Subspace Estimation and Decomposition for Large Millimeter-Wave MIMO systems

Hadi Ghauch, *Student Member, IEEE,* Taejoon Kim, *Member, IEEE,* Mats Bengtsson, *Senior Member, IEEE,* and Mikael Skoglund, *Senior Member, IEEE*

***Abstract*—Channel estimation and precoding in hybrid analog-digital millimeter-wave (mmWave) MIMO systems is a fundamental problem that has yet to be addressed, before any of the promised gains can be harnessed. For that matter, we propose a method (based on the well-known Arnoldi iteration) exploiting channel reciprocity in TDD systems and the sparsity of the channel's eigenmodes, to estimate the right (resp. left) singular subspaces of the channel, at the BS (resp. MS). We first describe the algorithm in the context of conventional MIMO systems, and derive bounds on the estimation error in the presence of distortions at both BS and MS. We later identify obstacles that hinder the application of such an algorithm to the hybrid analog-digital architecture, and address them individually. In view of fulfilling the constraints imposed by the hybrid analog-digital architecture, we further propose an iterative algorithm for subspace decomposition, whereby the above estimated subspaces, are approximated by a cascade of analog and digital precoder / combiner. Finally, we evaluate the performance of our scheme against the perfect CSI, fully digital case (i.e., an equivalent conventional MIMO system), and conclude that similar performance can be achieved, especially at medium-to-high SNR (where the performance gap is less than 5%), however, with a drastically lower number of RF chains (~ 4 to 8 times less).**

***Keywords*—Millimeter wave MIMO systems, sparse channel estimation, hybrid architecture, hybrid precoding, subspace decomposition, Arnoldi iteration, subspace estimation, echo-based channel estimation.**

I. Introduction

With the global volume of mobile data expected to increase *by an order of magnitude between* 2013 *and* 2019, and the volume corresponding to mobile devices outweighing that of all other devices [1], mobile network operators have the monumental task of meeting this exponentially increasing demand. Given that spectrum is a scarce and precious resource, future communication systems have to exhibit unparalleled spectral efficiency. Though earlier results date back to [2], [3], communication systems in the millimeter wave (mmWave) spectrum have been receiving growing interest over the past years. mmWave communication systems have the distinct advantage of exploiting the *huge amounts of unused (and possibly unlicensed) spectrum* in those bands - around 200 times more than conventional cellular systems. Moreover, the corresponding antennae size and spacing become small enough, such that *tens-to-hundreds of antennas can be fitted on conventional hand-held devices*, thereby enabling gigabit-per-second communication.

However, the large number of RF chains required to drive the increasing number of antennas, inevitably incurs a tremendous increase in power consumption (namely by the analog-to-digital converters), as well as added hardware cost. One elegant and promising solution to remedy this inherent problem is to offload part of the precoding / processing to the *analog domain*, via analog precoding (resp.combining), i.e., a network of phase shifters to linearly process the signal at the BS (resp. MS). This so-called problem of *analog and digital co-design* for beamforming and precoding in low-frequency regime was first investigated in [4], [5]. This architecture was later studied within the context of higher frequency (mmWave) systems in [6]–[8] - under the name of *hybrid precoding / architecture* - for the precoding problem. A similar setup for the case of beamforming was considered in [9]–[11].

However, several fundamental challenges have to resolved before any of the promised gains can be harnessed, namely, estimating the (large) mmWave channel, and designing the analog / digital precoders and combiners accordingly. We underline the fact that classical training schemes developed for MIMO systems are not applicable for that particular case. Moreover, note that our proposed technique encompasses both beamforming and precoding, i.e., it does not depend on the number of streams.

After a series of approximations to the mutual information, and taking into account precoding only, [6] derived an optimality condition relating the analog and digital precoders to the optimal unconstrained precoder (i.e., the right singular vectors of the channel), by assuming *full CSI at both the BS and MS*. This assumption was later relaxed in [7] where an algorithm for estimating the dominant propagation paths was proposed, based on the previously proposed concept of *hierarchical codebooks sounding* in [10], [11]. However, the algorithm requires *a priori knowledge of the number of propagation paths* (i.e. the propagation environment), its *performance is affected by the sparsity level of the channel*, and exhibits relatively elevated complexity. Finally, it appears rather inefficient to estimate the entire channel, while only a few eigenmodes are needed for transmission: this is particularly relevant in mmWave MIMO channels, since the majority of eigenmodes have negligible power.

The approach we present here attempts to address the above limitations. The proposed algorithm is based on the well known

H. Ghauch, M. Bengtsson, M. Skoglund, and are with the School of Electrical Engineering and the ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden. E-mails: ghauch@kth.se, mats.bengtsson@ee.kth.se, skoglund@kth.se

T. Kim is with Department of Electronic Engineering, City University of Hong Kong, Kowloon Tong, Hong Kong. E-mail: taejokim@cityu.edu.hk

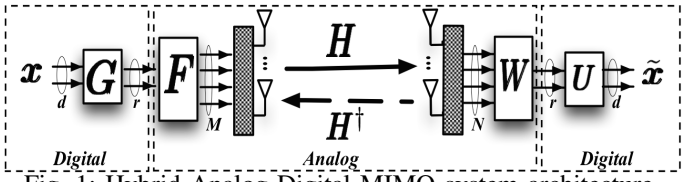

Fig. 1: Hybrid Analog-Digital MIMO system architecture

Arnoldi Iteration, exploits channel reciprocity inherent in TDD MIMO systems to gradually build an orthonormal basis for the corresponding Krylov subspace, and *directly estimates the desired left / right singular modes of the channel, rather than the entire channel*. We then propose an iterative method for subspace decomposition, to *approximate the estimated right (resp. left) singular subspace by a cascade of analog and digital precoder (resp. combiner)*, while taking into account the hardware constraints of this so-called hybrid analog-digital architecture. The subspace estimation (SE) algorithm is based on *BS-initiated echoing*, whereby the BS sends along some beamforming vector, and the MS echoes its received signal back to the BS (using amplify-and-forward), thereby enabling the BS to obtain an estimate of the effective uplink-downlink channel. We first detail the algorithm in the context of conventional MIMO, taking into account distortions in the the system (e.g., noise, or other disturbances), derive bounds on the estimation error, and highlight its desirable features. We then adapt its structure, to fit the many operational constraints dictated by the hybrid analog-digital architecture. Although the main results of the paper were earlier presented in [12], we provide in this work an in-depth look at our proposed methods, and derive several performance results.

In the following, we use bold upper-case letters to denote matrices, and bold lower-case letters denote vectors. Furthermore, for a given matrix $\boldsymbol{A}$, $[\boldsymbol{A}]_{i:j}$ denotes the matrix formed by taking columns i to j, of $\boldsymbol{A}$, tr($\boldsymbol{A}$) denotes its trace, $\|\boldsymbol{A}\|_F^2$ its Frobenius norm, $|\boldsymbol{A}|$ its determinant, $\boldsymbol{A}^\dagger$ its conjugate transpose. $[\boldsymbol{A}]_{i,j} = a_{i,j}$ denotes element (i,j) in $\boldsymbol{A}$, $\boldsymbol{a}_i$ its ith column, and $[\boldsymbol{a}]_i = a_i$ element i in vector $\boldsymbol{a}$. $[\boldsymbol{A}]_{SL}$ and $[\boldsymbol{A}]_U$ represent the matrix formed by the strictly lower and upper triangular matrix of a square matrix $\boldsymbol{A}$, respectively. $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix, diag($\boldsymbol{x}$) is a diagonal matrix with elements of $\boldsymbol{x}$ on its diagonal, $\Re(x)$ the real part of x, qr($\boldsymbol{U}$) is the QR decomposition of $\boldsymbol{U}$. Finally, we let $\{n\} \triangleq \{1, ..., n\}$, and $\mathcal{S}_{p,q} = \left\{ \boldsymbol{X} \in \mathbb{C}^{p \times q} \mid \ |\boldsymbol{X}_{ij}| = 1/\sqrt{p} \ , \ \forall (i,k) \in \{p\} \times \{q\} \right\}$.

II. System Model

A. Signal Model

Assume a single user MIMO system with M and N antennas at the BS and MS, respectively, where each is equipped with r RF chains, and sends d independent data streams (where we assume that $d \leq r \leq \min(M, N)$). The downlink (DL) received signal is given by

$$\boldsymbol{y}^{(r)} = \boldsymbol{H}\boldsymbol{F}\boldsymbol{G}\boldsymbol{x}^{(t)} + \boldsymbol{n}^{(r)} \tag{1}$$

where $\boldsymbol{H} \in \mathbb{C}^{N \times M}$ is the complex channel - assumed to be slowly block-fading, $\boldsymbol{F} \in \mathbb{C}^{M \times r}$ is the analog precoder, $\boldsymbol{G} \in \mathbb{C}^{r \times d}$ the digital precoder, $\boldsymbol{y}^{(r)}$ the N-dimensional signal at the MS antennas, $\boldsymbol{x}^{(t)}$ is the d-dimensional transmit signal with covariance matrix $E[\boldsymbol{x}^{(t)}\boldsymbol{x}^{(t)^\dagger}] = \boldsymbol{I}_d$ and $\boldsymbol{n}^{(r)}$ is the AWGN noise at the MS, with $E[\boldsymbol{n}^{(r)}\boldsymbol{n}^{(r)^\dagger}] = \sigma_{(r)}^2 \boldsymbol{I}_N$. Note that $^{(t)}$ and $^{(r)}$ subscripts/superscripts denote quantities at the BS and MS, respectively, rather than actual number. Both the analog precoder and combiner are constrained to have constant modulus elements (since the latter represent phase shifters), i.e., $\boldsymbol{F} \in \mathcal{S}_{M,r}$ and $\boldsymbol{W} \in \mathcal{S}_{N,r}$ (also referred to as the *constant-modulus* or *constant-envelope* constraint). Moreover, the gain that transmit symbol $x_i^{(t)}$ experiences is assumed to be bounded, i.e., $\|\boldsymbol{F}\boldsymbol{g}_i\|_2^2 \leq \rho^2 \ , \ \forall \ i \in \{d\}$ (where we assume that $\rho = 1$ w.l.o.g.) resulting in the following constraint, $\|\boldsymbol{F}\boldsymbol{G}\|_F^2 \leq d$. With that in mind, the received signal after filtering in the DL is given as,

$$\tilde{\boldsymbol{x}} = \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{y}^{(r)} = \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{H}\boldsymbol{F}\boldsymbol{G}\boldsymbol{x}^{(t)} + \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{n}^{(r)} \tag{2}$$

where $\boldsymbol{W} \in \mathbb{C}^{N \times r}$ and $\boldsymbol{U} \in \mathbb{C}^{r \times d}$ are the analog and digital combiners, respectively[1]. We also assume a TDD system, where channel reciprocity holds. Finally, we denote the SVD of $\boldsymbol{H}$ as,

$$\boldsymbol{H} = [\boldsymbol{\Phi}_1, \quad \boldsymbol{\Phi}_2] \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Gamma}_1^\dagger \\ \boldsymbol{\Gamma}_2^\dagger \end{bmatrix} = \boldsymbol{\Phi}_1 \boldsymbol{\Sigma}_1 \boldsymbol{\Gamma}_1^\dagger + \boldsymbol{\Phi}_2 \boldsymbol{\Sigma}_2 \boldsymbol{\Gamma}_2^\dagger \tag{3}$$

where $\boldsymbol{\Gamma}_1 \in \mathbb{C}^{M \times d}$ and $\boldsymbol{\Phi}_1 \in \mathbb{C}^{N \times d}$ are unitary, and $\boldsymbol{\Sigma}_1 = \text{diag}(\sigma_1, ..., \sigma_d)$ is diagonal with the d largest singular values of $\boldsymbol{H}$ (in decreasing order).

B. Motivation

We use the following expression as a performance metric,

$$R = \log_2 \left| \boldsymbol{I}_d + \boldsymbol{H}_e \boldsymbol{H}_e^\dagger (\sigma_{(r)}^2 \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{W}\boldsymbol{U})^{-1} \right| \tag{4}$$

where $\boldsymbol{H}_e = \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{H}\boldsymbol{F}\boldsymbol{G}$, $\frac{1}{\sigma_{(r)}^2} \triangleq$ SNR, and we assume uniform power allocation (no waterfilling). As we will discuss below, the value of the expression in (4) is related to achievable rates over the considered hybrid analog-digital MIMO link; in particular R becomes an *achievable rate* in the scenario that both the BS and MS are provided perfect knowledge of $\boldsymbol{H}$.

Using Hadamard's inequality, it can be easily verified that the optimal $\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{W}, \boldsymbol{U}$ that maximize R in (4), are the ones that diagonalize the effective channel $\boldsymbol{H}_e$, i.e.,

$$R^\star \triangleq \max \ R(\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{W}, \boldsymbol{U}) = \log_2 \left| \boldsymbol{I}_d + \boldsymbol{\Sigma}_1^2 \right| \tag{5}$$

Following the above discussion on the achievability of R, $R^\star$ is the *maximum achievable rate* over the precoders and combiners, when $\boldsymbol{H}$ is known to both BS and MS. Similarly to [13] we formulate our problem in terms of the distortion between $R^\star$ and R (albeit the latter has to be *instantaneous*). Moreover, it can be verified that the optimal $\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{W}, \boldsymbol{U}$ that achieve $R^\star$, are such that $\boldsymbol{F}\boldsymbol{G} \subseteq \text{span}(\boldsymbol{\Gamma}_1)$ and $\boldsymbol{W}\boldsymbol{U} \subseteq \text{span}(\boldsymbol{\Phi}_1)$, i.e. there is no loss in generality by assuming the latter.

