CHANNEL COVARIANCE IDENTIFICATION IN FDD MASSIVE MIMO SYSTEMS

José P. González-Coma[⋆] *Pedro Suárez-Casal*[†] *Paula M. Castro*[†] *Luis Castedo*[⋆] *Michael Joham*[‡]

[⋆] University of A Coruña, CITIC, Spain
[†] University of A Coruña, Department of Computer Engineering, Spain
[‡] Technische Universität München, Associate Institute for Signal Processing, Germany

ABSTRACT

Many channel estimation methods in Massive MIMO FDD systems usually rely on the knowledge of the channel covariance matrix to operate. However, in real scenarios, this covariance is not known beforehand and, hence, it should be estimated. In this work, we investigate different existing techniques for covariance identification to achieve full knowledge of this matrix with very short training sequences. Moreover, we propose a modification of the spatial smoothing approach with the goal of improving the quality of the channel covariance identification.

***Index Terms*—** Covariance identification, Massive MIMO, FDD, MUSIC

1. INTRODUCTION

Massive Multiple-Input and Multiple-Output (MIMO) is a solid candidate radio technology for future wireless communications due to its ability to provide large data rates [1]. The use of large antenna arrays characteristic of this technology entails numerous challenges related to the radiating system dimension. One of the main challenges is channel estimation in Frequency-Division Duplex (FDD) mode since a large number of antennas requires long training sequences [1].

In recent literature, different methods have been proposed to overcome these limitations with FDD such as exploiting the geometric characteristics of the channel [2], or grouping users with either similar local scatterers [3] or similar channel covariance matrices [4, 5]. Other solutions reduce the training overhead relying on the channel temporal correlation [6]. In general, these works assume the channel is sparse in some sense. Although it is generally assumed that the channel covariance matrix is known in advance, in practical circumstances it should also be estimated. The recent work [7] considers covariance identification in Time-Division Duplex (TDD) mode based on the assumption of channel reciprocity between the downlink and the uplink. Finally, [8, 9] consider the FDD case by exploiting the weaker assumption of angular reciprocity.

In this work, we will focus on covariance identification for FDD leveraging only on the assumption that the channel coefficients arise from a stationary process whose statistics slowly change with respect to the channel coherence time. We will also assume that the covariance matrix has a Toeplitz structure, which is satisfied by typical antenna arrangements like Uniform Linear Arrays (ULAs). Under this general assumption, training sequences can be designed according to the structure of a sparse ruler, which allows for employing training sequences with lengths in the order of the square root of the number of antennas [10]. Moreover, the common assumption of sparsity, used for channel estimation, is no longer necessary to obtain reasonable training lengths in covariance identification.

Another approach to covariance identification is based on the estimation of the Angle of Departures (AoDs) of the waves transmitted by the ULAs. Recent work addressing this problem finds solutions based on the MUltiple SIgnal Classification (MUSIC) algorithm [11–13]. However, for this scenario, it is particularly interesting to increase the angular resolution by means of spatial smoothing [14], thus allowing for very short training sequences. Moreover, the methods to obtain the power associated with each propagation path usually assume that there is a full-rank dictionary matrix [11] which may not be available in practice.

Therefore, in this work, we propose a variation of the spatial smoothing method presented in [15] that improves the quality of the AoDs estimation. In addition, our strategy exploits the covariance matrix structure to obtain a more general solution that does not impose a requirement of a full-rank dictionary matrix.

2. SYSTEM MODEL

Let us consider the downlink of a multiuser Massive MIMO system with M transmit antennas and several single-antenna receivers operating in FDD mode. Let $\boldsymbol{h}_k \in \mathbb{C}^{M\times 1}$ represent the channel response to an arbitrary user during the k-th training period, with $k \in \{1,\ldots,K\}$ and K being the to-

This work has been funded by Xunta de Galicia (ED431C 2016- 045, ED341D R2016/012, ED431G/01), AEI of Spain (TEC2015-69648-REDC, TEC2016-75067-C4-1-R), and ERDF funds (AEI/FEDER, EU).

