Tensor-Based Compressed Estimation of Frequency-Selective mmWave MIMO Channels

Daniel C. Araújo and André L. F. de Almeida
Wireless Telecommunications Research Group (GTEL)
Federal University of Ceará, Fortaleza, Brazil
Email: {araujo, andre}@gtel.ufc.br

Abstract—**This paper develops a novel channel estimation technique for frequency-selective mmWave MIMO channels using a hybrid analog-digital architecture. By adopting a tensor formalism to model the effective channel, we link the channel estimation problem to the theory of multi-way compressive sensing of sparse tensors via Parallel Factors (PARAFAC) analysis. By leveraging on this link, a joint estimation of the compressed channel bases (spatial transmit, spatial receive and delay) can be obtained by means of an alternating least squares algorithm. Once these bases are estimated, the channel parameters are extracted by solving a simpler compressive sensing (CS) problem for each basis. Some useful bounds on the minimum number of beams and pilot sequence length can be derived from Kruskal's uniqueness conditions for sparse PARAFAC models. Remarkable channel estimation performance is obtained with short pilot sequences and very few beams, as shown in our simulation results.**

I. Introduction

Beamforming for 5G mobile communication systems promises a great increase in wireless data rates by leveraging on the massive number of antennas at the base station (BS). Massive multiple-input multiple-output (MIMO) has the potential to provide extremely high energy and spectrum efficiencies required to 5G networks [1]. This potential is achieved if the system (i) implements digital beamforming and (ii) has accurate channel state information (CSI) available [2].

The implementation of a complete digital architecture for massive MIMO systems is a tremendous challenge [3]. A dedicated radio frequency (RF) chain is used at each antenna, including a power amplifier (low-noise amplifier at the receiver) and a digital to analog converter (DAC) (analog to digital converter (ADC) at the receiver) [3], [4]. However, for a very large number of antennas, the power consumption levels of high resolution ADC and DAC become prohibitive [3], [5]. To face such a problem, the use of hybrid beamforming (HB) architectures in massive MIMO systems has called attention, especially in millimeter wave (mmWave) systems. Essentially, a hybrid architecture combines analog and digital signal processing by using a limited number of RF chains (much smaller than the number of antennas). The digital part performs baseband processing using microprocessors, while the analog part is implemented at the RF domain by using phase-shifter networks [3], [5], [6].

To foresee the expected massive MIMO gains, a HB transceiver must have partial or complete channel state information (CSI) knowledge. Usually, channel estimation is accomplished by resorting to pilot-assisted (beam training) schemes [6], [7]. The channel acquisition problem plays an important role in large array systems due to the pilot overhead. Some works, such as [7]–[9] have addressed the estimation of a massive MIMO channel by exploiting the intrinsic channel sparsity. With the growing interest in hybrid architectures, the development of bandwidth-efficient channel estimation schemes capable of operating with a minimum pilot overhead is a hot topic, especially in wideband systems where channel frequency selectivity poses an additional challenge.

The works [3], [6] discuss the problem of channel estimation for HB architectures, where solutions based on compressive sensing (CS) are developed. In [10], a solution to deal with frequency selective channels is proposed. Therein, the authors propose to stack the time-domain samples of the received signal during multiple frames, each one associated with a given transmit beam. Using a traditional (vector-based) CS framework, the receiver is capable of extracting the channel parameters (angle of departure (AoD) and path gain). Accurate channel estimates can be obtained at the price of a high computational complexity due to the large dimension of the CS problem at hand. When considering multiple receive antennas, the approach of [10] results in even higher complexity, which may be prohibitive in a practical setting. It is worth noting, however, that the frequency-selective mmWave channel model admits a sparse representation in three dimensions (spatial transmit, spatial receive and delay) and can be modeled as a sparse (compressible) low-rank tensor.

In this work, we exploit both sparse and multidimensional structures of the frequency-selective mmWave MIMO channel. By using a tensor modeling formalism, we recast the channel estimation problem as a multi-way compressed sensing problem [11]. Based on this link, we present a tensor-based channel estimation method consisting of two phases. First, a joint estimation of the compressed channel bases (spatial transmit, spatial receive and delay) can be obtained by means of an alternating least squares algorithm. Then, the channel parameters are extracted by solving a simpler compressive sensing (CS) problem for each basis (i.e. per tensor dimension). We also derive useful conditions on the minimum number of beams and pilot sequence length that guarantee the unique recovery of the channel bases, by exploiting Kruskal's uniqueness conditions for sparse Parallel Factor (PARAFAC) models [11].

