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

Daniel C. Araújo
Research Group of Wireless Communication (GTEL)
Department of Teleinformatics
Federal University of Ceará
Ceará, Fortaleza 60020-181
Email: araujo@gtel.ufc.br

André Lima F. de Almeida
Research Group of Wireless Communication (GTEL)
Department of Teleinformatics
Federal University of Ceará
Ceará, Fortaleza 60020-181
Email: andre@gtel.ufc.br

***Abstract*—This paper develops a novel hybrid analog-digital frequency selective channel estimation technique assuming multiple antennas at the transmitter and multiple antennas at the receiver. We link the estimation problem to the Parallel Factors (PARAFAC) analysis by modeling the effective frequency-selective MIMO as a third order tensor. We leverage on this link to derive a solution that jointly estimates the transmit-receive spatial characteristics and the delay pattern of the multidimensional channel. The proposed method relies on the joint sparsity of the spatial and delay domains, thereby affording the system to use short pilot sequences and very few beams to accurate estimate channel. Moreover, we exploit the fundamental results of Kruskal's uniqueness for tensor compressive sensing [1] to derive useful bounds on the minimum number of beams and pilot sequence length. The simulation results show that the proposed solution can achieve very accurate estimates if the system meets Kruskal conditions.**

I. Introduction

Beamforming for 5G mobile communication systems promises to enable a great increase in wireless data rate due to the massive number of antennas intended to be implemented at the base station (BS). Massive multiple-input multiple-output (MIMO) has the potential to provide extremely high energy and spectrum efficiency required by the 5G networks [2]. Such a potential is achieved if the system (i) implements digital beamforming and (ii) has channel state information (CSI) available [3].

The implementation of a complete digital architecture for massive MIMO systems is a tremendous challenge [4]. For each antenna element, it has associated a dedicated radio frequency (RF) chain, which includes power amplifier (low-noise amplifier at the receiver), digital to analog converter (DAC) (analog-digital converter (ADC) at the receiver) and so on [4], [5]; therefore, when the number of antennas increases the power consumption from high resolution ADC and DAC becomes prohibitive [4], [6]. The use of hybrid beamforming (HB) to deploy massive MIMO has called attention, for its architecture can be implemented with limited number of RF chains, i.e. the number of DAC (ADC at the receiver) is reduced. Essentially, the hybrid architecture has a digital part which performs the baseband processing using microprocessors whereas, the analog part is implemented at the RF frequency by using a phase-shifter network [4], [6], [7].

To design the HB and provide the expected massive MIMO gains, the transceiver must have partial or complete CSI knowledge. The channel acquisition phase is crucial due to the large system overhead which grows proportionally to the number of antennas. Therefore, the development of a CSI estimator that fits to a hybrid massive MIMO transceiver must take into account the digital-analog architecture and provide reliable estimates with short pilot sequences [4], [6]. The use of large bandwidth in the next generation of wireless systems poses an additional challenge to problem, since the channel is frequency selective.

The works [4], [7] discuss the problem of channel estimation for HB architectures, and solutions based on compressive sensing (CS) are proposed. Recently, [8] presents a solution that deals with frequency selective channels. The idea consists of stacking the received signal of multiple frames each one associated to a given transmit beam. Using the CS framework, the receiver is capable of extracting the angle of departure (AoD) and the path gain associated to each path. Although this technique provides very accurate estimates as shown in [8], the extension for multiple antennas at the receiver becomes prohibitive since the dimension of the stacked received signal vector is very large which leads to high computational complexity of the sparse recovery algorithm.

The proposed solution consists of modeling the frequency selective channel as a third-order tensor [9], [10]. Such mathematical formalism is the natural extension of vector and matrices for multidimensional spaces and enables us to exploit well established models that carrier with them conditions to guarantee uniqueness in tensor problems. We explore herein a novel approach for estimating the massive MIMO frequency selective channels that relies on the tensor compressive sensing problem, originally presented in [1]. More specifically, this paper establishes a link between the fundamental Kruskal's uniqueness results for compressive PARAFAC analysis to the compressive estimation of frequency-selective mmWave MIMO channels.

