Multiuser Detection for Time-Frequency-Packed Systems

Multi-User Detection for Time-Frequency-Packed Systems

Giulio Colavolpe, *Senior Member*, *IEEE*, Tommaso Foggi, Amina Piemontese, Alessandro Ugolini, Ling Liu, Jilong Han

Abstract

We consider time-frequency-packed systems in an additive white Gaussian noise scenario. With respect to systems employing time packing, thus introducing intentional intersymbol interference only, a further improvement of the spectral efficiency is obtained by packing the adjacent carriers closer in frequency. The adoption of a multiuser detector can allow coping with the resulting intentional intercarrier interference.

The optimal detector for time-frequency-packed systems has a complexity that increases exponentially with the number of interferers in both time and frequency dimensions and thus becomes unmanageable. We propose a suboptimal receiver, which is derived by using the framework based on factor graphs and the sum-product algorithm, with a complexity that is linear in the number of adjacent carriers and takes also advantage of a multi-user channel shortener to further reduce the complexity.

We assess the performance by means of computer simulations and show that, when compared with other suboptimal receivers available in the literature and with the optimal detector, the proposed receiver results very promising in terms of trade-off between performance and computational complexity.

G. Colavolpe, T. Foggi, A. Piemontese, and A. Ugolini are with Department of Engineering and Architecture and the CNIT research Unit, University of Parma, I-43124 Parma, Italy (e-mail: name.surname@unipr.it).

L. Liu and J. Han are with Optical Research Department, Huawei Technologies, Shenzhen, 518129, China (email: ling.liuling@huawei.com; hanjilong@huawei.com).

This work was supported by Huawei Technologies Co. Ltd. under grant n. YBN2019055044.

Index Terms

Multi-user detection, MAP symbol detection, time-frequency packing, channel shortening, faster-than-Nyquist signaling.

I. INTRODUCTION

Faster-than-Nyquist (FTN) signaling [1] is a technique aiming at reducing the symbol time below the Nyquist limit up to the smallest value giving no reduction of the minimum Euclidean distance with respect to the Nyquist case, so that, asymptotically, the intersymbol interference (ISI)-free uncoded bit-error-rate (BER) performance is reached when the optimal detector is used, and the spectral efficiency is improved through the reduction of the symbol time. This concept has been extended to multicarrier transmissions in [2], where intentional intercarrier interference (ICI) is also introduced by reducing the frequency separation among carriers. From a practical point of view, FTN requires an optimal detector whose complexity, however, easily becomes unmanageable. No hints are provided in the original papers on how to perform the optimization in the more practical scenario where a reduced-complexity receiver is employed. Time-frequency packing (TFP) [3], which represents an evolution of the FTN technique, is an effective technique which has been developed for the purpose of optimizing the symbol time and the carrier spacing with the aim of maximizing the spectral efficiency for a given receiver complexity. In this way, a double goal is obtained: the complexity is selected in advance according to the constraints at the receiver and a figure of merit more effective than the uncoded BER, i.e., the spectral efficiency, is selected for system optimization. The advantages of this technique have been demonstrated in many scenarios (see the tutorial paper [4] and references therein).

In the original paper [3], the idea behind TFP has been demonstrated by using the simplest possible receiver, i.e., a symbol-by-symbol detector neglecting both ICI and ISI. In other words, instead of the optimal receiver for the actual channel, the optimal receiver for a simplified auxiliary channel is adopted, for which the combined effect of ISI and ICI is modeled as a zero-

mean random variable independent of the additive thermal noise. More complex suboptimal receivers, partially coping with the intentional interference, are adopted in other papers for the same additive white Gaussian noise (AWGN) scenario (see, e.g., [5]) or other more sophisticated scenarios (see, e.g., [6]–[8]).

In all these papers, the suboptimal receiver has been designed to cope with (part of) the ISI only. In fact, coping with the ICI requires the adoption of a multi-user receiver, not considered in the above mentioned papers for complexity reasons.[1] As far as the suboptimal receivers coping with part of the ISI are concerned, many algorithms are available in the literature (see [9]–[20] and references therein). Some of the most effective ones are based on the output of the whitened matched filter (WMF) (the so called Forney observation model). However, when pulses are tightly packed in time the whitening filter (WF) becomes unstable [21]. Let us consider, in fact, a shaping pulse having a root-raised-cosine (RRC) spectrum with roll-off α and let us define T_N and T the symbol time ensuring the fulfillment of the Nyquist condition and the actual one, respectively. The time compression factor τ is defined as the ratio T/T_N. As shown in [21], the necessary condition for a stable WF is

$$\frac{T}{T_N} \geq \frac{1}{1+\alpha}\,.$$

Since we are typically interested in lower values of the time compression factor, suboptimal detection algorithms for the ISI channels working on the output of the matched filter (MF) have to be preferred. In this class of algorithms, we mention those described in [19], [20]. In particular, in [20] the use of a channel shortener (CS), i.e., a proper discrete-time linear filter working on the samples at the output of the MF is proposed, plus a maximum a-posteriori (MAP) symbol detector which copes with only a portion of the original ISI intentionally introduced at the transmitter. This receiver architecture is particularly effective since the CS is designed by

[1]In this paper, we will use the terms, "channels", "users", and "carriers" interchangeably. For this reason, the algorithms coping with the ICI are defined, as usual in the literature, "multi-user" although "multi-carrier" would be more appropriate.

maximizing the information rate for a given number of states of the MAP symbol detector. For this reason, it has been adopted in [6], [7], [22].

The main contribution of this paper consists of the introduction of a novel suboptimal receiver architecture characterized by linear complexity with the number of carriers, which enhances the one based on independent single-user detectors (SUDs). In particular, we propose a novel multi-user detector (MUD) and a two-dimensional CS. The proposed MUD is obtained by means of the application of the sum-product algorithm (SPA) on a properly defined factor graph (FG) [23]. The considered factor graph represents the joint a-posteriori probability mass function of the transmitted symbols and requires proper manipulations to avoid short cycles. It achieves linear complexity with the number of carriers. The resulting suboptimal MUD is based on many SUDs, one per each carrier, exchanging soft information through proper additional factor nodes. The proposed multi-user CS is designed in order to reduce, in the time dimension, not only the ISI that the MAP detector takes into account (i.e., the self interference of each carriers) but also the ICI, so as to obtain a further complexity saving.

The paper is organized as follows. In Section II a general architecture for TFP system is described. Then, in Section III the MUD receiver is detailed, and in particular the suboptimal detector and the multi-user CS are derived, whereas in Section IV a few results are given in order to assess the detector performance. Finally, conclusions are drawn in Section V.

II. Single-user system architecture

We first describe the generic architecture for a time-frequency-packed system using SUDs, which is taken as a baseline. The system architecture in case of use of a MUD is considered later. The baseband equivalent of the transmitter architecture is shown in Fig. 1. A stream of information bits is split into U streams $\{a_k^{(\ell)}\}_{\ell=1}^{U}$ which are independently encoded (ENC). Gray mapping is used to obtain code symbols $\{x_k^{(\ell)}\}_{\ell=1}^{U}$, belonging to a zero-mean M-ary complex constellation $\mathcal{X}$, which are linearly modulated by using the shaping pulse $p(t)$ (MOD). Every

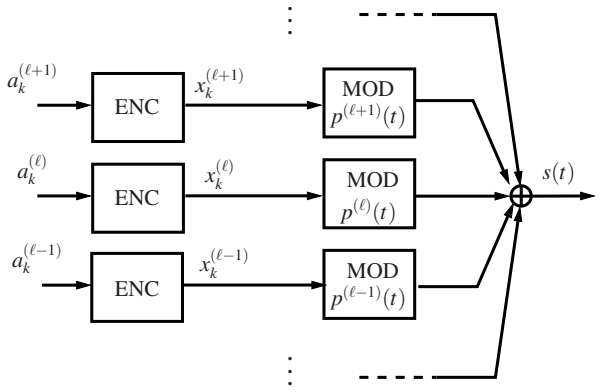

Figure 1. Transmitter architecture.

codeword, assumed of length K symbols, employs a different carrier.[2] Without loss of generality, we assume that the carriers are equally spaced in frequency and F is the frequency separation between two adjacent carriers. The complex envelope of the transmitted signal can thus be expressed as

$$s(t) = \sum_{\ell=1}^{U} \sum_{k=0}^{K-1} x_k^{(\ell)} p^{(\ell)}(t - kT) \tag{1}$$

where

$$p^{(\ell)}(t) = p(t) e^{j2\pi f^{(\ell)} t}$$

and $f^{(\ell)} - f^{(\ell-1)} = F$. In (1) the base pulse $p(t)$ is thus regularly shifted, in the time and frequency domains, of multiples of T seconds and F Hz respectively, on a rectangular lattice. The base pulse $p(t)$ can be a pulse with RRC spectrum having roll-off factor α, a properly optimized pulse [24], or any other pulse, chosen for reasons of implementation constraints.