Notation: Scalars are denoted by lower-case letters (a, b), vectors by bold upper-case letters $(\mathbf{a}, \mathbf{b}, ..)$, matrix by bold upper-case letters $(\mathbf{A}, \mathbf{B}, ..)$, and tensors are defined by calligraphic upper-case letters $(\mathcal{A}, \mathcal{B}, ...)$. $\mathbf{a}(i)$ denotes the ith entry of vector $\mathbf{a} \in \mathbb{C}^{I}$, $\mathbf{A}(i, j)$ denotes the (i, j)th entry of matrix $\mathbf{A} \in \mathbb{C}^{I\times J}$, while $\mathcal{A}(i, j, k)$ represents the (i, j, k)th entry of tensor $\mathcal{A} \in \mathbb{C}^{I\times J\times K}$. The vec$\{\cdot\}$ operator vectorizes a matrix by stacking its columns, $\otimes$ stands for the Kronecker product, and $\odot$ is the Khatri-Rao product (column-wise Kronecker product). The 1-mode unfolding of tensor $\mathcal{Y} \in \mathbb{C}^{I_1\times I_2\times I_3}$ is denoted as $\mathbf{Y}_1 \in \mathbb{C}^{I_2 I_3\times I_1}$. The 2-mode (resp. 3-mode) unfolding of this tensor is denoted as $\mathbf{Y}_2 \in \mathbb{C}^{I_3 I_1\times I_2}$ (resp.

978-1-5386-1251-4/17/$31.00 ©2017 IEEE

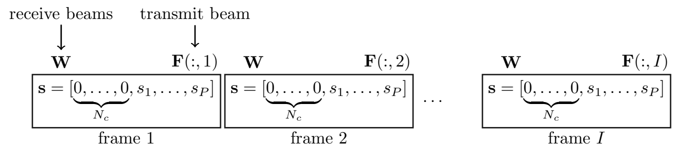

Fig. 1. Pilot transmission structure. The receiver uses the same set of beams while the transmitter changes its beams from frame-to-frame.

$\mathbf{Y}_3 \in \mathbb{C}^{I_1 I_2 \times I_1}$). The n-mode product between a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ and a matrix $\mathbf{X} \in \mathbb{C}^{R_n \times I_n}$ is defined as $\mathcal{Y} = \mathcal{A} \times_n \mathbf{X}$, so that $\mathbf{Y}_n = \mathbf{A}_n \mathbf{X}^T$, $n = 1, 2, 3$.

II. System Model

Consider a single user MIMO system where the transmitter and the receiver are equipped with N_t and N_r antennas, respectively. Assume that both transmitter and receiver employ a hybrid beamforming structure using M_t and M_r RF chains, respectively. The transmitter uses multiple beams to send a length-P pilot sequence over I spatial directions, while the receiver uses the Q RF chains to collect the incoming training signals from these different directions. The spatial filters at the transmitter and receiver are denoted as $\mathbf{F} = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} \in \mathbb{C}^{N_t \times I}$ and $\mathbf{W} = \mathbf{W}_{\text{RF}}\mathbf{W}_{\text{BB}}\ \mathbb{C}^{N_r \times Q}$, where $I \leq M_t$ and $Q \leq M_r$ denote the number of beams used in the same communication resource at the transmitter and receiver sides, respectively. The channel matrix is defined as a summation of N_c delay tap matrices $\mathbf{H}_d$, $d = \{0, 1, \ldots, N_c - 1\}$. The received signal is expressed as

$$\mathbf{r}_{i,n} = \sqrt{\rho} \sum_{d=0}^{N_c-1} \mathbf{H}_d \mathbf{F}(:, i) s_{n-d} + \mathbf{z}_{i,n}, \tag{1}$$

where i is the beam index associated with a given transmit frame, s_n is the nth non-zero instance of the training frame

$$\mathbf{s} = [\underbrace{0, \ldots, 0}_{N_c-1}, s_1, s_2, \ldots, s_P] \tag{2}$$

of length $N = P + N_c - 1$, ρ denotes the average received power, and $\mathbf{z}_n \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ the circularly symmetric complex Gaussian noise vector. As mentioned in [10], the $N_c - 1$ zero padding allows for RF reconfiguration between frames, while avoiding interframe interference.

