Low-Rank Channel Estimation for mm-Wave Multiple Antenna Systems using Joint Spatio-Temporal Covariance Matrix

Kun Chen-Hu[1], Dirk T.M. Slock[2] and Ana Garcia Armada[1]
[1]Department of Signal Theory and Communications, Universidad Carlos III de Madrid (Spain)
[2]Department of System of Communications, Eurecom (France)
E-mail: kchen@tsc.uc3m.es, slock@eurecom.fr and agarcia@tsc.uc3m.es

***Abstract*—Millimeter-Wave (mm-Wave) and very large multiple antenna systems (VLMAS) are two key technologies in the deployment of Fifth Generation (5G) mobile communication systems. In order to exploit all the benefits of VLMAS, spatial and temporal (ST) features must be estimated and exploited to compute the precoding/decoding matrices. In the literature, a practical channel estimation approach is proposed by assuming that the spatial features are completely unknown, leading to non-parametric estimation in which antenna array calibration is not required. Additionally, when the signal-to-noise ratio (SNR) is not so high, a low-rank (LR) version of the estimated channel is proposed that provides better performance than the full-rank (FR) one in terms of bias-variance trade-off in the mean squared error (MSE). However, previous work assumes that spatial and temporal characteristics of the channel can be estimated separately. Then, the performance is degraded in realistic channels. In this paper, we propose an alternative way to characterize the FR estimated channel using a joint ST covariance matrix, combined with a low-complexity semi-parametric spatial response and delay estimation technique. Moreover, we propose an automatic rank-selector (ARS) based on the MSE in order to provide the best LR channel estimation for each scenario. Numerical results show that the proposed technique outperforms existing approaches in the literature.**

I. Introduction

The new Fifth Generation (5G) of mobile communications is starting to be deployed. This new air interface will revolutionize the speed of data links, enabling new applications with a great data consumption. In order to accomplish this goal, millimeter-Wave (mm-Wave) [1] frequency bands are ideal for 5G due to the fact that an abundant spectrum is available. However, the transmission at these bands suffer from significantly higher path loss and susceptibility to blockage. Hence, very large multiple antenna systems (VLMAS) [2] are required in order to improve the reliability of the link. Furthermore, in order to fully-exploit all the benefits of VLMAS, coherent demodulation schemes are adopted where the precoding/decoding matrices, for compensating the effects of the channel and interference, are computed thanks to an accurate knowledge of the channel in the spatial and temporal (ST) domains.

In the literature, there are several efficient channel estimation techniques for mm-Wave VLMAS, such as: compressive sensing [3] and subspace methods [4]. However, a common issue of all techniques consists in assuming a parametric approach for the direction of arrival (DoA) of each tap of the frequency-selective channel. This fact implies the need of a calibration process of the system. Morever, previous works does not adapt well to the different signal-to-noise ratio (SNR) conditions that may affect the dynamic propagation environment in mm-Wave. Even though the VLMAS will compensate the large and changing path loss, the channel must be estimated before this compensation is effective.

One of the most appealing proposal given in [5], where it proposes a subspace method assuming that the spatial features are completely unknown, leading to non-parametric estimation in which antenna array calibration is not required. Furthermore, it exploits a low-rank (LR) algebraic structure of the channel by projecting the estimated channel on the ST covariance matrices. When the signal-to-noise ratio (SNR) is not so high, the LR version of the estimated channel provides a better performance than the full-rank (FR) one in terms of bias-variance trade-off in the mean squared error (MSE). However, this method is not accurate enough due to the fact that they assumed that there is no correlation between spatial and temporal channel behavior. This assumption holds if and only if the receiving signal of all taps of the multi-path channel, with different time of arrival (ToA), have the same DoA at the base station (BS), which is not true in realistic scenarios.

In our work, we will improve [5], starting from the same system model and assumptions, improving the performance of the LR estimated channel. A multi-slot scenario is used, where the slow-varying components of the channel response (angles and delays) are estimated, while the fast-varying ones (fading coefficients) are tracked at each slot. Unlike [5], we obtain a joint ST covariance matrix, which is capable of obtaining the full ST features of the VLMAS. Moreover, given the computed joint ST covariance matrix and applying a subspace method, we propose three estimation techniques of DoA and ToA considering a semi-parametric spatial response and delay estimation. The first method is based on one dimension (1D) multiple signal classification (MUSIC) [6] algorithm, that it is a low-complexity method that provides a very good performance in high SNR scenarios. The second method is based on iterative maximum likelihood estimation (MLE) [7], and it provides a very good performance in any scenario. However, its complexity is higher than the previous one. The third method is the hybrid one, where it combines the previous two methods and is capable of obtaining a good performance with a reduced complexity. Finally, we also propose an automatic rank selection (ARS) method which is able to select the best LR channel for each scenario based on the computation of the mean square error (MSE).

The remainder of the paper is organized as follows. Section II provides the model of the considered VLMAS. Section III describes the estimation of the joint ST covariance matrix. Section IV provide the different methods for the estimation of DoA and ToA. Section V describes the ARS method. Section VI presents some numerical results to verify our theoretical analysis and provides a better understanding of the system performance. Finally, in section VII, some conclusions are pointed out.

Notation: matrices, vectors and scalar quantities are denoted by boldface uppercase, boldface lowercase, and normal letters, respectively. $\mathbf{I}_M$ is the identity matrix of size $(M \times M)$. $\mathbf{0}_{M \times N}$ is the zero matrix of size $(M \times N)$. $\text{tr}\{\cdot\}$ denotes the matrix trace operation. The superscript $(\cdot)^H$ denotes Hermitian. $\otimes$ denotes the Kronecker product of two matrices. $\odot$ denotes the Khatri-Rao product of two matrices. $\mathbb{E}\{\cdot\}$ represents the expected value. $\text{Var}\{\cdot\}$ denotes the variance. $\mathcal{CN}(0,\sigma^2)$ represents the circularly-symmetric and zero-mean complex normal distribution with variance σ^2. $||\cdot||_F^2$ denotes the squared Frobenius norm. $\mathbf{P}_A = \mathbf{A}\left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H$is the matrix projector of $\mathbf{A}$ and $\mathbf{P}_A^{\perp} = \mathbf{I}_M - \mathbf{P}_A$ is its orthogonal matrix projector.