Notation: A scalar is denoted in italic, e.g. a. A column vector is a bold lowercase letter, e.g. $\mathbf{a}$ whose ith entry is $\mathbf{a}(i)$. A matrix is denoted by a bold uppercase letter, e.g. $\mathbf{A}$ with (i,j)th entry $\mathbf{A}(i,j)$; $\mathbf{A}(:,j)$ is the jth column of $\mathbf{A}$. A third-order tensor is a caligrafic uppercase letter, e.g. $\mathcal{A}$ with (i,j,k)th entry $\mathcal{A}(i,j,k)$. The vec$\{.\}$ operator stacks the columns of its argument into one big column vector; $\otimes$

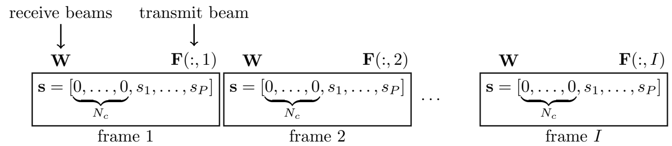

Fig. 1. Each frame carries the same pilot sequence. The receiver uses the same set of beams while the transmitter changes the beam from frame-to-frame.

stands for the Kronecker product, $\odot$ is the Khatri-Rao product (column-wise Kronecker product) $\times_i$ is the mode product over the ith tensor dimension [9], [10].

II. System Model

Consider a single user MIMO system with transmitter and receiver 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 spatial filter at the transmitter and receiver are $\mathbf{F} = \mathbf{F}_{\text{RF}}\mathbf{F}_{\text{BB}} \in \mathbb{C}^{N_t \times I}$ 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, respectively. The transmitter uses multiple beams to send a single pilot sequence with length P into I directions, and the receiver uses the multiple RF chains to collect the incoming training signal from different directions. 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 variable ρ denotes the average received power and $\mathbf{z}_n \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ the circularly symmetric complex Gaussian distributed noise vector, therefore, the received signal is expressed as

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

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

$$\mathbf{s} = [0, \ldots, 0, s_1, s_2, \ldots, s_P] \quad (2)$$

of length $N = P + N_c - 1$. Each transmit frame is associated to a specific beam pattern, i.e. assuming a codebook with I beam patterns, the system needs to I training frames to use beams specified by the codebook. Each frame carriers a pilot sequence with P symbols. Fig. 1 shows the frame configuration of the system.

The receiver applies the combiner $\mathbf{W}(:, q)$ over the training frame, so that the combiner output is expressed as

$$y_{q,i,n} = \sqrt{\rho}\mathbf{W}(:, q)^T \sum_{d=0}^{N_c-1} \mathbf{H}_d \mathbf{F}(:, i) s_{n-d} + z_{q,i,n}, \quad (3)$$

where $z_{q,i,n} = \mathbf{w}_q^H \mathbf{z}_{i,n}$ The output signal can be described in terms of tensor notation [9], [10]

$$y_{q,i,n} = \mathcal{H} \times_1 \mathbf{W}(:, q)^T \times_2 \mathbf{F}(:, i)^T \times_3 \mathbf{S}(:, n)^T, \quad (4)$$

where $\mathcal{H} \in \mathbb{C}^{N_r \times N_t \times N_c}$ is generated from the concatenation in the third dimension of the N_c delay taps, and $\mathbf{S} \in \mathbb{C}^{N_c \times P}$ is the convolution matrix containing the pilot sequence

$$\mathbf{S}^T = \begin{bmatrix} s_1 & 0 & \ldots & 0 \\ s_2 & s_1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ s_P & \ldots & \ldots & s_N \end{bmatrix}.$$

Assuming a collection of transmitter and receiver beams, i.e. $i = \{1, \ldots, I\}$ and $q = \{1, \ldots, Q\}$, respectively, and the N time instants within the training frame, we can express the signal model as the third-order tensor [10], [11]

$$\mathcal{Y} = \mathcal{H} \times_1 \mathbf{W}^T \times_2 \mathbf{F}^T \times_3 \mathbf{S}^T + \mathcal{Z} \times_1 \mathbf{W}^T. \quad (5)$$

There are two basic multiway models largely used in the literature: Tucker3 and PARAFAC. The first is essential for data compression as shown in [12], but its uniqueness cannot be ensured in general. The second is identifiable under mild conditions related with the concept of the *Kruskal-rank* [10], [13].

