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 ($\sim$4 to 8 times less).**

***Index Terms*—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 radio frequency (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 the base station (BS) (resp. mobile station (MS) (as shown in Fig. 1). 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 Multiple-input Multiple-output (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 (excluding the receive combiners), [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 channel state information (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

Manuscript received May 30, 2015; revised November 04, 2015, February 15, 2016, and February 24, 2016; accepted February 29, 2016. Date of publication March 23, 2016; date of current version April 14, 2016. The work of T. Kim was supported by the Research Grant Council, Hong Kong under Project No. CityU 11201015. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Nuria Gonzlez-Prelcic.

H. Ghauch, M. Bengtsson, and M. Skoglund are with the School of Electrical Engineering, ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm SE-100 44, Sweden (e-mail: ghauch@kth.se; mats.bengtsson@ee.kth.se; skoglund@kth.se).

T. Kim is with the State Key Laboratory of Millimeter Wave and the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: taejokim@cityu.edu.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2016.2538178

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 *Arnoldi Iteration*, exploits channel reciprocity inherent in Time-Division Duplexing (TDD) MIMO systems to gradually build an orthonormal basis for the corresponding Krylov subspace, and *directly estimates the dominant 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. While we feel that aspects such as complexity, overhead and numerical stability are best left for future works, we do shed light on each of them. 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 $\boldsymbol{A}$, $\mathrm{tr}(\boldsymbol{A})$ 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) of $\boldsymbol{A}$, $\boldsymbol{a}_i$ the ith of column $\boldsymbol{A}$, 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, $\mathrm{diag}(\boldsymbol{x})$ is a diagonal matrix with elements of $\boldsymbol{x}$ on its diagonal, $\Re(x)$ the real part of x, $\sigma_{\max}[\boldsymbol{U}]$ / $\sigma_{\min}[\boldsymbol{U}]$ the maximum/minimum singular value of $\boldsymbol{U}$. Moreover, $\hat{\boldsymbol{U}} = \mathrm{qr}(\boldsymbol{U})$ refers to the semi-unitary matrix returned by the QR algorithm, with $\boldsymbol{U}^\dagger\boldsymbol{U} = \boldsymbol{I}$. Finally, we let $\{n\} \triangleq \{1, \ldots, n\}$, and $\mathcal{S}_{p,q} = \{\boldsymbol{X} \in \mathbb{C}^{p\times q} \mid |\boldsymbol{X}_{ij}| = 1/\sqrt{p},\ \forall (i,k) \in \{p\} \times \{q\}\}$.

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

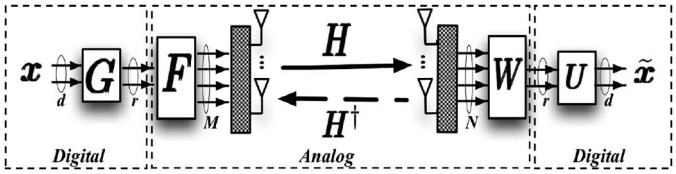


Fig. 1. Hybrid Analog-Digital MIMO system architecture

we assume that $d \le r \le \min(M,N)$). 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 (as shown in Fig. 1), $\boldsymbol{y}^{(r)}$ 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 $E[\boldsymbol{n}^{(r)}\boldsymbol{n}^{(r)^\dagger}] = \sigma^2_{(r)}\boldsymbol{I}_N$. Note that $^{(t)}$ and $^{(r)}$ subscripts/superscripts denote quantities at the BS and MS, respectively. 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). We adopt a *total power constraint* on the effective precoder, i.e., $\|\boldsymbol{F}\boldsymbol{G}\|_F^2 \le d$, a widespread one in the hybrid analog-digital precoding literature [6], [7].

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,

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

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

B. Motivation

Keeping in line with previous work in that area, our aim is to design the precoders and combiners as follows,

[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 $E[\boldsymbol{n}^{(t)}\boldsymbol{n}^{(t)^\dagger}] = \sigma^2_{(t)}\boldsymbol{I}_M$.

$$(\boldsymbol{F}^\star, \boldsymbol{G}^\star) = \begin{cases} \min\limits_{\boldsymbol{F},\,\boldsymbol{G}} \ \|\boldsymbol{\Gamma}_1 - \boldsymbol{F}\boldsymbol{G}\|_F^2 \\ \text{s. t. } \|\boldsymbol{F}\boldsymbol{G}\|_F^2 \le d, \ \ \boldsymbol{F} \in \mathcal{S}_{M,d} \end{cases}$$

$$(\boldsymbol{W}^\star, \boldsymbol{U}^\star) = \begin{cases} \min\limits_{\boldsymbol{W},\,\boldsymbol{U}} \ \|\boldsymbol{\Phi}_1 - \boldsymbol{W}\boldsymbol{U}\|_F^2 \\ \text{s. t. } \ \ \boldsymbol{W} \in \mathcal{S}_{N,d} \end{cases} \tag{4}$$

The latter design criterion has been quite prevalent in earlier works relating to the hybrid analog-digital architecture, and applied rather successfully in [6], [7], [13], [14]. After a series of approximations to the mutual information in [6], it was shown that the optimal precoders, $\boldsymbol{F}, \boldsymbol{G}$, are formulated in exactly the same fashion as above (though their formulation did not include receive combining).

Moreover, we use the following expression as a performance metric (i.e., the "user-rate" corresponding to a given choice of precoders and combiners),

$$R = \log_2 \left| \boldsymbol{I}_d + \boldsymbol{H}_e \boldsymbol{H}_e^\dagger (\sigma_{(r)}^2 \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{W} \boldsymbol{U})^{-1} \right| \tag{5}$$

where $\boldsymbol{H}_e = \boldsymbol{U}^\dagger \boldsymbol{W}^\dagger \boldsymbol{H} \boldsymbol{F} \boldsymbol{G}$, $\frac{1}{\sigma_{(r)}^2} \triangleq \text{SNR}$ is the signal-to-noise ratio. Moreover we assume, for simplicity, that uniform power allocation is performed (no waterfilling), keeping in mind that a power allocation matrix $\boldsymbol{\Lambda}$ can be easily incorporated in the expression. Although not directly optimized, the above expression was used in [6], within the context of hybrid analog-digital precoding. As we will discuss below, the value of the expression in (5) 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 $\boldsymbol{H}$.

In a nutshell, (4) boils down to finding $\boldsymbol{F}\boldsymbol{G}$ (resp. $\boldsymbol{W}\boldsymbol{U}$) that "best" approximate $\boldsymbol{\Gamma}_1$ (resp. $\boldsymbol{\Phi}_1$). Moreover, if there exists optimal precoders and combiners that make the distances in (4) zero, then they must satisfy

$$\boldsymbol{F}^\star \boldsymbol{G}^\star = \boldsymbol{\Gamma}_1, \quad \boldsymbol{W}^\star \boldsymbol{U}^\star = \boldsymbol{\Phi}_1.$$

We denote by $R^\star$ the resulting "user-rate" that is obtained by plugging in the above precoders/combiners in (5). Then $R^\star$ can be expressed as,

$$R^\star \triangleq R(\boldsymbol{F}^\star, \boldsymbol{G}^\star, \boldsymbol{W}^\star, \boldsymbol{U}^\star) = \log_2 \left| \boldsymbol{I}_d + \text{SNR}\ \boldsymbol{\Sigma}_1^2 \right| \tag{6}$$

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. We underline the fact that R in (5) depends on the subspace spanned by the precoders/combiners, rather than the Euclidean distance between the right/left dominant subspace and the precoder/combiner, i.e., (4). However, optimizing metrics that involve span or chordal distances, is not straightforward. We thus emphasize that attempts at directly maximizing R in (5) are outside the scope of this work: rather, the focus is put on proposing mechanisms for subspace estimation and decomposition, and analyzing their performance.

