TENSOR-BASED ESTIMATION OF MMWAVE MIMO CHANNELS WITH CARRIER FREQUENCY OFFSET

Lucas N. Ribeiro, André L. F. de Almeida[*]

Universidade Federal do Ceará
Wireless Telecom Research Group (GTEL)
Emails: {nogueira, andre}@gtel.ufc.br

Nitin J. Myers, Robert W. Heath Jr.[†]

The University of Texas at Austin
Dept. of Electrical and Computer Engineering
Emails: {nitinjmyers, rheath}@utexas.edu

ABSTRACT

Millimeter wave multiple-input-multiple-output (MIMO) achieves the best performance when reliable channel state information is used to design the beams. Most channel estimation methods proposed in the literature, however, ignore practical hardware impairments such as carrier frequency offset (CFO) and may fail under such impairment. In this paper, we present a joint CFO and channel estimation method based on tensor modeling and compressed sensing. Simulation results indicate that the proposed method yields better channel recovery performance than the benchmark and that it is more robust to a small number of channel measurements.

Index Terms— MmWave, MIMO, tensor, compressed sensing, channel estimation

1. INTRODUCTION

Millimeter wave (mmWave) multiple-input-multiple-output (MIMO) is a fundamental technology for future mobile communications systems. It is expected to enable data transfer in the order of gigabits per second, therefore significantly increasing system capacity. Although beam training already enables satisfactory data transfer performance [1, 2], considering the full channel state information (CSI) could significantly enhance spectral efficiency [3]. In this sense, the channel estimation problem becomes critical.

Some proposed mmWave channel estimation methods take CFO into account. In [4], a CFO-aware iterative channel estimation method based on the least squares (LS) solution is presented for a single-input-single-output mmWave system based on single-carrier frequency-domain equalization. In [5], a beam alignment technique which considers the effects of CFO is presented. This method relies on beam directions hashing, which allows to easily track energy changes across directions and find the correct alignment. A channel estimation technique robust to both CFO and phase noise is introduced in [6]. Therein, the received signal is modeled as a tensor and the channel parameters are estimated by a customized orthogonal matching pursuit (OMP) algorithm. However, the methods presented in [5, 6] are applicable only to analog filtering-based systems. To fill this gap out, the work of [7] presented a joint CFO and channel estimation method based on maximum likelihood and OMP for a hybrid analog/digital (A/D) system.

In this paper, we solve the CFO-distorted channel estimation problem for a hybrid A/D point-to-point massive MIMO system. We assume a time-variant small-scale fading gain, unlike the channel models in [6, 7]. The evolution of the phase errors during some time slots allows us to model the antenna measurements at the receiver as a multidimensional (tensor) signal. With such a model, we propose a tensor-based joint CFO and channel estimation technique. An appealing property of tensor methods is the relaxed conditions on tensor decomposition uniqueness compared to matrices. This property, determined by Kruskal's Theorem [8] for the canonical polyadic (CP) decomposition, has been shown to bring performance improvements, including reduction of training samples [9] and improved parameter estimation accuracy [10]. Simulation results indicate that the presented tensor-based method offers better channel estimation performance and exhibits robustness to CFO for a relatively small number of channel measurements.

The following notation is adopted throughout the paper: Lowercase bold-face letters $\boldsymbol{x}$ denotes vectors, upper-case letters matrices, e.g. $\boldsymbol{X}$, uppercase calligraphic letters tensors, e.g. $\mathcal{X}$. The (i,j)th entry of $\boldsymbol{X}$ is written as $[\boldsymbol{X}]_{i,j}$. The transpose and the conjugate transpose (Hermitian) of $\boldsymbol{X}$ are denoted by $\boldsymbol{X}^{\mathsf{T}}$, $\boldsymbol{X}^{\mathsf{H}}$, respectively. The M-dimensional identity matrix is represented by $\boldsymbol{I}_M$ and the M-dimensional null matrix by $\boldsymbol{0}_M$. $\mathcal{I}_{L\times L\times L}$ denotes the L-dimensional third-order identity tensor [11]. The matrix operator $\mathrm{Diag}(\cdot)$ transforms an input vector into a diagonal matrix. The Moore-Penrose pseudo-inverse is denoted by $(\cdot)^{\dagger}$. The Kronecker product, the Khatri-Rao product and the n-mode product are referred to as $\otimes$, $\odot$ and $\times_n$, respectively. The Euclidean norm, the statistical expected value operator and the determinant are respectively denoted as $\|\cdot\|_2$, $\mathbb{E}[\cdot]$ and $\det(\cdot)$. Complex Gaussian distribution of independent random variables with null mean and covariance matrix σ^2 is denoted as $\mathcal{CN}(\boldsymbol{0}, \sigma\boldsymbol{I})$. For more details on tensor notation, we refer the reader to [11, 12].

