Noisy Beam Alignment Techniques for Reciprocal MIMO Channels

Dennis Ogbe, *Student Member, IEEE,* David J. Love, *Fellow, IEEE,* and Vasanthan Raghavan, *Senior Member, IEEE*

arXiv:1609.03601v3 [cs.IT] 26 Jul 2017

***Abstract*—Future multi-input multi-output (MIMO) wireless communications systems will use beamforming as a first-step towards realizing the capacity requirements necessitated by the exponential increase in data demands. The focus of this work is on beam alignment for time-division duplexing (TDD) systems, for which we propose a number of novel algorithms. These algorithms seek to obtain good estimates of the optimal beamformer/combiner pair (which are the dominant singular vectors of the channel matrix). They are motivated by the power method, an iterative algorithm to determine eigenvalues and eigenvectors through repeated matrix multiplication. In contrast to the basic power method which considers only the most recent iteration and assumes noiseless links, the proposed techniques consider information from all the previous iterations of the algorithm and combine them in different ways. The first technique (*Sequential Least-Squares method*) sequentially constructs a least-squares estimate of the channel matrix, which is then used to calculate the beamformer/combiner pair estimate. The second technique (*Summed Power method*) aims to mitigate the effect of noise by using a linear combination of the previously tried beams to calculate the next beam, providing improved performance in the low-SNR regime (typical for mmWave systems) with minimal complexity/feedback overhead. A third technique (*Least-Squares Initialized Summed Power method*) combines the good performance of the first technique at the high-SNR regime with the low-complexity advantage of the second technique by priming the summed power method with initial estimates from the sequential method.**

***Index Terms*—Beam alignment, beamforming, channel reciprocity, TDD, channel estimation, massive MIMO, mmWave MIMO, power method.**

I. Introduction

ADVANCED multi-input multi-output (MIMO) systems will be among the most important technologies to realize the ever-increasing data rate demands of 5G wireless communication networks [1,2]. The two most promising MIMO applications[1], millimeter-wave (mmWave) MIMO [3]–[6] and massive MIMO [7]–[9], rely on utilizing large beamforming gains to realize the large data rate requirements set for future 5G networks. In mmWave systems, beamforming will be used to compensate for the increased path and penetration losses in the 25–100 GHz band [10,11], whereas massive MIMO systems will multiplex signals of different users via multi-user beamforming [12,13] in sub-6 GHz bands.

Many recent works such as [14]–[17] study the information theoretic limits of beamforming with practical mmWave hardware constraints. However, the substantial gains promised by these studies can be realized only if sufficient channel state information (CSI) is available at the communication nodes. In current state-of-the-art systems, this information is acquired by the use of channel sounding sequences and feedback [18]–[22]. The use of a large number of antenna elements in mmWave and massive MIMO systems will make CSI acquisition via the traditional approach impractical [23]–[25]. Further, in mmWave channels with a relatively small coherence period, it is not possible to *simultaneously* estimate all the elements of the channel matrix due to hardware constraints that render per-antenna sampling inefficient.

One way to circumvent this problem is to exploit the reciprocal nature[2] of wireless channels using time-division duplexing (TDD) systems. Channel reciprocity reduces the overall resources spent on channel sounding since CSI about the channel in one direction can be used to adapt to the channel in the reverse direction. Without readily available channel estimates, communication nodes are forced to obtain their optimal beamformer/combiner pair by sounding different beams during a beam alignment phase [26]. Furthermore, since it is desirable to minimize the usage of time and power resources of the beam alignment phase relative to actual data transmission [23]–[25], it is necessary to employ greedy strategies that maximize the signal-to-noise-ratio (SNR) during each time slot.

One approach to this goal is to leverage the underlying sparse structure [25,29]–[32] or the directional structure [33]–[37] of mmWave channels via the use of low-complexity beamforming approaches. The focus of this work is on another approach that leverages greedy TDD-based beamforming. Many recent works such as [27,28,36,38]–[40] have pursued this approach. The common theme that ties these works is the fact that repeated conjugation, normalization, and retransmission of an arbitrarily initialized beamforming vector

D. Ogbe and D. J. Love are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47906, USA (e-mail: {dogbe, djlove}@purdue.edu).

Vasanthan Raghavan is with Qualcomm, Inc., Bridgewater, NJ 08807, USA (email: vasanthan_raghavan@ieee.org)

This material is based upon work supported in part by the National Science Foundation under Grant No. CNS-1642982. A version of this paper has been published at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2017, New Orleans, LA.

[1]We use the terms massive and mmWave MIMO in the sense of the common understanding at 3GPP 5G-NR with massive MIMO typically corresponding to sub-6 GHz systems and mmWave MIMO typically corresponding to over-25 GHz systems.

[2]This work assumes that the radio-frequency (RF) circuit asymmetries in the uplink and downlink have been compensated via calibration and hence does not consider these aspects.