Moreover, 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 $\|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{F}\boldsymbol{G}\|_F^2$ and $\|\tilde{\boldsymbol{\Phi}}_1 - \boldsymbol{W}\boldsymbol{U}\|_F^2$ (while taking into consideration the constraints inherent to the hybrid analog-digital architecture).

III. Eigenvalue Algorithms and Subspace Estimation

A. Subspace 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,

$$\boldsymbol{p}_{l+1} = \boldsymbol{H}^\dagger \boldsymbol{H} \boldsymbol{p}_l / \|\boldsymbol{H}^\dagger \boldsymbol{H} \boldsymbol{p}_l\|_2; \ \boldsymbol{q}_{l+1} = \boldsymbol{H}\boldsymbol{p}_l \tag{7}$$

It was noted in [15] 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*. The authors of [15] 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 [15], [16]).

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 dominant 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 (because the latter estimates the dominant 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 [15].

TABLE I
ARNOLDI PROCEDURE

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 $t_{k,l} = \boldsymbol{q}_k^\dagger \boldsymbol{p}_l,\ k = 1, \ldots, l$
  2. $\boldsymbol{r}_l = \boldsymbol{p}_l - \sum_{k=1}^{l} t_{k,l}\boldsymbol{q}_k$
  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

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 [17], their use in the area of channel/subspace estimation is limited to equalization for doubly selective OFDM channels [18], and channel estimation in CDMA systems [19]. Algorithms falling into that category iteratively build a *basis for the Krylov subspace*, $\mathcal{K}^m = span\{\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}, \ldots, \boldsymbol{A}^{m-1}\boldsymbol{x}\}$, 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 [20]). 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

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

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 $\boldsymbol{Q}_m^\dagger \boldsymbol{A} \boldsymbol{Q}_m = \boldsymbol{T}_m$. We then say that the eigenvalues/eigenvectors of $\boldsymbol{T}_m$ are called *Ritz eigenvalues/eigenvectors*, and approximate the eigenvalues/eigenvectors of $\boldsymbol{A}$. The main idea behind processes such as the Arnoldi (and Lanczos) is to find the dominant eigenpairs of $\boldsymbol{A}$, by finding the eigenpairs of $\boldsymbol{T}_m$.

We note that the Arnoldi algorithm is a generalization of the Lanczos algorithm for the non-symmetric case, i.e., the latter is specifically tailored for cases where $\boldsymbol{A} \succeq \boldsymbol{0}$ (this is clearly the case in this work, since $\boldsymbol{A} = \boldsymbol{H}^\dagger \boldsymbol{H}$). This being said, the reason for not using the Lanczos iteration is that in practice, noise that is inherent to the echoing process, makes the Lanczos algorithm not applicable: namely, the requirement that $\boldsymbol{T}_m$ is tridiagonal, is violated.

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 A-F like fashion, over the uplink (UL) channel (following the process proposed in [21], and detailed in Sect. III-A), i.e.,

TABLE II
SUBSPACE ESTIMATION USING ARNOLDI ITERATION (SE-ARN)

procedure $\tilde{\boldsymbol{\Gamma}}_1, \tilde{\boldsymbol{\Sigma}}_1 =$ 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 $t_{k,l} = \boldsymbol{q}_k^\dagger \boldsymbol{p}_l \quad ,\forall\ k = 1, \ldots, l$
    2.b $\boldsymbol{r}_l = \boldsymbol{p}_l - \sum_{k=1}^{l} \boldsymbol{q}_k t_{k,l}$
    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 $\tilde{\boldsymbol{\Gamma}}_1$
  $\boldsymbol{T}_m = \tilde{\boldsymbol{\Theta}} \tilde{\boldsymbol{\Lambda}} \tilde{\boldsymbol{\Theta}}^{-1}$
  $\tilde{\boldsymbol{\Gamma}}_1 = \mathrm{qr}(\boldsymbol{Q}_m \tilde{\boldsymbol{\Theta}}_{1:d})$
  $[\tilde{\boldsymbol{\Sigma}}_1]_{i,i} = \sqrt{|[\tilde{\boldsymbol{\Lambda}}]_{i,i}|}, \forall\ i$
end procedure

$$\begin{aligned} DL:&\ \ \boldsymbol{s}_l = \boldsymbol{H}\boldsymbol{q}_l + \boldsymbol{w}_l^{(r)} \\ UL:&\ \ \boldsymbol{p}_l = \boldsymbol{H}^\dagger \boldsymbol{s}_l + \boldsymbol{w}_l^{(t)} = \boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l + \boldsymbol{H}^\dagger \boldsymbol{w}_l^{(r)} + \boldsymbol{w}_l^{(t)} \\ &\ \ \ \ = \boldsymbol{H}^\dagger \boldsymbol{H}\boldsymbol{q}_l + \tilde{\boldsymbol{w}}_l \end{aligned} \tag{8}$$

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 (8). 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,

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

$$\tilde{\boldsymbol{W}}_m = [\tilde{\boldsymbol{w}}_1, \ldots, \tilde{\boldsymbol{w}}_m],\ \tilde{\boldsymbol{E}}_m = [\boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m]_{SL} \tag{9}$$

where $\boldsymbol{r}_l$ is given in Step 2.b (Table II). Note that similarly to the conventional Arnoldi Iteration, $\tilde{\boldsymbol{T}}_m$ is an the upper Hessenberg matrix. It then follows from the above definitions that

$$\boldsymbol{T}_m = \tilde{\boldsymbol{T}}_m + [\boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m]_U. \tag{10}$$

This can be easily verified by plugging in Step 1.b into 2.a in Table II.

At the output of the SE-ARN procedure, the dominant eigenpairs of $\boldsymbol{H}^\dagger \boldsymbol{H}$ are approximated by those 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 $\tilde{\boldsymbol{\Theta}}$ is the (possibly non-orthonormal) set of eigenvectors. Then, it can easily be shown that $\tilde{\boldsymbol{\Gamma}}_1 = \mathrm{qr}(\boldsymbol{Q}_m[\tilde{\boldsymbol{\Theta}}]_{1:d})$ are the Ritz eigenvectors of $\boldsymbol{H}^\dagger \boldsymbol{H}$, where $[\tilde{\boldsymbol{\Theta}}]_{1:d}$ has as columns the eigenvectors of $\boldsymbol{T}_m$ associated with the d largest eigenvalues (in

magnitude).[3] Moreover, $\tilde{\boldsymbol{\Sigma}}_1$, the Ritz eigenvalues of $\boldsymbol{H}^\dagger\boldsymbol{H}$, come for free once the Ritz eigenvectors are obtained (Table II). Note that the latter procedure results in the BS obtaining $\tilde{\boldsymbol{\Gamma}}_1$, and consequently $\tilde{\boldsymbol{\Sigma}}_1$, 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,

$(P1)$:

$$\boldsymbol{Q}_m^\dagger \boldsymbol{A}\boldsymbol{Q}_m = \tilde{\boldsymbol{T}}_m - \tilde{\boldsymbol{E}}_m \triangleq \boldsymbol{C}_m, \tag{11}$$

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,

$$\|\boldsymbol{A}\boldsymbol{\theta}_i^{(m)} - \lambda_i^{(m)}\boldsymbol{\theta}_i^{(m)}\|_2^2 \leq c_m^{(i)} + \|\boldsymbol{I}_M - \boldsymbol{Q}_m\boldsymbol{Q}_m^\dagger\|_F^2\|\tilde{\boldsymbol{W}}_m\|_F^2, \tag{12}$$

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 A. ■

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 [20, Sect. 6.2]. Lemma 1 establishes the fact that each eigenpair $(\lambda_i^{(m)}, \boldsymbol{s}_i^{(m)})$ of $\boldsymbol{C}_m$, is associated with one eigenpair $(\lambda_i^{(m)}, \boldsymbol{\theta}_i^{(m)})$ of $\boldsymbol{A}$.[4]

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* $\boldsymbol{C}_m$.

