



Article

FDD Channel Estimation Via Covariance Estimation in Wideband Massive MIMO Systems

José P. González-Coma ^{*,†}, Pedro Suárez-Casal [†] and Paula M. Castro and Luis Castedo

Department of Computer Engineering & CITIC Research Center, University of A Coruña, 15001 Galicia, Spain; pedro.scasal@udc.es (P.S.-C.); paula.castro@udc.es (P.M.C.); luis.castedo@udc.es (L.C.)

* Correspondence: jose.gcoma@udc.es

† These authors contributed equally to this work.

Received: 10 January 2020 ; Accepted: 6 February 2020; Published: 10 February 2020



Abstract: A method for channel estimation in wideband massive Multiple-Input Multiple-Output systems using hybrid digital analog architectures is developed. The proposed method is useful for Frequency-Division Duplex at either sub-6 GHz or millimeter wave frequency bands and takes into account the beam squint effect caused by the large bandwidth of the signals. To circumvent the estimation of large channel vectors, the posed algorithm relies on the slow time variation of the channel spatial covariance matrix, thus allowing for the utilization of very short training sequences. This is possible due to the exploitation of the channel structure. After identifying the channel covariance matrix, the channel is estimated on the basis of the recovered information. To that end, we propose a novel method that relies on estimating the tap delays and the gains as associated with each path. As a consequence, the proposed channel estimator achieves low computational complexity and significantly reduces the training overhead. Moreover, our numerical simulations show better performance results compared to the minimum mean-squared error solution.

Keywords: mmWave; massive MIMO; hybrid architectures; FDD; channel estimation; beam squint

1. Introduction

Massive Multiple-Input Multiple-Output (MIMO) and millimeter wave (mmWave) technologies are promising candidates to satisfy the demands of future wideband wireless communication systems. A common feature of these technologies is the deployment of antenna arrays with a large number of elements while keeping the size of the antenna aperture small. This way, advantages such as large beamforming gains necessary to compensate the propagation losses at mmWave frequencies, or the ability to support many spatially multiplexed streams, are efficiently achieved [1,2]. In addition, it has been shown that in multi-cell massive MIMO systems the capacity increases without bound as the number of antennas increases, even under pilot contamination, if the channel covariance matrices of the contaminating users are asymptotically linearly independent [3]. Measurements campaigns have proved that this is usually the case in practice [4].

Deploying massive antenna arrays raises concerns in terms of hardware cost and power consumption, especially at mmWave frequencies. To alleviate these requirements, hybrid architectures with an analog preprocessing network operating in the Radio Frequency (RF) domain have been proposed to reduce the number of complete RF chains [2]. Hybrid approaches imply stringent limitations, such as the reduced spatial dimensionality and the lack of flexibility of the analog part, usually implemented with a Phase Shifter (PS) network. Thus, results obtained for fully digital scenarios are not applicable to the hybrid case in general, and new precoding or combining solutions are needed.

The benefits of large antenna arrays rely on the accuracy of the Channel State Information (CSI). Considering wideband systems with multicarrier modulations, this information should be acquired

for the whole frequency band. However, acquiring the CSI is challenging in this scenario, especially when the system operates in Frequency-Division Duplex (FDD) mode. In this case, if the channel is sounded in the downlink, the length of the training sequence depends on the number of transmit antennas [5]. Thus, most prior art on channel estimation focused on systems operating on Time-Division Duplex (TDD) mode [6]. Moreover, the aforementioned restrictions related to hybrid architectures are even more challenging on wideband scenarios, since the analog network is frequency-flat [7]. This is incompatible with methods independently designing the spatial signature of the training sequences for each subcarrier, such as those employed in [8,9], or fully digital narrowband solutions. Therefore, hybrid approaches usually compensate these limitations with longer training overheads [10]. This problem is even more relevant if large bandwidths are considered due to the beam squint effect [11], which causes that the channel statistics change in each subcarrier.

A common approach to reduce the training overhead is to assume that the channel is sparse [12–15]. Some authors conducting measurement experiments reported sparsity at mmWave frequencies, with a few channel propagation paths and small angular spreads [16,17]. Nevertheless, measurement campaigns on sub-6 GHz frequencies with local scatterers around the Base Station (BS) show that these channels are not sparse in general [10,18]. As a consequence, dropping the sparsity assumption enables a wider applicability, and allows us to circumvent the power leakage typically arising due to basis mismatching.

Other authors consider that the channel second order statistics are known in advance. Unfortunately, these channel statistics have to be acquired too in practical settings. In case the channel is almost spatially uncorrelated or Rician-distributed, it has been proved that it is a good approximation to estimate only the diagonal elements of the covariance matrix of the channel [19,20]. Under correlated channels with Rayleigh fading, the training sequences are designed according to the main eigenvectors of the downlink channel covariance [21,22]. However, these training sequences cannot be used in hybrid architectures given their hardware constraints. In TDD scenarios, this covariance can be computed from the uplink [23,24] and extrapolated to the downlink by means of channel reciprocity. However, uplink and downlink channels are different in FDD mode, and in that case angular reciprocity is assumed [25–29]. However, the computational complexity of these solutions for fully-digital narrowband scenarios is high, and they require an additional training period in the downlink to obtain the instantaneous channel realizations.

Channel covariance estimation techniques usually rely on the assumption of Toeplitz covariance matrices, a circumstance that occurs in usual antenna arrangements, such as Uniform Linear Array (ULA), and channel assumptions, such as Rayleigh fading and the one-ring channel model [19,21,27]. This structure can be exploited to design short training sequences based on sparse rulers. Previous works designed these rulers using coprime arrays [30,31], but this strategy cannot attain the shortest possible training lengths for any number of antennas. There are works that consider more efficient ruler designs for small covariance matrices [32,33], but they lack an efficient method to design these training sequences for an arbitrary number of antennas.

Regarding frequency selective channels, solutions have been proposed [15,34,35]. In [34], authors address covariance estimation in TDD mode for hybrid architectures. The proposed algorithm is an application of Orthogonal Matching Pursuit (OMP) and solves the Multiple Measurement Vector (MMV) problem by allowing for a dynamic sensing matrix. In addition, this method is extended to wideband scenarios under the assumption of common subcarrier support, as done in [8,14,15]. Conversely, authors consider approximately frequency flat second order channel statistics in [35]. These assumptions are impractical for scenarios where the signal bandwidth is as large as several GHz [10] because of the beam squint effect. Recently, refs. [28,29] considered the wideband scenario under beam squint and relied on angle reciprocity and channel sparsity to estimate the downlink channel in FDD. However, this approach shares the same drawbacks as other works based on similar assumptions, namely, large computational complexity and additional training in the downlink, as explained above.

In this work, we propose a downlink channel estimation method for wideband massive MIMO systems over hybrid architectures based on the explicit acquisition of the channel covariance matrix. With the aim of achieving low training overheads, the proposed method employs training sequences based on sparse rulers. The considered setup typically arises in mmWave frequencies, but the proposed framework is general enough to be applicable in sub-6 GHz bands since channel sparsity is not required. Indeed, our approach only assumes (1) the channel is wide-sense stationary over a period of time, and (2) the channel covariance matrix has a Toeplitz structure. Moreover, we consider the nonlinear spatial transformation caused by the beam squint effect. In the following, we summarize the contributions of our work:

- We propose a wideband downlink channel covariance estimation method for hybrid FDD massive MIMO systems that can be utilized for mmWave frequencies. The covariance matrices for different subcarriers lie in distinct subspaces due to the beam squint effect. The proposed covariance estimation algorithm implements dictionary rotations to overcome this difficulty and exploit the common structure of the wideband covariance matrices.
- Differently from the prior art, and with the aim of reducing the training overhead, Wichmann sparse rulers [36] are used to design the training sequences. Prior work made use of coprime arrays, but this solution cannot attain the lowest training lengths for a given number of transmit antennas. Other works that relied on sparse rulers did not provide methods to design the training sequences for a number of antennas arbitrarily large.
- We analyze the limitations of the proposed method based on Multiple Signal Classification (MUSIC): (1) The low rank characteristic of the sample covariance matrices limits the applicability of MUSIC-like techniques. The use of sparse rulers makes it possible to extend the proposed method using spatial smoothing techniques and allows for even shorter training sequences; (2) Prior knowledge on the number of channel paths. We show numerically that the number of channel paths can be estimated without significant performance losses and negligible computational cost. (3) The dictionary dependency. We provide a discussion on dictionary sizes, posing that moderated sizes provide enough angular resolution. Furthermore, since sparsity is not a requirement for the proposed method, we show empirically how to avoid the power leakage arising when the Angle of Departures (AoDs) do not lie on the dictionary.
- We propose a low complexity channel estimator, leveraging on the structure of the wideband channels to avoid the individual per-subcarrier procedure. This estimator obtains the delays and gains corresponding to each channel propagation path exploiting underlying common information for all the subcarriers. Moreover, the estimation is performed using the training sequence employed for the acquisition of the covariance matrices, and it does not require an additional training stage.

Notation

In this work, scalars are represented by lower case letters, vectors by bold lower case letters, and matrices by bold capital letters. The superscripts T and H are the transpose and the Hermitian operators, respectively. The operator $|\cdot|$ with a scalar argument represents the absolute value, $\|\cdot\|$ with a vector argument represents its norm and $\|\cdot\|_F$ with a matrix argument the Frobenius norm. The operator $\lfloor \cdot \rfloor$ represents the element-wise floor operation. $A \otimes B$ is the Kronecker product of matrices A and B , and $A \circ B$ the Khatri–Rao product (or the column-wise Kronecker product) of matrices A and B . $\text{diag}(\cdot)$ is the diagonal matrix with the arguments in its main diagonal, and $\text{blockdiag}(\cdot)$ constructs a diagonal supermatrix in which the diagonal elements are given by the matrices in the argument, and the off-diagonal elements are zero matrices. $\mathbf{0}$ and $\mathbf{1}$ denote the vectors of zeros and ones of the right dimension, respectively. The constant c is the speed of light.

2. System Model

We consider a hybrid FDD massive MIMO system with M transmit antennas and N_{RF} RF chains communicating with several single-antenna users over a multipath channel. To overcome the channel frequency selectivity, we assume an Orthogonal Frequency-Division Multiplexing (OFDM) modulation with N_c subcarriers and a cyclic prefix large enough to avoid intercarrier and intersymbol interference. One of the aims of this work is to estimate the downlink channel covariance; hence, assuming that the channel is locally stationary, we will consider several blocks $k \in \{1, \dots, K\}$ with the duration of the channel coherence time. At the beginning of each block, the BS transmits a training sequence of length T_{tr} OFDM symbols to sound both the covariance and the channel. The remaining time is devoted to data transmission. That is, we collect K wideband observations of size T_{tr} during K training periods. Since the block duration is limited by the channel coherence time, the training sequence length must satisfy $T_{tr} \ll M$ due to the large antenna array deployed at the BS. Indeed, we circumvent the requirement of training sequence lengths larger than the number of paths L , i.e., we propose solutions to the scenario $T_{tr} \leq L$.

Finally, we will assume that the users feed the observations back to the BS using, e.g., scalar quantization [8,37]. With the received feedback, the BS estimates the channel covariance matrices for all users and frequency subcarriers. This scheme allows for tracking the channel covariance matrices using subsequent received feedback due to the slow variation of the channel statistics. As previously mentioned, these slow changes make channel covariance acquisition simpler than estimating the channel itself. Next, with the aid of the obtained channel covariance matrices, we estimate the wideband channels employing the K aforementioned observations.

