Efficient Message-Passing Detection for Multi-Satellite Systems Using OTFS Modulation

Elisa Conti, Amina Piemontese, Tommaso Foggi, Giulio Colavolpe, and Armando Vannucci
CNIT Research Unit
University of Parma, 43124 Parma, Italy
{elisa.conti, amina.piemontese, tommaso.foggi, giulio.colavolpe, armando.vannucci}@unipr.it

Abstract— In this study, we analyze a low Earth orbit multi-satellite communication system using orthogonal time frequency space (OTFS) modulation. In this context, the choice of the OTFS modulation is driven by its robustness against high Doppler shifts in doubly-selective channels. This scenario is motivated by the need for a more stable and reliable system, which can be achieved through diversity, i.e., allowing each user to be jointly served by multiple satellites. However, numerous challenges arise due to the need to properly combine the different received signals, each one characterized by its own Doppler shift, delay and phase. This paper focuses on soft-output data detection algorithms and proposes a novel message passing (MP)-based approach that leverages both the channel sparsity in the Doppler-delay domain and the particular structure assumed by the equivalent channel matrix, derived from the compact block-wise input-output relation, according to the Forney observation model for linear modulations over additive white Gaussian noise channels. The proposed approach is the first one able to significantly reduce the complexity by acting on the choice of interferers, organized in diagonals, and on the schedule, by prioritizing the strongest elements. We assess the detector's performance by evaluating its pragmatic capacity, i.e., the achievable rate of the channel induced by the signal constellation and the detector soft-output. Simulation results demonstrate that the proposed solution achieves an optimal trade-off between complexity and performance compared to the alternative state-of-the-art solutions.

Table of Contents

1. Introduction

Non-terrestrial networks (NTN) will gain fundamental importance in the advancement of future 6G wireless systems. High-altitude platforms and space-born satellites will integrate and complement terrestrial networks ensuring seamless connectivity and supporting the rapidly growing data rate demands [?], [?]. Among the various kinds of NTN communications, low Earth orbit (LEO) satellites, which are located at an altitude of 500-2000 km, have attracted particular interest. Unlike traditional geostationary and medium Earth orbit satellites, which operate at altitudes of 35786 km and 5000-20000 km, respectively, LEO satellites offer lower propagation latency and reduced path loss.

LEO satellites mega-constellations can be exploited to enable macro-diversity schemes allowing each user to be jointly served by multiple satellites. This enhances the system reliability, especially in scenarios where the line-of-sight link between the satellite and the user terminal (UT) is obstructed by physical impairments on the ground. Given that satellites occupy different positions in the sky and follow independent trajectories, shadowing effects impact satellite-to-UT links in a mutually independent way. As a results, diversity is essential not only for ensuring a uniform throughput but also for drastically reducing the link outage probability. However, signals from diverse satellites typically reach the UT at distinct time epochs and are also affected by different Doppler shifts and phases. Therefore, while higher diversity can significantly improve performance, the UT must effectively combine the received information to benefit from multi-satellite configurations.

To fully exploit the advantages of diversity, the adoption of a proper modulation scheme plays a crucial role. The orthogonal time frequency space (OTFS) modulation (see [?],[?] and references therein) provides inherent robustness to Doppler shifts and suitability to sparse channels in the Doppler-delay domain, as the one under consideration. In fact, high Doppler shifts directly derive from the relative motion of LEO satellites with respect to UT on ground.

Several studies have already shown the advantages of OTFS with respect to other modulation formats. For instance, [?] offers a comparison between orthogonal frequency division multiplexing (OFDM) and OTFS in sparse channel conditions (e.g., outdoor scenarios with few reflectors). The analysis conducted in [?],[?] focuses instead on the application of OTFS in NTN and demonstrates the increased robustness of OTFS compared to OFDM in handling the significant Doppler effects typical of satellite communications. Other relevant contributions address the integration of OTFS with (massive) MIMO satellite systems [?], [?].

This paper addresses the use of OTFS for LEO multi-satellite communications, considering a scenario similar to that in [?]. In particular, it focuses on the analysis of the detection strategies under the assumption of perfect channel state information (CSI) at the receiver. We propose a novel soft-output algorithm based on message-passing (MP) [?]. Differently from the low-complexity approaches available in the literature [?],[?], the presented algorithm leverages the particular structure assumed by the channel matrix in the Doppler-delay domain. Moreover, it introduces an innovative scheduling approach, where the strongest interferers are prioritized

in the detection process. We consider separate detection and decoding, i.e., without "turbo" reprocessing of the decoder output, and test the system performance by evaluating the pragmatic capacity, i.e., the achievable rate of the channel induced by the signal constellation at its input and the detector soft-output [?],[?]. This performance metric represents the information theoretical rates achievable through separate detection and decoding when a particular signal constellation and soft-output symbol detector are used. As a result, it offers greater practical relevance compared to the typically examined uncoded bit error rate. The simulated satellite link scenarios are based on the Starlink constellation, considering various numbers of satellites in visibility with the UT. For comparison purposes, we tested (i) the linear minimum mean square error (MMSE) block equalizer, which is characterized by an extremely high complexity due to a large-dimensional matrix inversion, (ii) a low-complexity algorithm taken from the existing literature based on Gaussian interference approximation [?], and (iii) an MP-based detection scheme firstly proposed in [?] and applied to OTFS in [?]. Numerical results demonstrate that multi-satellite diversity effectively enhances the system performance and that the proposed solution reaches, and in some scenarios outperforms, the MMSE detector with a considerably lower complexity. Moreover, given the constraint on the computational load, significant performance advantages are observed with respect to the algorithms proposed in [?] and [?].

The rest of this paper is organized as follows. In Section 2, we present the system model of the analyzed scenario. Section 3 is dedicated to the description of the detectors considered for comparison purposes. The proposed algorithm is summarized and discussed in Section 4. In Section 5, we report and comment the numerical results and, finally, Section 6 provides some concluding remarks.

2. System Model

We consider a multi-satellite communication system consisting of LEO satellites transmitting the same OTFS-modulated signal to a single-antenna UT on ground. The analyzed scenario is outlined in Figure 1. Because of the relative movement of each satellite with respect to the UT, only P satellites are supposed to be in visibility with the UT, where P can be much smaller than the total number of satellites composing the constellation. The signals transmitted from the P satellites undergo different propagation delays $\{\tau_p\}_{p=1}^{P}$ and different Doppler shifts $\{\nu_p\}_{p=1}^{P}$ due to the different distances of the satellites from the UT and the different relative speeds. We will assume that the transmission system is designed such that the propagation delays and the Doppler shifts can be considered as constant for the duration of a transmitted frame.

Within the OTFS modulation, data-symbols $\{x[k,l]\}$ (belonging to a finite alphabet $\mathcal{A}$) for $k = 0, 1, \ldots, N-1$ and $l = 0, 1, \ldots M-1$ are arranged into an $N \times M$ grid in the Doppler-delay domain. These symbols are assumed to be spaced by $1/NT$ in the Doppler domain and $1/M\Delta f$ in the delay domain, where T is the symbol duration and Δf is the subcarrier spacing, i.e., $\Delta f = B/M$ being B the total available bandwidth.

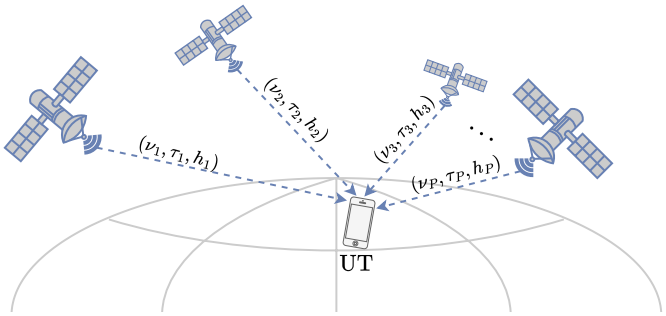


Figure 1. A representation of the analyzed scenario where a number of satellites serve the same UT to realize macro-diversity.