Signal $s(t)$ is then transmitted over an AWGN channel. The complex envelope of the received signal $r(t)$ can be thus expressed as $r(t) = s(t) + w(t)$, where $w(t)$ is a complex-valued Gaussian white noise process with independent components, each with two-sided power spectral density

[2]The values of K and U are design parameters that can be arbitrarily selected, as, in general, transmitted streams could belong to one or several users. For details on the TFP technique please see [3]–[8].

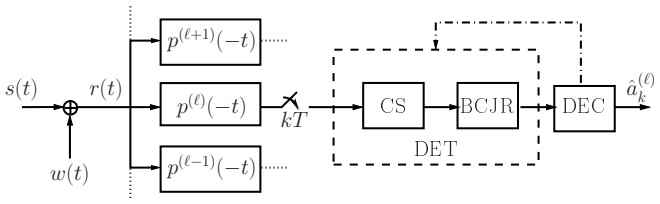

Figure 2. Receiver architecture for single-user detection.

N_0. In the SUD architecture, shown in Fig. 2, independent receivers are used for each stream. In each of them, a filter matched to the frequency shifted replica of the base pulse is first employed. The samples at the output of the MF, taken at discrete-time instants kT are sent to a soft-output constrained-complexity detector (DET) which exchanges soft information with a decoder (DEC). Iterative detection and decoding can be possibly considered (dash-dotted line). In our numerical examples, we will employ, at the transmitter, low-density parity-check (LDPC) codes whose coded bits are properly mapped onto M-ary constellations. Regarding the soft-output detector, many options are considered in [5]. However, the solution guaranteeing the best performance is represented by a proper discrete-time linear filter, a so-called *channel shortener* (CS) [20], [25], followed by a Bahl, Cocke, Jelinek, Raviv (BCJR) algorithm [26] working on a trellis with a constrained number of states S. The number of states of the BCJR algorithm is a-priori chosen according to the available complexity at the receiver. When $S = 1$, i.e., when the BCJR degenerates into a symbol-by-symbol detector, the optimal CS results to be a minimum mean square error (MMSE) equalizer [20], [25].

III. Multi-user system architecture and detector

We now consider a receiver architecture making use of a MUD. In this case, the main problem is related to the complexity which, for the optimal MUD, becomes exponential with the number of interfering carriers. For this reason, we propose an innovative *suboptimal* receiver whose complexity increases only linearly with the number of carriers. We first review the optimal

MUD and then introduce the proposed suboptimal MUD, for TFP transmissions. These detectors employ, as a sufficient statistic, the output of a bank of filters matched to the transmitted base pulses (Ungerboeck model [19], [27]–[29]). Indeed, an alternative sufficient statistic can be obtained by using a multidimensional WF, whose implementation is, however, possible only when the time and frequency compression factors are very limited [30]. Since it is our aim to go beyond these limits, the design of MUD receivers based on the samples at the output of the bank of MFs is of paramount importance.

We will denote by $y_k^{(\ell)}$ the output at time kT of the filter matched to the pulse $p^{(\ell)}(t)$ and define

$$\begin{aligned}\boldsymbol{x}^{(\ell)} &= (x_0^{(\ell)}, x_1^{(\ell)}, \ldots, x_{K-1}^{(\ell)})^T \\ \boldsymbol{x} &= (\boldsymbol{x}^{(1)T}, \boldsymbol{x}^{(2)T}, \ldots, \boldsymbol{x}^{(U)T})^T \\ \boldsymbol{y}^{(\ell)} &= (y_0^{(\ell)}, y_1^{(\ell)}, \ldots, y_{K-1}^{(\ell)})^T \\ \boldsymbol{y} &= (\boldsymbol{y}^{(1)T}, \boldsymbol{y}^{(2)T}, \ldots, \boldsymbol{y}^{(U)T})^T .\end{aligned}$$

We consider MAP symbol detection of symbols $\boldsymbol{x}$ that requires the evaluation of the a-posteriori probabilities (APPs) $P(x_k^{(\ell)}|\boldsymbol{y})$ for all values of k, ℓ, and $x_k^{(\ell)}$, given the observation of the received sequence $\boldsymbol{y}$. As usual in iterative detection and decoding, the transmitted symbols are assumed to be independent, so that the probability mass function of the transmitted symbols can be factorized as

$$P(\boldsymbol{x}) = \prod_{\ell=1}^{U} \prod_{k=0}^{K-1} P(x_k^{(\ell)}) .$$

This assumption is enabled by the presence of an interleaver at the output of each encoder or the use of an LDPC, as assumed in the numerical results section.

In the following, we assume that the reader is familiar with the FG/SPA framework. We start by computing the joint APP $P(\boldsymbol{x}|\boldsymbol{y})$ of the symbols given the received sequence $\boldsymbol{y}$. We then derive a few corresponding FGs and apply to them the SPA, which provides the exact (in the

case of absence of cycles) or approximate (in the presence of graph's cycles) marginal APPs $\{P(x_k^{(\ell)}|\boldsymbol{y})\}$.

The joint APP $P(\boldsymbol{x}|\boldsymbol{y})$ can be expressed as [19]

$$P(\boldsymbol{x}|\boldsymbol{y}) \propto P(\boldsymbol{x}) \exp\left(\frac{2\Re[\boldsymbol{x}^H\boldsymbol{y}] - \boldsymbol{x}^H\boldsymbol{G}\boldsymbol{x}}{2N_0}\right) \tag{2}$$

where matrix $\boldsymbol{G}$ is a block matrix and we denote by $\boldsymbol{G}^{(\ell,m)}$, $\ell, m = 1, \ldots, U$, the (ℓ, m) submatrix with K rows and K columns, accounting for the interference between users (carriers) ℓ and m. Its entries are

$$G_{k,n}^{(\ell,m)} = p^{(\ell)}(t - nT) \otimes p^{(m)}(-t)\big|_{t=kT}\,,$$
$$k, n = 0, 1, \cdots, K-1$$

where "$\otimes$" denotes convolution. It is easy to show that these matrices have a Toeplitz structure. Hence, we can define $G_{k,n}^{(\ell,m)} = g_{k-n}^{(\ell,m)}$. Matrix $\boldsymbol{G}$ is also Hermitian, i.e., $\boldsymbol{G} = \boldsymbol{G}^H$. Thus, $\boldsymbol{G}^{(\ell,m)} = \boldsymbol{G}^{(m,\ell)H}$.

When Nyquist pulses are employed, no ISI among symbols of the same user arises and the submatrices $\boldsymbol{G}^{(\ell,m)}$ are diagonal. On the other hand, when the different subcarriers do not overlap, all matrices $\boldsymbol{G}^{(\ell,m)}$ with $\ell \neq m$ have zero entries. On the contrary, when frequency packing is adopted, overlap in frequency is allowed. Interference among adjacent users is thus intentionally introduced in order to increase the bandwidth efficiency, provided that the detector can cope with it. As a consequence, the number of non-zero off-diagonal elements in the matrix $\boldsymbol{G}$ is not negligible and MUD becomes necessary. Hence, matrix $\mathbf{G}$ takes into account all the interference that arises both within the same channel and across adjacent channels. The exponential term in

(2) can be further factorized as

$$\exp\left(\frac{2\Re[\boldsymbol{x}^H\boldsymbol{y}]-\boldsymbol{x}^H\boldsymbol{G}\boldsymbol{x}}{2N_0}\right)$$
$$=\left[\prod_{\ell=1}^{U}\exp\left(\frac{2\Re[\boldsymbol{x}^{(\ell)H}\boldsymbol{y}^{(\ell)}]-\boldsymbol{x}^{(\ell)H}\boldsymbol{G}^{(\ell,\ell)}\boldsymbol{x}^{(\ell)}}{2N_0}\right)\right]$$
$$\cdot\left[\prod_{\ell=1}^{U}\prod_{m>\ell}\exp\left(\frac{-\boldsymbol{x}^{(\ell)H}\boldsymbol{G}^{(\ell,m)}\boldsymbol{x}^{(m)}-\boldsymbol{x}^{(m)H}\boldsymbol{G}^{(m,\ell)}\boldsymbol{x}^{(\ell)}}{2N_0}\right)\right].$$