Copyright © 2017 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org. This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication. The final version of record is available at http://dx.doi.org/10.1109/TSP.2017.2715001

	Computational Count	**Feedback**
BIMA (see [27])	$k_{\mathsf{max}} \cdot \mathcal{O}(M)$	-
BSM (see [28])	$k_{\mathsf{max}} \cdot \mathcal{O}(M)$	-
Sequential Least-Squares	$k_{\mathsf{max}} \cdot \mathcal{O}(M^3)$	$k_{\mathsf{max}} \cdot B \cdot (M_r + M_t)$
Summed Power	$k_{\mathsf{max}} \cdot \mathcal{O}(M)$	-
Least-Squares Initialized Summed Power	$k_{\mathsf{switch}} \cdot \mathcal{O}(M^3) + (k_{\mathsf{max}} - k_{\mathsf{switch}}) \cdot \mathcal{O}(M)$	$k_{\mathsf{switch}} \cdot B \cdot (M_r + M_t)$

TABLE I. Computational complexity and feedback requirements of different beam alignment techniques.

through a reciprocal MIMO channel (with no noise) is akin to performing the power method[3] on the channel matrix.

Beam alignment algorithms based on the power method are attractive due to their simplicity and low computational complexity. However, simple implementations like the ones proposed in [27,36,38] are likely to perform poorly in the low-SNR regime [36]. Other approaches for finding good beams using the power method have been proposed in [28] and [42]. These techniques offer improvements on the robustness and speed of convergence of the basic power method at the cost of additional complexity. The main idea behind these improved techniques is to combine previous estimates of the optimal beams with the received information during each time slot. In addition to these techniques, recent works such as [43] study the application of the more general Arnoldi iteration to the beam alignment problem. Furthermore, feedback-based beam alignment techniques for frequency-division duplexing (FDD) systems, which represent the majority of currently deployed commercial systems, have been studied in [44] and [45].

Building on [28,42,43], this paper presents multiple novel techniques for the TDD MIMO beam alignment problem in reciprocal channels. These techniques improve upon the performance of the simple power method-based algorithms, especially in low-SNR environments which are typical of practical mmWave systems [36]. The first technique, labeled the *sequential least-squares method*, is based on constructing a least-squares estimate of the channel matrix sequentially using the previously-used sounding beams. The channel estimates at each iteration can then be used to compute the next sounding beamformer/combiner pair, which is exchanged through a feedback[4] link. The second technique, labeled the *summed power method*, does not require a feedback link and computes a normalized running sum of the previous beamformers, thus gaining greater robustness against noise through averaging.

The first technique achieves better performance in the high-SNR regime at the cost of additional complexity and feedback[5] overhead. On the other hand, the second technique achieves better performance in the low-SNR regime and yet does not need significant complexity/feedback overhead. However, this technique has deteriorating performance as the SNR increases due to continued noise averaging. To enjoy the complementary advantages of both techniques, we propose a third technique, labeled the *least-squares initialized summed power method*, that switches from the first technique to the second technique after a certain number of iterations. By appropriately choosing the switching point k_{switch}, significant performance improvement can be realized in the high-SNR regime with a small increase in feedback and computational complexity. The motivation behind the third technique is that the high-SNR performance of a beam alignment algorithm critically depends on the beam initialization. By choosing this initialization from a scheme that rejects noise near-optimally, we are able to prime a low-complexity scheme and improve performance. Thus, the proposed approaches in this paper provide useful low-complexity solutions for realizing the large beamforming gains of mmWave systems.

Table I compares the computational complexity and feedback requirements of the techniques proposed in this paper with those from [27] and [28] over a run of k_{max} iterations of each algorithm. In this table, $M = \max(M_r, M_t)$ with M_r and M_t standing for the receive and transmit antenna dimensions, respectively and $\mathcal{O}(\cdot)$ stands for the big-O notation: $f(x) = \mathcal{O}(g(x))$ as $x \to \infty$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} \leq K_u$ for some $K_u < \infty$. The column labeled "Computational Count" lists the approximate number of complex-valued arithmetic operations required during the beam alignment phase. The column labeled "Feedback" lists the approximate total number of feedback bits exchanged during the beam alignment phase, where B is the number of bits per complex-valued element of the beamforming vectors.

Organization: This paper is organized as follows. Section II provides an overview of the system model and sets up the beam alignment techniques discussed in the rest of the paper. Sections III-V elaborate on the power methods proposed in this work. Simulation results illustrating the advantages of the proposed techniques are presented in Section VI with concluding remarks provided in Section VII.

Notations: The following notations are used in the paper. Bold upper-case and lower-case letters (such as $\mathbf{A}$ and $\mathbf{a}$) denote matrices and column vectors, respectively. The operators $(\cdot)^{\mathsf{T}}$, $\overline{(\cdot)}$, $(\cdot)^*$ and $(\cdot)^\dagger$ denote matrix transposition, element-wise complex conjugation, matrix Hermitian transposition and Moore-Penrose pseudoinverse operations, respectively. $\|\cdot\|_2$ denotes the vector ℓ_2-norm and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. $\mathbf{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathbb{C}^{n\times m}$, $\mathbb{C}^n$ and $\mathrm{E}\{\cdot\}$ stand for the space of $n \times m$ complex matrices, $n \times 1$ complex vectors and the expectation operator, respectively.