[1] Similarly, exploiting channel reciprocity, the uplink received signal is given by $\boldsymbol{y}^{(t)} = \boldsymbol{H}^\dagger \boldsymbol{W}\boldsymbol{U}\boldsymbol{x}^{(r)} + \boldsymbol{n}^{(t)}$ where $\boldsymbol{y}^{(t)}$ is the M-dimensional signal at the BS and $\boldsymbol{n}^{(t)}$ is the AWGN noise at the BS, such that $E[\boldsymbol{n}^{(t)}\boldsymbol{n}^{(t)^\dagger}] = \sigma_{(t)}^2 \boldsymbol{I}_N$

Proposition 1. *For the distortion $D \triangleq R^\star - R$, the following holds,*

$$D \leq \sum_{i=1}^{d}(1+\sigma_i^2) + k_o\epsilon^2\left(\|\boldsymbol{\Gamma}_1 - \boldsymbol{FG}\|_F^2 + \|\boldsymbol{\Phi}_1 - \boldsymbol{WU}\|_F^2\right)/2$$

Proof: Refer to Appendix A ∎

This result does imply that by finding $\boldsymbol{FG}$ (resp. $\boldsymbol{WU}$) that "best" approximate $\boldsymbol{\Gamma}_1$ (resp. $\boldsymbol{\Phi}_1$), we would be minimizing a bound on the distortion. However, since we assume that no channel information is available at neither the BS, nor the MS, our aim is *firstly to obtain an estimate of the subspaces in question*, i.e. $\tilde{\boldsymbol{\Phi}}_1 \approx \boldsymbol{\Phi}_1$ at the MS, and $\tilde{\boldsymbol{\Gamma}}_1 \approx \boldsymbol{\Gamma}_1$ at the BS. We then propose methods that optimize the precoders and combiners to *accurately approximate the estimated subspaces*, by providing means to solve problems such as $\|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{FG}\|_F^2$ and $\|\tilde{\boldsymbol{\Phi}}_1 - \boldsymbol{WU}\|_F^2$ (while taking into consideration the constraints inherent to the hybrid analog-digital architecture).

III. Eigenvalue Algorithms and Subspace Estimation

A. Subsapce Estimation vs. Channel Estimation

The aim of subspace estimation (SE) methods in MIMO systems is to estimate a predetermined *low-dimensional subspace of the channel*, required for transmission. We illustrate this in the context of conventional MIMO systems, i.e., where precoders/combiners are fully digital. For the sake of exposition, we start with a simple toy example, where noiseless single-stream transmission is assumed (and ignoring any physical constraints). The BS selects a random unit-norm beamforming vector, $\boldsymbol{p}_1$, and then sends $\boldsymbol{p}_1 x^{(t)}$, where $x^{(t)} = 1$. The received signal, $\boldsymbol{q}_1 = \boldsymbol{H}\boldsymbol{p}_1$, is echoed back to the BS (in effect, this implies that the signal is complex conjugated before being sent), in an Amplify-and-Forward (A-F) like fashion.[2] Then, exploiting channel reciprocity, the received signal at the BS is first normalized, i.e., $\boldsymbol{p}_2 = \boldsymbol{H}^\dagger\boldsymbol{q}_1/\|\boldsymbol{H}^\dagger\boldsymbol{q}_1\|_2 = \boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{p}_1/\|\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{p}_1\|_2$, and then echoed back to the MS. This simple procedure is done iteratively, and the resulting sequences $\{\boldsymbol{p}_l\}$ at the BS, and $\{\boldsymbol{q}_l\}$ at the MS, are defined as follows,

$$\boldsymbol{p}_{l+1} = \boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{p}_l/\|\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{p}_l\|_2; \quad \boldsymbol{q}_{l+1} = \boldsymbol{H}\boldsymbol{p}_l \tag{6}$$

It was noted in [14] that using the Power Method (PM), one can show that as $l \to \infty$, $\boldsymbol{p}_l \to \boldsymbol{\gamma}_1$ and $\boldsymbol{q}_l \to \sigma_1\boldsymbol{\phi}_1$, implying that this seemingly simple "ad-hoc" procedure will converge to the *maximum eigenmode transmission*. In the latter work, the authors also generalized the latter method to multistream transmission, i.e., by estimating $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Phi}_1$, using the Orthogonal / Subspace Iteration (which was dubbed Two-way QR (TQR) in [14], [15]).

We note that SE schemes such as the ones described above, offer the following distinct advantage over classical *pilot-based channel estimation*: in spite of the large number of transmit and receive antennas, SE methods can estimate the desired left / right singular subspaces with a relatively low communication overhead, when the latter have small dimension (relative to the channel dimensions). Consequently, subspace estimation is much more efficient than channel estimation, especially in large low-rank MIMO systems such as mmWave channels (since the latter estimates the desired low-dimensional subspace instead of the whole channel). For the reason above, our proposed algorithm falls under the umbrella of SE methods. We first describe this algorithm in the context of "classical" MIMO systems, and later adapt it to the hybrid analog-digital architecture.

[2]This mechanism for MIMO subspace estimation, where the MS echoes back the transmitted signal using A-F, was first reported in [14].

Set m ($m \leq M$); $\boldsymbol{q}_1$ = random unit-norm ; $\boldsymbol{Q} = [\boldsymbol{q}_1]$
for $l = 1, 2, ..., m$ **do**
1.a $\boldsymbol{p}_l = \boldsymbol{A}\boldsymbol{q}_l$
1.b $t_{m,l} = \boldsymbol{q}_m^\dagger\boldsymbol{p}_l,\ m = 1, \ldots, l$
2. $\boldsymbol{r}_l = \boldsymbol{p}_l - \sum_{m=1}^{l} t_{m,l}\boldsymbol{q}_m$
3. $t_{l+1,l} = \|\boldsymbol{r}_l\|_2$; **if** $(t_{l+1,l} = 0)$ **stop**
4. $\boldsymbol{Q} = [\boldsymbol{Q},\ \boldsymbol{q}_{l+1} = \boldsymbol{r}_l/t_{l+1,l}]$
end for

TABLE I: Arnoldi Procedure

B. Arnoldi Iteration for Subspace Estimation

Despite the fact that Krylov subspace methods (such as the Arnoldi and Lanczos Iterations for symmetric matrices) are among the most common methods for eigenvalue problems [16], their use in the area of channel / subspace estimation is limited to equalization for doubly selective OFDM channels [17], and channel estimation in CDMA systems [18]. Algorithms falling into that category iteratively build a *basis for the Krylov subspace*, $\mathcal{K}^m = span\{\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}, ..., \boldsymbol{A}^{m-1}\boldsymbol{x}\}$, one vector at a time. We use one of many variants of the so-called *Arnoldi Iteration / Procedure*, and a simplified version of the latter is shown in Table I (as presented in [19]). The algorithm returns $\boldsymbol{Q}_m = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m] \in \mathbb{C}^{M\times m}$ and an upper Hessenberg matrix $\boldsymbol{T}_m \in \mathbb{C}^{m\times m}$, such that

$$\boldsymbol{Q}_m^\dagger\boldsymbol{A}\boldsymbol{Q}_m = \boldsymbol{T}_m,\ \boldsymbol{Q}_m^\dagger\boldsymbol{Q}_m = \boldsymbol{I}_d.$$

It can be shown that the algorithm iteratively builds $\boldsymbol{Q}_m$, an orthonormal basis for $\mathcal{K}^m$ (when roundoff errors are neglected), and that $\boldsymbol{Q}_m^\dagger\boldsymbol{A}\boldsymbol{Q}_m = \boldsymbol{T}_m$ implying that the eigenvalues of $\boldsymbol{T}_m$ are the same as the eigenvalues of $\boldsymbol{A}$. The main idea behind this process is to find the desired eigenpairs of $\boldsymbol{A}$, by finding the eigenpairs of $\boldsymbol{T}_m$. Our goal in this section is to first apply the above algorithm to estimate the d largest eigenvectors of $\boldsymbol{A} = \boldsymbol{H}^\dagger\boldsymbol{H}$ at the BS (which are exactly $\boldsymbol{\Gamma}_1$), by implementing a *distributed version of the Arnoldi process*, that exploits the channel reciprocity inherent to TDD systems. Moreover, we extend the original formulation of the algorithm to incorporate a *distortion variable* (representing noise, or other distortions, as will be done later).

It becomes clear at this stage, that the BS requires knowledge of the sequence $\{\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{q}_l\}_{l=1}^m$, needed for the matrix-vector product in step 1 (Table I): the latter can be accomplished by obtaining an estimate $\boldsymbol{p}_l$, of $\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{q}_l,\ l \in \{m\}$. Without any explicit CSI at neither the BS nor the MS, we exploit the reciprocity of the medium to obtain such an estimate, via *BS-initiated echoing*: the BS sends $\boldsymbol{q}_l$ over the DL channel, the MS echoes back the received signal in an

procedure $\tilde{\boldsymbol{\Gamma}}_1 = \text{SE-ARN}\ (\boldsymbol{H},\ d)$
Set m ($m \leq M$); Random unit-norm $\boldsymbol{q}$; $\boldsymbol{Q} = [\boldsymbol{q}_1]$
for $l = 1, 2, ..., m$ **do**
// BS-initiated echoing: estimate $\boldsymbol{H}^\dagger \boldsymbol{H} \boldsymbol{q}_l$
1.a $\boldsymbol{s}_l = \boldsymbol{H}\boldsymbol{q}_l + \boldsymbol{w}_l^{(r)}$
1.b $\boldsymbol{p}_l = \boldsymbol{H}^\dagger \boldsymbol{s}_l + \boldsymbol{w}_l^{(t)}$
// Gram-Schmidt orthogonalization
2.a $t_{m,l} = \boldsymbol{q}_m^\dagger \boldsymbol{p}_l \quad ,\forall\ m = 1, \ldots, l$
2.b $\boldsymbol{r}_l = \boldsymbol{p}_l - \sum_{m=1}^{l} \boldsymbol{q}_m t_{m,l}$
2.c $t_{l+1,l} = \|\boldsymbol{r}_l\|_2$
// Update $\boldsymbol{Q}$
3.a $\boldsymbol{Q} = [\boldsymbol{Q},\ \boldsymbol{q}_{l+1} = \boldsymbol{r}_l / t_{l+1,l}]$
end for
// Compute $\tilde{\boldsymbol{\Gamma}}_1$
$\boldsymbol{T}_m = \tilde{\boldsymbol{\Theta}} \tilde{\boldsymbol{\Lambda}} \tilde{\boldsymbol{\Theta}}^{-1}$
$\tilde{\boldsymbol{\Gamma}}_1 = \boldsymbol{Q}_m \tilde{\boldsymbol{\Theta}}_{1:d}$
$\tilde{\boldsymbol{\Gamma}}_1 = \text{qr}(\tilde{\boldsymbol{\Gamma}}_1)$
end procedure

TABLE II: Subspace Estimation using Arnoldi Iteration (SE-ARN)

A-F like fashion, over the uplink (UL) channel (following the process proposed in [20], and detailed in Sect. III-A), i.e.,

$$\begin{aligned} DL:&\quad \boldsymbol{s}_l = \boldsymbol{H}\boldsymbol{q}_l + \boldsymbol{w}_l^{(r)} \\ UL:&\quad \boldsymbol{p}_l = \boldsymbol{H}^\dagger \boldsymbol{s}_l + \boldsymbol{w}_l^{(t)} = \boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l + \boldsymbol{H}^\dagger \boldsymbol{w}_l^{(r)} + \boldsymbol{w}_l^{(t)} \\ &\quad\ = \boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l + \tilde{\boldsymbol{w}}_l \end{aligned} \tag{7}$$

where $\boldsymbol{s}_l$ is the received signal in the DL, $\boldsymbol{w}_l^{(t)}$ and $\boldsymbol{w}_l^{(r)}$ are distortions at the BS and MS, respectively (representing noise for example).

After the echoing phase, the BS has an estimate, $\boldsymbol{p}_l$, of $\boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l$, as seen from (7). The remainder of the algorithm follows the conventional Arnoldi Iteration, and is shown in the Subspace Estimation using Arnoldi (SE-ARN) procedure (Table II). In addition to $\boldsymbol{T}_m$ at the output of the algorithm, we define the matrices, $\tilde{\boldsymbol{T}}_m$, $\tilde{\boldsymbol{W}}_m$ and $\tilde{\boldsymbol{E}}_m$, as follows,

$$[\tilde{\boldsymbol{T}}_m]_{i,l} = \begin{cases} \boldsymbol{q}_i^\dagger \boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l, & \text{if } l \leq m,\ \forall i \leq l \\ \|\boldsymbol{r}_l\|_2, & \text{if } l < m,\ i = l+1 \\ 0, & \text{otherwise} \end{cases}$$

$$\tilde{\boldsymbol{W}}_m = [\tilde{\boldsymbol{w}}_1, ..., \tilde{\boldsymbol{w}}_m],\ \tilde{\boldsymbol{E}}_m = [\boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m]_{SL} \tag{8}$$

where $\tilde{\boldsymbol{T}}_m$ is the upper Hessenberg matrix obtained by the conventional Arnoldi process, i.e., when the distortion $\tilde{\boldsymbol{w}}_l$ is ignored. At the output of the SE-ARN procedure, the desired eigenpairs of $\boldsymbol{H}^\dagger \boldsymbol{H}$ are approximated by that of $\boldsymbol{T}_m$ as follows. Let $\boldsymbol{T}_m = \tilde{\boldsymbol{\Theta}} \tilde{\boldsymbol{\Lambda}} \tilde{\boldsymbol{\Theta}}^{-1}$ be eigenvalue decomposition of $\boldsymbol{T}_m$, where $\boldsymbol{\Theta}$ is the (possibly non-orthonormal) set of eigenvectors. Then, it can easily be shown that $\tilde{\boldsymbol{\Gamma}}_1 = \boldsymbol{Q}_m [\tilde{\boldsymbol{\Theta}}]_{1:d}$ are the Ritz eigenvectors of $\boldsymbol{H}^\dagger \boldsymbol{H}$ where $[\tilde{\boldsymbol{\Theta}}]_{1:d}$ has as columns the eigenvectors of $\boldsymbol{T}_m$ with the d largest eigenvalues (in magnitude). Note that the latter procedure results in the BS obtaining $\tilde{\boldsymbol{\Gamma}}_1$, and consequently $\tilde{\boldsymbol{\Sigma}}_1$[3], using the so-called BS-initiated echoing. This same procedure can be applied using MS-initiated echoing, to estimate $\tilde{\boldsymbol{\Phi}}_1$ (i.e., the eigenvectors of $\boldsymbol{H}\boldsymbol{H}^\dagger$), at the MS.