2.1. Channel Model

We represent the channel response to an arbitrary user using a model similar to that in [34,38]. We assume a multipath channel model with N taps, where the gains for all antennas in the f -th subcarrier during the k -th coherence block are given by

$$\mathbf{h}_k(f) = \sum_{n=0}^{N-1} \mathbf{h}_{n,k}(f) e^{-j2\pi f n / N_c}, \quad f = 0, \dots, N_c - 1, \quad (1)$$

where $\mathbf{h}_{n,k}(f) \in \mathbb{C}^M$ are the channel coefficients of the n -th tap of the channel. Each tap, in turn, is decomposed into L paths as

$$\mathbf{h}_{n,k}(f) = \sum_{l=1}^L g_{l,k} p_{rc}(nT_s - \tau_l) \mathbf{a}(\vartheta_l, f), \quad (2)$$

where the $\mathbf{a}(\vartheta_l, f) \in \mathbb{C}^M$ are the steering vectors for the AoD of the l -th path ϑ_l , $g_{l,k} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_l^2)$ and τ_l are the complex channel gain and the relative delay corresponding to the l -th path, respectively, and $p_{rc}(t)$ is the transmit filter as associated with the Digital to Analog Converter (DAC) sampled with time interval T_s . We assume that the channel gains for different paths are statistically independent, i.e., $E[g_l g_{l'}^*] = \sigma_l^2 \delta(l - l')$.

Regarding the steering vectors, we consider a ULA configuration together with the so-called beam squint effect that arises for the typical antenna apertures in massive MIMO and large relative signal bandwidths B/f_c [11]. We express this effect as a rotation of the steering vectors with respect to the central frequency f_c , i.e.,

$$\mathbf{a}(\vartheta_l, f) = \mathbf{\Gamma}(\vartheta_l, f) \mathbf{a}(\vartheta_l), \quad (3)$$

where $\mathbf{a}(\vartheta_l) \in \mathbb{C}^M$ is the steering vector for ϑ_l in a ULA, whose elements are $[\mathbf{a}(\vartheta_l)]_m = e^{-j\frac{2\pi}{\lambda} d(m-1) \sin \vartheta_l}$, $m = 1, \dots, M$, with d is the distance between two consecutive array elements, λ

is the carrier wavelength, and $\mathbf{\Gamma}(\vartheta_l, f)$ is a diagonal matrix with the beam squint rotation coefficients. Under the common assumption of inter-antenna distance $d = \frac{c}{2f_c}$, the elements of $\mathbf{\Gamma}(\vartheta_l, f)$ read as

$$[\mathbf{\Gamma}(\vartheta_l, f)]_{m,m} = e^{j\pi \sin \vartheta_l (m-1) \left(1 + \frac{\Delta_f}{N_c} \frac{B}{f_c}\right)}, \quad m = 1, \dots, M, \quad (4)$$

where $\Delta_f \frac{B}{N_c}$ is the frequency offset of the f -th subcarrier with respect to f_c , and $\Delta_f = f - \frac{N_c - 1}{2}$ is its relative position to the central one. This frequency-dependent rotations allows for exploiting the common structure of the wideband covariance matrices, since the beam squint effect for the subcarriers is isolated as a rotation of the frequency-independent vectors $\mathbf{a}(\vartheta_l)$.

Combining Equations (1) and (2), the channel frequency response for the f -th subcarrier reads as

$$\mathbf{h}_k(f) = \sum_{l=1}^L g_{l,k} \beta_l(f) \mathbf{a}(\vartheta_l, f) \quad (5)$$

where N is the number of delay taps and $\beta_l(f) = \sum_{n=0}^{N-1} p_{rc}(nT_s - \tau_l) e^{-j2\pi f n / N_c}$.

Since the actual AoD of the channel are unknown, the linear combination of steering vectors in (5) can be approximated by means of the steering vectors as associated with G predefined angles $\theta_1, \dots, \theta_G$ comprised in a frequency-dependent dictionary $\mathbf{A}(f) = [\mathbf{a}(\theta_1, f), \dots, \mathbf{a}(\theta_G, f)]$, that is,

$$\mathbf{h}_k(f) = \sum_{l=1}^L g_{l,k} \beta_l(f) \mathbf{a}(\vartheta_l, f) \approx \mathbf{A}(f) \mathbf{B}(f) \mathbf{g}_k, \quad (6)$$

where $\mathbf{g}_k \in \mathbb{C}^G$ is a vector with L non-zero entries in the positions corresponding to the AoDs, i.e., $\mathbf{g}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{G})$ with $\mathbf{G} = \text{diag}(\sigma_1^2, \dots, \sigma_G^2)$, and $\mathbf{B}(f) = \text{diag}(\beta_1(f), \dots, \beta_G(f))$. Note that equality holds for (6) if the AoDs lie on the dictionary. According to the assumptions regarding the channel statistics, we obtain $\mathbf{h}_k(f) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{C}_{h(f)})$. Moreover, we use the approximation in (6) yielding

$$\mathbf{C}_{h(f)} \approx \mathbf{A}(f) \mathbf{B}(f) \mathbf{G} \mathbf{B}^H(f) \mathbf{A}^H(f) = \mathbf{A}(f) \tilde{\mathbf{G}}(f) \mathbf{A}^H(f), \quad (7)$$

with $\mathbf{C}_{h(f)}$ being Toeplitz, and $\tilde{\mathbf{G}}(f) = \mathbf{B}(f) \mathbf{G} \mathbf{B}^H(f)$ the effective gains for the f -th subcarrier. In the following, we will consider that the AoDs belong to the dictionary, i.e., we assume $\mathbf{C}_{h(f)} = \mathbf{A}(f) \tilde{\mathbf{G}}(f) \mathbf{A}^H(f)$. The consequences of this assumption are discussed in Section 4.4.

2.2. Channel Observations

We aim to estimate the channel response using hybrid architectures to reduce the power consumption and implementation costs. This is achieved by using a number of RF chains significantly lower than the number of antennas, i.e., $N_{\text{RF}} \ll M$. As a counterpart, the feasible hybrid precoding vectors are restricted to the subspace imposed by the hardware limitations. Considering hybrid precoding, the signal received by the users during the k -th training block in the f -th subcarrier is

$$\boldsymbol{\varphi}_k(f) = \mathbf{T}(f) \mathbf{h}_k(f) + \mathbf{v}_k(f) = \mathbf{X}(f) \mathbf{P} \mathbf{h}_k(f) + \mathbf{v}_k(f), \quad f = 0, \dots, N_c - 1, \quad k = 1, \dots, K, \quad (8)$$

where $\mathbf{v}_k(f) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_{v(f)}^2 \mathbf{I})$ is the noise, $\mathbf{T}(f) = \mathbf{X}(f) \mathbf{P} \in \mathbb{C}^{T_{\text{tr}} \times M}$ is a matrix whose rows are the training vectors for each of the T_{tr} training channel uses, such that $\mathbf{T}^T(f) = [\mathbf{t}_1(f), \dots, \mathbf{t}_{T_{\text{tr}}}(f)]$, $\mathbf{t}_t(f) \in \mathbb{C}^M$, $\mathbf{X}(f) = \text{diag}(x_1(f), \dots, x_{T_{\text{tr}}}(f))$ contains the transmitted pilot symbols in the f -subcarrier satisfying $|x_t(f)| = 1$, and $\mathbf{P}^T = [\mathbf{p}_1, \dots, \mathbf{p}_{T_{\text{tr}}}]$ is the hybrid precoder. Therefore, each of these training vectors is the result of the product

$$\mathbf{t}_t(f) = \mathbf{p}_t x_t(f) = \mathbf{P}_{A,t} \mathbf{p}_{D,t} x_t(f), \quad t = 1, \dots, T_{\text{tr}}, \quad (9)$$

where \mathbf{p}_t is the configuration of the precoders during the t -th training period of k -th training block, and $\mathbf{P}_{A,t} \in \mathbb{C}^{M \times N_{\text{RF}}}$ and $\mathbf{p}_{D,t} \in \mathbb{C}^{N_{\text{RF}}}$ denote the analog and digital precoding matrices. Recall that $\mathbf{P}_{A,t}$

is flat in the frequency domain. Hence, it is not possible to design independent training sequences for each subcarrier.

After the transmission of the T_{tr} vectors and the subsequent feedback stage, the channel observation $\varphi_k(f)$ of the k -th training period will be available at the BS. Since the channel and the noise vectors are statistically independent, the covariance of the observations is given by

$$\mathbf{C}_{\varphi(f)} = \mathbf{T}(f)\mathbf{C}_{h(f)}\mathbf{T}(f)^H + \sigma_{v(f)}^2\mathbf{I}. \quad (10)$$

By considering the channel covariance matrix model in (7), we rewrite the previous equation as

$$\mathbf{C}_{\varphi(f)} = \mathbf{\Psi}(f)\tilde{\mathbf{G}}(f)\mathbf{\Psi}^H(f) + \sigma_{v(f)}^2\mathbf{I}, \quad (11)$$

where we introduce the measurement matrices $\mathbf{\Psi}(f) = [\psi_1(f), \dots, \psi_G(f)] = \mathbf{T}(f)\mathbf{A}(f)$. When the K samples of the random variable of (8) are available at the BS, $\mathbf{C}_{\varphi(f)}$ is estimated. Furthermore, given that both $\mathbf{\Psi}(f)$ and $\sigma_{v(f)}$ are known, determining $\mathbf{C}_{\varphi(f)}$ is equivalent to estimating $\tilde{\mathbf{G}}(f)$.

3. Training Design

This section is devoted to the design of training sequences to estimate the channel covariance, together with bounds on their length to guarantee its correct identification under the assumption of a Toeplitz structure.

Toeplitz covariance matrices of size $M \times M$ have M different coefficients and a training sequence of length T_{tr} allows for defining $(T_{\text{tr}}^2 + T_{\text{tr}})/2$ equations. This establishes the following trivial lower bound on the training length: $T_{\text{tr}} \geq \sqrt{2M + 1/4} - 1/2$. A necessary condition to achieve this bound is that the training sequence should allow for measuring the correlation between pairs of antennas for any separation. This has a strong connection with complete sparse rulers, defined mathematically as an ordered integer sequence of with T elements $\mathcal{R} = (r_1, \dots, r_T)$, such that $r_1 = 0$ and any $z \leq r_T$ can be written as $z = r_i - r_j$ for some $r_i, r_j \in \mathcal{R}$. In this case, we speak of a complete sparse ruler of length r_T , and a perfect ruler is a complete ruler that contains the minimum possible number of elements for a given r_T . It has been shown that a training sequence built from a perfect sparse ruler is sufficient for determining Toeplitz covariance matrices [32,33]. In that case, the length of the training sequence is equal to the number of elements of the ruler, i.e., $T_{\text{tr}} = T$, and the last element of the ruler is $r_T = M - 1$.

In the following, we summarize some results found in the literature for sparse rulers. First, there exist bounds on the size of perfect rulers for a given length [39].