[3]The power method is a result from numerical linear algebra which provides a simple algorithm to find the dominant eigenvector(s)/eigenspace of a matrix [41].

[4]Due to the small packet overheads, the feedback link is assumed to be ideal: error-free and incurring no delay.

[5]Nevertheless, the feedback link in itself is not onerous given that mmWave links are expected to support Gbps rates.

II. System Model

We consider a multi-antenna communication system such as the one shown in Fig. 1, consisting of two transceivers (communication nodes), with M_t antennas at node 1 and M_r antennas at node 2. The two nodes communicate over a channel $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$. We also assume that $\mathbf{H}$ is reciprocal, i.e., the channel matrix from node 2 to node 1 (uplink) is the transpose of the channel matrix from node 1 to node 2 (downlink). For a transmission on the downlink channel, the transmit data at node 1 is precoded by a unit-norm transmit beamforming vector $\mathbf{f} = \begin{bmatrix} f_1 & f_2 & \dots & f_{M_t} \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^{M_t}$, sent over the channel, and combined at node 2 with a unit-norm receive combiner $\mathbf{z} = \begin{bmatrix} z_1 & z_2 & \dots & z_{M_r} \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^{M_r}$. Hence, for a data symbol $s_o[k]$ sent on the downlink channel, we obtain the received symbol

$$r_o[k] = \sqrt{\rho_o}\, \mathbf{z}^* \mathbf{H} \mathbf{f} s_o[k] + n_o[k], \tag{1}$$

where ρ_o is the downlink SNR and $n_o[k] \sim \mathcal{CN}(0,1)$ is additive Gaussian noise, which we assume to be independent and identically distributed (i.i.d.) spatially as well as temporally. Similarly, for a data symbol $s_e[k]$ sent on the uplink channel, node 1 obtains the received symbol

$$r_e[k] = \sqrt{\rho_e}\, \mathbf{f}^{\mathsf{T}} \mathbf{H}^{\mathsf{T}} \overline{\mathbf{z}} s_e[k] + n_e[k]. \tag{2}$$

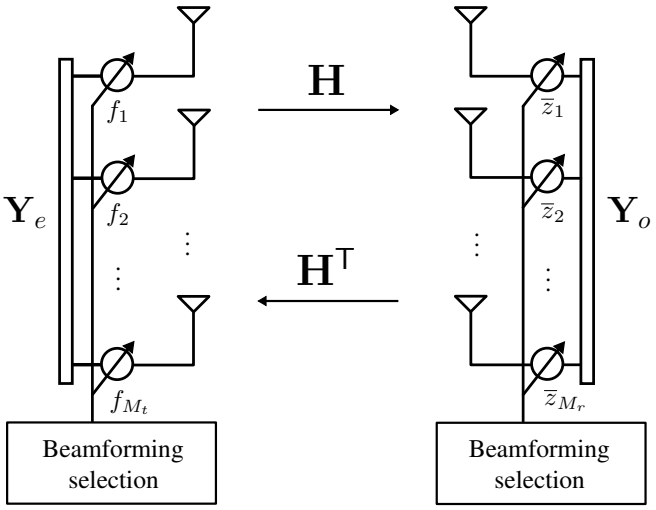


Fig. 1. Communication node 1 transmits data over the downlink channel $\mathbf{H}$ to node 2, while node 2 transmits data over the uplink channel $\mathbf{H}^{\mathsf{T}}$.

In both (1) and (2), we denote $|\mathbf{z}^* \mathbf{H} \mathbf{f}|^2 = |\mathbf{f}^{\mathsf{T}} \mathbf{H}^{\mathsf{T}} \overline{\mathbf{z}}|^2$ as the effective channel gain, which we want to maximize in order to achieve reliable communications and the highest possible data rates in both directions. We denote the vectors that achieve this as $\mathbf{f}_{\mathsf{opt}}$ and $\mathbf{z}_{\mathsf{opt}}$, respectively. It is well-known from [46,47] that the effective channel gain is maximized when $\mathbf{f}$ and $\mathbf{z}$ are the right- and left-singular vectors of $\mathbf{H}$ corresponding to the largest singular value of $\mathbf{H}$ and that its maximum achievable value is $\|\mathbf{H}\|_2^2 = \lambda_{\mathsf{max}}(\mathbf{H}^* \mathbf{H})$. Further, we assume that neither node has knowledge of the channel. It is therefore impossible for either node to compute the estimates of $\mathbf{f}_{\mathsf{opt}}$ and $\mathbf{z}_{\mathsf{opt}}$ using the singular value decomposition (SVD) of their channel estimate. Instead, as mentioned earlier, these estimates are obtained iteratively.