C. Perturbation Analysis

In what follows, we extend some of the known properties of the conventional Arnoldi iteration, to account for the estimation error, emanating from the distortion variable.

Lemma 1. *For the output of the Arnoldi process the following holds,*
$(P1)$:

$$\boldsymbol{Q}_m^\dagger \boldsymbol{A} \boldsymbol{Q}_m = \tilde{\boldsymbol{T}}_m - \tilde{\boldsymbol{E}}_m \triangleq \boldsymbol{C}_m, \tag{9}$$

where $\boldsymbol{C}_m = \boldsymbol{S}_m \boldsymbol{\Lambda}_m \boldsymbol{S}_m^{-1}$ *is such that* $[\boldsymbol{\Lambda}]_{i,i} \geq 0$ *and* $\boldsymbol{S}_m^{-1} = \boldsymbol{S}_m^\dagger$

$(P2)$: *Let* $(\lambda_i^{(m)}, \boldsymbol{s}_i^{(m)})$ *be any eigenpair of* $\boldsymbol{C}_m$. *Then* $(\lambda_i^{(m)},\ \boldsymbol{\theta}_i^{(m)} \triangleq \boldsymbol{Q}_m \boldsymbol{s}_i^{(m)})$ *is an approximate Ritz eigenpair for* $\boldsymbol{A}$. *Furthermore, the approximation error is such that,*

$$\|\boldsymbol{A}\boldsymbol{\theta}_i^{(m)} - \lambda_i^{(m)} \boldsymbol{\theta}_i^{(m)}\|_2^2 \leq c_m^{(i)} + \|\boldsymbol{I}_M - \boldsymbol{Q}_m \boldsymbol{Q}_m^\dagger\|_F^2 \|\tilde{\boldsymbol{W}}_m\|_F^2, \tag{10}$$

where $c_m^{(i)} = ([\tilde{\boldsymbol{T}}_m]_{m+1,m} | [\boldsymbol{s}_i^{(m)}]_m|)^2$.

$(P3)$: *As* $m \to M$, $\|\boldsymbol{A}\boldsymbol{\theta}_i^{(m)} - \lambda_i^{(m)} \boldsymbol{\theta}_i^{(m)}\|_2^2 \to 0$, *implying that the eigenpairs of* $\boldsymbol{C}_m$ *perfectly approximate the eigenpairs of* $\boldsymbol{A}$*(up to round-off errors).*

Proof: The proof is shown in Appendix B. ■

We underline the fact that if the distortion variable $\tilde{\boldsymbol{W}}_m$ is zero, the above derivations reduce to the well-known results on the Arnoldi process [19, Sect. 6.2]. Lemma 1 establishes the fact that each eigenpair $(\lambda_i^{(m)}, \boldsymbol{s}_i^{(m)})$ of $\boldsymbol{C}_m$[4], is associated with one eigenpair $(\lambda_i^{(m)}, \boldsymbol{\theta}_i^{(m)})$ of $\boldsymbol{A}$.[5]

Thus, one might be tempted to conclude at this point, that by computing the eigenpairs of $\boldsymbol{C}_m$, one can *perfectly estimate* the eigenpairs of $\boldsymbol{A}$, despite the presence of the distortion variable $\tilde{\boldsymbol{W}}_m$. However, the fact remains that $\boldsymbol{C}_m \triangleq \tilde{\boldsymbol{T}}_m - \tilde{\boldsymbol{E}}_m$ *cannot be computed*, mainly because $\tilde{\boldsymbol{E}}_m$ is not known to the BS. As a result, $\boldsymbol{T}_m$ *at the output of the Arnoldi process will be used instead to approximate the eigenpairs of* $\boldsymbol{A}$. Now that we established that the eigenpairs of $\boldsymbol{C}_m$ approximate that of $\boldsymbol{A}$, the natural question is *how close are the eigenpairs of* $\boldsymbol{T}_m$*, to that of* $\boldsymbol{C}_m$. Noting that $\boldsymbol{T}_m = \boldsymbol{C}_m + \boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m$, where $\boldsymbol{C}_m$ is the matrix in question, and $\boldsymbol{P}_m \triangleq \boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m$ is the perturbation, we apply the Bauer-Fike Theorem [21, Th. 7.2.2] to bound the difference in eigenvalues.

Lemma 2. *Every eigenvalue* $\tilde{\lambda}$ *of* $\boldsymbol{T}_m = \boldsymbol{C}_m + \boldsymbol{P}_m$ *satisfies*

$$|\tilde{\lambda} - \lambda| \leq \sqrt{m}\ \|\tilde{\boldsymbol{W}}_m\|_F,$$

where λ *is an eigenvalue of* $\boldsymbol{C}_m$.

[3]Once $\tilde{\boldsymbol{\Gamma}}_1$ is obtained, then estimating the eigenvalues of $\boldsymbol{H}^\dagger \boldsymbol{H}$, i.e., $\tilde{\boldsymbol{\Sigma}}_1$, needed for waterfilling power allocation, comes for free.

[4]Note that one can show that $\boldsymbol{C}_m$ is positive semidefinite, i.e., $\boldsymbol{C}_m = \boldsymbol{S}_m \boldsymbol{\Lambda}_m \boldsymbol{S}_m^\dagger$, where $[\boldsymbol{\Lambda}]_{i,i} \geq 0$, and $\boldsymbol{S}_m$ unitary.

[5]Though $(P3)$ in Lemma 1 implies that the error in approximating the eigenpairs of $\boldsymbol{A}$ with those of $\boldsymbol{C}_m$ vanishes as $m \to M$, our simulations will later show that very good approximations can be obtained, even for $m \ll M$.

Proof: Refer to Appendix C ∎

Summarizing thus far, Lemma 1 showed that the eigenpairs of $\boldsymbol{A}$ can be approximated by the eigenvalues of $\boldsymbol{C}_m$, with arbitrarily small error. However, since the latter is not available, we approximate the eigenpairs of $\boldsymbol{C}_m$ (and consequently of $\boldsymbol{A}$) by that of $\boldsymbol{T}_m$, the upper Hessenberg matrix at the output of the Arnoldi process. Finally, Lemma 2 established the fact that this approximation error is upper bounded by the magnitude of the perturbation itself, for the eigenvalues

IV. Hybrid Analog-Digital Precoding for mmWaveMIMO systems

In this section we turn our attention to applying the above framework for subspace estimation and precoding, to the hybrid analog-digital architecture. As this section will gradually reveal, several obstacles have to be overcome for that matter. We start by presenting some preliminaries that will be used throughout this section.

A. Preliminaries: Subspace Decomposition

We will limit our discussion to the digital and analog precoder, keeping in mind that the same applies to the digital and analog combiner. In conventional MIMO systems, the estimates of the right and left singular subspace, $\tilde{\boldsymbol{\Gamma}}_1$ and $\tilde{\boldsymbol{\Phi}}_1$, obtained using SE-ARN, can directly be used to diagonalize the channel. However, the hybrid analog-digital architecture entails a cascade of analog and digital precoder. Recalling the result of Proposition 1, finding $\boldsymbol{F}$, $\boldsymbol{G}$ such that $\|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{F}\boldsymbol{G}\|_F^2$ is minimized is equivalent to minimizing a bound on the distortion D. Equivalently, $\tilde{\boldsymbol{\Gamma}}_1$ has to be decomposed into $\boldsymbol{F}\boldsymbol{G}$ - hence the term Subspace Decomposition (SD), as follows,

$$\begin{cases} \min\limits_{\boldsymbol{F},\,\boldsymbol{G}} \;\; h_0(\boldsymbol{F},\boldsymbol{G}) = \|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{F}\boldsymbol{G}\|_F^2 \\ \text{s. t.} \;\; h_1(\boldsymbol{F},\boldsymbol{G}) = \|\boldsymbol{F}\boldsymbol{G}\|_F^2 \leq d \\ \qquad \boldsymbol{F} \in \mathcal{S}_{M,d} \end{cases} \tag{11}$$

We underline the fact that the authors in [6] arrived to the same formulation as (11) after a series of approximations to the mutual information. To a certain extent, (11) is reminiscent of formulations arising from areas such as blind source separation, (sparse) dictionary learning, and vector quantization [22], [23]. Though there is a battery of algorithms and techniques that have been developed to tackle such problems, the additional hardware constraint on $\boldsymbol{F}$ makes the use of such tools not possible. As a result, we will resort to developing our own algorithm. In spite of the non-convex and non-separable nature of the above quadratically-constrained quadratic program, we propose an iterative method that attempts to solve for an approximate solution.

1) Block Coordinate Descent for Subspace Decomposition: In this part, we further assume that only d of the r available RF chains are used, i.e., $\boldsymbol{F} \in \mathbb{C}^{M\times d}$ and $\boldsymbol{G} \in \mathbb{C}^{d\times d}$ (the reason for that will become clear later in this section). The coupled nature of the objective and constraints in (11) suggests a Block Coordinate Descent (BCD) approach. The main challenges arise from the coupled nature of the variables in the constraint (since the latter makes convergence claims of BCD, not possible [24]), and from the hardware constraint on $\boldsymbol{F}$. We will show that a BCD approach implicitly enforces the power constraint in (11), and consequently the latter can be dropped without changing the problem.

Our approach consists in relaxing the hardware constraint on $\boldsymbol{F}$, and then applying a Block Coordinate Descent (BCD) approach to alternately optimize $\boldsymbol{F}$ and $\boldsymbol{G}$ (while projecting each of the obtained solutions for $\boldsymbol{F}$ on $\mathcal{S}$). For that matter, we first define the *Euclidean projection* on the set $\mathcal{S}$ in the following proposition.

Proposition 2. *Let $\boldsymbol{X} \in \mathbb{C}^{M\times d}$ be defined as $[\boldsymbol{X}]_{i,k} = |x_{i,k}|\, e^{j\phi_{i,k}},\ \forall(i,k)$, and*

$$\boldsymbol{Y} = \Pi_{\mathcal{S}}[\boldsymbol{X}] \triangleq \operatorname*{argmin}_{\boldsymbol{U}\in\mathcal{S}_{M,d}} \|\boldsymbol{U} - \boldsymbol{X}\|_F^2$$

denote its (unique) Euclidean projection on the set $\mathcal{S}_{M,d}$. Then $[\boldsymbol{Y}]_{i,k} = (1/\sqrt{M})\, e^{j\phi_{i,k}}, \forall(i,k)$.

Proof: Refer to Appendix E ∎

The latter result implies that given an arbitrary $\boldsymbol{F}$, finding the closest point to $\boldsymbol{F}$, lying in $\mathcal{S}_{M,d}$ simply reduces to *setting the magnitude of each element in $\boldsymbol{F}$, to $1/\sqrt{M}$.*

Neglecting the constraint on $\boldsymbol{F}$ in (11), one can indeed show that for fixed $\boldsymbol{G}$ (resp. $\boldsymbol{F}$), the resulting subproblem is convex in $\boldsymbol{F}$ (resp. $\boldsymbol{G}$). With this in mind, our aim is to produce a *sequence of updates, $\{\boldsymbol{F}_k, \boldsymbol{G}_k\}_k$ such that the sequence $\{h_0(\boldsymbol{F}_k, \boldsymbol{G}_k)\}_k$ is non-increasing* (keeping in mind that monotonicity cannot be shown due to the coupling in the power constraint). Thus, given $\boldsymbol{G}_k$, each of the updates, $\boldsymbol{F}_{k+1}$ and $\boldsymbol{G}_{k+1}$, are defined as as follows,

$$(J1) \quad \boldsymbol{F}_{k+1} \triangleq \min_{\boldsymbol{F}} \;\; h_0(\boldsymbol{F}) = \|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{F}\boldsymbol{G}_k\|_F^2$$

$$(J2) \quad \boldsymbol{G}_{k+1} \triangleq \min_{\boldsymbol{G}} \;\; h_0(\boldsymbol{G}) = \|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{F}_{k+1}\boldsymbol{G}\|_F^2$$

Both $(J1)$ and $(J2)$ are instances of a non-homogeneous (unconstrained) convex QCQP that can easily be solved (globally) by finding stationary points of their respective cost functions, to yield,

$$\boldsymbol{F}_{k+1} = \tilde{\boldsymbol{\Gamma}}_1 \boldsymbol{G}_k^\dagger (\boldsymbol{G}_k \boldsymbol{G}_k^\dagger)^{-1} \tag{12}$$

$$\boldsymbol{G}_{k+1} = (\boldsymbol{F}_{k+1}^\dagger \boldsymbol{F}_{k+1})^{-1} \boldsymbol{F}_{k+1}^\dagger \tilde{\boldsymbol{\Gamma}}_1 \tag{13}$$