The values of T and Δf are usually selected in such a way that $\max_p\{\tau_p\} < T$ and $\max_p\{\nu_p\} < \Delta f$. It is worth to underline that we do not make the simplifying unrealistic assumption that Doppler and delay shifts are integer multiples of the symbol grid. Symbols are then converted to the time-frequency domain through the so-called inverse symplectic finite Fourier transform (ISFFT), i.e.,

$$X[n,m] = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} x[k,l] e^{\jmath 2\pi \left(\frac{nk}{N} - \frac{ml}{M}\right)} \tag{1}$$

for $n = 0, \ldots, N-1$ and $m = 0, \ldots, M-1$. Consequently, the continuous-time transmitted signal $s(t)$ can be expressed as

$$s(t) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} X[n,m] p_{\text{tx}}(t - nT) e^{\jmath 2\pi m \Delta f (t-nT)}\,, \tag{2}$$

i.e., symbol $X[n,m]$ is transmitted at time n and over subcarrier m, and $p_{\text{tx}}(t)$ is a transmit shaping pulse. The transform applied in (2), which generates the signal $s(t)$ from symbols $\{X[n,m]\}$, is usually called Heisemberg transform in the OTFS literature. The general expression in (2) can be specialized to represent any of the multicarrier modulation formats available in the literature by selecting a particular shaping pulse and a proper value of T and Δf. For instance, setting $\Delta f = 1/T$ and adopting a rectangular shaping pulse with duration T, (2) represents a classical OFDM modulation with properly precoded information symbols.

The analyzed scenario is equivalent to the one where the information bearing signal is transmitted through a multipath time-varying channel modeled as

$$h(t,\tau) = \sum_{p=1}^{P} h_p e^{\jmath 2\pi \nu_p t} \delta(\tau - \tau_p) \tag{3}$$

where $\{h_p\}_{p=1}^{P}$ are the complex channel gains accounting for the different path losses, the complex envelope of the received signal $r(t)$ at the UT can be expressed as

$$r(t) = \sum_{p=1}^{P} h_p s_p(t - \tau_p) e^{\jmath 2\pi \nu_p t} + w(t)\,, \tag{4}$$

where $s_p(t)$ is the complex envelope of the signal transmitted by the p-th satellite and $w(t)$ models the complex additive white Gaussian noise (AWGN) whose real and imaginary components have power spectral density σ^2. The signal $s_p(t)$ transmitted by the p-th satellite includes the factors $\widetilde{\nu}_p$, $\widetilde{\theta}_p$ and $\widetilde{\tau}_p$ required for the compensation of the frequency offset, phase and delay, respectively, and can thus be expressed as

$$s_p(t) = e^{j(2\pi\widetilde{\nu}_p t+\widetilde{\theta}_p)} s(t-\widetilde{\tau}_p)\,, \tag{5}$$

where $s(t)$ is the ordinary information-bearing waveform sent by all the P satellites serving the considered UT.

At the receiver side, the resulting signal at the output of a filter matched to $p_{\rm rx}(t)e^{j2\pi ft}$, where $p_{\rm rx}(t)$ is a receive shaping pulse, is

$$Y(t,f) = \int r(t')p^*_{\rm rx}(t'-t)e^{-j2\pi f(t'-t)}\,dt'\,. \tag{6}$$

By sampling $Y(t,f)$ at $t=nT$ and $f=m\Delta f$, we obtain the samples $\{Y[n,m]\}$. In the OTFS literature, the filtering of the received signal with the bank of matched filters plus sampling is usually called Wigner transform. Finally, through the symplectic finite Fourier transform (SFFT) it is possible to obtain the received samples $\{y[k,l]\}$, for $k=0,\ldots,N-1$ and $l=0,\ldots,M-1$ in the Doppler-delay domain:

$$y[k,l] = \frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} Y[n,m]e^{j2\pi\left(-\frac{nk}{N}+\frac{ml}{M}\right)}\,. \tag{7}$$

In the following, we will assume $p_{\rm rx}(t)=p_{\rm tx}(t)$, $\Delta f = 1/T$, and that $p_{\rm tx}(t)$ is a rectangular pulse with duration T. Under these assumptions, the noise samples affecting the useful signal in $\{Y[n,m]\}$ are white. Moreover, since the SFFT does not color the noise, the same consideration holds also for $\{y[k,l]\}$. These noise samples will be omitted for the sake of notational simplicity.

Under the further assumption of absence of interblock interference, the received samples $y[k,l]$ can be expressed as [?], [?]

$$y[k,l] = \sum_{k',l'} x[k',l']g_{k,k'}[l,l']\,, \tag{8}$$

where the intersymbol interference (ISI) coefficient of the Doppler-delay pair $[k',l']$ seen by sample $[k,l]$ is given by [?], [?]

$$g_{k,k'}[l,l'] = \sum_{p=1}^{P} h_p e^{j2\pi\nu_p\tau_p}\mathbf{\Psi}^p_{k,k'}[l,l']\,, \tag{9}$$

where the expression for the entries of the channel matrix $\mathbf{\Psi}^p_{k,k'}$ can be found in [?], [?].

It follows from (8) that each symbol $x[k,l]$ in the Doppler-delay grid can, in principle, interfere with any other symbol within the same OTFS block. We therefore obtain a linear system characterized by a two-dimensional ISI. This is the reason for which the receiver has to process the entire OTFS block and could not take into account a subset of samples only. The magnitude of $\mathbf{\Psi}^p_{k,k'}[l,l']$ depends on $\{k,l,k',l'\}$ through the differences $k-k'$ and $l-l'$. Defining the Dirichlet kernel function

$$D_n(x) = \frac{1-e^{j2\pi x}}{1-e^{j2\pi x/n}}\,,$$

we can express

$$\left|\mathbf{\Psi}^p_{k,k'}[l,l']\right| \simeq \frac{1}{NM}\left|D_N(k'-k+\nu_p NT)\right| \cdot\left|D_M(l-l'-\tau_p M\Delta f)\right|\,.$$

The input-output equation (8) can be organized in matrix form. Writing the $N\times M$ matrices of transmitted symbols and received samples as NM-dimensional column vectors (stacking the columns of the corresponding matrices on top of each other), we obtain the block-wise input-output relation in the form [?]

$$\boldsymbol{y} = \mathbf{\Psi}\boldsymbol{x} + \boldsymbol{w}\,, \tag{10}$$

where, defining $h'_p = h_p e^{j2\pi\nu_p\tau_p}$,

$$\mathbf{\Psi} = \sum_{p=1}^{P} h'_p\mathbf{\Psi}_p \tag{11}$$

and matrices $\{\mathbf{\Psi}_p\}$ have dimension $NM\times NM$ and are obtained by properly stacking the entries of the $M\times M$ matrices $\mathbf{\Psi}^p_{k,k'}$ for $k,k'=0,\ldots,N-1$, while $\boldsymbol{w}$ denotes the additive white Gaussian noise (AWGN) with zero mean and covariance matrix $2\sigma^2\mathbb{I}_{NM}$, where $\mathbb{I}_{NM}$ is the $NM\times NM$ identity matrix.

The channel effect on each transmitted symbol in the Doppler-delay grid can be described as follows. As an example, consider the transmission over the time-frequency selective channel in (3) of an OTFS block composed by all zero symbols except one (placed anywhere in the grid) with enough energy to be distinguishable. The received signal from the p-th satellite shows the transmitted symbol (circularly) shifted in the grid according to the pair $(\nu_p,\ \tau_p)$. Moreover, if the involved channel parameters are integer multiples of the grid resolution, the whole received energy is confined in a single grid position. Otherwise, in the general and realistic scenario of Doppler and delay with a fractional component, an energy leakage, which is dictated by the Dirichlet kernel functions included in the channel matrix $\mathbf{\Psi}$, appears in the positions surrounding the peak value. Due to the system's linearity, the overall effect of the multiple paths results in the superposition of the shifts associated with each component. In Figure 2, an example of the symbol shift in a multipath scenario is shown.