As shown in [28], [29], we can express

$$\exp\left(\frac{2\Re[\boldsymbol{x}^{(\ell)H}\boldsymbol{y}^{(\ell)}]-\boldsymbol{x}^{(\ell)H}\boldsymbol{G}^{(\ell,\ell)}\boldsymbol{x}^{(\ell)}}{2N_0}\right)$$
$$=\prod_{k=0}^{K-1}\exp\left(\frac{2}{N_0}\Re\left[y_k x_k^{(\ell)*}-\frac{1}{2}\left|x_k^{(\ell)}\right|^2 g_0^{(\ell,\ell)}+\right.\right.$$
$$\left.\left.-x_k^{(\ell)*}\sum_{n=1}^{L}g_n^{(\ell,\ell)}x_{k-n}^{(\ell)}\right]\right),$$

where L is the channel correlation length which depends on the amount of time packing — the lower the time compression factor, the larger the value of L. Similarly, it is possible to show that (see Appendix A)

$$\exp\left(\frac{-\boldsymbol{x}^{(\ell)H}\boldsymbol{G}^{(\ell,m)}\boldsymbol{x}^{(m)}-\boldsymbol{x}^{(m)H}\boldsymbol{G}^{(m,\ell)}\boldsymbol{x}^{(\ell)}}{2N_0}\right)$$
$$=\prod_{k=0}^{K-1}\exp\left(\frac{1}{N_0}\Re\left[x_k^{(\ell)*}g_0^{(\ell,m)}x_k^{(m)}+\right.\right.$$
$$\left.\left.+\sum_{q=1}^{L}\left(x_k^{(\ell)*}g_q^{(\ell,m)}x_{k-q}^{(m)}+x_k^{(m)*}g_{-q}^{(\ell,m)*}x_{k-q}^{(\ell)}\right)\right]\right).$$

We finally define by

$$\sigma_k^{(\ell)}=(x_{k-1}^{(\ell)},x_{k-2}^{(\ell)},\ldots,x_{k-L}^{(\ell)}),$$

$$\boldsymbol{\sigma}_k=(\sigma_k^{(1)},\sigma_k^{(2)},\ldots,\sigma_k^{(U)}),$$

$$\boldsymbol{\sigma}=\{\boldsymbol{\sigma}_k\}$$

the state of the memory for user ℓ at time instant k, the set of the states of all users at time k, and the set of all the states, respectively, and by

$$\boldsymbol{x}_k = (x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(U)})$$

a vector containing the symbols transmitted by all users at time k. The following definitions, instead, represent various functions involving states and symbols, which will become useful in the description of the MUD receivers.

$$H_k^{(\ell)}(x_k^{(\ell)}, \sigma_k^{(\ell)}) = \exp\left(\frac{2}{N_0}\Re\left[y_k x_k^{(\ell)*} - \frac{1}{2}\left|x_k^{(\ell)}\right|^2 g_0^{(\ell,\ell)} - x_k^{(\ell)*}\sum_{n=1}^{L} g_n^{(\ell,\ell)} x_{k-n}^{(\ell)}\right]\right)$$

$$F_k^{(\ell,m)}(x_k^{(\ell)}, x_k^{(m)}, \sigma_k^{(\ell)}, \sigma_k^{(m)}) = \exp\left(\frac{1}{N_0}\Re\left[x_k^{(\ell)*} g_0^{(\ell,m)} x_k^{(m)} + \sum_{q=1}^{L}\left(x_k^{(\ell)*} g_q^{(\ell,m)} x_{k-q}^{(m)} + x_k^{(m)*} g_{-q}^{(\ell,m)*} x_{k-q}^{(\ell)}\right)\right]\right)$$

$$H_k(\boldsymbol{x}_k, \boldsymbol{\sigma}_k) = \left[\prod_{\ell=1}^{U} H_k^{(\ell)}(x_k^{(\ell)}, \sigma_k^{(\ell)})\right]$$

$$F_k(\boldsymbol{x}_k, \boldsymbol{\sigma}_k) = \left[\prod_{\ell=1}^{U}\prod_{m>\ell} F_k^{(\ell,m)}(x_k^{(\ell)}, x_k^{(m)}, \sigma_k^{(\ell)}, \sigma_k^{(m)})\right].$$

At this point, we can describe the optimal and suboptimal MUD receivers.

A. *Optimal MUD*

With the above definitions, we can express

$$P(\boldsymbol{x}, \boldsymbol{\sigma}|\boldsymbol{y}) \propto P(\boldsymbol{\sigma}_0)\prod_{k=0}^{K-1} P(\boldsymbol{x}_k) H_k(\boldsymbol{x}_k, \boldsymbol{\sigma}_k) \cdot F_k(\boldsymbol{x}_k, \boldsymbol{\sigma}_k) I(\boldsymbol{x}_k, \boldsymbol{\sigma}_k, \boldsymbol{\sigma}_{k+1}) \quad (3)$$

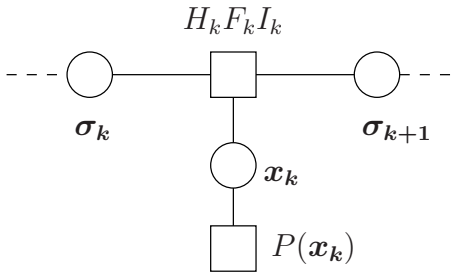

Figure 3. FG corresponding to eq. (3).

where

$$P(\boldsymbol{x}_k) = \prod_{\ell=1}^{U} P(x_k^{(\ell)})$$

and

$$I(\boldsymbol{x}_k, \boldsymbol{\sigma}_k, \boldsymbol{\sigma}_{k+1}) = P(\boldsymbol{\sigma}_{k+1}|\boldsymbol{\sigma}_k, \boldsymbol{x}_k)$$

is an indicator function, equal to one if $\boldsymbol{x}_k$, $\boldsymbol{\sigma}_k$, and $\boldsymbol{\sigma}_{k+1}$ are in a one-to-one correspondence and to zero otherwise. In other words, state $\boldsymbol{\sigma}_k$ is defined as

$$\boldsymbol{\sigma}_k = (\boldsymbol{x}_{k-1}, \boldsymbol{x}_{k-2}, \ldots, \boldsymbol{x}_{k-L})\,.$$

When a new block $\boldsymbol{x}_k$ of U symbols, one per stream, arrives, the state $\boldsymbol{\sigma}_{k+1}$ is uniquely determined. Hence, given the pair $(\boldsymbol{\sigma}_k, \boldsymbol{x}_k)$ only one state $\boldsymbol{\sigma}_{k+1}$ is compatible with this pair, giving the value 1 of the indicator function. Factorization (3) corresponds to the Wiberg-type graph [23], [31] shown in Fig. 3.

The application of the SPA algorithm to this FG provides the optimal (because of the absence of cycles) APPs $P(\boldsymbol{x}_k|\boldsymbol{y})$ from which the APPs $P(x_k^{(\ell)}|\boldsymbol{y})$ can be computed through a further marginalization. The message-passing resulting algorithm takes the form of a forward-backward BCJR algorithm. Its complexity is, as known, proportional to the number of states, i.e., proportional to M^{UL}.

B. Subptimal MUD

The suboptimal MUD detector is based on a different graph coming from the same factorization. In fact, the joint APP mass function in (3) can be equivalently expressed as

$$P(\boldsymbol{x},\boldsymbol{\sigma}|\boldsymbol{y}) \propto \left[\prod_{\ell=1}^{U} P(\sigma_0^{(\ell)})\right] \left[\prod_{k=0}^{K-1} F_k(\boldsymbol{x}_k,\boldsymbol{\sigma}_k) \cdot \prod_{\ell=1}^{U} P(x_k^{(\ell)}) H_k^{(\ell)}(x_k^{(\ell)},\sigma_k^{(\ell)}) I(x_k^{(\ell)},\sigma_k^{(\ell)},\sigma_{k+1}^{(\ell)})\right] \tag{4}$$

$$= \left[\prod_{\ell=1}^{U} P(\sigma_0^{(\ell)})\right] \left[\prod_{\ell=1}^{U}\prod_{k=0}^{K-1} P(x_k^{(\ell)}) H_k^{(\ell)}(x_k^{(\ell)},\sigma_k^{(\ell)}) I(x_k^{(\ell)},\sigma_k^{(\ell)},\sigma_{k+1}^{(\ell)}) \prod_{m>\ell} F_k^{(\ell,m)}(x_k^{(\ell)},x_k^{(m)},\sigma_k^{(\ell)},\sigma_k^{(m)})\right] \tag{5}$$

where

$$I^{(\ell)}(x_k^{(\ell)},\sigma_k^{(\ell)},\sigma_{k+1}^{(\ell)}) = P(\sigma_{k+1}^{(\ell)}|\sigma_k^{(\ell)},x_k^{(\ell)})\,.$$

The FG corresponding to (4), shown in Fig. 4, has cycles of length four, which make the convergence of the SPA unlikely, since they are too short. We can remove these short cycles in the original graph by stretching [23] the variables $\sigma_k^{(\ell)}$ in $(x_k^{(\ell)},\sigma_k^{(\ell)})$. In other words, instead of representing variable $x_k^{(\ell)}$ alone, we define a new variable given by the couple $(x_k^{(\ell)},\sigma_k^{(\ell)})$, thus allowing to remove the edges connecting node F_k with variable nodes $\sigma_k^{(\ell)}$ [23]. We remark here that this transformation does not involve approximations—the resulting graph still preserves all the information of the original graph.

