Tensor-Based Receiver for Joint Channel, Data, and Phase-Noise Estimation in MIMO-OFDM Systems

Bruno Sokal, *Student Member, IEEE*, Paulo R. B. Gomes, André L. F. de Almeida, *Senior Member, IEEE*, and Martin Haardt, *Fellow, IEEE*

***Abstract*—Phase-noise is a system impairment caused by the mismatch between the oscillators at the transmitter and the receiver. In OFDM systems, this induces inter-carrier-interference (ICI) by rotating the transmitted symbols. Thus it can cause severe system performance degradation. To reduce its effects, the phase-noise must be estimated or compensated. In this work, we propose a two-stage tensor-based receiver for a joint channel, phase-noise (PN), and data estimation in MIMO-OFDM systems. In the first stage, we show that the received signal at the pilot subcarriers can be modeled as a third-order PARAFAC tensor. Based on this model, we propose two algorithms for channel and phase-noise estimation at the pilot subcarriers. The first algorithm, based on the BALS (Bilinear Alternating Least Squares), is an iterative algorithm that estimates the channel gains and the phase-noise impairments. The second is a closed-form algorithm based on the LS-KRF (Least Squares - Khatri-Rao Factorization) that estimates the channel gains and the phase-noise terms through multiple rank-one factorizations. Both algorithms achieve similar performance, but in terms of computational complexity, we show that the LS-KRF becomes more attractive than the BALS as the number of receive antennas is increased. The second stage consists of data estimation, for which we propose a ZF (Zero-Forcing) receiver that capitalizes on the PARATuck tensor structure of the received signal at the data subcarriers using the Selective Kronecker Product (SKP) operator. Our numerical simulations show that the proposed receiver achieves an improved performance compared to the state-of-art receivers in terms of symbol error rate (SER) and normalized mean square error (NMSE) of the estimated channel and phase-noise matrices.**

***Index Terms*—Joint channel and data estimation, MIMO-OFDM systems, phase-noise estimation, selective Kronecker product (SKP), tensor-based modeling.**

Manuscript received August 1, 2020; revised December 11, 2020 and January 23, 2021; accepted February 11, 2021. Date of publication February 24, 2021; date of current version March 29, 2021. This work was supported in part by the Coordenação de Aperfeicoamento de Pessoal de Nivel Superior -Brasil (CAPES) - Finance Code 001, and CNPq (Proc. 306616/2016-5), in part by CAPES/PROBRAL Proc. 88887.144009/2017-00, CAPES/PRINT Proc. 88887.311965/2018-00, and in part by the German Research Foundation (DFG) under Grant HA 2239/14-1 (AdAMMM). The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Sergiy Vorobyov. *(Corresponding author: Bruno Sokal.)*

Bruno Sokal and André L. F. de Almeida are with the Wireless Telecom Research Group (GTEL), Department of Teleinformatics Engineering, Federal University of Ceará, Fortaleza, CE 60455970, Brazil (e-mail: brunosokal@gtel.ufc.br; andre@gtel.ufc.br).

Paulo R. B. Gomes is with the Federal Institute of Education, Science and Technology of Ceará, Tauá, CE 63660-000, Brazil (e-mail: paulo.gomes@ifce.edu.br).

Martin Haardt is with the Comunication Research Laboratory (CRL), Ilmenau University of Technology, D-98684 Ilmenau, Germany (e-mail: martin.haardt@tu-ilmenau.de).

Digital Object Identifier 10.1109/JSTSP.2021.3061917

I. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) has been adopted in 3G/4G LTE services due to its well known robustness to multi-path propagation, frequency-selective fading, as well as low complexity implementation and equalization [1], [2]. However, these advantages rely on a perfect system synchronization to keep intact the orthogonality among the subcarriers. In practical scenarios, synchronization impairments lead to inter-carrier-interference (ICI) that degrades the ultimate system performance. The two most common impairments that destroy subcarrier orthogonality is the carrier-frequency-offset (CFO), which is caused by the Doppler frequency shift effect induced by the mobility of the user, as well as the phase-noise (PN) originated by hardware imperfections at the transmitter and receiver oscillators [1].

The PN compensation problem has been extensively studied in the past years [3]–[12]. For instance, in [3] and [4], techniques for PN compensation are proposed for single-input single-output (SISO) systems. However, they assume a perfect knowledge of the channel state information (CSI) at the receiver, which may not be feasible in practice. In MIMO systems, the PN compensation becomes more challenging due to the presence of multiple independent phase-noise processes, i.e., one for each transmit and receive antenna. The authors of [5] propose a novel placement of pilot subcarriers in the preamble and data portions of the MIMO-OFDM frame for joint channel and PN estimation. The authors in [6], [7] propose compensation schemes based on the knowledge of the statistical model for the PN process.

In [8], a detailed study on the variance of the ICI as a function of the phase-noise is provided, and an algorithm to compensate the so-called common phase error (CPE) and the ICI is formulated. The work [9] proposes a phase-noise compensation method based on channel estimation via a linear time domain interpolation. In [10], the authors propose Least Squares (LS) and weighted LS (WLS) methods for data, channel, and phase-noise tracking over a OFDM frame.

More recently, [11] presents a pilot signal design scheme for PN mitigation in millimeter wave (mmWave) MIMO-OFDM systems, where the combined channel and phase-noise term is estimated and compensated for data detection. The work [12] proposes a PN compensation technique for high-frequency MIMO-OFDM systems using an LS method.

The use of tensor decompositions for modeling MIMO systems has been growing [13]–[24], and more recently, has resulted in proposals of tensor-based receivers for MIMO-OFDM systems [25]–[29], However, these works do not take into account system impairments such as the phase-noise that leads to ICI. As discussed in these many references, the main reasons for using tensor-based modeling and signal processing are their ability to exploit the inherent multidimensional structure of the transmitted and/or received signals (e.g., time, frequency, space, polarization, etc.), as well as their built-in identifiability properties that allow us to derive blind or semi-blind receiver algorithms and operate under more flexible conditions than matrix-based processing receivers. These key features are a natural consequence of the uniqueness properties of tensor decompositions compared to matrix decompositions [30].

In this paper, we propose a new tensor-based method for frequency-selective MIMO-OFDM channel, phase-noise, and data estimation. By assuming that the PN is approximately constant over a very short symbol length, and motivated by the multidimensional structure of the received signal, a third-order PARAFAC (Parallel Factors) tensor model is developed for the received signal at the pilot subcarriers [31]. This model is exploited to separate the PN contribution from the channel. Then, we propose a two-stage receiver for joint channel, PN, and data estimation. In the first stage, we propose two algorithms to estimate the channel at the pilot subcarriers jointly with the PN impairments. The first one is an iterative solution based on the bilinear ALS (BALS), which consists of estimating the channel and phase-noise by solving two LS problems in an alternating way. The second one is a closed-form algorithm based on the LS-KRF solution that solves multiple rank-one factorizations to jointly estimate the channel gains and phase-noise terms. The channel at the data subcarrieris is then obtained via interpolation. In the second stage, a ZF equalizer that exploits a PARATuck tensor structure of the signal is used to estimate the transmitted data. The identifiability conditions and computational complexity for both processing stages are also discussed. Our simulation results show the effectiveness and high accuracy of the proposed receiver for the joint estimation of the channel, data, and PN impairments. To the best of our knowledge, this is the first work that links tensor decomposition to the joint channel, PN, and data estimation problem in MIMO-OFDM systems. We can summarize our main contributions as follows:

1) We derive a new tensor decomposition based formulation to solve the problem of joint channel, phase-noise, and data estimation in MIMO-OFDM systems, by assuming different phase-noise perturbations for every pair of transmit and receive antenna. The proposed approach allows to separate the phase-noise from the channel gains, even without any CSI knowledge;
2) We propose a two-stage receiver to jointly estimate the channels, phase-noise terms, and data symbols. The first stage estimates the channel and the phase-noise (via BALS or LS-KRF), while the second stage extracts the symbol estimates from a selective Kronecker product (SKP) formulation by means of a zero forcing filtering.

The rest of the paper is organized as follows. Section II introduces the main tensor operations used in this paper. The system model is described in Section III. The tensor-based formulation of the received signal is developed in Section IV. The iterative pilot-assisted channel and PN estimation algorithm (BALS) is presented in Section V-A, while the closed-form algorithm (LS-KRF) is introduced in Section V-B. The data estimation procedure exploiting the SKP operator is presented in Section V-D. Simulations and discussions are provided in Section VI, and the paper is concluded in Section VII.

A. Notation and Tensor Pre-Requisites

Scalars are represented as non-bold lower-case letters a, column vectors as lower-case boldface letters $\boldsymbol{a}$, matrices as upper-case boldface letters $\boldsymbol{A}$, and tensors as calligraphic upper-case letters $\boldsymbol{\mathcal{A}}$. The superscripts $\{\cdot\}^{\mathrm{T}}$, $\{\cdot\}^{*}$, $\{\cdot\}^{\mathrm{H}}$ and $\{\cdot\}^{+}$ stand for transpose, conjugate, conjugate transpose and pseudo-inverse operations, respectively. The operator $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm of a matrix or tensor, $\mathbb{E}\{\cdot\}$ is the expectation operator. The operator $\mathrm{diag}(\boldsymbol{a})$ converts $\boldsymbol{a}$ into a diagonal matrix, while $D_i(\boldsymbol{A})$ forms a diagonal matrix from the i-th row of $\boldsymbol{A}$. Moreover $\mathrm{vec}(\boldsymbol{A})$ converts $\boldsymbol{A} \in \mathbb{C}^{I_1 \times R}$ to a column vector $\boldsymbol{a} \in \mathbb{C}^{I_1 R \times 1}$ by stacking its columns on top of each other, while the $\mathrm{unvec}(\cdot)$ operator is the inverse of the vec operation. The symbol $\circ$ denotes the outer product operator. $\boldsymbol{A}_{.r} \in \mathbb{C}^{I \times 1}$ represents the r-th column of $\boldsymbol{A} \in \mathbb{C}^{I \times R}$. The Khatri-Rao product, also known as the column wise Kronecker product, between two matrices $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R] \in \mathbb{C}^{I_1 \times R}$ and $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_R] \in \mathbb{C}^{I_2 \times R}$, symbolized by $\diamond$, is defined as $\boldsymbol{A} \diamond \boldsymbol{B} = [\boldsymbol{a}_1 \otimes \boldsymbol{b}_1, \ldots, \boldsymbol{a}_R \otimes \boldsymbol{b}_R] \in \mathbb{C}^{I_2 I_1 \times R}$.

II. Tensor Background

In this section, some tensor preliminaries are provided, by focusing on the main operations and properties that will be useful in the rest of the paper. For more general formulations, the reader may refer to [30] and [32].

Consider a set of matrices $\{\boldsymbol{X}_{i_3}\} \in \mathbb{C}^{I_1 \times I_2}$, for $i_3 = 1, \ldots, I_3$. We can concatenate all I_3 matrices to form the third-order tensor $\boldsymbol{\mathcal{X}} = [\boldsymbol{X}_1 \sqcup_3 \boldsymbol{X}_2 \sqcup_3 \ldots \sqcup_3 \boldsymbol{X}_{I_3}] \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, where $\sqcup_3$ indicates a concatenation in the third dimension. We can interpret $\boldsymbol{X}_{i_3}$ as the i_3-th frontal slice of $\boldsymbol{\mathcal{X}}$, defined as $\boldsymbol{\mathcal{X}}_{..i_3} = \boldsymbol{X}_{i_3}$ where the ".." indicates that the dimensions I_1 and I_2 are fixed. The tensor $\boldsymbol{\mathcal{X}}$ can be "matricized" by letting one dimension vary along the rows and the remaining two dimensions along the columns. From $\boldsymbol{\mathcal{X}}$, we can form three different matrices, referred to as the n-mode unfoldings (for $n = \{1, 2, 3\}$ in this case), $[\boldsymbol{\mathcal{X}}]_{(1)} = [\boldsymbol{\mathcal{X}}_{..1}, \ldots, \boldsymbol{\mathcal{X}}_{..I_3}] \in \mathbb{C}^{I_1 \times I_2 I_3}$, $[\boldsymbol{\mathcal{X}}]_{(2)} = [\boldsymbol{\mathcal{X}}^{\mathrm{T}}_{..1}, \ldots, \boldsymbol{\mathcal{X}}^{\mathrm{T}}_{..I_3}] \in \mathbb{C}^{I_2 \times I_1 I_3}$ and $[\boldsymbol{\mathcal{X}}]_{(3)} = [\mathrm{vec}(\boldsymbol{\mathcal{X}}_{..1}), \ldots, \mathrm{vec}(\boldsymbol{\mathcal{X}}_{..I_3})]^{\mathrm{T}} \in \mathbb{C}^{I_3 \times I_1 I_2}$. Consider the tensor $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{X}} \times_1 \boldsymbol{A} \in \mathbb{C}^{R \times I_2 \times I_3}$, where $\boldsymbol{A} \in \mathbb{C}^{R \times I_1}$, the operator "$\times_1$" defines the 1-mode product which represents a linear combination of the columns of $\boldsymbol{A}$ with the first dimension of $\boldsymbol{\mathcal{X}}$. We can define this 1-mode product in terms of the 1-mode unfolding of $\boldsymbol{\mathcal{Y}}$ as $[\boldsymbol{\mathcal{Y}}]_{(1)} = \boldsymbol{A}[\boldsymbol{\mathcal{X}}]_{(1)} \in \mathbb{C}^{R \times I_2 I_3}$.

Throughout this paper, we shall make use of the following properties:

$$\mathrm{vec}\left(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}\right) = \left(\boldsymbol{C}^{\mathrm{T}} \otimes \boldsymbol{A}\right) \mathrm{vec}\left(\boldsymbol{B}\right), \tag{1}$$

$$\mathrm{vec}\left(\boldsymbol{A}\mathrm{diag}\left(\boldsymbol{b}\right)\boldsymbol{C}\right) = \left(\boldsymbol{C}^{\mathrm{T}} \diamond \boldsymbol{A}\right) \boldsymbol{b}, \tag{2}$$

$$\boldsymbol{a}^{\mathrm{T}} \diamond \boldsymbol{B} = \boldsymbol{B}\mathrm{diag}(\boldsymbol{a}), \tag{3}$$

$$\boldsymbol{a} \otimes \boldsymbol{b} = \mathrm{vec}\left(\boldsymbol{b} \circ \boldsymbol{a}\right), \tag{4}$$

where the vectors and matrices involved have compatible dimensions in each case.