Theorem 1. (see [39]) Denoting $l(r)$ as the minimum number of marks of a complete sparse ruler of length r , $\lim_{r \rightarrow \infty} \frac{l^2(r)}{r}$ exists and

$$2.434 \leq \lim_{r \rightarrow \infty} \frac{l^2(r)}{r} \leq \frac{375}{112} \approx 3.348. \quad (12)$$

Hence, it is possible to identify a Toeplitz covariance matrix using sparse rulers with a training length that satisfies $\sqrt{2.434M} \leq T_{\text{tr}} \leq \sqrt{3.348M}$ when M approaches infinity.

Utilizing the bounds provided by (12), we evaluate the efficiency of the training sequences employed in previous works on AoD estimation [30,31]. Let us start by introducing a trivial example of a complete sparse ruler of length m^2 , for any arbitrary m . Using the $2m - 1$ marks $(0, 1, 2, \dots, m - 1, 2m - 1, 3m - 1, \dots, m^2 - 1)$ provides a length of $m^2 - 1$ different differences. This complete ruler achieves a value of $\lim_{m \rightarrow \infty} \frac{(2m-1)^2}{m^2} = 4$, larger than the bound of (12). A similar result is obtained with the use of coprime arrays [30,31]. This method is based on two co-prime numbers p and q , i.e., $\text{mcd}(p, q) = 1$, and on generating the ruler as the set $\{mp : 0 \leq m < q\} \cup \{nq : 0 \leq n < p\}$. This method achieves a training length of $T_{\text{tr}} = p + q$ and allows for identifying pq different lags.

Thus, the limiting function reads as $\frac{(p+q)^2}{pq}$. Setting $p = q = \sqrt{M}$ provides the bound $\lim_{M \rightarrow \infty} \frac{(2\sqrt{M})^2}{M} = 4$. Hence, coprime arrays cannot improve the asymptotic efficiency of the previous trivial complete ruler.

In the literature, we can find some techniques to build sparse rulers for arbitrary lengths satisfying the bound in (12). The following section summarizes a method to obtain rulers that are close to that bounds with very low computational complexity.

3.1. Wichmann Rulers

Wichmann [36] proposed a direct method to obtain sparse rulers of any size that, denoted as the differences between consecutive terms, have the form

$$\mathcal{W} = (\underbrace{1, \dots, 1}_r, \underbrace{r+1, 2r+1, \dots, 2r+1}_r, \underbrace{4r+3, \dots, 4r+3}_s, \underbrace{2r+2, \dots, 2r+2}_{r+1}, \underbrace{1, \dots, 1}_r) \quad (13)$$

for some positive integers r and s such that $2r - 2 \leq s \leq 2r + 4$. The number of elements of this ruler is $l(r, s) = 4r + s + 3$ and its length $n(r, s) = 4r^2 + 8r + 3 + (4r + 3)s$. Hence, if the elements of \mathcal{W} are denoted as $w_i, i = 1, \dots, l(r, s)$, the sparse ruler is $\mathcal{R} = (r_1, \dots, r_{l(r,s)+1})$ with $r_i = r_{i-1} + w_i, i = 2, \dots, l(r, s) + 1$ and $r_1 = 0$.

With a change of variable $s = s' + (2r - 2)$, the length expression changes to

$$\begin{aligned} n(r, s') &= 12r^2 + 6r - 3 + (4r + 3)s' \\ &= n_{10}(r) + (4r + 3)s', \quad s' = 0, \dots, 6. \end{aligned} \quad (14)$$

The parameter r roughly adjusts the length to intervals starting at $n_{10}(r)$. Then, for each r value, the s' parameter finely adjusts length in steps of size $4r + 3$. This allows for finding parameters r and s' that generate a sparse ruler whose length is as close as possible to a given value.

Finally, the ruler can be completed up to the desired length, relying on the following observation: it is always possible to build a ruler that contains all differences in the interval $[a, b]$ with a set of the form $\mathcal{R}' = (0, 1, 2, \dots, c, b - c(c + 1), \dots, b - 2c, b - c, b)$, with $c = \lfloor \sqrt{b - a} \rfloor$. The cardinality of this set is at most $2c + 3$. In this case, given a Wichmann ruler $\mathcal{R} = (r_1, \dots, r_{l(r,s)+1})$ of length $n(r, s)$, and a set \mathcal{R}' that contains all differences in the interval $[n(r, s), M]$, a new ruler extended to length M can be found by $\mathcal{R} \cup \mathcal{R}'$.

3.2. Training Based on Sparse Rulers

A training sequence to estimate a Toeplitz covariance matrix can be built from a sparse ruler. In that case, the length of the ruler must be equal to $M - 1$, the number of correlation coefficients. Given a sparse ruler of length $M - 1$ with elements $(r_1, \dots, M - 1) \in \mathbb{Z}^{\text{Tr}}$, a training sequence for a Toeplitz covariance matrix can be designed as $T^T(f) = [e_{r_1+1}, \dots, e_M]$, where e_i is the M -length indicator column vector of zeros and a one in the i -th element.

In the case of Wichmann rulers, it is not possible to obtain them for arbitrary lengths, but instead a length close to a given value. Hence, to obtain a ruler close to length $M - 1$, the r parameter can be found as the largest such that $n_{10}(r) \leq M - 1$, which is $r^* = \lfloor \frac{\sqrt{12M+33}-3}{12} \rfloor$. The other parameter is the largest s' such that $n(r^*, s') \leq M - 1$, implying that $s'^* = \lfloor \frac{M-1-n_{10}(r^*)}{4r^*+3} \rfloor$. The length of the ruler can be completed up to $M - 1$ by the method detailed in the previous section. Since the Wichmann rulers always contain entries from 0 to r , and the s parameter increases length in steps of $4r + 3$, this completion procedure will add always at most $\lfloor \frac{4r+3}{r} \rfloor = 5$ marks to the ruler, and the minimum training period will be in the interval $T_{\text{tr}} \in [l(r^*, s'^*), l(r^*, s'^*) + 5]$. Figure 1 shows the training length obtained with this strategy compared to the upper and lower bounds in (12). It also shows the length achieved with the coprime arrays strategy.

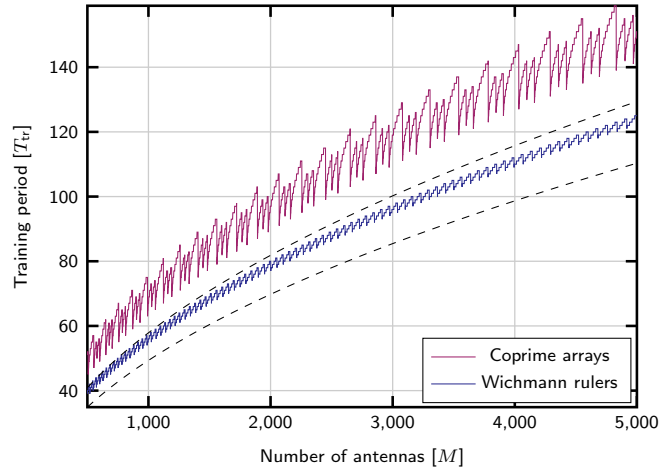


Figure 1. Training length vs. number of antennas with Wichmann rulers. Dashed lines show as asymptotic upper and lower bounds on the minimum training size based on (12).

In the following, we show that the mild assumption $N_{\text{RF}} \geq 2$ is enough to obtain the sparse ruler training sequence $T(f)$ according to (9). Let $\phi_r, r = 1, \dots, N_{\text{RF}}$ be blueeligible arbitrary phases for the PSs in the analog processing network, and $\mathbf{P}_{A,t}^T = [\mathbf{p}_{A,t,1}, \dots, \mathbf{p}_{A,t,M}]$, with the $\mathbf{p}_{A,t,m}^T \in \mathbb{C}^{N_{\text{RF}}}$ is the m -th row of the analog precoding matrix for the t -th channel training use. Since the desired training based on sparse rulers must satisfy $\mathbf{t}_t(f) = \mathbf{e}_{r_t} x_t(f)$, we obtain $\mathbf{P}_{A,t} \mathbf{p}_{D,t}(f) = \mathbf{e}_{r_t}$ as

$$\mathbf{p}_{A,t,m}^T = \begin{cases} [e^{j\phi_1}, e^{j\phi_1}, e^{j\phi_2}, \dots, e^{j\phi_{N_{\text{RF}}-1}}], & \text{if } [e_{r_t}]_m = 1 \\ [e^{j\phi_1}, e^{j(\phi_1+\pi)}, e^{j\phi_2}, \dots, e^{j\phi_{N_{\text{RF}}-1}}], & \text{otherwise,} \end{cases} \quad (15)$$

with the digital part as $\mathbf{p}_{D,t}(f) = \frac{1}{2} e^{-j\phi_1} [1, 1, \underbrace{0, \dots, 0}_{N_{\text{RF}}-2}]^T$. Note that only two elements of $\mathbf{p}_{A,t,m}^T$ are employed to achieve $\mathbf{P}_{A,t} \mathbf{p}_{D,t}(f) = \mathbf{e}_{r_t}$. Indeed, $N_{\text{RF}} = 2$ is enough to apply (15).

4. Covariance Matrix Estimation

In this section, we propose algorithmic solutions to estimate the downlink covariance in wideband massive MIMO systems under the beam squint effect. It is apparent that per-subcarrier (or per-sub-band) covariance estimation is not a practical approach since it does not exploit the structure of the wideband channel. Furthermore, system scalability is compromised when the number of subcarriers is large. The proposed method is based on the estimation of the subspace spanned by the steering vectors corresponding to the L AoD of the channel propagation paths. Well-known estimation methods like MUSIC have been employed to estimate a single signal subspace. However, the beam squint effect results in different subspaces for the channel covariance matrices at distinct frequencies. Resorting to the rotation matrices in (4), we extend MUSIC to the considered setup.

4.1. Subspace Estimation

To estimate the subspace, let us start by computing the eigendecomposition of the sample covariance matrices $\hat{\mathbf{C}}_{\varphi(f)} = \mathbf{U}(f) \mathbf{\Lambda}(f) \mathbf{U}^H(f)$ for all the frequencies f . Given that the covariance converges to the second order moment of the channel for a large number of samples, we have $\hat{\mathbf{C}}_{\varphi(f)} \rightarrow \mathbf{C}_{\varphi(f)}$ when $K \rightarrow \infty$. Next, the eigenvectors corresponding to the L largest eigenvalues are discarded to obtain the basis $\bar{\mathbf{U}}(f)$ that spans the noise subspace of the f subcarrier. Due to the beam squint effect, the relationship among dictionary matrices for different frequencies is nonlinear. Therefore, it is not feasible to perform this step jointly for all the subcarriers unless the relative signal bandwidth B/f_c is very small. A remarkable feature of this subspace discrimination is its lack

of dependency with the SNR, as long as the noise is spatially white. Under such a assumption, for the approximation in (11), we get

$$\mathbf{C}_{\varphi(f)} = \mathbf{\Psi}(f)\tilde{\mathbf{G}}(f)\mathbf{\Psi}^H(f) + \sigma_{v(f)}^2\mathbf{I}_{T_{tr}} = \mathbf{U}(f)(\mathbf{\Lambda}(f) + \sigma_{v(f)}^2\mathbf{I}_{T_{tr}})\mathbf{U}^H(f). \quad (16)$$

Therefore, it is crucial to force this condition by applying a pre-whitening filter, if necessary. According to this observation, the search is performed in the effective subspace, contrarily to OMP-based methods like Covariance OMP (COMP) [34]. Thus, we identify the support $\mathcal{S} = \{\theta_{s_1}, \dots, \theta_{s_L}\}$, with $s_i \in \{1, \dots, G\} \forall i$, with the following wideband estimator function

$$\mathcal{J}_i = \sum_{f=1}^{N_c} \frac{1}{\|\boldsymbol{\psi}_i^H(f)\bar{\mathbf{U}}(f)\|_2^2}, \quad i = 1, \dots, G, \quad (17)$$

where $\boldsymbol{\psi}_i(f) = \mathbf{T}(f)\mathbf{a}(\theta_i, f)$. Note that this function greatly increases the robustness of the decision, compared to a per-carrier estimator because it exploits the common structure for all the frequencies, given that the vectors $\mathbf{a}(\theta_i, f)$ are generated from a common dictionary as shown in (3). Finally, the support \mathcal{S} is identified as the L largest values of \mathcal{J}_i .