The FG corresponding to this operation has shortest cycles of length twelve and is depicted in Fig. 5 in the case of a system with three users. Obviously, since cycles are still present, the SPA applied to this graph is iterative and leads to an approximate computation of the APPs $P(x_k^{(\ell)}|\boldsymbol{y})$. We remark that the proposed suboptimal MUD is based on the exact joint APP $P(\boldsymbol{x},\boldsymbol{\sigma}|\boldsymbol{y})$ also used by the optimal MUD although factorized in a different way and it is suboptimal only

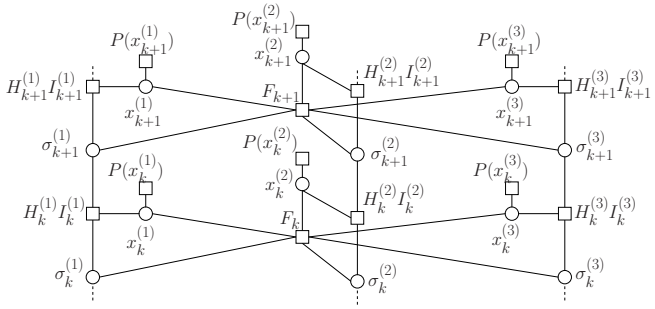

Figure 4. FG corresponding to (4) for $U = 3$.

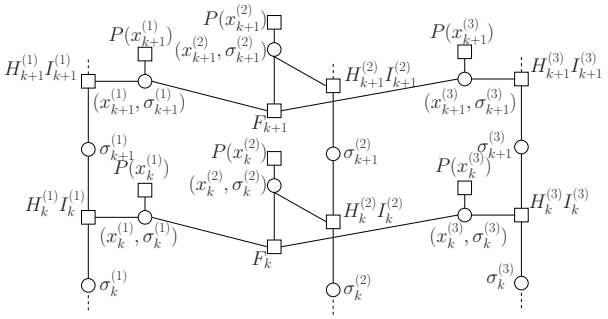

Figure 5. FG corresponding to (4) for $U = 3$, after stretching the variables $\sigma_k^{(\ell)}$ in $(x_k^{(\ell)}, \sigma_k^{(\ell)})$.

because the corresponding FG has cycles. However, the absence of short cycles allows us to obtain very good approximations, as demonstrated by the excellent performance of the algorithm.

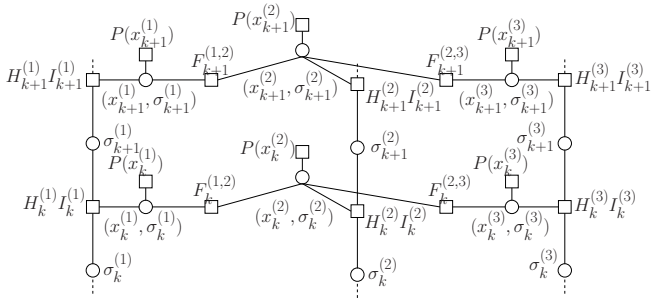

Figure 6. FG corresponding to the approximation (6) and for $U = 3$.

We can introduce a further simplification when the interference among non-adjacent users

is not present (this happens when $2F > (1+\alpha)/T_N$) or negligible. Under this condition, (5) becomes

$$P(\boldsymbol{x},\boldsymbol{\sigma}|\boldsymbol{y}) \propto \left[\prod_{\ell=1}^{U} P(\sigma_0^{(\ell)})\right] \left[\prod_{\ell=1}^{U}\prod_{k=0}^{K-1} P(x_k^{(\ell)}) H_k^{(\ell)}(x_k^{(\ell)},\sigma_k^{(\ell)}) \right.$$
$$\left. \cdot I(x_k^{(\ell)},\sigma_k^{(\ell)},\sigma_{k+1}^{(\ell)}) F_k^{(\ell,\ell+1)}(x_k^{(\ell)},x_k^{(\ell)},\sigma_k^{(\ell)},\sigma_k^{(\ell)})\right] \tag{6}$$

The corresponding FG is shown in Fig. 6 for $U = 3$. As the reader can observe by comparing Fig. 5 and Fig. 6, this further factorization in (6) implies that node F_k is split in two nodes ($U-1$ nodes, in general), $F_k^{(1,2)}$ and $F_k^{(2,3)}$, each of them connecting only two adjacent users.

Let us consider the computational complexity and look, for example, at the FG in Fig. 6. If the factor nodes $F_k^{(\ell,\ell+1)}$ were removed, we would have obtained U single-user detectors which neglect the interference. Hence, these nodes are in charge of the interference mitigation. Their introduction enhances a receiver based on independent SUDs in such a way correlation of the adjacent channels is taken into account. The complexity related to these nodes grows **linearly** and not exponentially with U. Thus the overall complexity is proportional to UM^L. The computational aspects of the described algorithm will be extensively discussed in Section IV.

C. Multi-user channel shortening

In this section, we derive a multi-user channel shortening multidimensional filter. This will help to further reduce the complexity of the adopted suboptimal MUD receiver. As mentioned, the complexity of the MUD approach increases exponentially with the channel correlation length, L. When L is very large, as happens when TFP is adopted, the receiver complexity becomes unfeasible. Hence, the need to resort to reduced-complexity receiver strategies. The most trivial solution would be to simply truncate the channel response to the first $L_r < L$ coefficients, which would, however, cause the effect of all the remaining channel coefficients to be completely neglected, causing a significant performance loss. The CS technique, instead, allows to compute a reduced channel response, with length L_r, taking into account all L channel coefficients.

Hence, the effect of the whole channel response is included in the new reduced response, used for detection. For a desired channel length L_r (and, hence, for a desired receiver complexity), the CS algorithm is able to compute the channel response and the corresponding filter, which can be demonstrated to maximize the channel information rate [20].

The proposed multi-user channel shortening filter is a direct extension of the adaptive channel shortening design strategy proposed in [25] for a single-user ISI channel. We first recall the design proposed in [25], then we describe the extension to the multi-user case. In order to compute the optimal CS according to [20], the following steps must be taken.

- Compute the sequence $\{b_k\}_{k=-L_r}^{L_r}$ as

$$b_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{N_0}{G(\omega)+N_0} e^{j\omega k} \mathrm{d}\omega \,,$$

  where $G(\omega)$ is the DTFT of the channel response g_k and L_r is the desired length of the shortened channel response.

- Compute the value

$$\mathcal{C} = b_0 - \mathbf{b}\mathbf{B}^{-\mathbf{1}}\mathbf{b}^{\dagger} \,,$$

  where $\mathbf{b} = [b_1, b_2, \ldots, b_{L_r}]$, and $\mathbf{B}$ is an $L_r \times L_r$ Toeplitz matrix with elements $(\mathbf{B})_{kj} = b_{j-k}$ for k and j between 1 and L_r.

- Define the vector $\mathbf{u} = \frac{1}{\sqrt{\mathcal{C}}}[1, -\mathbf{b}\mathbf{B}^{-\mathbf{1}}]$ and compute the optimal shortened response as

$$g_k^r = \sum_{n=\max(0,k)}^{\min(L_r, L_r+k)} u_n u_{n-1}^* - \delta_k \,,$$

  where δ_k is the Kronecker delta function.

- Finally, the optimal channel shortener is computed as

$$H^r(\omega) = \frac{G^r(\omega)+1}{G(\omega)+N_0} \,, \tag{7}$$

  where $G^r(\omega)$ is the DTFT of $\{g_k^r\}_{k=-L_r}^{L_r}$.