The final published paper is available online at: https://doi.org/10.1109/GlobalSIP.2018.8646448

tal number of channel blocks used to estimate the channel covariance. We assume a block-fading channel where each channel realization remains constant during a channel block and is statistically independent from one channel block to another. Furthermore, the training sequence $\boldsymbol{X} \in \mathbb{C}^{T_{\text{tr}} \times M}$ is transmitted for each block, where T_{tr} is the number of channel uses devoted to the transmission of the training sequence. The remaining channel uses within the block are dedicated to data transmission. Finally, we will consider that $T_{\text{tr}} \ll M$, hence the columns of $\boldsymbol{X}$ are non-orthogonal.

We further consider the channel response vector $\boldsymbol{h}$ results from a linear combination of L array response vectors $\boldsymbol{a}(\alpha_l)$, each one associated to an Angle of Arrival (AoA) α_l. This can be approximated by means of a steering vector dictionary $\boldsymbol{A} = [\boldsymbol{a}(\theta_1), \ldots, \boldsymbol{a}(\theta_G)] \in \mathbb{C}^{M \times G}$ for a set of G predefined angles, that is,

$$\boldsymbol{h} = \sum_{l=1}^{L} g_l \boldsymbol{a}(\alpha_l) \approx \sum_{i=1}^{G} \tilde{g}_i \boldsymbol{a}(\theta_i), \tag{1}$$

where $\tilde{g}_i$ are the channel gains with L nonzero elements, and $E\left[\tilde{g}_i \tilde{g}_j^*\right] = \sigma_i^2 \delta_{i-j}$. Note that equality in (1) holds if the AoAs α_l lie on the angles chosen for the dictionary $\{\theta_1, \ldots, \theta_G\}$. Hence, the channel covariance matrix can be written as

$$\boldsymbol{C_h} \approx \sum_{i=1}^{G} \sigma_i^2 \boldsymbol{a}(\theta_i) \boldsymbol{a}^H(\theta_i) = \boldsymbol{A} \boldsymbol{D} \boldsymbol{A}^H, \tag{2}$$

where the diagonal matrix $\boldsymbol{D} = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_G^2)$ contains the channel gain variances. Observe that the former matrix $\boldsymbol{C_h}$ is Toeplitz due to the assumed array geometry.

When the vector training sequence is transmitted, the received signal at the single-antenna user is

$$\boldsymbol{\phi} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{v} \quad \in \mathbb{C}^{T_{\text{tr}} \times 1}, \tag{3}$$

where $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{C_v})$ is the Additive White Gaussian Noise (AWGN). Since $\boldsymbol{h}$ and $\boldsymbol{v}$ are independent, and if the training sequence is the same for all training periods, the covariance of $\boldsymbol{\phi}$ can be written as

$$\boldsymbol{C_\phi} = \boldsymbol{X} \boldsymbol{C_h} \boldsymbol{X}^H + \boldsymbol{C_v} \approx \boldsymbol{F} \boldsymbol{D} \boldsymbol{F}^H + \boldsymbol{C_v}. \tag{4}$$

where we have used the approximation in (2) and $\boldsymbol{F} = \boldsymbol{X}\boldsymbol{A} = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_G]$. Assuming $\boldsymbol{C_v}$ known, estimating $\boldsymbol{C_\phi}$ is equivalent to estimating $\boldsymbol{D}$. Moreover, we will assume that the noise is spatially white, i.e. $\boldsymbol{C_v} = \sigma_{\boldsymbol{v}}^2 \boldsymbol{I}$. Finally, notice that K realizations of (3), denoted as $\{\boldsymbol{\phi}_k = \boldsymbol{X}\boldsymbol{h}_k + \boldsymbol{v}_k\}_{k=1}^K$, will be available at the receiver. Since $\boldsymbol{h}$ is stationary, these K training blocks will be used to estimate the channel covariance $\boldsymbol{C_h}$.

3. COVARIANCE IDENTIFICATION

In this section, we will consider several methods for covariance identification. Each algorithm is particularly suited for some scenario, depending on the noise covariance, channel sparsity, and computational complexity.