3. State-of-art detection algorithms

In the following, the soft-output state-of-art algorithms for OTFS detection, which are tested for comparison purposes, will be presented. The first one belongs to the family of linear equalizers, it shows a good performance but it is characterized by an extremely high complexity. Differently, the other soft-output algorithms that

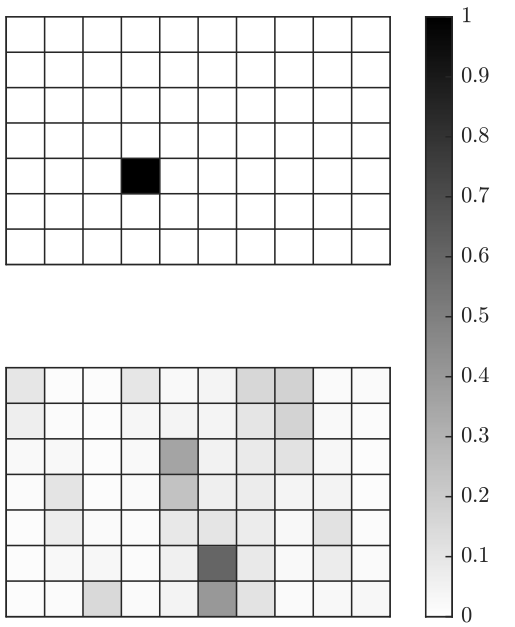


Figure 2. Channel effect on a generic transmitted symbol. In the top subfigure, the transmitted symbol (magnitude) in the Doppler-delay grid. In the bottom subfigure, the received energy distribution after the propagation through the multipath channel with $P = 5$ satellites in visibility with the UT.

we analyzed [?], [?] exploit the framework based on Factor Graphs (FGs) and the Sum Product Algorithm (SPA) [?] as other recently proposed approaches aiming to a complexity reduction [?]. The performance will be assessed in Section 5 in terms of pragmatic capacity.

A. Linear MMSE Block-Wise Equalization

Denoting with x_k the k-th entry of the column vector $\boldsymbol{x}$ in (10), the detector provides at its output the noisy estimate $\hat{x}_k$, $k = 1, \ldots, NM$. This detector is derived under the assumption that $\boldsymbol{x}$ is a Gaussian vector. The optimal estimator $\hat{\boldsymbol{x}}$ of $\boldsymbol{x}$ according to the MMSE criterion results to be linear with the observation $\boldsymbol{y}$ and has expression

$$\hat{\boldsymbol{x}} = \boldsymbol{\Psi}^H \left[\boldsymbol{\Psi}\boldsymbol{\Psi}^H + 2\sigma^2 \mathbb{I}_{NM} \right]^{-1} \boldsymbol{y} \,, \tag{12}$$

where $(\cdot)^H$ denotes the Hermitian, i.e., transpose conjugate, operator. Finally, for the computation of the pragmatic capacity we need the a-posteriori probability mass function (pmf) of the transmitted symbols x_k given the detector soft-output $\hat{x}_k$, i.e., $P(x_k \,|\, \hat{x}_k)$. To this purpose, the estimate $\hat{x}_k$ is treated as the output of a (virtual) AWGN channel. We can thus express the k-th symbol as a function of the relative estimate $\hat{x}_k$ as

$$x_k = \hat{x}_k + \epsilon_k \,, \tag{13}$$

where ϵ_k is the zero-mean estimation error whose variance $\sigma^2_{\epsilon,k}$ corresponds to the element in position (k,k) of the following error covariance matrix:

$$\boldsymbol{C}_\epsilon = \left(\mathbb{I}_{NM} + \frac{\boldsymbol{\Psi}^H \boldsymbol{\Psi}}{2\sigma^2} \right)^{-1} . \tag{14}$$

The complexity of this detector is $\mathcal{O}\big((NM)^3\big)$, since its implementation requires the inversion of the $NM \times NM$ matrix in (12). In a time-varying scenario, when the channel matrix $\boldsymbol{\Psi}$ changes at every block, and for typical values of N and M (e.g., $N = 50$ and $M = 64$ as the ones used in the simulations), this complexity cannot be afforded.

B. Matrix $\boldsymbol{\Psi}$-Based Algorithm ($\mathrm{MP}_{\boldsymbol{\Psi}}$)

In the context of FG-based algorithms, depending on how the conditional pmf of the modulation symbols $\boldsymbol{x}$ given the received samples $\boldsymbol{y}$

$$P(\boldsymbol{x} \,|\, \boldsymbol{y}) \propto p(\boldsymbol{y} \,|\, \boldsymbol{x}) P(\boldsymbol{x}) \tag{15}$$

is factorized, different algorithms can be derived.

The MP algorithm of [?], referred to as $\mathrm{MP}_{\boldsymbol{\Psi}}$, builds its FG starting from the following factorization of (15)

$$P(\boldsymbol{x} \,|\, \boldsymbol{y}) \propto \prod_{k=1}^{NM} p(y_k \,|\, \boldsymbol{x}) P_k(x_k) \,, \tag{16}$$

which is based on the assumption that the modulation symbols x_k are treated by the detector as independent and identically distributed (i.i.d.) with given (typically uniform) a-priori pmfs $P_k(x_k)$, $k = 1, \ldots, NM$. We define the factor nodes (FNs)

$$Q_k(\boldsymbol{x}) \triangleq p(y_k \,|\, \boldsymbol{x}) = \exp\left(-\frac{\|y_k - \boldsymbol{\Psi}_k \boldsymbol{x}\|^2)}{2\sigma^2} \right) , \tag{17}$$

where $\boldsymbol{\Psi}_k$ denotes the k-th row of the channel matrix $\boldsymbol{\Psi}$. The resulting FG is shown in Figure 3. This FG has girth (i.e., minimum length of the cycles) equal to 4, as highlighted in Figure 3 and, as it is known from graphs theory [?], the cycles' length influences the accuracy of the approximation of the a-posteriori marginal pmf. As directly follows from the FNs definition (17), the j-th variable node (VN) is connected to a number of FNs equal to the number of nonzero elements in the j-th column of $\boldsymbol{\Psi}$ and the k-th FN is connected to all the symbols corresponding to the positions of the nonzero elements of the k-th row of $\boldsymbol{\Psi}$. Since in the general case of delay and Doppler shifts with a fractional component the channel matrix is (almost) full, the FG presents an extremely high number of edges. Denoting with l_k the number of nonzero elements in the k-th row of $\boldsymbol{\Psi}$ which is, therefore, the degree of the FN $Q_k(\boldsymbol{x})$, the exact application of the SPA would require to sum over $l_k - 1$ discrete variable taking values in $\mathcal{A}$. Thus, the computational load is $|\mathcal{A}|^{l_k - 1}$ which, especially for large constellations and/or real channel conditions, is highly impractical. In [?], the authors propose to adopt a Gaussian approximation for the interfering symbols in the computation at the FN $Q_k(\boldsymbol{x})$, which decreases the computational load.