2. SYSTEM MODEL

Let us consider a narrowband point-to-point mmWave hybrid A/D MIMO system model. We consider the narrowband assumption for simplicity purpose although it is not practical for mmWave systems. Yet, we believe the proposed tensor approach will make it possible to consider frequency selective channels in future work. The transmitter (TX) employs a uniform linear array (ULA) of N_t antennas connected to L_t RF chains, while the receiver (RX) captures the transmitted signals with a ULA of N_r antennas connected to L_r RF chains. We consider simultaneous transmission of N_s data streams to the RX. To this end, the TX applies a precoding filter $\bar{\boldsymbol{F}} \in \mathbb{C}^{N_t\times N_s}$ and the RX a combining filter $\bar{\boldsymbol{W}} \in \mathbb{C}^{N_r\times N_s}$. The hybrid A/D transceiver architecture splits the precoding filter into a RF analog precoder $\boldsymbol{F}_{\mathrm{RF}} \in \mathbb{C}^{N_t\times L_t}$ and a baseband digital precoder $\boldsymbol{F}_{\mathrm{BB}} \in \mathbb{C}^{L_t\times N_s}$ such that $\bar{\boldsymbol{F}} = \boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}$. Likewise, the combining filter is implemented as the cascade of $\boldsymbol{W}_{\mathrm{RF}} \in \mathbb{C}^{N_r\times L_r}$ and $\boldsymbol{W}_{\mathrm{BB}} \in \mathbb{C}^{L_r\times N_s}$ with $\bar{\boldsymbol{W}} = \boldsymbol{W}_{\mathrm{RF}}\boldsymbol{W}_{\mathrm{BB}}$. We consider that the analog filters are implemented by a phase-shifting network, hence the

[*]This work is partially supported by the Brazilian National Council for Scientific and Technological Development CNPq, CAPES/PROBRAL Proc. no. 88887.144009/2017-00, and FUNCAP.

[†]This material is based upon work supported in part by the National Science Foundation under Grant No. CNS-1702800 and ECCS-1711702

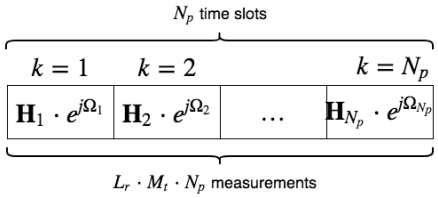

Fig. 1. At each kth time slot, the TX employs M_t precoders and the RX makes L_r measurements. The channel matrix and CFO are assumed to be constant within a time slot.

elements of the corresponding matrices are constrained to have constant modulus [13, 14].

We consider the narrowband clustered model contributing with L propagation paths [15] to model mmWave channels. We assume that the geometrical information (angles of arrival and departure) remain constant over N_p time slots, while the small-scale fading changes for every kth time slot, $k = 1, \ldots, N_p$. Therefore, the $(N_r \times N_t)$-dimensional channel matrix at the kth time slot can be written as

$$\boldsymbol{H}_k = \sqrt{\tfrac{N_t N_r}{L}} \sum_{\ell=1}^{L} \alpha_{k,\ell} \boldsymbol{a}_{\mathrm{r}}(\theta_\ell^{\mathrm{r}}) \boldsymbol{a}_{\mathrm{t}}(\theta_\ell^{\mathrm{t}})^{\mathsf{H}} = \boldsymbol{A}_{\mathrm{r}} \boldsymbol{\Gamma}_k \boldsymbol{A}_{\mathrm{t}}^{\mathsf{H}}, \quad (1)$$

where $\alpha_{k,\ell}$ denotes the small-scale fading complex gain relative to path ℓ, $\boldsymbol{a}_{\mathrm{r}}(\theta_\ell^{\mathrm{r}}) \in \mathbb{C}^{N_r}$ and $\boldsymbol{a}_{\mathrm{t}}(\theta_\ell^{\mathrm{t}}) \in \mathbb{C}^{N_t}$ the steering vectors evaluated at elevation θ_ℓ^{r} (θ_ℓ^{t}) arrival (departure) angle at the receiver (transmitter), respectively. The receive and transmit array manifold matrices are written as $\boldsymbol{A}_{\mathrm{r}} = [\boldsymbol{a}_{\mathrm{r}}(\theta_1^{\mathrm{r}}), \ldots, \boldsymbol{a}_{\mathrm{r}}(\theta_L^{\mathrm{r}})]$ and $\boldsymbol{A}_{\mathrm{t}} = [\boldsymbol{a}_{\mathrm{t}}(\theta_1^{\mathrm{t}}), \ldots, \boldsymbol{a}_{\mathrm{t}}(\theta_L^{\mathrm{t}})]$, respectively. The diagonal matrix $\boldsymbol{\Gamma}_k = \sqrt{\frac{N_t N_r}{L}} \mathrm{Diag}(\alpha_{k,1}, \ldots, \alpha_{k,L})$ bears the small-scale fading complex gains. We assume that each path has the same average power, i.e., $\alpha_{k,\ell}$ is modeled as a circular symmetric Gaussian random variable with zero mean and unit variance. In practice, the narrowband assumption is quite limiting and simplifies the CFO problem. In spite of that, we consider this assumption to validate the applicability of our tensor formulation. More practical scenarios, including wideband communications, are envisaged as future work.

We assume that the angles of arrival and departure are on a uniform grid with G_r and G_t points. Without loss of generality, we consider $G_r = N_r$ and $G_t = N_t$. Under these assumptions, the channel matrix $\boldsymbol{H}_k$ can be approximated as $\boldsymbol{H}_k \approx \boldsymbol{\Psi}_{\mathrm{r}} \tilde{\boldsymbol{\Gamma}}_k \boldsymbol{\Psi}_{\mathrm{t}}^{\mathsf{H}}$, where $\boldsymbol{\Psi}_{\mathrm{r}}$ and $\boldsymbol{\Psi}_{\mathrm{t}}$ stand for N_r- and N_t-dimensional discrete Fourier matrices [9], and $\tilde{\boldsymbol{\Gamma}}_k \in \mathbb{C}^{N_r \times N_t}$ is a L-sparse matrix whose non-zero elements correspond to the angles of arrival and departure pairs. To deal with out-of-grid angles, one could, for example, increase the number of grid points.