3.1. ML Estimation

Maximum Likelihood (ML) estimation is based on maximizing the log-likelihood function. For given $\boldsymbol{\phi}_k$, the minus log-likelihood function of $\boldsymbol{D}$ is

$$L\left(\boldsymbol{\phi}_k; \boldsymbol{D}\right) = \log\det\left(\boldsymbol{C_\phi}\right) + \operatorname{tr}\left(\boldsymbol{C_\phi}^{-1} \boldsymbol{\phi}_k \boldsymbol{\phi}_k^H\right). \tag{5}$$

and the joint likelihood function is $L(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K; \boldsymbol{D}) = \sum_{k=1}^K L(\boldsymbol{\phi}_k; \boldsymbol{D})$ because $\{\boldsymbol{\phi}_k\}_{k=1}^K$ are independent. The ML estimate of $\boldsymbol{D}$ is the result of the following optimization problem

$$\min_{\boldsymbol{D}} \; L(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K; \boldsymbol{D}) \quad s.t. \;\; \sigma_l^2 \geq 0, \forall l \in [1, G]. \tag{6}$$

The log-likelihood function (5) can be expressed for a block of K snapshots as

$$\frac{1}{K} L\left(\boldsymbol{\Phi}; \boldsymbol{D}\right) = \log\det\left(\boldsymbol{C_\phi}\right) + \operatorname{tr}\left(\boldsymbol{C_\phi}^{-1} \hat{\boldsymbol{C}}_{\boldsymbol{\phi}}\right), \tag{7}$$

where $\hat{\boldsymbol{C}}_{\boldsymbol{\phi}} = \frac{1}{K} \boldsymbol{\Phi}\boldsymbol{\Phi}^H$ with $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K]$. This is a difficult problem because it involves the sum of a concave and a convex function, and it has been subject of extensive research.

Multiple algorithms have been proposed to solve (6) (e.g., [16]) but one of the most robust is the LIKelihood-based Estimation of Sparse parameters (LIKES) algorithm [17] which is based on the majorization-minimization principle. This algorithm assumes that the covariance can be written as the linear combination of a set of matrices, in the present case $\boldsymbol{C_\phi} \approx \sum_{i=1}^G \sigma_i^2 \boldsymbol{f}_i \boldsymbol{f}_i^H + \sigma_{\boldsymbol{v}}^2 \boldsymbol{I}$, and estimates the eigenvalues σ_i^2. However, the LIKES algorithm suffers from a large computational complexity.

3.2. OMP-based estimator

Compressive sensing algorithms have been extensively analyzed in the context of channel estimation for Massive MIMO FDD. An example of compressive covariance identification is the variation of the Orthogonal Matching Pursuit (OMP) algorithm known as Covariance OMP (COMP) [18]. This algorithm relies on the more general assumption that $\boldsymbol{D}$ is a sparse hermitian matrix, instead of diagonal, and the adaptation of OMP to the quadratic case. The COMP algorithm greedily selects columns of $\boldsymbol{F}$ to obtain $\boldsymbol{F}_{\mathcal{S}} = [\boldsymbol{f}_{\mathcal{S}_1}, \ldots, \boldsymbol{f}_{\mathcal{S}_L}]$, which stacks the columns for a given integer index subset $\mathcal{S} \subseteq [1, G]$. Then, for fixed $\boldsymbol{F}_{\mathcal{S}}$, the channel gain variances are estimated by solving

$$\hat{\boldsymbol{D}}_{\mathcal{S}} = \underset{\boldsymbol{D}:\boldsymbol{D}=\boldsymbol{D}^H}{\arg\min} \; \left\| \hat{\boldsymbol{C}}_{\boldsymbol{\phi}} - \boldsymbol{F}_{\mathcal{S}} \boldsymbol{D} \boldsymbol{F}_{\mathcal{S}}^H \right\|_F^2 \tag{8}$$

$$= \boldsymbol{F}_{\mathcal{S}}^{\dagger} \hat{\boldsymbol{C}}_{\boldsymbol{\phi}} (\boldsymbol{F}_{\mathcal{S}}^H)^{\dagger}. \tag{9}$$

Finally, $\hat{\boldsymbol{D}}$ is computed as a matrix of zeros except for elements at the rows and colums selected by $\mathcal{S}$ which are set to $[\hat{\boldsymbol{D}}]_{\mathcal{S},\mathcal{S}} = \hat{\boldsymbol{D}}_{\mathcal{S}}$.