4.2. Estimating Gain Variances

Once the angles are determined, we construct the matrices $\mathbf{\Psi}_{\mathcal{S}}(f) = \mathbf{T}(f)[\mathbf{a}(\theta_{s_1}, f), \dots, \mathbf{a}(\theta_{s_L}, f)]$. Recall that the path delays τ_l of (2) are unknown, as well as the matrices $\mathbf{B}(f)$ in (7). Therefore, we aim at estimating the per-carrier gain variances $\tilde{\mathbf{G}}(f)$ in (7). In particular, we denote by $\hat{\mathbf{G}}_{\mathcal{S}}(f)$ the estimated gain variances as associated with the angles in the support \mathcal{S} . Then, it is possible to use the estimator in [40], i.e., $\hat{\mathbf{G}}_{\mathcal{S}}(f) = \mathbf{\Psi}_{\mathcal{S}}^{\dagger}(f)(\hat{\mathbf{C}}_{\varphi(f)} - \sigma_{v(f)}^2\mathbf{I})(\mathbf{\Psi}_{\mathcal{S}}^{\dagger}(f))^H$. This solution clearly depends on the rank of $\mathbf{\Psi}_{\mathcal{S}}(f)$, and cannot be applied in the following situations:

1. $L > T_{tr}$. This scenario corresponds to short training sequences compared to the number of propagation paths. Under this assumption, the pseudo-inverse $\mathbf{\Psi}_{\mathcal{S}}^{\dagger}(f)$ cannot recover the L variance gains.
2. $L \leq T_{tr}$. In this case, the rank of $\mathbf{\Psi}_{\mathcal{S}}(f)$ depends on the choice of the dictionary matrix, \mathbf{A} , for a proper training matrix \mathbf{T} . Recall that the dictionary contains non-orthogonal vectors when M is finite. When the distance between two consecutive angles in the dictionary is small, a situation that arises when G is large, this may lead to dictionary matrices such that $\text{rank}(\mathbf{A}) \leq \min\{M, G\}$, thus making $\mathbf{\Psi}_{\mathcal{S}}^H(f)\mathbf{\Psi}_{\mathcal{S}}(f)$ ill-conditioned.

To circumvent (2), one possibility is to reduce the dictionary size (see Section 4.4). Alternatively, we propose to exploit the diagonal structure of $\tilde{\mathbf{G}}(f)$ and estimate it as

$$\text{vec}(\hat{\mathbf{G}}_{\mathcal{S}}(f)) = (\mathbf{\Psi}_{\mathcal{S}}^*(f) \circ \mathbf{\Psi}_{\mathcal{S}}(f))^{\dagger} \text{vec}(\hat{\mathbf{C}}_{\varphi(f)} - \sigma_{v(f)}^2\mathbf{I}_{T_{tr}}), \quad (18)$$

with \circ the Khatri–Rao product (or the column-wise Kronecker product). The product $\mathbf{\Psi}_{\mathcal{S}}^*(f) \circ \mathbf{\Psi}_{\mathcal{S}}(f)$ produces L -rank matrices when the columns of $\mathbf{\Psi}_{\mathcal{S}}(f)$ are not co-linear (see Appendix A).

Regarding the channel gain variances, it is again possible to exploit the channel structure to increase the estimation accuracy. Consider the relationship between the values of $\beta_l(f)$ across frequencies, in particular,

$$\beta_l(f) = \sum_{n=0}^{N-1} p_{rc}(nT_s - \tau_l) e^{-j2\pi fn/N_c} = \sum_{n=0}^{N-1} p_{rc}(nT_s - \tau_l) \left(e^{-j2\pi(-f)n/N_c} \right)^* = \beta_l^*(-f), \quad (19)$$

and, denoting $f' = -f + N_c$, we obtain $\beta_l(f) = \beta_l^*(f')$. Recall now that the L non-zero diagonal elements of $\tilde{\mathbf{G}}(f)$ are $\beta_l(f)\sigma_l\beta_l^*(f) = |\beta_l(f)|^2\sigma_l, l = 1, \dots, L$. This implies that $\tilde{\mathbf{G}}(f) = \tilde{\mathbf{G}}(f')$, and we can increase the robustness of the estimated gain variances in (18) with

$$\text{vec}(\hat{\mathbf{G}}_S(f)) = \frac{1}{2}((\mathbf{\Psi}_S^*(f) \circ \mathbf{\Psi}_S(f))^\dagger \text{vec}(\hat{\mathbf{C}}_{\varphi(f)} - \sigma_{v(f)}^2 \mathbf{I}) + (\mathbf{\Psi}_S^*(f') \circ \mathbf{\Psi}_S(f'))^\dagger \text{vec}(\hat{\mathbf{C}}_{\varphi(f')} - \sigma_{v(f')}^2 \mathbf{I})). \quad (20)$$

Observe that not all the gains can be calculated by means of the latter expression. In the case of odd N_c , $f = 0$ has to be determined using (18), whereas $f = 0$ and $f = N_c/2$ have to be computed via (18) for N_c even.

The proposed method, labeled as DR-W-MUSIC (Dictionary-Rotation Wideband MUSIC), is summarized in Algorithm 1.

Algorithm 1 Dictionary-Rotation Wideband MUSIC

```

1:  $\mathbf{g} \leftarrow [\sigma_1^2, \dots, \sigma_G^2]^T$ 
2: for all  $f \in \{1, \dots, N_c\}$  do
3:    $\hat{\mathbf{C}}_{\varphi(f)} \leftarrow \frac{1}{K} \sum_{k=1}^K \boldsymbol{\varphi}_k(f) \boldsymbol{\varphi}_k^H(f)$ 
4:    $\bar{\mathbf{U}}(f) \leftarrow$  compute noise basis (16)
5: end for
6: for all  $i \in \{1, \dots, G\}$  do
7:    $\mathcal{J}_i \leftarrow 0$ 
8:   for all  $f \in \{1, \dots, N_c\}$  do
9:      $\mathbf{a}(\theta_i, f) \leftarrow \mathbf{\Gamma}(\theta_i, f) \mathbf{a}(\theta_i)$  rotate dictionary
10:     $\boldsymbol{\psi}_i(f) \leftarrow \mathbf{T}(f) \mathbf{a}(\theta_i, f)$ 
11:     $\mathcal{J}_i \leftarrow \mathcal{J}_i + \|\boldsymbol{\psi}_i^H(f) \bar{\mathbf{U}}(f)\|_2^{-2}$ 
12:   end for
13: end for
14:  $\mathcal{S} \leftarrow L$  larger values on  $\mathcal{J}_i$ 
15: for all  $f \in \{1, \dots, N_c/2\}$  do
16:    $\mathbf{\Psi}_S(f) \leftarrow \mathbf{T}(f)[\mathbf{a}(\theta_{s_1}, f), \dots, \mathbf{a}(\theta_{s_L}, f)]$ 
17:    $\mathbf{\Psi}_S(f') \leftarrow \mathbf{T}(f')[\mathbf{a}(\theta_{s_1}, f'), \dots, \mathbf{a}(\theta_{s_L}, f')]$ 
18:    $\text{vec}(\hat{\mathbf{G}}_S(f)), \text{vec}(\hat{\mathbf{G}}_S(f')) \leftarrow$  compute using (20)
19: end for

```

4.3. Dictionary-Rotation Wideband Spatial Smoothing

Note that Algorithm 1 requires $\text{rank}(\hat{\mathbf{C}}_{\varphi}(f)) > L$, i.e., the number of estimated paths (or angles) is bounded by the number of training periods and the number of observations, i.e., $L \leq \min\{T_{\text{tr}}, K\}$. Solutions to address rank deficient scenarios have been proposed in [41,42]. In particular, these solutions apply when $\text{rank}(\hat{\mathbf{C}}_{\varphi}(f)) < L$. Nevertheless, in the proposed scenario, it is desirable to reduce the training sequence length T_{tr} as much as possible. Unfortunately, we obtain a more involved setup for $\text{rank}(\hat{\mathbf{C}}_{\varphi}) = T_{\text{tr}} < L$. This scenario can be addressed by using spatial smoothing.

Spatial Smoothing (SS) is an array processing technique based on nested arrays with different antenna separation [43]. An effect similar to that of the nested arrays is obtained thanks to the use of sparse rulers [44]. Again, the proposed methods addressed the identification of signals lying in a single subspace, and we have to resort to dictionary rotations to exploit the wideband channel structure.

We start by introducing the sample covariance matrix vectorization $\mathbf{y}(f) = \text{vec}(\hat{\mathbf{C}}_{\varphi(f)})$, i.e.,

$$\mathbf{y}(f) = (\mathbf{\Psi}^*(f) \circ \mathbf{\Psi}(f)) \text{diag}(\tilde{\mathbf{G}}(f)) + \sigma_{v(f)}^2 \mathbf{e}', \quad (21)$$

where $e' = [e_1^T, \dots, e_{T_{tr}}^T]^T$. Observe that, due to the utilization of a perfect sparse ruler training $T(f)$, $(\Psi^*(f) \circ \Psi(f))$ contains the products corresponding to the spatial differences $z \in \{-M+1, M-1\}$. Conventional SS in [44] discards the repeated distances. However, this means employing $2M-1$ elements out of T_{tr}^2 from $y(f)$ in (22). Hence, we propose to refine the estimation by averaging over all the samples from the same inter-antenna distance. To that end, let us introduce $y_z^j(f)$, $j = 1, \dots, J_z$, $z = -M+1, \dots, M-1$ as the j -th sample of the z -th difference in $y(f)$. Note that J_z represents the number of samples available for the z -th difference, and its values depend on the sparse ruler structure and T_{tr} . Next, we introduce the reduced vector $\check{y}(f) \in \mathbb{C}^{2M-1}$ with sorted elements $[\check{y}(f)]_z = \frac{1}{J_z} \sum_{j=1}^{J_z} y_z^j(f)$ corresponding to the $2M-1$ distances. Accordingly, we define the matrix $\mathbf{Z}(f)$ containing $2M-1$ sorted rows from $(\Psi^*(f) \circ \Psi(f))$. Thus, the reduced vector is

$$\check{y}(f) = \mathbf{Z}(f) \text{diag}(\check{\mathbf{G}}(f)) + \sigma_{v(f)}^2 e_M. \quad (22)$$