Hardware imperfections such as carrier frequency offset (CFO) may distort the transmit/receive signals, and, thus, standard channel estimation methods may not be able to deal with these frequency offsets and fail to properly estimate the channel's parameters. To tackle this issue, we consider a CFO term in our signal model, which is presented in the following. For modeling simplicity reasons, we consider CFO only at the receive side. Also, we assume that a single oscillator drives all the RF chains at a given end. As a result, there is a unique CFO in the system.

3. CHANNEL ESTIMATION

Let us consider that the downlink channel estimation (from TX to RX) occurs during N_p consecutive time slots, as illustrated in Figure 1. At each slot $k \in \{1, \ldots, N_p\}$, the TX employs a sequence of M_t precoding vectors (each corresponding to a measurement) in successive time instants. The RX applies simultaneously its L_r combiners, each one corresponding to a RF chain, to measure the received signals. To simplify the analysis, we assume the TX sends the same symbol $s = 1$ in all transmission, as in [16]. In the kth time slot, the channel measurement acquired by the receiver for the pth combiner $\boldsymbol{w}_p \in \mathbb{C}^{N_r}$ and the qth precoder $\boldsymbol{f}_q \in \mathbb{C}^{N_t}$ is

$$y_{p,q,k} = \boldsymbol{w}_p^{\mathsf{H}} \boldsymbol{H}_k \boldsymbol{f}_q e^{j\Omega_k} + \boldsymbol{w}_p^{\mathsf{H}} \boldsymbol{b}_{q,k}, \quad k \in \{1, \ldots, N_p\}, \quad (2)$$

where $\Omega_k = 2\pi \frac{f_{\mathrm{CFO}}}{F_s} k$ denotes the error phase term due to CFO, F_s the sampling frequency, and $\boldsymbol{b}_{q,k} \in \mathbb{C}^{N_r}$ the additive white Gaussian vector noise term, with $\boldsymbol{b}_{q,k} \sim \mathcal{CN}(\boldsymbol{0}_{N_r}, \sigma_n^2 \boldsymbol{I}_{N_r})$. After M_t transmissions, the RX accumulates $M_t L_r$ measurements into a matrix:

$$\boldsymbol{Y}_k = \boldsymbol{W}^{\mathsf{H}} \boldsymbol{C}_k \boldsymbol{F} + \boldsymbol{W}^{\mathsf{H}} \boldsymbol{B}_k \in \mathbb{C}^{L_r \times M_t}, \quad k \in \{1, \ldots, N_p\}, \quad (3)$$

where $\boldsymbol{C}_k = \boldsymbol{H}_k e^{j\Omega_k}$ denotes the effective channel matrix, $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{L_r}] \in \mathbb{C}^{N_r \times L_r}$ the combining matrix, $\boldsymbol{F} = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_{M_t}] \in \mathbb{C}^{N_t \times M_t}$ the precoding matrix, and $\boldsymbol{B}_k = [\boldsymbol{b}_{1,k}, \ldots, \boldsymbol{b}_{M_t,k}] \in \mathbb{C}^{N_r \times M_t}$ the additive noise matrix. Now, stacking (3) for all N_p time slots into $\bar{\boldsymbol{Y}} = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{N_p}] \in \mathbb{C}^{L_r \times M_t N_p}$, we have

$$\bar{\boldsymbol{Y}} = \boldsymbol{W}^{\mathsf{H}} \bar{\boldsymbol{C}} (\boldsymbol{I}_{N_p} \otimes \boldsymbol{F}) + \boldsymbol{W}^{\mathsf{H}} \bar{\boldsymbol{B}} \quad (4)$$

$$\bar{\boldsymbol{C}} = [\boldsymbol{C}_1, \ldots, \boldsymbol{C}_{N_p}], \; \bar{\boldsymbol{B}} = [\boldsymbol{B}_1, \ldots, \boldsymbol{B}_{N_p}] \in \mathbb{C}^{N_r \times M_t N_p}$$

Due to the structure of $\bar{\boldsymbol{C}}$, which comes from the fading, one cannot directly apply compressed sensing based approaches to (1). To exploit this structure, we identify (4) as a tensor unfolding, and then apply tensor techniques to estimate the channel considering CFO. It is important to note that model (2) implies the CFO at the RX remains approximately constant while the TX switches between the M_t precoders. We assume that the phase errors induced by CFO are invariant in a given time slot. Modeling phase errors across slots is more important as they increase at a rate that is M_t times higher than the in-slot phase errors. The assumption allows us to develop a tractable tensor formulation and is validated through simulations.