A. Selective Kronecker Product

Different from the standard Krocneker product, the Selective Kronecker Product (SKP) "selects" the modes in which the "spreading" of products will occur [33]. It can also be viewed as a tensor rearrangement of the elements of the standard Kronecker product. For example, the Kronecker product between two matrices $\boldsymbol{X} \in \mathbb{C}^{M\times N}$ and $\boldsymbol{Y} \in \mathbb{C}^{R\times S}$ is given by $\boldsymbol{Z} = \boldsymbol{X} \otimes \boldsymbol{Y} \in \mathbb{C}^{RM\times SN}$, but we may want to spread only the second dimension of $\boldsymbol{X}$ with the first dimension of $\boldsymbol{Y}$, which yields a tensor $\boldsymbol{\mathcal{Z}} = \boldsymbol{X} \otimes_2^1 \boldsymbol{Y} \in \mathbb{C}^{M\times RN\times S}$. Note that the lower index in the operator "$\otimes_2^1$" indicates which dimension of the matrix on the left ($\boldsymbol{X}$) will spread, while the upper index indicates the spreading dimension of the matrix on the right ($\boldsymbol{Y}$). In the matrix case, the relationship between the standard Kronecker operator and the SKP operator is $\otimes = \otimes_{1,2}^{1,2}$.

Let us consider two third-order tensors $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1\times I_2\times I_3}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{J_1\times J_2\times J_3}$. The standard Kronecker product between these tensors results in a third-order tensor $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} \otimes \boldsymbol{\mathcal{B}} \in \mathbb{C}^{J_1I_1\times J_2I_2\times J_3I_3}$ which, in the matrix slice notation, is given by

$$\boldsymbol{\mathcal{C}}_{..p} = \boldsymbol{\mathcal{A}}_{..i_3} \otimes \boldsymbol{\mathcal{B}}_{..j_3} \in \mathbb{C}^{J_1I_1\times J_2I_2}, \tag{5}$$

where $i_3 = \{1,\ldots,I_3\}$, $j_3 = \{1,\ldots,J_3\}$ and $p = j_3 + (i_3 - 1)J_3$ for $p = \{1,\ldots,J_3,\ldots,J_3I_3\}$. However, note that, as in the matrix case presented above, we may want to select the spreading dimensions of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$. For example, taking $\boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{A}} \otimes_{1,2}^{1,2} \boldsymbol{\mathcal{B}} \in \mathbb{C}^{J_1I_1\times J_2I_2\times J_3\times I_3}$ results in a fourth-order tensor in which the first and second modes of the tensors $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ are merged, while their third modes are kept as separate modes of the resulting tensor. This operation can be written in the matrix slice notation as

$$\boldsymbol{\mathcal{D}}_{..i_3j_3} = \boldsymbol{\mathcal{A}}_{..i_3} \otimes \boldsymbol{\mathcal{B}}_{..j_3} \in \mathbb{C}^{J_1I_1\times J_2I_2}. \tag{6}$$

Although the right-hand side of equations (5) and (6) are equal, in the latter case the third dimension of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ are not merged, yielding a fourth-order tensor $\boldsymbol{\mathcal{D}}$ as the result. Otherwise stated, the SKP operator has a representation flexibility compared to the standard Kronecker product by defining which modes are combined. Note that, for the third-order case, the equivalence between the standard Kronecker operator and the SKP operator is given by $\otimes = \otimes_{1,2,3}^{1,2,3}$. The SKP operator will play a key role in Section V-D, where the proposed ZF receiver will be developed.

B. Tensor Models

The two most common tensor decompositions are the Tucker [34] and the parallel factor (PARAFAC) [31] decompositions. Some other tensor decompositions can be seen as different combinations of these two, such as the PARATuck 2 [35], the nested Tucker [23], nested PARAFAC [22], and others [13], [36]. In this work, we focus on the Tucker, PARAFAC, and PARATuck models that are useful in our context.

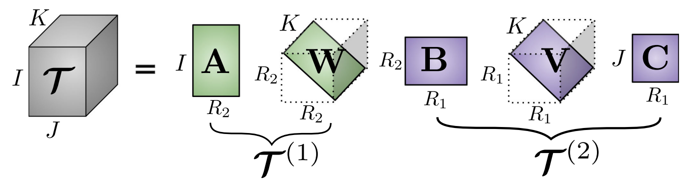


Fig. 1. 3D illustration of a PARATuck tensor.

1) Tucker Decomposition: A third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1\times I_2\times I_3}$ admitting a Tucker decomposition can be written as

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \boldsymbol{A}^{(1)} \times_2 \boldsymbol{A}^{(2)} \times_3 \boldsymbol{A}^{(3)}, \tag{7}$$

where $\boldsymbol{A}^{(n)} \in \mathbb{C}^{I_n\times R_n}$, for $n = \{1,2,3\}$, are the factor matrices and $\boldsymbol{\mathcal{G}} \in \mathbb{C}^{R_1\times R_2\times R_3}$ is the core tensor. We can express the three matrix unfoldings of $\boldsymbol{\mathcal{X}}$ as $[\boldsymbol{\mathcal{X}}]_{(1)} = \boldsymbol{A}^{(1)}[\boldsymbol{\mathcal{G}}]_{(1)}(\boldsymbol{A}^{(3)} \otimes \boldsymbol{A}^{(2)})^{\mathrm{T}} \in \mathbb{C}^{I_1\times I_2I_3}$, $[\boldsymbol{\mathcal{X}}]_{(2)} = \boldsymbol{A}^{(2)}[\boldsymbol{\mathcal{G}}]_{(2)}(\boldsymbol{A}^{(3)} \otimes \boldsymbol{A}^{(1)})^{\mathrm{T}} \in \mathbb{C}^{I_2\times I_1I_3}$ and $[\boldsymbol{\mathcal{X}}]_{(3)} = \boldsymbol{A}^{(3)}[\boldsymbol{\mathcal{G}}]_{(3)}(\boldsymbol{A}^{(2)} \otimes \boldsymbol{A}^{(1)})^{\mathrm{T}} \in \mathbb{C}^{I_3\times I_1I_2}$.

2) PARAFAC Decomposition: The PARAFAC decomposition expresses a tensor as a sum of rank-one components. It is a rank-revealing decomposition that can also be viewed as a special case of the Tucker decomposition. A tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{J_1\times J_2\times J_3}$ admits a PARAFAC decomposition if

$$\begin{aligned}\boldsymbol{\mathcal{Y}} &= \sum_{r=1}^{R} \boldsymbol{B}_{.r}^{(1)} \circ \boldsymbol{B}_{.r}^{(2)} \circ \boldsymbol{B}_{.r}^{(3)} \\ &= \boldsymbol{\mathcal{I}}_{3,R} \times_1 \boldsymbol{B}^{(1)} \times_2 \boldsymbol{B}^{(2)} \times_3 \boldsymbol{B}^{(3)},\end{aligned} \tag{8}$$

where $\boldsymbol{B}^{(n)} \in \mathbb{C}^{J_n\times R}$, $n = \{1,2,3\}$, are the factor matrices of the decomposition, and $\boldsymbol{\mathcal{I}}_{3,R} \in \mathbb{R}^{R\times R\times R}$ is the identity tensor which, analogous to the identity matrix, has elements equal to one when the indices are the same, and zeros otherwise. Here, R is the tensor rank which is the minimum number of rank-one tensors required to perfectly represent $\boldsymbol{\mathcal{Y}}$ in a linear combination.

A sufficient condition for the uniqueness of the PARAFAC decomposition of a third-order tensor is given by the well-known Kruskal's uniqueness theorem [37], [38],

$$k_{\boldsymbol{B}^{(1)}} + k_{\boldsymbol{B}^{(2)}} + k_{\boldsymbol{B}^{(3)}} \geq 2R + 2, \tag{9}$$

where $k_{\boldsymbol{B}^{(n)}}$ is the Kruskal rank of $\boldsymbol{B}^{(n)}$ in [37]. From equation (8), we can write the three unfoldings of $\boldsymbol{\mathcal{Y}}$ as $[\boldsymbol{\mathcal{Y}}]_{(1)} = \boldsymbol{B}^{(1)}(\boldsymbol{B}^{(3)} \diamond \boldsymbol{B}^{(2)})^{\mathrm{T}} \in \mathbb{C}^{J_1\times J_2J_3}$, $[\boldsymbol{\mathcal{Y}}]_{(2)} = \boldsymbol{B}^{(2)}(\boldsymbol{B}^{(3)} \diamond \boldsymbol{B}^{(1)})^{\mathrm{T}} \in \mathbb{C}^{J_2\times J_1J_3}$ and $[\boldsymbol{\mathcal{Y}}]_{(3)} = \boldsymbol{B}^{(3)}(\boldsymbol{B}^{(2)} \diamond \boldsymbol{B}^{(1)})^{\mathrm{T}} \in \mathbb{C}^{J_3\times J_1J_2}$.

3) PARATuck Decomposition: The PARATuck decomposition [35], also known as the tensor slice-wise multiplication [39] combines the properties of the PARAFAC and Tucker decompositions. The PARATuck decomposition of a third-order tensor $\boldsymbol{\mathcal{T}} \in \mathbb{C}^{I\times J\times K}$ can be written in slice-wise notation as

$$\begin{aligned}\boldsymbol{\mathcal{T}}_{..k} &= \boldsymbol{\mathcal{T}}_{..k}^{(1)} \cdot \boldsymbol{\mathcal{T}}_{..k}^{(2)} \in \mathbb{C}^{I\times J} \\ &= \boldsymbol{A}D_k\left(\boldsymbol{W}\right)\boldsymbol{B}D_k\left(\boldsymbol{V}\right)\boldsymbol{C}^{\mathrm{T}},\end{aligned}$$

where $\boldsymbol{A}^{I\times R_2}$, $\boldsymbol{B} \in \mathbb{C}^{R_1\times R_1}$, $\boldsymbol{C} \in \mathbb{C}^{J\times R_2}$, $\boldsymbol{W} \in \mathbb{C}^{K\times R_2}$ and $\boldsymbol{V} \in \mathbb{C}^{K\times R_1}$ are the corresponding factor matrices. This decomposition is illustrated in Figure 1.

Let us define $\boldsymbol{t}_k = \text{vec}(\boldsymbol{\mathcal{T}}_{..k}) \in \mathbb{C}^{IJ\times 1}$. Applying property (1) multiple times, as well as property (2) and (3), yields

$$\begin{aligned}\boldsymbol{t}_k &= (\boldsymbol{C} \otimes \boldsymbol{A}) \,\text{vec}\left(D_k(\boldsymbol{W})\, \boldsymbol{B} D_k(\boldsymbol{V})\right) \\ &= (\boldsymbol{C} \otimes \boldsymbol{A}) \left(D_k(\boldsymbol{V}) \otimes D_k(\boldsymbol{W})\right) \text{vec}(\boldsymbol{B}) \\ &= \left(\text{vec}(\boldsymbol{B})^{\text{T}} \diamond (\boldsymbol{C} \otimes \boldsymbol{A})\right) (\boldsymbol{V}_{k.} \otimes \boldsymbol{W}_{k.})^{\text{T}} \\ &= (\boldsymbol{C} \otimes \boldsymbol{A})\, \text{diag}(\boldsymbol{b})\, (\boldsymbol{V}_{k.} \otimes \boldsymbol{W}_{k.})^{\text{T}},\end{aligned}$$

where $\boldsymbol{b} = \text{vec}(\boldsymbol{B}) \in \mathbb{C}^{R_2R_1\times 1}$. Collecting the K vectors $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_K$ yields the 3-mode unfolding $[\boldsymbol{\mathcal{T}}]_{(3)} = [\boldsymbol{t}_1, \ldots \boldsymbol{t}_K]^{\text{T}} \in \mathbb{C}^{K\times IJ}$, which admits the following factorization

$$[\boldsymbol{\mathcal{T}}]^{\text{T}}_{(3)} = (\boldsymbol{C} \otimes \boldsymbol{A})\, \text{diag}(\boldsymbol{b}) \left(\boldsymbol{V}^{\text{T}} \diamond \boldsymbol{W}^{\text{T}}\right) \in \mathbb{C}^{IJ\times K}. \quad (10)$$

We are interested in writing $\boldsymbol{\mathcal{T}}$ using the n-mode product notation, which will be useful in Sections IV and V. To this end, let us reshape the diagonal matrix $\text{diag}(\boldsymbol{b}) \in \mathbb{C}^{R_2R_1\times R_2R_1}$ as the tensor $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{R_2\times R_1\times R_2R_1}$. Using the SKP operator introduced in Section II-A, we can write $\boldsymbol{\mathcal{B}}$ as

$$\boldsymbol{\mathcal{B}} = \underbrace{\left(\boldsymbol{\mathcal{I}}_{3,R_1} \otimes^{2,3}_{2,3} \boldsymbol{\mathcal{I}}_{3,R_2}\right)}_{\in \mathbb{R}^{R_2\times R_1\times R_2R_1\times R_2R_1}} \times_4 \boldsymbol{b}^{\text{T}} \in \mathbb{C}^{R_2\times R_1\times R_2R_1}, \quad (11)$$

where $\boldsymbol{\mathcal{I}}_{3,R_1} \in \mathbb{R}^{R_1\times R_1\times R_1}$ and $\boldsymbol{\mathcal{I}}_{3,R_2} \in \mathbb{R}^{R_2\times R_2\times R_2}$ are identity tensors. Using this definition, we can rewrite the PARATuck decomposition of $\boldsymbol{\mathcal{T}}$ as

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{B}} \times_1 \boldsymbol{A} \times_2 \boldsymbol{C} \times_3 \left(\boldsymbol{V}^{\text{T}} \diamond \boldsymbol{W}^{\text{T}}\right)^{\text{T}} \in \mathbb{C}^{I\times J\times K}. \quad (12)$$

Comparing (12) with equation (7), it is clear that the PARATuck decomposition can be interpreted as a special Tucker decomposition that has a constrained (sparse) core tensor $\boldsymbol{\mathcal{B}}$ defined in (11). Note also that $[\boldsymbol{\mathcal{B}}]_{(3)} = \text{diag}(\boldsymbol{b})$.