In fact, the approach proposed in [?] relies on the following approximation. When computing the message from FN $Q_k(\boldsymbol{x})$ to VN x_j, the k-th received sample, which can be expressed as

$$y_k = \psi_{k,j} x_j + \sum_{l \in n(k) \setminus \{j\}} \psi_{k,l} x_l + w_k \tag{18}$$

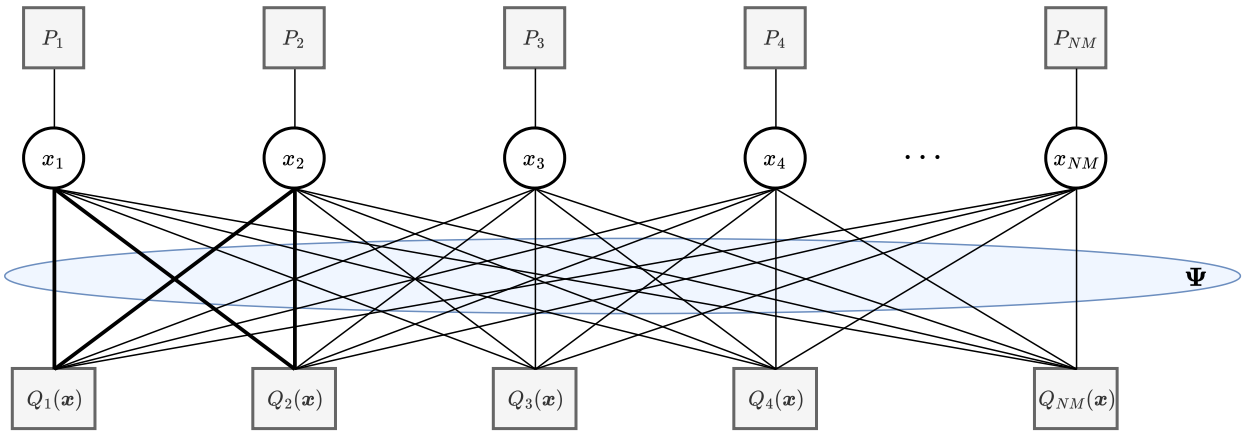


Figure 3. Structure of the FG for the Matrix $\boldsymbol{\Psi}$-based algorithm.

where w_k is the k-th component of the AWGN vector $\boldsymbol{w}$, $\psi_{k,j}$ represents the element in the k-th row and j-th column of $\boldsymbol{\Psi}$ and $n(k)$ is the set of VNs connected to FN $Q_k(\boldsymbol{x})$, is assumed to be

$$\tilde{y}_k = \psi_{k,j} x_j + \zeta_{k \to j} \tag{19}$$

where $\zeta_{k\to j}$ is the approximation of the interference term, modeled as a Gaussian random variable. In this way, a significant performance reduction can be obtained, achieving an overall complexity of $\mathcal{O}(n_{iter} NML|\mathcal{A}|)$, where n_{iter} denotes the number of allowed inner detector iterations, and L the number of elements of each row and column of $\boldsymbol{\Psi}$ which are considered for the detection process. It is worth specifying that, in order to increase the sparsity of $\boldsymbol{\Psi}$ and, therefore, decrease the number of edges in the graph of Figure 3, the detector proposed in [?] considers the delay shifts as integer multiples of the grid resolution. For this reason, the number L of nonzero elements per row/column is limited and much smaller than the channel matrix dimension. However, the above assumption is generally not satisfied by real channels and the matrix $\boldsymbol{\Psi}$ is typically full. Moreover, making the detector work on an approximated channel matrix constructed by rounding the delay shifts to integer multiples of the grid leads to a significant performance degradation. Therefore, in the simulation results presented in Section 5, we let the $\text{MP}_{\boldsymbol{\Psi}}$ detector work on the exact matrix $\boldsymbol{\Psi}$ and we set the value of L according to the complexity constraints and the choice of the interfering terms per symbol is done by selecting the most significant entries in terms of energy.

C. Matrix $\boldsymbol{G}$-Based Algorithm ($\text{MP}_{\boldsymbol{G}}$)

Among the MP algorithms working on FGs, an effective one, referred to as $\text{MP}_{\boldsymbol{G}}$, is proposed in [?] and recently applied to OTFS in [?]. Differently from the $\text{MP}_{\boldsymbol{\Psi}}$ algorithm, it works on a processed version of $\boldsymbol{\Psi}$.

This algorithm is derived from a different factorization of the conditional probability density function (pdf) of the received samples $\boldsymbol{y}$ given the symbols $\boldsymbol{x}$. In fact, starting from the system model in (10) and assuming the channel matrix $\boldsymbol{\Psi}$ to be perfectly known, (15) can be computed as

$$p\left(\boldsymbol{y}|\boldsymbol{x}\right) \propto \exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{\Psi}\boldsymbol{x}\|^2}{2\sigma^2}\right) \tag{20}$$

$$\propto \exp\left(\frac{2\Re\left\{\boldsymbol{x}^H\boldsymbol{z}\right\} - \boldsymbol{x}^H\boldsymbol{G}\boldsymbol{x}}{2\sigma^2}\right), \tag{21}$$

where

$$\boldsymbol{z} = \boldsymbol{\Psi}^H\boldsymbol{y}\,, \tag{22}$$

$$\boldsymbol{G} = \boldsymbol{\Psi}^H\boldsymbol{\Psi}\,. \tag{23}$$

From (21), it is possible to notice that $\boldsymbol{z}$ is a sufficient statistic for symbols detection. The algorithm is thus developed on the Ungerboeck observation model since it performs the detection process operating on $\boldsymbol{G}$. Expressing in (21) the matrices products explicitly in terms of their elements, it is possible to factorize the conditional pmf in (15) as

$$P(\boldsymbol{x}\,|\,\boldsymbol{y}) \propto \prod_{k=1}^{NM}\left[P_k\left(x_k\right)F_k\left(x_k\right)\prod_{l<k} I_{k,l}\left(x_k,x_l\right)\right], \tag{24}$$

where we defined

$$F_k\left(x_k\right) \triangleq \exp\left[\frac{1}{\sigma^2}\Re\left\{z_k x_k^* - \frac{1}{2}G_{k,k}\left|x_k\right|^2\right\}\right], \tag{25}$$

$$I_{k,l}\left(x_k,x_l\right) \triangleq \exp\left[-\frac{1}{\sigma^2}\Re\left\{G_{k,l}x_l x_k^*\right\}\right]. \tag{26}$$

The corresponding FG is shown in Figure 4. The resulting $\text{MP}_{\boldsymbol{G}}$ algorithm consists in a direct application of the SPA to the obtained graph. The details of the resulting MP algorithm can be found in [?]. An important property of this approach is that the FG has girth 6 as highlighted in Figure 4, therefore larger than that of [?] working on the FG of Figure 3. Moreover, the complexity of the $\text{MP}_{\boldsymbol{G}}$ algorithm is linear in the number of interferers per symbol L [?] and quadratic in the constellation size $|\mathcal{A}|$, i.e., the overall computational load is $\mathcal{O}(n_{iter}NML|\mathcal{A}|^2)$. As discussed for the $\text{MP}_{\boldsymbol{\Psi}}$, in the simulation results reported in Section 5 we selected the L terms bringing the highest amount of interference, where L is set such that the complexity constraints are satisfied.

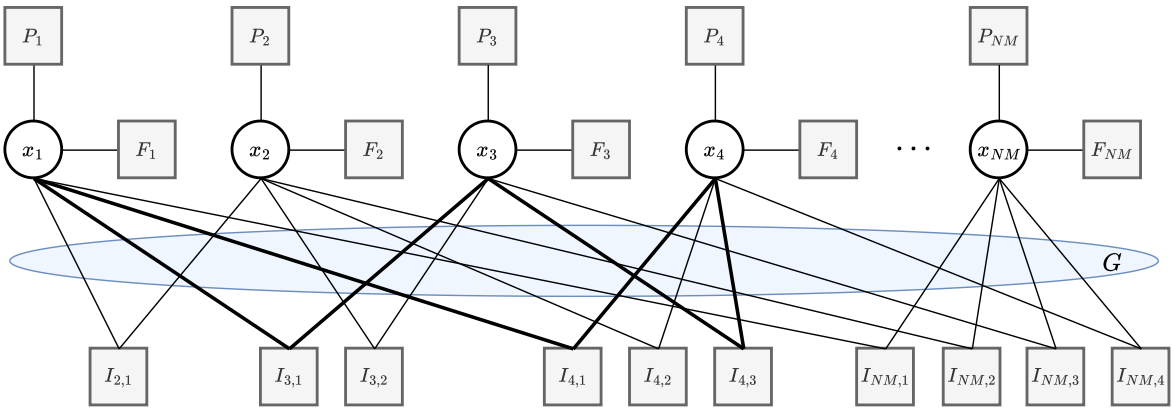


Figure 4. Structure of the FG for the Matrix $\boldsymbol{G}$-based algorithm.