To see that (4) is a tensor unfolding, let us first concatenate (3) into a third-order tensor $\mathcal{Y} \in \mathbb{C}^{L_r \times M_t \times N_p}$. It means that, instead of stacking the $\boldsymbol{Y}_k$ matrices into a wide matrix, we stack them into a rectangular cuboid, such that $[\mathcal{Y}]_{:,:,k} = \boldsymbol{Y}_k$, $k \in \{1, \ldots, N_p\}$. Likewise, we can concatenate the effective channels into a tensor $\mathcal{C} \in \mathbb{C}^{N_r \times N_t \times N_p}$, such that $[\mathcal{C}]_{:,:,k} = \boldsymbol{C}_k$, $k \in \{1, \ldots, N_p\}$. Thus, using n-mode tensor products [11], it can be shown that $\mathcal{Y}$ and $\mathcal{C}$ are written as

$$\mathcal{Y} = \mathcal{C} \times_1 \boldsymbol{W}^{\mathsf{H}} \times_2 \boldsymbol{F}^{\mathsf{T}} + \mathcal{B}, \quad (5)$$

$$\mathcal{C} = \mathcal{I}_{L \times L \times L} \times_1 \boldsymbol{A}_{\mathrm{r}} \times_2 \boldsymbol{A}_{\mathrm{t}}^{*} \times_3 \boldsymbol{\Gamma}, \quad (6)$$

where $\mathcal{B} = \bar{\mathcal{B}} \times_1 \boldsymbol{W}^{\mathsf{H}}$ is the filtered additive noise tensor, $\bar{\mathcal{B}}$ is the tensor obtained by reshaping $\bar{\boldsymbol{B}}$, and $\boldsymbol{\Gamma} \in \mathbb{C}^{N_p \times L}$ denotes the effective small-scale fading gain matrix, defined as $[\boldsymbol{\Gamma}]_{k,\ell} = \sqrt{\frac{N_t N_r}{L}} e^{j\Omega_k} \alpha_{k,\ell}$. We identify (5) and (6) as Tucker and rank-L CP models, respectively. In fact, the 1-mode unfolding of (5) is exactly (4).

In the following, we describe a method to estimate the channel parameters $\boldsymbol{A}_{\mathrm{r}}$, $\boldsymbol{A}_{\mathrm{t}}$ and $\boldsymbol{\Gamma}$ from the received signal tensor (5) over N_p time slots. First, we need to rewrite (5) in a convenient manner. In this direction, let us replace (6) into (5) to get the following CP model for the measured signal tensor $\mathcal{Y}$:

$$\mathcal{Y} = \mathcal{I}_{L \times L \times L} \times_1 \boldsymbol{W}^{\mathsf{H}} \boldsymbol{A}_{\mathrm{r}} \times_2 \boldsymbol{F}^{\mathsf{T}} \boldsymbol{A}_{\mathrm{t}}^{*} \times_3 \boldsymbol{\Gamma} + \mathcal{B}. \quad (7)$$

Due to the sparse spatial response, the array manifold matrices can be expanded into their sparse representation [17]: $\boldsymbol{A}_{\mathrm{r}} = \boldsymbol{\Psi}_{\mathrm{r}}\boldsymbol{S}_{\mathrm{r}}$ and $\boldsymbol{A}_{\mathrm{t}} = \boldsymbol{\Psi}_{\mathrm{t}}\boldsymbol{S}_{\mathrm{t}}$, where $\boldsymbol{S}_{\mathrm{r}} \in \mathbb{C}^{G_r \times L}$ and $\boldsymbol{S}_{\mathrm{t}} \in \mathbb{C}^{G_t \times L}$ denote sparse coefficient matrices [9]. Hence, (7) can be rewritten as a rank-L CP model with factor matrices $\boldsymbol{Q}_{(1)} \in \mathbb{C}^{L_r \times L}$, $\boldsymbol{Q}_{(2)} \in \mathbb{C}^{M_t \times L}$ and $\boldsymbol{Q}_{(3)} \in \mathbb{C}^{N_p \times L}$:

$$\mathcal{Y} = \mathcal{I}_{L\times L\times L} \times_1 \underbrace{\boldsymbol{W}^{\mathsf{H}}\boldsymbol{\Psi}_{\mathrm{r}}\boldsymbol{S}_{\mathrm{r}}}_{\boldsymbol{Q}_{(1)}} \times_2 \underbrace{\boldsymbol{F}^{\mathsf{T}}\boldsymbol{\Psi}_{\mathrm{t}}\boldsymbol{S}_{\mathrm{t}}}_{\boldsymbol{Q}_{(2)}} \times_3 \underbrace{\boldsymbol{\Gamma}}_{\boldsymbol{Q}_{(3)}} + \mathcal{B}. \quad (8)$$

For later use, let us write the 3-mode unfolding of $\mathcal{Y}$ as follows [11]:

$$[\mathcal{Y}]_{(3)} = \boldsymbol{\Gamma}\left[(\boldsymbol{W}^{\mathsf{H}}\boldsymbol{A}_{\mathrm{r}}) \odot (\boldsymbol{F}^{\mathsf{T}}\boldsymbol{A}_{\mathrm{t}}^{*})\right]^{\mathsf{T}} + [\mathcal{B}]_{(3)} \in \mathbb{C}^{N_p \times L_r M_t}, \quad (9)$$

where $[\mathcal{B}]_{(3)} \in \mathbb{C}^{N_p \times L_r M_t}$ is the 3-mode unfolding of $\mathcal{B}$.