These differences in $\check{y}(f)$ 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 $\check{y}_m(f) \in \mathbb{C}^M$, the spatial smoothed matrix is obtained as follows: $\check{\mathbf{Y}}(f) = \frac{1}{M} \sum_{m=1}^M \check{y}_m(f) \check{y}_m^H(f)$. As shown in [43], $\check{\mathbf{Y}}(f) = \check{\mathbf{Y}}^{1/2}(f) \check{\mathbf{Y}}^{1/2}(f)$ where the matrix $\check{\mathbf{Y}}^{1/2}(f)$ can be rewritten as

$$\check{\mathbf{Y}}^{1/2}(f) = \frac{1}{\sqrt{M}} (\mathbf{A}(f) \check{\mathbf{G}}(f) \mathbf{A}^H(f) + \sigma_{v(f)}^2 \mathbf{I}_M). \quad (23)$$

Due to the beam squint effect, (23) has to be computed individually for each subcarrier f . Nevertheless, we desire to achieve the benefits of exploiting the structure of the wideband channel. In this regard, we employ $\sqrt{M} \check{\mathbf{Y}}^{1/2}(f) = \mathbf{U}(f) \mathbf{\Lambda}(f) \mathbf{U}^H(f)$ as the sample covariance matrix in line 3 of Algorithm 1. Then, following Algorithm 1 steps, the L eigenvectors corresponding to the L largest eigenvalues are discarded to obtain an incomplete basis $\bar{\mathbf{U}}(f)$ spanning the noise subspace. Hence, the wideband estimator incorporating the dictionary rotation capability of line 11 is updated as $\mathcal{J}_i = \mathcal{J}_i + \|\mathbf{a}^H(\theta_i, f) \bar{\mathbf{U}}(f)\|_2^{-2}$, where θ_i is the i -th angle contained in the dictionary matrix \mathbf{A} . To estimate the variance gains, we follow a similar procedure as that in Section 4.2 to derive the robust estimator

$$\text{vec}(\hat{\mathbf{G}}_S(f)) = \frac{1}{2} ((\mathbf{A}_S^*(f) \circ \mathbf{A}_S(f))^\dagger \text{vec}(\check{\mathbf{Y}}(f)) + (\mathbf{A}_S^*(f') \circ \mathbf{A}_S(f'))^\dagger \text{vec}(\check{\mathbf{Y}}(f'))),$$

with $\mathbf{A}_S(f) = [\mathbf{a}(\theta_{s_1}, f), \dots, \mathbf{a}(\theta_{s_L}, f)]$ and $\check{\mathbf{Y}}(f) = \sqrt{M} \check{\mathbf{Y}}^{1/2}(f) - \sigma_{v(f)}^2 \mathbf{I}_M$. This procedure replaces lines 16 to 18, and completes the posed DR-W-SS method.

4.4. Dictionary Size

In this section, we provide some insight regarding the choice of the dictionary size G . Let us start by determining a sufficient condition for unique angular identification. To that end, we first define the Kruskal rank of a matrix \mathbf{X} , $\text{krank}(\mathbf{X})$. If $\text{krank}(\mathbf{X}) = x$, then x is the maximum number such that any x columns of \mathbf{X} are linearly independent.

Theorem 2. *In the noiseless scenario, and assuming AoD lying on the dictionary $\mathbf{A}(f)$ and $\text{rank}(\hat{\mathbf{C}}_{\varphi(f)}) = L$, the inequality $L < \text{krank}(\Psi(f))$ guarantees the unique identification of the L AoD [41]. See [45] for the proof.*

Therefore, when $\text{krank}(\mathbf{A})$ is small, the probability of identifying false directions increases. To illustrate the dependence with G , we compute in Figure 2 an upper bound of the Kruskal rank for G and M . This upper bound is computationally feasible and takes into account only z consecutive columns, a reasonable approximation due to the structure of \mathbf{A} . As shown in the figure, large G values complicate the AoD detection, which suggests to reduce the dictionary size. However, this strategy entails a power leakage due to basis mismatching, that is, a performance loss due to off-grid AoDs. However, the basis mismatching is moderate in practical settings due to the finite angular resolution

of the ULA. For instance, a typical inter-antenna spacing $d = \lambda/2$ entails an angular resolution of about $\frac{2}{M}$ radians [46], leading to $G \approx \frac{\Delta\theta\pi}{360^\circ} M$ for an angular range $\Delta\theta$. That is, $G \approx M$ provides enough resolution to cover the sector $\Delta\theta = 120^\circ$. In addition, an alternative to increasing the dictionary size is to overestimate the number of channel propagation paths L . This strategy is explored in Section 6.2.

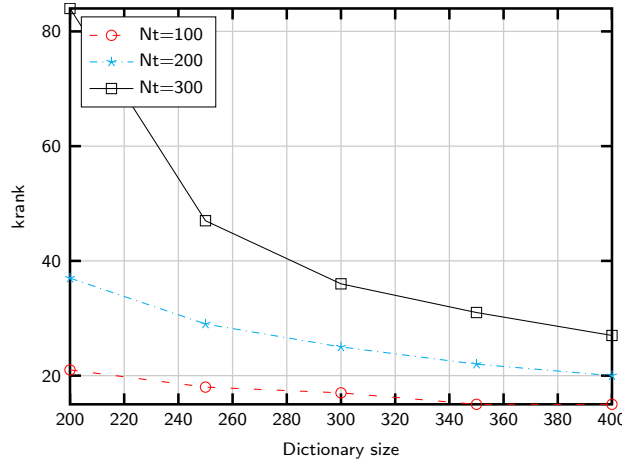


Figure 2. Kruskal rank upper bound for different array and dictionary sizes.

5. Delay-Gain Channel Estimator

As a means to estimate the channel, we propose a (DG) Delay-Gains estimator that exploits the structure of the wideband channels. This method employs the observations for all the subcarriers $\varphi_k(f)$ to estimate the frequency independent gains of each propagation path g_k (cf. (6)). Thus, even in the case where certain frequencies present very poor SNR, their corresponding channels can be estimated by using the information from the remaining frequencies.

The covariance estimation procedure provides the measurement matrix $\Psi_S(f)$ and the estimated gain variances $\hat{G}_S(f)$ for all the frequencies. Using this information, we estimate the delay matrices $\mathbf{B}(f)$ and the gain vector \mathbf{g}_k from the observations gathered in the covariance estimation stage. Next, a robust low complexity estimator is calculated for the frequency selective channels.

First, notice that the observations in (8) can be rewritten as $\varphi_k(f) = \Psi_S(f)\mathbf{B}_S(f)\mathbf{g}_{S,k} + \mathbf{v}_k(f)$, where $\Psi_S(f)$ was introduced in Section 4.2, $\mathbf{B}_S(f) = \text{diag}(\beta_{s_1}(f), \dots, \beta_{s_L}(f))$ and $\mathbf{g}_{S,k}$ contains the gains for the positions s_1, \dots, s_L of \mathbf{g}_k . Then, we focus on the unknown vector $\zeta_k(f) = \mathbf{B}_S(f)\mathbf{g}_{S,k}$ that contains the channel gains affected by the shaping pulse. Accordingly, the linear Minimum Mean Squared Error (MMSE) estimator of $\zeta_k(f)$ reads as

$$\hat{\zeta}_k(f) = \hat{G}_S(f)\Psi_S^H(f) \left(\Psi_S(f)\hat{G}_S(f)\Psi_S^H(f) + \sigma_{v(f)}^2 \mathbf{I}_{T_{tr}} \right)^{-1} \varphi_k(f). \quad (24)$$

Once $\hat{\zeta}_k(f)$ is determined, we have to identify the matrices $\mathbf{B}_S(f)$ before performing the estimation of $\mathbf{g}_{S,k}$ jointly using the vectors $\hat{\zeta}_k(f)$ for all the frequencies. Since $\mathbf{B}_S(f)$ contains samples of the shaping filter $p_{rc}(\cdot)$ in the frequency domain, we obtain $\mathbf{B}_S(f)$ by estimating the delays corresponding to the L channel paths. In particular, to determine delay for the l -th path τ_l , we collect the entries as associated with the l -th path for all the frequencies, i.e., $[\hat{\zeta}_k(f)]_l$ for $f = 1, \dots, N_c$, in a vector. Then, we multiply this vector times the first N columns of the Discrete Fourier Transform (DFT) matrix, $\mathbf{F} \in \mathbb{C}^{N_c \times N}$, that is

$$\zeta_{l,k} = \mathbf{F}^H \left[[\hat{\zeta}_k(1)]_l, \dots, [\hat{\zeta}_k(N_c)]_l \right]^T. \quad (25)$$

Observe now that $\zeta_{l,k} = g_{l,k} \mathbf{p}_{rc}(\tau_l) + \mathbf{w}_k$ contains the aforementioned N stacked pulse samples $\mathbf{p}_{rc}(\tau_l) = [p_{rc}(-\tau_l), p_{rc}(T_s - \tau_l), \dots, p_{rc}((N-1)T_s - \tau_l)]^T$, and the estimation noise $\mathbf{w}_k \sim \mathcal{N}_C(\mathbf{0}, \sigma_w^2 \mathbf{I}_{N_c})$. Unfortunately, these samples are scaled by the unknown complex-valued channel gain g_l of (2). Note that we assume that the error is statistically independent for different frequencies. Hence, for given shaping filter $p_{rc}(t)$, we propose the following estimator function for τ

$$\mathcal{D}(\tau) = \sum_{k=1}^K \frac{|\mathbf{p}_{rc}^T(\tau) \zeta_{l,k}|}{\|\mathbf{p}_{rc}(\tau)\|_2 \|\zeta_{l,k}\|_2}, \quad (26)$$

where $\mathbf{p}_{rc}(\tau) = [p_{rc}(-\tau), p_{rc}(T_s - \tau), \dots, p_{rc}((N-1)T_s - \tau)]^T$ contains the pulse samples for the delay τ . Unfortunately, $\mathcal{D}(\tau)$ is a non-monotonic function. Thus, we minimize $\mathcal{D}(\tau)$ over τ , e.g., by linear search over $[0, (N-1)T_s]$ to obtain the delay estimate $\hat{\tau}_l$. In the high SNR regime, it is enough to search within the interval corresponding to the two samples of $\mathbf{p}_{rc}(\tau)$ with larger absolute values.