We can now notice that the channel shortening filter in (7) can be split into two filters, one with frequency response

$$\frac{1}{G(\omega)+N_0}\,, \tag{8}$$

which does not depend on the target length L_r, and one with frequency response $G^r(\omega)+1$ and length L_r. The filter (8) minimizes the MSE, so, combined with the front-end stage, it represents an MMSE filter. We can also notice that the sequence $\{b_k\}_{k=-L_r}^{L_r}$ is the autocorrelation of the error between the transmitted symbols and the output of the MMSE filter [25]. It is worth mentioning that the algorithm relies on a training phase in which known symbols, for example a preamble or pilot fields, are transmitted and used to compute the CS based on the error at the output of the MMSE filter.

Let us now consider the extension to a multi-user case. As we will see, the steps are an extension to the design strategy of [25]. The received signal is passed through a bank of filters matched to the transmit pulses, as shown in Fig. 7. The resulting outputs are then filtered by means of a bank of MMSE filters, and the error vectors, $\mathbf{e}_u$, $u = 1, 2, \ldots, U$ are computed. We denote with the superscript $\tilde{\cdot}$ tensors with size $U \times U \times V$. For the proposed algorithm, the first two dimensions of the tensor are equal, and will represent the number of carriers, while the third dimension will be related to the length of the reduced channel response, L_r. We also use the notation $\tilde{\mathbf{a}}(i)$ to denote a matrix with size $U \times U$ which is extracted from $\tilde{\mathbf{a}}$ by taking the elements of $\tilde{\mathbf{a}}$ whose 3rd index is equal to i. To compute a multi-user CS, we can take the following steps.

- Compute a tensor $\tilde{\mathbf{b}}$ with size $U \times U \times L_r + 1$. The elements of the tensor along the 3rd dimension can be seen as vectors of size L_r+1 which represent the cross correlation among all the error vectors. The elements on the main diagonal are the autocorrelations of the error, while the off-diagonal elements represent the mixed correlation terms.

- Define a matrix $\mathbf{B}$, with size $UL_r \times UL_r$ as a block matrix with elements

$$\mathbf{B} = \begin{bmatrix} \tilde{\mathbf{b}}(0) & \tilde{\mathbf{b}}(1) \dots \tilde{\mathbf{b}}(L_r - 1) \\ \tilde{\mathbf{b}}^*(1) & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{b}}^*(L_r - 1) & \dots \quad \dots & \tilde{\mathbf{b}}(0) \end{bmatrix} .$$

- Compute the matrix, with size $U \times U$

$$\mathcal{C} = \tilde{\mathbf{b}}(0) - \tilde{\mathbf{b}}(L_r)\mathbf{B}^{-1}\tilde{\mathbf{b}}^{\dagger}(L_r) .$$

- Compute the matrix, with size $U \times U$

$$\mathbf{U}_0 = \mathrm{Chol}(\mathcal{C}^{-1}) ,$$

  where the $\mathrm{Chol}(\cdot)$ operator represents the Cholesky decomposition.

- Define a tensor $\tilde{\mathbf{u}}_m$ with size $U \times U \times L_r + 1$. Its elements can be computed as

$$\tilde{\mathbf{u}}_m(0) = \mathbf{U}_0 ,$$
$$\tilde{\mathbf{u}}_m(j) = -\mathbf{U}_0 \sum_{i=1}^{L_r} \tilde{\mathbf{b}}(L_r)\mathbf{B}^{-1}(i,j) , \quad j = 1, \dots, L_r ,$$

  where $\mathbf{B}^{-1}(i,j)$ is the submatrix in position (i,j), with size $U \times U$.

- Compute a tensor $\tilde{\mathbf{u}}_h$ with size $U \times U \times L_r + 1$, whose elements are defined as

$$\tilde{\mathbf{u}}_h(j) = \tilde{\mathbf{u}}_m^{\dagger}(L_r - j) , \quad j = 0, \dots, L_r .$$

- Compute the CS as a tensor with size $U \times U \times 2L_r + 1$

$$\tilde{\mathbf{H}}_r(j) = \sum_{i=\max(0, j-L_r)}^{j} \tilde{\mathbf{u}}_h(j-i)\tilde{\mathbf{u}}_m(i) , \quad j = 0, \dots, 2L_r$$

  and the shortened channel response as another tensor with the same size as the CS, whose values are equal to those of $\tilde{\mathbf{H}}_r$, except for

$$\tilde{\mathbf{G}}_r(L_r + 1) = \tilde{\mathbf{H}}_r(L_r + 1) - \mathrm{I}(U) ,$$

  where $\mathrm{I}(U)$ is an identity matrix with size U.

The final MUD receiver architecture is depicted in Fig. 7.

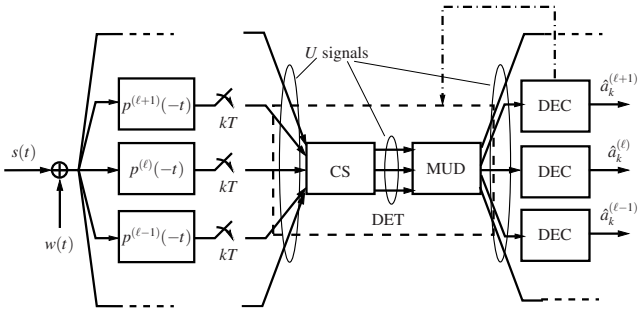

Figure 7. System architecture for multi-user detection.

IV. Numerical Results

A general performance benchmark irrespective of the employed code is the achievable information rate. When a detector is optimal for an auxiliary channel, one can resort to the mismatched detection principle [32] and compute, by using the simulation technique in [33], a lower bound of the channel information rate, achievable by the proposed receiver. Since the proposed suboptimal detector is not optimal for any auxiliary channel, we assess the performance in terms of BER, for a specific code.

Hence, we consider the following scenario. At the transmitter side, a quaternary phase-shift keying (QPSK) modulation with a RRC shaping pulse having roll-off $\alpha = 0.1$ is adopted for all carriers. The amounts of time and frequency packing are defined through the time compression factor $\tau = T/T_N$ and the parameter $\nu = FT_N/(1+\alpha)$, respectively, where $\nu < 1$ means that two adjacent carriers overlap.[3] In order to maintain the complexity of the simulations manageable, we considered just 3 adjacent channels. We selected, as examples, different values of T and F. The choice of the pairs (T, F) we adopted in the simulation results is arbitrary, since an optimization would be different for each considered detector; we just paid attention to choose T and F in order to limit the complexity. Moreover, a comparison from the point of view of

[3]Notice that T_N, representing the Nyquist condition spacing, has been normalized to one.

TFP by using the BER as a performance metric would be unaffordable, since given the receiver complexity, the optimal (T, F) pair would depend on the SNR working point and the designed codes. We thus decided to fix the frequency spacing to values where the SUD has no error floor in the BER performance and we added a slightly higher time packing, since it can be easily managed by the SUD detectors.

At the receiver side, the proposed suboptimal MUD strategy is compared with a SUD approach, not dealing with intercarrier interference, with a soft interference cancellation MMSE (sic-MMSE) detector [12], and with the optimal MUD.[4] We compute the performance of both the SUD and the MUD approaches with and without CS; in the former case, the value of the reduced channel length is fixed to $L_r = 2$, as the value of L_r depends on the amount of packing considered, therefore it derives from the trade-off between complexity and performance. The performance of the sic-MMSE is reported as a function of the parameter W, which is related to the length of the sliding window employed in the implementation of a low-complexity version of the MUD (see [12] for details). We also designed two specific codes for the scenarios at hand (i.e., we considered TFP corresponding to $T = F = 0.75$), the first one with rate 7/10 and a short length equal to 4000 bits, and the second one with rate 4/9 and length equal to 16200 bits (derived from the DVB-S2X [34] equivalent-rate code).[5] The design of the codes is based on the heuristic technique for the optimization of the degree distribution of the LDPC variable and check nodes proposed in [35], by exploiting the extrinsic information transfer (EXIT) charts. The code nodes distributions are reported in Table I, using the notation in [36]. The parity check matrix of the codes was then built through the progressive edge-growth (PEG) algorithm [37].

In the following figures, the performance of the different detectors is reported in terms of BER, as a function of E_b/N_0. In all simulations, the decoder run 40 inner iterations, whereas

[4]The comparison with non-TFP systems is not considered here since widely investigated in previous papers, see, e.g., [7].

[5]The short code was designed and employed in order to speed up the simulation time.