II. System Model

We consider a time-slotted wireless communication system where a single-antenna user equipment (UE) transmits data to a BS equipped with N_r antennas. The received signal at the BS in the l-th pilot sequence $\mathbf{Y}(l)\,(N_r \times N_s)$ out of N_l pilot sequences is given by

$$\mathbf{Y}(l) = \mathbf{H}(l)\,\mathbf{X}(l) + \mathbf{N}(l)\,, \tag{1}$$

where $\mathbf{H}(l)\,(N_r \times N_t)$ is the block-fading frequency-selective single-input multiple-output (SIMO) channel with temporal support N_t, $\mathbf{X}(N_t \times N_s)$ denotes the training sequence of length N_s organized in a form of N_t one sample-delayed replicas and $\mathbf{N}(l)\,(N_r \times N_s) = \begin{bmatrix}\mathbf{n}_1(l) & \cdots & \mathbf{n}_{N_r}(l)\end{bmatrix}$ denotes the additive white Gaussian noise (AWGN), and we assume that the noise is temporally and spatially uncorrelated. Hence, $\mathbb{E}\left\{\mathbf{n}_i(l)\,\mathbf{n}_{i+m}^H(l)\right\} = \delta(m)\,\sigma_n^2\mathbf{I}_{N_r}$, where we consider that all antennas have the same noise power σ_n^2.

According to [5], the channel $\mathbf{H}(l)$ can be modeled as

$$\mathbf{H}(l) = \mathbf{A}\mathbf{D}(l)\,\mathbf{G}^T, \tag{2}$$

where

$$\mathbf{D}(l)\,(N_p \times N_p) = \text{diag}\left(\mathbf{d}(l)\right), \tag{3}$$

and $\mathbf{d}(l)$ is a vector that contains the N_p coefficients of the multi-path channel distributed according to

$$\mathbf{d}(l)\,(N_p \times 1) \sim \mathcal{CN}(\mathbf{0}_{N_p\times 1}, \mathbf{\Sigma}_d), \tag{4}$$

$$\mathbf{\Sigma}_d = \mathbb{E}\left\{\mathbf{d}(l)\left(\mathbf{d}(l)\right)^H\right\} = \text{diag}\left(\left[\sigma_1^2, \cdots, \sigma_{N_p}^2\right]\right). \tag{5}$$

Note that σ_i^2 denotes the average power of the i-th tap of the power delay profile (PDP). Additionally,

$$\mathbf{A}\,(N_r \times N_p) = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_{N_p}\end{bmatrix}, \tag{6}$$

$$\mathbf{G}\,(N_t \times N_p) = \begin{bmatrix}\mathbf{g}_1 & \cdots & \mathbf{g}_{N_p}\end{bmatrix}, \tag{7}$$

where $\mathbf{a}_i = \mathbf{a}(\theta_i)$ denotes the array response of the BS at the i-th tap, and $\mathbf{g}_i = \mathbf{g}(\tau_i) = \left[\mathbf{g}_\alpha(\tau - \tau_i)\right]_{\downarrow\alpha}$ represents the convolution of the transmitter pulse and the matched filter at the receiver side of the i-th tap, where α is the decimation factor. Note that, we assume that the shape of $\mathbf{g}_i$ is perfectly known. However, τ_i and $\mathbf{a}_i$ are completely unknown and they must be estimated.

Given (1), where $\mathbf{X}(l)$ denotes a pilot sequence, the estimated channel is given as

$$\widehat{\mathbf{H}}(l) = \mathbf{R}_{yx}(l)\,\mathbf{R}_{xx}^{-1} = \mathbf{H}(l) + \Delta\mathbf{H}(l)\,, \tag{8}$$

$$\mathbf{R}_{yx}(l) = \frac{1}{N_s}\mathbf{Y}(l)\,\mathbf{X}^H(l)\,, \tag{9}$$

$$\mathbf{R}_{xx} = \frac{1}{N_s}\mathbf{X}(l)\,\mathbf{X}^H(l)\,. \tag{10}$$

The MSE of the channel estimation $\sigma_{\Delta h}^2$ can be derived as

$$\begin{aligned}\sigma_{\Delta h}^2 &= \mathbb{E}\left\{\left\|\widehat{\mathbf{H}}(l) - \mathbf{H}(l)\right\|_F^2\right\} = \\ &= \text{tr}\left(\mathbb{E}\left\{\Delta\mathbf{h}(l)\,\Delta\mathbf{h}^H(l)\right\}\right) = \frac{N_r\sigma_n^2}{N_s}\text{tr}\left(\mathbf{R}_{xx}^{-1}\right),\end{aligned} \tag{11}$$

where $\Delta\mathbf{h}(l) = \text{vec}\left(\widehat{\mathbf{H}}(l) - \mathbf{H}(l)\right)$.