III. Channel Estimation via Multiway Compressive Sensing

In this section, we propose a channel estimator method that leverages on the PARAFAC uniqueness properties and the joint sparse of the transmitter-receiver angular and delay domains.

A. Sparse PARAFAC Formulation

Consider a geometric channel model for a frequency selective channel

$$\begin{aligned} \mathbf{H} &= \sum_{d=0}^{N_c-1} \mathbf{H}_d \\ &= \sum_{d=0}^{N_c-1} \alpha_d g(dT_s - \tau_d) \mathbf{v}_d(\theta) \mathbf{v}_d(\phi)^H, \end{aligned} \quad (6)$$

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

The channel model in (6) can be conveniently represented as a third-order tensor by concatenating the $\mathbf{H}_d$ to form a cube. This results in a PARAFAC model to which the core tensor is the identity tensor $\mathcal{I}$, the factor matrices $\mathbf{V}_r \in \mathbb{C}^{N_r \times N_c}$ and $\mathbf{V}_t \in \mathbb{C}^{N_t \times N_c}$ contain the steering vectors $\mathbf{v}_d(\theta)$ and $\mathbf{v}_d(\phi)\ \forall\ d = \{0, \ldots, N_c - 1\}$, and $\mathbf{D} = \text{diag}\{\alpha_0 g(nT_s - \tau_0), \ldots, \alpha_{N_c-1} g(nT_s - \tau_{N_c-1})\}$. The spatio-delay channel response can be written as

$$\mathcal{H} = \mathcal{I} \times_1 \mathbf{V}_r \times_2 \mathbf{V}_t^* \times_3 \mathbf{D}. \quad (7)$$

Substituting (7) into (5), the received signal fits into a PARAFAC model up to an error accounted in $\mathcal{E} = \mathcal{Z} \times \mathbf{W}^H$

$$\mathcal{Y} = \mathcal{I} \times_1 \mathbf{W}^T \mathbf{V}_r \times_2 \mathbf{F}^T \mathbf{V}_t^* \times_3 \mathbf{S}^T \mathbf{D} + \mathcal{E}. \quad (8)$$

Comparing (7) and (8), the matrices $\mathbf{W}^T$, $\mathbf{F}^T$, and $\mathbf{S}^T$ compress 1-mode, 2-mode, and 3-mode, respectively. Therefore, the received tensor signal is the compressed version of the channel. Assuming that the tensor channel $\mathcal{H}$ admits a sparse representation, the received signal can again be rewritten as

$$\mathcal{Y} = \mathcal{I} \times_1 \mathbf{W}^T \mathbf{\Phi}_1 \mathbf{B}_1 \times_2 \mathbf{F}^T \mathbf{\Phi}_2 \mathbf{B}_2 \times_3 \mathbf{S}^T \mathbf{\Phi}_3 \mathbf{B}_3 + \mathcal{E}. \quad (9)$$

where $\mathbf{\Phi}_1 \in \mathbb{C}^{N_r \times N_r}$, $\mathbf{\Phi}_2 \in \mathbb{C}^{N_t \times N_t}$, and $\mathbf{\Phi}_3 \in \mathbb{C}^{T \times T}$, are orthogonal bases to the tensor modes, and $\mathbf{B}_1 \in \mathbb{C}^{N_t \times N_c}$, $\mathbf{B}_2 \in \mathbb{C}^{N_r \times N_c}$, and $\mathbf{B}_3 \in \mathbb{C}^{T \times N_c}$ are sparse matrices associated to each tensor mode. Given that $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{B}_3$ are sparse, consider that m_1 (m_2, m_3) is an upper bound on the number of non-zero elements per column of $\mathbf{B}_1$ (respectively $\mathbf{B}_2, \mathbf{B}_3$).