For that purpose, we first show the following,

$$\begin{aligned}\boldsymbol{C}_m + \boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m &= \left(\tilde{\boldsymbol{T}}_m - \tilde{\boldsymbol{E}}_m\right) + \boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m \\ &= \tilde{\boldsymbol{T}}_m + \left(\boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m - [\boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m]_{SL}\right) \\ &= \tilde{\boldsymbol{T}}_m + \left[\boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m\right]_U \triangleq \boldsymbol{T}_m \end{aligned} \tag{13}$$

where the first equality follows from the definition of $\boldsymbol{C}_m$, and the last one from (10). Thus $\boldsymbol{C}_m$ can be viewed as the matrix in question, and $\boldsymbol{P}_m \triangleq \boldsymbol{Q}_m^\dagger\tilde{\boldsymbol{W}}_m$ a perturbation matrix. We then apply the Bauer-Fike Theorem [22, Th. 7.2.2] [22] 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 $\boldsymbol{C}_m$.

Proof: Refer to Appendix B ■

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 those 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, for the eigenvalues, is upper bounded by the magnitude of the perturbation itself. We note that the relevant "error-metric" here is the distance between the true subspace $\boldsymbol{\Gamma}_1$, and estimated subspace $\tilde{\boldsymbol{\Gamma}}_1 \propto \boldsymbol{Q}_m\tilde{\boldsymbol{\Theta}}_{1:d}$ (Table II). This does suggest that the estimation error is dependent on $\tilde{\boldsymbol{\Theta}}_{1:d}$, the eigenvectors of $\boldsymbol{T}_m$. However, performing a similar sensitivity analysis on the eigenvectors is much more involved, since the sensitivity of eigenvectors generally depends on the clustering of eigenvalues.

IV. Hybrid Analog–Digital Precoding for mmWave MIMO 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. Thus, $\tilde{\boldsymbol{\Gamma}}_1$ has to be decomposed into $\boldsymbol{F}\boldsymbol{G}$ (hence the term *Subspace Decomposition (SD)*), as follows,

$$\begin{cases} \min\limits_{\boldsymbol{F},\,\boldsymbol{G}} & h_0(\boldsymbol{F},\boldsymbol{G}) = \|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{F}\boldsymbol{G}\|_F^2 \\ \text{s. t.} & h_1(\boldsymbol{F},\boldsymbol{G}) = \|\boldsymbol{F}\boldsymbol{G}\|_F^2 \leq d \\ & \boldsymbol{F} \in \mathcal{S}_{M,d} \end{cases} \tag{14}$$

[3]Note that, to be exact, the Ritz eigenvectors do not contain any estimation noise. That being said, we stick to this nomenclature, with a slight abuse of definition.

[4]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 $m \ll M$.

We underline the fact that the authors in [6] arrived to the same formulation as (14), and proposed a variation on the well-known Orthogonal Matching Pursuit (OMP), to tackle it. The same framework was recently extended in [14] to relax the need for dictionaries based on the array response matrix. An alternate decomposition was proposed by [23], where the optimization metric is the user rate. Both works were published after the initial submission of our paper.

Within the context of hybrid precoding, the authors in [4] showed that there exists (non-unique) $\boldsymbol{F} \in \mathcal{S}_{M,r}, \boldsymbol{g} \in \mathbb{C}^{r\times 1}$ such that $\tilde{\boldsymbol{\Gamma}}_1 = \boldsymbol{F}\boldsymbol{g}$, if and only if $r \geq 2$. This was extended in [14] where it was shown that there exists $\boldsymbol{F} \in \mathcal{S}_{M,r}, \boldsymbol{G} \in \mathbb{C}^{r\times d}$ such that $\tilde{\boldsymbol{\Gamma}}_1 = \boldsymbol{F}\boldsymbol{G}$, if $r \geq 2d$. We note that for such cases, the cost function in (14) is zero, and we refer to such cases as *optimal decomposition* -whose performance we evaluate in the numerical results section: although the aforementioned schemes use all the available RF chains for the decomposition (and our decomposition uses a subset of the RF chains), the sum-rate performance is actually the same.

To a certain extent, (14) is reminiscent of formulations arising from areas such as blind source separation, (sparse) dictionary learning, and vector quantization [24], [25]. Though there is a battery of algorithms and techniques that have been developed to tackle such problems, the additional hardware constraint on $\boldsymbol{F}$, i.e. $\boldsymbol{F} \in \mathcal{S}_{M,r}$ 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 determine an approximate solution.

1) Block Co-Oordinate 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 (14) 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 [26]), and from the hardware constraint on $\boldsymbol{F}$. We will show that a BCD approach implicitly enforces the power constraint in (14), 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 1: 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

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

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: The proof is straightforward variation on previous results such as [4]. ■

TABLE III
BLOCK COORDINATE DESCENT FOR SUBSPACE DECOMPOSITION (BCD-SD)

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

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 (14), 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,

$$(J1)\ \ \boldsymbol{F}_{k+1} \triangleq \min_{\boldsymbol{F}}\ h_0(\boldsymbol{F}) = \|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{F}\boldsymbol{G}_k\|_F^2$$
$$(J2)\ \ \boldsymbol{G}_{k+1} \triangleq \min_{\boldsymbol{G}}\ h_0(\boldsymbol{G}) = \|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{F}_{k+1}\boldsymbol{G}\|_F^2$$

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

$$\boldsymbol{F}_{k+1} = \tilde{\boldsymbol{\Gamma}}_1 \boldsymbol{G}_k^\dagger \left(\boldsymbol{G}_k \boldsymbol{G}_k^\dagger\right)^{-1} \tag{15}$$

$$\boldsymbol{G}_{k+1} = \left(\boldsymbol{F}_{k+1}^\dagger \boldsymbol{F}_{k+1}\right)^{-1} \boldsymbol{F}_{k+1}^\dagger \tilde{\boldsymbol{\Gamma}}_1 \tag{16}$$

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 (16) is invertible, almost surely: in fact, our numerical results show that the incurred performance loss is quite negligible.

Moreover, note that the solution in (15) does not necessarily satisfy the hardware constraint on $\boldsymbol{F}$. Thus, the result of Proposition 1 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 (14), we plug (16) into the following (dropping all subscripts for simplicity),

$$\begin{aligned}\|\boldsymbol{F}\boldsymbol{G}\|_F^2 &= \operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_1^\dagger \boldsymbol{F} \underbrace{(\boldsymbol{F}^\dagger\boldsymbol{F})^{-1}\boldsymbol{F}^\dagger\boldsymbol{F}}_{=\boldsymbol{I}_d}(\boldsymbol{F}^\dagger\boldsymbol{F})^{-1}\boldsymbol{F}^\dagger\tilde{\boldsymbol{\Gamma}}_1\right) \\ &\leq \operatorname{tr}\left((\boldsymbol{F}^\dagger\boldsymbol{F})^{-1}\boldsymbol{F}^\dagger\boldsymbol{F}\right)\operatorname{tr}\left(\tilde{\boldsymbol{\Gamma}}_1\tilde{\boldsymbol{\Gamma}}_1^\dagger\right) = d\end{aligned} \tag{17}$$

where we assumed that $\|\tilde{\boldsymbol{\Gamma}}_1\|_F^2 = 1$ w.l.o.g., and used the fact that $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) \leq \operatorname{tr}(\boldsymbol{A})\operatorname{tr}(\boldsymbol{B})$ for $\boldsymbol{A}, \boldsymbol{B} \succeq \boldsymbol{0}$. Note that the above relation holds for any arbitrary full-rank $\boldsymbol{F}$, and thus, the power constraint is satisfied even after applying the projection step.