In the proposed techniques, both nodes cooperatively determine $\mathbf{f}_{\mathsf{opt}}$ and $\mathbf{z}_{\mathsf{opt}}$ during a beam training phase by exploiting the channel's reciprocity property. To model this, our system operates on a *ping-pong* observation framework, which divides each discrete channel use into two time slots. During slot 1 (*ping*), node 1 sends a training symbol to node 2 on the downlink channel $\mathbf{H}$. During slot 2 (*pong*), node 2 sends a training symbol back to node 1 on the uplink channel $\mathbf{H}^{\mathsf{T}}$. Since the two nodes are exchanging training symbols that are known to both sides, we focus on the received signal vectors after correlating with the known training data. Hence, the observation at slot 1 (at node 2) during the k-th channel use is given as

$$\mathbf{y}_o[k] = \sqrt{\rho_o}\, \mathbf{H} \mathbf{f}[k] + \mathbf{n}_o[k]. \tag{3}$$

In (3), the term $\mathbf{f}[k]$ denotes an estimate of $\mathbf{f}_{\mathsf{opt}}$ at training phase time-index k and $\mathbf{n}_o[k] \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is a complex Gaussian noise vector of size M_r. Due to the reciprocity of the uplink and downlink channels, the observation at slot 2 (at node 1) is given as

$$\mathbf{y}_e[k] = \sqrt{\rho_e}\, \mathbf{H}^{\mathsf{T}} \overline{\mathbf{z}}[k] + \mathbf{n}_e[k]. \tag{4}$$

Similar to (3), ρ_e denotes the uplink SNR, $\mathbf{z}[k]$ denotes an estimate of $\mathbf{z}_{\mathsf{opt}}$ at training phase time-index k and $\mathbf{n}_e[k] \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is a complex Gaussian noise vector of size M_t.

The proposed techniques in this work use all of the ping-pong observations to determine a good estimate to the optimal beamforming vectors $\mathbf{f}_{\mathsf{opt}}$ and $\mathbf{z}_{\mathsf{opt}}$ in a greedy manner, i.e., each time-index yields the current choice based on all previously collected observations.

III. Power Method Using A Sequentially Estimated Channel Matrix

A. Batch Least-Squares Estimator

In the first scheme, since the channel matrix is not known at either node, the nodes construct a least-squares estimate of $\mathbf{H}$ before each ping-pong time slot using all of the previous estimates of $\mathbf{f}_{\mathsf{opt}}$ and $\mathbf{z}_{\mathsf{opt}}$. These estimates are then used to compute the next state of their beamforming vectors.

In particular, using all observations up to time slot k, we can write (3) and (4) in matrix form as

$$\mathbf{Y}_{o,k} = \sqrt{\rho_o}\, \mathbf{H} \mathbf{F}_k + \mathbf{N}_{o,k} \tag{5}$$

and

$$\mathbf{Y}_{e,k} = \sqrt{\rho_e}\, \mathbf{H}^{\mathsf{T}} \overline{\mathbf{Z}}_k + \mathbf{N}_{e,k}. \tag{6}$$

In (5) and (6), $\mathbf{F}_k = [\mathbf{f}[0]\ \mathbf{f}[1]\ \dots\ \mathbf{f}[k]]$ and $\mathbf{Z}_k = [\mathbf{z}[0]\ \mathbf{z}[1]\ \dots\ \mathbf{z}[k]]$ contain all of the estimates of $\mathbf{f}_{\mathsf{opt}}$ and $\mathbf{z}_{\mathsf{opt}}$ up to time-index k. Also, $\mathbf{Y}_{o,k} = [\mathbf{y}_o[0]\ \mathbf{y}_o[1]\ \dots\ \mathbf{y}_o[k]]$ and $\mathbf{Y}_{e,k} = [\mathbf{y}_e[0]\ \mathbf{y}_e[1]\ \dots\ \mathbf{y}_e[k]]$ contain all of the observed signal vectors, respectively. On the other hand, $\mathbf{N}_{o,k} = [\mathbf{n}_o[0]\ \mathbf{n}_o[1]\ \dots\ \mathbf{n}_o[k]]$ and $\mathbf{N}_{e,k} = [\mathbf{n}_e[0]\ \mathbf{n}_e[1]\ \dots\ \mathbf{n}_e[k]]$ contain all of the noise vectors, respectively.

Based on this information, node 1 constructs an estimate of the channel by solving the least-squares problem

$$\widehat{\mathbf{H}}_{e,k} = \underset{\widetilde{\mathbf{H}} \in \mathbb{C}^{M_r \times M_t}}{\operatorname{argmin}} \left(\left\| \mathbf{Y}_{e,k-1}^{\mathsf{T}} - \sqrt{\rho_e}\, \mathbf{Z}_{k-1}^{*} \widetilde{\mathbf{H}} \right\|_F^2 \right). \quad (7)$$

Similarly, node 2 constructs an estimate of the channel by solving

$$\widehat{\mathbf{H}}_{o,k} = \underset{\widetilde{\mathbf{H}} \in \mathbb{C}^{M_r \times M_t}}{\operatorname{argmin}} \left(\left\| \mathbf{Y}_{o,k} - \sqrt{\rho_o}\, \widetilde{\mathbf{H}} \mathbf{F}_k \right\|_F^2 \right). \quad (8)$$