We note that our earlier assumption that only d of the RF chains are used here (i.e. $\boldsymbol{G}$ is square), guarantees that $(\boldsymbol{G}_l \boldsymbol{G}_l^\dagger)$ in (13) is invertible: in fact, our numerical results show that the incurred performance loss is quite negligible. Moreover, note that the solution in (12) does not necessarily satisfy the hardware constraint on $\boldsymbol{F}$. Thus, the result of Proposition 2 can be used to compute the projection of $\boldsymbol{F}$ on $\mathcal{S}_{M,d}$. To prove our earlier observation that the optimal updates $\boldsymbol{F}_{k+1}$ and $\boldsymbol{G}_{k+1}$ satisfy the power constraint in (11), we plug (13) into the following (dropping all subscripts for simplicity),

$$\begin{aligned} \|\boldsymbol{F}\boldsymbol{G}\|_F^2 &= \operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_1^\dagger \boldsymbol{F} \underbrace{(\boldsymbol{F}^\dagger\boldsymbol{F})^{-1}\boldsymbol{F}^\dagger\boldsymbol{F}}_{=\boldsymbol{I}_d} (\boldsymbol{F}^\dagger\boldsymbol{F})^{-1}\boldsymbol{F}^\dagger \tilde{\boldsymbol{\Gamma}}_1 \right) \\ &\leq \operatorname{tr}\left((\boldsymbol{F}^\dagger\boldsymbol{F})^{-1}\boldsymbol{F}^\dagger\boldsymbol{F}\right) \operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_1 \tilde{\boldsymbol{\Gamma}}_1^\dagger\right) = d \end{aligned} \tag{14}$$

procedure $[\boldsymbol{F}, \boldsymbol{G}] = \text{BCD-SD}\ (\tilde{\boldsymbol{\Gamma}}_1)$
  Start with arbitrary $\boldsymbol{F}_0$
  for $k = 0, 1, 2, \ldots$ **do**
    $\boldsymbol{G}_{k+1} \leftarrow \boldsymbol{F}_k^{\dagger} \boldsymbol{F}_k^{-1} \boldsymbol{F}_k^{\dagger} \tilde{\boldsymbol{\Gamma}}_1$
    $\boldsymbol{F}_{k+1} \leftarrow \Pi_{\mathcal{S}}[\tilde{\boldsymbol{\Gamma}}_1 \boldsymbol{G}_{k+1}^{\dagger} (\boldsymbol{G}_{k+1} \boldsymbol{G}_{k+1}^{\dagger})^{-1}]$
  end for
end procedure

TABLE III: Block Coordiate Descent for Subspace Decomposition (BCD-SD)

where we assumed that $\|\tilde{\boldsymbol{\Gamma}}_1\|_F^2 = 1$ w.l.o.g. The above shows that if BCD is used, then the power constraint in (11) is always enforced. The corresponding method is termed Block Coordinate Descent for Subspace Decomposition (BCD-SD), and is shown in Table III. We underline the fact that due to the projection step, one cannot show that the sequence $\{h_o(\boldsymbol{F}_k, \boldsymbol{G}_k)\}_k$ is non-increasing.

Remark 1. It can be easily verified that the optimal $\boldsymbol{F}^{\star}, \boldsymbol{G}^{\star}$ that maximize the R in (4) are such that $\|\boldsymbol{F}^{\star}\boldsymbol{G}^{\star}\| = d$. Though the optimal solution to (11) is not invariant to scaling, as far as the performance metric in (4) is concerned, there in no loss in optimality in scaling the solution given by BCD-SD, to fulfill the power constraint with equality.

2) One-dimensional case: Note that echoing - the mechanism at the heart of our proposed approach, relies on the BS being able to send any vector $\boldsymbol{q}_l$, to be echoed back by the MS (Table II). For the hybrid analog-digital architecture, this translates into the BS being able to (accurately) approximate $\boldsymbol{q}_l$ by $\boldsymbol{f}_l g_l$, where $\boldsymbol{f}_l$ is a vector, g_l is a scalar. As a result, subspace decomposition for the one-dimensional case is of great interest here. When $d = 1$, (11) reduces to the problem below,

Lemma 3. *Consider single dimension SD problem,*

$$\begin{cases} \min\limits_{\boldsymbol{f},\ g} \ h_o(\boldsymbol{f},\ g) = \|\boldsymbol{f}\|_2^2\ g^2 - 2g\Re(\boldsymbol{f}^{\dagger}\tilde{\boldsymbol{\gamma}}_1) \\ \text{s. t. } [\boldsymbol{f}]_i = 1/\sqrt{M}\ e^{j\phi_i}, \forall i \end{cases} \tag{15}$$

where $g \in \mathbb{R}_+$ and $[\tilde{\boldsymbol{\gamma}}_1]_i = r_i e^{j\theta_i}$. Then the problem admits a globally optimum solution given by, $[\boldsymbol{f}^{\star}]_i = 1/\sqrt{M}\ e^{j\theta_i}, \forall i$ and $g^{\star} = \|\tilde{\boldsymbol{\gamma}}_1\|_1/\sqrt{M}$

Proof: Refer to Appendix D ∎

Similarly to (14), it can be verified that a power constraint is indeed implicitly verified. Moreover, the approximation error $\boldsymbol{e} \triangleq \tilde{\boldsymbol{\gamma}}_1 - \boldsymbol{f}g$ is such that,

$$[\boldsymbol{e}]_i = |r_i - \|\tilde{\boldsymbol{\gamma}}_1\|_1/M| e^{j\theta_i},\ \forall i \in \{M\}. \tag{16}$$

We note that when considering the effective beamformer, i.e., $\boldsymbol{f}g$, the solution given by Lemma 3 is to some extent reminiscent of equal gain transmission in [25], [26], in terms of the optimal phases.

Note that the decomposition can be written in a simple form. Given a vector $\tilde{\boldsymbol{\gamma}}_1$, its globally optimal decomposition (from the perspective of (11)) is given as,

$$\tilde{\boldsymbol{\gamma}}_1 \approx g_1^{\star}\boldsymbol{f}_1^{\star} \triangleq (\|\tilde{\boldsymbol{\gamma}}_1\|_1/\sqrt{M})\ \Pi_{\mathcal{S}}[\tilde{\boldsymbol{\gamma}}_1].$$

This can be generalized to obtain an alternate method to BCD-SD, by decomposing $\tilde{\boldsymbol{\Gamma}}_1$, in a *column-wise* fashion,

$$\tilde{\boldsymbol{\Gamma}}_1 = [\tilde{\boldsymbol{\gamma}}_1, \cdots, \tilde{\boldsymbol{\gamma}}_d] \approx [g_1^{\star}\boldsymbol{f}_1^{\star}, \cdots, g_d^{\star}\boldsymbol{f}_d^{\star}]$$
$$\triangleq (1/\sqrt{M})\ [\Pi_{\mathcal{S}}[\tilde{\boldsymbol{\gamma}}_1], \cdots, \Pi_{\mathcal{S}}[\tilde{\boldsymbol{\gamma}}_d]]\ \text{diag}(\|\tilde{\boldsymbol{\gamma}}_1\|_1, \cdots, \|\tilde{\boldsymbol{\gamma}}_d\|_1)$$

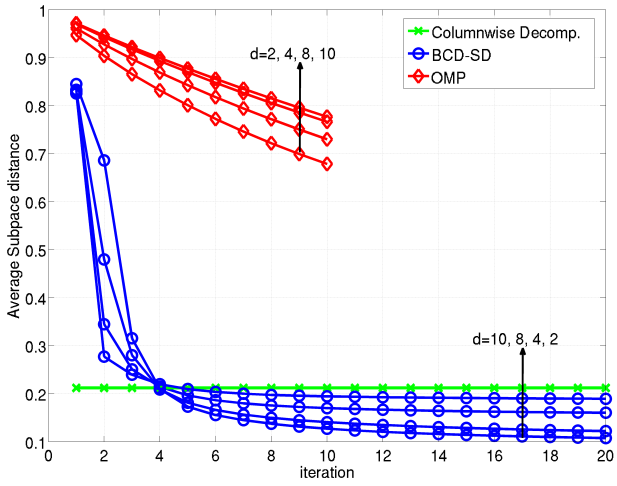

Fig. 2: Average subspace distance $\|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{F}\boldsymbol{G}\|_F^2$, for our proposed method and OMP

3) Numerical Results: As mentioned earlier, (11) was formulated and solved in [6], using a variation on the well-known Orthogonal Matching Pursuit (OMP), by recovering $\boldsymbol{F}$ in a greedy manner, then updating the estimate of $\boldsymbol{G}$ in a least squares sense. We thus compare its average performance with our proposed method, for a case where $\tilde{\boldsymbol{\Gamma}}_1 \in \mathbb{C}^{M \times d}$ is such that $M = 64, r = 10$ (for several values of d). The reason for the massive performance gap in Fig. 2 is that BCD-SD attempts to find a locally optimal solution to (11) (though this cannot be shown due to the coupled variables). Moreover, OMP is halted after r iterations, since it recovers the columns of $\boldsymbol{F}$ one at time, whereas our proposed method runs until reaching a stable point. Interestingly, despite its extreme simplicity, the column-wise decomposition offers a surprisingly good performance (as seen in Fig. 2).

B. Echoing in the Hybrid Analog-Digital Architecture

It is clear by now that the gist behind the schemes described in this work, is to obtain an estimate of $\{\boldsymbol{H}^{\dagger}\boldsymbol{H}\boldsymbol{q}_l\}_{l=1}^m$ at the BS, by exploiting channel reciprocity, using BS-initiated echoing described in (7). However, in the case of the hybrid analog-digital architecture, there are several issues that prevent the application of the latter procedure. Firstly, the digital beamforming vector $\boldsymbol{q}_l$ needs to be approximated by a cascade of analog and digital beamformer, using the decomposition in Sect. IV-A, i.e., $\boldsymbol{q}_l = \tilde{\boldsymbol{f}}_l \tilde{g}_l + \boldsymbol{e}_l$, where $\boldsymbol{e}_l$ is the approximation error given in (16). Moreover, the BS-initiated echoing relies on the MS being able to amplify-and-forward its received signal: this is clearly *not possible* using the hybrid analog-digital architecture. In addition, neither the BS nor MS can

digitally process the received signal at the antennas: only after the application the analog precoder / combiner (and possibly the digital precoder / combiner) can the baseband signal be digitally manipulated [6], [10].

With this in mind, we *emulate* the A-F step in BS-initiated echoing, (7), as follows. $\boldsymbol{q}_l$ is decomposed into $\tilde{\boldsymbol{f}}_l\tilde{g}_l$ at the BS and sent over the DL. The MS linearly processes the received signal in the downlink, with the analog combiner, i.e., $\boldsymbol{s}_l = \boldsymbol{W}_l^\dagger(\boldsymbol{H}\tilde{\boldsymbol{f}}_l\tilde{g}_l)$, and same filter is used as the analog precoder, to process the transmit signal in the UL, i.e., $\boldsymbol{W}_l\boldsymbol{s}_l$. Finally, the received signal at the BS is processed with the analog precoder, $\boldsymbol{F}_l$. The resulting estimate, $\boldsymbol{p}_l$, at the BS is,

$$\boldsymbol{p}_l = \boldsymbol{F}_l^\dagger\boldsymbol{H}^\dagger\boldsymbol{W}_l\boldsymbol{W}_l^\dagger\boldsymbol{H}(\boldsymbol{q}_l - \boldsymbol{e}_l) \tag{17}$$

Note that the above process is possible using the hybrid analog-digital architecture. In this section, we opt to ignore noise at both the BS and MS, and focus on other sources of distortion that can be dealt with, such as the decomposition error, $\boldsymbol{e}_l$. It is clear from (17) that $\boldsymbol{p}_l$ is no longer a "good" estimate of $\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{q}_l$, for the reasons stated below.

1. *Analog-Processing Impairments (API):* Processing the signal at the MS with the analog combiner / precoder $\boldsymbol{W}_l$ greatly distorts the the singular values / vectors of the effective channel. Moreover, processing the received signal at the BS with the analog combiner $\boldsymbol{F}_l \in \mathbb{C}^{M\times r}$ implies that $\boldsymbol{p}_l$ is now a low-dimensional observation of the desired M-dimensional quantity $\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{q}_l$ (since $r < M$).
2. *Decomposition-Induced Distortions (DID):* The error from decomposing $\boldsymbol{q}_l$ at the BS, $\boldsymbol{e}_l$, further distorts the estimate (as seen in (17)).

The above impairments are a byproduct of shifting the massive digital precoding burden to the analog domain. In what follows, these impairments will individually be investigated and addressed.

1) Cancellation of Analog-Processing Impairments: Our proposed method for mitigating analog-processing impairments (API) relies on the simple idea of taking multiple measurements at both the BS and MS, and linearly combining them, such that $\boldsymbol{W}_l\boldsymbol{W}_l^\dagger$ and $\boldsymbol{F}_l\boldsymbol{F}_l^\dagger$ approximate an identity matrix.

In the DL, $\boldsymbol{q}_l$ is approximated by $\tilde{\boldsymbol{f}}_l\tilde{g}_l$, and $\tilde{\boldsymbol{f}}_l\tilde{g}_l$ is sent over the DL channel[6], K_r times (where $K_r = N/r$), each linearly processed with an analog combiner $\{\boldsymbol{W}_{l,k} \in \mathbb{C}^{N\times r}\}_{k=1}^{K_r}$, to obtain the digital samples $\{\boldsymbol{s}_{l,k}\}_{k=1}^{K_r}$ (this process is shown in Table (IV)). Moreover, the analog combiners are taken from the columns of a Discrete Fourier Transform (DFT) matrix, i.e,

$$[\boldsymbol{W}_{l,1}, ..., \boldsymbol{W}_{l,K_r}] = \boldsymbol{D}_r, \tag{18}$$

where $\boldsymbol{D}_r \in \mathbb{C}^{N\times N}$ is a normalized $N \times N$ DFT matrix (i.e., where each column has unit norm and satisfies the unit-modulus constraint). The same analog combiners, $\{\boldsymbol{W}_{l,k}\}_k$, are used to linearly combine $\{\boldsymbol{s}_{l,k}\}_k$, to form $\tilde{\boldsymbol{s}}_l$. We dub this procedure Repetition-Aided (RAID) Echoing, and the aforementioned DL phase, is shown in Table IV. The resulting signal at the MS, $\tilde{\boldsymbol{s}}_l$, can be rewritten as,

$$\tilde{\boldsymbol{s}}_l = \left(\sum_{k=1}^{K_r}\boldsymbol{W}_{l,k}\boldsymbol{W}_{l,k}^\dagger\right)\boldsymbol{H}(d\tilde{\boldsymbol{f}}_l\tilde{g}_l) = d\boldsymbol{H}\tilde{\boldsymbol{f}}_l\tilde{g}_l \tag{19}$$

where equality follows from our earlier definition of $\{\boldsymbol{W}_{l,k}\}_k$ as columns of DFT matrices in (18). Note that *the effect of processing the received signal with the analog combiner has been completely suppressed*. Now, $\tilde{\boldsymbol{s}}_l$ is normalized, and echoed back in the UL direction.