3.3. MUSIC

MUSIC algorithm is a well-known AoA estimation algorithm based on the orthogonality of signal and noise subspaces. In our particular scenario, the MUSIC algorithm will be employed to identify the angles corresponding to the L channel paths. The first step of MUSIC consists in computing the eigendecomposition of the sample covariance matrix $\hat{\boldsymbol{C}}_{\boldsymbol{\phi}} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$ assuming that this covariance converges to the channel second order moment for a large number of samples, i.e., $\hat{\boldsymbol{C}}_{\boldsymbol{\phi}} \rightarrow \boldsymbol{C}_{\boldsymbol{\phi}}$ when $K \rightarrow \infty$. Next, the eigenvectors corresponding to the L largest eigenvalues are discarded, that is, we obtain the basis spanning the noise subspace $\bar{\boldsymbol{U}}$.

The second step of MUSIC performs the support identification. To that end, we define the estimator function as

$$\mathcal{J}_i = \frac{1}{\|\boldsymbol{f}_i^H \bar{\boldsymbol{U}}\|_2^2}, \tag{10}$$

The angles are identified as the L largest values of $\mathcal{J}_i$. Next, we construct the matrix $\boldsymbol{F}_{\mathcal{S}}$ containing the L columns of $\boldsymbol{F}$ corresponding to the estimated angles θ_i. Once the angles are determined, it is possible to obtain the associated gain variances. If the matrix $\boldsymbol{F}$ is full rank, these gains can be estimated as in [11], i.e.,

$$\hat{\boldsymbol{D}}_{\mathcal{S}} = \boldsymbol{F}_{\mathcal{S}}^{\dagger}(\hat{\boldsymbol{C}}_{\boldsymbol{\phi}} - \sigma_{\boldsymbol{v}}^2 \boldsymbol{I})(\boldsymbol{F}_{\mathcal{S}}^H)^{\dagger}. \tag{11}$$

This simple and intuitive approximation has some limitations. For short training sequences, $T_{\text{tr}} < L$, the pseudo-inverse does not exist. This scenario will be analyzed in the ensuing section. Moreover, for $T_{\text{tr}} > L$ it is possible that the calculation of (11) results in numerical difficulties. Recall that, for finite M, the dictionary contains non-orthogonal vectors. When the distance between two consecutive angles in the dictionary is small, which arises when G is large, this may lead to dictionary matrices such that $\operatorname{rank}(\boldsymbol{F}) \leq \min\{M, G\}$, thus making it possible to obtain matrices $\boldsymbol{F}_{\mathcal{S}}$ with arbitrarily small minimum singular value.

To circumvent this limitation, the dictionary size can be reduced to increase the distance between consecutive angles. However, this would also reduce the angular resolution resulting in poorer covariance estimations. Alternatively, we propose to exploit the diagonal structure of $\boldsymbol{D}$ and perform the estimation by means of

$$\operatorname{diag}(\hat{\boldsymbol{D}}_{\mathcal{S}}) = (\boldsymbol{F}_{\mathcal{S}}^* \circ \boldsymbol{F}_{\mathcal{S}})^{\dagger} \operatorname{vec}(\hat{\boldsymbol{C}}_{\boldsymbol{\phi}} - \sigma_{\boldsymbol{v}}^2 \boldsymbol{I}), \tag{12}$$

where the Khatri-Rao product, i.e. the column-wise Kronecker product, produces matrices of rank L when the columns of $\boldsymbol{F}_{\mathcal{S}}$ are not co-linear.

3.4. Spatial Smoothing (SS)

A well-known limitation of MUSIC is that it can only work if the sample covariance has a rank equal to at least the number of elements to estimate. That is, the number of paths (or angles) to be estimated is bounded by the number of snapshots and the number of observations, i.e., $L \leq \min\{T_{\text{tr}}, K\}$. Improvements to MUSIC were proposed in [12,13] to overcome the rank deficiency limitation. In particular, these solutions apply when the number of snapshots is not sufficient, or when the snapshots are linearly dependent, that is, $\operatorname{rank}(\hat{\boldsymbol{C}}_{\boldsymbol{\phi}}) < L$. Nevertheless, in the proposed scenario, the number of snapshots is not a stringent limitation since the covariance slowly changes over time. Hence, it is possible to employ more observations to obtain L linear independent received signals $\boldsymbol{\phi}$.