To reconstruct the channel coefficients $\mathcal{C}$ from $\mathcal{Y}$, one needs first to estimate the sparse coefficient matrices $\boldsymbol{S}_{\mathrm{r}}$, $\boldsymbol{S}_{\mathrm{t}}$, rebuild the array manifold matrices $\boldsymbol{A}_{\mathrm{r}}$, $\boldsymbol{A}_{\mathrm{t}}$, and then estimate the effective fading matrix $\boldsymbol{\Gamma}$. Let $\hat{\boldsymbol{S}}_{\mathrm{r}}$, $\hat{\boldsymbol{S}}_{\mathrm{t}}$ and $\hat{\boldsymbol{\Gamma}}$ denote the estimated matrices. Then, we rebuild the array manifold matrices as $\hat{\boldsymbol{A}}_{\mathrm{r}} = \boldsymbol{\Psi}_{\mathrm{r}}\hat{\boldsymbol{S}}_{\mathrm{r}}$ and $\hat{\boldsymbol{A}}_{\mathrm{t}} = \boldsymbol{\Psi}_{\mathrm{t}}\hat{\boldsymbol{S}}_{\mathrm{t}}$ and the effective fading matrix can be estimated by solving a LS problem, as will be shown in Section 3.2. The channel coefficients can thus be estimated as

$$\hat{\mathcal{C}} = \mathcal{I}_{L\times L\times L} \times_1 \underbrace{\boldsymbol{\Psi}_{\mathrm{r}}\hat{\boldsymbol{S}}_{\mathrm{r}}}_{\hat{\boldsymbol{A}}_{\mathrm{r}}} \times_2 \underbrace{\boldsymbol{\Psi}_{\mathrm{t}}\hat{\boldsymbol{S}}_{\mathrm{t}}}_{\hat{\boldsymbol{A}}_{\mathrm{t}}} \times_3 \hat{\boldsymbol{\Gamma}}. \quad (10)$$

Under some conditions one can uniquely factorize $\mathcal{Y}$ into (8) and estimate its factor matrices $\boldsymbol{Q}_{(1)}$, $\boldsymbol{Q}_{(2)}$ and $\boldsymbol{Q}_{(3)}$ up to trivial permutation and scale ambiguities [12]. In the next section, we discuss these conditions in more detail.

3.1. Uniqueness

The uniqueness of the CP model (8) is related to Kruskal's Theorem [8]. In [18], this Theorem is applied to 3rd order CP tensors whose factor matrices are sparse. In (8), however, only the the factor matrices corresponding to the two first tensor modes are sparse. Therefore, we adapt the results of [18] to our scenario. Let m_r and m_t denote upper bounds on the number of nonzero elements per column of $\boldsymbol{S}_{\mathrm{r}}$ and $\boldsymbol{S}_{\mathrm{t}}$, respectively. Also, let $k_{\boldsymbol{S}_{\mathrm{r}}}$, $k_{\boldsymbol{S}_{\mathrm{t}}}$ and $k_{\boldsymbol{\Gamma}}$ denote the Kruskal rank of the matrices indicated in the sub-index. Model (8) is essentially unique (i.e., its factor matrices can be uniquely recovered up to scalar and permutation ambiguities [12]) if the sufficient conditions below is satisfied:

$$\begin{gathered}\min(L_r, k_{\boldsymbol{S}_{\mathrm{r}}}) + \min(M_t, k_{\boldsymbol{S}_{\mathrm{t}}}) + \min(N_p, k_{\boldsymbol{\Gamma}}) \leq 2L + 2 \quad (11)\\ L_r \geq 2m_r, \quad M_t \geq 2m_t.\end{gathered}$$

From the definition of the fading matrix $\boldsymbol{\Gamma}$ in (6) and the i.i.d. fading assumptions, we have that $k_{\boldsymbol{\Gamma}} = \min(N_p, L)$. Since, in practice, the number of multipaths is typically smaller than the number N_p of frame time slots, then $k_{\boldsymbol{\Gamma}} = L$ and the uniqueness condition (11) simplifies to:

$$\begin{gathered}\min(L_r, k_{\boldsymbol{S}_{\mathrm{r}}}) + \min(M_t, k_{\boldsymbol{S}_{\mathrm{t}}}) \leq L + 2 \quad (12)\\ L_r \geq 2m_r, \quad M_t \geq 2m_t.\end{gathered}$$

A discussion on the minimum number of measurements can be drawn from (12) following the same reasoning presented in [9]. Note that condition (12) provides a trade-off between the number L_r of RF chains at RX and the number M_t of beams at TX ensuring the recovery of the channel parameters from the (compressed) measurement tensor $\mathcal{Y}$ given in (8).

3.2. CP-OMP Algorithm

Assuming that the uniqueness conditions given in (12) are satisfied, one can estimate the factor matrices $\boldsymbol{Q}_{(1)}$, $\boldsymbol{Q}_{(2)}$, and $\boldsymbol{Q}_{(3)}$ up to scaling and permutation by applying a CP decomposition algorithm such as the alternating least squares algorithm (see [12, 19, 20]). Once these matrices are estimated, we need to estimate $\boldsymbol{S}_{\mathrm{r}}$ and $\boldsymbol{S}_{\mathrm{t}}$. To this end, we formulate the following compressed recovery problems:

$$\boldsymbol{s}_{\mathrm{r,opt}} = \arg\min_{\boldsymbol{s}_{\mathrm{r}}} \|\boldsymbol{s}_{\mathrm{r}}\|_0 \quad \text{s.t.} \quad \left\|\boldsymbol{q}_{(1)} - [\boldsymbol{I}_L \otimes (\boldsymbol{W}^{\mathsf{H}}\boldsymbol{\Psi}_{\mathrm{r}})]\boldsymbol{s}_{\mathrm{r}}\right\|_2 \leq \sigma, \quad (13)$$

$$\boldsymbol{s}_{\mathrm{t,opt}} = \arg\min_{\boldsymbol{s}_{\mathrm{t}}} \|\boldsymbol{s}_{\mathrm{t}}\|_0 \quad \text{s.t.} \quad \left\|\boldsymbol{q}_{(2)} - [\boldsymbol{I}_L \otimes (\boldsymbol{F}^{\mathsf{T}}\boldsymbol{\Psi}_{\mathrm{t}})]\boldsymbol{s}_{\mathrm{t}}\right\|_2 \leq \sigma, \quad (14)$$

where $\boldsymbol{s}_{\mathrm{r}} = \mathrm{vec}(\boldsymbol{S}_{\mathrm{r}})$, $\boldsymbol{s}_{\mathrm{t}} = \mathrm{vec}(\boldsymbol{S}_{\mathrm{t}})$, $\boldsymbol{q}_{(1)} = \mathrm{vec}(\boldsymbol{Q}_{(1)})$, $\boldsymbol{q}_{(2)} = \mathrm{vec}(\boldsymbol{Q}_{(2)})$, and σ is a small positive threshold, which may be different for (13) and (14). Since these problems are non-convex, they can be relaxed to ℓ_1 norm minimization problems [17] and solved using any standard compressed recovery algorithm. In this work, we consider the well-known OMP, which gives us the following estimates: $\hat{\boldsymbol{S}}_{\mathrm{r}} = \mathrm{unvec}(\boldsymbol{s}_{\mathrm{r,opt}})$ and $\hat{\boldsymbol{S}}_{\mathrm{t}} = \mathrm{unvec}(\boldsymbol{s}_{\mathrm{t,opt}})$. Then, we rebuild the array manifold matrices as $\hat{\boldsymbol{A}}_{\mathrm{r}} = \boldsymbol{\Psi}_{\mathrm{r}}\hat{\boldsymbol{S}}_{\mathrm{r}}$ and $\hat{\boldsymbol{A}}_{\mathrm{t}} = \boldsymbol{\Psi}_{\mathrm{t}}\hat{\boldsymbol{S}}_{\mathrm{t}}$. To find an estimate of $\boldsymbol{\Gamma}$, we formulate the following LS problem considering the unfolding representation (9):

$$\min_{\boldsymbol{\Gamma}} \left\|[\mathcal{Y}]_{(3)} - \boldsymbol{\Gamma}\left[(\boldsymbol{W}^{\mathsf{H}}\hat{\boldsymbol{A}}_{\mathrm{r}}) \odot (\boldsymbol{F}^{\mathsf{T}}\hat{\boldsymbol{A}}_{\mathrm{t}}^{*})\right]^{\mathsf{T}}\right\|_{\mathrm{F}}, \quad (15)$$

where $[\mathcal{Y}]_{(3)} = [\mathrm{vec}(\boldsymbol{Y}_1), \ldots, \mathrm{vec}(\boldsymbol{Y}_{N_p})]^{\mathsf{T}} \in \mathbb{C}^{N_p \times L_r M_t}$. The solution to (15) is given by $\hat{\boldsymbol{\Gamma}} = [\mathcal{Y}]_{(3)}\left[(\boldsymbol{W}^{\mathsf{H}}\hat{\boldsymbol{A}}_{\mathrm{r}}) \odot (\boldsymbol{F}^{\mathsf{T}}\hat{\boldsymbol{A}}_{\mathrm{t}}^{*})\right]^{\mathsf{T}\dagger}$.

Once all parameters have been estimated, one can reconstruct the channel with (10). The proposed channel estimation method is referred to as CP-OMP, and it is summarized in Algorithm 1. Note that the knowledge of the channel rank L is assumed to compute the CP decomposition of $\mathcal{Y}$. Any CP model order estimation technique, e.g., CORCONDIA [21], can be used to this end.

Algorithm 1 CP-OMP

Require: $\mathcal{Y}$, $\boldsymbol{W}$, $\boldsymbol{F}$, $\boldsymbol{\Psi}_{\mathrm{r}}$, $\boldsymbol{\Psi}_{\mathrm{t}}$, L

1: Estimate $\boldsymbol{Q}_{(1)}$, $\boldsymbol{Q}_{(2)}$, and $\boldsymbol{Q}_{(3)}$ from $\mathcal{Y}$ by applying a CP decomposition algorithm (e.g. [12, 19, 20]);
2: Solve (13) and (14) by OMP to get $\hat{\boldsymbol{S}}_{\mathrm{r}}$ and $\hat{\boldsymbol{S}}_{\mathrm{t}}$, respectively;
3: Build up $\hat{\boldsymbol{A}}_{\mathrm{r}} = \boldsymbol{\Psi}_{\mathrm{r}}\hat{\boldsymbol{S}}_{\mathrm{r}}$ and $\hat{\boldsymbol{A}}_{\mathrm{t}} = \boldsymbol{\Psi}_{\mathrm{t}}\hat{\boldsymbol{S}}_{\mathrm{t}}$;
4: From (15), obtain an LS estimate of $\boldsymbol{\Gamma}$;
5: Reconstruct the channel tensor from (10).