III. System Model

Let us consider a MIMO-OFDM system with M_T transmit antennas, M_R receive antennas, and F subcarriers. The total duration of an OFDM frame corresponds to K blocks of size L_P symbols each. We assume that each transmit and receive antenna is equipped with its own independent oscillator, so that the PN is assumed to be different between the antennas and is constant within a block k, with $k = \{1, \ldots, K\}$. Note that the assumption that the transmit/receive antennas are subject to different PN perturbations has been addressed in the literature, such as in [2], [10]–[12]. As mentioned in [2], [40], this assumption copes with realistic mmWave massive MIMO setups, including distributed MIMO scenarios.

Figure 2 illustrates the a MIMO system with antenna-dependent phase-noise impairments. Let $h_{m_r,b}[n]$ and $s_{m_t,i}[n]$ be the discrete-time channel and the transmitted symbol, for $m_r = \{1, \ldots, M_R\}$, $m_t = \{1, \ldots, M_T\}$, $i = \{1, \ldots, L_P\}$. Then, the sampled discrete-time received signal at the period n, for $n = \{0, \ldots, F-1\}$, can be represented as

$$\begin{aligned}&y_{m_r,i,k}[n] \\ &= e^{j\phi^{[\text{r}]}_{k,m_r}[n]} h_{m_r,m_t}[n] \circledast e^{j\phi^{[\text{t}]}_{k,m_t}[n]} s_{i,m_t}[n] + v_{m_r,i,k}[n],\end{aligned} \quad (13)$$

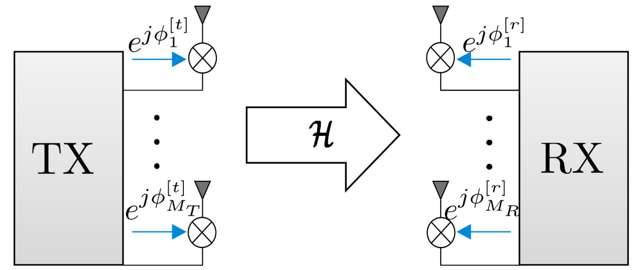


Fig. 2. MIMO system with antenna-dependent phase-noise impairments.

where "$\circledast$" stands for time-domain convolution, $\phi^{(\text{r})}_{k,m_r}[n]$ and $\phi^{(\text{t})}_{k,m_t}[n]$ are the PN impairment. Both are modeled as a Wiener process, i.e., $\phi^{[\text{x}]}_{k,m_{\text{x}}}[n+1] = \phi^{[\text{x}]}_{k,m_{\text{x}}}[n] + \epsilon^{[\text{x}]}_{k,m_{\text{x}}}$, with $\epsilon^{[\text{x}]}_{k,m_{\text{x}}} \sim \mathcal{N}(0, \sigma^2_{\phi^{[\text{x}]}})$, for x $\in \{\text{r}, \text{t}\}$. [1] Moreover, $v_{m_r,i,k}[n]$ is the Additive White Gaussian Noise (AWGN) at the receiver $\sim \mathcal{CN}(0, \sigma^2_v)$. Applying the discrete Fourier transform (DFT), the received signal at the f-th subcarrier is given by

$$Y_{m_r,i,k}[f] = \sum_{n=0}^{F-1} y_{m_r,i,k}[n] e^{-2j\pi f n/F}$$

$$\begin{aligned}Y_{m_r,i,k}[f] &= \Phi^{[\text{r}]}_{k,m_r}[0] H_{m_r,m_t}[f] \Phi^{[\text{t}]}_{k,m_t}[0] S_{m_t,i}[f] \\ &\quad + V_{m_r,i,k}[f]\end{aligned} \quad (14)$$

$$+ \underbrace{\sum_{p=0,p\neq f}^{F-1} \Phi^{[\text{r}]}_{k,m_r}[f-p] H_{m_r,m_t}[p] \sum_{q=0,q\neq f}^{F-1} \Phi^{[t]}_{k,m_t}[p-q] S_{m_t,i}[q]}_{\text{ICI components}}$$

$$\begin{aligned}Y_{m_r,i,k}[f] &= \Phi^{[\text{r}]}_{k,m_r}[0] H_{m_r,m_t}[f] \Phi^{[\text{t}]}_{k,m_t}[0] S_{i,m_t}[f] \\ &\quad + G_{m_r,i,k}[f] + V_{m_r,i,k}[f],\end{aligned} \quad (15)$$

where $f = \{0, \ldots, F-1\}$ is the subcarrier index in the frequency domain. The matrices $\Phi^{[\text{r}]}_{k,m_r}[0] = \frac{1}{F}\sum_{n=0}^{F-1} e^{j\phi^{[\text{r}]}_{k,m_r}[n]}_{k,m_r}$ and $\Phi^{[\text{t}]}_{k,m_t}[0] = \frac{1}{F}\sum_{n=0}^{F-1} e^{j\phi^{[\text{t}]}_{k,m_t}[n]}_{k,m_t}$ are the common phase error (CPE) generated by the receive and transmit antennas, respectively, and $G_{m_r,i,k}[f]$ is the inter-carrier-interference (ICI) term.[2] For convenience, the detailed steps from equation (13) to (15) are given in the Appendix A.

In this paper, we model the received signal $Y_{m_r,i,k}[f]$ as a fourth-order tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{M_R\times L_P\times K\times F}$ and make use of the proposed tensor notation to develop our receivers. Let us introduce a combiner matrix $\boldsymbol{W} \in \mathbb{C}^{M_R\times M_R}$ that is assumed to be fixed over the K blocks and all F subcarriers. Resorting to the

[1]The PN variance, $\sigma^2_{\phi^{[\text{x}]}}$ is usually defined as $\sigma^2_{\phi^{[\text{x}]}} = 4\pi\beta T_{sp}$, with β being the single-sided of the 3 dB bandwidth of the Lorentzian spectrum for $e^{j\phi^{[\text{x}]}[n]}$ and T_{sp} is the sampling period [12], [41]. However, in this paper we will assume a very high phase-noise variance $\sigma^2_{\phi^{[\text{x}]}} = 5\cdot 10^{-3}\text{rad}^2$.

[2]Note that the ICI structure in eq. (14) consists of a summation of phase-noise, channel frequency responses, and OFDM symbols, and can be modeled as an AWGN process $\sim \mathcal{CN}(0, \sigma^2_{\text{ICI}})$ [2], reducing the receiver complexity [11], [12]. In Appendix XI, we derive an expression to calculate σ^2_{ICI}, which mainly depends on the phase-noise variance $\sigma^2_{\phi^{[\text{x}]}}$ and the total number of subcarriers F.

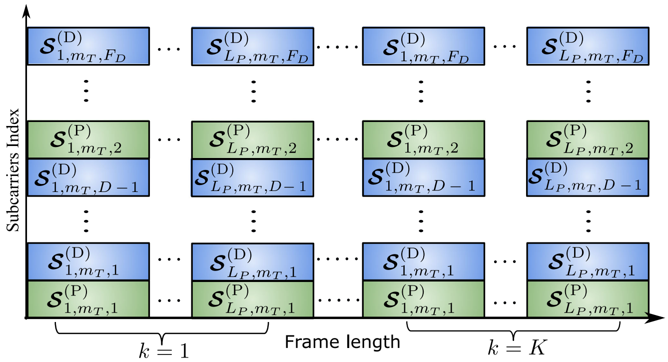


Fig. 3. Illustration of frame and block structures. The total frame length is $K \cdot L_P$.

slice-wise notation, the received signal tensor $\boldsymbol{\mathcal{Y}}_{..kf} \in \mathbb{C}^{M_R \times L_P}$ at the k-th block and f-th subcarrier can be written as

$$\boldsymbol{\mathcal{Y}}_{..kf} = \boldsymbol{W} D_k \left(\boldsymbol{\Phi}^{[\mathrm{r}]}[0]\right) \boldsymbol{\mathcal{H}}_{..f} D_k \left(\boldsymbol{\Phi}^{[\mathrm{t}]}[0]\right) \boldsymbol{\mathcal{S}}^{\mathrm{T}}_{..f} + \boldsymbol{W}\boldsymbol{\mathcal{G}}_{..kf} + \boldsymbol{W}\boldsymbol{\mathcal{V}}_{..kf}, \tag{16}$$

where $\boldsymbol{\Phi}^{[\mathrm{x}]}[0] \in \mathbb{C}^{K \times M_X}$, for $\mathrm{x} \in \{\mathrm{r}, \mathrm{t}\}$, is the phase-noise matrix formed by collecting the phase-noise CPE terms across the K blocks and the M_X antennas, i.e.,

$$\boldsymbol{\Phi}^{[\mathrm{x}]}[0] = \begin{bmatrix} \Phi^{(\mathrm{x})}_{1,1}[0] & \dots & \Phi^{(\mathrm{x})}_{1,M_X}[0] \\ \vdots & \ddots & \vdots \\ \Phi^{(\mathrm{x})}_{K,1}[0] & \dots & \Phi^{(\mathrm{x})}_{K,M_X}[0] \end{bmatrix}. \tag{17}$$

The third-order tensor $\boldsymbol{\mathcal{S}} \in \mathbb{C}^{L_P \times M_T \times F}$ is obtained by collecting the transmitted symbols $S_{i,m_t}[f]$ across all L_P symbol periods, M_T antennas and F subcarriers, for $i = \{1, \dots, L_P\}$, $m_t = \{1, \dots, M_T\}$, and $f = \{1, \dots, F\}$.

In equation (16), $\boldsymbol{\mathcal{H}}_{..f}$ represents the f-th frontal slice of the frequency-selective channel tensor $\boldsymbol{\mathcal{H}} \in \mathbb{C}^{M_R \times M_T \times F}$, which, in Appendix IX, is given by

$$\boldsymbol{\mathcal{H}}_{a,b,f} = \sum_{l=1}^{L} \bar{h}_{a,b}[l] e^{-2j\pi f \tau_l d_{\mathrm{F}}}, \tag{18}$$

where $\bar{h}_{a,b}[l]$ is the channel gain of the l-th propagation path, $l = \{1, \dots, L\}$, τ_l is the associated delay (in seconds), $d_{\mathrm{F}} = 1/T_s$ is the frequency spacing between two subcarriers (in Hertz), and T_s is the symbol duration.

In tensor notation, equation (16) can be considered as a special case of a fourth-order PARATuck model. Next, we will show how to recast such a model as an equivalent third-order PARAFAC model that will be exploited for the joint phase-noise and channel estimation.

IV. PARAFAC Pilot Modeling

The symbol tensor $\boldsymbol{\mathcal{S}} \in \mathbb{C}^{L_P \times M_T \times F}$ is given by the concatenation of a pilot tensor $\boldsymbol{\mathcal{S}}^{(\mathrm{P})} \in \mathbb{C}^{L_P \times M_T \times F_{\mathrm{P}}}$ and a data symbol tensor $\boldsymbol{\mathcal{S}}^{(\mathrm{D})} \in \mathbb{C}^{L_P \times M_T \times F_{\mathrm{D}}}$, such that $F = F_{\mathrm{P}} + F_{\mathrm{D}}$, as illustrated in Figure 3. This concatenation depends on the value of $D = F/F_{\mathrm{P}}$, which is the distance between the pilot and data subcarriers, counted in indices. We can write the concatenation of the pilot and symbol tensors using the slice notation as

$$\boldsymbol{\mathcal{S}} = \left[\boldsymbol{\mathcal{S}}^{(\mathrm{P})}_{..1} \sqcup_3 \boldsymbol{\mathcal{S}}^{(\mathrm{D})}_{..1:D-1} \sqcup_3 \dots \sqcup_3 \boldsymbol{\mathcal{S}}^{(\mathrm{P})}_{..F_{\mathrm{P}}} \sqcup_3 \boldsymbol{\mathcal{S}}^{(\mathrm{D})}_{..D+F_{\mathrm{P}}:F_{\mathrm{D}}}\right]. \tag{19}$$

Likewise, the channel tensor $\boldsymbol{\mathcal{H}}$ is divided into a pilot channel part $\boldsymbol{\mathcal{H}}^{(\mathrm{P})} \in \mathbb{C}^{M_R \times M_T \times F_{\mathrm{P}}}$ and a data channel part $\boldsymbol{\mathcal{H}}^{(\mathrm{D})} \in \mathbb{C}^{M_R \times M_T \times F_{\mathrm{D}}}$.