On the contrary, it is desirable to avoid the large training overhead. By reducing the training sequence length T_{tr}, the remaining of the channel coherence time can be used for data transmission. This strategy, however, may lead to scenarios where the rank of the sample covariance matrix $\operatorname{rank}(\hat{\boldsymbol{C}}_{\boldsymbol{\phi}}) = T_{\text{tr}}$ is smaller than the number of propagation paths, that is $T_{\text{tr}} < L$. In this involved setup, neither MUSIC nor COMP is applicable to identify the L spatial directions. We propose to address this interesting scenario by using spatial smoothing.

Spatial smoothing was originally based on nested arrays with different antenna separation [14]. A similar effect can be obtained with the use of sparse rulers of length G as training sequences [15], which is equivalent to deploying non-uniform distance antenna elements. Applied to our scenario, the training sequence is built as $\boldsymbol{X} = [\boldsymbol{e}_{r_1}, \ldots, \boldsymbol{e}_{r_T}]^T$, where $\boldsymbol{e}_i$ is a vector of zeros with a one in the i-th element and $\{r_1, \ldots, r_T\}$ are the marks of the ruler [10]. Thus, the sample covariance $\hat{\boldsymbol{C}}_{\phi}$ approximates the true covariance for K sufficiently large, and therefore we can write the approximation $\hat{\boldsymbol{C}}_{\boldsymbol{\phi}} \approx \boldsymbol{F}\boldsymbol{D}\boldsymbol{F}^H + \sigma_{\boldsymbol{v}}^2 \boldsymbol{I}$. This expression can be vectorized as

$$\boldsymbol{y} = \operatorname{vec}(\hat{\boldsymbol{C}}_{\boldsymbol{\phi}}) \approx (\boldsymbol{F}^* \circ \boldsymbol{F}) \operatorname{diag}(\boldsymbol{D}) + \sigma_{\boldsymbol{v}}^2 \boldsymbol{e}', \tag{13}$$

where $\boldsymbol{e}' = [\boldsymbol{e}_1^T, \ldots, \boldsymbol{e}_{T_{\text{tr}}}^T]^T$. Observe that $(\boldsymbol{F}^* \circ \boldsymbol{F})$ contains the products corresponding to the spatial differences $z \in \{-M+1, M-1\}$. Spatial smoothing discards the repeated distances and sorts the remaining ones in a matrix $\boldsymbol{B}$ containing $2M-1$ rows from $(\boldsymbol{F}^* \circ \boldsymbol{F})$. The reduced vector is then

$$\check{\boldsymbol{y}} \approx \boldsymbol{B} \operatorname{diag}(\boldsymbol{D}) + \sigma_{\boldsymbol{v}}^2 \boldsymbol{e}_M. \tag{14}$$

The $2M-1$ differences in the reduced vector $\check{\boldsymbol{y}} \in \mathbb{C}^{2M-1}$ are next seen as a phase shift of the M differences to be estimated. Considering M overlapping subarrays, such that the m-th subarray comprises the differences $\{-M+m, m-1\}$ in $\check{\boldsymbol{y}}_m \in \mathbb{C}^M$, the spatial smoothed matrix is obtained by averaging over the M subarrays

$$\check{\boldsymbol{Y}} = \frac{1}{M} \sum_{m=1}^{M} \check{\boldsymbol{y}}_m \check{\boldsymbol{y}}_m^H. \tag{15}$$

As shown in [14], $\check{\boldsymbol{Y}} = \check{\boldsymbol{Y}}^{1/2}\check{\boldsymbol{Y}}^{1/2}$, where $\check{\boldsymbol{Y}}^{1/2}$ is

$$\check{\boldsymbol{Y}}^{1/2} = \frac{1}{\sqrt{M}}(\boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^H + \sigma_{\boldsymbol{v}}^2 \mathbf{I}_M). \tag{16}$$

Since the result of applying spatial smoothing is a linear combination of the steering vectors, it is possible to identify the angles by using algorithms like MUSIC. After computing the eigendecomposition of $\sqrt{M}\check{\boldsymbol{Y}}^{1/2} = \boldsymbol{U}_k\boldsymbol{\Lambda}_k\boldsymbol{U}_k^H$, the L eigenvectors corresponding to the L largest eigenvalues are discarded. Thus, we obtain an incomplete basis $\bar{\boldsymbol{U}}_k$ spanning the noise subspace and the estimator function as

$$\mathcal{J}(\theta_i) = \frac{1}{\|\boldsymbol{a}(\theta_i)^H\bar{\boldsymbol{U}}\|_2^2}. \tag{17}$$

Following the procedure explained for the MUSIC algorithm, we estimate the gains using a similar procedure to (12).