The above shows that if BCD is used, then the power constraint in (14) is always enforced. The corresponding method is termed Block Coordinate Descent for Subspace Decomposition (BCD-SD), and is shown in Table III.

Remark 1: 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. Nevertheless, despite the fact that monotonic convergence of BCD-SD cannot be showed analytically, our simulations indicate that the latter is indeed the case, under normal operating conditions.

Remark 2: It can be easily verified that the optimal $\boldsymbol{F}^\star, \boldsymbol{G}^\star$ that maximize R in (5) are such that $\|\boldsymbol{F}^\star\boldsymbol{G}^\star\| = d$. Though the optimal solution to (14) is not invariant to scaling, as far as the performance metric in (5) 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 (e.g., our proposed mechanism in Table II) relies on the BS being able to send any vector $\boldsymbol{q}_l$, to be echoed back by the MS. 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 $d = 1$, (14) reduces to the problem below,

Lemma 3: Consider the single dimension SD problem,

$$\begin{cases} \min\limits_{\boldsymbol{f},\, g} h_o(\boldsymbol{f},\, g) = \|\boldsymbol{f}\|_2^2\, g^2 - 2g\Re(\boldsymbol{f}^\dagger \tilde{\boldsymbol{\gamma}}_1) \\ \text{s. t. } [\boldsymbol{f}]_i = 1/\sqrt{M}\; e^{j\phi_i}, \forall i \end{cases} \tag{18}$$

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 C ■

Similarly to (17), 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,

$$[\boldsymbol{e}]_i = |r_i - \|\tilde{\boldsymbol{\gamma}}_1\|_1/M| e^{j\theta_i},\; \forall i \in \{M\}. \tag{19}$$

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 [27], [28], in terms of the optimal phases. We recall that a similar hybrid beamforming setup was considered in [4] where the authors optimize $u, \boldsymbol{w}, \boldsymbol{f}, g$, to maximize the SNR as well as the spectral efficiency. Although our formulation optimizes the same quantities, the optimization metric we consider, the subspace distance, is different.

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 (14)) 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,

$$\begin{aligned} \tilde{\boldsymbol{\Gamma}}_1 &= [\tilde{\boldsymbol{\gamma}}_1, \ldots, \tilde{\boldsymbol{\gamma}}_d] \approx \left[g_1^\star \boldsymbol{f}_1^\star, \cdots, g_d^\star \boldsymbol{f}_d^\star\right] \\ &\triangleq (1/\sqrt{M})\; [\Pi_{\mathcal{S}}[\tilde{\boldsymbol{\gamma}}_1], \cdots, \Pi_{\mathcal{S}}[\tilde{\boldsymbol{\gamma}}_d]]\; \mathrm{diag}(\|\tilde{\boldsymbol{\gamma}}_1\|_1, \cdots, \|\tilde{\boldsymbol{\gamma}}_d\|_1) \end{aligned} \tag{20}$$

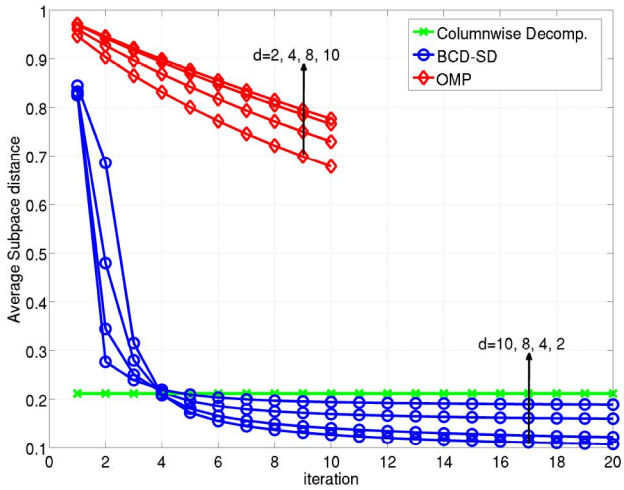


Fig. 2. Average subspace distance $\|\tilde{\boldsymbol{\Gamma}}_1 - \boldsymbol{F}\boldsymbol{G}\|_F^2$, for our proposed method and OMP

3) Numerical Results: As mentioned earlier, (14) 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 curves are averaged over 500 random realizations of $\boldsymbol{\Gamma}_1$ (the latter are random unitary matrices). Moreover, we follow the same setup for OMP as that of [6], namely, that the dictionary is designed based on the array response vectors (of size 256). The reason for the large performance gap in Fig. 2 is that BCD-SD attempts to find a locally optimal solution to (14) (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 a time, whereas our proposed method runs until reaching a stable point. With that in mind, although OMP might perform better in terms of approximating the span of $\boldsymbol{\Gamma}_1$, it is challenging to measure and optimize such metrics in practice. Moreover, we recall that in its original formulation in [6] OMP is indeed formulated to solve the problem at hand (i.e. (14)), and thus the comparison seems fair. Interestingly, despite its extreme simplicity, the column-wise decomposition in (20) 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 (8). 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 (19). 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, (8), 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,

$$\boldsymbol{p}_l = \boldsymbol{F}_l^\dagger\boldsymbol{H}^\dagger\boldsymbol{W}_l\boldsymbol{W}_l^\dagger\boldsymbol{H}(\boldsymbol{q}_l - \boldsymbol{e}_l) \tag{21}$$

Note that the above process is possible using the hybrid analog-digital architecture. Since noise is present in any uplink/downlink transmission, for clarity in what follows, we drop the noise-related terms from all equations. Needless to say, their effect is accounted for in the numerical results. It is clear from (21) 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 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 (21)).

The above impairments are a byproduct of shifting the burden of digital precoding, 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[5], 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,

$$[\boldsymbol{W}_{l,1}, \ldots, \boldsymbol{W}_{l,K_r}] = \boldsymbol{D}_r, \tag{22}$$

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,

$$\tilde{\boldsymbol{s}}_l = \left(\sum_{k=1}^{K_r}\boldsymbol{W}_{l,k}\boldsymbol{W}_{l,k}^\dagger\right)\boldsymbol{H}(d\tilde{\boldsymbol{f}}_l\tilde{g}_l) = d\boldsymbol{H}\tilde{\boldsymbol{f}}_l\tilde{g}_l, \tag{23}$$

where equality follows from our earlier definition of $\{\boldsymbol{W}_{l,k}\}_k$ in (22). 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, $[\boldsymbol{F}_{l,1}, \ldots, \boldsymbol{F}_{l,K_t}] = \boldsymbol{D}_t$. The process for the UL is also shown in Table IV. We combine its steps to rewrite $\boldsymbol{p}_l$ as,

$$\boldsymbol{p}_l = \left(\sum_{m=1}^{K_t}\boldsymbol{F}_{l,m}\boldsymbol{F}_{l,m}^\dagger\right)\boldsymbol{H}^\dagger(d\tilde{\boldsymbol{w}}_l\tilde{u}_l) = d\boldsymbol{H}^\dagger\tilde{\boldsymbol{w}}_l\tilde{u}_l \tag{24}$$