As a first step, the receiver extracts only the pilot preamble of each block. From (16), the contribution of the pilot part $\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}_{..kf_p} \in \mathbb{C}^{M_R \times L_P}$ associated with the k-th block at the f_p-th pilot subcarrier can be expressed as

$$\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}_{..kf_p} = \boldsymbol{W} D_k \left(\boldsymbol{\Phi}^{[r]}[0]\right) \boldsymbol{\mathcal{H}}^{(\mathrm{P})}_{..f_p} D_k \left(\boldsymbol{\Phi}^{[t]}[0]\right) \boldsymbol{\mathcal{S}}^{(\mathrm{P})\mathrm{T}}_{..f_p} + \boldsymbol{W}\boldsymbol{\mathcal{V}}^{(\mathrm{P})}_{..kf_p} + \boldsymbol{W}\boldsymbol{\mathcal{G}}^{(\mathrm{P})}_{..kf_p}. \tag{20}$$

Let us consider that the same pilot is transmitted across all pilot subcarriers, i.e.,

$$\boldsymbol{\mathcal{S}}^{(\mathrm{P})}_{..f_p} = \boldsymbol{S}^{(\mathrm{P})} \in \mathbb{C}^{L_P \times M_T} \quad \forall f_p \in F_{\mathrm{P}}. \tag{21}$$

Then, we can write equation (20) as

$$\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}_{..kf_p} = \boldsymbol{W} D_k \left(\boldsymbol{\Phi}^{[\mathrm{r}]}[0]\right) \boldsymbol{\mathcal{H}}^{(\mathrm{P})}_{..f_p} D_k \left(\boldsymbol{\Phi}^{[\mathrm{t}]}[0]\right) \boldsymbol{S}^{(\mathrm{P})\mathrm{T}} + \boldsymbol{W}\boldsymbol{\mathcal{V}}^{(\mathrm{P})}_{..kf_p} + \boldsymbol{W}\boldsymbol{\mathcal{G}}^{(\mathrm{P})}_{..kf_p}. \tag{22}$$

Defining $\boldsymbol{y}^{(\mathrm{P})}_{kf_p} = \mathrm{vec}(\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}_{..kf_p}) \in \mathbb{C}^{M_R L_P \times 1}$, $\boldsymbol{h}_{f_p} = \mathrm{vec}(\boldsymbol{\mathcal{H}}^{(\mathrm{P})}_{..f_p}) \in \mathbb{C}^{M_R M_T \times 1}$ and applying properties (1), (2), and (3) to equation (22), neglecting the noise and the ICI terms, yields

$$\begin{aligned} \boldsymbol{y}^{(\mathrm{P})}_{kf_p} &= \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right) \mathrm{vec}\left(D_k \left(\boldsymbol{\Phi}^{[\mathrm{r}]}[0]\right) \boldsymbol{\mathcal{H}}^{(\mathrm{P})}_{..f_p} D_k \left(\boldsymbol{\Phi}^{[\mathrm{t}]}[0]\right)\right) \\ &= \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right) \left(D_k \left(\boldsymbol{\Phi}^{[\mathrm{t}]}[0]\right) \otimes D_k \left(\boldsymbol{\Phi}^{[\mathrm{r}]}[0]\right)\right) \boldsymbol{h}^{(\mathrm{P})}_{f_p} \\ &= \left(\boldsymbol{h}^{(\mathrm{P})\mathrm{T}}_{f_p} \diamond \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right)\right) \left(\boldsymbol{\Phi}^{[\mathrm{t}]}[0]_{k.} \otimes \boldsymbol{\Phi}^{[\mathrm{r}]}[0]_{k.}\right)^{\mathrm{T}} \\ &= \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right) \mathrm{diag}\left(\boldsymbol{h}^{(\mathrm{P})}_{f_p}\right) \boldsymbol{\Phi}^{\mathrm{T}}_{k.} \end{aligned} \tag{23}$$

where $\boldsymbol{\Phi}_{k.} = \boldsymbol{\Phi}^{[\mathrm{t}]}[0]_{k.} \otimes \boldsymbol{\Phi}^{[\mathrm{r}]}[0]_{k.} \in \mathbb{C}^{1 \times M_R M_T}$ is the combined phase-noise. Collecting the vectorized received pilots $\boldsymbol{y}^{(\mathrm{P})}_{kf_p}$ for all the $k = 1, \dots, K$ blocks as the columns of the resulting matrix $\boldsymbol{Y}^{(\mathrm{P})}_{f_p} = [\boldsymbol{y}^{(\mathrm{P})}_{1,f_p}, \dots, \boldsymbol{y}^{(\mathrm{P})}_{K,f_p}]$, and according to the definition of the Khatri-Rao product, we have

$$\begin{aligned} \boldsymbol{Y}^{(\mathrm{P})}_{f_p} &= \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right) \mathrm{diag}\left(\boldsymbol{h}^{(\mathrm{P})}_{f_p}\right) \left[\boldsymbol{\Phi}^{\mathrm{T}}_{1.}, \dots, \boldsymbol{\Phi}^{\mathrm{T}}_{K.}\right] \\ &= \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right) \mathrm{diag}\left(\boldsymbol{h}^{(\mathrm{P})}_{f_p}\right) \boldsymbol{\Phi}^{\mathrm{T}} \in \mathbb{C}^{M_R L_P \times K}, \end{aligned} \tag{24}$$

where $\boldsymbol{\Phi} = (\boldsymbol{\Phi}^{[t]\mathrm{T}}[0] \diamond \boldsymbol{\Phi}^{[r]\mathrm{T}}[0])^{\mathrm{T}} \in \mathbb{C}^{K \times M_R M_T}$ is the combined phase-noise across all the blocks and antennas. According to [30], $\boldsymbol{Y}^{(\mathrm{P})}_{f_p}$ can be interpreted as the f_p-th frontal slice of the third-order tensor $\boldsymbol{\mathcal{Y}}^{(\mathrm{P})} \in \mathbb{C}^{M_R L_P \times K \times F_{\mathrm{P}}}$ which corresponds to the following PARAFAC decomposition

$$\boldsymbol{\mathcal{Y}}^{(\mathrm{P})} = \boldsymbol{\mathcal{I}}_{3,M_RM_T} \times_1 (\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}) \times_2 \boldsymbol{\Phi} \times_3 [\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)}, \quad (25)$$

where $[\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)} \in \mathbb{C}^{F_\mathrm{P} \times M_RM_T}$ is the 3-mode unfolding of the channel tensor $\boldsymbol{\mathcal{H}}^{(\mathrm{P})} \in \mathbb{C}^{M_R \times M_T \times F_\mathrm{P}}$. The 1-mode, 2-mode and 3-mode unfoldings of $\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}$, denoted by $[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}]_{(1)} \in \mathbb{C}^{M_RL_P \times KF_\mathrm{P}}$, $[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}]_{(2)} \in \mathbb{C}^{K \times M_RL_PF_\mathrm{P}}$ and $[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}]_{(3)} \in \mathbb{C}^{F_\mathrm{P} \times M_RL_PK}$ respectively, admit the following factorizations:

$$\left[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}\right]_{(1)} = \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right)\left([\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)} \diamond \boldsymbol{\Phi}\right)^{\mathrm{T}}, \quad (26)$$

$$\left[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}\right]_{(2)} = \boldsymbol{\Phi}\left([\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)} \diamond \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right)\right)^{\mathrm{T}}, \quad (27)$$

$$\left[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}\right]_{(3)} = [\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)}\left(\boldsymbol{\Phi} \diamond \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right)\right)^{\mathrm{T}}. \quad (28)$$

Equations (26) to (28) are the basis for the first-stage of the proposed receiver, where two different algorithms can be used to estimate the combined phase-noise and the pilot channel.

V. Proposed Receiver

The proposed tensor-based receiver is composed of two processing stages. In the first stage, two different algorithms are derived to estimate the channel at the pilot subcarriers and the PN impairments. The first algorithm is based on BALS (Bilinear Alternating Least Squares), and consists of estimating the channel from the pilot subcarriers and the PN terms in an iterative way, while the second one has a closed-form solution based on the LS-KRF (Least Squares Khatri-Rao Factorization) for simultaneous channel and phase-noise estimation. Once the channel coefficients associated with the pilot subcarriers are estimated, the channel coefficients at the data subcarriers are obtained by interpolating the estimated channel at the pilot subcarriers. In the second stage, the proposed ZFSKP receiver, a Zero-Forcing (ZF) approach based on the SKP formulation, is used to estimate the data symbols. These two stages are detailed below.

A. *Stage 1: Channel and PN Estimation via BALS*

The first stage of the proposed receiver consists of jointly estimating the channel and PN impairments from the received signal tensor in (25), which can be done by exploiting the unfoldings $[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}]_{(2)}$ and $[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}]_{(3)}$ from equations (27) and (28), respectively. To this end, we make use of the BALS algorithm that alternates between the estimation of the factor matrices $\boldsymbol{\Phi}$ and $[\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)}$ by optimizing the following two least squares (LS) criteria:

$$\hat{\boldsymbol{\Phi}} = \underset{\boldsymbol{\Phi}}{\operatorname{argmin}} \left\| \left[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}\right]_{(2)} - \boldsymbol{\Phi}\left([\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)} \diamond \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right)\right)^{\mathrm{T}} \right\|^2_{\mathrm{F}},$$

$$\left[\hat{\boldsymbol{\mathcal{H}}}\right]^{(\mathrm{P})}_{(3)} = \underset{[\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)}}{\operatorname{argmin}} \left\| \left[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}\right]_{(3)} - \left[\hat{\boldsymbol{\mathcal{H}}}\right]^{(\mathrm{P})}_{(3)}\left(\boldsymbol{\Phi} \diamond \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right)\right)^{\mathrm{T}} \right\|^2_{\mathrm{F}},$$

the solutions of which are given, respectively, by

$$\hat{\boldsymbol{\Phi}} = \left[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}\right]_{(2)} \left[\left([\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)} \diamond \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right)\right)^{\mathrm{T}}\right]^{+}, \quad (29)$$

$$\left[\hat{\boldsymbol{\mathcal{H}}}\right]^{(\mathrm{P})}_{(3)} = \left[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}\right]_{(3)} \left[\left(\boldsymbol{\Phi} \diamond \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right)\right)^{\mathrm{T}}\right]^{+}. \quad (30)$$

Due to the knowledge of $\boldsymbol{W}$ and $\boldsymbol{S}^{(\mathrm{P})}$ at the receiver, each iteration of the BALS algorithm contains only two updating steps. At each step, the fitting error is minimized with respect to one factor matrix by fixing the other to its value obtained at the previous updating step. This procedure is repeated until the convergence of the BALS stage at the i-th iteration. The convergence is declared when $|e_{(i)} - e_{(i-1)}| \leq 10^{-6}$, where $e_{(i)}$ denotes the residual error calculated as

$$e_{(i)} = \frac{\left\|\boldsymbol{\mathcal{Y}}^{(\mathrm{P})} - \hat{\boldsymbol{\mathcal{Y}}}^{(\mathrm{P})}_{(i)}\right\|^2_{\mathrm{F}}}{\left\|\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}\right\|^2_{\mathrm{F}}}, \quad (31)$$

where $\hat{\boldsymbol{\mathcal{Y}}}^{(\mathrm{P})}_{(i)}$ is the reconstructed tensor $\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}$ computed from the estimated factor matrices at the end of the i-th iteration.

Note that, for the initialization of Algorithm 1, we consider that the phase-noise variance is known at the receiver. This assumption is realistic if we note that the phase-noise variance parameters (β and T_{sp}) are usually known at the receiver, or can be considered known in advance, since they are parameters intrinsic to the oscillators and fixed by the system design [10].

B. *Stage 1: Channel and PN Estimation Via LS-KRF*

The proposed LS-KRF algorithm estimates the channel at the pilot subcarriers and the phase-noise contributions in a closed-form manner. Transposing the unfolding $[\boldsymbol{\mathcal{Y}}]^{(\mathrm{P})}_{(1)}$ in equation (26), and applying a pseudo-inverse at the right, we have

$$[\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)} \diamond \boldsymbol{\Phi} \approx \left[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}\right]^{\mathrm{T}}_{(1)} \left(\boldsymbol{S}^{(\mathrm{P})\mathrm{T}} \otimes \boldsymbol{W}^{\mathrm{T}}\right)^{+}. \quad (32)$$

Note that the uniqueness of the Khatri-Rao product term in (32) requires that $\boldsymbol{S}^{(\mathrm{P})\mathrm{T}} \otimes \boldsymbol{W}$ to be full-column rank, which implies $L_P \geq M_T$.

Let us define $\boldsymbol{Z} \approx [\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)} \diamond \boldsymbol{\Phi} \in \mathbb{C}^{KF_\mathrm{P} \times M_RM_T}$ and note that $\boldsymbol{Z}_{.r} \approx [\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3).r} \otimes \boldsymbol{\Phi}_{.r}$, with $r = \{1, \ldots, M_RM_T\}$. From property (4), we have

$$\mathrm{vec}\left(\boldsymbol{\Phi}_{.r} \circ [\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3).r}\right) \approx \boldsymbol{Z}_{.r} \in \mathbb{C}^{KF_\mathrm{P} \times 1}. \quad (33)$$

Applying the unvec$(\cdot)$ operator, we have

$$\overline{\boldsymbol{Z}}^{(r)} \approx \boldsymbol{\Phi}_{.r} \cdot [\boldsymbol{\mathcal{H}}]^{(\mathrm{P})\mathrm{T}}_{(3).r} \in \mathbb{C}^{K \times F_\mathrm{P}}. \quad (34)$$

Hence, an estimation of $\boldsymbol{\Phi}_{.r}$ and $[\boldsymbol{\mathcal{H}}]^{(\mathrm{P})\mathrm{T}}_{(3).r}$ can be obtained, respectively, from the dominant left and right singular vectors of $\overline{\boldsymbol{Z}}^{(r)} = \boldsymbol{U}^{(r)}\boldsymbol{\Sigma}^{(r)}\boldsymbol{V}^{(r)\mathrm{H}}$, i.e.,

$$\hat{\boldsymbol{\Phi}}_{.r} = \left(\boldsymbol{\Sigma}^{(r)}_{1,1}\right)^{1/2} \boldsymbol{U}^{(r)}_{.1} \quad (35)$$

$$\left[\hat{\boldsymbol{\mathcal{H}}}\right]^{(\mathrm{P})}_{(3).r} = \left(\boldsymbol{\Sigma}^{(r)}_{1,1}\right)^{1/2} \boldsymbol{V}^{(r)*}_{.1}, \quad (36)$$

Algorithm 1: Combined BALS and ZFSKP.