III. Estimation of the joint ST covariance matrix

Given (2), the vectored version of it is given by

$$\mathbf{h}(l)\,(N_h \times 1) = \text{vec}\left(\mathbf{H}(l)\right) = \left(\mathbf{G}\odot\mathbf{A}\right)\mathbf{d}(l) = \mathbf{S}\mathbf{d}(l)\,, \tag{12}$$

where $N_h = N_r N_t$ and

$$\mathbf{S}\,(N_h \times N_p) = \begin{bmatrix}\mathbf{g}_1\otimes\mathbf{a}_1 & \ldots & \mathbf{g}_{N_p}\otimes\mathbf{a}_{N_p}\end{bmatrix}. \tag{13}$$

The joint ST covariance matrix of $\widehat{\mathbf{h}}(l) = \mathbf{h}(l) + \Delta\mathbf{h}(l)$ can be defined as

$$\mathbf{\Sigma}_{\widehat{h}} = \mathbb{E}\left\{\widehat{\mathbf{h}}(l)\,\widehat{\mathbf{h}}^H(l)\right\} = \mathbf{\Sigma}_h + \mathbf{\Sigma}_{\Delta h}, \tag{14}$$

where

$$\begin{aligned}\mathbf{\Sigma}_h &= \mathbb{E}\left\{\mathbf{h}(l)\,\mathbf{h}^H(l)\right\} = \mathbf{S}\mathbf{\Sigma}_d\mathbf{S}^H = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H = \\ &= \sum_{i=1}^{N_h}\lambda_i v_i v_i^H = \sum_{i=1}^{N_p}\lambda_i v_i v_i^H = \mathbf{V}_S\mathbf{\Lambda}_S\mathbf{V}_S^H,\end{aligned} \tag{15}$$

where $\mathbf{V}$ and $\mathbf{\Lambda}$ denote the matrices of eigen-vectors and eigen-values, respectively. Note that, only N_p out of N_h eigen-values are different from zero. This fact enables the reduction of the noise by using a subspace technique.

Applying $\mathbf{V}^H\mathbf{V} = \mathbf{I}_{N_h}$, (14) can be developed as

$$\mathbf{\Sigma}_{\widehat{h}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H + \frac{\sigma^2_{\Delta h}}{N_h}\mathbf{I}_{N_h} = \mathbf{V}\left(\mathbf{\Lambda} + \frac{\sigma^2_{\Delta h}}{N_h}\mathbf{I}_{N_h}\right)\mathbf{V}^H. \quad (16)$$

Regarding (15), $\mathbf{V} = \begin{bmatrix}\mathbf{V}_S & \mathbf{V}_\Delta\end{bmatrix}$ and

$$\mathbf{\Lambda} + \frac{\sigma^2_{\Delta h}}{N_h}\mathbf{I}_{N_h} = \begin{bmatrix}\mathbf{\Lambda}_S & \mathbf{0}_{N_p\times(N_h-N_p)} \\ \mathbf{0}_{(N_h-N_p)\times N_p} & \mathbf{0}_{(N_h-N_p)\times(N_h-N_p)}\end{bmatrix} + \\ + \frac{\sigma^2_{\Delta h}}{N_h}\begin{bmatrix}\mathbf{I}_{N_p} & \mathbf{0}_{N_p\times(N_h-N_p)} \\ \mathbf{0}_{(N_h-N_p)\times N_p} & \mathbf{I}_{(N_h-N_p)}\end{bmatrix}. \quad (17)$$

Therefore, (16) can be derived as

$$\mathbf{\Sigma}_{\widehat{h}} = \mathbf{V}_S\left(\mathbf{\Lambda}_S + \frac{\sigma^2_{\Delta h}}{N_h}\mathbf{I}_{N_p}\right)\mathbf{V}_S^H + \frac{\sigma^2_{\Delta h}}{N_h}\mathbf{V}_\Delta\mathbf{V}_\Delta^H = \\ = \mathbf{V}_S\mathbf{\Lambda}'_S\mathbf{V}_S^H + \frac{\sigma^2_{\Delta h}}{N_h}\mathbf{V}_\Delta\mathbf{V}_\Delta^H. \quad (18)$$

Inspecting (12) and (18), we can see that $\mathbf{S}$ and $\mathbf{V}_S$ belong to the same subspace, due to the fact that both of them contain the full information of all DoA and ToA of each tap.

However, in a realistic scenario, we do not have as many slots as needed in order to compute (16). Hence, given N_l slots, the estimated version of (16) is given by

$$\widehat{\mathbf{\Sigma}}_{\widehat{h}} = \frac{1}{N_l}\sum_{l=1}^{N_l}\widehat{\mathbf{h}}(l)\,\widehat{\mathbf{h}}^H(l) = \widehat{\mathbf{V}}\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{V}}^H = \\ = \widehat{\mathbf{V}}_S\widehat{\mathbf{\Lambda}}'_S\widehat{\mathbf{V}}_S^H + \widehat{\mathbf{V}}_\Delta\widehat{\mathbf{\Lambda}}_\Delta\widehat{\mathbf{V}}_\Delta^H. \quad (19)$$

IV. Estimation of DoA and ToA

In order to obtain the DoA and the ToA of each tap, we must obtain $\widehat{\mathbf{S}}$, which requires to solve the following Least-Square (LS) minimization problem

$$\min_{\mathbf{a}_i,\tau_i,\mathbf{d}(l)}\sum_{l=1}^{N_l}\left|\left|\widehat{\mathbf{h}}(l) - \mathbf{S}\mathbf{d}(l)\right|\right|_F^2. \quad (20)$$

Placing $\widehat{\mathbf{d}}(l) = \left(\mathbf{S}^H\mathbf{S}\right)^{-1}\mathbf{S}^H\widehat{\mathbf{h}}(l)$ in (20), the argument of the minimization problem can be transformed as

$$\sum_{l=1}^{N_l}\left|\left|\widehat{\mathbf{h}}(l) - \mathbf{P}_S\widehat{\mathbf{h}}(l)\right|\right|_F^2 = \sum_{l=1}^{N_l}\left|\left|\mathbf{P}_S^\perp\widehat{\mathbf{h}}(l)\right|\right|_F^2 = \\ = \sum_{l=1}^{N_l}\text{tr}\left(\widehat{\mathbf{h}}^H(l)\,\mathbf{P}_S^\perp\widehat{\mathbf{h}}(l)\right). \quad (21)$$