A quite similar process is used in the UL: $\tilde{\boldsymbol{s}}_l$ is first decomposed into $\tilde{\boldsymbol{w}}_l\tilde{u}_l$, d RF chains are used to send it over the UL, K_t times (where $K_t = M/r$), and each observation is linearly processed with an analog combiner $\{\boldsymbol{F}_{l,m} \in \mathbb{C}^{M\times r}\}_{m=1}^{K_t}$. The resulting digital samples $\{\boldsymbol{z}_{l,m}\}_{m=1}^{K_t}$ are again linearly combined with the same $\{\boldsymbol{F}_{l,m}\}_m$, to obtain the desired estimate $\boldsymbol{p}_l$. Similar to the DL case, the analog combiners are taken from the columns of a Discrete Fourier Transform (DFT) matrix, i.e, $[\boldsymbol{F}_{l,1}, ..., \boldsymbol{F}_{l,K_t}] = \boldsymbol{D}_t$. The process for the UL is also shown in Table IV. We combine its steps to rewrite $\boldsymbol{p}_l$ as,

$$\boldsymbol{p}_l = \left(\sum_{m=1}^{K_t}\boldsymbol{F}_{l,m}\boldsymbol{F}_{l,m}^\dagger\right)\boldsymbol{H}^\dagger(d\tilde{\boldsymbol{w}}_l\tilde{u}_l) = d\boldsymbol{H}^\dagger\tilde{\boldsymbol{w}}_l\tilde{u}_l \tag{20}$$

At the output of the RAID procedure, the BS has the following $\boldsymbol{p}_l$,

$$\begin{aligned}\boldsymbol{p}_l &= d\boldsymbol{H}^\dagger\tilde{\boldsymbol{w}}_l\tilde{u}_l = d\boldsymbol{H}^\dagger(\tilde{\boldsymbol{s}}_l - \boldsymbol{e}_l^{(r)}) = d\boldsymbol{H}^\dagger(d\boldsymbol{H}\tilde{\boldsymbol{f}}_l\tilde{g}_l - \boldsymbol{e}_l^{(r)})\\ &= d^2\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{q}_l - d^2\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{e}_l^{(t)} - d\boldsymbol{H}^\dagger\boldsymbol{e}_l^{(r)}\end{aligned} \tag{21}$$

Note that $\boldsymbol{e}_l^{(t)} = \boldsymbol{q}_l - \tilde{\boldsymbol{f}}_l\tilde{g}_l$ (resp. $\boldsymbol{e}_l^{(r)} = \tilde{\boldsymbol{s}}_l - \tilde{\boldsymbol{w}}_l\tilde{u}_l$) is the error emanating from approximating $\boldsymbol{q}_l$ (resp. $\tilde{\boldsymbol{s}}_l$) at the BS (resp. MS), that we dub *BS-side* (resp. MS-side) *decomposition-induced distortion (DID)*. It is quite insightful to compare $\boldsymbol{p}_l$ in the latter equation with (17). We can clearly see that impairments originating from processing the received signals with both $\boldsymbol{W}_l$ and $\boldsymbol{F}_l$, have completely been suppressed. In (21), $\boldsymbol{p}_l$ indeed is the desired estimate, i.e., $\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{q}_l$, corrupted by distortions emanating for the BS-side decomposition, $\boldsymbol{e}_l^{(t)}$, and

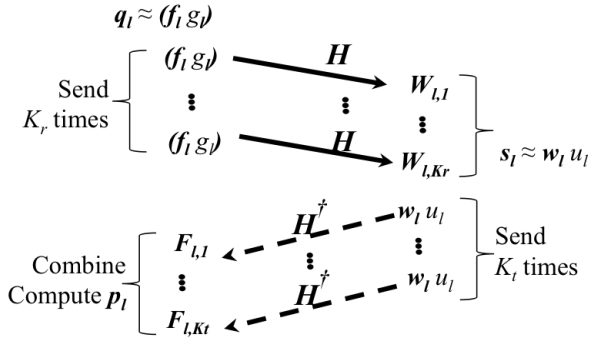

Fig. 3: Repetition-aided (RAID) echoing for the hybrid analog-digital architecture

[6]When sending $\tilde{\boldsymbol{f}}_l\tilde{g}_l$ over the DL, we can use d RF chains, i.e.,

$$\boldsymbol{F}_l\boldsymbol{G}_l\ \mathbf{1}_d = [\tilde{\boldsymbol{f}}_l, \cdots, \tilde{\boldsymbol{f}}_l]\ \mathrm{diag}(\tilde{g}_l, \cdots, \tilde{g}_l)\ \mathbf{1}_d = d\tilde{\boldsymbol{f}}_l\tilde{g}_l$$

thereby resulting in an array gain factor of d. Moreover, since we know from (14) that $\|\tilde{\boldsymbol{f}}_l\tilde{g}_l\|_2^2 \leq 1$, indeed this transmission scheme satisfies the power constraint. We also make use of this observation in the UL sounding.

// DL phase
$\boldsymbol{q}_l = \tilde{\boldsymbol{f}}_l \tilde{g}_l + \boldsymbol{e}_l^{(t)}$
$\boldsymbol{s}_{l,k} = \boldsymbol{W}_{l,k}^{\dagger} \boldsymbol{H} (d\tilde{\boldsymbol{f}}_l \tilde{g}_l),\ \forall k \in \{K_r\}$
$\tilde{\boldsymbol{s}}_l = \sum_{k=1}^{K_r} \boldsymbol{W}_{l,k} \boldsymbol{s}_{l,k}$
// UL phase
$\tilde{\boldsymbol{s}}_l = \tilde{\boldsymbol{w}}_l \tilde{u}_l + \boldsymbol{e}_l^{(r)}$
$\boldsymbol{z}_{l,m} = \boldsymbol{F}_{l,m}^{\dagger} \boldsymbol{H}^{\dagger} (d\tilde{\boldsymbol{w}}_l \tilde{u}_l),\ \forall m \in \{K_t\}$
$\boldsymbol{p}_l = \sum_{m=1}^{K_t} \boldsymbol{F}_{l,m} \boldsymbol{z}_{l,m}$

TABLE IV: Repetition-Aided (RAID) echoing

the MS side decomposition, $\boldsymbol{e}_l^{(r)}$ (both investigated later in the next subsection). Both UL and DL phases of he process are illustrated in Fig. 3, and detailed in Table IV.

Remark 2. Note that employing this process reduces the hybrid analog-digital architecture into a conventional MIMO channel: any transmitted vector in the DL, $(\tilde{\boldsymbol{f}}_l \tilde{g}_l)$, can be received in a "MIMO-like" fashion, as seen from (19), at a cost of K_r channel uses (the same holds for the UL, as seen from (20)).

It can be seen from the above, that in the DL (resp. UL), d RF chains are active at the BS (resp. MS), while all r RF chains are used at the MS (resp. BS), to minimize the overhead. With this in mind, it can be seen that the associated overhead with each echoing, $\Omega = (M+N)/r$ (channel uses), will decrease as more RF chains are used.

2) Imperfect Compensation of Analog-Processing Impairments: Though the above method perfectly removes all artifacts of analog processing, the overhead is proportional to $(M+N)/r$. A natural question is whether a similar result can still be achieved when $\boldsymbol{D}_r$ and $\boldsymbol{D}_t$ are truncated matrices i.e. when $K_r < N/r$ and $K_t < M/r$. Perfect cancellation of API relies on a careful choice of the analog precoder / combiner for each measurement, by picking $\{\boldsymbol{W}_{l,k}\}_{k=1}^{K_r}$ and $\{\boldsymbol{F}_{l,m}\}_{m=1}^{K_t}$ to span all the columns of (square) DFT matrices. We investigate the effect of picking $\boldsymbol{D}_r$ and $\boldsymbol{D}_t$ as truncated matrices, i.e. when $K_r < N/r$ and $K_t < M/r$. Focusing our discussion on just analog precoders for brevity, we seek to find a (tall) matrix $\tilde{\boldsymbol{D}}_t \in \mathbb{C}^{M \times (\eta M)}$, $\eta < 1$, such that,

$$\begin{cases} \min\limits_{\tilde{\boldsymbol{D}}_t} \| \frac{1}{M} \boldsymbol{I}_M - \tilde{\boldsymbol{D}}_t \tilde{\boldsymbol{D}}_t^{\dagger} \|_F^2 \\ \text{s. t. } \ \tilde{\boldsymbol{D}}_t \in \mathcal{S}_{M,\ \eta M}\ . \end{cases} \tag{22}$$

Due to the apparent difficulty of the problem, one can resort to *stochastic optimization* tools, e.g. simulated annealing: this approach is ideal for the design of $\tilde{\boldsymbol{D}}_t$ (and $\tilde{\boldsymbol{D}}_r$ as well), since it is completely independent of all parameters (except M, N and η), and can thus be computed off-line and stored for later use. Then, the resulting overhead would be reduced to $\Omega = \eta \frac{M+N}{r}$. Further investigations along this line are outside the scope of this work, but we opted to include them briefly, for completeness.

3) Decomposition-Induced Distortions: We investigate the effect of BS-side DID, $\boldsymbol{e}_l^{(t)}$, and MS-side DID, $\boldsymbol{e}_l^{(r)}$, that distort $\boldsymbol{p}_l$, at the output of the RAID procedure in (21). It can be easily verified that $\boldsymbol{e}_l^{(t)}$ only distorts the magnitude of $\boldsymbol{H}^{\dagger}\boldsymbol{H}\boldsymbol{q}_l$, not its phase, and consequently its effect is minimal and can be neglected. Since this claim cannot be made for the MS-side DID, $\boldsymbol{e}_l^{(r)}$, we provide a mechanism for mitigating the latter.

Our intuition for compensating for MS-side DID is to obtain an estimate of $\boldsymbol{H}^{\dagger}\boldsymbol{e}_l^{(r)}$ at the BS, and combine it with $\boldsymbol{p}_l$ in (21). Though at a first glance, one obvious solution is for the MS to send $\boldsymbol{e}_l^{(r)}$ in the UL, the fact is, $\boldsymbol{e}_l^{(r)}$ does not satisfy the hardware constraints. Letting

$$\boldsymbol{\psi}_l = \Pi_{\mathcal{S}}[\boldsymbol{e}_l^{(r)}] = (1/\sqrt{N})[e^{j\phi_1}, \cdots, e^{j\phi_N}]^T,$$

we recall that $\boldsymbol{\psi}_l$ has the same phases as $\boldsymbol{e}_l^{(r)}$. In a nutshell, our proposed method is to first approximate $\boldsymbol{e}_l^{(r)}$ by $\alpha_l \boldsymbol{\psi}_l$ at the MS, send $\boldsymbol{\psi}_l$ over the UL, and feedback α_l(note that $\boldsymbol{\psi}_l$ can be directly sent in the UL using the analog combiner). Further investigations reveal that indeed this mechanism works. However, since it comes at an additional overhead cost of $(M+N)/r$, we opt not to include them in the RAID procedure.

C. Proposed Algorithms

Combining the results of the previous subsections, we can now formulate our algorithm for Subspace Estimation and Decomposition (SED) for the hybrid analog-digital architecture (shown in Algorithm 1): estimates of the right / left singular subspaces, $\tilde{\boldsymbol{\Gamma}}_1$ and $\tilde{\boldsymbol{\Phi}}_1$, can be obtained by using the SE-ARN procedure (Sect. III), keeping in mind that the *echoing phase (Steps 1.a and 1.b) is now replaced by the RAID echoing procedure* (Sect. IV-B3). Then, the multi-dimensional subspace decomposition procedure, BCD-SD in Sect. IV-A, is then used to approximate each of the estimated singular spaces, by a cascade of analog and digital precoder / combiner.

Algorithm 1 Subspace Estimation and Decomposition (SED) for Hybrid Analog-Digital Architecture

// Estimate $\tilde{\boldsymbol{\Gamma}}_1$ *and* $\tilde{\boldsymbol{\Phi}}_1$
$\tilde{\boldsymbol{\Gamma}}_1 =$ SE-ARN ($\boldsymbol{H}$, d)
$\tilde{\boldsymbol{\Phi}}_1 =$ SE-ARN ($\boldsymbol{H}^{\dagger}$, d)
// Decompose $\tilde{\boldsymbol{\Gamma}}_1$ *and* $\tilde{\boldsymbol{\Phi}}_1$
[$\boldsymbol{F}$, $\boldsymbol{G}$] = BCD-SD ($\tilde{\boldsymbol{\Gamma}}_1$, ρ)
[$\boldsymbol{W}$, $\boldsymbol{U}$] = BCD-SD ($\tilde{\boldsymbol{\Phi}}_1$, ρ)

Note that previously proposed algorithms within this context such as the PM and TQR in [14], are no longer applicable here: indeed both rely on the MS being able to amplify-and-forward its received signal at the antennas, back to the BS, to form estimates for $\{\boldsymbol{H}^{\dagger}\boldsymbol{H}\boldsymbol{q}_l\}_l$. Clearly this modus operandi cannot be supported by the hybrid analog-digital architecture. Interestingly, it is possible to apply elements from the RAID echoing structure that we developed, effectively modifying the original echoing structure of the latter schemes, and adapting them to the hybrid analog-digital architecture (as shown in Algorithm 2). Due to limited space, we will skip the details behind the algorithm steps, most of them have been already elaborated.