Recall that the channel gains are multiplied times $\mathbf{B}_S(f)$ in the unknown vector $\zeta_k(f)$ of (24). Thus, employing the estimated delays $\hat{\tau}_l$, we compute $\hat{\mathbf{B}}_S(f)$ for all frequencies (see (6)). To eventually estimate the channel gains $g_{S,k}$, we compute the minimum variance unbiased estimator [47] using information from all the frequencies, i.e., $\hat{\mathbf{g}}_{S,k} = \hat{\mathbf{B}}^\dagger \hat{\zeta}_k$, where $\hat{\mathbf{B}} = [\hat{\mathbf{B}}_S^T(1), \dots, \hat{\mathbf{B}}_S^T(N_c)]^T$ and $\hat{\zeta}_k = [\hat{\zeta}_k^T(1), \dots, \hat{\zeta}_k^T(N_c)]^T$. Finally, by invoking (6), the estimated channel yields

$$\hat{\mathbf{h}}_k(f) = \mathbf{A}_S(f) \hat{\mathbf{B}}_S(f) \hat{\mathbf{g}}_{S,k}. \quad (27)$$

6. Simulation Results

The following setup is considered for the numerical experiments. We assume a training sequence \mathbf{T} generated from a Wichmann ruler of length T_{tr} . The number of transmit antennas is $M = 200$, while we consider a hybrid architecture using $N_{RF} = 2$ RF chains, $N_c = 64$ subcarriers, and a dictionary of size $G = 400$ with equally spaced angles within the range $[0, \pi]$. We consider the channel covariance model of (6) with L propagation paths uniformly distributed in $[0, \pi]$; different number of training periods K , and $N = 8$ random delay taps uniformly distributed in $[0, (N-1)T_s]$, with $T_s = 1/B$. The numerical results are averaged over 200 channel covariance realizations for independent users, and over the different subcarriers N_c . To evaluate the accuracy of the covariance identification, we employ the Normalized Mean Squared Error (NMSE) metric defined as $\text{NMSE} = \frac{\|\hat{\mathbf{C}}_h - \mathbf{C}_h\|_F^2}{\|\mathbf{C}_h\|_F^2}$ with the Frobenius norm $\|\cdot\|_F$. In the case of channel estimation, we use the efficiency $\eta \in [0, 1]$ given by $\eta = \frac{|\hat{\mathbf{h}}^H \mathbf{h}|}{\|\hat{\mathbf{h}}\| \|\mathbf{h}\|}$, which is similar to the figures of merit in e.g., [26,34,48]. This insightful metric evaluates the signal power lost by beamforming using the estimated channels. Note that this metric is averaged over the number of channel blocks.

As a benchmark, we include an LS (known θ) curve that individually estimates the gain variances for each subcarrier using (18) assuming that the channel angles are known. That is, it does not exploit the symmetry with respect to the central frequency of (19) to estimate the variance gain.

6.1. Wideband Covariance Estimation

Figure 3a shows the NMSE obtained with the proposed algorithms. The training sequence length is set to $T_{tr} = 25$ and the channel parameters are adjusted according to massive MIMO settings with $f_c = 5$ GHz, $B = 20$ MHz, $L = \{15, 30\}$ channel propagation paths, and a SNR of 15 dB. Using the same training sequence, we compare our methods with COMP in [34], which supports wideband estimation in the absence of beam squint. For $L = 15$, the proposed methods achieve better performance than COMP when the number of training blocks K is large, but the accuracy of the estimation deteriorates when $K < 20$. This behavior is motivated by the lack of similarity between the sample covariance matrix $\hat{\mathbf{C}}_{\varphi(f)}$ and the actual channel covariance matrix $\mathbf{C}_{\varphi(f)}$ that

makes difficult the discrimination of the signal and noise subspaces (see Algorithm 1). Nevertheless, values of K above 50 are feasible in practical settings, since K is not restricted by the channel coherence time. We also include the case of $L = 30$ channel paths in Figure 3a. Under such as sumption, COMP and DR-W-MUSIC do not apply due to the lack of angular sparsity, and the limited rank of the sample covariance matrices, respectively. Conversely, the DR-W-SS approach performs close to the benchmark with known angles, denoted by LS (known θ).

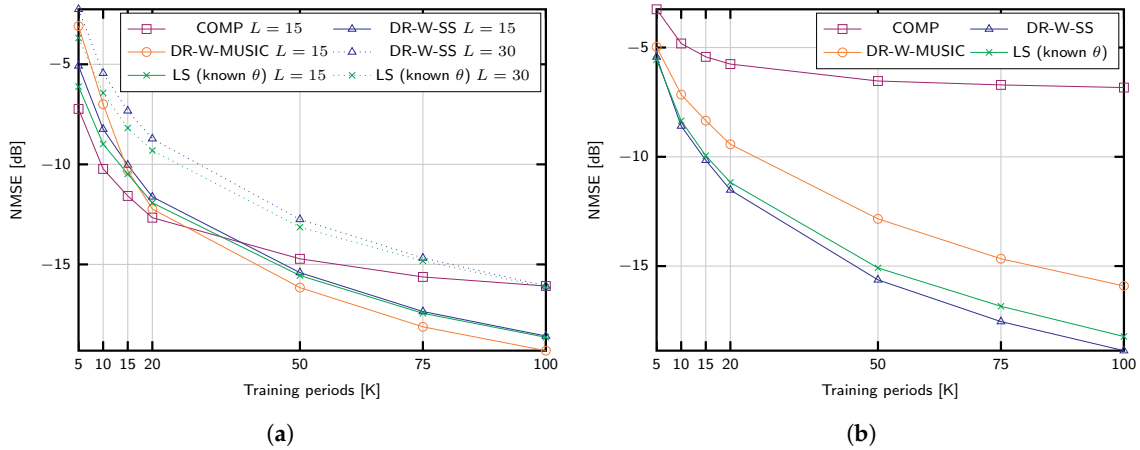


Figure 3. NMSE of different covariance estimation strategies for $M = 200$ antennas: (a) SNR = 15 dB, and $L = \{15, 30\}$ propagation paths; (b) SNR = 0 dB, and $L = 5$ paths.

Figure 3a shows the NMSE for a mmWave setup with $L = 5$, SNR = 0 dB, a small signal bandwidth $B = 200$ MHz and $f_c = 28$ GHz. The robustness against the noise of the proposed strategies is remarkable, especially in the case of DR-W-SS. On the contrary, COMP does not achieve good performance results in this SNR regime compared to the proposed methods.

To evaluate the performance impact of the training sequence length T_{tr} , we fix the parameters to $M = 100$, $L = 50$, SNR = 15 dB and $K = 50$ in Figure 4a. The training sequence is shorter than the number of paths and takes the values 19, 25, 32, 40, 50, to show the capability of DR-W-SS in the challenging scenario $L > T_{tr}$. As shown, the proposed method performs better than the LS benchmark with known support \mathcal{S} .

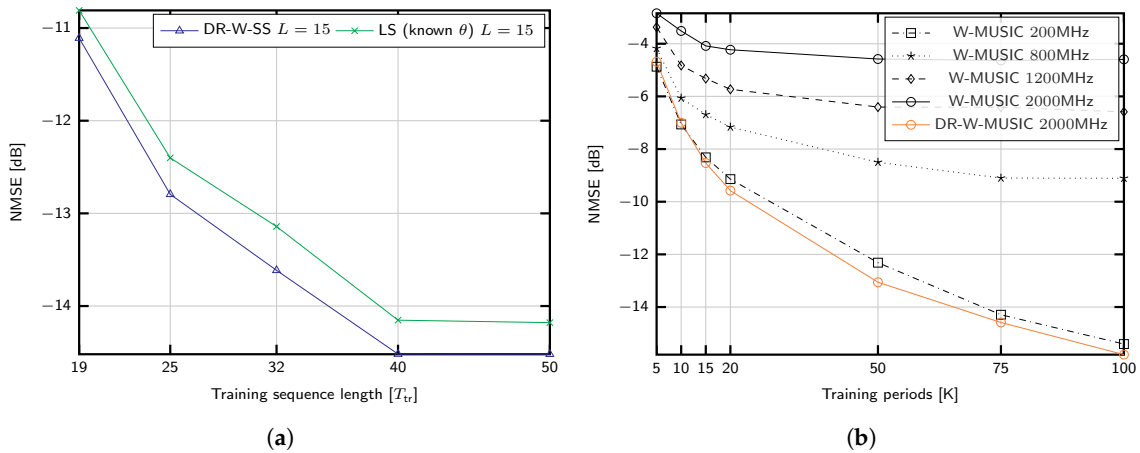


Figure 4. NMSE of covariance estimation for (a) $M = 100$ antennas, $L = 50$ propagation paths, $K = 50$, SNR = 15 dB, and different training sequence lengths; (b) $M = 200$ antennas, $L = 5$ propagation paths, and different signal bandwidths B for $f_c = 28$ GHz.

We now evaluate the consequences of neglecting the beam squint effect for the setup considered in Figure 3a, that is, $f_c = 28$ GHz and different signal bandwidths, $B = 200, 800, 1200, 2000$ MHz. The methods referred to as W-MUSIC consist of Algorithm 1 ignoring the dictionary rotation. Figure 4b exhibits the performance impact of beam squint due to the lack of a common subspace for different frequencies. As shown, the proposed dictionary rotation is able to exploit the structure even in the case of $B = 2000$ MHz.

6.2. Covariance Subspace Estimation

In the previous numerical examples, the number of paths L was assumed to be known. Subspace estimation methods, like the one in [49], can be used to estimate L by comparing the differences between consecutive eigenvalues of the sample covariance matrix with a threshold ϵ . In the wideband case, we employ the information of all frequencies jointly, to increase the detection accuracy. Moreover, the computational cost over DR-W-MUSIC and DR-W-SS is negligible while it is significant for other approaches not performing a subspace discrimination.

Figure 5a shows the robustness against uncertainty in the number of channel propagation paths L . We employ Algorithm 2 in [49] and empirically adjust the parameter ϵ for the different estimation methods. For the considered scenario with SNR = 0 dB, we set the threshold as $\epsilon = 0.025$ for DR-W-MUSIC and $\epsilon = 0.01$ for DR-W-SS. Remarkably, the performance obtained with subspace estimation for DR-W-SS is similar to that of the strategies with known L . In particular, for DR-W-MUSIC, overestimating the number of channel paths L improves the performance. This observation reveals that the support identified by Algorithm 1 with known L was incorrect. Recall that the support \mathcal{S} comprises the angles corresponding to the L larger values on the estimator function \mathcal{J}_i . Under the strong noise conditions considered in Figure 5a, some angles belonging to the covariance support might be ignored (see line 14 of Algorithm 1). Instead, step 14 selects angles in the dictionary whose steering vectors are linearly dependent. Thus, overestimating L enables a better detection of the channel covariance subspace where angles with lower \mathcal{J}_i are included in \mathcal{S} . In addition, we compare the effect of discarding the repeated distances in DR-W-SS, as performed in conventional SS [44]. In Figure 5a, this strategy is labeled DR-W-SS Discard. As shown, for $T_{tr} = 50$, the performance loss is significant.

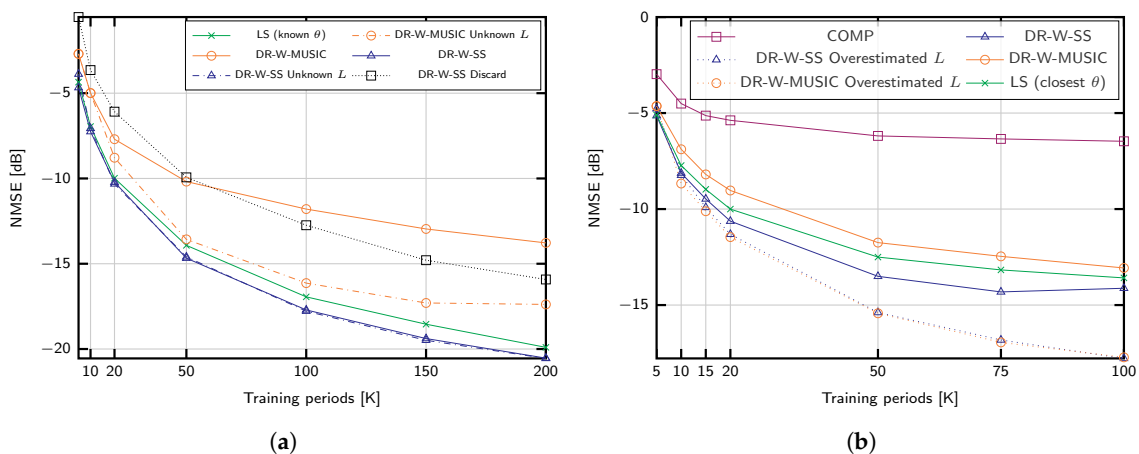


Figure 5. NMSE of covariance estimation for (a) $M = 200$ antennas, $L = 15$ propagation paths, $T_{tr} = 50$, SNR = 0 dB, and subspace estimation; (b) $M = 200$, SNR = 0 dB, $L = 5$ propagation paths, and off-grid AoDs.