Using (21) and the property $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$, (20) can be simplified as

$$\min_{\mathbf{a}_i,\tau_i}\ \text{tr}\left(\mathbf{P}_S^\perp\sum_{l=1}^{N_l}\widehat{\mathbf{h}}(l)\,\widehat{\mathbf{h}}^H(l)\right) = \min_{\mathbf{a}_j,\tau_j}\ \text{tr}\left(\mathbf{P}_S^\perp\widehat{\Sigma}_{\widehat{h}}\right). \quad (22)$$

Inspecting (22), we can see that the true $\mathbf{S}$ will guarantee that projecting the estimated joint ST covariance matrix $\widehat{\mathbf{\Sigma}}_{\widehat{h}}$ in the orthogonal projector of $\mathbf{S}$ ($\mathbf{P}_S^\perp$) will produce a global minimum.

A. MUSIC-1D method

Given (22), we realize that it is very complicated to optimize it in terms of $\mathbf{S}$. Hence, as we mentioned before, using the property that $\mathbf{S}$ and $\widehat{\mathbf{V}}_S$ belong to the same subspace, an equivalent minimization problem to (22) can be formulated as

$$\min_{\mathbf{a}_i,\tau_i}\left|\left|\mathbf{S} - \mathbf{P}_{\widehat{\mathbf{V}}_S}\mathbf{S}\right|\right|_F^2, \ \text{s.t.}\ ||\mathbf{a}_i|| = 1. \quad (23)$$

Note that we added a constraint to our minimization problem in order to avoid the trivial solution.

The argument of (23) can be simplified as

$$\left|\left|\mathbf{S} - \mathbf{P}_{\widehat{\mathbf{V}}_S}\mathbf{S}\right|\right|_F^2 = \left|\left|\mathbf{P}_{\widehat{\mathbf{V}}_\Delta}\mathbf{S}\right|\right|_F^2 = \text{tr}\left(\mathbf{S}^H\mathbf{P}_{\widehat{\mathbf{V}}_\Delta}\mathbf{S}\right) = \\ = \text{tr}\left(\mathbf{P}_{\widehat{\mathbf{V}}_\Delta}\mathbf{S}\mathbf{S}^H\right), \quad (24)$$

where we use that

$$\mathbf{P}^\perp_{\widehat{\mathbf{V}}_S} = \mathbf{I}_{N_h} - \widehat{\mathbf{V}}_S\widehat{\mathbf{V}}_S^H = \widehat{\mathbf{V}}_\Delta\widehat{\mathbf{V}}_\Delta^H = \mathbf{P}_{\widehat{\mathbf{V}}_\Delta}. \quad (25)$$

Inspecting in (24), we can clearly see that it is quadratic in $\mathbf{S}$, reducing the complexity of the minimization problem. Moreover, the trace operator decouples (24) into the sum of N_p cost functions as

$$\sum_{i=1}^{N_p} f(\mathbf{a}_i,\tau_i) = \text{tr}\left(\mathbf{S}^H\mathbf{P}_{\widehat{\mathbf{V}}_\Delta}\mathbf{S}\right) = \sum_{i=1}^{N_p}\mathbf{s}_i^H\mathbf{P}_{\widehat{\mathbf{V}}_\Delta}\mathbf{s}_i. \quad (26)$$

Using the definition given in (13), $f(\mathbf{a}_i,\tau_i)$ can be manipulated as

$$f(\mathbf{a}_i,\tau_i) = \left(\mathbf{g}_i^H\otimes\mathbf{a}_i^H\right)\mathbf{P}_{\widehat{\mathbf{V}}_\Delta}(\mathbf{g}_i\otimes\mathbf{a}_i) = \\ = \mathbf{a}_i^H\left(\mathbf{g}_i^H\otimes\mathbf{I}_{N_r}\right)\mathbf{P}_{\widehat{\mathbf{V}}_\Delta}(\mathbf{g}_i\otimes\mathbf{I}_{N_r})\,\mathbf{a}_i = \mathbf{a}_i^H\mathbf{F}(\tau_i)\,\mathbf{a}_i. \quad (27)$$

Hence, (23) can be simplified as

$$\min_{\mathbf{a}_i,\tau_i}\sum_{i=1}^{N_p}\mathbf{a}_i^H\mathbf{F}(\tau_i)\,\mathbf{a}_i, \ \text{s.t.}\ ||\mathbf{a}_i|| = 1. \quad (28)$$

As we mentioned before, we assume that the $\mathbf{a}_i$ is completely generic. Hence, (24) can be solved by the eigen-value and eigen-vector decomposition as

$$\mathbf{F}(\tau_i)\,\mathbf{V}_F = \mathbf{V}_F\Lambda_F, \quad (29)$$

$$\mathbf{\Lambda}_F = \text{diag}\left(\begin{bmatrix}\lambda_1^F & \cdots & \lambda_{N_r}^F\end{bmatrix}\right), \ \mathbf{V}_F = \begin{bmatrix}\mathbf{v}_1^F & \cdots & \mathbf{v}_{N_r}^F\end{bmatrix}, \quad (30)$$

where the optimum vector for $\mathbf{a}_i$ is given by the eigen-vector of $\mathbf{F}(\tau_i)$ which corresponds to the lowest eigen-value as

$$\mathbf{a}_i^{opt} = \mathbf{v}_u^F, \quad u = \underset{1<u<N_r}{\text{argmin}}\begin{bmatrix}\lambda_1^F & \cdots & \lambda_{N_r}^F\end{bmatrix}. \quad (31)$$

Substituting $\mathbf{a}_i^{opt}$ in (28), we can estimate all the values of $\widehat{\tau}_i$ performing one dimensional search as

$$\widehat{\tau}_i = \min_{\tau_i}\ f\left(\mathbf{a}_i^{opt},\tau_i\right) = \min_{\tau_i}\ f(\tau_i), \ \forall i\in\{1,\ldots,N_p\}. \quad (32)$$

Once we have obtained all values of $\widehat{\tau}_i$, we can compute its corresponding vector $\widehat{\mathbf{a}}_i^{opt}$ and the matrix $\widehat{\mathbf{S}}$.