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

$$\begin{aligned}\boldsymbol{p}_l &= d\boldsymbol{H}^\dagger\tilde{\boldsymbol{w}}_l\tilde{u}_l = d\boldsymbol{H}^\dagger\left(\tilde{\boldsymbol{s}}_l - \boldsymbol{e}_l^{(r)}\right) = d\boldsymbol{H}^\dagger\left(d\boldsymbol{H}\tilde{\boldsymbol{f}}_l\tilde{g}_l - \boldsymbol{e}_l^{(r)}\right)\\ &= d^2\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{q}_l - d^2\boldsymbol{H}^\dagger\boldsymbol{H}\boldsymbol{e}_l^{(t)} - d\boldsymbol{H}^\dagger\boldsymbol{e}_l^{(r)}\end{aligned} \tag{25}$$

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 (21). We can clearly see that

TABLE IV
REPETITION-AIDED (RAID) ECHOING

// 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
$\tilde{\boldsymbol{s}}_l = \tilde{\boldsymbol{w}}_l\tilde{u}_l + \boldsymbol{e}_l^{(r)}$
$\boldsymbol{z}_{l,m} = \boldsymbol{F}_{l,m}^\dagger\boldsymbol{H}^\dagger(d\tilde{\boldsymbol{w}}_l\tilde{u}_l),\ \forall m \in \{K_t\}$
$\boldsymbol{p}_l = \sum_{m=1}^{K_t}\boldsymbol{F}_{l,m}\boldsymbol{z}_{l,m}$

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

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

thereby resulting in an array gain factor of d. Moreover, since we know from (17) 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.

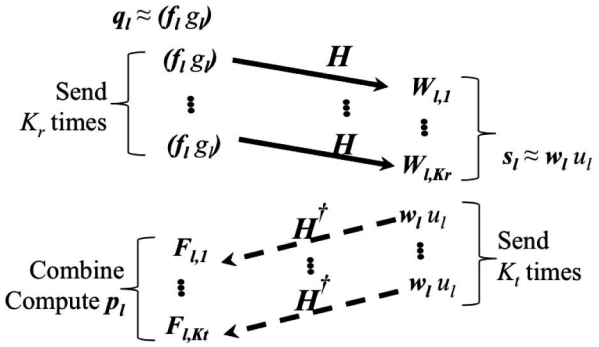


Fig. 3. Repetition-aided (RAID) echoing for the hybrid analog-digital architecture

impairments originating from processing the received signals with both $\boldsymbol{W}_l$ and $\boldsymbol{F}_l$, have completely been suppressed. In (25), $\boldsymbol{p}_l$ indeed is the desired estimate, i.e., $\boldsymbol{H}^\dagger \boldsymbol{H} \boldsymbol{q}_l$, corrupted by distortions emanating from the BS-side decomposition, $\boldsymbol{e}_l^{(t)}$, and 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 3: 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 (23), at a cost of K_r channel uses (the same holds for the UL, as seen from (24)).

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,

$$\begin{cases} \min\limits_{\tilde{\boldsymbol{D}}_t} \| \frac{1}{M} \boldsymbol{I}_M - \tilde{\boldsymbol{D}}_t \tilde{\boldsymbol{D}}_t^\dagger \|_F^2 \\ \text{s. t. } \tilde{\boldsymbol{D}}_t \in \mathcal{S}_{M,\, \eta M}. \end{cases} \tag{26}$$

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.

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

// Estimate $\tilde{\boldsymbol{\Gamma}}_1$ *and* $\tilde{\boldsymbol{\Phi}}_1$
$\tilde{\boldsymbol{\Gamma}}_1, \tilde{\boldsymbol{\Sigma}}_1 =$ 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 $(\tilde{\boldsymbol{\Gamma}}_1, \rho)$
$[\boldsymbol{W}, \boldsymbol{U}]$ = BCD-SD $(\tilde{\boldsymbol{\Phi}}_1, \rho)$
Perform waterfilling on $\tilde{\boldsymbol{\Sigma}}_1$

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* (Table IV. 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. {We highlight a desirable feature for the SED algorithm: the subspace estimation mechanism is totally decoupled from the subspace decomposition part, and thus any of the latter parts can be substituted, if desired.

Note that previously proposed algorithms within this context such as the PM and TQR in [15], are no longer applicable here: indeed both rely on the MS being able to amplify-and-forward its received signal at the antennas - 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).

Operationally, the proposed MTQR algorithm is the same as the Two-way QR (TQR) in [15], whereby $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Phi}_1$ are obtained iteratively: as $I \to \infty$, $\boldsymbol{X} \to \boldsymbol{\Gamma}_1$ (at BS) and $\boldsymbol{Y} \to \boldsymbol{\Phi}_1$ (at MS). At each iteration of the algorithm, the BS sends $\boldsymbol{X}$ in the downlink, and the QR algorithm is applied to the received signal. Then, the resulting signal is sent by the MS in the uplink, and the QR algorithm is applied at the BS to form $\boldsymbol{Y}$. While TQR assumes fully digital MIMO transmission, our contribution is to apply the RAID scheme, to make the transmission compatible with the hybrid analog-digital systems.

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 Hybrid Analog-Digital Architecture

for $l = 1, 2, \ldots, I$ **do**
  // Decompose each column of $\boldsymbol{X}$
  $[\boldsymbol{X}]_n \approx \tilde{\boldsymbol{f}}_n \tilde{g}_n, \forall n \in d$ (using Lemma 3)
  $\tilde{\boldsymbol{X}} = [\tilde{\boldsymbol{f}}_1 \tilde{g}_1 \cdots , \tilde{\boldsymbol{f}}_d \tilde{g}_d]$
  // Send $\tilde{\boldsymbol{X}}$ *in DL, one column at a time*
  $\boldsymbol{T}_k = \boldsymbol{W}_k^{\dagger} \boldsymbol{H} \tilde{\boldsymbol{X}}, \forall k \in \{K_r\}$
  $\boldsymbol{Y} = \sum_{k=1}^{K_r} \boldsymbol{W}_k \boldsymbol{T}_k$; $\boldsymbol{Y} = \text{qr}(\boldsymbol{Y})$
  // Decompose of $\boldsymbol{Y}$
  $[\boldsymbol{Y}]_n \approx \tilde{\boldsymbol{w}}_n \tilde{u}_n, \forall n \in d$ (using Lemma 3)
  $\tilde{\boldsymbol{Y}} = [\tilde{\boldsymbol{w}}_1 \tilde{u}_1 \cdots , \tilde{\boldsymbol{w}}_d \tilde{u}_d]$
  // Send $\tilde{\boldsymbol{Y}}$ *in UL, one column at a time*
  $\boldsymbol{S}_k = \boldsymbol{F}_k^{\dagger} \boldsymbol{H}^{\dagger} \tilde{\boldsymbol{Y}}, \forall k \in \{K_t\}$
  $\boldsymbol{Z} = \sum_{k=1}^{K_t} \boldsymbol{F}_k \boldsymbol{S}_k$; $\boldsymbol{X} = \text{qr}(\boldsymbol{Z})$
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 dominant eigenvalues of $\boldsymbol{C}_m$, and consequently of $\boldsymbol{A}$ (Sect. III-B).