4. The Proposed Algorithm

In this section, we propose a new low-complexity MP algorithm which, similarly to the MP$_{\boldsymbol{G}}$ approach discussed in Section 3C, works on the processed channel matrix $\boldsymbol{G}$. Differently from the existing solutions in the OTFS literature (see [?], [?], [?] and references therein), this algorithm takes advantage of the specific structure assumed by the matrix $\boldsymbol{G}$ which is:

- Sparse, as it represents a processed version of the channel matrix in the Doppler-delay domain;
- Arranged in diagonals: its entries (i.e., the coefficients responsible for determining the amount of interference) are organized in a relatively ordered way. Specifically, the most relevant ones, i.e., the terms with the largest magnitudes, are located along diagonals. Each diagonal consists of elements with comparable energy, with only a few exceptions. This structure is also partially present in $\boldsymbol{\Psi}$, though to a lesser extent.

In fact, from (23) and (11), we can write $\boldsymbol{G}$ as

$$\boldsymbol{G} = \left(\sum_{p=1}^{P} h'_p \boldsymbol{\Psi}_p\right)^H \left(\sum_{t=1}^{P} h'_t \boldsymbol{\Psi}_t\right) \tag{27}$$

$$= \underbrace{\sum_{p=1}^{P} \left|h'_p\right|^2 \boldsymbol{\Psi}_p^H \boldsymbol{\Psi}_p}_{\boxed{1}:\text{Main Diagonal}} + \underbrace{\sum_{p=1}^{P} \sum_{t\neq p} (h'_p)^* h'_t \boldsymbol{\Psi}_p^H \boldsymbol{\Psi}_t}_{\boxed{2}:\text{Subdiagonals}} . \tag{28}$$

The first term in (28) identifies the main diagonal. In fact, it can be demonstrated that

$$\boldsymbol{\Psi}_p^H \boldsymbol{\Psi}_p = \mathbb{I}_{NM} , \quad \forall p = 1, \ldots, P . \tag{29}$$

Moreover, equation (29) is verified for any pair (ν_p, τ_p), no matter if they are or not integer multiples of the grid spacings. The second term in (28) is the sum of $P(P-1)$ matrices which give rise to the subdiagonals.

In the special case where τ_p and ν_p are integer multiples of the symbol grid, the matrix $\boldsymbol{\Psi}_p$ associated with the path p has one nonzero element in each row and column, all placed along (typically incomplete) diagonals. Since the product $\boldsymbol{\Psi}_p^H \boldsymbol{\Psi}_t$ in (28) shifts the diagonal positions without altering the number of elements per row/column, $\boldsymbol{G}$ results to be composed by $P(P-1)+1$ nonzero elements in each row and column. From (11), it is clear that, under these unrealistic assumptions, $\boldsymbol{\Psi}$ is sparser than $\boldsymbol{G}$, as it contains only P nonzero entries per row and column.

However, in the general case of real channels where (ν_p, τ_p) are not integer multiples of the grid granularity, the above considerations about the precise matrix structure are not valid anymore. In fact, the energy is not strictly confined along diagonals but rather spreads to positions near the peak value. Therefore, working on $\boldsymbol{G}$ can be beneficial because the product of $\boldsymbol{\Psi}$ with its Hermitian allows to obtain a more defined and compact matrix structure. An example of this behavior can be seen in Figure 5 where, on the left we reported the structure of the channel matrix $\boldsymbol{\Psi}$ by showing a map of the energy distribution and on the right that of the related matrix $\boldsymbol{G}$.

The proposed algorithm thus takes advantage of the structure assumed by the matrix $\boldsymbol{G}$ and can be summarized as follows. Firstly, the energy of each diagonal is computed as

$$E_j = \sum_{k=1}^{NM-j} \left|G_{j+k,k}\right| , \quad j = 1, \ldots, NM-1 . \tag{30}$$

Notice that, since $\boldsymbol{G}$ is Hermitian, considering the upper triangular submatrix would yield the same result. Then, select the first D diagonals with the highest energy to be used in the detection process, where D is chosen in accordance to the complexity constraints or/and optimized in order to achieve the best trade-off between complexity and performance. The greatest D energy values are sorted in decreasing order and collected in the vector $\overline{\boldsymbol{E}}$. The detection algorithm operates on the reduced Hermitian matrix $\boldsymbol{G}^r$, which is composed by D nonzero diagonals in the lower triangular submatrix and the corresponding D in the upper one, and has the

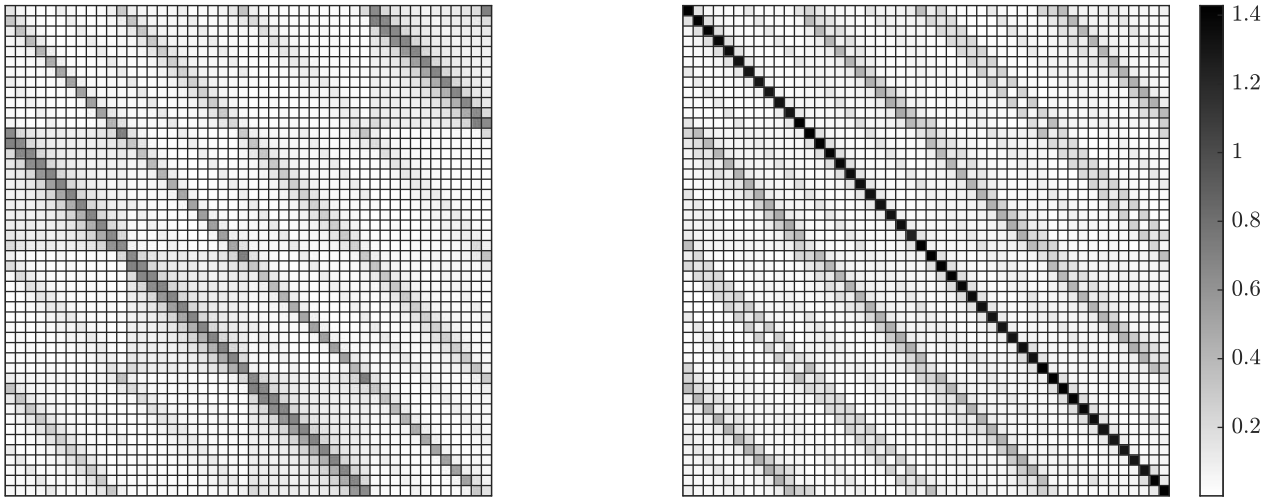
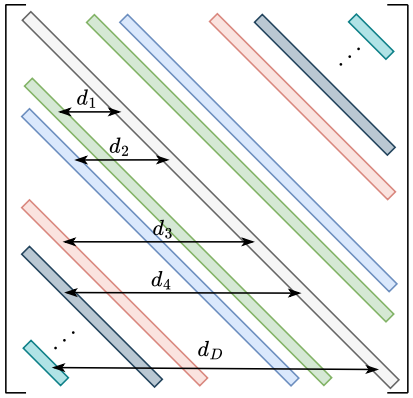


Figure 5. Comparison of the energy distribution for the matrix $\boldsymbol{\Psi}$ (on the left) and $\boldsymbol{G}$ (on the right).