B. Iterative MLE method

In order to obtain the best performance of the estimated channel, we propose an iterative method based on MLE. The argument of (22) can be manipulated as

$$\mathrm{tr}\left(\mathbf{P}_S^{\perp}\widehat{\Sigma}_{\widehat{h}}\right)=\mathrm{tr}\left(\widehat{\Sigma}_{\widehat{h}}\right)-\mathrm{tr}\left(\mathbf{P}_S\widehat{\Sigma}_{\widehat{h}}\right). \quad (33)$$

Hence, (22) can be transformed into

$$\max_{\mathbf{a}_i,\tau_i}\ \mathrm{tr}\left(\mathbf{P}_S\widehat{\Sigma}_{\widehat{h}}\right),\ \ \text{s.t.}\ ||\mathbf{a}_i||=1. \quad (34)$$

We propose an alternative way of defining $\mathbf{P}_S$ as

$$\mathbf{P}_S=\mathbf{P}_{S_{\bar{i}}}+\mathbf{P}_{p_i}, \quad (35)$$

where $\mathbf{S}_{\bar{i}}\left(N_h\times N_p-1\right)$ is $\mathbf{S}$ deleting its i-th column and $\mathbf{P}_{p_i}$ denotes the matrix projector of $\mathbf{p}_i\left(N_h\times 1\right)=\mathbf{P}_{S_{\bar{i}}}^{\perp}s_i$.

This alternative way of defining $\mathbf{P}_S$ is forcing us to sequentially compute each column of $\widehat{\mathbf{S}}$ and at the same time canceling the effects of the other taps for each iteration. Hence, the argument of (34) for each iteration and the i-th column of $\widehat{\mathbf{S}}$ is transformed into

$$\mathrm{tr}\left(\frac{\mathbf{P}_{S_{\bar{i}}}^{\perp}s_i s_i^H\mathbf{P}_{S_{\bar{i}}}^{\perp}}{s_i^H\mathbf{P}_{S_{\bar{i}}}^{\perp}s_i}\widehat{\mathbf{\Sigma}}_{\widehat{h}}\right)=\frac{s_i^H\mathbf{P}_{S_{\bar{i}}}^{\perp}\widehat{\mathbf{\Sigma}}_{\widehat{h}}\mathbf{P}_{S_{\bar{i}}}^{\perp}s_i}{s_i^H\mathbf{P}_{S_{\bar{i}}}^{\perp}s_i}. \quad (36)$$

Note that, this method implies that we have to execute the minimization problem (36) N_p times for each iteration, and in each iteration we improve the estimation of $\mathbf{S}$. Hence, if we compare the complexity of this method with the previous one, the complexity is increased by $N_p\times N_i$ times where N_i denotes the number of iterations.

Given (36), using again the definition of (13), we obtain that

$$\max_{\mathbf{a}_i,\tau_i} f_i\left(\mathbf{a}_i,\tau_i\right)\ \ \text{s.t.}\ ||\mathbf{a}_i||=1, \quad (37)$$

where

$$f_i\left(\mathbf{a}_i,\tau_i\right)=\frac{\mathbf{a}_i^H\mathbf{F}_i^n\left(\tau_i\right)\mathbf{a}_i}{\mathbf{a}_i^H\mathbf{F}_i^d\left(\tau_i\right)\mathbf{a}_i}, \quad (38)$$

$$\mathbf{F}_i^n\left(\tau_i\right)=\left(\mathbf{g}_i^H\otimes\mathbf{I}_{N_r}\right)\mathbf{P}_{S_{\bar{i}}}^{\perp}\widehat{\mathbf{\Sigma}}_{\widehat{h}}\mathbf{P}_{S_{\bar{i}}}^{\perp}\left(\mathbf{g}_i\otimes\mathbf{I}_{N_r}\right), \quad (39)$$

$$\mathbf{F}_i^d\left(\tau_i\right)=\left(\mathbf{g}_i^H\otimes\mathbf{I}_{N_r}\right)\mathbf{P}_{S_{\bar{i}}}^{\perp}\left(\mathbf{g}_i\otimes\mathbf{I}_{N_r}\right), \quad (40)$$

Finally, we assume again that $\mathbf{a}_i$ is generic, therefore, (36) can be solved by the generalized eigen-value and eigen-vector decomposition as

$$\mathbf{F}_i^n\left(\tau_i\right)\mathbf{V}_F=\mathbf{F}_i^d\left(\tau_i\right)\mathbf{V}_F\Lambda_F, \quad (41)$$

where the optimum vector for $\mathbf{a}_i$ is given by the eigen-vector which corresponds to the highest eigen-value as

$$\mathbf{a}_i^{opt}=\mathbf{v}_u^F,\quad u=\underset{1<u<N_r}{\mathrm{argmax}}\left[\lambda_1^F\quad\cdots\quad\lambda_{N_r}^F\right]. \quad (42)$$

Substituting $\mathbf{a}_i^{opt}$ in (38), we can estimate only one $\widehat{\tau}_i$ searching in $f_i\left(\tau_i\right)$. Hence, the expression is given as

$$\widehat{\tau}_i=\max_{\tau_i}\ f_i\left(\mathbf{a}_i^{opt},\tau_i\right)=\max_{\tau_i}\ f_i\left(\tau_i\right). \quad (43)$$

The estimation of $\left\{\widehat{\mathbf{a}}_i,\widehat{\tau}_i\right\}\ \ \forall 1\leq i\leq N_p$ must be improved in the following iterations.