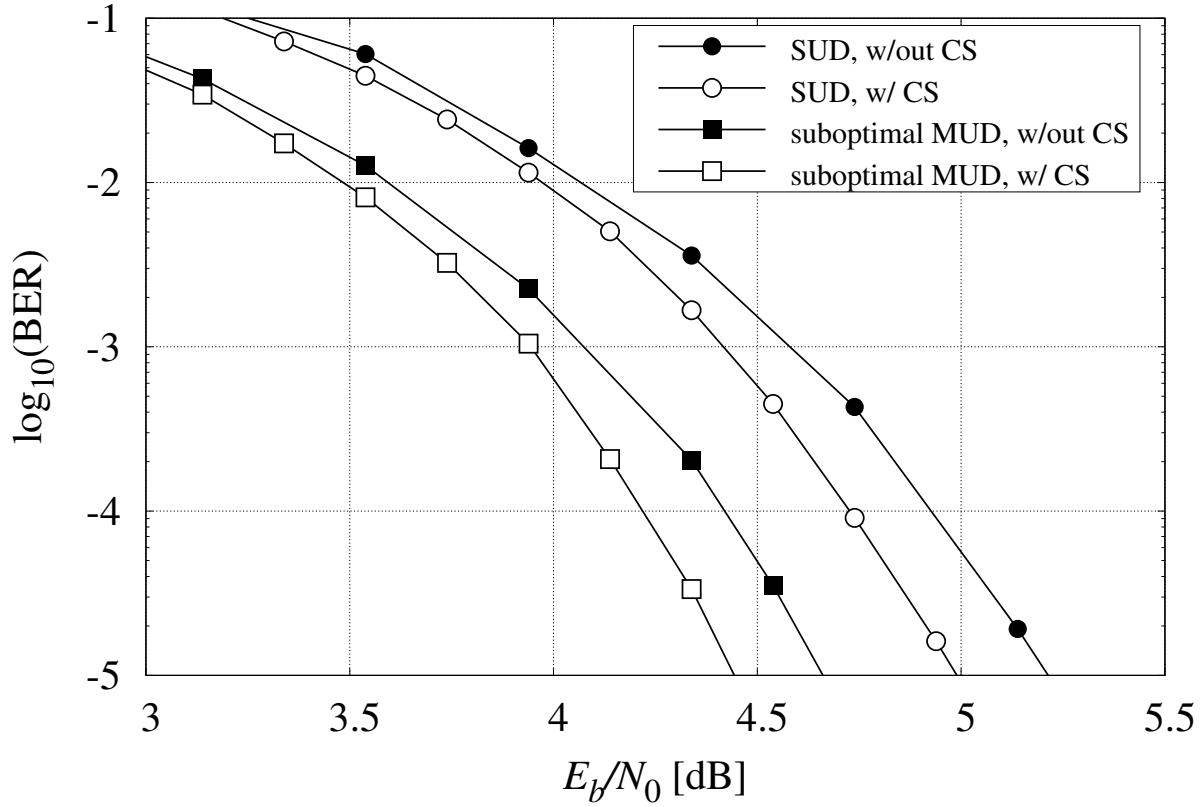

Figure 8. BER curves of the single-user (SUD) and multi-user (MUD) detectors, with $T = 0.75,\ F = 0.9$ and the 7/10 code.

Table I

VARIABLE AND CHECK NODES DEGREE DISTRIBUTIONS OF THE ad-hoc DESIGNED CODES (FRACTIONS).

rate	check nodes degree distrib.	variable nodes degree distrib.
7/10	13: 0.0133	20: 0.062
	12: 0.9867	3: 0.488
		2: 0.45
4/9	6: 0.001	8: 0.0476
	5: 0.999	3: 0.3024
		2: 0.65

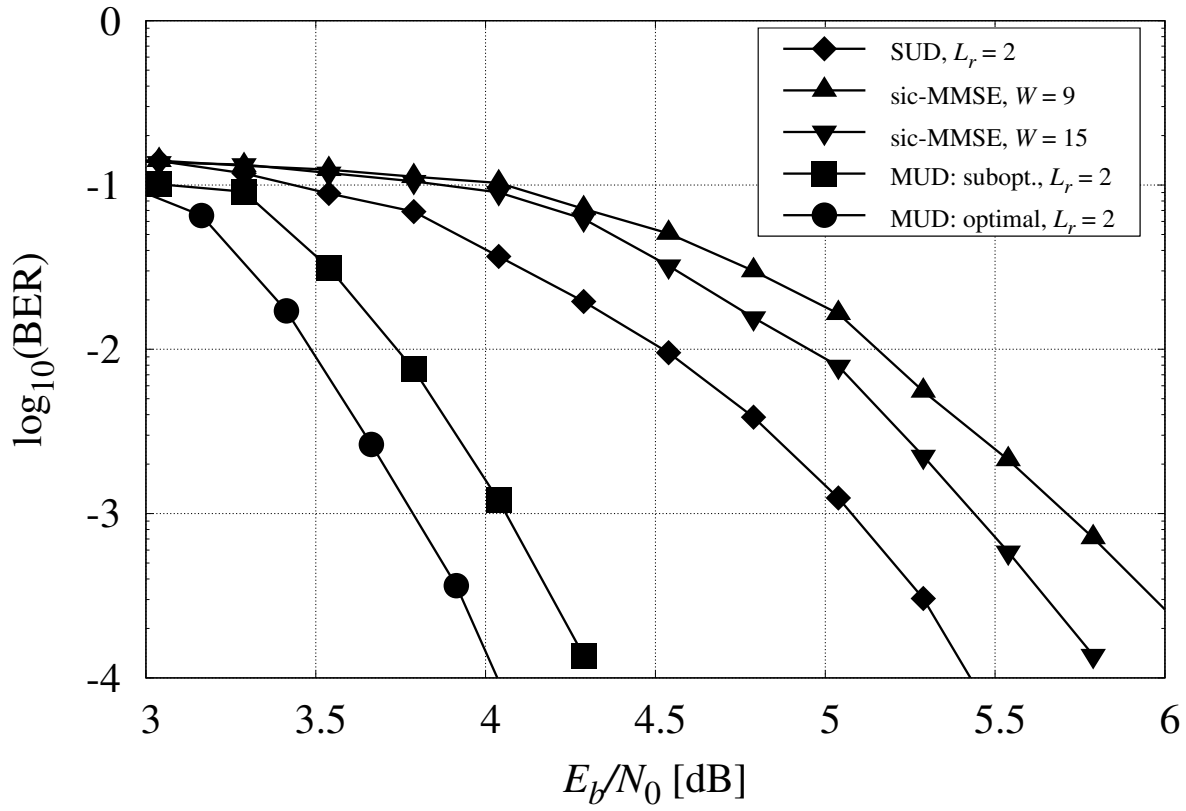

Figure 9. BER curves of the single-user (SUD) and multi-user (MUD) detectors, with $T = 0.7,\ F = 0.9$ and the 7/10 code.

there were 10 decoder-detector iterations. We first compared SUD and suboptimal MUD in case of $T = 0.75,\ F = 0.9$, with the 7/10 code. Fig. 8 shows that a slight improvement with respect to SUD can be obtained by both shortening the channel and by employing the suboptimal MUD, at the expense of an increased processing complexity. Then, Figs. 9–10 show the comparison of SUD, sic-MMSE, suboptimal and optimal MUD with $T = 0.7,\ F = 0.9$ and the 7/10 code, and with $T = 0.6,\ F = 0.8$ and the 4/9 code, respectively. In these scenarios the inter-channel interference due to adjacent channels is quite severe, therefore SUD and sic-MMSE entail a remarkable penalty, whereas the suboptimal MUD performance is about half dB away from the optimal one, proving that a significant complexity saving can be obtained at the expense of a