To recover the CSI, both the channel and received signal tensors must obey the uniqueness condition of the PARAFAC model. The *Kruskal-rank* of $\mathbf{V}_r$, denoted as $k_{\mathbf{V}_r}$, is the maximum k such that any k columns of $\mathbf{V}_r$ are linearly independent ($k_{\mathbf{V}_r} < r_{\mathbf{V}_r} \equiv \text{rank}(\mathbf{V}_r)$). Given the channel tensor $\mathcal{H}$, if $k_{\mathbf{V}_r} + k_{\mathbf{V}_t} + k_{\mathbf{D}} \geq 2N_c + 2$, then $(\mathbf{V}_r, \mathbf{V}_t, \mathbf{D})$ are unique up to a common column permutation and scaling [10]. The uniqueness condition of the received tensor signal is given by [1]

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

$$\min(Q, k_{\mathbf{B}_1}) + \min(I, k_{\mathbf{B}_2}) + \min(P, k_{\mathbf{B}_3}) \geq 2N_c + 2, \quad (10)$$

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 the Theorem 3.1, the system can properly choose the number of receiver beams, transmit beams, pilot symbols that guarantee uniqueness of the model. Assume that sparse matrices meet the condition $r_{\mathbf{B}_1} = k_{\mathbf{B}_1}$, $r_{\mathbf{B}_2} = k_{\mathbf{B}_2}$, and $r_{\mathbf{B}_3} = k_{\mathbf{B}_3}$, i.e. the factor matrices can not have a column that is a scaled version of another one. This means that the two distinct paths can not have the neither the same spatial signal signature nor the same propagation delay. Under such an assumption we can state important corollaries.

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

Another implication of the Theorem 3.1 asserts that 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 has available the channel $\mathcal{H}$. It is important to highlight that if we decide to neglect the channel low-rank structure and attempt to estimate the sparse vector formed by $N_c m_1 m_2 m_3$ non-zero elements, then $QIP \geq 2N_c m_1 m_2 m_3$ must hold. Assuming $m_1 = m_2 = m_3 = m$, the vector problem formulation needs $2N_c m^3$ measurements while the problem exploiting the PARAFAC framework needs $8m^3$. Therefore, the number of measurements does not depend on the number of paths, but only the sparsity on the factor matrices. In principle, if we know that $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$, and $\mathbf{\Phi}_3$ returns the sparsest factor matrices, the receiver measures the channel $\mathcal{H}$ using the fewest amount of samples, so the system overhead achieves the its lower-bound. Of course, more accurate estimations are possible the more samples are collected as we show in the results.

B. Algorithm Description

We exploit the principle of alternating least-square (ALS) to fit the compressed PARAFAC model obtained from noisy observations. After this, we compute separately the minimum l_1-norm solution for each factor matrix.

The idea behind the ALS consist of updating a subset of PARAFAC parameters according. We define $\mathbf{A} = \mathbf{W}^T \mathbf{\Phi}_1 \mathbf{B}_1$, $\mathbf{B} = \mathbf{F}^T \mathbf{\Phi}_2 \mathbf{B}_2$, and $\mathbf{C} = \mathbf{S}^T \mathbf{\Phi}_3 \mathbf{B}_3$ as the factor matrices of $\mathcal{Y}$. We obtain them using least square (LS) criterion conditioned on estimation of the remaining parameters. More specifically, we exploit the unfolding representations

$$\mathbf{Y}_1 = (\mathbf{B} \odot \mathbf{C})\mathbf{A}^T \quad (11)$$

$$\mathbf{Y}_2 = (\mathbf{C} \odot \mathbf{A})\mathbf{B}^T \quad (12)$$

$$\mathbf{Y}_3 = (\mathbf{A} \odot \mathbf{B})\mathbf{C}^T \quad (13)$$

and recover $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ by calculating the pseudoinverse of the matrix resulting from the Katri-rao product. For instance, the factor matrix of 1-mode is $\hat{\mathbf{A}}^T = (\mathbf{B} \odot \mathbf{C})^\dagger \mathbf{Y}_1$; similarly, $\hat{\mathbf{B}}$ and $\hat{\mathbf{C}}$ are calculated. Repeat this procedure until the error $e = \|\hat{\mathbf{Y}}_1 - \mathbf{Y}_1\|_F^2 / \|\mathbf{Y}_1\|_F^2 < \sigma^2$; this criterion asserts the algorithm convergence. We summarize the steps of ALS in the Table 1.