C. Hybrid method

The two previous methods have their own strengths and weaknesses. MUSIC-1D is a low-complexity method which only has a great performance in high SNRs. Iterative MLE provides a great performance in any scenario with a considerable complexity. Therefore, we propose the hybrid one which consists in the combination of both of them in order to trade-off performance and complexity.

Firstly, we use the MUSIC-1D method in order to compute a preliminary $\widehat{\mathbf{S}}$. Then, we use the iteration MLE method in order to improve the estimation. Note that, $\widehat{\mathbf{S}}$ provided by MUSIC-1D method may have $N_p'\neq N_p$ columns. If $N_p'>N_p$ we keep only the best N_p ones which corresponds to the lowest values of $f\left(\widehat{\tau}_i\right)$. However, if $N_p'<N_p$, the iterative MLE method will compute first the missing columns of it, and then, it will improve the estimation in the following iterations. As we can see later in the numerical results, only one additional iteration of MLE is needed in order to achieve the convergence of $\widehat{\mathbf{S}}$.

V. Automatic rank selection (ARS)

Once we have $\widehat{\mathbf{S}}$, we propose an ARS algorithm based on the selection of the best columns of $\widehat{\mathbf{S}}$ in order to minimize the MSE. Note that the number of selected columns (rank) is denoted by N_k, where $1\leq N_k<N_p$. Therefore, (12) can be decomposed as

$$\mathbf{h}\left(l\right)=\mathbf{S}\mathbf{d}\left(l\right)=\begin{bmatrix}\mathbf{S}_k & \mathbf{S}_{\bar{k}}\end{bmatrix}\begin{bmatrix}\mathbf{d}_k\left(l\right)\\ \mathbf{d}_{\bar{k}}\left(l\right)\end{bmatrix}, \quad (44)$$

where $\mathbf{S}_k\left(N_h\times N_k\right)$ and $\mathbf{d}_k\left(l\right)\left(N_k\times 1\right)$ denote the LR version of $\mathbf{S}$ and $\mathbf{d}\left(l\right)$, respectively; and $\mathbf{S}_{\bar{k}}\left(N_h\times N_p-N_k\right)$ and $\mathbf{d}_{\bar{k}}\left(l\right)\left(N_p-N_k\times 1\right)$ are the complementary ones. Note that, the N_k columns of $\mathbf{S}_k$ correspond with the strongest coefficients of $\mathbf{\Sigma}_d$ (see (5)).

In order to reduce the complexity of the algorithm, we analyze the MSE to obtain some insight on the best rank to use. For the ease of the analysis, we assume that $\widehat{\mathbf{S}}=\mathbf{S}$. Hence, the MSE of the FR projected version of the estimated channel is given by

$$\sigma_{S\Delta h}^2=\mathbb{E}\left\{\left|\left|\mathbf{P}_S\widehat{\mathbf{h}}-\mathbf{h}\right|\right|_F^2\right\}=\mathbb{E}\left\{||\mathbf{P}_S\Delta\mathbf{h}||_F^2\right\}=\frac{N_p}{N_h}\sigma_{\Delta h}^2, \quad (45)$$

and the MSE of the LR projected version can be expressed as

$$\sigma_{S_k\Delta h}^2=\mathbb{E}\left\{\left|\left|\mathbf{P}_{S_k}\widehat{\mathbf{h}}-\mathbf{h}\right|\right|_F^2\right\}, \quad (46)$$

where its argument can be derived as

$$\begin{aligned}\mathbf{P}_{S_k}\widehat{\mathbf{h}}-\mathbf{h}&=-\left(\mathbf{S}_k\mathbf{d}_k\left(l\right)+\mathbf{S}_{\bar{k}}\mathbf{d}_{\bar{k}}\left(l\right)\right)+\\&+\left(\mathbf{S}_k\mathbf{d}_k\left(l\right)+\mathbf{P}_{S_k}\mathbf{S}_{\bar{k}}\mathbf{d}_{\bar{k}}\left(l\right)+\mathbf{P}_{S_k}\Delta\mathbf{h}\right)=\\&=-\mathbf{P}_{S_k}^{\perp}\mathbf{S}_{\bar{k}}\mathbf{d}_{\bar{k}}\left(l\right)+\mathbf{P}_{S_k}\Delta\mathbf{h}.\end{aligned} \quad (47)$$

Assuming that the channel estimation error (noise) is independent of the channel itself, (46) can be derived as

$$\begin{aligned}\sigma_{S_k\Delta h}^2 &= \mathbb{E}\left\{\left|\left|\mathbf{P}_{S_k}^{\perp}\mathbf{S}_{\bar{k}}\mathbf{d}_{\bar{k}}\left(l\right)\right|\right|_F^2\right\} + \frac{N_k}{N_h}\sigma_{\Delta h}^2 = \\ &= \mathbb{E}\left\{f_{\text{bias}}\left(N_k, l\right)\right\} + \frac{N_k}{N_h}\sigma_{\Delta h}^2.\end{aligned} \tag{48}$$

Inspecting (48), its second term can be easily computed given the SNR. However, the first term involves an expectation operator and it will cause a negative impact in the system in terms of latency. Therefore, we propose to use directly $f_{\text{bias}}\left(N_k, l\right)$ for each iteration and each rank, instead of computing its expected value.