D. Bounds on Eigenvalue Perturbation

It can be clearly seen that *the iterative nature of Algorithm 2 makes the application of Lemma 2, to quantify the impact of*

Algorithm 2 Modified Two-way QR (MTQR) for the hybrid analog-digital architecture

for $l = 1, 2, ..., I$ **do**
  // Approximate columns of $\boldsymbol{X}_l$
  $[\boldsymbol{X}_l]_n \approx \tilde{\boldsymbol{f}}_{l,n}\tilde{g}_{l,n},\ \forall n \in \{d\}$
  $\tilde{\boldsymbol{X}}_l = [\ \tilde{\boldsymbol{f}}_{l,1}\tilde{g}_{l,1}\ , \cdots,\ \tilde{\boldsymbol{f}}_{l,d}\tilde{g}_{l,d}\ ]$
  // Send $\tilde{\boldsymbol{X}}_l$ *in DL, one column at time*
  $\boldsymbol{T}_{l,k} = \boldsymbol{W}_{l,k}^{\dagger}\boldsymbol{H}\tilde{\boldsymbol{X}}_l,\ \forall k \in \{K_r\}$
  $\boldsymbol{Y}_l = \sum_{k=1}^{K_r} \boldsymbol{W}_{l,k}\boldsymbol{T}_{l,k}\ ;\ \boldsymbol{Y}_{l+1} = \text{qr}(\boldsymbol{Y}_l)$
  // Approximate columns of $\boldsymbol{Y}_{l+1}$
  $[\boldsymbol{Y}_{l+1}]_n \approx \tilde{\boldsymbol{w}}_{l+1,n}\tilde{u}_{l+1,n},\ \forall n \in \{d\}$
  $\tilde{\boldsymbol{Y}}_{l+1} = [\ \tilde{\boldsymbol{w}}_{l+1,1}\tilde{u}_{l+1,1}\ , \cdots,\ \tilde{\boldsymbol{w}}_{l+1,d}\tilde{u}_{l+1,d}\ ]$
  // Send $\tilde{\boldsymbol{Y}}_{l+1}$ *in UL, one column at time*
  $\boldsymbol{S}_{l,k} = \boldsymbol{F}_{l,k}^{\dagger}\boldsymbol{H}^{\dagger}\tilde{\boldsymbol{Y}}_{l+1},\ \forall k \in \{K_t\}$
  $\boldsymbol{Z}_l = \sum_{k=1}^{K_t} \boldsymbol{F}_{l,k}\boldsymbol{S}_{l,k}\ ;\ \boldsymbol{X}_{l+1} = \text{qr}(\boldsymbol{Z}_l)$
end for

decomposition and approximation errors, not possible. On the other hand, for Algorithm 1, the fact that each $\boldsymbol{H}^{\dagger}\boldsymbol{H}\boldsymbol{q}_l$ is only corrupted by two sources of DID, $\boldsymbol{e}_l^{(r)}$ and $\boldsymbol{e}_l^{(r)}$, makes the latter possible. With that in mind, *we specialize the result of Sect. III-B and Lemma 2 (developed for generic MIMO systems) to the case of Algorithm 1 in the hybrid analog-digital architecture*. We thus relate the eigenvalues of $\boldsymbol{T}_m$ at the output of SE-ARN, to the desired eigenvalues of $\boldsymbol{C}_m$, and consequently of $\boldsymbol{A}$ (Sect.III-B).

Corollary 1. *Every eigenvalue* $\tilde{\lambda}$ *of* $\boldsymbol{T}_m$ *satisfies*

$$|\tilde{\lambda} - \lambda| \leq m\|\boldsymbol{H}\|_F^2\ (3 + \frac{1}{d\|\boldsymbol{H}\|_F})$$

where λ *in an eigenvalue of* $\boldsymbol{C}_m$.

Proof: Refer to Appendix F ∎

Moreover, recall that as $m \to M$, λ is an eigenvalue of $\boldsymbol{A}$ (Lemma 1 - P3). Thus, this result directly relates the eigenvalues of $\boldsymbol{T}_m$, to that of $\boldsymbol{A}$: though this holds asymptotically in m, our simulations will show that good approximations can still be obtained, even for $m \ll M$. Note that we have ignored the effect of DID compensation, within the RAID echoing process, for convenience. As a result, the above bound is a "pessimistic" performance measure.

E. Practical Implementation Aspects

We evaluate the *communication overhead* of both schemes, in number of channel uses, keeping in mind that the actual overhead will be dominated by the latter. Algorithm 1 requires K_t+K_r channel uses per iteration, to estimate $\tilde{\boldsymbol{\Gamma}}_1$, and K_t+K_r to estimate $\tilde{\boldsymbol{\Phi}}_1$, for a total of

$$\Omega_{SED} = 2m\ \frac{M+N}{r}, \tag{23}$$

m being the number of iterations for the Arnoldi process. Letting I denote the number of iterations for MTQR, the number of channel uses required for Algorithm 2 is,

$$\Omega_{MTQR} = dI\ \frac{M+N}{r} \tag{24}$$

F. Discussion

We have presented an approach to maximizing the metric R defined in (4). As mentioned earlier, the value of the objective function is in general not an achievable rate for our system. However, optimizing similar expressions related to achievable rates has been proved to give good results in previous work on transmission with partial CSI [27]. Since any rate achievable with partial CSI, cannot be larger than the corresponding rate achievable with perfect CSI, this criterion always provides an upper bound on the achievable rates in our system. Hence, in our approach, if the proposed algorithms result in values for R that are closing in on the perfect CSI upper bound, then the scheme is performing optimally (in the sense of achievable rates). With that in mind, and letting $\tilde{\boldsymbol{H}}$ be the channel estimate resulting from our proposed methods, we use the following, as our performance metric in the simulations,

$$\tilde{R} = \log_2 \left| \boldsymbol{I}_d + \frac{1}{\sigma_{(r)}^2}\boldsymbol{U}^{\dagger}\boldsymbol{W}^{\dagger}\tilde{\boldsymbol{H}}\boldsymbol{F}\boldsymbol{G}\boldsymbol{G}^{\dagger}\boldsymbol{F}^{\dagger}\tilde{\boldsymbol{H}}^{\dagger}\boldsymbol{W}\boldsymbol{U}(\boldsymbol{U}^{\dagger}\boldsymbol{W}^{\dagger}\boldsymbol{W}\boldsymbol{U})^{-1} \right| .$$

V. Numerical Results

A. Simulation Setup

In this section, we numerically evaluate the performance of our algorithms, in the context of a single-user MIMO link. We adopt the prevalent physical representation of sparse mmWave channels adopted in the literature, e.g., [6], [7] , where only L scatterers are assumed to contribute to the received signal - an inherent property of the poor scattering nature in mmWave channels,

$$\boldsymbol{H} = \sqrt{\frac{MN}{L}} \sum_{i=1}^{L} \beta_i\ \boldsymbol{a}_r(\chi_i^{(r)})\boldsymbol{a}_t^{\dagger}(\chi_i^{(t)}) \tag{25}$$

where $\chi_i^{(r)}$ and $\chi_i^{(t)}$ are angles of arrival at the MS, and angles of departure at the BS (AoA / AoD) of the i^{th} path, respectively (both assumed to be uniform over $[-\pi/2,\ \pi/2]$), β_i is the complex gain of the i^{th} path such that $\beta_i \sim \mathcal{CN}(0,1),\ \forall i$. Finally, $\boldsymbol{a}_r(\chi_i^{(r)})$ and $\boldsymbol{a}_t(\chi_i^{(t)})$ are the array response vectors at both the MS and BS, respectively. For simplicity, we will use uniform linear arrays (ULAs), where we assume that the inter-element spacing is equal to half of the wavelength. In what follows, we also assume that $M/r = 8$ and $N/r = 4$, i.e., as M, N increase, so does the number of RF chains.

1) Benchmarks / Upper bounds: We use the Adaptive Channel Estimation (ACE) method (Algorithm 2 in [7]) as a benchmark, to estimate the mmWave channel. It is based on sounding of *hierarchical codebooks* at the BS, feedback of the best codebook indexes by the MS, and finding the analog / digital precoders and combiners using OMP [6]. Moreover, the authors characterized the resulting communication overhead Ω_{ACE}, as a function of the codebook resolution. We used the corresponding MATLAB implementation that was provided by

the authors. We adjust the number of iterations for both our proposed schemes, and the codebook resolution of benchmark scheme, such that $\Omega_{SED} = \Omega_{MTQR} \triangleq \Omega_o \approx \Omega_{ACE}$. Note that we do not assume any quantization for phases of the RF filters.

We also compare the algorithms' performance against the "optimal performance", $R^\star$ in (5), where full CSIT/CSIR is assumed, fully digital precoding is employed, and the optimal precoders are used. All curves are averaged over 500 channel realizations.

Remark 3. Note that if one want to use "classical" pilot-based channel estimation to estimate the DL channel, i.e., a pilot sequence of minimum length M, then the same repetition-based framework that was used in RAID echoing, has to be used to cancel the effect of $\boldsymbol{W}$ from the effective channel estimate: it can be easily seen that the resulting total (both DL and UL) number of pilots slots would be $2MN/r^2$, thereby making the latter method infeasible.

B. Performance Evaluation

We start by investigating the performance of our schemes against the above benchmarks, for the case where $M = 128, N = 64, L = 3$, and $m = 3d$, for two cases: $d = 1$ and $d = 2$ where the resulting overhead is $\Omega_o = 72$ and $\Omega_o = 144$, respectively. It can be seen from Fig. 4 that both *proposed schemes exhibit relatively similar performances*, that are in turn very *close to the optimal performance bound* $R^\star$ (especially at medium-to-high SNR). This indeed suggests that the multiplexing gain achieved by conventional MIMO systems can still be maintained in the hybrid analog-digital architecture, albeit at a much lower cost: the number of *required RF chains can be drastically decreased*, resulting in savings in terms of cost and power consumption. Moreover, we observe a sharp and significant performance gap between both our schemes, and the benchmark from [7], over all SNR ranges (the gap being more significant in the low-SNR regime).

Though the performance of Algorithm 2 seems to be somewhat better, we opt to focus on Algorithm 1, since it is the main focus of the current work. With that in mind, we next investigate its scalability: we scale up M and N (assuming $N = M/2$, for simplicity), while keeping everything else fixed, i.e., $d = 2, m = 6$, and consequently $\Omega_o = 144$. In doing that, we noticed that the complexity of the benchmark scheme was *prohibitively high*, thus preventing us from investigating its scalability: we were unable to get any results for systems larger than 128×64 (our initial investigations suggest that this is chiefly due to the SVD that is applied to the large estimated channel). On the other hand, both our algorithms exhibit no such problems since all the computations that they involve are matrix-vectors / matrix-matrix operations. Consequently, the *complexity gap between Algorithm 1 and the benchmark increases drastically, as* M, N grow.

Fig 5 clearly shows that Algorithm 1 is able to harness the significant array gain inherent to large antenna systems (by closely following the optimal performance bound, $R^\star$, with a small constant gap), while keeping the overhead strikingly small. Though the performance might not be good enough to offset the overhead, for the 16×8 case, it surely does for the 256×128. Moreover, note that that gap between the optimal performance and Algorithm 1 is quite small (across the entire SNR range) for small systems dimensions, and quite small even for large values of M (at high SNR). The key to this rather impressive result is to have M/r and N/r fixed, as M, N increase.