Stage 1: PN and Channel estimation via the BALS Algorithm

1: Generate $\hat{\phi}^{[r]}[n] \sim \mathcal{N}(0, \sigma^2_{\phi^{[r]}})$ and $\hat{\phi}^{[t]}[n] \sim \mathcal{N}(0, \sigma^2_{\phi^{[t]}})$ using the Wiener model;
2: Set $\hat{\boldsymbol{\Phi}}^{[r]}_{k,m_r}[0] = 1/F \sum_{n=0}^{F-1} e^{j\hat{\phi}^{[r]}_{k,m_r}[n]}$ and $\hat{\boldsymbol{\Phi}}^{[t]}_{k,m_t}[0] = 1/F \sum_{n=0}^{F-1} e^{j\hat{\phi}^{[t]}_{k,m_t}[n]}$;
3: Iteration $i = 0$;
4: Set $\hat{\boldsymbol{\Phi}}_{(0)} = \hat{\boldsymbol{\Phi}}^{[t]\mathrm{T}}[0] \diamond \hat{\boldsymbol{\Phi}}^{[r]\mathrm{T}}[0]$;
5: $i = i + 1$;
6: Compute an estimation for $[\hat{\boldsymbol{\mathcal{H}}}]^{(\mathrm{P})}_{(3)(i)}$ using

$$\left[\hat{\boldsymbol{\mathcal{H}}}\right]^{(\mathrm{P})}_{(3)(i)} = \left[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}\right]_{(3)} \left[\left(\hat{\boldsymbol{\Phi}}_{(i-1)} \diamond \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right)\right)^{\mathrm{T}}\right]^{+},$$

7: Compute an estimation for $\hat{\boldsymbol{\Phi}}_{(i)}$ using

$$\hat{\boldsymbol{\Phi}}_{(i)} = \left[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}\right]_{(2)} \left[\left(\left[\hat{\boldsymbol{\mathcal{H}}}\right]^{(\mathrm{P})}_{(3)(i)} \diamond \left(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}\right)\right)^{\mathrm{T}}\right]^{+},$$

8: Reconstruct tensor $\hat{\boldsymbol{\mathcal{Y}}}^{(\mathrm{P})}_{(i)}$ and compute the residual error according to equation (31);
9: Return to step 5 and repeat until convergence;
10: Estimate the channel at the data subcarriers $[\hat{\boldsymbol{\mathcal{H}}}]^{(\mathrm{D})}_{(3)}$ by interpolating $[\hat{\boldsymbol{\mathcal{H}}}]^{(\mathrm{P})}_{(3)}$;

Stage 2: Data estimation via the ZFSKP approach

11: Reshape the elements of $\mathrm{diag}(\mathrm{vec}([\hat{\boldsymbol{\mathcal{H}}}]^{(\mathrm{D})\mathrm{T}}_{(3)}))$ into $\overline{\boldsymbol{\mathcal{H}}}^{(\mathrm{D})}$ using equation (41).
12: Estimate the transmitted data using eq. (42).

$r = 1, \ldots, M_R M_T$. The estimation of the entire matrices $\hat{\boldsymbol{\Phi}}$ and $[\hat{\boldsymbol{\mathcal{H}}}]^{(\mathrm{P})}_{(3)}$ requires the computation of $M_R M_T$ rank-one matrix approximations. Instead of computing SVDs, efficient solutions based on the power method can be used [42].

C. Scaling Ambiguity in the Estimated Matrices At Stage 1

It is important to mention that, after the BALS and the LS-KRF in Stage 1, the estimated matrices are linked to the true ones by the following relationship:

$$\left[\hat{\boldsymbol{\mathcal{H}}}\right]^{(\mathrm{P})}_{(3)} = [\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)} \boldsymbol{\Lambda}, \hat{\boldsymbol{\Phi}} = \boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1}, \tag{37}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda}) \in \mathbb{C}^{M_R M_T \times M_R M_T}$, with $\boldsymbol{\lambda} = [\alpha_1, \ldots, \alpha_{M_R M_T}]$ being a vector containing $M_R M_T$ scaling factors. To solve these scaling ambiguities, one row of $\boldsymbol{\Phi}$ or $[\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)}$ has to be known, which physically means that the phase-noise process at a certain block k is known, or, alternatively, that the CSI at a certain pilot subcarrier f_p is available. However, since we are interested in data estimation, the scaling factors affecting the channel matrix are canceled with the ones acting over the phase-noise matrix, i.e., $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1} = \boldsymbol{I}_{M_R M_T}$. Thus, the role of the BALS and the LS-KRF algorithms is to decouple and refine the estimates of the channel and phase-noise matrices before combining them for the data estimation, according to eq. (42).

Algorithm 2: Combined LS-KRF and ZFSKP.

Stage 1: PN and Channel estimation via the LS-KRF Algorithm

1: Compute an estimate of $\boldsymbol{Z}$ as

$$\hat{\boldsymbol{Z}} = \left[\boldsymbol{\mathcal{Y}}^{(\mathrm{P})}\right]^{\mathrm{T}}_{(1)} \left(\boldsymbol{S}^{(\mathrm{P})\mathrm{T}} \otimes \boldsymbol{W}^{\mathrm{T}}\right)^{+}$$

2: **for** $r = 1 : M_R M_T$ **do**
3: Define $\overline{\boldsymbol{Z}} = \mathrm{unvec}(\hat{\boldsymbol{Z}}_{.r}) \in \mathbb{C}^{K \times F_{\mathrm{P}}}$
4: Compute the SVD of $\overline{\boldsymbol{Z}} = \boldsymbol{U}^{(r)} \boldsymbol{\Sigma}^{(r)} \boldsymbol{V}^{(r)\mathrm{H}}$
5: Estimate $\hat{\boldsymbol{\Phi}}_{.r} = (\boldsymbol{\Sigma}^{(r)}_{1,1})^{1/2} \boldsymbol{U}^{(r)}_{.1}$
6: Estimate $[\hat{\boldsymbol{\mathcal{H}}}]^{(\mathrm{P})}_{(3).r} = (\boldsymbol{\Sigma}^{(r)}_{1,1})^{1/2} \boldsymbol{V}^{(r)*}_{.1}$
7: **end for**
8: Estimate the channel at the data subcarriers $[\hat{\boldsymbol{\mathcal{H}}}]^{(\mathrm{D})}_{(3)}$ by interpolating $[\hat{\boldsymbol{\mathcal{H}}}]^{(\mathrm{P})}_{(3)}$;

Stage 2: Data estimation via the ZFSKP approach

9: Reshape the elements of $\mathrm{diag}(\mathrm{vec}([\hat{\boldsymbol{\mathcal{H}}}]^{(\mathrm{D})\mathrm{T}}_{(3)}))$ into $\overline{\boldsymbol{\mathcal{H}}}^{(\mathrm{D})}$ using equation (41).
10: Estimate the transmitted data using eq. (42).

D. *Stage 2*: Data Estimation Using the ZFSKP Approach

We derive an SKP-based ZF approach, namely ZFSKP, to estimate the transmitted symbol tensor from the data subcarriers. Let us consider the received signal at the k-th frame and f_d-th data subcarrier in equation (16). By defining $\boldsymbol{y}^{(\mathrm{D})}_{kf_d} \doteq \mathrm{vec}(\boldsymbol{\mathcal{Y}}^{(\mathrm{D})}_{..kf_d}) \in \mathbb{C}^{M_R L_P \times 1}$ and $\boldsymbol{h}^{(\mathrm{D})}_{f_d} = \mathrm{vec}(\boldsymbol{\mathcal{H}}^{(\mathrm{D})}_{..f_d}) \in \mathbb{C}^{M_R M_T \times 1}$, and neglecting the noise and the ICI terms (for notational convenience), we have

$$\begin{aligned}\boldsymbol{y}^{(\mathrm{D})}_{kf_d} &= \mathrm{vec}\left(\boldsymbol{W} D_k\left(\boldsymbol{\Phi}^{[\mathrm{r}]}[0]\right) \boldsymbol{\mathcal{H}}^{(\mathrm{D})}_{..f_d} D_k\left(\boldsymbol{\Phi}^{[\mathrm{t}]}[0]\right) \boldsymbol{\mathcal{S}}^{(\mathrm{D})\mathrm{T}}_{..f_d}\right)\\ &= \left(\boldsymbol{\mathcal{S}}^{(\mathrm{D})}_{..f_d} \otimes \boldsymbol{W}\right) \mathrm{diag}(\boldsymbol{h}^{(\mathrm{D})}_{f_d}) \left(\boldsymbol{\Phi}^{[\mathrm{t}]}[0]_{k.} \otimes \boldsymbol{\Phi}^{[\mathrm{r}]}[0]_{k.}\right)^{\mathrm{T}}.\end{aligned}$$

By collecting data during the K frames, we obtain

$$\boldsymbol{Y}^{(\mathrm{D})}_{f_d} = \left(\boldsymbol{\mathcal{S}}^{(\mathrm{D})}_{..f_d} \otimes \boldsymbol{W}\right) \mathrm{diag}(\boldsymbol{h}^{(\mathrm{D})}_{f_d}) \boldsymbol{\Phi}^{\mathrm{T}} \in \mathbb{C}^{M_R L_P \times K}. \tag{38}$$

Note that (38) differs from (24) since the term $\boldsymbol{\mathcal{S}}^{(\mathrm{D})}_{..f_d}$ varies over the F_{D} subcarriers, which means that equation (38) does not fit a PARAFAC model. However, by collecting all the F_{D} terms in $\boldsymbol{Y}^{(\mathrm{D})} = [\boldsymbol{Y}^{(\mathrm{D})}_1 \ldots \boldsymbol{Y}^{(\mathrm{D})}_{F_{\mathrm{D}}}] \in \mathbb{C}^{M_R L_P \times K F_{\mathrm{D}}}$, we get

$$\begin{aligned}\boldsymbol{Y}^{(\mathrm{D})} &= \left[\left(\boldsymbol{\mathcal{S}}^{(\mathrm{D})}_{..1} \otimes \boldsymbol{W}\right), \ldots, \left(\boldsymbol{\mathcal{S}}^{(\mathrm{D})}_{..F_{\mathrm{D}}} \otimes \boldsymbol{W}\right)\right]\\ &\quad \cdot \mathrm{diag}\left(\begin{bmatrix}\boldsymbol{h}^{(\mathrm{D})}_1\\ \vdots\\ \boldsymbol{h}^{(\mathrm{D})}_{F_D}\end{bmatrix}\right) \begin{bmatrix}\boldsymbol{\Phi} & & \\ & \ddots & \\ & & \boldsymbol{\Phi}\end{bmatrix}^{\mathrm{T}}\\ &= \left([\boldsymbol{\mathcal{S}}]^{(\mathrm{D})}_{(1)} \otimes \boldsymbol{W}\right) \mathrm{diag}\left(\mathrm{vec}([\boldsymbol{\mathcal{H}}]^{(\mathrm{D})\mathrm{T}}_{(3)})\right) (\boldsymbol{I}_{F_{\mathrm{D}}} \otimes \boldsymbol{\Phi})^{\mathrm{T}}.\end{aligned} \tag{39}$$

The received data matrix $\boldsymbol{Y}^{(\mathrm{D})} \in \mathbb{C}^{M_R L_P \times K F_\mathrm{D}}$ can be interpreted as a generalized unfolding of the fourth-order received data tensor $\boldsymbol{\mathcal{Y}}^{(\mathrm{D})} \in \mathbb{C}^{M_R \times L_P \times K \times F_\mathrm{D}}$ given by

$$\boldsymbol{\mathcal{Y}}^{(\mathrm{D})} = \overline{\boldsymbol{\mathcal{H}}}^{(\mathrm{D})} \times_1 \boldsymbol{W} \times_2 [\boldsymbol{\mathcal{S}}]_{(1)}^{(\mathrm{D})} \times_3 \boldsymbol{\Phi} \times_4 \boldsymbol{I}_{F_\mathrm{D}}, \tag{40}$$

where $\overline{\boldsymbol{\mathcal{H}}}^{(\mathrm{D})}$ is the core tensor, which corresponds to a reshaping of the elements of $\mathrm{diag}(\mathrm{vec}([\boldsymbol{\mathcal{H}}]_{(3)}^{(\mathrm{D})\mathrm{T}})) \in \mathbb{C}^{M_R M_T F_\mathrm{D} \times M_R M_T F_\mathrm{D}}$ as a fourth-order tensor $\overline{\boldsymbol{\mathcal{H}}}^{(\mathrm{D})} \in \mathbb{C}^{M_R \times M_T F_\mathrm{D} \times M_R M_T \times F_\mathrm{D}}$ that can be constructed similar to the tensor $\boldsymbol{\mathcal{B}}$ in equation (11), using the SKP structure introduced in Section II-A as

$$\overline{\boldsymbol{\mathcal{H}}}^{(\mathrm{D})} = \left(\underbrace{\boldsymbol{\mathcal{I}}_{3,F_\mathrm{D}} \otimes_{2,3}^{2,4} (\boldsymbol{\mathcal{I}}_{3,M_T} \otimes_{2,3}^{2,3} \boldsymbol{\mathcal{I}}_{3,M_R})}_{\in \mathbb{R}^{M_R \times M_T F_\mathrm{D} \times M_R M_T \times M_R M_T F_\mathrm{D} \times F_\mathrm{D}}} \right) \times_4 \mathrm{vec}\left([\boldsymbol{\mathcal{H}}]_{(3)}^{(\mathrm{D})\mathrm{T}} \right)^{\mathrm{T}}. \tag{41}$$

From equation (40), an estimate of the data symbol tensor can be obtained as

$$\left[\hat{\boldsymbol{\mathcal{S}}}\right]_{(1)}^{(\mathrm{D})} = [\boldsymbol{\mathcal{Y}}]_{(2)}^{(\mathrm{D})} \left(\left[\hat{\overline{\boldsymbol{\mathcal{H}}}}\right]_{(2)}^{(\mathrm{D})} (\boldsymbol{I}_{F_\mathrm{D}} \otimes \hat{\boldsymbol{\Phi}} \otimes \boldsymbol{W})^{\mathrm{T}} \right)^{+}, \tag{42}$$

where $\hat{\boldsymbol{\Phi}}$ is the PN matrix estimated in the first stage, and $[\hat{\overline{\boldsymbol{\mathcal{H}}}}]_{(2)}^{(\mathrm{D})}$ is the 2-mode unfolding of the tensor $\overline{\boldsymbol{\mathcal{H}}}^{(\mathrm{D})}$ obtained at step 11 and 9 of Algorithms 1 and 2, respectively.

E. Uniqueness and Identifiability Conditions

In this section, we discuss the range of parameter settings that ensure the uniqueness and identifiability of the proposed receiver. First, recall equation (9), where the uniqueness condition for a third-order PARAFAC model is stated. Adapting this condition to our system context, we can rewrite equation (9) as

$$k_{(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W})} + k_{\boldsymbol{\Phi}} + k_{[\boldsymbol{\mathcal{H}}]_{(3)}^{(\mathrm{P})}} \geq 2(M_R M_T) + 2. \tag{43}$$

Since the factor $\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W} \in \mathbb{C}^{M_R L_P \times M_R M_T}$ is known at the receiver, we can ensure by a proper design such that $\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}$ has full rank, so that $k_{(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W})} = \mathrm{rank}(\boldsymbol{S}^{(\mathrm{P})} \otimes \boldsymbol{W}) = M_R M_T$. The factor $[\boldsymbol{\mathcal{H}}]_{(3)}^{(\mathrm{P})} \in \mathbb{C}^{F_\mathrm{P} \times M_R M_T}$ can also be assumed to be a full rank matrix under a rich scattering propagation environment, which implies that $k_{[\boldsymbol{\mathcal{H}}]_{(3)}^{(\mathrm{P})}} = \mathrm{rank}([\boldsymbol{\mathcal{H}}]_{(3)}^{(\mathrm{P})}) = M_R M_T$. Under these assumptions, we can deduce from (43) that $k_{\boldsymbol{\Phi}} \geq 2$ is enough to ensure the uniqueness of model (9) and, hence, to guarantee a joint channel and phase-noise recovery.