Note that there exists an asymmetry in the time-index between (7) and (8). The solutions to these least-squares problems ($\widehat{\mathbf{H}}_{e,k}$ and $\widehat{\mathbf{H}}_{o,k}$) are obtained using all of the previously observed outputs and beamforming vectors and are as follows:

$$\widehat{\mathbf{H}}_{e,k} = \frac{\left(\mathbf{Z}_{k-1}^{*}\right)^{\dagger} \mathbf{Y}_{e,k-1}^{\mathsf{T}}}{\sqrt{\rho_e}}, \quad (9)$$

$$\widehat{\mathbf{H}}_{o,k} = \frac{\mathbf{Y}_{o,k} \left(\mathbf{F}_k\right)^{\dagger}}{\sqrt{\rho_o}}. \quad (10)$$

In (9) and (10), the $(\cdot)^{\dagger}$ operation stands for the Moore-Penrose pseudoinverse[6] of the underlying matrix. Using the definitions of the pseudoinverse, we have the following simplifications:

$$\widehat{\mathbf{H}}_{e,k} = \frac{1}{\sqrt{\rho_e}} \cdot \begin{cases} \mathbf{Z}_{k-1} \left(\mathbf{Z}_{k-1}^{*} \mathbf{Z}_{k-1}\right)^{-1} \mathbf{Y}_{e,k-1}^{\mathsf{T}} & \text{if } k < M_r \\ \left(\mathbf{Z}_{k-1} \mathbf{Z}_{k-1}^{*}\right)^{-1} \mathbf{Z}_{k-1} \mathbf{Y}_{e,k-1}^{\mathsf{T}} & \text{if } k \geq M_r, \end{cases} \quad (11)$$

$$\widehat{\mathbf{H}}_{o,k} = \frac{1}{\sqrt{\rho_o}} \cdot \begin{cases} \mathbf{Y}_{o,k} \left(\mathbf{F}_k^{*} \mathbf{F}_k\right)^{-1} \mathbf{F}_k^{*} & \text{if } k < M_t \\ \mathbf{Y}_{o,k} \mathbf{F}_k^{*} \left(\mathbf{F}_k \mathbf{F}_k^{*}\right)^{-1} & \text{if } k \geq M_t. \end{cases} \quad (12)$$

Note that the second condition in both (11) and (12) has been separated (from the first) at the $k = M_r$ and $k = M_t$ cases artificially. Since the solutions in (9) and (10) use all the underlying data up to time-index k, we call this approach the *batch least-squares method*. Once $\widehat{\mathbf{H}}_{e,k}$ and $\widehat{\mathbf{H}}_{o,k}$ have been estimated, beamforming vector computation follows directly from the SVD theorem [41,48].

Lemma 1. *Let the SVD of a matrix $\mathbf{A}$ be denoted as $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{*}$. We can obtain a multiple of the i-th left-singular vector of $\mathbf{A}$ by multiplying $\mathbf{A}$ with its i-th right-singular vector, i.e., $\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$. Here, σ_i is the i-th singular value. Similarly, we can obtain a multiple of the i-th right-singular vector by multiplying $\mathbf{A}^{*}$ with its i-th left-singular vector, i.e., $\mathbf{A}^{*}\mathbf{u}_i = \sigma_i \mathbf{v}_i$.*

Applying Lemma 1, we note that node 1 can compute its k-th estimate for $\mathbf{f}_{\mathsf{opt}}$ as

$$\mathbf{f}[k] = \frac{\widehat{\mathbf{H}}_{e,k}^{*} \mathbf{z}[k-1]}{\left\| \widehat{\mathbf{H}}_{e,k}^{*} \mathbf{z}[k-1] \right\|_2}. \quad (13)$$

[6]Note that the expressions in (9) and (10) hold even in the case when $k < M_r$ since it can be shown that the left pseudoinverse of a "tall matrix", i.e., a $K_1 \times K_2$ matrix with $K_1 > K_2$ minimizes $\|\mathbf{A}\mathbf{C} - \mathbf{I}\|^2$, where $\mathbf{C}$ is optimized over all $K_2 \times K_1$ matrices.

Similarly, applying Lemma 1, we note that node 2 obtains its k-th estimate for $\mathbf{z}_{\mathsf{opt}}$ as

$$\mathbf{z}[k] = \frac{\widehat{\mathbf{H}}_{o,k} \mathbf{f}[k]}{\left\| \widehat{\mathbf{H}}_{o,k} \mathbf{f}[k] \right\|_2}. \quad (14)$$

Some comments are in order at this stage.

1) We have the following result on error covariance matrices with the batch estimators.

   Theorem 1. *If $k \geq \max(M_t, M_r)$, the error covariance matrices of the columns of $\widehat{\mathbf{H}}_{e,k}$ and $\widehat{\mathbf{H}}_{o,k}$ under the assumption of a channel $\mathbf{H}$ with independent and identically distributed (i.i.d.) entries are given as*

   $$\mathbf{C}_{e,k} = \frac{1}{\rho_e} \left(\mathbf{Z}_{k-1} \mathbf{Z}_{k-1}^{*}\right)^{-1} \quad (15)$$

   and

   $$\mathbf{C}_{o,k} = \frac{1}{\rho_o} \left(\mathbf{F}_k \mathbf{F}_k^{*}\right)^{-1}, \quad (16)$$

   respectively.