following structure:

$$\boldsymbol{G}^r = \tag{31}$$



where $d_i\,, i = 1, \ldots, D$ is the distance in terms of number of elements between the main diagonal and the i-th subdiagonal. Analytically,

$$G^r_{m,n} = \begin{cases} G_{m,n}, & \text{if } m = n \vee |m-n| \in \{d_i\}_{i=1}^{D} \\ 0, & \text{elsewhere} \end{cases} \tag{32}$$

where $G^r_{m,n}$ denotes the element in the m-th row and n-th column of $\boldsymbol{G}^r$. It is worth to underline that a longer distance from the main diagonal does not imply a reduction in the magnitude of its entries. With reference to the conditional pdf in (21), by substituting $\boldsymbol{G}^r$ in place of $\boldsymbol{G}$ and explicitly expressing the matrix operations in terms of their components, we obtain

$$\boldsymbol{x}^H \boldsymbol{z} = \sum_{k=1}^{NM} z_k x_k^*\,, \tag{33}$$

$$\boldsymbol{x}^H \boldsymbol{G}^r \boldsymbol{x} = \sum_{k=1}^{NM} G^r_{k,k}|x_k|^2 + \sum_{k=1}^{NM} x_k^* \sum_{l\neq k} G^r_{k,l} x_l \tag{34}$$

$$= \sum_{k=1}^{NM} G_{k,k}|x_k|^2 + \sum_{k=1}^{NM}\sum_{j=1}^{D} 2\Re\{G_{k,k-d_j} x_{k-d_j} x_k^*\}\,. \tag{35}$$

Therefore, we can rewrite the approximate a-posteriori pmf as

$$\tilde{P}(\boldsymbol{x}|\boldsymbol{y}) \propto \prod_{k=1}^{NM} \left[P_k(x_k) F_k(x_k) \prod_{j=1}^{D} I_{k,k-d_j}(x_k, x_{k-d_j}) \right], \tag{36}$$

where the factor $F_k(x_k)$ is defined in (25) and

$$I_{k,k-d_j}(x_k, x_{k-d_j}) \triangleq \exp\left(-\frac{1}{\sigma^2} \Re\left\{ G_{k,k-d_j} x_{k-d_j} x_k^* \right\} \right). \tag{37}$$

Considering the generic VN x_k, a portion of the FG corresponding to (36) is shown in Figure 6.

Each couple of diagonals in the matrix $\boldsymbol{G}^r$ gives rise to a set of d_j parallel branches implementing independently the BCJR algorithm with unitary memory. In fact, each BCJR involves NM/d_j symbols and takes as input information a partial marginal distribution $T_k^{(d_j)}(x_k)$ defined as

$$T_k^{(d_j)}(x_k) = P_k(x_k) F_k(x_k) \prod_{i\neq j} \underbrace{\alpha_{in}^{(d_i)}(x_k)\beta_{in}^{(d_i)}(x_k)}_{O_k^{(d_i)}(x_k)} \tag{38}$$

where $\alpha_{in}^{(d_i)}(x_k)$ and $\beta_{in}^{(d_i)}(x_k)$ are the incoming messages to the VN x_k in Figure 6, and their product, $O_k^{(d_i)}(x_k)$, represents the output information about the k-th symbol generated by the BCJR corresponding to the diagonal at distance d_j. Notice that in (38), the messages $\alpha_{in}^{(d_i)}(x_k)$ and $\beta_{in}^{(d_i)}(x_k)$ correspond to their most recent updates and, in order to account for the extrinsic information on the VN x_k, we exclude from the product those messages that refer to the considered j-th BCJR. For easier visualization, we report in Figure 7 a section of the FG of Figure 6, partially reordered for the sake of readability,

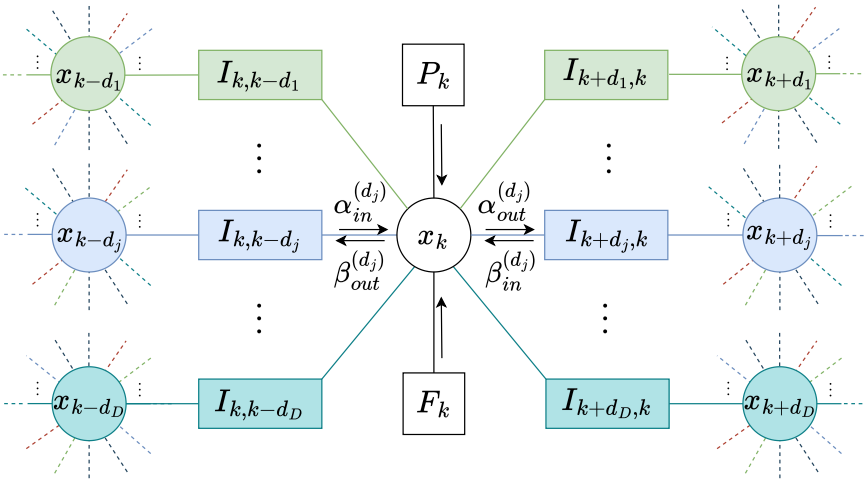


Figure 6. Partial structure of the FG for the proposed algorithm.

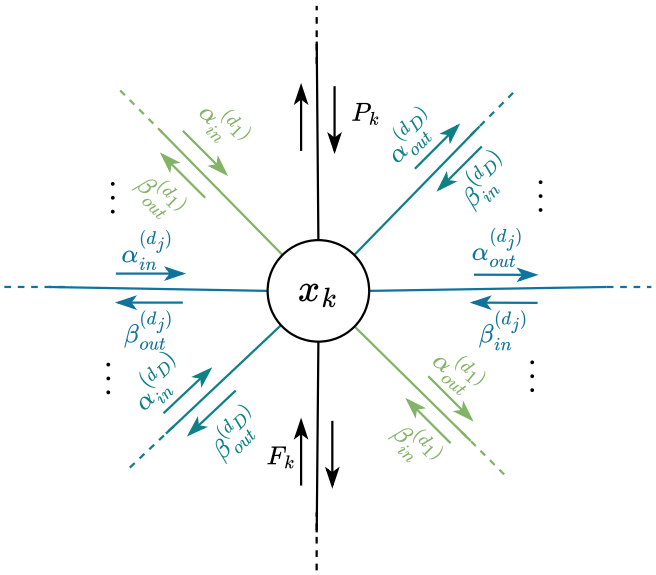


Figure 7. Overview of the outgoing and incoming messages to the VN x_k.

where the incoming and outgoing messages related to VN x_k are highlighted. With reference to the messages in Figure 7, by applying the SPA updating rules, $\alpha_{in}^{(d_j)}$, $\alpha_{out}^{(d_j)}$, $\beta_{in}^{(d_j)}$ and $\beta_{out}^{(d_j)}$ can be recursively computed by means of the following forward and backward recursions:

$$\alpha_{in}^{(d_j)}(x_k) = \sum_{\sim\{x_k\}} \alpha_{out}^{(d_j)}(x_{k-d_j}) I_{k,k-d_j}(x_k, x_{k-d_j}) \quad (39)$$

$$\alpha_{out}^{(d_j)}(x_k) = \alpha_{in}^{(d_j)}(x_k) T_k^{(d_j)}(x_k) \quad (40)$$

$$\beta_{in}^{(d_j)}(x_k) = \sum_{\sim\{x_k\}} \beta_{out}^{(d_j)}(x_{k+d_j}) I_{k,k+d_j}(x_{k+d_j}, x_k) \quad (41)$$

$$\beta_{out}^{(d_j)}(x_k) = \beta_{in}^{(d_j)}(x_k) T_k^{(d_j)}(x_k)\,. \quad (42)$$

Concerning the scheduling, the set of BCJR algorithms that is activated first is the one involving the largest interference, i.e., $\overline{E}_M = \max(\overline{\boldsymbol{E}})$ (placed at distance $\overline{d}_M$ from the main diagonal), with input information

$$T_k^{(\overline{d}_M)}(x_k) = P_k(x_k) F_k(x_k)\,.$$

Then, we proceed following the decreasing energy order, i.e., the one dictated by the entries of $\overline{\boldsymbol{E}}$. This way, we are prioritizing the strongest diagonals of $\boldsymbol{G}^r$.