Finally, note that LS estimation steps 6 and 7 of Algorithm 1 require that $F L_P \geq M_T$ and $K L_P \geq M_T$, while the ZF estimation of the data symbols in equation (42) requires that $M_T \leq M_R K$.

F. Computational Complexity

Here we derive the computational complexity of the proposed receiver in terms of FLOPS (Floating-point Operations Per Second). Considering a matrix $\boldsymbol{A} \in \mathbb{C}^{M \times N}$, we know that the computation of its dominant singular value and singular vectors have a cost of $\mathcal{O}(l \cdot (N^2\, M + M^2\, N))$, where l is the maximum number of iterations of the power method (see [42]). In our experiments, we have noted that $l = 1$ is enough to achieve a good accuracy. Moreover, we know that $\mathcal{O}(N^2\, M)$ operations are required to compute the pseudo-inverse of $\boldsymbol{A}$, with $N < M$. Using these results, we find that Stage 1 via BALS has a complexity of order $\mathcal{O}(I \cdot 2((M_R M_T)^2 M_R L_P (F_\mathrm{P} + K)))$, where I is the number of iterations required to achieve the convergence. If the LS-KRF is used in Stage 1, we have a complexity of $\mathcal{O}(M_R M_T (F_\mathrm{P} K (F_\mathrm{P} + K)))$. Finally, the complexity associated with the data symbol estimation in Stage 2 is given by $\mathcal{O}((F_\mathrm{D} M_T)^3)$.

VI. Simulation Results

We evaluate the performance of the proposed receiver in terms of the normalized mean square error (NMSE) between the estimated and true frequency-selective MIMO channel and the combined phase-noise matrix, the symbol error rate (SER), and the number of iterations for convergence. Our simulation results represent an average over $M = 5000$ independent Monte Carlo runs. Each run corresponds to an independent realization of the channel, pilots, data, additive noise, ICI, and phase-noise. The transmitted symbols are normalized such that the $\mathrm{SNR} = \frac{1}{\sigma_v^2}$ is controlled by varying the noise power σ_v^2 of the AWGN tensor $\boldsymbol{\mathcal{V}}$, while the variance of the ICI σ_{ICI}^2 is kept constant, since it depends on the variance of the phase-noise and on the number of subcarriers, as detailed in the Appendix B.

The combiner matrix $\boldsymbol{W} \in \mathbb{C}^{M_R \times M_R}$ is designed as a DFT matrix. The NMSE is defined as

$$\mathrm{NMSE}(\hat{\boldsymbol{\mathcal{H}}}^{(\mathrm{D})}) = \frac{1}{M} \sum_{m=1}^{M} \frac{\|\boldsymbol{\mathcal{H}}^{(\mathrm{D})} - \hat{\boldsymbol{\mathcal{H}}}^{(\mathrm{D})}\|_\mathrm{F}}{\|\boldsymbol{\mathcal{H}}^{(\mathrm{D})}\|_\mathrm{F}}, \tag{44}$$

$$\mathrm{NMSE}(\hat{\boldsymbol{\Phi}}) = \frac{1}{M} \sum_{m=1}^{M} \frac{\|\boldsymbol{\Phi} - \hat{\boldsymbol{\Phi}}\|_\mathrm{F}}{\|\boldsymbol{\Phi}\|_\mathrm{F}}, \tag{45}$$

Unless stated otherwise, our simulations consider $F = 64$ subcarriers, of which $F_\mathrm{P} = 16$ are used for pilots, and $F_\mathrm{D} = 48$ for data transmission. Both pilots and data symbols follow a 4-QAM constellation. We also assume $M_T = L_P = 2$, and $M_R = K = 4$ frames. For the channel model in equation (18), the coefficients $\{h[l]\}$ are i.i.d. complex Gaussian random variables with zero mean and unit variance. The number of propagation paths is equal to $L = 16$, and the frequency spacing between two subcarriers is $d_\mathrm{F} = 15$ kHz. The path delays are chosen as $\tau_l = \mathrm{DS} \cdot v_l$, where v_l is a random variable that follows a normal distribution with $v_l \sim \mathcal{N}(0, 1)$ and $\mathrm{DS} = 1\,\mu$s is the channel delay spread. The variance of the phase-noise at each transmit and receive antenna is assumed to be $\sigma_{\phi^{(\mathrm{x})}}^2 = 5 \cdot 10^{-3}$, for $\mathrm{x} \in \{\mathrm{r}, \mathrm{t}\}$, thus, for $F = 64$, we have an ICI power of $\sigma_{\mathrm{ICI}}^2 \approx 0.04$, according to eq.(58). To estimate the channel coefficients at the data subcarriers from those at the pilot subcarriers, we consider cubic splines for the channel interpolation [43]. Indeed, in our simulations, the channel delay taps do not necessarily coincide with the sampling period, i.e., $\tau_l d_\mathrm{F} \neq \frac{l}{F}$ in eq. (18). In this situation, the spline-based interpolation provides a more accurate

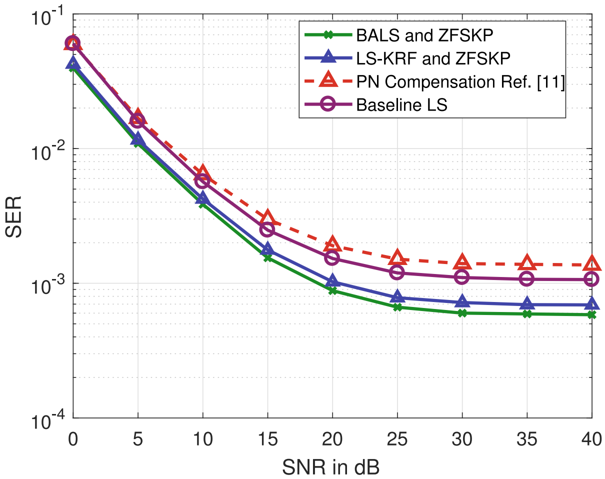


Fig. 4. SER $vs.$ SNR.

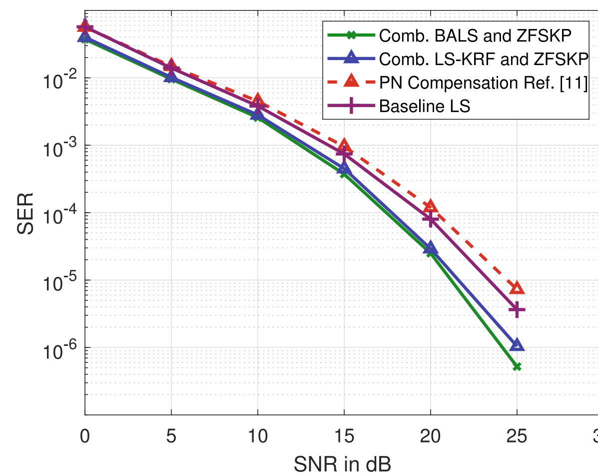


Fig. 5. SER $vs.$ SNR, for a perfect ICI suppression case.

reconstruction of the frequency-domain channel compared to the traditional DFT-based interpolation.

In the following experiments, we compare the performance of the proposed receivers, namely, the combined BALS and ZFSKP, as well as the combined LS-KRF and ZFSKP, with those of two different competitors: 1) the PN compensation scheme based on [11] and 2) the baseline LS estimator.

The competing receivers start from the LS estimate provided by eq. (32), i.e., $\boldsymbol{Z} \approx [\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)} \diamond \boldsymbol{\Phi} \in \mathbb{C}^{KF_\mathrm{P}\times M_R M_T}$. The PN compensation scheme of [11] considers $\boldsymbol{Z}$ as the effective pilot channel. An estimate of the effective data channel $\hat{\boldsymbol{\mathcal{Z}}}^{(\mathrm{D})} \in \mathbb{C}^{M_R\times M_T\times F_\mathrm{D}}$ is then obtained from the time-domain by averaging over the K blocks followed by interpolation. We use the following subcarrier-wise ZF for data estimation:

$$\hat{\boldsymbol{\mathcal{S}}}^{(\mathrm{D})}_{..f_\mathrm{D}} = \left(\boldsymbol{W}\hat{\boldsymbol{\mathcal{Z}}}^{(\mathrm{D})}_{..f_\mathrm{D}}\right)^{+} \bar{\boldsymbol{\mathcal{Y}}}^{(\mathrm{D})}_{..f_D},$$

where $\bar{\boldsymbol{\mathcal{Y}}}^{(\mathrm{D})}_{..f_D} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{\mathcal{Y}}^{(\mathrm{D})}_{..kf_D} \in \mathbb{C}^{M_R\times L_P}$. For the baseline LS estimator, we consider the solution of problem (32) in two different manners. If the pilot channel $[\boldsymbol{\mathcal{H}}]^{(\mathrm{P})}_{(3)}$ is known, an estimate of the phase-noise matrix $\boldsymbol{\Phi}$ can be extracted from $\boldsymbol{Z}$ using the Khatri-Rao factorization algorithm with one known factor, as discussed in [44]. On the other hand, if the phase-noise matrix is known, the same algorithm can be applied to estimate the pilot channel. Then, the data channel is estimated by interpolating the estimated pilot channel followed by ZF filtering

$$\hat{\boldsymbol{\mathcal{S}}}^{(\mathrm{D})}_{..f_D} = \left(\boldsymbol{W}(\bar{\hat{\boldsymbol{\Phi}}} \odot \hat{\boldsymbol{\mathcal{H}}}^{(\mathrm{D})}_{..f_D})\right)^{+} \bar{\boldsymbol{\mathcal{Y}}}^{(\mathrm{D})}_{..f_D},$$

where $\bar{\hat{\boldsymbol{\Phi}}} = \frac{1}{K}\sum_{k=1}^{K} \mathrm{unvec}(\hat{\boldsymbol{\Phi}}_{k.}) \in \mathbb{C}^{M_R\times M_T}$ is phase-noise matrix estimated by averaging it over the K blocks, $\hat{\boldsymbol{\mathcal{H}}}^{(\mathrm{D})}_{..f_D} \in \mathbb{C}^{M_R\times M_T}$ is the estimated data channel at the f_D data subcarrier, and "$\odot$" stands for the Hadamard product.

Figure 4 shows that the presence of a high phase-noise variance $\sigma^2_{\phi^{[\mathrm{x}]}} = 5\cdot 10^{-3}$ at the transmitter and the receiver introduces a significant ICI perturbation ($\sigma^2_{\mathrm{ICI}} \approx 0.04$), which explains the saturation of the curves at higher SNR values. We can also note that our proposed receivers achieve a target SER of 10^{-3} at an SNR of 20 dB, while the baseline LS receiver only achieves this target SER at an SNR close to 30 dB. Therefore, a significant SNR gain can be obtained by exploiting the tensor structure of the data model. More specifically, this remarkable gain comes from two factors: 1) the decoupling of the channel matrix from the phase-noise matrix, which allows us to successfully refine these estimates via BALS and LS-KRF, improving the data estimation performance; 2) the all-at-once estimation of the data symbols using the ZFSKP filter that exploits the tensor structure of the received signal through the combination of all the subcarriers.

In Figure 5, we show that our proposed receiver outperforms the baseline LS estimator and the phase-noise compensation scheme of [11] in the ideal case where the ICI is perfectly suppressed. More specifically, for a target SER of 10^{-3}, our solution provides an SNR gain of approximately 3 dB over the competing PN compensation scheme, and 1 dB over the baseline LS estimator. Again, this result can be explained by the fact that the Comb. BALS and ZFSKP, and the Comb. LS-KRF and ZFSKP receivers effectively exploit the tensor structure of the received signal, while the baseline LS solution and the PN compensation scheme perform the phase-noise compensation on a per subcarrier basis, thus ignoring the multidimensional structure of the problem.

In Figures 6 and 7, the performances of the two proposed receivers are depicted in comparison with the LS solution. The goal is to evaluate the closeness of the NMSE performance of the estimated phase-noise matrix and of the estimated data channel matrix with respect to the baseline LS receiver that assumes perfect knowledge of the channel (Figure 6) or perfect knowledge of the phase-noise (Figure 7). For the proposed receivers, we assume that the first row of the phase-noise matrix $\boldsymbol{\Phi}$ is known in order to eliminate the scaling ambiguity in the estimates provided by BALS and LS-KRF, as discussed in Section V-C. Please note that this assumption is not required for data estimation. Figure 6 shows that the proposed receivers perform very close to the baseline LS one that assumes a perfect knowledge of the channel. The same conclusions can be drawn

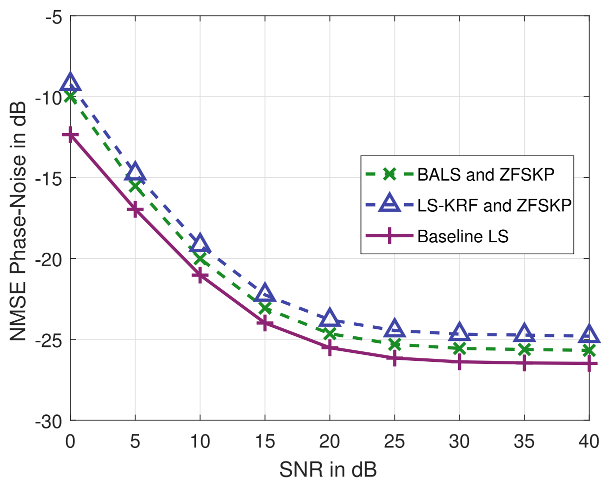


Fig. 6. NMSE of the phase-noise matrix $\mathbf{\Phi}$ $vs.$ SNR.