   For the proof, see Appendix A.
2) The proposed algorithm is valid for a general channel matrix $\mathbf{H}$ and the i.i.d. assumption has been made only in the context of Theorem 1. From Theorem 1, we note that the estimation error is monotonically decreasing in the SNRs, ρ_e and ρ_o. This shows that a reasonable channel estimate can be obtained in the medium- to high-SNR regimes. Nevertheless, the low-SNR regime is typical in mmWave systems, especially with self-blocking or blocking due to other humans, vehicles, buildings, foliage, etc. [11,36]. Thus, Section VI studies the performance of the different approaches proposed in this work as a function of the SNR as well as for both i.i.d. and sparse channel models.
3) While we need ρ_e and ρ_o to compute $\widehat{\mathbf{H}}_{e,k}$ and $\widehat{\mathbf{H}}_{o,k}$, the beamformer estimates do not depend on these quantities. Therefore, a mismatched estimate of ρ_e and ρ_o is still sufficient to implement the proposed scheme.
4) The computation of $\widehat{\mathbf{H}}_{e,k}$ and $\mathbf{f}[k]$ at node 1 requires the feedback of $\mathbf{z}[k-1]$ from node 2. Similarly, computation of $\widehat{\mathbf{H}}_{o,k}$ and $\mathbf{z}[k]$ at node 2 requires the feed forward of $\mathbf{f}[k]$ from node 1. While on a first glance this feedback and feed forward sounds onerous, given the Gbps rates that mmWave systems are expected to realize, these feedback overheads can be supported on either a lower frequency control/data channel or on a mmWave control channel. This feedback/feed forward has to be specified only over a large sub-band (a component carrier, for example) or on a wideband basis, further reducing the overhead. Thus, it makes sense to not dismiss this approach as impractical and study its performance gain relative to other competing approaches. This is the subject of Section VI. We will also consider other lower feedback overhead approaches in Sections IV and V.
5) Throughout this text, we are assuming that the initial transmit beam $\mathbf{f}[0]$ is a unit-norm complex random vector. An alternative approach which could be considered for channels with a large line-of-sight component

would be to initialize $\mathbf{f}[0]$ with an omni-directional beam pattern that approximates equal gain in every spatial direction. Omni-directional beams have been constructed and used in [26], but are out of the scope of this work.

The batch least-squares estimators are obtained by computing the Moore-Penrose pseudoinverse. The complexity in computing these estimators in (11) and (12) is limited to the inversion of a $\widetilde{k} \times \widetilde{k}$ matrix where $\widetilde{k} = \min(M_r, k)$ in the former case and $\widetilde{k} = \min(M_t, k)$ in the latter case. However, computation of the matrix to be inverted requires a multiplication count that scales with k and can hence be onerous.

B. Sequential Least-Squares Estimator (Optimal)

Following a similar approach to [49], we therefore propose a sequential algorithm that updates each previous channel estimate based on the current received signal vector. This approach minimizes computational burden as well as eliminates the need to store all of the previously received signal and beamforming vectors. Since (13) uses the conjugate transpose of the channel to compute a new beamformer, we use an algorithm that directly computes an estimate for $\widehat{\mathbf{H}}_{e,k}^*$ instead of $\widehat{\mathbf{H}}_{e,k}$. This choice is made here simply to make the derivation of the sequential formulas more consistent between the two nodes. In this setup, the sequential version of (9) (the channel estimator update) is given as

$$\widehat{\mathbf{H}}_{e,k}^* = \widehat{\mathbf{H}}_{e,k-1}^* + \left(\frac{\overline{\mathbf{y}}_e[k-1]}{\sqrt{\rho_e}} - \widehat{\mathbf{H}}_{e,k-1}^* \mathbf{z}[k-1] \right) \mathbf{K}_{e,k} \quad (17)$$

where

$$\mathbf{K}_{e,k} = \frac{\mathbf{z}^*[k-1]\mathbf{C}_{e,k-1}}{1 + \mathbf{z}^*[k-1]\mathbf{C}_{e,k-1}\mathbf{z}[k-1]} \quad (18)$$

and the covariance matrix update is given as

$$\mathbf{C}_{e,k} = \mathbf{C}_{e,k-1} \left(\mathbf{I} - \mathbf{z}[k-1]\mathbf{K}_{e,k} \right). \quad (19)$$

After obtaining $\widehat{\mathbf{H}}_{e,k}^*$, node 1 uses (13) to obtain the k-th estimate for $\mathbf{f}_{\mathsf{opt}}$. The value of this beamformer then needs to be fed back to node 2, where it will be used to obtain the next estimate for $\mathbf{z}_{\mathsf{opt}}$.