In Section 4.4, we discussed the advantages and practical orientation of moderate dictionary sizes. Nevertheless, we investigate the effect of off-grid AoDs in Figure 5b for the mmWave setup and $B = 200$ MHz. The error floor due to basis mismatch is clear. As an alternative to increase G ,

we propose to compensate the basis inaccuracy by overestimating the number of channel paths L . This way, part of the power leakage is recovered with the additional AoDs. These approaches are denoted DR-W-MUSIC Overestimated L and DR-W-SS Overestimated L . For the LS estimator with known angles, we choose the closest dictionary angles to the real ones.

6.3. Wideband Channel Estimation

Figure 6a shows the performance of the channel estimator with the covariance identified and the two different channel estimation methods: LMMSE [21,25,47]

$$\hat{\mathbf{h}}_k(f) = \hat{\mathbf{C}}_{h(f)} \mathbf{T}^H \left(\mathbf{T} \hat{\mathbf{C}}_{h(f)} \mathbf{T}^H + \sigma_{v(f)}^2 \mathbf{I}_{T_{tr}} \right)^{-1} \boldsymbol{\varphi}_k(f) \quad (28)$$

and the proposed DG estimator of (27). The number of channel blocks is set to $K = 100$, and the number of delay candidates τ considered in (26) is $20N$. Since the proposed DG method uses the information of all the subcarriers to produce the estimates, it is very robust against noise and results in a considerable performance gain. Indeed, more than 80% of the signal power is captured at -10 dB with the proposed strategies.

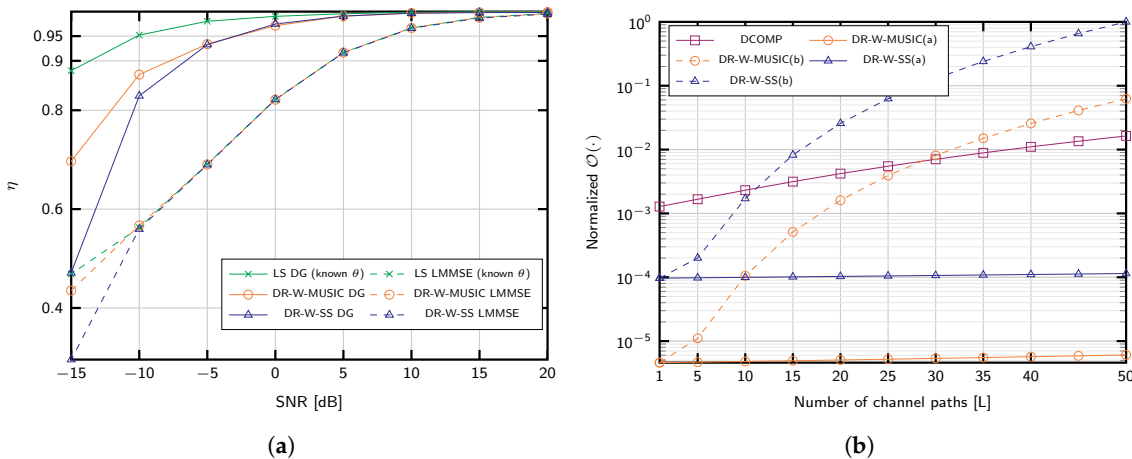


Figure 6. (a) efficiency of channel estimation for $M = 200$ antennas, $L = 5$ propagation paths, $T_{tr} = 25$, and $K = 100$ training periods; (b) normalized computational complexity for $M = 200$ antennas, dictionary size $G = 2M$, $K = 150$ training periods, training length $T_{tr} = 50$, $N_c = 64$ subcarriers, and different number of paths.

6.4. Computational Complexity

Table 1 summarizes the computational complexity of the proposed methods. Recall that a fast covariance estimation is important, but not as critical as the actual channel estimation. Indeed, we only employ a training stage contrarily to other approaches like [21,37].

For the favorable scenario, DR-W-MUSIC(a) and DR-W-SS(a), the gain variances are estimated using well-conditioned measurement matrices, while (18) with the Khatri–Rao product is considered for MUSIC(b) and DR-W-SS(b). The complexities have been calculated as summing $T_{tr} > L$ for DR-W-MUSIC and COMP. It is noticeable that neither DR-W-MUSIC nor COMP depend on the number of antennas, although both of them linearly depend on the dictionary size G . The number of channel taps N is also relatively small, and T is the number of τ candidates evaluated to estimate the delays using (26). Since COMP and DR-W-MUSIC apply to scenarios where L and T_{tr} are very small compared to M or G , the complexity of the latter is smaller in general. Furthermore, the number of steps performed by COMP is larger, and its computational burden depends on the number of training periods K .

Table 1. Complexities of estimation methods.

Covariance Estimation	
DR-W-MUSIC(a)	$\mathcal{O}(N_c(T_{tr}^3 + (T_{tr}^2 + T_{tr})G + L^2T_{tr} + 2LT_{tr}^2))$
DR-W-MUSIC(b)	$\mathcal{O}(N_c(T_{tr}^3 + (T_{tr}^2 + T_{tr})G + L^4T_{tr}^2 + L^2T_{tr}^2))$
DCOMP	$\mathcal{O}(N_c(2KT_{tr}^2G + KL^3T_{tr} + 2KT_{tr}^2(L^2 + L) + KLT_{tr}^3))$
DR-W-SS(a)	$\mathcal{O}(N_c(M^2 + M^3 + (M^2 + M)G + L^2M + 2LM^2))$
DR-W-SS(b)	$\mathcal{O}(N_c(M^2 + M^3 + (M^2 + M)G + L^4M^2 + L^2M^2))$
Computing Estimator	
DG	$\mathcal{O}(N_c(3L^2T_{tr} + T_{tr}^3 + NL + L^3 + ML^2 + ML) + KN TL)$
LMMSE (28)	$\mathcal{O}(N_c(3M^2T_{tr} + T_{tr}^3))$
Channel Estimation	
DG	$\mathcal{O}(N_c(LT_{tr} + ML + L^2))$
LMMSE (28)	$\mathcal{O}(N_cMT_{tr})$

When $L \ll M$, the computational complexity virtually depends on the number of antennas. Hence, DR-W-SS and DR-W-MUSIC establish a trade-off among T_{tr} , L and M , since we could use the less complex option by adjusting T_{tr} and K . The effect of increasing L is depicted in Figure 6b. Interestingly, in the region where the channel is sparse and Dynamic COMP (DCOMP) applies, $L < 10$, the computational complexity of the proposed approaches is lower than that of DCOMP.

We now differentiate between the operations computed once using the estimated covariance matrix, referred to as “Computing Estimator”, and the channel estimation for each observation, referred to as “Channel Estimation”. Remarkably, the complexity of computing the proposed estimator (DG) linearly depends on M , cf. (24)–(26) whereas for the LMMSE estimator this dependency is quadratic.

Regarding the actual channel estimation, in scenarios suitable for DR-W-MUSIC with $L < T_{tr}$, DG channel estimation is more efficient. This applies to mmWave frequencies and other massive MIMO scenarios, e.g., [8,37,50–53]. Conversely, LMMSE estimator might be less complex for large number of channel propagation paths, i.e., $L > \sqrt{M}$.

7. Conclusions

In this work, we proposed a covariance and channel estimation method for FDD wideband hybrid massive MIMO systems. The numerical experiments showed the good performance of the posed covariance estimation approaches compared to other methods in the literature, even when considering very short training sequence lengths. In addition, we developed a robust channel estimator with computational complexity depending linearly on the number of antennas. Both facts contributed to a reduction in the channel training overhead. Finally, we evaluated the performance of the proposed scheme for a wide variety of scenarios. The extension to Uniform Planar Array (UPA) antenna arrangements, and the use of shorter training sequences is an interesting research line for future works.

Author Contributions: Conceptualization, J.P.G.-C. and P.S.-C.; formal analysis, J.P.G.-C. and P.S.-C.; investigation, J.P.G.-C. and P.S.-C.; writing—original draft preparation, J.P.G.-C., P.S.-C., P.M.C. and L.C.; writing—review and editing, P.M.C. and L.C.; funding acquisition, P.M.C. and L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded by Xunta de Galicia (ED431G2019/01), Agencia Estatal de Investigación of Spain (TEC2016-75067-C4-1-R, RED2018-102668-T, PID2019-104958RB-C42), and ERDF funds of the EU (AEI/FEDER, UE).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

To prove the linear independence, we consider two cases: (1) If the columns of $\Psi_S(f)$ are linearly independent, the same property holds for the columns of $\Psi_S^*(f) \circ \Psi_S(f)$. (2) Consider that the i -th

column of $\Psi_{\mathcal{S}}(f)$, $\psi_{\mathcal{S}_i}$ for notation simplicity, is a linear combination of the columns whose indices are in the set $\mathcal{T} \subseteq \mathcal{S}$, i.e., $\psi_{\mathcal{S}_i} = \sum_{j=1}^{|\mathcal{T}|} c_j \psi_{\mathcal{T}_j}$ with $\psi_{\mathcal{T}_j} \neq \mathbf{0}, \forall j \in \mathcal{T}$. Then, for the i -th column of $(\Psi_{\mathcal{S}}^*(f) \circ \Psi_{\mathcal{S}}(f))$, we obtain

$$\sum_{j=1}^{|\mathcal{T}|} c_j^* \psi_{\mathcal{T}_j}^* \otimes \sum_{n=1}^{|\mathcal{T}|} c_n \psi_{\mathcal{T}_n} = \sum_{j=1}^{|\mathcal{T}|} \sum_{n=1}^{|\mathcal{T}|} c_j^* \psi_{\mathcal{T}_j}^* \otimes c_n \psi_{\mathcal{T}_n}. \quad (\text{A1})$$

Notice that it is not possible to write this column as a linear combination of the form $\sum_{j=1}^{|\mathcal{T}|} b_j \psi_{\mathcal{T}_j}^* \otimes \psi_{\mathcal{T}_j}$ unless $\psi_{\mathcal{T}_j} = \sum_{n=1}^{|\mathcal{T}|} c_n \psi_{\mathcal{T}_n}, \forall j$. This condition holds only if $\psi_{\mathcal{S}_i} = c_j \psi_{\mathcal{T}_j}$.