The proposed algorithm benefits from inner iterations. However, it achieves convergence very quickly: typically, just 3 iterations result to be sufficient. After the allowed iterations have been performed, the final estimate for the modulation symbols is computed as

$$P(x_k \,|\, \boldsymbol{y}) \propto P_k(x_k) F_k(x_k) \prod_{i=1}^{D} \alpha_{in}^{(d_i)}(x_k) \beta^{(d_i)}(x_k)\,. \quad (43)$$

Similarly to the graph of the MP$_{\boldsymbol{G}}$ [?], the FG on which the proposed algorithm operates has girth 6. However, although they are quite long, a known issue of the SPA, as applied to FGs with cycles, is the overestimation of the reliability of the propagated messages. A possible solution is the adoption in (25) and (37) of a value of σ^2 greater than the exact one. The effectiveness of this algorithm comes from its ability to leverage the matrix structure and operate along the diagonals, allowing for different optimizations of σ^2 based on the specific diagonal being considered. In particular, we found out that it is sufficient to optimize only two variances: one for the first BCJR set and another for the remaining $D-1$.

The overall computational load of the proposed approach is $\mathcal{O}(n_{iter} NML|\mathcal{A}|^2)$, where L is the number of interferers per symbol and depends linearly on the selected value D of diagonals.

It is possible to demonstrate that, being derived from the same channel matrix $\boldsymbol{G}$, the FG of the proposed algorithm can be traced back to that of the MP$_{\boldsymbol{G}}$. However, the primary distinction between the two algorithms

lies in their scheduling strategies. In fact, as will be shown in Section 5, adopting the proposed operating criterion allows to achieve a superior performance and faster convergence in all the analyzed scenarios.

5. Performance Analysis

We tested the performance of the analyzed detectors employing an OTFS-based system with carrier frequency 5 GHz, system bandwidth of 0.24 MHz, $M = 64$ and $N = 50$, resulting in a subcarrier spacing of 3750 Hz and a symbol duration of 267 μs. We assumed QPSK modulation for the information symbols and considered P satellites in visibility with the UT, belonging to the SpaceX Starlink constellation, whose orbital parameters are derived from [?]. The UT on ground has a fixed position. We assume that the P satellites perfectly compensate for delay and Doppler shifts at one point on the Earth, referred to as ideal UT position. At this location, the signal contributions from the transmitting satellites arrive simultaneously. Then, we consider an offset distance on the order of kilometers of the UT from the ideal position, such that the residual uncompensated delay and Doppler shifts satisfy $\max_p\{\tau_p\} < T$ and $\max_p\{\nu_p\} < \Delta f$. We assume that the considered UT is served by the P closest satellites, ordered by increasing slant range. In all the simulated scenarios, we assumed perfect CSI at the receiver. We evaluated the performance in terms of pragmatic capacity, measured in bit/s/Hz versus the signal-to-noise ratio (SNR) of the first path (i.e., the one characterized by the largest gain). The energy of all the considered paths is normalized to the shortest slant range attenuation. In Figure 8, we verified the performance gain related to the adoption of macro-diversity configurations. We considered different numbers of satellites in visibility with the UT ($P = 1, \ldots, 7$) and, at the receiver, we employed the MMSE detector. As demonstrated in Figure 8, when $P > 2$ a higher number of satellites serving the same UT simultaneously allows to significantly improve the system performance, thus confirming the advantages that diversity schemes applied to OTFS-modulated transmissions could bring. Increasing the number of contributions from 1 to 2 results in a performance gain only in the low-to-medium SNR range when the MMSE detector is employed, as shown in Figure 8. This performance degradation at high spectral efficiency is addressed adopting other kinds of detectors, as discussed below.

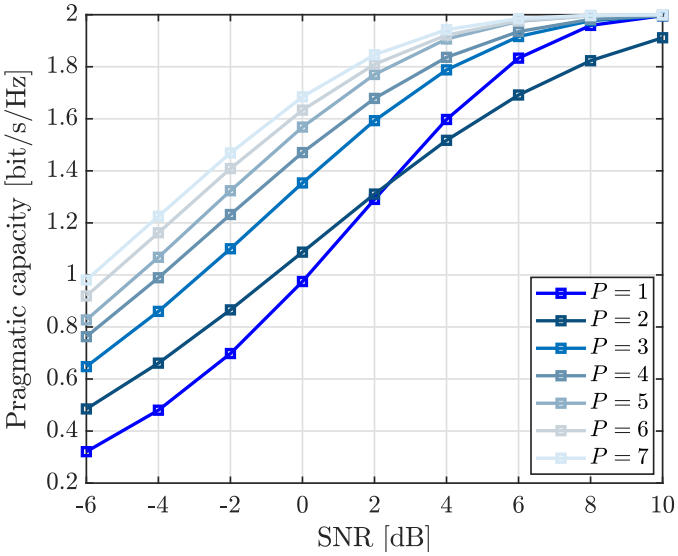


Figure 8. Pragmatic capacity exploiting multi-satellite diversity and using the MMSE detector.

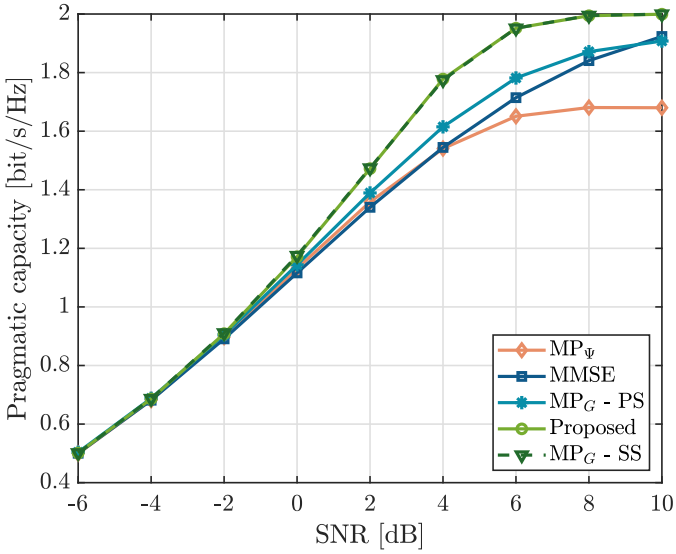


Figure 9. Performance comparison in terms of pragmatic capacity with $P = 2$ satellites in visibility with the UT.

The performance of the analyzed detectors is reported in Figures 9, 10. We tested the different receivers considering a fixed constraint on the computational load. The MP detectors perform a maximum of 3 inner iterations and the selection of the interfering terms for the detection process is based on their energy as discussed in Section 3. The MMSE detector is allowed to operate at full complexity as it is used only for reference but it cannot be practically implemented in most cases. Firstly, the scenario considered in Figure 9 foresees the presence of two transmitting satellites. In this context, the proposed detector and the MP$_{\boldsymbol{G}}$ [?] based on a serial-schedule (SS) achieve the same pragmatic capacity, outperforming the MP$_{\boldsymbol{\Psi}}$ [?] at fixed complexity and the MMSE. Moreover, we also tested the MP$_{\boldsymbol{G}}$ with a parallel-schedule (PS) and found out that this implementation requires a higher number of iterations for achieving convergence. In fact, even if it reaches the MMSE, it shows a significant performance loss with respect to the proposed algorithm, as can be observed in Figure 9.