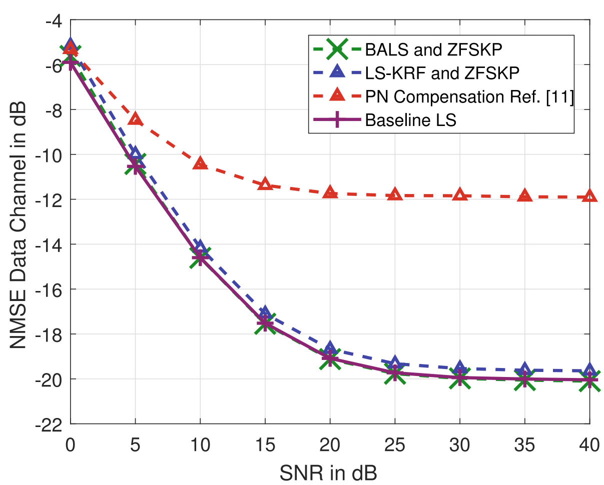


Fig. 7. NMSE of $\hat{\mathcal{H}}^{(\mathrm{D})}$ $vs.$ SNR.

from Figure 7, where the baseline LS receiver assumes a perfect knowledge of the phase-noise matrix.

Figure 8 compares the computational complexity between the BALS and LS-KRF algorithms for different numbers of pilot subcarriers. It is important to mention that, for the BALS algorithm, $I = 6$ iterations are needed to achieve the convergence for higher SNR values (> 10 dB). Note that the pseudo-inverse cost in step 1 in Algorithm 2 is not taken into account since the factors $\boldsymbol{S}^{(\mathrm{P})\mathrm{T}}$ and $\boldsymbol{W}$ are known, which means that it can be computed off-line. A trade-off between BALS and LS-KRF exists. The first is more attractive for a small number of receiving antennas (e.g., downlink), while the second is preferable when assuming a larger number of receiving antennas (e.g., uplink). Regarding the baseline LS estimator, we can observe that it is a cheaper algorithm compared to the BALS and the LS-KRF algorithms. However, as shown in Figure 4, our proposed receiver achieves a better SER performance. Also, the baseline LS estimator can be considered as a benchmark receiver since the pilot channel and phase-noise matrices are assumed to be known.

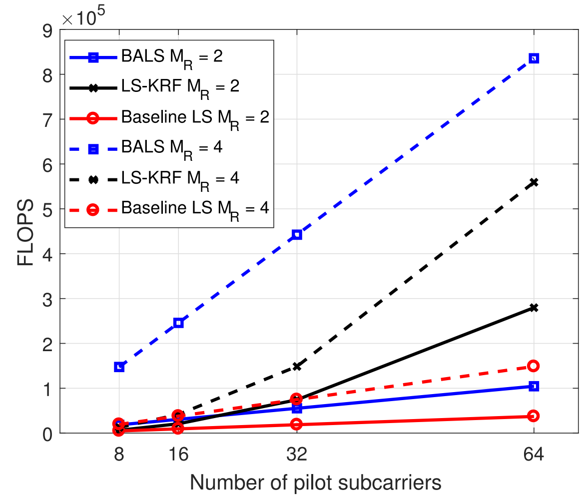


Fig. 8. Computational complexity of the BALS, LS-KRF, and Baseline LS algorithms for different scenarios.

VII. Conclusions and Future Perspectives

We have proposed a two-stage tensor-based receiver for joint channel, PN, and data estimation in MIMO-OFDM systems. For the first stage, we have derived two algorithms for channel and phase-noise estimation. The first one consists of an iterative solution based on the BALS, while the second algorithm has a closed-form solution based on the LS-KRF. Our numerical simulations show that both algorithms achieve a similar performance, but LS-KRF becomes preferable as compared to the BALS when the number of receiving antennas is increased due to its lower computational complexity. We also have proposed a ZFSKP equalizer that exploits the PARATuck tensor structure of the received signal via the Selective Kronecker Product (SKP) operator, allowing us to estimate the data on all subcarriers at once. In terms of performance, we have shown that our proposed receiver outperforms its competitors in challenging scenarios with a high phase-noise variance inducing a significant ICI power. A future perspective of this work is the extension of our system model to the massive MIMO setup by assuming hybrid analog/digital architectures, while taking into account a potential carrier frequency offset.

Appendix A
Time-to-Frequency Signal Development

Here, we show, step by step, how to derive equation (15) from (13), which is a key property to the tensor model in equation (16). The time-domain transmitted symbols are modelled as $s(t) = \sum_{f=0}^{F-1} S[f]e^{j2\pi\frac{f}{T_s}t}$, with $T_s = 1/d_{\mathrm{F}}$ being the symbol period and d_{F} the subcarrier spacing in Hertz. For notation simplicity, let us consider a scalar (single-antenna) version of the received signal, where $M_R = M_T = K = L_P = 1$ is assumed. The same derivation applies to the multiple-antenna case by introducing the transmit and receive antenna indices, without changing the developments. In order to simplify the notation, let us consider $c(t) = e^{j\phi^{[t]}(t)}s(t)$, where $\phi^{[t]}$ is the phase-noise term at the transmitter side. Assuming $t = (\frac{nT_s}{F} - \tau_l)$, the discrete-time received signal can be written as

$$
\begin{aligned}
y[n] &= \sum_{l=0}^{L-1} e^{j\phi^{[\mathrm{r}]}[n]} \bar{h}[l] c\left(\frac{nT_s}{F} - \tau_l\right) \\
&= \sum_{l=0}^{L-1} e^{j\phi^{[\mathrm{r}]}[n]} \bar{h}[l] \frac{1}{F} \sum_{p=0}^{F-1} C[p] e^{j2\pi \frac{p}{T_s}(\frac{nT_s}{F} - \tau_l)} \\
&= \frac{1}{F} \sum_{l=0}^{L-1} \sum_{p=0}^{F-1} e^{j\phi^{[\mathrm{r}]}[n]} \bar{h}[l] e^{-j2\pi p d_{\mathrm{F}} \tau_l} C[p] e^{j2\pi \frac{p}{n}/F} \\
&= \frac{1}{F} \sum_{p=0}^{F-1} e^{j\phi^{[\mathrm{r}]}[n]} H[p] C[p] e^{j2\pi \frac{p}{n}/F}. \qquad (46)
\end{aligned}
$$

Applying the DFT, we have

$$
\begin{aligned}
Y[f] &= \sum_{n=0}^{F-1} y[n] e^{-j2\pi f n/F} \qquad (47) \\
&= \sum_{n=0}^{F-1} \left(\frac{1}{F} \sum_{p=0}^{F-1} e^{j\phi^{[\mathrm{r}]}[n]} H[p] C[p] e^{j2\pi \frac{p}{n}/F} \right) e^{-j2\pi f n/F} \qquad (48) \\
&= \sum_{n=0}^{F-1} \sum_{p=0}^{F-1} \frac{1}{F} e^{j\phi^{[\mathrm{r}]}[n]} e^{j2\pi p n/F} e^{-j2\pi f n/F} H[p] \\
&\quad \sum_{q=0}^{F-1} \Phi^{[\mathrm{t}]}[p-q] S[q] \qquad (49) \\
&= \sum_{n=0}^{F-1} \sum_{p=0}^{F-1} \frac{1}{F} e^{j\phi^{[\mathrm{r}]}[n]} e^{-j2\pi (f-p) n/F} H[p] \\
&\quad \sum_{q=0}^{F-1} \Phi^{[\mathrm{t}]}[p-q] S[q] \qquad (50) \\
&= \sum_{p=0}^{F-1} \Phi^{[\mathrm{r}]}[f-p] H[p] \sum_{q=0}^{F-1} \Phi^{[\mathrm{t}]}[p-q] S_i[q] \qquad (51) \\
&= \Phi^{[\mathrm{r}]}[0] H[f] \Phi^{[\mathrm{t}]}[0] S[f] + \sum_{p=0, p\neq f}^{F-1} \Phi^{[\mathrm{r}]}[f-p] H[p] \\
&\quad \sum_{q=0, q\neq f}^{F-1} \Phi^{[t]}[p-q] S[q] \\
&= \Phi^{[\mathrm{r}]}[0] H[f] \Phi^{[\mathrm{t}]}[0] S[f] + G[f] \qquad (52)
\end{aligned}
$$

where $G[f]$ is the ICI term. To derive equation (49) from (48), we note that

$$
\begin{aligned}
C[p] &= \sum_{n=0}^{F-1} e^{j\phi^{[\mathrm{t}]}[n]} s[n] e^{-j2\pi p n/F} \\
&= \sum_{n=0}^{F-1} e^{j\phi^{[\mathrm{t}]}[n]} \frac{1}{F} \sum_{q=0}^{F-1} S[q] e^{-j2\pi (p-q) n/F} \\
&= \sum_{q=0}^{F-1} \sum_{n=0}^{F-1} \frac{1}{F} e^{j\phi^{[\mathrm{t}]}[n]} e^{-j2\pi (p-q) n/F} S[q] \\
&= \sum_{q=0}^{F-1} \Phi^{[\mathrm{t}]}[p-q] S[q].
\end{aligned}
$$

Applying the combiner $\boldsymbol{W} \in \mathbb{C}^{M_R \times M_R}$, and taking into account the AWGN tensor at the receiver $\boldsymbol{\mathcal{V}}$, the received signal at the k-th block and f-th subcarrier in matrix slice notation can be written as

$$
\begin{aligned}
\boldsymbol{\mathcal{Y}}_{..kf} &= \boldsymbol{W} D_k\left(\boldsymbol{\Phi}^{[\mathrm{r}]}[0]\right) \boldsymbol{\mathcal{H}}_{..f} D_k\left(\boldsymbol{\Phi}^{[\mathrm{t}]}[0]\right) \boldsymbol{\mathcal{S}}_{..f}^{\mathrm{T}} \\
&\quad + \boldsymbol{W} \boldsymbol{\mathcal{G}}_{..kf} + \boldsymbol{W} \boldsymbol{\mathcal{V}}_{..kf}, \qquad (53)
\end{aligned}
$$

where $\boldsymbol{\Phi}^{[\mathrm{x}]}[0] \in \mathbb{C}^{K \times M_X}$, for $\mathrm{x} \in \{\mathrm{r}, \mathrm{t}\}$, is given by

$$
\boldsymbol{\Phi}^{[\mathrm{x}]}[0] = \begin{bmatrix} \Phi_{1,1}^{(\mathrm{x})}[0] & \dots & \Phi_{1,M_X}^{(\mathrm{x})}[0] \\ \vdots & \ddots & \vdots \\ \Phi_{K,1}^{(\mathrm{x})}[0] & \dots & \Phi_{K,M_X}^{(\mathrm{x})}[0] \end{bmatrix} \qquad (54)
$$

Appendix B
Derivation of the ICI Power

As in many works [11], [12], the ICI can be approximated as a Gaussian random variable with zero mean and variance σ_{ICI}^2, i.e., $G[f] \sim \mathcal{CN}(0, \sigma_{\mathrm{ICI}}^2)$. Considering the ICI term in equation (52), we have $\sigma_{\mathrm{ICI}}^2 = \mathbb{E}[G[f]G[f]^*]$, which can be computed as

$$
\sigma_{\mathrm{ICI}}^2 = \mathbb{E}\left[\left|\sum_{p=0,p\neq f}^{F-1} \Phi^{[\mathrm{r}]}[f-p] H[p] \sum_{q=0,q\neq f}^{F-1} \Phi^{[\mathrm{t}]}[p-q] S[q]\right|^2\right].
$$

Assuming $\mathbb{E}[S[q]S[q]^*] = 1$ and $\mathbb{E}[H[p]H[p]^*] = 1$, we get

$$
\sigma_{\mathrm{ICI}}^2 = \mathbb{E}\left[\left|\sum_{p=0,p\neq f}^{F-1} \Phi^{[\mathrm{r}]}[f-p] \sum_{q=0,q\neq f}^{F-1} \Phi^{[\mathrm{t}]}[p-q]\right|^2\right]. \qquad (55)
$$

It is important to note that, when $q \neq p$, $\mathbb{E}[|\Phi^{[\mathrm{r}]}[f-p]\Phi^{[\mathrm{t}]}[p-q]|^2] \approx 0$, since the product between the non-DC terms (i.e., $p \neq f$ and $q \neq f$) of the phase-noise at the transmitter and the receiver is very small, even for $\sigma_{\phi^{[\mathrm{t}]}}^2 = \sigma_{\phi^{[\mathrm{r}]}}^2 = 5 \cdot 10^{-3}$. We can rewrite equation (55) as

$$
\begin{aligned}
\sigma_{\mathrm{ICI}}^2 &\approx \mathbb{E}\left[\left|\sum_{p=0,p\neq f}^{F-1} \Phi^{[\mathrm{r}]}[f-p] \Phi^{[\mathrm{t}]}[0]\right|^2\right] \qquad (56) \\
&\approx (1 - \mathbb{E}\left[\left|\Phi^{[\mathrm{r}]}[0]\right|^2\right]) \mathbb{E}\left[\left|\Phi^{[\mathrm{t}]}[0]\right|^2\right]. \qquad (57)
\end{aligned}
$$

Adopting the same approach as in [8], in our model the ICI power can be calculated as

$$
\sigma_{\mathrm{ICI}}^2 \approx \left(\mathrm{trace}\{\boldsymbol{R}_{\Phi\Phi}^{[\mathrm{r}]}\} - \boldsymbol{R}_{\Phi\Phi}^{[\mathrm{r}]}(0,0)\right) \boldsymbol{R}_{\Phi\Phi}^{[\mathrm{t}]}(0,0), \qquad (58)
$$

where $\boldsymbol{R}_{\Phi\Phi}^{[\mathrm{x}]} \in \mathbb{C}^{F \times F}$, $\mathrm{x} \in \{\mathrm{r}, \mathrm{t}\}$, is given by

$$\begin{aligned} \boldsymbol{R}_{\Phi\Phi}^{[\mathrm{x}]}(i,m) &= E\left[\Phi^{[\mathrm{x}]}[i]\Phi^{[\mathrm{x}]*}[m]\right] \\ &= \frac{1}{F^2}E\left[\sum_{k=0}^{F-1}\sum_{l=0}^{F-1} e^{j(\phi^{[\mathrm{x}]}[k]-\phi^{[\mathrm{x}]}[l])}e^{-j2\pi(ik-ml)/F}\right] \\ &= \frac{1}{F^2}\sum_{k=0}^{F-1}\sum_{l=0}^{F-1} E\left[e^{j(\Delta\phi^{[\mathrm{x}]}[k,l])}\right]e^{-j2\pi(ik-ml)/F}, \end{aligned} \tag{59}$$

with $i, m = \{0 \ldots F-1\}$. Note that $\mathrm{trace}(\boldsymbol{R}_{\Phi\Phi}^{[\mathrm{x}]}) = 1$.