References

- Rusek, F.; Persson, D.; Lau, B.K.; Larsson, E.G.; Marzetta, T.L.; Edfors, O.; Tufvesson, F. Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays. *IEEE Signal Process. Mag.* **2013**, *30*, 40–60. [\[CrossRef\]](#)
- Alkhateeb, A.; Mo, J.; González-Prelcic, N.; Heath, R.W. MIMO Precoding and Combining Solutions for Millimeter-Wave Systems. *IEEE Commun. Mag.* **2014**, *52*, 122–131. [\[CrossRef\]](#)
- Björnson, E.; Hoydis, J.; Sanguinetti, L. Massive MIMO Has Unlimited Capacity. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 574–590. [\[CrossRef\]](#)
- Gao, X.; Edfors, O.; Tufvesson, F.; Larsson, E.G. Massive MIMO in Real Propagation Environments: Do All Antennas Contribute Equally? *IEEE Trans. Commun.* **2015**, *63*, 3917–3928. [\[CrossRef\]](#)
- Hassibi, B.; Hochwald, B.M. How much training is needed in multiple-antenna wireless links? *IEEE Trans. Inf. Theory* **2003**, *49*, 951–963. [\[CrossRef\]](#)
- Marzetta, T.L. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Trans. Wirel. Commun.* **2010**, *9*, 3590–3600. [\[CrossRef\]](#)
- Mailloux, R. *Phased Array Antenna Handbook*; Artech House: Norwood, MA, USA, 2018.
- Gao, Z.; Dai, L.; Wang, Z.; Chen, S. Spatially Common Sparsity Based Adaptive Channel Estimation and Feedback for FDD Massive MIMO. *IEEE Trans. Signal Process.* **2015**, *63*, 6169–6183. [\[CrossRef\]](#)
- Luo, X.; Zhang, X.; Cai, P.; Shen, C.; Hu, D.; Qian, H. DL CSI acquisition and feedback in FDD massive MIMO via path aligning. In Proceedings of the Ninth International Conference on Ubiquitous and Future Networks (ICUFN), Milan, Italy, 4–7 July 2017; pp. 349–354. [\[CrossRef\]](#)
- Björnson, E.; der Perre, L.V.; Buzzi, S.; Larsson, E.G. Massive MIMO in Sub-6 GHz and mmWave: Physical, Practical, and Use-Case Differences. *CoRR* **2018**. [\[CrossRef\]](#)
- Van Trees, H.L. *Optimum Array Processing*; Wiley: Hoboken, NJ, USA, 2002. [\[CrossRef\]](#)
- Neumann, D.; Wiese, T.; Utschick, W. Learning the MMSE Channel Estimator. *IEEE Trans. Signal Process.* **2018**, *66*, 2905–2917. [\[CrossRef\]](#)
- Gao, Z.; Hu, C.; Dai, L.; Wang, Z. Channel Estimation for Millimeter-Wave Massive MIMO with Hybrid Precoding Over Frequency-Selective Fading Channels. *IEEE Commun. Lett.* **2016**, *20*, 1259–1262. [\[CrossRef\]](#)
- González-Coma, J.P.; Rodríguez-Fernández, J.; González-Prelcic, N.; Castedo, L.; Heath, R.W. Channel Estimation and Hybrid Precoding for Frequency Selective Multiuser mmWave MIMO Systems. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 353–367. [\[CrossRef\]](#)
- Haghighatshoar, S.; Caire, G. Massive MIMO Pilot Decontamination and Channel Interpolation via Wideband Sparse Channel Estimation. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 8316–8332. [\[CrossRef\]](#)
- Sayeed, A.M. Deconstructing multiantenna fading channels. *IEEE Trans. Signal Process.* **2002**, *50*, 2563–2579. [\[CrossRef\]](#)
- Sun, G.R.M.J.S.; Rappaport, T.S. A novel millimeter-wave channel simulator and applications for 5G wireless communications. In Proceedings of the International Conference on Communications (ICC), Washington, DC, USA, 2–7 July 2017.
- Gao, X.; Tufvesson, F.; Edfors, O.; Rusek, F. Measured propagation characteristics for very-large MIMO at 2.6 GHz. In Proceedings of the Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, USA, 4–7 November 2012; pp. 295–299. [\[CrossRef\]](#)

19. Sanguinetti, L.; Björnson, E.; Hoydis, J. Towards Massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination. *arXiv* **2019**, arXiv:1904.03406.
20. Özdoğan, O.; Björnson, E.; Larsson, E.G. Massive MIMO With Spatially Correlated Rician Fading Channels. *IEEE Trans. Commun.* **2019**, *67*, 3234–3250. [[CrossRef](#)]
21. Adhikary, A.; Nam, J.; Ahn, J.Y.; Caire, G. Joint Spatial Division and Multiplexing: The Large-Scale Array Regime. *IEEE Trans. Inf. Theory* **2013**, *59*, 6441–6463. [[CrossRef](#)]
22. Zhang, B.; Cimini, L.J.; Greenstein, L.J. Efficient Eigenspace Training and Precoding for FDD Massive MIMO Systems. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), Singapore, 4–10 December 2017; pp. 1–6. [[CrossRef](#)]
23. Neumann, D.; Joham, M.; Utschick, W. Covariance Matrix Estimation in Massive MIMO. *IEEE Signal Process. Lett.* **2018**, *25*, 863–867. [[CrossRef](#)]
24. Upadhyaya, K.; Vorobyov, S.A. Covariance Matrix Estimation for Massive MIMO. *IEEE Signal Process. Lett.* **2018**, *25*, 546–550. [[CrossRef](#)]
25. Xie, H.; Gao, F.; Jin, S.; Fang, J.; Liang, Y. Channel Estimation for TDD/FDD Massive MIMO Systems with Channel Covariance Computing. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 4206–4218. [[CrossRef](#)]
26. Zhang, X.; Zhong, L.; Sabharwal, A. Directional Training for FDD Massive MIMO. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 5183–5197. [[CrossRef](#)]
27. Khalilsarai, M.B.; Haghghatshoar, S.; Yi, X.; Caire, G. FDD Massive MIMO via UL/DL Channel Covariance Extrapolation and Active Channel Sparsification. *CoRR* **2018**. [[CrossRef](#)]
28. Jian, M.; Gao, F.; Tian, Z.; Jin, S.; Ma, S. Angle-Domain Aided UL/DL Channel Estimation for Wideband mmWave Massive MIMO Systems With Beam Squint. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 3515–3527. [[CrossRef](#)]
29. Wang, B.; Jian, M.; Gao, F.; Li, G.Y.; Lin, H. Beam Squint and Channel Estimation for Wideband mmWave Massive MIMO-OFDM Systems. *IEEE Trans. Signal Process.* **2019**, *67*, 5893–5908. [[CrossRef](#)]
30. Pal, P.; Vaidyanathan, P.P. Coprime sampling and the music algorithm. In Proceedings of the Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE), Sedona, AZ, USA, 4–7 January 2011; pp. 289–294. [[CrossRef](#)]
31. Haghghatshoar, S.; Caire, G. Low-Complexity Massive MIMO Subspace Estimation and Tracking From Low-Dimensional Projections. *IEEE Trans. Signal Process.* **2018**, *66*, 1832–1844. [[CrossRef](#)]
32. Romero, D.; Leus, G. Compressive covariance sampling. In Proceedings of the Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 2–7 February 2013; pp. 1–8. [[CrossRef](#)]
33. Shakeri, S.; Ariananda, D.D.; Leus, G. Direction of arrival estimation using sparse ruler array design. In Proceedings of the Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cesme, Turkey, 17–20 June 2012; pp. 525–529. [[CrossRef](#)]
34. Park, S.; Heath, R.W. Spatial Channel Covariance Estimation for the Hybrid MIMO Architecture: A Compressive Sensing-Based Approach. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 8047–8062. [[CrossRef](#)]
35. Haghghatshoar, S.; Khalilsarai, M.B.; Caire, G. Multi-Band Covariance Interpolation with Applications in Massive MIMO. *CoRR* **2018**. [[CrossRef](#)]
36. Wichmann, B. A Note on Restricted Difference Bases. *J. Lond. Math. Soc.* **2016**, *s1-38*, 465–466. [[CrossRef](#)]
37. Huang, W.; Huang, Y.; Xu, W.; Yang, L. Beam-Blocked Channel Estimation for FDD Massive MIMO With Compressed Feedback. *IEEE Access* **2017**, *5*, 11791–11804. [[CrossRef](#)]
38. Schniter, P.; Sayeed, A. Channel estimation and precoder design for millimeter-wave communications: The sparse way. In Proceedings of the 48th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 2–5 November 2014; pp. 273–277. [[CrossRef](#)]
39. Leech, J. On the Representation of $1, 2, \dots, n$ by Differences. *J. Lond. Math. Soc.*, **1956**, *s1-31*, 160–169. [[CrossRef](#)]
40. Schmidt, R. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280. [[CrossRef](#)]
41. Lee, K.; Bresler, Y.; Junge, M. Subspace Methods for Joint Sparse Recovery. *IEEE Trans. Inf. Theory* **2012**, *58*, 3613–3641. [[CrossRef](#)]
42. Kim, J.M.; Lee, O.K.; Ye, J.C. Compressive MUSIC: Revisiting the Link Between Compressive Sensing and Array Signal Processing. *IEEE Trans. Inf. Theory* **2012**, *58*, 278–301. [[CrossRef](#)]

43. Pal, P.; Vaidyanathan, P.P. Nested Arrays: A Novel Approach to Array Processing With Enhanced Degrees of Freedom. *IEEE Trans. Signal Process.* **2010**, *58*, 4167–4181. [[CrossRef](#)]
44. Ariananda, D.D.; Leus, G. Direction of arrival estimation for more correlated sources than active sensors. *Signal Process.* **2013**, *93*, 3435–3448. [[CrossRef](#)]
45. Wax, M.; Ziskind, I. On unique localization of multiple sources by passive sensor arrays. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 996–1000. [[CrossRef](#)]
46. Richards, M. *Fundamentals of Radar Signal Processing*; Professional Engineering; McGraw-Hill: New York, NY, USA, 2005.
47. Kay, S. *Fundamentals of Statistical Signal Processing*; Prentice-Hall PTR: Englewood Cliffs, NJ, USA, 1993.
48. Haghghatshoar, S.; Caire, G. Massive MIMO Channel Subspace Estimation From Low-Dimensional Projections. *IEEE Trans. Signal Process.* **2017**, *65*, 303–318. [[CrossRef](#)]
49. Tsai, C.R.; Liu, Y.H.; Wu, A.Y. Efficient Compressive Channel Estimation for Millimeter-Wave Large-Scale Antenna Systems. *IEEE Trans. Signal Process.* **2018**, *66*, 2414–2428. [[CrossRef](#)]
50. Shen, J.; Zhang, J.; Alsusa, E.; Letaief, K.B. Compressed CSI Acquisition in FDD Massive MIMO: How Much Training is Needed? *IEEE Trans. Wirel. Commun.* **2016**, *15*, 4145–4156. [[CrossRef](#)]
51. Zhao, P.; Wang, Z.; Sun, C. Angular domain pilot design and channel estimation for FDD massive MIMO networks. In Proceedings of the International Conference on Communications (ICC), Washington, DC, USA, 2–7 July 2017; pp. 1–6. [[CrossRef](#)]
52. Rao, X.; Lau, V.K.N. Distributed Compressive CSIT Estimation and Feedback for FDD Multi-User Massive MIMO Systems. *IEEE Trans. Signal Process.* **2014**, *62*, 3261–3271. [[CrossRef](#)]
53. Cheng, X.; Sun, J.; Li, S. Channel Estimation for FDD Multi-User Massive MIMO: A Variational Bayesian Inference-Based Approach. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 7590–7602. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).