In Figure 10, we analyzed the detectors performance when the same OTFS signal is transmitted simultaneously by $P = 5$ satellites. The MP$_{\boldsymbol{G}}$ has been tested with both schedules and it can be observed that, in this context, the implementation relying on the PS results to be much more effective than that based on the SS. Although this reversal in performance between SS and PS may at first surprise, the interplay between the FG complexity and the scheduling strategy is a known issue. Such that an ad-hoc scheduling is sometimes needed to implement an MP algorithm on the FG, modeling some specific

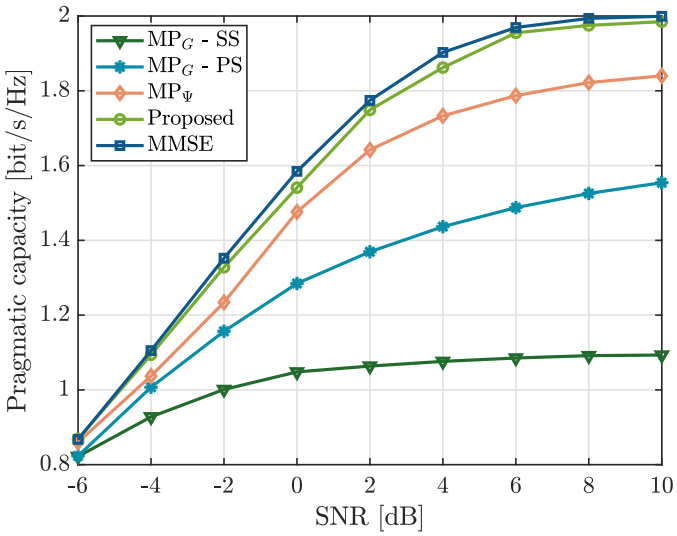


Figure 10. Performance comparison in terms of pragmatic capacity with $P = 5$ satellites serving simultaneously the UT.

problems that would otherwise become intractable [?], [?], [?]. However, in this case, both MP$_{\boldsymbol{G}}$ algorithms are outperformed by all the other algorithms, showing that a higher number of paths, and a resulting reduced channel sparsity, causes a performance degradation. The largest pragmatic capacity is achieved by the MMSE, which is practically reached by the proposed algorithm. On the contrary, the MP$_{\boldsymbol{\Psi}}$, which converges more slowly, requires a larger number of iterations to improve its performance. In the current analysis, which is carried out at equal complexity, the MP$_{\boldsymbol{\Psi}}$ algorithm saturates at approximately 1.8 bit/s/Hz and it is thus outperformed by the proposed algorithm.

6. Conclusions

In this paper, we proposed a low-complexity detection algorithm for OTFS modulation based on MP. We considered a LEO multi-satellite communication system, where the multipath channel is doubly-selective because of the high Doppler shifts and channel time-variance, typically encountered in such scenarios. The proposed algorithm leverages the sparsity and the particular structure of the channel matrix to enhance the detection efficiency. Moreover, it relies on an innovative schedule which, elaborating the strongest interferers first, allows to speed up the convergence process. Performance comparisons with state-of-the-art methods demonstrate that our approach offers an optimal trade-off between complexity and performance.

Acknowledgments

This work was supported by the European Union - Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, CUP D93C22000910001, partnership on "Telecommunications of the Future" (PE00000001 - program "RESTART") and from the PRIN 2022 project no. 2022BEXMXN_01 entitled "INSPIRE: Integrated Terrestrial/Space wireless networks for broadband connectivity and IoT services" (CUP: D53D23001150006) funded by the Italian Ministry of Universities and Research (MUR). This work has been also funded by the European Union Smart Networks and Services Joint Undertaking Project 5G-STARDUST under Grant Agreement 101096573. The views expressed are those of the authors and do not necessarily represent the project. The Commission is not liable for any use that may be made of any of the information contained therein.

Biography

Elisa Conti was born in Parma, Italy, in 1997. She received the B.Sc. degree in Computer, Electronic and Telecommunications Engineering and the M.Sc. degree in Communication Engineering (summa cum laude) in 2019 and 2021, respectively, from the University of Parma, Italy. She is currently pursuing the Ph.D. degree in Information Technology with the Department of Engineering and Architecture at the same university. Her main research interests focus on digital communications and signal processing, with particular emphasis on iterative joint detection and decoding techniques based on factor graphs.

Amina Piemontese received the Dr. Ing. degree in Telecommunications Engineering from the University of Parma, Italy, in 2006 and the Ph.D. degree in Information Technology from the University of Parma and from TELECOM Bretagne, Brest, France, in 2011. From 2011 to 2015 she held a postdoctorate position at the Department of Engineering and Architecture of the University of Parma, where she is currently an associate professor. From May 2015 to May 2020 she was with the Department of Electrical Engineering at Chalmers University of Technology, Gothenburg, Sweden. Her research activity includes various topics in digital communications, with particular emphasis on iterative joint detection and decoding algorithms, multiuser communications theory and information theory. She received the best paper award at the 5th Advanced Satellite Mobile Systems Conference and 11th International Workshop on Signal Processing for Space Communications (ASMS&SPSC 2010) and at the IEEE Wireless Communications and Networking Conference (WCNC), and the Marie Curie Individual Fellowship of the European Commission.

Tommaso Foggi received the master's degree in Telecommunication Engineering from the University of Parma in 2003 and the Ph.D. degree in Information Technology from the same University in 2008. From 2009 to 2018, he was a Research Engineer of National Inter-University Consortium for Telecommunications (CNIT) in the University of Parma research unit. He is now an Associate Professor at the Engineering and Architecture Department of the University of Parma. His main research interests include electronic signal processing for optical and satellite communication systems. He is author of tens of peer-reviewed papers and several patents, and he won the best paper award in the Optical Networks and Systems Symposium at the IEEE International Conference on Communications (ICC 2008), Beijing, China, May 2008. He was/is involved in many research projects funded by public authories like MIUR, ESA, EU or private companies like Ericsson, CGS, Inmarsat, Huawei.

Giulio Colavolpe received the Dr. Ing. degree in Telecommunications Engineering (cum laude) from the University of Pisa, Italy, in 1994 and the Ph.D. degree in Information Technologies from the University of Parma, Italy, in 1998. Since 1997, he has been at the University of Parma, Italy, where he is now Professor of Telecommunications at the Dipartimento di Ingegneria e Architettura (DIA). In 2000, he was Visiting Scientist at the Institut Eurécom, Valbonne, France. In 2013, he was a Visiting Scientist at the European Space Agency (ESTEC, Noordwijk, The Netherlands). His research interests include the design of digital communication systems, adaptive signal processing (with particular emphasis on iterative detection techniques for channels with memory), channel coding and information theory. His research activity has led to more than 200 papers in refereed journals and in leading international conferences, and 18 industrial patents. He received the best paper award at the 13th International Conference on Software, Telecommunications and Computer Networks (SoftCOM'05), Split, Croatia, September 2005, the best paper award for Optical Networks and Systems at the IEEE International Conference on Communications (ICC 2008), Beijing, China, May 2008, and the best paper award at the 5th Advanced Satellite Mobile Systems Conference and 11th International Workshop on Signal Processing for Space Communications (ASMS&SPSC 2010), Cagliari, Italy. He served as an Editor for IEEE Transactions on Wireless Communications, IEEE Transactions on Communications, and IEEE Wireless Communications Letters and as an Executive Editor for Transactions on Emerging Telecommunications Technologies (ETT).

Armando Vannucci received the Dr.Ing. degree in Electronic Engineering "cum laude" from the University of Rome "La Sapienza" in 1993, discussing a thesis about the digital analysis of speech signals applied to speech recognition. In the following, he was first with the INFO-COM Dept. at the University of Rome and then a Ph.D. Student at the University of Parma, joining an industrial research project on digital radio links, for which he holds two patents. He received a Ph.D. in Information Technology, discussing the thesis "Sequence estimation receivers for nonlinear transmission channels" at the Polytechnic of Turin. In 2002 he became an Assistant Professor at the University of Parma. Across time, his research interests included: variational techniques for the design of digital receivers; fiber optics transmission; Polarization Mode Dispersion; Optical Amplifiers; Digital Speech Processing. He took part in various research projects, both institutional and with industrial partners. He is the author of more than sixty scientific publications, half of which on international journals. Armando Vannucci has been a visiting scientist at the Alcatel Labs in Marcoussis, France, and at the Université Laval in Quèbec City, Canada and a visiting lecturer at the Hochschule Karlsruhe, Germany. He is the author of two textbooks and has taught several courses (undergraduate, graduate, master, Ph.D., summer schools) at the Universities of Rome, Urbino, and Parma, where he is currently in charge of "Signals and Systems".