VI. Numerical Results

In this section, we show several simulation results in order to provide a better understanding of our proposals.

A. *Simulation parameters*

In Table I we provide the numeric values to our defined parameters. The selected multi-path channel model is given in Table II. The SNR is defined as $\text{SNR} = \mathbb{E}\left\{||\mathbf{H}\left(\mathbf{l}\right)||_F^2\right\}/(N_r\sigma_n^2)$.

TABLE I
Simulation Parameters

$\mathbf{N}_r$	10 antennas	$\boldsymbol{\alpha}$ $\boldsymbol{\Delta}\mathbf{f}$	10×60KHz
$\mathbf{N}_l$	20 slots	**Constellation**	QPSK
$\mathbf{N}_s$	100 symbols	$\mathbf{N}_p$	5 taps
Tx-Rx pulse shape		Raised cosine filter roll-off 0.25	

TABLE II
PDP

$\boldsymbol{i}$	1	2	3	4	5
$\boldsymbol{\sigma_i^2}$[dB]	-1	0	-3	-5	-7
$\boldsymbol{\tau_i}$[ms]	0	0.5	1.7	2.4	5

According to [5], the antenna array response is given by

$$\left[\mathbf{a}\left(\theta_i\right)\right]_r = a_d\left(\theta_i, r\right)a_m\left(\theta_i, r\right), \quad 1 \leq i \leq N_p, \quad 1 \leq r \leq N_r, \tag{49}$$

where $a_d\left(\theta_i, r\right)$ and $a_m\left(\theta_i, r\right)$ are the directivity and the array mainfold, respectively. We assume that the structure of the antenna array at the BS is circular, hence, the directivity and mainfold are detailed as

$$a_d\left(\theta, r\right) = 1.8\cos^{1.6}\left(\theta - \frac{2\pi}{N_r}\left(r-1\right)\right), \tag{50}$$

$$a_m\left(\theta, r\right) = \exp\left(j\frac{d_r N_r}{\lambda}\left(1 - \cos\left(\theta - \frac{2\pi}{N_r}\left(r-1\right)\right)\right)\right), \tag{51}$$

where λ denotes the wavelength and $d_r = \lambda/2$ is the distance between two contiguous elements. The carrier frequency is $f_c = 28$GHz and DoA is modeled as $\theta \sim U\left[-\pi/3, \pi/3\right]$.

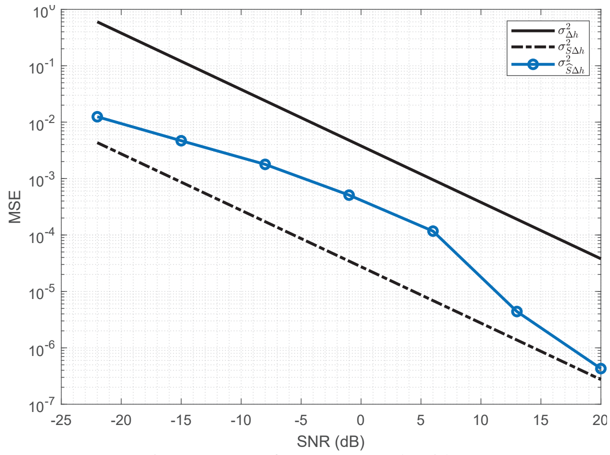

Fig. 1. MSE of MUSIC-1D algorithm.

B. *Results and discussion*

In Fig. 1, we plot the channel estimation error $\sigma_{\Delta h}^2$ and the idealistic FR channel estimation error $\sigma_{S\Delta h}^2$ ($\widehat{\mathbf{S}} = \mathbf{S}$) as reference and the MSE performance of the algorithm MUSIC-1D. We can see that the projected version of the estimated channel using the theoretical $\mathbf{S}$ significantly outperforms the non-projected one thanks to the subspace method. Moreover, regarding the MUSIC-1D performance, it clearly shows that the performance at low SNR regimes is compromised, obliging us to find other alternatives.

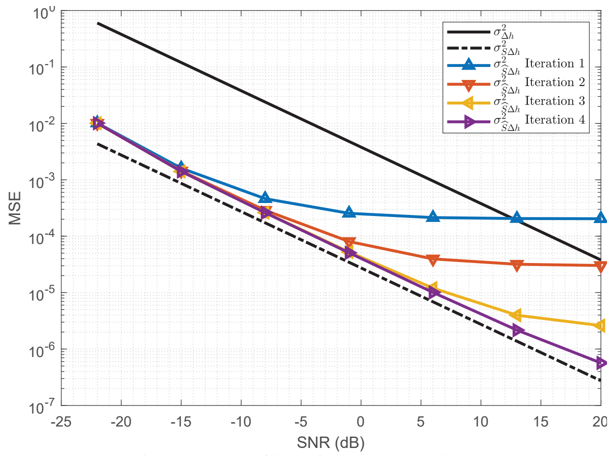

Fig. 2. MSE of iterative MLE algorithm.

In Fig. 2, we show the performance of the iterative MLE algorithm. We can see that this method can achieve a very good performance at the expense of iterating up to four times. Additionally, we can see that there is a small gap error between the fourth iteration of MLE and the $\sigma_{S\Delta h}^2$, which is due to the estimation error in $\widehat{\mathbf{S}}$.