A direct improvement of this method consists of employing all the differences in the vector $\boldsymbol{y}$ instead of discarding them. Notice that in (13) the vector $\boldsymbol{y} = \mathrm{vec}(\hat{\boldsymbol{C}}_{\boldsymbol{\phi}})$ has T_{tr}^2 elements, whereas in (14) only $2M-1$ elements are considered for $\check{\boldsymbol{y}}$. Let us introduce y_i^j as the element of $\boldsymbol{y}$ denoting the j-th snapshot of the i-th difference. In this case, we can redefine the vector $[\check{\boldsymbol{y}}]_i = \frac{1}{J}\sum_{j=1}^{J} y_i^j$, where J represents the number of snapshots available for the i-th difference, which depends on the structure of the sparse ruler. This modified spatial smoothing will be referred to as Improved Spatial Smoothing (ISS) in the following.

4. SIMULATION RESULTS

The following setup is considered for the numerical experiments. The number of transmit antennas is $M = 400$, and the dictionary size is $G = 400$ with equally spaced angles within the range $(\pi/32, 31\pi/32)$. We consider the channel covariance model of an ULA assuming, as in (1), L different channel paths. The training sequence was generated from a Wichmann ruler and its length is $T_{\mathrm{tr}} = 50$. The numerical results are averaged over 400 channel realizations.

Fig. 1 shows the Normalized Mean Squared Error (NMSE) for the channel covariance matrix, defined as

$$\frac{\|\hat{\boldsymbol{C}}_{\boldsymbol{h}} - \boldsymbol{C}_{\boldsymbol{h}}\|_F^2}{\|\boldsymbol{C}_{\boldsymbol{h}}\|_F^2},$$

and determined for the algorithms described in the previous section for a Signal-to-Noise Ratio (SNR) of 0 dB and 30 dB, and $K = \{5, 10, 20, 50, 100\}$. It is remarkable the robustness of MUSIC compared to COMP for the low SNR regime. ML is the best algorithm in terms of NMSE, but it is also the most computationally expensive. Moreover, MUSIC performs better if the number of snapshots is large enough. Taking into account the spatial smoothing, SS and ISS curves show the performance gain of our approach with respect to the standard one.

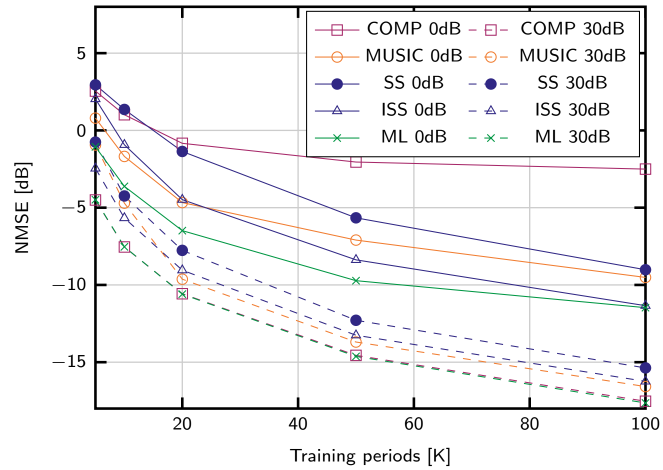

Fig. 1. NMSE of different covariance identification strategies for $N = 400$ antennas and $L = 15$.