Each transmit frame is associated with a specific beam pattern, i.e. assuming a codebook with I beam patterns, the system needs I training frames to use the beams specified by the codebook. Each frame carriers a pilot sequence with P symbols. Figure 1 shows the frame configuration of the system. By applying the spatial filters (combiners), we get

$$\begin{aligned} y_{q,i,n} &= \sqrt{\rho}\mathbf{W}(:, q)^T \mathbf{r}_{i,n} \\ &= \mathbf{W}(:, q)^T \sum_{d=0}^{N_c-1} \mathbf{H}_d \mathbf{F}(:, i) s_{n-d} + \tilde{z}_{q,i,n}, \end{aligned} \tag{3}$$

where $\tilde{z}_{q,i,n} = \mathbf{W}(:, q)^T \mathbf{z}_{i,n}$. Therefore, the output signal can be modeled as a third-order tensor $\mathcal{Y} \in \mathbb{C}^{Q \times I \times P}$, whose (q, i, n)th entry can be written as [12]

$$y_{q,i,n} = \mathcal{H} \times_1 \mathbf{W}(:, q)^T \times_2 \mathbf{F}(:, i)^T \times_3 \mathbf{S}(:, n)^T + \tilde{z}_{q,i,n}$$

where $\mathcal{H} \in \mathbb{C}^{N_r \times N_t \times N_c}$ is the channel tensor obtained by concatenating the N_c channel matrices $\{\mathbf{H}_d\}_{d=0}^{N_c-1}$ along its third mode, i.e. $\mathcal{H} = \mathbf{H}_0 \sqcup_3 \mathbf{H}_1 \sqcup_3 \ldots \sqcup_3 \mathbf{H}_{N_c-1}$, and $\mathbf{S} \in \mathbb{C}^{P \times N_c}$ is the convolution matrix containing the pilot sequence

$$\mathbf{S} = \begin{bmatrix} s_1 & 0 & \ldots & 0 \\ s_2 & s_1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ s_P & \ldots & \ldots & s_{P-N_c+1} \end{bmatrix}^T.$$

Collecting the N time-domain samples during I frames (transmit beams) from Q RF chains (receive beams), we get

$$\mathcal{Y} = \mathcal{H} \times_1 \mathbf{W}^T \times_2 \mathbf{F}^T \times_3 \mathbf{S}^T + \tilde{\mathcal{Z}} \tag{4}$$

There are two basic multiway models largely used in the literature: Tucker3 and PARAFAC. The first is essential for data compression as shown in [13], but its uniqueness cannot be ensured in general. The second is unique under mild conditions related to the *Kruskal-rank* concept [11], [14].

III. Channel Estimation via Multiway Compressive Sensing

In this section, We formulate a multidimensional sparse channel estimator for mmWave bands by leveraging on the uniqueness properties of sparse PARAFAC models.

A. Sparse PARAFAC Formulation

Consider a geometric channel model for a frequency selective scenario

$$\mathbf{H}_d = \sum_{l=0}^{L-1} \alpha_l g(dT_s - \tau_l) \mathbf{v}(\theta_l) \mathbf{v}(\phi_l)^H,\ d = 0, \ldots, N_c - 1,$$

where $g(\tau)$ denotes the system pulse shaping evaluated at τ, $\alpha_d \in \mathbb{C}$ is the complex gain associated with the dth path, $\tau_d \in \mathbb{R}$ is the delay of the dth path, $\theta \in [0, 2\pi)$ and $\phi \in [0, 2\pi)$ are the AoA and AoD, respectively, $\mathbf{v}(\theta) \in \mathbb{C}^{N_r}$ and $\mathbf{v}(\phi) \in \mathbb{C}^{N_t}$ are the steering vectors for the transmitter and receiver, respectively.