At node 2, the same sequential algorithm is used to solve the least-squares problem, and the update expression for $\widehat{\mathbf{H}}_{o,k}$ becomes

$$\widehat{\mathbf{H}}_{o,k} = \widehat{\mathbf{H}}_{o,k-1} + \left(\frac{\mathbf{y}_o[k]}{\sqrt{\rho_o}} - \widehat{\mathbf{H}}_{o,k-1}\mathbf{f}[k] \right) \mathbf{K}_{o,k} \quad (20)$$

where

$$\mathbf{K}_{o,k} = \frac{\mathbf{f}^*[k]\mathbf{C}_{o,k-1}}{1 + \mathbf{f}^*[k]\mathbf{C}_{o,k-1}\mathbf{f}[k]} \quad (21)$$

with the covariance matrix update

$$\mathbf{C}_{o,k} = \mathbf{C}_{o,k-1} \left(\mathbf{I} - \mathbf{f}[k]\mathbf{K}_{o,k} \right). \quad (22)$$

Node 2 then obtains $\mathbf{z}[k]$ from (14), which in turn is fed back to node 1 to compute $\mathbf{f}[k+1]$.

We observe that these sequential least-squares (SLS) estimators are only equivalent to their batch estimators when the beamformer matrices $\mathbf{F}_k$ and $\mathbf{Z}_k$ are of full column rank. That is, for $k \leq \text{rank}[\mathbf{H}]$, both nodes would need to compute their channel estimates using the batch approach. Theorem 2 establishes that the sequential approach is equivalent to using the batch estimator for all k.

Theorem 2. *The sequential least-squares estimator* $\widehat{\mathbf{H}}_{o,k}^{\text{Seq}}$ *is identical to the batch least-squares estimator* $\widehat{\mathbf{H}}_{o,k}^{\text{Batch}}$ *for* $k > r$ *if* $\widehat{\mathbf{H}}_{o,r}^{\text{Seq}} = \widehat{\mathbf{H}}_{o,r}^{\text{Batch}}$ *where* $r = rank[\mathbf{H}]$.

For the proof, see Appendix B.

Motivated by Theorem 2, we propose to initialize $\mathbf{f}[0]$ as a complex random unit-norm vector. We then use the batch estimator from (9) and (10) for $k \leq \text{rank}[\mathbf{H}]$ and switch to the sequential estimator for $k > \text{rank}[\mathbf{H}]$. Under these assumptions, the Gauss-Markov Theorem states that the least-squares estimator is the best linear unbiased estimator (BLUE) for the channel matrix $\mathbf{H}$ [49]. The asymptotic normality property of the least-squares estimator [50] then shows how our sequential estimates for the channel matrix converge to its true value. As the channel estimate becomes more accurate with the number of iterations, steps (13) and (14) essentially perform a two-iteration power method without noise, which converges at a rate of $(\sigma_1/\sigma_2)^2$ [41]. The description under Algorithm 1 gives a succinct summary of this technique (labeled as *SLS Estimator (Optimal)*) corresponding to stopping at k_{max} iterations, where k_{max} is chosen appropriately.

Algorithm 1 SLS Estimator (Optimal)

Initialize $\mathbf{f}[0]$ as a complex random unit-norm vector.
for all $k = 1, \ldots, k_{\mathsf{max}}$ **do**
  Node 2 receives $\mathbf{y}_o[k-1]$ as in (3) and gets $\mathbf{f}[k-1]$ from Node 1
  if $k \leq \text{rank}[\mathbf{H}]$ **then**
    Node 2 estimates $\widehat{\mathbf{H}}_{o,k-1}$ as in (12)
  else if $k > \text{rank}[\mathbf{H}]$ **then**
    Node 2 estimates $\widehat{\mathbf{H}}_{o,k-1}$ as in (20)
  end if
  Node 2 computes $\mathbf{z}[k-1]$ as in (14)
  Node 1 receives $\mathbf{y}_e[k-1]$ as in (4) and gets $\mathbf{z}[k-1]$ from Node 2
  if $k \leq \text{rank}[\mathbf{H}]$ **then**
    Node 1 estimates $\widehat{\mathbf{H}}_{e,k}$ as in (11)
  else if $k > \text{rank}[\mathbf{H}]$ **then**
    Node 1 estimates $\widehat{\mathbf{H}}_{e,k}$ as in (17)
  end if
  Node 1 computes $\mathbf{f}[k]$ as in (13)
end for

C. Sequential Least-Squares Estimator (Suboptimal)

For large antenna dimensions as is typical in mmWave systems, it can be computationally difficult to use the batch estimator for the first M_t iterations. In this case, we initialize the sequential least-squares estimator with an arbitrary initial covariance estimate. With such a choice, the following result shows that we are guaranteed to asymptotically approach the batch least-squares estimate.

Theorem 3. *The sequential least squares estimate* $\widehat{\mathbf{H}}_{o,k}^{\text{Seq}}$*, initialized with* $\mathbf{C}_{o,0} = \alpha\mathbf{I}$ *approaches the batch least-squares estimate* $\widehat{\mathbf{H}}_{o,k}^{\text{Batch}}$ *as* $\alpha \to \infty$.

For the proof, see Appendix C.