Fitting the PARAFAC model to the compressed tensor $\mathcal{Y}$, the next step consist of solving three compressive sensing problems to estimate $\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{B}_3$. The mathematical formulation for the 1-mode is

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

where $\hat{\mathbf{a}} \equiv \text{vec}\{\hat{\mathbf{A}}\}$ and $\mathbf{b}_1 \equiv \text{vec}\{\mathbf{B}_1\}$. The problem in (14) is a least absolute shrinkage and selection operator (lasso) problem and can be solved using convex optimization solvers [14]. The 2-mode and 3-mode can be formulated similarly as in (14). The outcome of the three problems gives a sparse PARAFAC model with the factor matrices $\mathbf{B}_1$, $\mathbf{B}_2$, and $\mathbf{B}_3$ that can be mapped to the tensor $\mathcal{H}$ by using the mode product with $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2$, and $\mathbf{\Phi}_3$, i.e.

$$\hat{\mathcal{H}} = \mathcal{I} \times_1 \mathbf{\Phi}_1 \mathbf{B}_1 \times_2 \mathbf{\Phi}_2 \mathbf{B}_2 \times_3 \mathbf{\Phi}_3 \mathbf{B}_3. \quad (15)$$

The matrices $\mathbf{W}$, $\mathbf{F}$, and $\mathbf{S}$ are the measurement matrices, i.e. each one performs measurements of the channel $\mathcal{H}$ associated to their own modes. The optimum design is desired so the maximum channel compression is achieved, however, such a problem is not investigated in this paper. For the ease of explanation, the receiver uses $\mathbf{W} = \mathbf{W}_{RF}$, and the transmitter employs $\mathbf{F} = \mathbf{F}_{RF}$. They are obtained from a Bernoulli distribution which meets the constant modulus restriction of the analog beamforming. Nevertheless such a distribution is convenient, it is known from the literature that the Gaussian complex distribution provides the best performance in sparse problem frameworks, as discussed in [15]. Although the use of such a distribution does not lead to a typical analog beamforming matrix, it is possible to use this set of matrices if the analog and digital beamformers are designed such that the resulting matrices $\mathbf{W}$ and $\mathbf{F}$ have entries that follow Gaussian complex distribution.

Algorithm 1 ALS description

Initialize factor matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$.
while $e > \sigma^2$ **do**
  $\hat{\mathbf{A}}^T \leftarrow (\mathbf{B} \odot \mathbf{C})^\dagger \mathbf{Y}_1$
  $\hat{\mathbf{B}}^T \leftarrow (\mathbf{C} \odot \hat{\mathbf{A}})^\dagger \mathbf{Y}_2$
  $\hat{\mathbf{C}}^T \leftarrow (\mathbf{A} \odot \hat{\mathbf{B}})^\dagger \mathbf{Y}_3$
  $\hat{\mathbf{Y}}_1 \leftarrow (\hat{\mathbf{B}} \odot \hat{\mathbf{C}})\hat{\mathbf{A}}^T$
  $e \leftarrow \|\hat{\mathbf{Y}}_1 - \mathbf{Y}_1\|_F^2 / \|\mathbf{Y}_1\|_F^2$
end while

IV. Results

In this section, the performance of the proposed method is evaluated in terms of the normalized mean square error (NMSE) which is defined as

$$\text{NMSE} = \sum_{v=1}^{V} \frac{\|\hat{\mathcal{H}}_v - \mathcal{H}\|_F^2}{\|\mathcal{H}\|_F^2}, \tag{16}$$

where v are the montecarlo simulations. First, we apply ALS, described in 1, to obtain PARAFAC decomposition of the compressed tensor $\mathcal{Y}$. Second, we apply matching pursuit algorithm (e.g. [16]) to solve for the sparse recovery of each factor matrix, from each the channel parameters can be extracted.

We consider a scenario with $N_t = 64$ transmitter antennas and $N_r = 16$ receiver antennas. Both arrays are linear with half length separation. The frame length is $N = 20$ and the delay tap length is assumed to be $N_c = 3$.