Let us define the matrices $\mathbf{V}_{\text{r}} \in \mathbb{C}^{N_r \times L}$ and $\mathbf{V}_{\text{t}} \in \mathbb{C}^{N_t \times L}$ concatenating the receive and transmit steering vectors respectively, and the matrix $\mathbf{D} \in \mathbb{C}^{N_c \times L}$, the (d, l)th entry of which is given by $\mathbf{D}(d, l) = \alpha_l g(dT_s - \tau_l)$. The dth channel matrix $\mathbf{H}_d$ can be compactly written as $\mathbf{H}_d = \mathbf{V}_{\text{r}} \text{diag}\{\mathbf{D}(d, :)\} \mathbf{V}_{\text{t}}^H$, where $\text{diag}\{\mathbf{D}(d, :)\}$ defines a diagonal matrix formed from the dth row of $\mathbf{D}$. This matrix corresponds to the d-th frontal slice of the channel tensor $\mathcal{H} \in \mathbb{C}^{N_r \times N_t \times N_c}$, which follows a rank-$L$ PARAFAC model, i.e.

$$\mathcal{H} = \mathcal{I}_{3,L} \times_1 \mathbf{V}_{\text{r}} \times_2 \mathbf{V}_{\text{t}}^* \times_3 \mathbf{D}, \tag{5}$$

where $\mathcal{I}_{3,L}$ is the third-order identity tensor of dimensions $L \times L \times L$. The uniqueness of the model is guaranteed under *Kruskal-rank* of $\mathbf{V}_{\text{r}}$, $\mathbf{V}_{\text{t}}$ and $\mathbf{D}$. The *Kruskal-rank* of $\mathbf{V}_{\text{r}}$, denoted as $k_{\mathbf{V}_{\text{r}}}$, is the maximum k such that any k columns of $\mathbf{V}_{\text{r}}$ are linearly independent ($k_{\mathbf{V}_{\text{r}}} < r_{\mathbf{V}_{\text{r}}} \equiv \text{rank}(\mathbf{V}_{\text{r}})$). Given the channel tensor $\mathcal{H}$, if $k_{\mathbf{V}_{\text{r}}} + k_{\mathbf{V}_{\text{t}}} + k_{\mathbf{D}} \geq 2L + 2$, then $(\mathbf{V}_{\text{r}}, \mathbf{V}_{\text{t}}, \mathbf{D})$ are unique up to a column permutation and scaling [11]. In this work, however, we do not explore directly the model (5) to estimate the channel parameters. Instead, we resort to a multi-way compressed representation of the channel. To this end, let us first substitute (5) into (4), yielding

$$\mathcal{Y} = \mathcal{I}_{3,L} \times_1 (\mathbf{W}^T \mathbf{V}_{\text{r}}) \times_2 (\mathbf{F}^T \mathbf{V}_{\text{t}}^*) \times_3 (\mathbf{S}^T \mathbf{D}) + \tilde{\mathcal{Z}} \tag{6}$$

Comparing (5) and (6), we can see that $\mathbf{W}$, $\mathbf{F}$, and $\mathbf{S}$ play the role of compression matrices associated with the first, second and third modes of the channel tensor $\mathcal{H}$, respectively. The received signal tensor $\mathcal{Y}$ can also be interpreted as the "effective" (compressed version of the) mmWave channel. We assume that the channel tensor $\mathcal{H}$ admits a sparse representation at each mode, which translates into imposing sparsity on the spatial receive, spatial transmit, and delay domains. Under this assumption, $\mathcal{Y}$ can be expressed as rewritten as

$$\mathcal{Y} = \mathcal{I}_{3,L} \times_1 (\mathbf{W}^T \mathbf{\Phi}_1 \mathbf{B}_1) \times_2 (\mathbf{F}^T \mathbf{\Phi}_2 \mathbf{B}_2) \times_3 (\mathbf{S}^T \mathbf{\Phi}_3 \mathbf{B}_3) + \tilde{\mathcal{Z}}, \quad (7)$$

where $\mathbf{\Phi}_1 \in \mathbb{C}^{N_r \times N_r}$, $\mathbf{\Phi}_2 \in \mathbb{C}^{N_t \times N_t}$, and $\mathbf{\Phi}_3 \in \mathbb{C}^{N_c \times N_c}$, are orthogonal bases to the spatial receive, spatial transmit and delay modes. Likewise, $\mathbf{B}_1 \in \mathbb{C}^{N_t \times L}$, $\mathbf{B}_2 \in \mathbb{C}^{N_r \times L}$, and $\mathbf{B}_3 \in \mathbb{C}^{N_c \times L}$ are sparse matrices associated with each mode of the channel tensor $\mathcal{H}$. Without lacking of generality, we assume $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$, and $\mathbf{\Phi}_3$ are Fourier matrices.