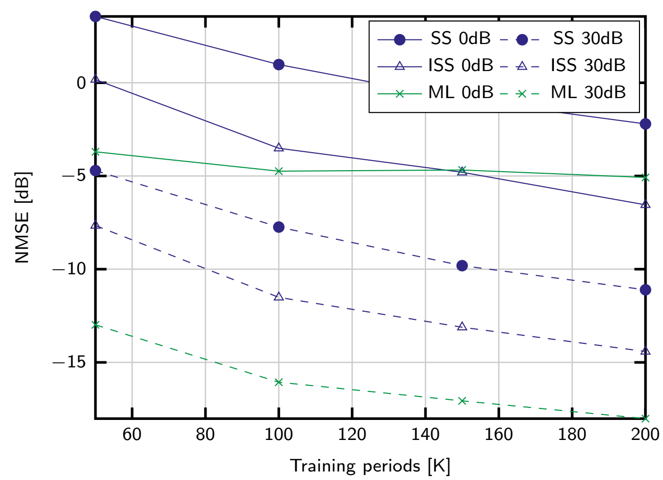

Fig. 2. NMSE of different covariance identification strategies for $N = 400$ antennas and $L = 70$.

Fig. 2 shows the NMSE for $L = 70$ channel propagation paths and SNR levels. In this case, $K = \{50, 100, 150, 200\}$ because the number of paths to estimate is much larger. MUSIC and COMP do not apply to this scenario due to the large number of paths. Therefore, we compare SS and ISS with ML, which again exhibits the best performance.

5. CONCLUSIONS

We have investigated several algorithms for covariance identification in a Massive MIMO FDD scenario. We have shown that it is possible to identify the covariance matrix even if the number of channel paths is much larger than the length of the training sequences. Finally, we have also modified the spatial smoothing method to substantially improve the estimation quality in terms of NMSE.

6. REFERENCES

[1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, January 2013.

[2] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially Common Sparsity Based Adaptive Channel Estimation and Feedback for FDD Massive MIMO," *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6169–6183, December 2015.

[3] X. Rao and V. K. N. Lau, "Distributed Compressive CSIT Estimation and Feedback for FDD Multi-User Massive MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3261–3271, June 2014.

[4] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, Oct 2013.

[5] B. Zhang, L. J. Cimini, and L. J. Greenstein, "Efficient Eigenspace Training and Precoding for FDD Massive MIMO Systems," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec 2017, pp. 1–6.

[6] Y. Han, J. Lee, and D. J. Love, "Compressed sensing-aided downlink channel training for fdd massive mimo systems," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 2852–2862, July 2017.

[7] D. Neumann, M. Joham, and W. Utschick, "Covariance Matrix Estimation in Massive MIMO," *IEEE Signal Processing Letters*, vol. 25, no. 6, pp. 863–867, June 2018.

[8] H. Xie, F. Gao, S. Jin, J. Fang, and Y. Liang, "Channel Estimation for TDD/FDD Massive MIMO Systems With Channel Covariance Computing," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 4206–4218, June 2018.

[9] Mahdi Barzegar Khalilsarai, Saeid Haghighatshoar, Xinping Yi, and Giuseppe Caire, "FDD massive MIMO via UL/DL channel covariance extrapolation and active channel sparsification," *CoRR*, vol. abs/1803.05754, 2018.

[10] D. Romero and G. Leus, "Compressive covariance sampling," in *2013 Information Theory and Applications Workshop (ITA)*, Feb 2013, pp. 1–8.

[11] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, March 1986.

[12] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3613–3641, June 2012.

[13] J. M. Kim, O. K. Lee, and J. C. Ye, "Compressive music: Revisiting the link between compressive sensing and array signal processing," *IEEE Transactions on Information Theory*, vol. 58, no. 1, pp. 278–301, Jan 2012.

[14] P. Pal and P. P. Vaidyanathan, "Nested Arrays: A Novel Approach to Array Processing With Enhanced Degrees of Freedom," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4167–4181, August 2010.

[15] Dyonisius Dony Ariananda and Geert Leus, "Direction of arrival estimation for more correlated sources than active sensors," *Signal Processing*, vol. 93, no. 12, pp. 3435 – 3448, 2013.

[16] D. Romero and G. Leus, "Wideband spectrum sensing from compressed measurements using spectral prior information," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6232–6246, Dec 2013.

[17] P. Babu and P. Stoica, "Sparse spectral-line estimation for nonuniformly sampled multivariate time series: Spice, likes and msbl," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Aug 2012, pp. 445–449.

[18] Sungwoo Park and Robert W. Heath Jr., "Spatial Channel Covariance Estimation for the Hybrid MIMO Architecture: A Compressive Sensing Based Approach," *CoRR*, vol. abs/1711.04207, 2017.