4. SIMULATION RESULTS

In this section, we present some results from computer simulations carried out to evaluate the performance of the proposed channel estimation method. As a benchmark, we consider the OMP algorithm applied independently to each time slot, which can be considered as a "naive" solution since it ignores the CFO diversity across the different slots. As figures of merit, we consider the normalized mean square error (NMSE) of the estimated channel and the achievable rate, which is given by

$$R = \log_2 \det\left[\boldsymbol{I}_{N_s} + \tfrac{\mathrm{SNR}}{N_s}\boldsymbol{W}_{\mathrm{BB}}^{\mathsf{H}}\boldsymbol{W}_{\mathrm{RF}}^{\mathsf{H}}\boldsymbol{H}\boldsymbol{F}_{\mathrm{RF}}\boldsymbol{F}_{\mathrm{BB}}\boldsymbol{F}_{\mathrm{BB}}^{\mathsf{H}}\boldsymbol{F}_{\mathrm{RF}}^{\mathsf{H}}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{W}_{\mathrm{RF}}\boldsymbol{W}_{\mathrm{BB}}\right]$$

The figures of merit are calculated relative to the actual channel matrix taken at the first time slot ($k = 1$), and the transceiver filters are

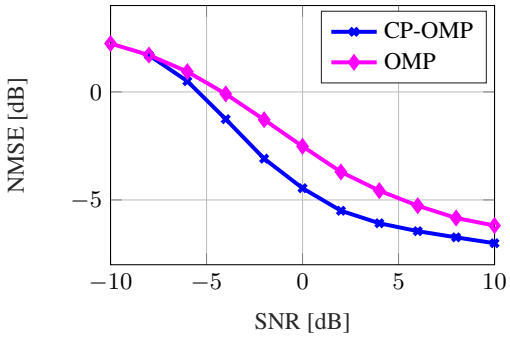

Fig. 2. NMSE vs. SNR. $M_t = 32$ transmissions.

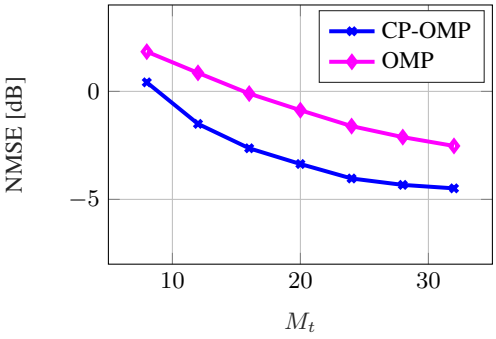

Fig. 3. NMSE vs. M_t. SNR $= 0\,\mathrm{dB}$.

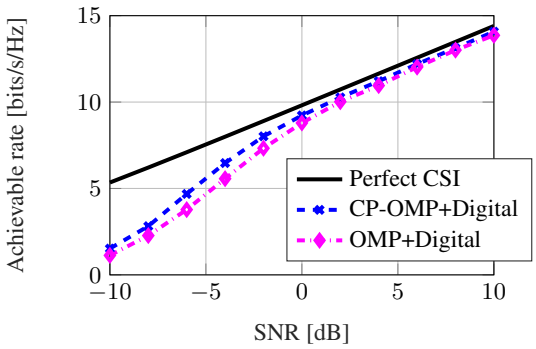

Fig. 4. Achievable rate vs. SNR. $M_t = 32$ transmissions.

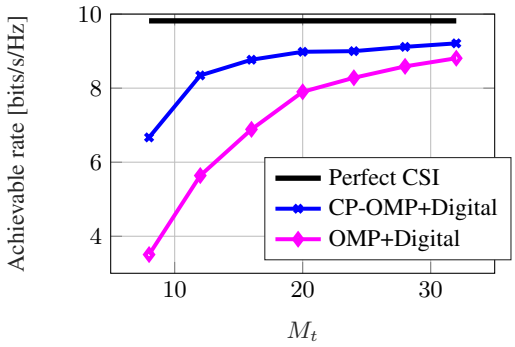

Fig. 5. Achievable rate vs. M_t. SNR $= 0\,\mathrm{dB}$.

computed with respect to the estimated channel matrices. Figures 2–5 were obtained by averaging the results of 1000 independent experiments.

The simulation parameters are: $N_r = 16$, $N_t = 64$ antennas, $L_r = L_t = 6$ RF chains, $N_s = 2$ data streams, $N_p = 10$ time slots, $L = 8$ multipaths, $f_c = 28\,\mathrm{GHz}$ carrier frequency, 10 ppm CFO, which gives $f_{\mathrm{CFO}} = 280\,\mathrm{kHz}$, $F_s = 2\,\mathrm{MHz}$ sampling frequency. During the channel estimation phase, the elements of combiner $\boldsymbol{W}$ and precoder $\boldsymbol{F}$ matrices are respectively generated as $\frac{1}{\sqrt{N_r}}e^{j\varphi_r}$ and $\frac{1}{\sqrt{N_t}}e^{j\varphi_t}$, with phase φ_r, φ_t randomly (discrete uniform distribution) selected from the set of $N_q = 32$ quantized phase shifts $\{0, \frac{2\pi}{N_q}, \ldots, \frac{2\pi(N_q-1)}{N_q}\}$.