Let us define m_1 (resp. m_2, m_3) as an upper bound on the number of non-zero elements per column of $\mathbf{B}_1$ (resp. $\mathbf{B}_2, \mathbf{B}_3$). More specifically, m_1 and m_2 are the maximum number of receive and transmit beams contained in $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$, respectively, that describe the spatial response of a channel path, while m_3 is the maximum number of delayed pulse shaping responses contained in $\mathbf{\Phi}_3$ that describe the time-domain impulse response of a channel path. To estimate the channel tensor, the receiver explores the multi-way compressed representation given in (7), where the sparse factor matrices $\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{B}_3$ are the parameters of interest. Their uniqueness are linked to the theorem stated in [15], as follows:

Theorem 3.1: Considering the upper bounds m_1, m_2, and m_3 on the number of nonzero elements per column of $\mathbf{B}_1, \mathbf{B}_2$, and $\mathbf{B}_3$, respectively, if

$$\min(Q, k_{\mathbf{B}_1}) + \min(I, k_{\mathbf{B}_2}) + \min(P, k_{\mathbf{B}_3}) \geq 2L + 2, \quad (8)$$

and $Q \geq 2m_1$, $I \geq 2m_2$, $P \geq 2m_3$, then the matrices $\mathbf{B}_1, \mathbf{B}_2$, and $\mathbf{B}_3$ are almost sure identifiable.

Exploiting Theorem 3.1, the system can properly choose the number of receive beams, transmit beams, and pilot symbols that guarantee model uniqueness. Assume that $r_{\mathbf{B}_1} = k_{\mathbf{B}_1}$, $r_{\mathbf{B}_2} = k_{\mathbf{B}_2}$, and $r_{\mathbf{B}_3} = k_{\mathbf{B}_3}$, which means that $\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{B}_3$ do not contain proportional columns. In our context, this assumption implies the channel paths have different spatial signatures and different propagation delays. Under these constraints, we have the following:

1) If $Q \geq L$ and $I \geq L$, then $P \geq 2m_3$ pilot symbols per frame are enough to estimate L paths. Thus, the overhead per frame can be reduced at the minimum of $2m_3$.
2) If $Q \geq L$ and $P \geq L$, then $I \geq 2m_2$ transmit beams are sufficient to estimate L paths. Since the transmit beams are associated with the training frames, the total number of frames is reduced to $2m_2$.
3) If $I \geq L$ and $P \geq L$, then $Q \geq 2m_1$ receive beams are sufficient to estimate L paths. This configuration can be useful for scenarios where the receiver has a limited number of RF chains. This means the system must increase either the number of beams or the pilot sequence length.

Theorem 3.1 also asserts the matrices $\mathbf{B}_1, \mathbf{B}_2$, and $\mathbf{B}_3$ are indentifiable from the received signal $\mathcal{Y}$ if $Q > k_{\mathbf{B}_1}$, $I > k_{\mathbf{B}_2}$, $P > k_{\mathbf{B}_3}$ as if the receiver had available the channel $\mathcal{H}$. Note that if we neglect the tensor structure of the channel by using a vector-CS problem, then $QIP \geq 2Lm_1m_2m_3$ measurements are required to estimate the vectorized channel. If we assume $m_1 = m_2 = m_3 = m$, the vector-based CS formulation would need $2Lm^3$ measurements while the tensor-based one requires $8m^3$. Therefore, in our approach, the number of measurements does not depend on the number of paths, but rather on the sparsity of the factor matrices representing the spatial transmit, spatial receive and delay domains. In principle, if we know that $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$, and $\mathbf{\Phi}_3$ return the sparsest factor matrices, the system overhead achieves the its lower-bound to measure the channel tensor $\mathcal{H}$ using the minimum number of samples.