We also evaluate the performance of Algorithm 1 in a more realistic manner, by adopting the Spatial Channel Model (SCM) detailed in [28], [29], and modifying its parameters to emulate mmWave channels: the number of paths is set to 4, the carrier frequency to 60 GHz, the mobile speed is 5 km/h, the BS / MS array is modified to implement ULAs,

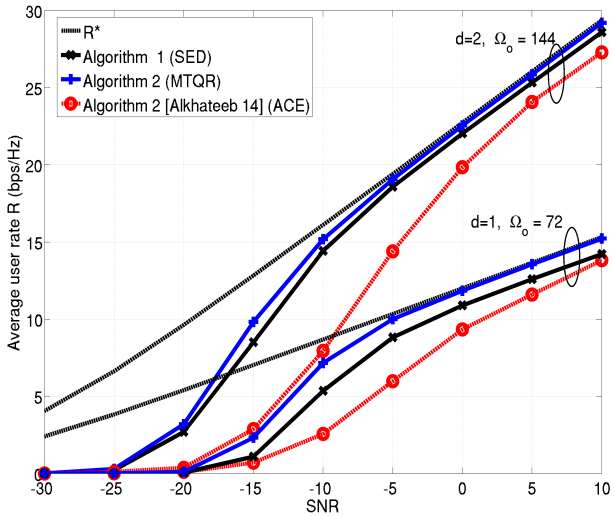

Fig. 4: Average performance of proposed schemes ($M = 128, N = 64, d = 2, L = 3, m = 6$)

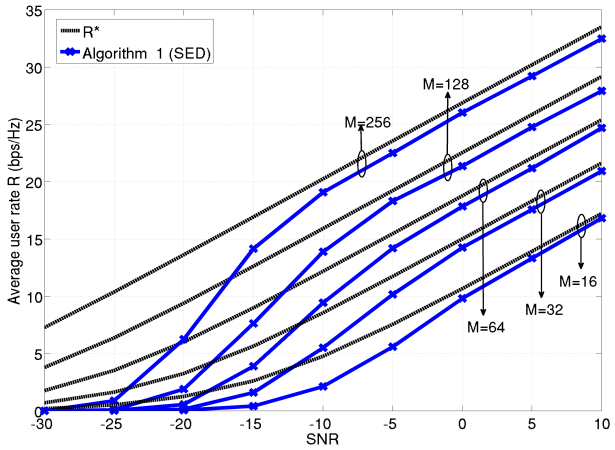

Fig. 5: Average performance for different M, N ($N = M/2, d = 2, L = 4, m = 6, \Omega_o = 144$)

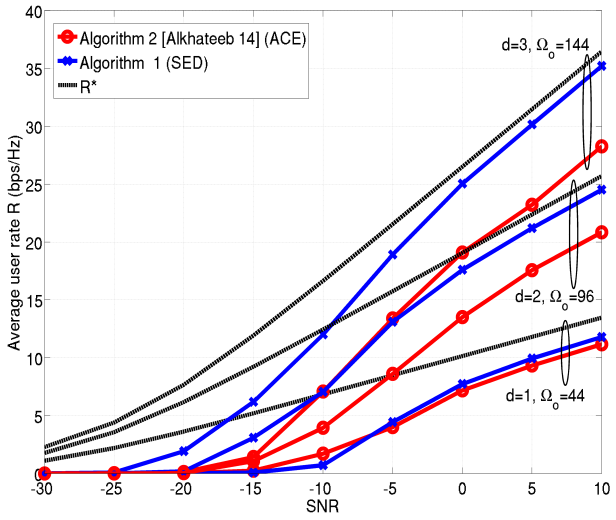

Fig. 6: Average performance of proposed schemes over SCM channels ($M = 64, N = 32, m = 2d$)

and an 'urban micro' scenario is selected, where small a Ω_o is desired. Fig. 6 shows the average performance of such a system, with $M = 64, N = 32, m = 2d$, for several values of d (each resulting in different values for Ω_o). Though both our algorithm, and the benchmark exhibit similar performances for $d = 1$, this gap increases with d, e.g. for $d = 3$ this performance gap is quite significant. Moreover, we can clearly see that Algorithm 1 yields a relatively high throughput in this realistic simulation setting (especially for $d = 3$), while still keeping the overhead at a relatively low level.

VI. Conclusion

We proposed an algorithm for blindly estimating the left and right singular subspace of a mmWave MIMO channel, by exploiting channel reciprocity that is inherent to TDD systems. Though the algorithm is a perfect match for conventional (large) MIMO systems, we extend it to operate under the constraints dictated by the hybrid analog-digital architecture, and show via simulations that it is ideal for large MIMO channels, with low rank, e.g., mmWave channels. Finally, our simulations showed that a similar performance to the ideal case (fully digital perfect CSI) can be achieved, with a only a few RF chains, thereby resulting in significant saving in energy and cost, over conventional MIMO systems.

Appendix

A. Proof of Proposition 1

Assuming w.l.o.g. that $\boldsymbol{FG} \subseteq \text{span}(\boldsymbol{\Gamma}_1)$ and $\boldsymbol{WU} \subseteq \text{span}(\boldsymbol{\Phi}_1)$, then $\boldsymbol{H}_e = \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger (\boldsymbol{\Phi}_1 \boldsymbol{\Sigma}_1 \boldsymbol{\Gamma}_1^\dagger) \boldsymbol{FG}$.

$$
\begin{aligned}
D &\triangleq \log_2 \left| \boldsymbol{I}_d + \boldsymbol{\Sigma}_1^2 \right| - \log_2 \left| \boldsymbol{I}_d + \boldsymbol{H}_e \boldsymbol{H}_e^\dagger (\sigma_{(r)}^2 \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{WU})^{-1} \right| \\
&\overset{(a)}{\lessapprox} \text{tr}(\boldsymbol{I}_d + \boldsymbol{\Sigma}_1^2) - \text{tr}\left(\boldsymbol{H}_e \boldsymbol{H}_e^\dagger (\sigma_{(r)}^2 \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{WU})^{-1} \right) \\
&\overset{(b)}{\leq} c - (\sigma_{max}^2[\boldsymbol{WU}]/\sigma_{(r)}^2) \text{tr}\left(\boldsymbol{H}_e \boldsymbol{H}_e^\dagger \right) \leq c - (\delta/\sigma_{(r)})^2 \text{tr}\left(\boldsymbol{H}_e \boldsymbol{H}_e^\dagger \right) \\
&\overset{(c)}{=} c - (\delta/\sigma_{(r)})^2 \text{tr}\left(\boldsymbol{\Sigma}_1 (\boldsymbol{\Gamma}_1^\dagger \boldsymbol{FGG}^\dagger \boldsymbol{F}^\dagger \boldsymbol{\Gamma}_1) \boldsymbol{\Sigma}_1 (\boldsymbol{\Phi}_1^\dagger \boldsymbol{WUU}^\dagger \boldsymbol{W}^\dagger \boldsymbol{\Phi}_1) \right) \\
&\overset{(b)}{\leq} c - k_o \text{tr}\left[(\boldsymbol{\Phi}_1^\dagger \boldsymbol{WUU}^\dagger \boldsymbol{W}^\dagger \boldsymbol{\Phi}_1)(\boldsymbol{\Gamma}_1^\dagger \boldsymbol{FGG}^\dagger \boldsymbol{F}^\dagger \boldsymbol{\Gamma}_1) \right] \triangleq D_1
\end{aligned}
$$

where $c = \text{tr}(\boldsymbol{I}_d + \boldsymbol{\Sigma}_1^2)$ and $k_o = (\delta \sigma_d / \sigma_{(r)})^2$. Note that (a) follows from $\log |\boldsymbol{A}| \leq \text{tr}(\boldsymbol{A})$ and $\log |\boldsymbol{I} + \boldsymbol{A}| \approx \text{tr}(\boldsymbol{A}), \forall\ \boldsymbol{A} \succeq \boldsymbol{0}$ (Taylor expansion of $\log |\boldsymbol{I} + \boldsymbol{A}|$), (b) form the fact that $\text{tr}(\boldsymbol{AB}) \geq \lambda_{min}[\boldsymbol{A}] \text{tr}(\boldsymbol{B}), \forall\ \boldsymbol{A}, \boldsymbol{B} \succeq \boldsymbol{0}$, and (c) from the circular invariance of the trace. We use the tr() lower bound to further bound D_1 in two ways,

$$
\begin{aligned}
D_1 &\leq c - k_o \sigma_{\min}^2[\boldsymbol{\Phi}_1^\dagger \boldsymbol{WU}]\ \|\boldsymbol{G}^\dagger \boldsymbol{F}^\dagger \boldsymbol{\Gamma}_1\|_F^2 \leq c - k_o \epsilon^2\ \|\boldsymbol{G}^\dagger \boldsymbol{F}^\dagger \boldsymbol{\Gamma}_1\|_F^2 \\
D_1 &\leq c - k_o \sigma_{\min}^2[\boldsymbol{\Gamma}_1^\dagger \boldsymbol{FG}]\ \|\boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{\Phi}_1\|_F^2 \leq c - k_o \epsilon^2\ \|\boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{\Phi}_1\|_F^2
\end{aligned}
$$

and combine them to yield, a bound on D,

$$
\begin{aligned}
D &\leq c + k_o \epsilon^2 \left(-\|\boldsymbol{G}^\dagger \boldsymbol{F}^\dagger \boldsymbol{\Gamma}_1\|_F^2 - \|\boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{\Phi}_1\|_F^2 \right)/2 \\
&\leq c + k_o \epsilon^2 \left(-\Re\text{tr}(\boldsymbol{G}^\dagger \boldsymbol{F}^\dagger \boldsymbol{\Gamma}_1) - \Re\text{tr}(\boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{\Phi}_1) \right)/2 \\
&< c + k_o \epsilon^2 \left(\|\boldsymbol{\Gamma}_1 - \boldsymbol{FG}\|_F^2 + \|\boldsymbol{\Phi}_1 - \boldsymbol{WU}\|_F^2 \right)/2
\end{aligned}
$$

B. Proof of Lemma 1

$(P1)$: Combining steps (2.b) and (3.a) in the SE-ARN procedure, we write,

$$
\boldsymbol{A}\boldsymbol{q}_l + \tilde{\boldsymbol{w}}_l = \sum_{i=1}^{l+1} [\tilde{\boldsymbol{T}}_m]_{i,l}\ \boldsymbol{q}_i + \sum_{i=1}^{l} [\boldsymbol{E}_m]_{i,l}\ \boldsymbol{q}_i\ ,\ \forall l \in \{m\},
$$

We can rewrite the latter equation in matrix form, using the definitions of $\tilde{\boldsymbol{T}}_m, \tilde{\boldsymbol{W}}_m$ given in (8),

$$
\boldsymbol{A}\boldsymbol{Q}_m + \tilde{\boldsymbol{W}}_m = \boldsymbol{Q}_m \tilde{\boldsymbol{T}}_m + [\tilde{\boldsymbol{T}}_m]_{m+1,m}\ \boldsymbol{q}_{m+1} \boldsymbol{b}_m^\dagger + \boldsymbol{Q}_m \boldsymbol{E}_m \tag{26}
$$

where $\boldsymbol{b}_m$ is the m^{th} elementary vector, and $\boldsymbol{E}_m = [\boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m]_U$. We can further simplify the above, using the fact that $\boldsymbol{Q}_m^\dagger \boldsymbol{Q}_m = \boldsymbol{I}_m$ and $\boldsymbol{Q}_m^\dagger \boldsymbol{q}_{m+1} = \boldsymbol{0}$,

$$
\boldsymbol{Q}_m^\dagger \boldsymbol{A} \boldsymbol{Q}_m + \boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m = \tilde{\boldsymbol{T}}_m + \boldsymbol{E}_m
$$

Using the definition of $\boldsymbol{E}_m$, we write,

$$
\begin{aligned}
\boldsymbol{Q}_m^\dagger \boldsymbol{A} \boldsymbol{Q}_m &= \tilde{\boldsymbol{T}}_m + [\boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m]_U - \boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m \\
&= \tilde{\boldsymbol{T}}_m - \tilde{\boldsymbol{E}}_m \triangleq \boldsymbol{C}_m
\end{aligned}
$$

where $\tilde{\boldsymbol{E}}_m = [\boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m]_{SL}$, as defined in (8).

$(P2)$: Noting that $\tilde{\boldsymbol{T}}_m + \boldsymbol{E}_m = \boldsymbol{C}_m + \boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m$, we rewrite (26) as,

$$
\boldsymbol{A}\boldsymbol{Q}_m - \boldsymbol{Q}_m \boldsymbol{C}_m = [\tilde{\boldsymbol{T}}_m]_{m+1,m}\ \boldsymbol{q}_{m+1} \boldsymbol{b}_m^\dagger - (\boldsymbol{I}_M - \boldsymbol{Q}_m \boldsymbol{Q}_m^\dagger) \tilde{\boldsymbol{W}}_m
$$

Multiplying the latter equation by $\boldsymbol{s}_i^{(m)}$, and using the fact that $\boldsymbol{C}_m\boldsymbol{s}_i^{(m)} = \lambda_i^{(m)}\boldsymbol{s}_i^{(m)}$, and $\boldsymbol{Q}_m\boldsymbol{s}_i^{(m)} = \boldsymbol{\theta}_i^{(m)}$

$$\begin{aligned}\boldsymbol{A}\boldsymbol{\theta}_i^{(m)} - &\lambda_i^{(m)}\boldsymbol{\theta}_i^{(m)} \\ &= [\tilde{\boldsymbol{T}}_m]_{m+1,m}\ \boldsymbol{q}_{m+1}\boldsymbol{b}_m^\dagger\boldsymbol{s}_i^{(m)} - (\boldsymbol{I}_M - \boldsymbol{Q}_m\boldsymbol{Q}_m^\dagger)\tilde{\boldsymbol{W}}_m\boldsymbol{s}_i^{(m)}\end{aligned}$$

Finally, the desired residual is upper bounded as,

$$\begin{aligned}&\|\boldsymbol{A}\boldsymbol{\theta}_i^{(m)} - \lambda_i^{(m)}\boldsymbol{\theta}_i^{(m)}\|_2^2 \\ &\quad\le ([\tilde{\boldsymbol{T}}_m]_{m+1,m}|\boldsymbol{b}_m^\dagger\boldsymbol{s}_i^{(m)}|)^2 + \|(\boldsymbol{I}_M - \boldsymbol{Q}_m\boldsymbol{Q}_m^\dagger)\tilde{\boldsymbol{W}}_m\boldsymbol{s}_i^{(m)}\|_F^2 \\ &\quad\le ([\tilde{\boldsymbol{T}}_m]_{m+1,m}|[\boldsymbol{s}_i^{(m)}]_m|)^2 + \|\boldsymbol{I}_M - \boldsymbol{Q}_m\boldsymbol{Q}_m^\dagger\|_F^2\|\tilde{\boldsymbol{W}}_m\|_F^2\end{aligned}$$

where the last inequality follows from $\|\boldsymbol{B}_1\boldsymbol{B}_2\boldsymbol{x}\|_2^2 \le \|\boldsymbol{B}_1\|_F^2.\|\boldsymbol{B}_2\|_F^2.\|\boldsymbol{x}\|_2^2$

$(P3)$: The proof immediately follows by noting that $\|\boldsymbol{I}_M - \boldsymbol{Q}_m\boldsymbol{Q}_m^\dagger\|_F^2 \to 0$ and $[\tilde{\boldsymbol{T}}_m]_{m+1,m} \to 0$, as $m \to M$, thereby implying that $\|\boldsymbol{A}\boldsymbol{\theta}_i^{(M)} - \lambda_i^{(M)}\boldsymbol{\theta}_i^{(M)}\|_2^2 \ll 1$.