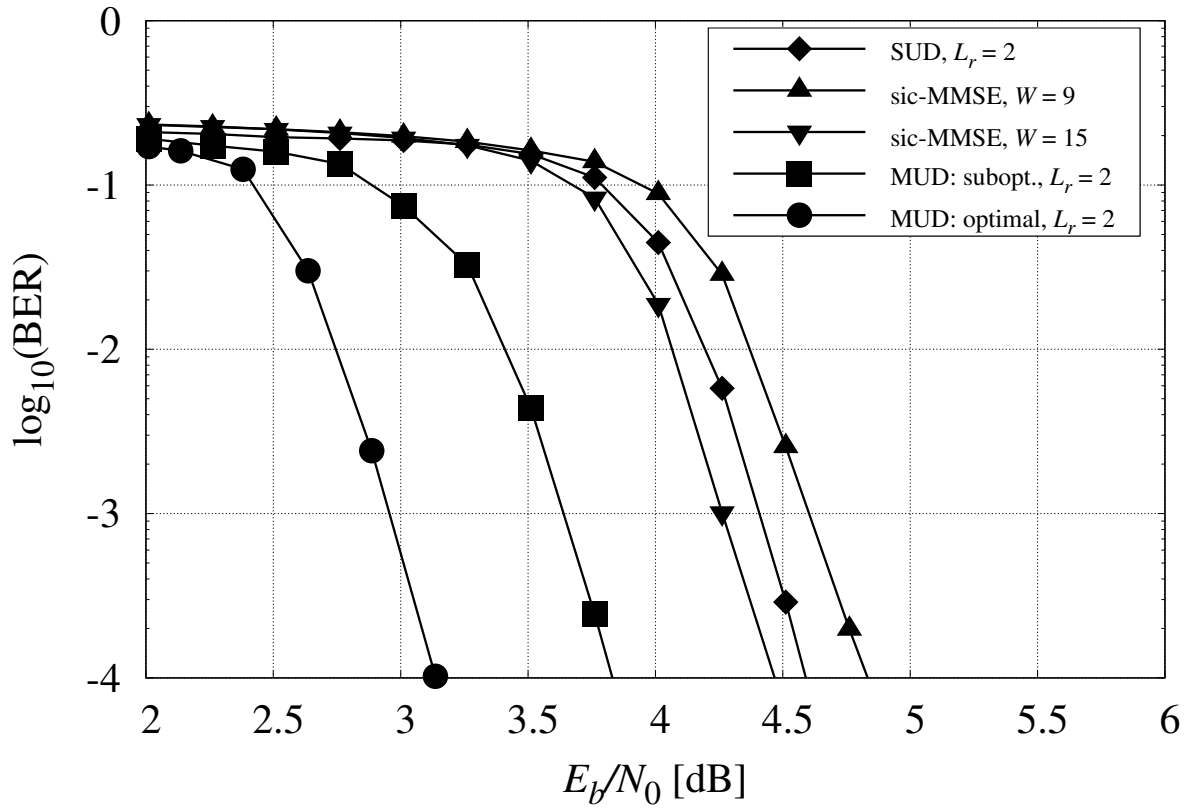

Figure 10. BER curves of the single-user (SUD) and multi-user (MUD) detectors, with $T = 0.6,\ F = 0.8$ and the 4/9 code.

limited performance penalty.

As far as the comparison between the optimal and the proposed MUD algorithms is concerned, it is worth considering complexity issues. The most demanding operation, that affects the complexity computation of the detection processing, is the max-star calculation, which is needed to update the branch metrics of the logarithmic version of the MAP trellis processing, defined as [38]

$$\max^*(a, b) \doteq \max(a, b) + \log(1 + \exp(-|a - b|))$$

$$= \log(\exp(a) - \exp(b)).$$

The complexity of the optimal and proposed MUDs can be summarized as follows: the optimal MUD is $O(M^{U(L_r+1)})$, whereas the suboptimal MUD is $O(UM^{Lr+1} + (U-1)M^{2(L_r+1)})$, since in addition to the trellis processing, the computation of nodes between adjacent channels must be taken into account. Evaluating a numerical example, if we consider a QPSK modulation ($M = 4$) with 3 channels and $L_r = 2$, we get that the optimal MUD has a complexity proportional to $4^{3\times3} = 262144$ max-star operations, whereas the suboptimal MUD complexity is proportional to $3 \times 4^3 + 2 \times 4^6 = 8384$ max-star operations, hence a remarkable complexity reduction clearly emerges.

Then, as far as the complexity comparison among all considered receivers is concerned, the number of sums, multiplications, and look-up table accesses is reported in Table II, per received symbol and per detector iteration; by considering the parameters employed in the presented results, i.e., $L_r = 2$, $U = 3$, $W = 9,\ 15$, an example of effective number of operations is reported in Table III. It must be noticed that, although reporting an exact number of operations, this table should be interpreted qualitatively more than quantitatively, as many of the operations needed for each algorithm may be implemented in different ways and with different complexities. Though, it clearly stands out that the suboptimal MUD entails a remarkable saving, and that it mainly involves additions and look-up table (LUT) access, whereas the sic-MMSE, involving

Table II

COMPLEXITY COMPARISON OF OPTIMAL, SUBOPTIMAL MUD AND SIC-MMSE DETECTOR, PER SYMBOL AND PER ITERATION.

	Additions	Multiplications	LUT accesses
Opt. MUD	$M^{U(L_r+1)} \times (1 + U \times 4)$	$M^{U(L_r+1)}$	$M^{U(L_r+1)} \times U$
Subopt. MUD	$M^{Lr+1} \times (1 + U \times 4) + 10(U-1) \times M^{2(L_r+1)}$	M^{Lr+1}	$U \times M^{(L_r+1)} + 2(U-1) \times M^{2(L_r+1)}$
sic-MMSE	$12 \times W$	$18 \times W + 21 \times W^2 + 9 \times W^3$	–

matrix inversions, is mostly implemented through multiplications, therefore a fair comparison is quite difficult and significantly affected by the specific implementation.

As mentioned in the introduction, there exist several different techniques with different complexities that likewise show different performances. For scenarios with limited ICI, so when the performance loss of the SUD with respect to the MUD is limited, clearly the proposed scheme is useless. It is plausible that, given the numerous scenarios identified by specific values of time and frequency packings, the trade-off between complexity and performance would benefit one technique over the others case-by-case; it is not possible, mainly due to computational load, to carry out an exhaustive analysis of such comparisons over a wide range of different regions. Nevertheless, the proposed suboptimal MUD has shown very good performance in the considered scenarios. Moreover, different combinations of the presented or cited techniques could be investigated, including hybrid detection/cancellation schemes. They are left for future works.

Table III

COMPLEXITY EXAMPLE.

	Additions	Multiplications	LUT accesses
Opt. MUD	3407872	262144	786432
Subopt. MUD	82752	64	16576
MMSE $W = 9$	108	8424	-
MMSE $W = 15$	180	35370	-

V. CONCLUSION

In this paper, we presented a novel suboptimal algorithm for MUD receivers, specifically tailored for TFP systems, and derived by applying the SPA to a proper FG. The complexity of the detector scales linearly in the number of users, instead of exponentially as the conventional

optimal detector. Then, we extended the single-user CS technique to the multi-user scenario at hand, obtaining a further performance gain. This detector was demonstrated to perform better than the SUD detector over a linear AWGN channel, thus allowing the full exploitation of the TFP principle.

APPENDIX

In this appendix, we consider the exponential term

$$\exp\left(\frac{-\boldsymbol{x}^{(\ell)H}\boldsymbol{G}^{(\ell,m)}\boldsymbol{x}^{(m)} - \boldsymbol{x}^{(m)H}\boldsymbol{G}^{(m,\ell)}\boldsymbol{x}^{(\ell)}}{2N_0}\right).$$

At the numerator, we have

$$\begin{aligned}
&\boldsymbol{x}^{(\ell)H}\boldsymbol{G}^{(\ell,m)}\boldsymbol{x}^{(m)} + \boldsymbol{x}^{(m)H}\boldsymbol{G}^{(m,\ell)}\boldsymbol{x}^{(\ell)} \\
&= \sum_{k=0}^{K-1}\sum_{n=0}^{K-1}\left[x_k^{(\ell)*}g_{k-n}^{(\ell,m)}x_n^{(m)} + x_k^{(m)*}g_{k-n}^{(m,\ell)}x_n^{(\ell)}\right] \\
&= \sum_{k=0}^{K-1}\sum_{n=0}^{K-1}\left[x_k^{(\ell)*}g_{k-n}^{(\ell,m)}x_n^{(m)} + x_k^{(m)*}g_{n-k}^{(\ell,m)*}x_n^{(\ell)}\right] \qquad (9)
\end{aligned}$$

having exploited the fact that $\boldsymbol{G}^{(\ell,m)} = \boldsymbol{G}^{(m,\ell)H}$. Arranging all terms in square brackets into a matrix $\boldsymbol{A}$, whose generic element is

$$\boldsymbol{A}_{k,n} = x_k^{(\ell)*}g_{k-n}^{(\ell,m)}x_n^{(m)} + x_k^{(m)*}g_{n-k}^{(\ell,m)*}x_n^{(\ell)}$$

the double summation in (9) can be interpreted as the sum of all the elements of this matrix. It is straightforward to show that matrix $\boldsymbol{A}$ is Hermitian. Hence, the sum of all its elements can be computed as the sum of all the elements of its main diagonal, plus two times the real part of the sum of all its elements below the main diagonal. This allows to express the kth term of the sum over k, as a function of the kth symbol and the previous ones, introducing a sort of