Algorithm 1 ALS description

```
Initialize the factor matrices A_1, A_2, and A_3.
while e > σ^2 do
    Â_1^T ← (Â_2 ⊙ Â_3)^† Y_1
    Â_2^T ← (Â_3 ⊙ Â_1)^† Y_2
    Â_3^T ← (A_1 ⊙ Â_2)^† Y_3
    Ŷ_1 ← (Â_2 ⊙ Â_3) Â_1^T
    e ← ||Ŷ_1 − Y_1||_F^2 / ||Y_1||_F^2
end while
```

B. Algorithm Description

We propose a two-stage approach to reconstruct the mmWave channel tensor (5) from its multi-way compressed representation (7). First, the alternating least squares (ALS) algorithm is used to fit a PARAFAC model to (7) in the presence of noise. As a second stage, we compute the minimum l_1-norm solutions for the sparse factor matrices $\mathbf{B}_1, \mathbf{B}_2$, and $\mathbf{B}_3$. Let us define compressed factor matrices of $\mathcal{Y}$ as

$$\mathbf{A}_1 = \mathbf{W}^T \mathbf{\Phi}_1 \mathbf{B}_1, \;\; \mathbf{A}_2 = \mathbf{F}^T \mathbf{\Phi}_2 \mathbf{B}_2, \;\; \mathbf{A}_3 = \mathbf{S}^T \mathbf{\Phi}_3 \mathbf{B}_3 \quad (9)$$

We exploit the unfolding representations of (7) given by

$$\mathbf{Y}_1 = (\mathbf{A}_2 \odot \mathbf{A}_3) \mathbf{A}_1^T \quad (10)$$
$$\mathbf{Y}_2 = (\mathbf{A}_3 \odot \mathbf{A}_1) \mathbf{A}_2^T \quad (11)$$
$$\mathbf{Y}_3 = (\mathbf{A}_1 \odot \mathbf{A}_2) \mathbf{A}_3^T \quad (12)$$

The ALS algorithm makes use of (10), (11), and (12) to update the factor matrices $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{A}_3$ in the least squares (LS) sense. Each ALS iteration has three steps. At each time, one matrix $\mathbf{A}_i$ is updated by fixing the other two to their values estimated in previous steps. The procedure is repeated until $e = \|\hat{\mathbf{Y}}_1 - \mathbf{Y}_1\|_F^2 / \|\mathbf{Y}_1\|_F^2 < \sigma^2$, where e is the reconstruction error computed at every iteration, and σ is a threshold. The ALS algorithm is summarized in Algorithm 1.

Upon convergence of the ALS algorithm, the second step consists of solving three CS problems to estimate the channel parameters (AoA, AoD, and delay) from $\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{B}_3$, respectively, as follows

$$\min_{\mathbf{b}_1} \|\hat{\mathbf{a}}_1 - \left(\mathbf{I} \otimes \mathbf{W}^T \mathbf{\Phi}_1\right) \mathbf{b}_1\|_2 + \beta \|\mathbf{b}_1\|_1, \quad (13)$$
$$\min_{\mathbf{b}_2} \|\hat{\mathbf{a}}_2 - \left(\mathbf{I} \otimes \mathbf{F}^T \mathbf{\Phi}_2\right) \mathbf{b}_2\|_2 + \beta \|\mathbf{b}_2\|_1, \quad (14)$$
$$\min_{\mathbf{b}_3} \|\hat{\mathbf{a}}_3 - \left(\mathbf{I} \otimes \mathbf{S}^T \mathbf{\Phi}_3\right) \mathbf{b}_3\|_2 + \beta \|\mathbf{b}_3\|_1, \quad (15)$$

where $\hat{\mathbf{a}}_n \equiv \text{vec}\{\hat{\mathbf{A}}_n\}$ and $\mathbf{b}_n \equiv \text{vec}\{\mathbf{B}_n\}$, $n = 1, 2, 3$. The problem in (13) is a least absolute shrinkage and selection operator (LASSO) problem, which can be handled with convex optimization solvers [16]. Once the sparse factor matrices

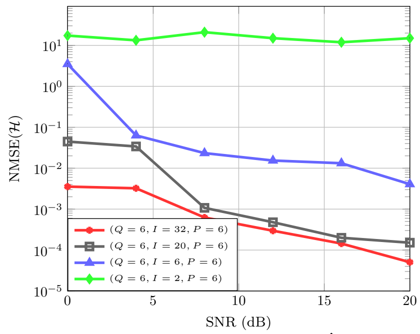

Fig. 2. NMSE of the estimated channel tensor $\hat{\mathcal{H}}$ for different number of transmit beams.