Fig. 2 shows the estimation performance for $I = \{2, 6, 20, 32\}$ number of beams. The algorithms fails in estimating the channel for $I = 2$, for the corollary 2 is not satisfied. The number of taps $N_c = 3$ requires that the number of transmit beams $I \geq 6$, otherwise the uniqueness of the model is not guaranteed. Assuming such condition is achieved, the estimation improves the more transmit beams are used which leads to use of more training frames. Thus, the performance improvement comes at the cost of more overhead as expected.

Fig 3 shows another approach to improve the channel estimation. Instead of increasing the number of transmit beams, it is possible to use more receive receive beams Q that measures that channel within the frame period. This will require more RF chains to be used at the same time, so there is a trade-off among hardware complexity, system overhead, and accuracy so that system can offer good spectral efficiency (reduced overhead) but using transceivers with reduced number of combiners. Nevertheless the increasing of the number of transmit beams can afford to reduce the pilot sequence length, it cannot be $P < 6$ as it violates the first corollary; therefore, the uniqueness of the PARAFAC model does not hold.

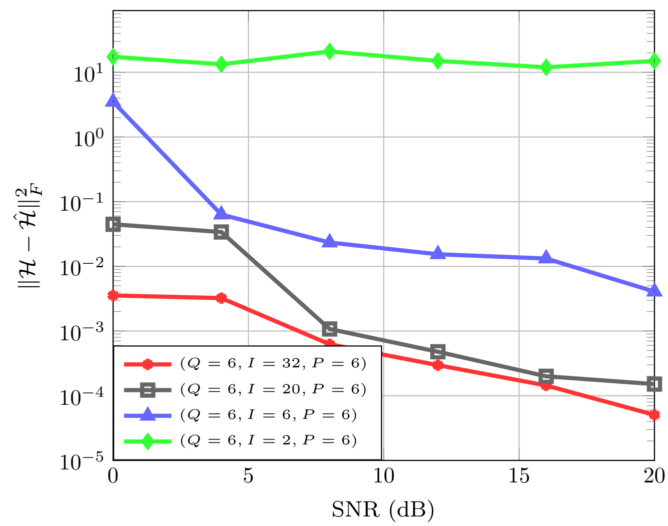

Fig. 2. The plots show the tensor-based algorithm performance when we vary the number of transmit beams.

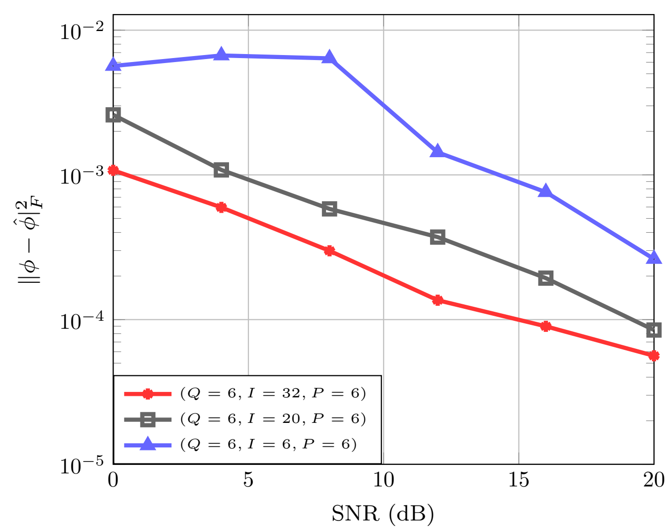

Fig. 3. The plots show the estimation error of the AoD

V. Conclusion

In this paper, we investigated the problem of estimating frequency selective channels under the assumption that transmitter and receiver have a hybrid architecture. The proposed solution exploits jointly the intrinsic sparsity of a three dimensional channel. Using the sparse PARAFAC model, we provides useful bounds on the minimum number of transmit beams, receive beams, and length of the pilot sequence that guarantee the uniqueness of the model. This offers flexibility to the system as it can conveniently control the system overhead according to the number of RF chains and spectral efficiency by choosing the number of beams and pilot sequence length. In future works, we intend to extend the proposed tensor-based compressive channel estimator to time varying channels whose time response changes from frame-to-frame.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

[10] 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.

[11] A. Cichocki, R. Zdunek, A. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley, 2009. [Online]. Available: http://books.google.com.br/books?id=8S4MRXMRdcMC

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

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

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

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

[16] 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