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

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

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

Proof: Refer to Appendix D ■

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

$$\Omega_{SED} = 2m \, \frac{M+N}{r}, \tag{27}$$

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,

$$\Omega_{MTQR} = dI \, \frac{M+N}{r} \tag{28}$$

It should be emphasized here that our main focus in this work is to investigate the principle of subspace estimation employing numerical techniques, and through simulations describe the performance gain that can be expected by taking on such an approach. Hence, our major concern is not to investigate a stable and low-complexity technique that can be readily implemented in practice. We will, however, provide suggestions on what can be done to enhance the stability of the devised schemes, while admitting that many of the problems connected with practical implementation of the proposed method are subject to further study. Generally, it is known that the Arnoldi (and Lanczos) algorithm may suffer from numerical stability issues. Though analytically speaking, the basis $\boldsymbol{Q}_m$ is easily shown to be orthonormal, in practice, however, errors resulting from floating-point operations lead to a loss in orthogonality (the extent to which it happens is dependent on the application) [20, Sec. 7.3]. Moreover, for our algorithm, noise inherent to the echoing process will further amplify this effect. One of the widely adopted fixes for this matter is the Implicitly Restarted Arnoldi algorithm [20, Sec. 7.3]. We did experiment with such an algorithm, and though it does enhance the numerical stability of the algorithm, the resulting overhead is increased by a large factor. This issue is critical for the SED algorithm (that employs the RAID echoing), since it renders real-world implementation quite impractical. Moreover, there are many problems connected with practical implementations of the Restarted Arnoldi method, that are subject to further study. Other methods that might enhance the stability the Arnoldi Iteration, such as deflation techniques, have been reported in [29].

F. Discussion

We have presented an approach to maximizing the metric R defined in (5). 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 [30]. 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 the above in mind, we use the following, as our performance metric in the simulations,

$$\tilde{R} = \log_2 \left| \boldsymbol{I}_d + \frac{1}{\sigma_{(r)}^2} \boldsymbol{U}^{\dagger}\boldsymbol{W}^{\dagger}\boldsymbol{H}\boldsymbol{F}\boldsymbol{G}\boldsymbol{G}^{\dagger}\boldsymbol{F}^{\dagger}\boldsymbol{H}^{\dagger}\boldsymbol{W}\boldsymbol{U}(\boldsymbol{U}^{\dagger}\boldsymbol{W}^{\dagger}\boldsymbol{W}\boldsymbol{U})^{-1} \right|. \tag{29}$$

In that sense, $\tilde{R}$ is the 'user rate' that is based on the actual channel $\boldsymbol{H}$, and the precoders/combiners that are in turn designed based on the estimated channel.

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,

$$\boldsymbol{H} = \sqrt{\frac{MN}{L}} \sum_{i=1}^{L} \beta_i \, \boldsymbol{a}_r \left(\chi_i^{(r)}\right) \boldsymbol{a}_t^{\dagger} \left(\chi_i^{(t)}\right) \tag{30}$$

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 performance of the algorithms against the "optimal performance", $R^{\star}$ in (6), 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 4: 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 $\Omega_o = 144$ channel uses, respectively. It can be seen from Fig. 4 that both

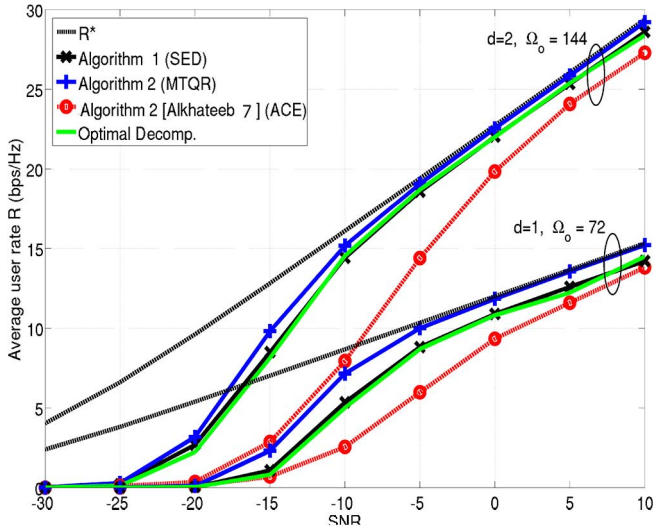


Fig. 4. Average sum-rate of proposed schemes ($M = 128, N = 64, d = 2, L = 3, m = 6$)

proposed schemes exhibit relatively similar performances, that are in turn very *close to the optimal performance bound* $R^{\star}$ (especially above -10 dB). 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). We also evaluate the so-called optimal decomposition schemes [4] [14] that can exactly decompose $\boldsymbol{\Gamma}_1$ into $\boldsymbol{FG}$ (discussed in Sec. IV). Thus, the curves labeled 'Optimal Decomp.' refer to the case where the optimal decomposition is used in conjunction with SED. Fig 4 clearly reveals that the ability to optimally decompose the estimated subspaces does not bring about additional gains. We note that the tiny mismatch between 'Optimal Decomp.' and Algorithm 1 is due to simulation resolution.

We attempt to shed light on the stability of the proposed algorithms, as the number of paths in the mmWave channel, L, increases (where we set $M = 64, N = 32, d = 2, m = 6$). For clarity we restrict the result to the low SNR regime. Though a degradation in the performance of both algorithms is expected, as L increases, Fig. 5 clearly indicates that the latter degradation is not quite significant. Though not visible here, our simulations show that this degradation is not present in the medium-to-high SNR region. As expected, this technique is best used for channels with a few paths, e.g., mmWave channels.

We investigate the performance of both SED and MTQR in terms of average subspace angle, $\theta = \mathbb{E}[\alpha(\boldsymbol{\Gamma}_1, \tilde{\boldsymbol{\Gamma}}_1)]$ where $\alpha(\boldsymbol{\Gamma}_1, \tilde{\boldsymbol{\Gamma}}_1)$ (radians) is defined as the subspace angle between $\boldsymbol{\Gamma}_1$ and $\tilde{\boldsymbol{\Gamma}}_1$ (implemented by computing the principal angles of the latter subspaces). As shown in Fig. 6, both schemes exhibit a similar behavior of better estimation accuracy, as the SNR increases.

Remark 5: Though the performance of Algorithm 2 seems to be better, Fig. 4–6 both suggest that this gap is quite narrow.

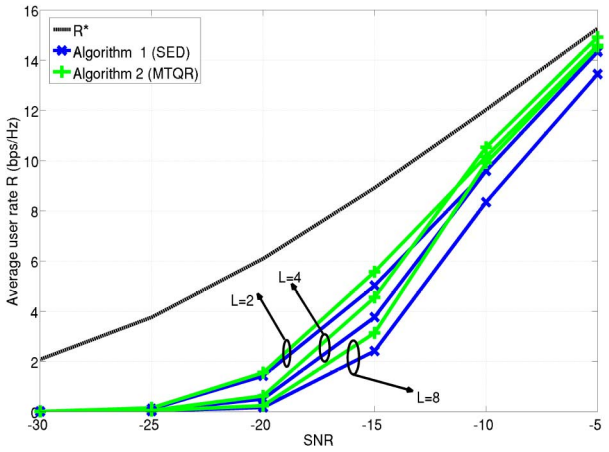


Fig. 5. Effect of number of paths L, on the average user rate ($M = 64, N = 32, d = 2, m = 6$)