$\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{B}_3$ are estimated, the channel tensor $\mathcal{H}$ can be reconstructed by the following multi-way mapping

$$\hat{\mathcal{H}} = \mathcal{I}_{3,L} \times_1 (\mathbf{\Phi}_1 \mathbf{B}_1) \times_2 (\mathbf{\Phi}_2 \mathbf{B}_2) \times_3 (\mathbf{\Phi}_3 \mathbf{B}_3). \tag{16}$$

Remark: It is worth noting that the transmit/receive spatial filters and pilot sequence define the structure of the measurement matrices $\mathbf{W}$, $\mathbf{F}$, and $\mathbf{S}$ of our multi-way CS model. An optimum design of these matrices yields the maximum channel compression and, consequently, the optimum use of the system resources. This problem is not investigated in this paper, and we assume $\mathbf{W} = \mathbf{W}_{RF}$ and $\mathbf{F} = \mathbf{F}_{RF}$. In this work, the entries of these matrices are drawn from a Bernoulli distribution which meets the constant modulus restriction of the analog beamforming network. Although such a distribution is convenient, it is known that the complex Gaussian distribution provides the best performance in terms of sparse recovery [17].

IV. Simulation Results

The performance of the proposed channel estimation algorithm is evaluated in terms of the normalized mean square error (NMSE) defined as

$$\text{NMSE}(\mathcal{H}) = \sum_{r=1}^{R} \frac{\|\hat{\mathcal{H}}_r - \mathcal{H}\|_F^2}{\|\mathcal{H}\|_F^2}, \tag{17}$$

where r is the number of Monte Carlo runs. The ALS implementation is the plain vanilla algorithm described in 1. Although more efficient implementations are present in the literature, such as in [11], [18]–[20], our focus is not on the algorithm design itself, but rather on the modeling of the problem. The factor matrices estimated by the ALS algorithm are affected by column permutation and scaling ambiguities. However, these ambiguities are irrelevant to our problem, since we are interested in the extraction the channel parameters after the sparse recovery step. To this end, a traditional matching pursuit algorithm [21] is used. We consider a scenario with $N_t = 64$, $N_r = 16$, $N_c = 20$, and $L = 5$. Both arrays are linear with half wavelength spaced antennas.

Figure 2 shows the estimation performance for different numbers $I = \{2, 6, 20, 32\}$ of transmit beams (frames). Note that the proposed algorithm fails in estimating the channel for $I = 2$, since the second corollary of Theorem 3.1 is not satisfied in this case. More specifically, since we assume $L = 5$, the number of transmit beams should satisfy $I \geq 6$ to ensure a unique channel recovery. It can also be seen in 2 that the estimation accuracy increases as more transmit beams are used, which also implies more training frames, i.e. an increased training overhead. Figures 3, 4, and 5 illustrate the algorithm

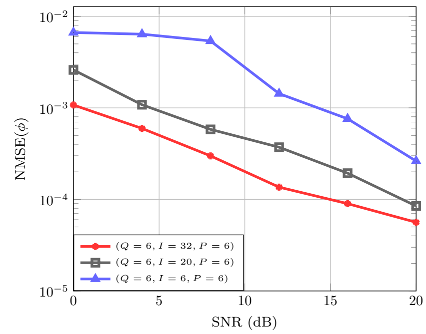

Fig. 3. NMSE of the estimated AoDs.

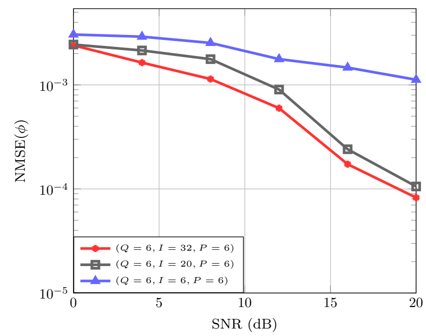

Fig. 4. NMSE of the estimated AoAs.