In Fig. 3, we provide the performance of the hybrid solution. It shows that the first iteration is better than the regular MUSIC-1D we mentioned before, MUSIC-1D may provide $N_p' \neq N_p$ and, when $N_p' < N_p$ the iterative MLE may compute the missing columns of $\widehat{\mathbf{S}}$. Furthermore, the hybrid method

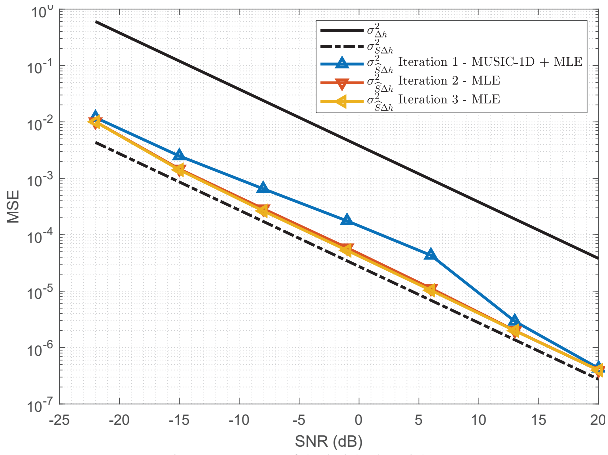

Fig. 3. MSE of hybrid algorithm.

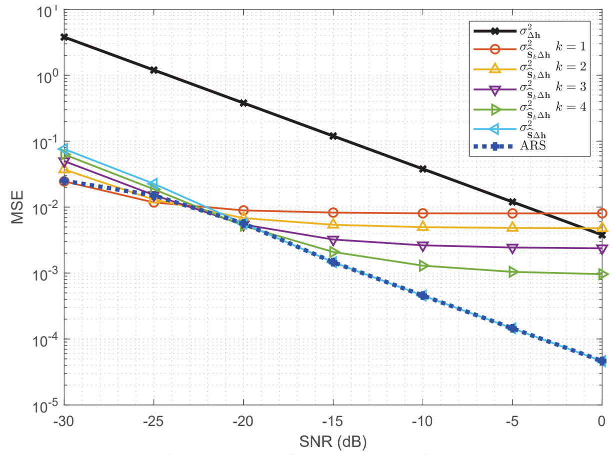

Fig. 5. MSE of the ARS algorithm.

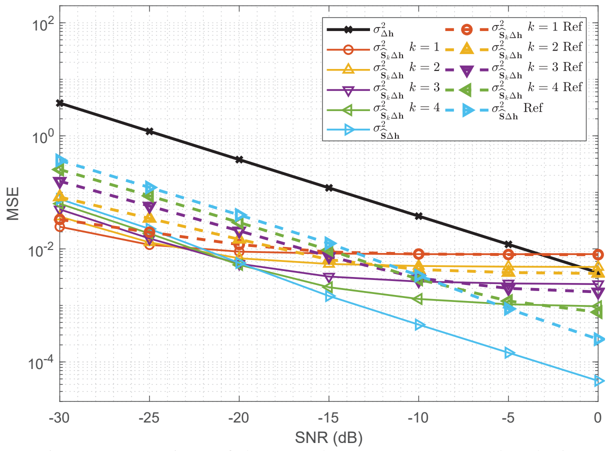

Fig. 4. Comparison of the MSE between our proposal and [5].

only requires an additional iteration in order to saturate the performance, showing the superiority of this method which is capable of trading-off the performance and complexity.

For the hybrid method with two iterations, in Fig. 4, we plot a performance comparison of our proposed scheme against the proposed in [5]. Note that in order to implement [5] and make a fair comparison, we have considered the absence of any interference source, so we have removed its whitening part. We can see that our proposed scheme significantly outperforms [5] in extremely low SNR scenarios and for moderate SNR, our FR also outperforms the referenced one. Additionally, for our proposed hybrid method, the gap of all LR estimated channels between using $\mathbf{S}$ (idealistic case), as a benchmark, and $\widehat{\mathbf{S}}$ (realistic case) is negligible. We omit the simulation results for the sake of space.

Finally, we provide the performance of the ARS algorithm. We can see that using the instantaneous error $f_{\text{bias}}(N_k, l)$, that involves the error estimation in $\widehat{\mathbf{d}}_k(l)$, is good enough to select always the best LR estimated channel.

VII. Conclusions

In the present work, we have proposed a novel method to perform the channel estimation in VLMAS. Our method is based on the estimation of a joint ST covariance matrix; and then, applying a subspace method to obtain the DoA and ToA of each tap using an effective searching method. We have shown that our proposal outperforms the existing techniques, either FR and LR ones and we also provided a novel ARS which is capable of selecting the best LR channel for each scenario providing the lowest MSE. Thus, the proposed schemes are suitable for the dynamic SNR environment in mmWaves.

Acknowledgment

This work has been partly funded by projects TERESA-ADA (TEC2017-90093-C3-2-R) (MINECO/AEI/FEDER, UE) and the fellowship of University Carlos III of Madrid under the program "Ayudas para la Movilidad del Programa Propio de Investigacion" granted to K. Chen-Hu.

References

[1] B. Yang, Z. Yu, J. Lan, R. Zhang, J. Zhou, and W. Hong, “Digital beamforming-based massive MIMO transceiver for 5G millimeter-wave communications,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 7, pp. 3403–3418, July 2018.

[2] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of Massive MIMO: Benefits and challenges,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, Oct 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] S. Haghighatshoar and G. Caire, “Massive MIMO channel subspace estimation from low-dimensional projections,” *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 303–318, Jan 2017.

[5] G. Soatti, A. Murtada, M. Nicoli, J. Gambini, and U. Spagnolini, “Low-rank channel and interference estimation in mm- wave massive antenna arrays,” in *2018 26th European Signal Processing Conference (EUSIPCO)*, 09 2018, pp. 922–926.

[6] Z. Guo, X. Wang, and W. Heng, “Millimeter-wave channel estimation based on 2-D beamspace MUSIC method,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5384–5394, Aug 2017.

[7] M. Abuthinien, S. Chen, and L. Hanzo, “Semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems,” *IEEE Signal Processing Letters*, vol. 15, pp. 202–205, 2008.