C. Proof of Lemma 2

The proof follows from a direct application of the Bauer-Fike Theorem [21, Th. 7.2.2]. Let $\boldsymbol{C}_m = \boldsymbol{S}_m\boldsymbol{\Lambda}_m\boldsymbol{S}_m^{-1}$ be the diagonalizable matrix in question, and $\boldsymbol{T}_m = \boldsymbol{C}_m + \boldsymbol{P}_m$ the "perturbed" matrix. Then, every eigenvalue $\tilde{\lambda}$ of $\boldsymbol{T}_m$ satisfies,

$$|\tilde{\lambda} - \lambda|^2 \le \|\boldsymbol{S}_m\|_2^2.\|\boldsymbol{S}_m^{-1}\|_2^2.\|\boldsymbol{P}_m\|_2^2 = \|\boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m\|_2^2$$

where λ is an eigenvalue of $\boldsymbol{C}_m$, and $\|\boldsymbol{B}\|_2 \triangleq \sigma_{\max}(\boldsymbol{B})$ is the vector-induced matrix 2-norm. The last equality follows from the fact that $\boldsymbol{S}_m$ is unitary, as discussed in Lemma 1. Using the fact that $\|\boldsymbol{B}\|_2 \le \|\boldsymbol{B}\|_F$, we rewrite the last equation,

$$\begin{aligned}|\tilde{\lambda} - \lambda|^2 \le \|\boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m\|_F^2 &= \sum_{(i,j)=1}^{m} |\boldsymbol{q}_i^\dagger\tilde{\boldsymbol{w}}_j|^2 \le \sum_{(i,j)=1}^{m} \|\boldsymbol{q}_i\|_2^2\|\tilde{\boldsymbol{w}}_j\|_2^2 \\ &= \sum_{(i,j)=1}^{m} \|\tilde{\boldsymbol{w}}_j\|_2^2 = m\ \|\tilde{\boldsymbol{W}}_m\|_F^2\end{aligned}$$

This concludes the proof.

D. Proof of Lemma 3

Note that there is not loss in optimality by assuming the $g \in \mathbb{R}_+$. Moreover, exploiting the structure of h_o, the globally optimal solution can be found by optimizing for $\boldsymbol{f}$, assuming g is fixed (and vice) versa, i.e.,

$$\boldsymbol{f}^\star \overset{\triangle}{=} \underset{\boldsymbol{f}}{\text{argmin}}\ g^2(\boldsymbol{f}^\dagger\boldsymbol{f}) - 2g\Re(\boldsymbol{f}^\dagger\tilde{\boldsymbol{\gamma}}_1),\ \text{s. t. } [\boldsymbol{f}]_i = 1/\sqrt{M}\ e^{j\phi_i}$$

$$\begin{aligned}\overset{(a)}{\Leftrightarrow}\ \{\phi_i^\star\} &= \underset{\{\phi_i\}}{\text{argmax}}\ 1/\sqrt{M}\ \Re\left(\sum_{i=1}^{M} r_i\ e^{j(\theta_i-\phi_i)}\right) \\ \{\phi_i^\star\} &= \underset{\{\phi_i\}}{\text{argmax}}\ \sum_{i=1}^{M}\Re\left(e^{j(\theta_i-\phi_i)}\right) = \{\theta_i\}\end{aligned}$$

where (a) follows from applying the one-to-one mapping $[\boldsymbol{f}]_i \to 1/\sqrt{M}\ e^{j\phi_i}, \forall i$. Thus, $[\boldsymbol{f}^\star]_i = 1/\sqrt{M}\ e^{j\theta_i}, \forall i$. Plugging $\boldsymbol{f}^\star$ into the original problem, the optimization of g is a simple unconstrained quadratic problem,

$$g^\star \overset{\triangle}{=} \underset{g}{\text{argmin}}\ g^2 - 2g(\|\tilde{\boldsymbol{\gamma}}_1\|_1/\sqrt{M}) = \|\tilde{\boldsymbol{\gamma}}_1\|_1/\sqrt{M} \qquad (27)$$

E. Proof of Proposition 2

Since $\boldsymbol{Y} \in \mathcal{S}_{M,d}$ by definition (i.e., $|[\boldsymbol{Y}]_{i,k}| = 1/\sqrt{M}$) the problem just reduces to finding the phase of each element in $\boldsymbol{Y}$. Thus,

$$\begin{aligned}\boldsymbol{Y} = \Pi_{\mathcal{S}}[\boldsymbol{X}] &\overset{\triangle}{=} \underset{\boldsymbol{U}\in\mathcal{S}_{M,d}}{\text{argmin}}\ \|\boldsymbol{U} - \boldsymbol{X}\|_F^2 \\ &\overset{(a)}{\Leftrightarrow} \underset{\{\theta_{i,k}\}}{\text{argmin}}\ \sum_{i,k} |(1/\sqrt{M})e^{j\theta_{i,k}} - x_{ik}e^{j\phi_{i,k}}|^2 \\ &\Leftrightarrow \{\theta_{i,k}^\star\} = \{\phi_{i,k}^\star\}\end{aligned}$$

where (a) follows from the fact that $\boldsymbol{U}_{i,k} = (1/\sqrt{M})e^{j\theta_{i,k}}, \forall \boldsymbol{U} \in \mathcal{S}_{M,d}$. Thus, we conclude that $[\boldsymbol{Y}]_{i,k} = (1/\sqrt{M})\ e^{j\phi_{i,k}}, \forall (i,k)$. Furthermore, it follows from this formulation that this projection is unique (despite the non-convexity of $\mathcal{S}_{M,d}$).

F. Proof of Corrollary 1

The proof consists of finding a closed-from expression for $\tilde{\boldsymbol{W}}_m$ as a function of $\boldsymbol{e}_l^{(t)}$ and $\boldsymbol{e}_l^{(r)}$, and applying the result of Lemma 2. Note that $\tilde{\boldsymbol{w}}_l$ in (7) can represent any distortion, and by comparing $\boldsymbol{p}_l$ in both (7) and (21), can infer that $\tilde{\boldsymbol{w}}_l = -\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{e}_l^{(t)} - (1/d)\boldsymbol{H}^\dagger\boldsymbol{e}_l^{(r)}$. Thus, $\tilde{\boldsymbol{W}}_m$ in (8) can be written as,

$$\begin{aligned}\tilde{\boldsymbol{W}}_m &= -\boldsymbol{H}^\dagger\boldsymbol{H}[\boldsymbol{e}_1^{(t)}, \cdots, \boldsymbol{e}_m^{(t)}] - (1/d)\boldsymbol{H}^\dagger[\boldsymbol{e}_1^{(r)}, \cdots, \boldsymbol{e}_m^{(r)}] \\ &\triangleq -\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{E}^{(t)} - (1/d)\boldsymbol{H}^\dagger\boldsymbol{E}^{(r)}\end{aligned}$$

Then applying some basic inequalities of the Frobenius norm,

$$\|\tilde{\boldsymbol{W}}_m\|_F \le \|\boldsymbol{H}\|_F^2.\|\boldsymbol{E}^{(t)}\|_F + (1/d)\|\boldsymbol{H}\|_F.\|\boldsymbol{E}^{(r)}\|_F \qquad (28)$$

On the other hand, recall that $\boldsymbol{e}_l^{(t)} = \boldsymbol{q}_l - \tilde{\boldsymbol{f}}_l\tilde{g}_l$ and $\boldsymbol{e}_l^{(r)} = \tilde{\boldsymbol{s}}_l - \tilde{\boldsymbol{w}}_l\tilde{u}_l$. Thus, using the results of Sec. IV-A2,

$$\begin{aligned}\|\boldsymbol{e}_l^{(t)}\|_2 &\le \|\boldsymbol{q}_l\|_2 + \|\tilde{\boldsymbol{f}}_l\tilde{g}_l\|_2 \le 2 \\ \|\boldsymbol{e}_l^{(r)}\|_2 &\le \|d\boldsymbol{H}\tilde{\boldsymbol{f}}_l\tilde{g}_l\|_2 + \|\tilde{\boldsymbol{w}}_l\tilde{u}_l\|_2 \le 1 + d\|\boldsymbol{H}\|_F\end{aligned}$$

and it follows that

$$\|\boldsymbol{E}^{(t)}\|_F \le 2\sqrt{m},\ \|\boldsymbol{E}^{(r)}\|_F \le \sqrt{m}(1 + d\|\boldsymbol{H}\|_F) \qquad (29)$$

The upper bound follows by combining (28) and (29).

REFERENCES

[1] "On the pulse of the networked society," *Ericsson Mobility Report*, Jun 2014.

[2] K. Ohata, K. Maruhashi, J.-i. Matsuda, M. Ito, W. Domon, and S. Yamazaki, "A 500Mbps 60GHz-band transceiver for IEEE 1394 wireless home networks," in *Microwave Conference, 2000. 30th European*, pp. 1–4, Oct 2000.

[3] K. Ohata, K. Maruhashi, M. Ito, S. Kishimoto, K. Ikuina, T. Hashiguchi, K. Ikeda, and N. Takahashi, "1.25 Gbps wireless Gigabit ethernet link at 60 GHz-band," in *Radio Frequency Integrated Circuits (RFIC) Symposium, 2003 IEEE*, pp. 509–512, June 2003.

[4] X. Zhang, A. Molisch, and S.-Y. Kung, "Variable-phase-shift-based rf-baseband codesign for mimo antenna selection," *Signal Processing, IEEE Transactions on*, vol. 53, pp. 4091–4103, Nov 2005.

[5] V. Venkateswaran and A.-J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *Signal Processing, IEEE Transactions on*, vol. 58, pp. 4131–4143, Aug 2010.

[6] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *Wireless Communications, IEEE Transactions on*, vol. 13, pp. 1499–1513, March 2014.

[7] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 8, pp. 831–846, Oct 2014.

[8] J. Nsenga, A. Bourdoux, and F. Horlin, "Mixed analog/digital beamforming for 60 GHz MIMO frequency selective channels," in *Communications (ICC), 2010 IEEE International Conference on*, pp. 1–6, May 2010.

[9] Y. Tsang, A. Poon, and S. Addepalli, "Coding the beams: Improving beamforming training in mmwave communication system," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pp. 1–6, Dec 2011.

[10] J. Wang, Z. Lan, C.-W. Pyo, T. Baykas, C.-S. Sum, M. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada, and S. Kato, "Beam codebook based beamforming protocol for multi-gbps millimeter-wave wpan systems," *Selected Areas in Communications, IEEE Journal on*, vol. 27, pp. 1390–1399, October 2009.

[11] S. Hur, T. Kim, D. Love, J. Krogmeier, T. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *Communications, IEEE Transactions on*, vol. 61, pp. 4391–4403, October 2013.

[12] H. Ghauch, M. Bengtsson, T. Kim, and M. Skoglund, "Subspace estimation and decomposition for hybrid analog-digital millimeter-wave MIMO systems," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*.

[13] T. Kim, D. Love, and B. Clerckx, "Mimo systems with limited rate differential feedback in slowly varying channels," *Communications, IEEE Transactions on*, vol. 59, pp. 1175–1189, April 2011.

[14] T. Dahl, N. Christophersen, and D. Gesbert, "Blind MIMO eigenmode transmission based on the algebraic power method," *Signal Processing, IEEE Transactions on*, vol. 52, pp. 2424–2431, Sept 2004.

[15] T. Dahl, S. Pereira, N. Christophersen, and D. Gesbert, "Intrinsic subspace convergence in TDD MIMO communication," *Signal Processing, IEEE Transactions on*, vol. 55, pp. 2676–2687, June 2007.

[16] D. S. Watkins, *The Matrix Eigenvalue Problem: GR and Krylov Subspace Methods*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1 ed., 2007.

[17] T. Hrycak, S. Das, G. Matz, and H. Feichtinger, "Low complexity equalization for doubly selective channels modeled by a basis expansion," *Signal Processing, IEEE Transactions on*, vol. 58, pp. 5706–5719, Nov 2010.

[18] M. Torlak and O. Ozdemir, "A Krylov subspace approach to blind channel estimation for CDMA systems," in *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*, vol. 1, pp. 674–678 vol.1, Nov 2002.

[19] Y. Saad, "Numerical Methods for Large Eigenvalue Problems," *Manchester University Press*, no. Second Edition, pp. 1–337, 2011.

[20] L. Withers, R. Taylor, and D. Warme, "Echo-MIMO: A two-way channel training method for matched cooperative beamforming," *Signal Processing, IEEE Transactions on*, vol. 56, pp. 4419–4432, Sept 2008.

[21] G. H. Golub and C. F. Van Loan, *Matrix computations (3rd ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.

[22] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1758–1789, 2013.

[23] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, pp. 4311–4322, Nov 2006.

[24] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization," *SIAM Journal on Optimization*, vol. 23, sep 2012.

[25] D. Love and R. Heath, "Equal gain transmission in multiple-input multiple-output wireless systems," *Communications, IEEE Transactions on*, vol. 51, pp. 1102–1110, July 2003.

[26] X. Zheng, Y. Xie, J. Li, and P. Stoica, "Mimo transmit beamforming under uniform elemental power constraint," *Signal Processing, IEEE Transactions on*, vol. 55, pp. 5395–5406, Nov 2007.

[27] D. Baum and H. Bolcskei, "Information-theoretic analysis of MIMO channel sounding," *Information Theory, IEEE Transactions on*, vol. 57, pp. 7555–7577, Nov 2011.

[28] "Spatial channel model for multiple input multiple output (MIMO) simulations," *3GPP TR 25.996 V10.0*, Mar 2011.

[29] J. Salo, G. Del Galdo, J. Salmi, P. Kysti, M. Milojevic, D. Laselva, and C. Schneider, "MATLAB implementation of the 3GPP Spatial Channel Model (3GPP TR 25.996)." On-line, Jan. 2005. http://www.tkk.fi/Units/Radio/scm/.