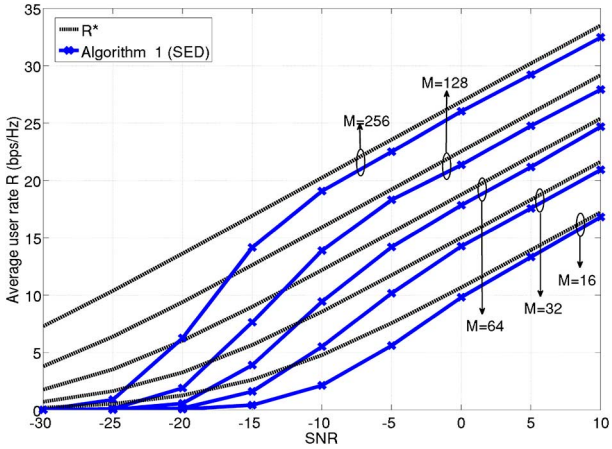


Fig. 7. Average user-rate for different M, N ($N = M/2, d = 2, L = 4, m = 6, \Omega_o = 144$)

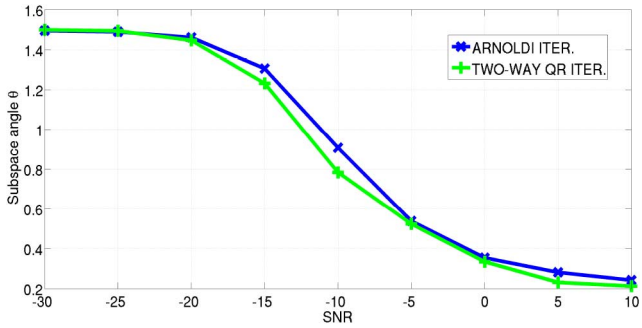


Fig. 6 Average subspace angle ($M = 64, N = 32, d = 3, L = 4, m = 6$)

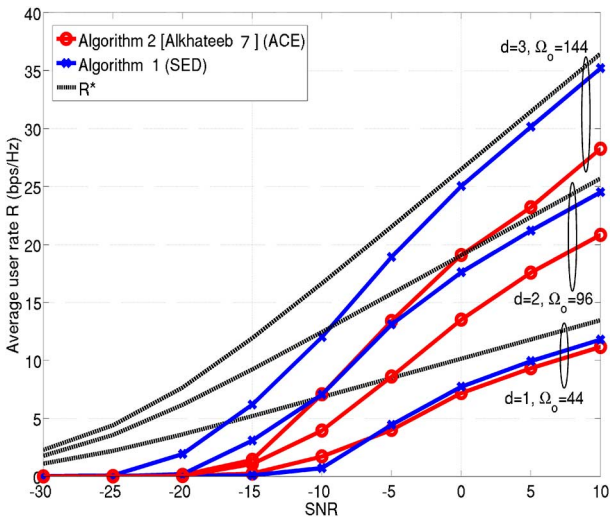


Fig. 8. Average user-rate of proposed schemes over SCM channels ($M = 64, N = 32, m = 2d$)

Moreover, both algorithms seem to exhibit very similar behavior. With that in mind, and for the sake of clarify of our results, we opt to focus on Algorithm 1, the main object of investigation in this work.

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 [7] was *prohibitively high*, thus preventing us from investigating its scalability: we were unable to get any results for systems larger than 128×64 . 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 7 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 remarkably 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 the 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 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 [31] [32], and modifying its parameters to emulate mmWave channels: the number of paths is set to 4, the carrier frequency to 60 GHz, the BS/MS array is modified to implement ULAs, and an 'urban micro' scenario is selected, where a small Ω_o is desired. Fig. 8 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.

Evidently, increasing m (the number of iterations for the Arnoldi) has the effect enhancing the estimation accuracy (and increasing the communication overhead as well (27)). The marginal improvement brought about by increasing m,

is decreasing, and thus our simulations indicated that setting $2d \leq m \leq 3d$ provides a good trade-off.

C. Discussions

A few remarks are in order at this stage, regarding similarities and differences between our two proposed algorithms. As discussed in Remark 5, when the communication overhead is normalized, both SED and MTQR exhibit a similar behavior and performance profile, across the entire SNR range (with a relatively small performance gap): indeed they can be used interchangeably with no change at all in the operational requirements. However, as this work shows, we have an accurate analytical description of the behavior of SED: the Arnoldi algorithm was adapted to the subspace estimation part (with some analytical performance guarantees), and BCD-SD to mathematically describe the decomposition algorithm. In contrast, MTQR is a (heuristic) variation on the original TQR, whose behavior we have not modeled analytically.

One of the conclusions suggested by all the above results, is the fact that the low-SNR performance of the proposed schemes is rather poor. However, interestingly, Figs. 4–8 unambiguously point out that this is the case for the benchmark scheme as well (ACE in [6]): one might be tempted to conjecture at this point that this low-SNR behavior is an inherent aspect of mmWave channel estimation. Initial investigations reveal that, if more RF chains (more than r) can be employed during the RAID echoing phase, the low-SNR performance can be greatly boosted.

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 extended it to operate under the constraints dictated by the hybrid analog-digital architecture, and showed via simulations that it is a good fit 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 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 (9),

$$\boldsymbol{A}\boldsymbol{Q}_m + \tilde{\boldsymbol{W}}_m = \boldsymbol{Q}_m \tilde{\boldsymbol{T}}_m + [\tilde{\boldsymbol{T}}_m]_{m+1,m}\, \boldsymbol{q}_{m+1}\boldsymbol{b}_m^\dagger + \boldsymbol{Q}_m \boldsymbol{E}_m \tag{31}$$

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 (9).

$(P2)$: Noting that $\tilde{\boldsymbol{T}}_m + \boldsymbol{E}_m = \boldsymbol{C}_m + \boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m$, we rewrite (31) 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 $\boldsymbol{Q}_m \boldsymbol{s}_i^{(m)} = \boldsymbol{\theta}_i^{(m)}$

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

Finally, the desired residual is upper bounded as,

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

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

$(P3)$: 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 $\|\boldsymbol{A}\boldsymbol{\theta}_i^{(M)} - \lambda_i^{(M)} \boldsymbol{\theta}_i^{(M)}\|_2^2 \ll 1$.

B. Proof of Lemma 2

The proof follows from a direct application of the Bauer-Fike Theorem [22, 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,

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

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 $\|\boldsymbol{B}\|_2 \leq \|\boldsymbol{B}\|_F$, we rewrite the last equation,

$$|\tilde{\lambda} - \lambda|^2 \leq \|\boldsymbol{Q}_m^\dagger \tilde{\boldsymbol{W}}_m\|_F^2 \leq \|\boldsymbol{Q}_m\|_F^2 \|\tilde{\boldsymbol{W}}_m\|_F^2 = m\|\tilde{\boldsymbol{W}}_m\|_F^2$$

This concludes the proof.

C. 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.,

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

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

$$\{\phi_i^\star\} = \underset{\{\phi_i\}}{\text{argmax}}\ \sum_{i=1}^{M} \Re\left(e^{j(\theta_i-\phi_i)}\right) = \{\theta_i\}$$

where (a) follows from applying the one-to-one mapping $[\boldsymbol{f}]_i \rightarrow 1/\sqrt{M}\ e^{j\phi_i}, \forall i$. 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,

$$g^\star \triangleq \underset{g}{\text{argmin}}\ g^2 - 2g(\|\tilde{\boldsymbol{\gamma}}_1\|_1/\sqrt{M}) = \|\tilde{\boldsymbol{\gamma}}_1\|_1/\sqrt{M} \qquad (32)$$

D. 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 (8) can represent any distortion, and by comparing $\boldsymbol{p}_l$ in both (8) and (25), 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 (9) can be written as,

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

Then using properties of the Frobenius norm,

$$\|\tilde{\boldsymbol{W}}_m\|_F \le \|\boldsymbol{H}\|_F^2 \|\boldsymbol{E}^{(t)}\|_F + (1/d)\|\boldsymbol{H}\|_F \|\boldsymbol{E}^{(r)}\|_F \qquad (33)$$

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

$$\|\boldsymbol{E}^{(t)}\|_F \le 2\sqrt{m},\ \|\boldsymbol{E}^{(r)}\|_F \le \sqrt{m}(1 + d\|\boldsymbol{H}\|_F) \qquad (34)$$

The upper bound follows by combining (33) and (34).