We first investigate the NMSE performance of CP-OMP and compare it to the benchmark. In Figure 2, the NMSE is computed as a function of the SNR for $M_t = 32$ transmissions. This result indicates that CP-OMP exhibits superior channel estimation performance from -8 to $10\,\mathrm{dB}$. At $0\,\mathrm{dB}$ SNR, for example, the performance difference is $2\,\mathrm{dB}$. To study how does the number M_t of transmissions affects the NMSE, it is calculated for varying M_t and $0\,\mathrm{dB}$ SNR in Figure 3. As expected from CS theory, the NMSE decreases as M_t grows. We observe that for the given parameters, there is a performance gap of roughly $2\,\mathrm{dB}$ from $M_t = 12$ to $M_t = 32$. This performance improvement provided by CP-OMP can be explained as follows: (i) the proposed method exploits the CFO diversity by combining data embedded in all N_p time slots, while the benchmark considers the information of only a single time slot. (ii) The CP model of the measurement signal tensor allows to breakdown a N_rN_t-dimensional CS problem into N_r- and N_t-dimensional CS sub-problems. Note that the high-dimensional CS problem is more sensitive to noise than the lower dimensional CP-OMP problems. This comes from the fact that CP-OMP begins with a CP decomposition algorithm, which already rejects some noise.

We also conduct computer simulations to assess the rate performance of our mmWave MIMO system model with imperfect CSI. In Figure 4, the achievable rate is calculated in terms of SNR. As a reference, we plot the achievable rate with perfect CSI knowledge. For $M_t = 32$ transmissions, this figure reveals that CP-OMP and the benchmark OMP exhibits close performance. In particular, we note that both methods perform close to the perfect CSI upper bound for high SNR. However, what is the influence of M_t on the achievable rate performance? To find out, we compute the achievable rate as a function of this parameter in Figure 5. We note that, for relatively small transmission numbers M_t, the performance gap between the channel estimation increases. For example, for $M_t = 8$ transmissions, the performance gap is 3 bits/s/Hz. We conclude that our CP-OMP approach is more robust to a small number of channel measurements, represented by M_t in this scenario.

5. CONCLUSION

This paper tackled the mmWave MIMO channel estimation problem under CFO impairments. We proposed a tensor-based channel estimation method called CP-OMP which takes the CFO diversity into account to ameliorate performance. Our simulation results indicate that CP-OMP yields better NMSE and achievable rate performance than the matrix-based OMP benchmark. Also, our results suggest that CP-OMP is more robust to CFO for a small number of channel measurements. Perspectives of this work include the generalization of the proposed tensor-based channel estimator to cope with frequency-selective channels and other hardware impairments.

6. REFERENCES

[1] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, 2016.

[2] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.

[3] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of massive MIMO*, Cambridge University Press, 2016.

[4] C. Zhang, Z. Xiao, L. Su, L. Zeng, and D. Jin, "Iterative channel estimation and phase noise compensation for SC-FDE based mmWave systems," in *Proc. of the 2015 IEEE International Conference on Communications (ICC)*, pp. 2133–2138.

[5] O. Abari, H. Hassanieh, M. Rodriguez, and D. Katabi, "Millimeter wave communications: From point-to-point links to agile network connections," in *Proc. of the 15th ACM Workshop on Hot Topics in Networks*, 2016, pp. 169–175.

[6] N. J. Myers and R. W. Heath, "A compressive channel estimation technique robust to synchronization impairments," in *Proc. of the 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2017, pp. 1–5.

[7] J. Rodriguez-Fernandez, N. Gonzalez-Prelcic, and R. W. Heath Jr., "Channel estimation for millimeter wave MIMO systems in the presence of CFO uncertainties," in *Proc. of the 2018 IEEE International Conference on Communications (ICC)*, pp. 1–6.

[8] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of n-way arrays," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 14, no. 3, pp. 229–239, 2000.

[9] D. C. Araújo and A. L. F. de Almeida, "Tensor-based compressed estimation of frequency-selective mmWave MIMO channels," in *Proc. of the 2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 1–5.

[10] C. E. R. Fernandes, G. Favier, and J. C. M. Mota, "Blind channel identification algorithms based on the PARAFAC decomposition of cumulant tensors: the single and multiuser cases," *Signal Processing*, vol. 88, no. 6, pp. 1382–1401, 2008.

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

[12] P. Comon, X. Luciani, and A. L. De Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 23, no. 7-8, pp. 393–405, 2009.

[13] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.

[14] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, "Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 298–312, May 2018.

[15] H. Xie, F. Gao, and S. Jin, "An overview of low-rank channel estimation for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7313–7321, 2016.

[16] R. Méndez-Rial, C. Rusu, N. G. Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, vol. 4, no. 8, pp. 247–267, 2016.

[17] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Communications Magazine*, vol. 48, no. 11, 2010.

[18] N. D. Sidiropoulos and A. Kyrillidis, "Multi-way compressed sensing for sparse low-rank tensors," *IEEE Signal Processing Letters*, vol. 19, no. 11, pp. 757–760, 2012.

[19] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, 2017.

[20] L. Sorber, M. Van Barel, and L. De Lathauwer, "Structured data fusion," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 586–600, 2015.

[21] R. Bro and H. A. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 17, no. 5, pp. 274–286, 2003.