causality in the expression, fundamental for a trellis definition:

$$
\begin{aligned}
&\boldsymbol{x}^{(\ell)H}\boldsymbol{G}^{(\ell,m)}\boldsymbol{x}^{(m)} + \boldsymbol{x}^{(m)H}\boldsymbol{G}^{(m,\ell)}\boldsymbol{x}^{(\ell)} \\
=&\sum_{k=0}^{K-1} x_k^{(\ell)*} g_0^{(\ell,m)} x_k^{(m)} + x_k^{(m)*} g_0^{(\ell,m)*} x_k^{(\ell)} \\
&+ 2\Re\left[\sum_{k=1}^{K-1}\sum_{n=0}^{k-1}\left(x_k^{(\ell)*} g_{k-n}^{(\ell,m)} x_n^{(m)} + x_k^{(m)*} g_{n-k}^{(\ell,m)*} x_n^{(\ell)}\right)\right] \\
=&2\Re\left[\sum_{k=0}^{K-1} x_k^{(\ell)*} g_0^{(\ell,m)} x_k^{(m)}\right] \\
&+ 2\Re\left[\sum_{k=1}^{K-1}\sum_{q=1}^{k}\left(x_k^{(\ell)*} g_q^{(\ell,m)} x_{k-q}^{(m)} + x_k^{(m)*} g_{-q}^{(\ell,m)*} x_{k-q}^{(\ell)}\right)\right] \\
=&2\Re\left[\sum_{k=0}^{K-1} x_k^{(\ell)*} g_0^{(\ell,m)} x_k^{(m)}\right] \\
&+ 2\Re\left[\sum_{k=0}^{K-1}\sum_{q=1}^{L}\left(x_k^{(\ell)*} g_q^{(\ell,m)} x_{k-q}^{(m)} + x_k^{(m)*} g_{-q}^{(\ell,m)*} x_{k-q}^{(\ell)}\right)\right] .
\end{aligned}
$$

The result is thus

$$
\begin{aligned}
&\exp\left(\frac{-\boldsymbol{x}^{(\ell)H}\boldsymbol{G}^{(\ell,m)}\boldsymbol{x}^{(m)} - \boldsymbol{x}^{(m)H}\boldsymbol{G}^{(m,\ell)}\boldsymbol{x}^{(\ell)}}{2N_0}\right) \\
&= \prod_{k=0}^{K-1} \exp\left(\frac{1}{N_0}\Re\left[x_k^{(\ell)*} g_0^{(\ell,m)} x_k^{(m)}\right.\right. \\
&\left.\left. + \sum_{q=1}^{L}\left(x_k^{(\ell)*} g_q^{(\ell,m)} x_{k-q}^{(m)} + x_k^{(m)*} g_{-q}^{(\ell,m)*} x_{k-q}^{(\ell)}\right)\right]\right) .
\end{aligned}
$$

REFERENCES

[1] J. E. Mazo, "Faster-than-Nyquist signaling," *Bell System Tech. J.*, vol. 54, pp. 1450–1462, Oct. 1975.

[2] F. Rusek and J. B. Anderson, "The two dimensional Mazo limit," in *Proc. IEEE International Symposium on Information Theory*, Adelaide, Australia, Nov. 2005, pp. 970–974.

[3] 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.

[4] A. Modenini, F. Rusek, and G. Colavolpe, "Faster-than-Nyquist signaling for next generation communication architectures," in *Proc. 22nd European Signal Processing Conf. (EUSIPCO'14)*, Lisbon, Portugal, 2014.

[5] A. Modenini, G. Colavolpe, and N. Alagha, "How to significantly improve the spectral efficiency of linear modulations through time-frequency packing and advanced processing," in *Proc. IEEE Intern. Conf. Commun.*, Ottawa, Canada, Jun. 2012, pp. 3299–3304.

[6] A. Piemontese, A. Modenini, G. Colavolpe, and N. Alagha, "Improving the spectral efficiency of nonlinear satellite systems through time-frequency packing and advanced processing," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3404–3412, Aug. 2013.

[7] G. Colavolpe and T. Foggi, "Time-frequency packing for high capacity coherent optical links," *IEEE Trans. Commun.*, vol. 62, pp. 2986–2995, Aug. 2014.

[8] 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 Signal Processing Mag.*, vol. 31, no. 6, pp. 80–93, Nov. 2014.

[9] V. Franz and J. B. Anderson, "Concatenated decoding with a reduced-search BCJR algorithm," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 186–195, Feb. 1998.

[10] G. Colavolpe, G. Ferrari, and R. Raheli, "Reduced-state BCJR-type algorithms," *IEEE J. Select. Areas Commun.*, vol. 19, no. 5, pp. 848–859, May 2001.

[11] Y. Ma, N. Wu, A. Zhang, B. Li, and L. Hanzo, "Generalized approximated message passing equalization for multi-carrier faster-than-nyquist signaling," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2021.

[12] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, pp. 1046–1061, Jul. 1999.

[13] M. Tüchler, A. C. Singer, and R. Koetter, "Minimum mean square error equalization using *a priori* information," *IEEE Trans. Signal Processing*, vol. 50, pp. 673–683, Mar. 2002.

[14] M. Sikora and D. J. Costello, Jr., "A new SISO algorithm with application to turbo equalization," in *Proc. IEEE International Symposium on Information Theory*, Adelaide, Australia, Sep. 2005, pp. 2031–2035.

[15] D. Fertonani, A. Barbieri, and G. Colavolpe, "Reduced-complexity BCJR algorithm for turbo equalization," *IEEE Trans. Commun.*, vol. 55, no. 12, pp. 2279–2287, Dec. 2007.

[16] F. Rusek, M. Loncar, and A. Prlja, "A comparison of Ungerboeck and Forney models for reduced-complexity ISI equalization," in *Proc. IEEE Global Telecommun. Conf.*, Washington, DC, U.S.A., Nov. 2007.

[17] Q. Guo and L. Ping, "LMMSE turbo equalization based on factor graph," *IEEE J. Select. Areas Commun.*, vol. 26, pp. 311–319, Feb. 2008.

[18] J. B. Anderson, A. Prlja, and F. Rusek, "New reduced state space BCJR algorithms for the ISI channel," in *Proc. IEEE International Symposium on Information Theory*, Seoul, Korea, 2009.

[19] G. Colavolpe, D. Fertonani, and A. Piemontese, "SISO detection over linear channels with linear complexity in the number of interferers," *IEEE J. Sel. Topics in Signal Proc.*, vol. 5, pp. 1475–1485, Dec. 2011.

[20] F. Rusek and A. Prlja, "Optimal channel shortening for MIMO and ISI channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 810–818, Feb. 2012.

[21] 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, Oct. 2017.

[22] T. Foggi, "On performance limits for spectrally efficient optical transmission techniques in short-haul metro/access links," *Journal of Lightwave Technology*, vol. 38, no. 3, pp. 661–667, 2020.

[23] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[24] A. Modenini, F. Rusek, and G. Colavolpe, "Optimal transmit filters for ISI channels under channel shortening detection," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4997–5005, Dec. 2013.

[25] ——, "Adaptive rate-maximizing channel-shortening for ISI channels," *IEEE Commun. Letters*, vol. 19, no. 12, pp. 2090–2093, Dec. 2015.

[26] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.

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

[28] G. Colavolpe and A. Barbieri, "On MAP symbol detection for ISI channels using the Ungerboeck observation model," *IEEE Commun. Letters*, vol. 9, no. 8, pp. 720–722, Aug. 2005.

[29] F. Rusek, G. Colavolpe, and C. E. Sundberg, "40 years with the Ungerboeck model: Overlooked opportunities," *IEEE Signal Processing Mag.*, vol. 32, no. 3, pp. 156–161, May 2015.

[30] M. Jana, L. Lampe, and J. Mitra, "Precoded time-frequency-packed multicarrier faster-than-Nyquist transmission," in *Proc. IEEE Intern. Work. on Signal Processing Advances for Wireless Commun.*, Cannes, France, July 2019.

[31] N. Wiberg, "Codes and decoding on general graphs," Ph.D. dissertation, Linköping University (Sweden), 1996.

[32] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[33] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[34] ETSI EN 302 307-2 Digital Video Broadcasting (DVB), Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications, Part II: S2-Extensions (DVB-S2X), Available on ETSI web site (http://www.etsi.org).

[35] 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.

[36] 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.

[37] 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.

[38] P. Roberston, E. Villebrun, and P. Hoeher, "Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding," *European Trans. Telecommun.*, vol. 8, no. 2, pp. 119–125, March/April 1997.