012.
[7] N. Sambo, G. Meloni, F. Paolucci, F. Cugini, M. Secondini, F. Fresi, L. Poti, and P. Castoldi, “Programmable transponder, code and differentiated filter configuration in elastic optical networks,” J. Lightwave Tech., vol. 32, pp. 2079–2086, June 2014.
[8] G. Colavolpe and T. Foggi, “Time-frequency packing for high capacity coherent optical links,” IEEE Trans. Commun., vol. 62, pp. 2986–2995, Aug. 2014.
[9] J. E. Mazo, “Faster-than-Nyquist signaling,” Bell System Tech. J., vol. 54, pp. 1450–1462, Oct. 1975.
[10] F. Rusek and J. B. Anderson, “The two dimensional mazo limit,” in Proc. IEEE International Symposium on Information Theory, pp. 970–974, Sept. 2005.
[11] J.-X. Cai, C. R. Davidson, A. Lucero, H. Zhang, D. G. Foursa, O. V. Sinkin, W. W. Patterson, A. N. Pilipetskii, G. Mols, and N. S. Bergano, “20 Tbit/s transmission over 6800 km with sub-Nyquist channel spacing,” J. Lightwave Technol., vol. 30, pp. 651–657, Feb. 15 2012.
[12] M. Secondini, E. Forestieri, and G. Prati, “Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps,” Journal of Lightwave Technology, vol. 31, no. 23, pp. 3839–3852, 2013.
[13] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[14] N. Merhav, G. Kaplan, A. Lapidoth, and S. S. Shitz, “On information rates for mismatched decoders,” *IEEE Trans. Inform. Theory*, vol. 40, pp. 1953–1967, Nov. 1994.

[15] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, “Simulation-based computation of information rates for channels with memory,” *IEEE Trans. Inform. Theory*, vol. 52, pp. 3498–3508, Aug. 2006.

[16] G. Ungerboeck, “Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems,” *IEEE Trans. Commun.*, vol. 22, pp. 624–636, May 1974.

[17] G. Colavolpe and A. Barbieri, “On MAP symbol detection for ISI channels using the Ungerboeck observation model,” *IEEE Commun. Letters*, pp. 720–722, August 2005.

[18] R. Bahl, J. Cocke, F. Jelinek, and R. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate,” *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–284, Mar. 1974.

[19] A. Modenini, F. Rusek, and G. Colavolpe, “Optimal transmit filters for ISI channels under channel shortening detection,” *IEEE Trans. Commun.*, 2013.

[20] S. ten Brink, G. Kramer, and A. Ashikhmin, “Design of low-density parity-check codes for modulation and detection,” *IEEE Trans. Commun.*, vol. 52, pp. 670–678, Apr. 2004.

[21] T. Richardson, A. Shokrollahi, and R. Urbanke, “Design of capacity-approaching irregular low-density parity check codes,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.

[22] H. Xiao and A. H. Banihashemi, “Improved progressive-edge-growth (PEG) construction of irregular LDPC codes,” *IEEE Commun. Letters*, vol. 8, pp. 715–717, Dec. 2004.

[23] X. Hu, E. Eleftheriou, and D. M. Arnold, “Regular and irregular progressive edge-growth Tanner graphs,” *IEEE Trans. Inform. Theory*, vol. 51, pp. 386–398, Jan. 2005.

[24] “Digital video broadcasting (DVB); second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broad-band satellite applications (DVB-S2),” Standard EN 302 307.

[25] G. Colavolpe, T. Foggii, E. Forestieri, and G. Prati, “Robust multilevel coherent optical systems with linear processing at the receiver,” *J. Lightwave Technol.*, vol. 27, pp. 2357–2369, July 2009.

[26] F. Fresi, M. Secondini, G. Berrettini, G. Meloni, and L. Poti, “Impact of optical and electrical narrowband spectral shaping in faster than Nyquist Tb superchannel,” *IEEE Photon. Technol. Lett.*, vol. 25, no. 23, pp. 2301–2303, 2013.

[27] A. Abrardo and G. Cincotti, “Design of AWG devices for all-optical time-frequency packing,” *J. Lightwave Tech.*, vol. 32, pp. 1951–1959, May 2014.

[28] U. Mengali and M. Morelli, “Data-aided frequency estimation for burst digital transmission,” *IEEE Trans. Commun.*, vol. 45, pp. 23–25, January 1997.

[29] G. Colavolpe, A. Barbieri, and G. Caire, “Algorithms for iterative decoding in the presence of strong phase noise,” *IEEE J. Select. Areas Commun.*, vol. 23, pp. 1748–1757, Sept 2005.

[30] J. Hagenauer, “The turbo principle: Tutorial introduction and state of the art,” in *Proc. International Symposium on Turbo Codes and Related Topics*, pp. 1–11. 1997.

[31] F. Rusek and A. Prlja, “Optimal channel shortening for MIMO and ISI channels,” *IEEE Trans. on Wireless Commun.*, vol. 11, pp. 810–818, February 2012.

[32] M. Secondini, D. Marsella, and E. Forestieri, “Enhanced split-step fourier method for digital backpropagation,” in *Proc. European Conf. on Optical Commun. (ECOC)*, Sept. 21–25, 2014.