ACKNOWLEDGMENT

We are deeply indebted to the reviewers, whose invaluable comments greatly improved the manuscript.

REFERENCES

[1] P. Cerwall, "On the pulse of the networked society," Ericsson Mobility Report, Nov. 2015.

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

[3] K. Ohata *et al.*, "1.25 Gbps wireless Gigabit ethernet link at 60 GHz-band," in *Proc. IEEE Radio Freq. Integr. Circuits (RFIC) Symp.*, Jun. 2003, pp. 509–512.

[4] X. Zhang, A. Molisch, and S.-Y. Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, 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," *IEEE Trans. Signal Process.*, vol. 58, no. 8, 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," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[7] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, 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 *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2010, pp. 1–6.

[9] Y. Tsang, A. Poon, and S. Addepalli, "Coding the beams: Improving beamforming training in mmwave communication system," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM'11)*, Dec. 2011, pp. 1–6.

[10] J. Wang *et al.*, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 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," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[12] H. Ghauch, M. Bengtsson, T. Kim, and M. Skoglund, "Subspace estimation and decomposition for hybrid analog-digital millimeter-wave MIMO systems," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2015, pp. 395–399.

[13] O. Ayach, R. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 3724–3729.

[14] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, and R. Heath, "Dictionary-free hybrid precoders and combiners for mmwave MIMO systems," in *Proc. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 151–155.

[15] T. Dahl, N. Christophersen, and D. Gesbert, "Blind MIMO eigenmode transmission based on the algebraic power method," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2424–2431, Sep. 2004.

[16] T. Dahl, S. Pereira, N. Christophersen, and D. Gesbert, "Intrinsic subspace convergence in TDD MIMO communication," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2676–2687, Jun. 2007.

[17] D. S. Watkins, *The Matrix Eigenvalue Problem: GR and Krylov Subspace Methods*, 1st ed. Philadelphia, PA, USA: SIAM, 2007.

[18] T. Hrycak, S. Das, G. Matz, and H. Feichtinger, "Low complexity equalization for doubly selective channels modeled by a basis expansion," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 5706–5719, Nov. 2010.

[19] M. Torlak and O. Ozdemir, "A Krylov subspace approach to blind channel estimation for CDMA systems," in *Proc. Conf. Rec. 36th Asilomar Conf. Signals Syst. Comput.*, Nov. 2002, vol. 1, pp. 674–678.

[20] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, 2nd ed. Manchester, UK: Manchester Univ. Press, 2011, pp. 1–337.

[21] L. Withers, R. Taylor, and D. Warme, "Echo-MIMO: A two-way channel training method for matched cooperative beamforming," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4419–4432, Sep. 2008.

[22] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996.

[23] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale MIMO systems," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 2929–2933.

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

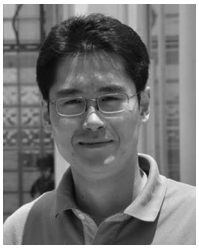
[25] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[26] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, pp. 1123–1153, Sep. 2012.

[27] D. Love and R. Heath, "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1102–1110, Jul. 2003.

[28] X. Zheng, Y. Xie, J. Li, and P. Stoica, "MIMO transmit beamforming under uniform elemental power constraint," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 5395–5406, Nov. 2007.
[29] D. C. Sorensen, "Implicitly restarted Arnoldi/Lanczos methods for large scale eigenvalue calculations," Parallel Numerical Algorithms, vol. 4, 1997, Springer Netherlands, Dordrecht, pp. 119–165.
[30] D. Baum and H. Bolcskei, "Information-theoretic analysis of MIMO channel sounding," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7555–7577, Nov. 2011.
[31] "Spatial channel model for multiple input multiple output (MIMO) simulations," 3GPP TR 25.996 V10.0, Mar. 2011.
[32] J. Salo *et al.*, "MATLAB implementation of the 3GPP spatial channel model," 3GPP TR 25.996, Jan. 2005 [Online]. Available: http://www.tkk.fi/Units/Radio/scm/.

Hadi Ghauch (S'06) received the M.S. degree in information networking from Carnegie Mellon University, Pittsburgh, PA, USA, in 2011, and the Ph.D. degree from the Communication Theory Group, Royal Institute of Technology (KTH), Stockholm, Sweden. He was a Visiting Graduate Research Assistant at the City University of Hong Kong, from June 2014 to December 2014. Using tools from matrix theory and optimization, his focus is developing algorithms for optimizing the performance of multiantenna multiuser cellular systems. He serves as a Reviewer for several IEEE journals including the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON COMMUNICATIONS, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.



Taejoon Kim (S'08–M'11) received the B.S. degree from Sogang University (with highest Hons.), in 2002 and M.S. degree from KAIST, Korea, in 2004, both in electrical engineering and Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, 2011. From 2004 to 2006, he was with Electronics and Telecommunications Research Institute (ETRI), Korea. From 2011 to 2012, he was with the Nokia Research Center (NRC), Berkeley, CA, USA as a Senior Researcher. Before joining City University of Hong Kong as an assistant professor in 2013, he was a postdoctoral researcher in the Communication Theory group at KTH, Sweden. He was the recipient of the Best Paper Award in IEEE PIMRC 2012. His research interests are in the design and analysis of adaptive communication systems.

Mats Bengtsson (M'00–SM'06) received the M.S. degree in computer science from Linköping University, Linköping, Sweden, in 1991, and the Tech.Lic. and Ph.D. degrees in electrical engineering from Royal Institute of Technology (KTH), Stockholm, Sweden, in 1997 and 2000, respectively. From 1991 to 1995, he was with Ericsson Telecom AB Karlstad. He currently holds a position as an Associate Professor with the Signal Processing Laboratory, School of Electrical Engineering, KTH. His research interests include statistical signal processing and its applications to antenna-array processing and communications, radio resource management, and propagation channel modeling. He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2007 to 2009. He is a Member of the IEEE SPCOM Technical Committee.

Mikael Skoglund (S'93–M'97–SM'04) received the Ph.D. degree from Chalmers University of Technology, Gothenburg, Sweden, in 1997. In 1997, he joined the Royal Institute of Technology (KTH), Stockholm, Sweden, where he was appointed to the Chair in Communication Theory in 2003. At KTH, he heads the Communication Theory Division and he is the Assistant Dean for Electrical Engineering. He is also a founding Faculty Member of the ACCESS Linnaeus Center and the Director for the Center Graduate School. His research interests include source channel coding, coding and transmission for wireless communications, Shannon theory, and statistical signal processing. He has authored and coauthored more than 100 journal and 250 conference papers, and holds 6 patents. He has served on numerous technical program committees for IEEE sponsored conferences. From 2003 to 2008, he was an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and from 2008 to 2012, he was on the Editorial Board for the IEEE TRANSACTIONS ON INFORMATION THEORY.