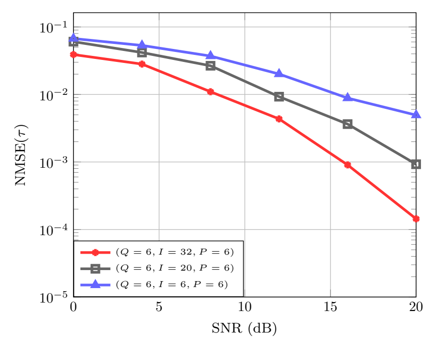

Fig. 5. NMSE of the esimated path delays.

performance in terms of AoD, AoD, and delay estimation. The proposed algorithm provides accurate estimates with very few measurements by exploiting the low-rank tensor structure of the effective mmWave channel.

V. Conclusion

In this paper, we investigated the problem of frequency-selective channel estimation in mmWave MIMO systems using a hybrid beamforming architecture at the transmitter and receiver. The proposed solution jointly exploits the intrinsic sparsity and multidimensional nature of the channel by resorting to a multi-way CS model. Based on this link, bounds on the minimum number of transmit beams, receive beams, and pilot sequence length can be obtained from Kruskal's uniqueness conditions for sparse PARAFAC models. In future works, we shall extend the proposed estimator to time-varying channels.

Acknowledgment

The authors acknowledge the technical and financial support from Ericsson Research, Sweden and from the Ercisson Innovation Center, Brazil, under EDB/UFC.44.

References

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[2] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1259–1262, June 2016.

[3] A. Alkhateeb, O. El Ayach, G. Leus, and R. 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, Oct 2014.

[4] V. Venkateswaran and A. J. van der Veen, "Partial beamforming to reduce adc power consumption in antenna array systems," in *2008 IEEE 9th Workshop on Signal Processing Advances in Wireless Communications*, Jul. 2008, pp. 146–150.

[5] A. Alkhateeb, Y. H. Nam, J. Zhang, and R. W. Heath, "Massive MIMO combining with switches," *IEEE Wireless Communications Letters*, vol. 5, no. 3, pp. 232–235, June 2016.

[6] R. W. Heath, N. Gonzlez-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, April 2016.

[7] J. Choi, D. Love, and P. Bidigare, "Downlink Training Techniques for FDD Massive MIMO Systems: Open-Loop and Closed-Loop Training With Memory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 802–814, Oct 2014.

[8] J. W. Choi and B. Shim, "Compressive sensing based pilot reduction technique for massive MIMO systems," in *Information Theory and Applications Workshop (ITA), 2015*, Feb. 2015, pp. 115–118.

[9] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, "Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO," *IEEE Transactions on Communications*, vol. 64, no. 2, pp. 601–617, Feb. 2016.

[10] K. Venugopal, A. Alkhateeb, R. W. Heath, and N. G. Prelcic, "Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 6493–6497.

[11] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 810–823, Mar 2000.

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

[13] M. Duarte and R. Baraniuk, "Kronecker compressive sensing," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 494 –504, Feb. 2012.

[14] A. K. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis with Applications in the Chemical Sciences*. John Wiley & Sons, Ltd., 2004.

[15] 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, Nov 2012.

[16] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[17] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[18] M. Rajih and P. Comon, "Enhanced line search: A novel method to accelerate PARAFAC," in *2005 13th European Signal Processing Conference*, Sept 2005, pp. 1–4.

[19] B. W. Bader, T. G. Kolda *et al.*, "Matlab tensor toolbox version 2.6," Available online, February 2015. [Online]. Available: http://www.sandia.gov/ tgkolda/TensorToolbox/

[20] B. W. Bader and T. G. Kolda, "Efficient MATLAB computations with sparse and factored tensors," *SIAM Journal on Scientific Computing*, vol. 30, no. 1, pp. 205–231, December 2007.

[21] J. A. Tropp and D. Needell, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *CoRR*, vol. abs/0803.2392, 2008. [Online]. Available: http://arxiv.org/abs/0803.2392