References

[1] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM Wireless Communications With MATLAB*. Hoboken, NJ, USA: Wiley, 2010.

[2] T.-J. Lee and Y.-C. Ko, "Channel estimation and data detection in the presence of phase noise in MIMO-OFDM systems with independent oscillators," *IEEE Access*, vol. 5, pp. 9647–9662, 2017.

[3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[4] K. Y. Kim, Q. Zou, H. J. Choi, and A. H. Sayed, "An efficient carrier phase synchronization technique for high-order M-QAM-OFDM," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3789–3794, Aug. 2008.

[5] I. Ngebani, I. Zibani, E. Matlotse, K. Tsamaase, and J. Chuma, "Joint channel and phase noise estimation in MIMO-OFDM systems," in *Proc. IEEE Radio Antenna Days Indian Ocean*, 2016, pp. 1–2.

[6] A. Ishaque and G. Ascheid, "On the efficient mitigation of phase noise in MIMO-OFDM receivers," in *Proc. Int. Symp. Signals, Syst., Electron.*, 2012, pp. 1–6.

[7] N. Hadaschik *et al.*, "Improving MIMO phase noise estimation by exploiting spatial correlations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, pp. iii/833–iii/836.

[8] D. Petrovic, W. Rave, and G. Fettweis, "Effects of phase noise on OFDM systems with and without PLL: Characterization and compensation," *IEEE Trans. Commun.*, vol. 55, no. 8, pp. 1607–1616, Aug. 2007.

[9] Q. Zou, A. Tarighat, and A. H. Sayed, "Compensation of phase noise in OFDM wireless systems," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5407–5424, Nov. 2007.

[10] H. Mehrpouyan, A. A. Nasir, S. D. Blostein, T. Eriksson, G. K. Karagiannidis, and T. Svensson, "Joint estimation of channel and oscillator phase noise in MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4790–4807, Sep. 2012.

[11] Y.-P. Zhang, J. Liu, S. Feng, and P. Zhang, "Pilot design for phase noise mitigation in millimeter wave MIMO-OFDM systems," in *Proc. IEEE 85th Veh. Technol. Conf.*, 2017, pp. 1–6.

[12] H. Huang, W. G. Wang, and J. He, "Phase noise and frequency offset compensation in high frequency MIMO-OFDM system," in *Proc. IEEE Int. Conf. Commun.*, 2015, pp. 1280–1285.

[13] A. L. F. de Almeida, G. Favier, and J. C. M. Mota, "A constrained factor decomposition with application to MIMO antenna systems," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2429–2442, Jun. 2008.

[14] G. Favier, M. N. da Costa, A. L. F. de Almeida, and J. M. T. Romano, "Tensor space-time (TST) coding for MIMO wireless communication systems," *Signal Process.*, vol. 92, no. 4, pp. 1079–1092, 2012.

[15] A. L. F. de Almeida and G. Favier, "Double Khatri-Rao space-time-frequency coding using semi-blind PARAFAC based receiver," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 471–474, May 2013.

[16] K. Liu, J. P. C. da Costa, H. So, and A. L. F. de Almeida, "Semi-blind receivers for joint symbol and channel estimation in space-time-frequency MIMO-OFDM systems," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5444–5457, Nov. 2013.

[17] A. L. F. de Almeida, G. Favier, and L. R. Ximenes, "Space-time-frequency (STF) MIMO communication systems with blind receiver based on a generalized PARATUCK2 model," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 1895–1909, Apr. 2013.

[18] F. Roemer and M. Haardt, "Tensor-based channel estimation and iterative refinements for two-way relaying with multiple antennas and spatial reuse," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5720–5735, Nov. 2010.

[19] C. A. R. Fernandes, A. L. F. de Almeida, and D. B. da Costa, "Unified tensor modeling for blind receivers in multiuser uplink cooperative systems," *IEEE Signal Process. Lett.*, vol. 19, no. 5, pp. 247–250, May 2012.

[20] Y. Rong, M. R. A. Khandaker, and Y. Xiang, "Channel estimation of dual-hop MIMO relay system via parallel factor analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2224–2233, Jun. 2012.

[21] A. L. F. de Almeida, C. A. Fernandes, and D. B. da Costa, "Multiuser detection for uplink DS-CDMA amplify-and-forward relaying systems," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 697–700, Jul. 2013.

[22] L. R. Ximenes, G. Favier, and A. L. F. de Almeida, "Semi-blind receivers for non-regenerative cooperative MIMO communications based on nested PARAFAC modeling," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4985–4998, Sep. 2015.

[23] G. Favier, C. A. R. Fernandes, and A. L. F. de Almeida, "Nested tucker tensor decomposition with application to MIMO relay systems using tensor space-time coding (TSTC)," *Signal Process.*, vol. 128, pp. 318–331, 2016.

[24] B. Sokal, A. L. de Almeida, and M. Haardt, "Semi-blind receivers for MIMO multi-relaying systems via rank-one tensor approximations," *Signal Process.*, vol. 166, 2020, Art. no. 107254.

[25] K. Naskovska, M. Haardt, and A. L. de Almeida, "Generalized tensor contraction with application to Khatri-Rao coded MIMO OFDM systems," in *Proc. IEEE 7th Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, 2017, pp. 1–5.

[26] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1524–1538, Jul. 2017.

[27] K. Naskovska, M. Haardt, and A. L. F. de Almeida, "Generalized tensor contractions for an improved receiver design in MIMO-OFDM systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 3186–3190.

[28] D. C. Araújo, A. L. F. De Almeida, J. P. Da Costa, and R. T. de Sousa, "Tensor-based channel estimation for massive MIMO-OFDM systems," *IEEE Access*, vol. 7, pp. 42133–42147, 2019.

[29] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Tensor-based channel estimation for millimeter wave MIMO-OFDM with dual-wideband effects," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4218–4232, Jul. 2020.

[30] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[31] R. A. Harshman *et al.*, "Foundations of the parafac procedure: Models and conditions for an "explanatory" multimodal factor analysis," UCLA Working Papers in Phonetics, vol. 16, pp. 1–84, 1970.

[32] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[33] B. Sokal, M. Haardt, and A. L. F. de Almeida, "Semi-blind receiver for two-hop MIMO relaying systems via selective Kronecker product modeling," in *Proc. IEEE 8th Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, 2019, pp. 714–718.

[34] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[35] G. Favier, M. N. Da Costa, A. L. F. De Almeida, and J. M. T. Romano, "Tensor space-time (TST) coding for MIMO wireless communication systems," *Signal Process.*, vol. 92, no. 4, pp. 1079–1092, 2012.

[36] I. V. Oseledets, "Tensor-train decomposition," *SIAM J. Sci. Comput.*, vol. 33, no. 5, pp. 2295–2317, 2011.

[37] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, no. 2, pp. 95–138, 1977.

[38] A. Stegeman and N. D. Sidiropoulos, "On kruskal's uniqueness condition for the candecomp/parafac decomposition," *Linear Algebra Appl.*, vol. 420, no. 2, pp. 540–552, 2007.

[39] K. Naskovska, "Advanced tensor based signal processing techniques for wireless communication systems and biomedical signal processing," Ph.D. dissertation, Technische Universitat Ilmenau, Germany, 2019.

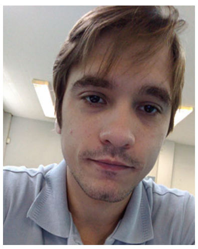
[40] X. Chen, H. Wang, W. Fan, Y. Zou, A. Wolfgang, T. Svensson, and J. Luo *et al.*, "Phase noise effect on MIMO-OFDM systems with common and independent oscillators," *Wireless Commun. Mobile Comput.*, vol. 2017, pp. 1–12, 2017.

[41] K. Wang, L. Jalloul, and A. Gomaa, "Phase noise compensation using limited reference symbols in 3GPP LTE downlink," 2017, *arXiv:1711.10064*.

[42] W. Sun, L. Huang, H.-C. So, and J. Wang, "Orthogonal tubal rank-1 tensor pursuit for tensor completion," *Signal Process.*, vol. 157, pp. 213–224, 2019.

[43] J. Ahlberg, J. Walsh, R. Bellman, and E. N. Nilson, *The Theory of Splines and Their Applications*. New York, NY, USA: Academic Press, 1967.

[44] F. Roemer, "Advanced algebraic concepts for efficient multi-channel signal processing," Ph.D. dissertation, Ilmenau Univ. Technol., Germany, 2013.



Bruno Sokal (Student Member, IEEE) received the B.Sc. degree in telecommunications engineering from the Federal Institution of Science, Education and Technology of Ceará, Fortaleza, Brazil, in 2015 and the M.Sc. degree in teleinformatics engineering in 2017 from the Federal University of Ceará (UFC), Fortaleza, Brazil, where he is currently working toward the D.Sc. degree in teleinformatics engineering with the emphasis in signal processing for communications. From 2017 to 2019, he was a Researcher with the Wireless Telecom Research Group (GTEL), UFC, where he worked on the Project Remote area Access Network for the 5th GEneration (5G RANGE). His research focuses on tensor signal processing applied to wireless networks and future generation mobile broadband networks.

Paulo R. B. Gomes received the Diploma degree in physics (with Magna cum Laude) from the Federal University of Ceará (UFC), Fortaleza, Brazil, in 2011, the Diploma degree in telecommunications engineering from the Federal Institute of Education, Science and Technology of Ceará, Fortaleza, Brazil, in 2014, and the M.Sc. and Ph.D. degrees in teleinformatics engineering from UFC in 2014 and 2018, respectively. He is currently a Postdoctoral Researcher with the Wireless Telecommunications Research Group Laboratory. His research focuses on tensor signal processing, with applications to sensor arrays and wiereless communication systems.

André L. F. de Almeida (Senior Member, IEEE) received the double Ph.D. degrees in sciences and teleinformatics engineering from the University of Nice Sophia Antipolis, Nice, France, and the Federal University of Ceará, Fortaleza, Brazil, in 2007. He is currently an Associate Professor with the Department of Teleinformatics Engineering, Federal University of Ceará. From 2007 to 2008, he was a Teacher with the University of Nice Sophia Antipolis. His research interest mainly include multilinear algebra and tensor decompositions with applications to wireless communication systems and signal processing. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2012 to 2016 and the IEEE SIGNAL PROCESSING LETTERS from 2016 to 2020. He is currently a Senior Area Editor of the IEEE SIGNAL PROCESSING LETTERS and an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is an Elected Member of the Sensor Array and Multichannel, Technical Committee of the IEEE Signal Processing Society (SPS), and an Elected Member of the EURASIP Signal Processing for Multi-Sensor Systems Technical Area Committee. He was with organization and chairing several IEEE SPS conferences, which include CAMSAP-2017, GlobalSIP-2018 and 2019, SAM 2020, and CAMSAP 2021. Since January 2018, he has been a Research Fellow of the CNPq (the Brazilian National Council for Scientific and Technological Development) and an Affiliate Member of the Brazilian Academy of Sciences, Rio de Janeiro, Brazil. He was the recipient of the multiple times Visiting Professor positions at the University of Nice Sophia-Antipolis, France, in 2012, 2013, 2015, 2018, and 2019.

Martin Haardt (Fellow, IEEE) received the Electrical Engineering degree from Ruhr-University Bochum, Bochum,

Germany, and Purdue University, West Lafayette, IN, USA, the M.S. (Diplom-Ingenieur) degree from Ruhr-University Bochum in 1991, and the Ph.D. (Doktor-Ingenieur) degree from the Munich University of Technology, Munich, Germany, in 1996. Since 2001, he has been a Full Professor with the Department of Electrical Engineering and Information Technology and the Head of the Communications Research Laboratory, Ilmenau University of Technology, Ilmenau, Germany. In 1997, he was with Siemens Mobile Networks, Munich, Germany, where he was responsible for strategic research for third generation mobile radio systems. From 1998 to 2001, he was the Director for International Projects and University Cooperations in the mobile infrastructure business of Siemens, Munich, Germany, where his work was focused on mobile communications beyond the third generation. His research interests include wireless communications, array signal processing, high-resolution parameter estimation, and tensor-based signal processing. Since 2019, he has been a Senior Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2002 to 2006 and 2011 to 2015, the IEEE SIGNAL PROCESSING LETTERS from 2006 to 2010, the *Research Letters in Signal Processing* from 2007 to 2009, the *Hindawi Journal of Electrical and Computer Engineering* since 2009, the *EURASIP Signal Processing Journal* from 2011 to 2014, and the Guest Editor of the EURASIP *Journal on Wireless Communications and Networking*. From 2011 to 2019, he was an Elected Member of the Sensor Array and Multichannel technical committee of the IEEE Signal Processing Society, where he was the Vice Chair from 2015 to 2016, the Chair from 2017 to 2018, and the Past Chair in 2019. Since 2020, he has been an Elected Member of the Signal Processing Theory and Methods technical committee of the IEEE Signal Processing Society. He was the Technical Co-Chair of the PIMRC 2005 in Berlin, Germany, ISWCS 2010 in York, U.K., the European Wireless 2014 in Barcelona, Spain, and the Asilomar Conference on Signals, Systems, and Computers 2018, USA, and the General Co-Chair of WSA 2013 in Stuttgart, Germany, ISWCS 2013 in Ilmenau, Germany, CAMSAP 2013 in Saint Martin, French Antilles, WSA 2015 in Ilmenau, SAM 2016 in Rio de Janeiro, Brazil, CAMSAP 2017 in Curacao, Dutch Antilles, SAM 2020 in Hangzhou, China, and the Asilomar Conference on Signals, Systems, and Computers 2021, USA. He was the recipient of the 2009 Best Paper Award from the IEEE Signal Processing Society, the Vodafone for outstanding research in mobile communications, the ITG Best Paper Award from the Association of Electrical Engineering, Electronics, and Information Technology (VDE), and the Rohde & Schwarz Outstanding Dissertation Award.