Using Theorem 3, the alternative algorithm (labeled as *SLS Estimator (Suboptimal)*) also requires us to initialize $\mathbf{f}[0]$ as a complex random unit-norm vector. The nodes then transmit this vector across $\mathbf{H}$ according to (3) and (4) and compute their initial rank-1 channel estimates and beamforming vectors as follows:

$$\widehat{\mathbf{H}}_{o,0} = \frac{\mathbf{y}_o[0]\mathbf{f}^*[0]}{\sqrt{\rho_o}} \tag{23}$$

$$\mathbf{z}[0] = \frac{\widehat{\mathbf{H}}_{o,0}\mathbf{f}[0]}{\left\|\widehat{\mathbf{H}}_{o,0}\mathbf{f}[0]\right\|_2} = \frac{\mathbf{y}_o[0]}{\|\mathbf{y}_o[0]\|_2} \tag{24}$$

$$\widehat{\mathbf{H}}_{e,1}^* = \frac{\overline{\mathbf{y}}_e[0]\mathbf{z}^*[0]}{\sqrt{\rho_e}} \tag{25}$$

$$\mathbf{f}[1] = \frac{\widehat{\mathbf{H}}_{e,1}^*\mathbf{z}[0]}{\left\|\widehat{\mathbf{H}}_{e,1}^*\mathbf{z}[0]\right\|_2}. \tag{26}$$

The nodes then initialize $\mathbf{C}_{o,0} = \mathbf{C}_{e,1} = \alpha\mathbf{I}$ for an appropriately chosen α. The nodes then use the sequential formulas (17)-(22) to estimate their beamformers. To conclude this section, Algorithm 2 provides a brief summary of this technique corresponding to k_{max} iterations.

Algorithm 2 SLS Estimator (Suboptimal)

Initialize $\mathbf{f}[0]$ as a complex random unit-norm vector and obtain $\widehat{\mathbf{H}}_{o,0}$, $\mathbf{z}[0]$, $\widehat{\mathbf{H}}_{e,1}$ and $\mathbf{f}[1]$ as in (23)-(26).
Initialize $\mathbf{C}_{o,0} = \mathbf{C}_{e,1} = \alpha\mathbf{I}$ for an appropriate α.
for all $k = 1, \ldots, k_{\mathsf{max}}$ **do**
  Node 2 receives $\mathbf{y}_o[k]$ as in (3) and gets $\mathbf{f}[k]$ from Node 1
  Node 2 estimates $\widehat{\mathbf{H}}_{o,k}$ as in (20)
  Node 2 computes $\mathbf{z}[k]$ as in (14)
  Node 1 receives $\mathbf{y}_e[k]$ as in (4) and gets $\mathbf{z}[k]$ from Node 2
  Node 1 estimates $\widehat{\mathbf{H}}_{e,k+1}$ as in (17)
  Node 1 computes $\mathbf{f}[k+1]$ as in (13)
end for

IV. Summed Power Method

We now propose an alternate approach, labeled the *summed power method*, to align the beams at the two nodes. The main idea behind this scheme is that both nodes calculate their next beamformers as a function of the running sum of their previously received vectors, effectively averaging out noise in the estimation process. This low-complexity approach adds only one additional vector addition per iteration at each node when compared to the simple power method [27,36]. Additionally, there is no need for a feedback link, as neither node needs to have knowledge of the other node's beamformer.

As described in Section II, both nodes exchange training symbols according to (3) and (4). However, instead of simply conjugating and retransmitting their received vector as in the simple power method, both nodes obtain their next beamformers from a running sum of all of their previous received vectors. At each time-index k, node 1 computes its next beamformer as

$$\mathbf{f}[k+1] = \alpha_k\left[\overline{\mathbf{y}}_e[k] + \overline{\mathbf{y}}_e[k-1] + \cdots + \overline{\mathbf{y}}_e[0]\right] \tag{27}$$

$$= \alpha_k\,\mathbf{s}_e[k]. \tag{28}$$

Similarly, node 2 computes its next beamformer as

$$\mathbf{z}[k+1] = \beta_k\left[\mathbf{y}_o[k] + \mathbf{y}_o[k-1] + \cdots + \mathbf{y}_o[0]\right] \tag{29}$$

$$= \beta_k\,\mathbf{s}_o[k]. \tag{30}$$

In (28) and (30), $\mathbf{s}_e[k]$ and $\mathbf{s}_o[k]$ are the state vectors at each node which hold the running sum of the received vectors. The terms α_k and β_k are normalization factors ensuring the unit-norm constraint and are given as

$$\alpha_k = \frac{1}{\|\mathbf{s}_e[k]\|_2} \tag{31}$$

and

$$\beta_k = \frac{1}{\|\mathbf{s}_o[k]\|_2}. \tag{32}$$

Algorithm 3 provides an overview of the proposed technique.

Algorithm 3 Summed Power Method

Initialize $\mathbf{f}[0]$ and $\mathbf{z}[0]$ as complex random unit-norm vectors.
for all $k = 1, \ldots, k_{\mathsf{max}}$ **do**
  Node 2 receives $\mathbf{y}_o[k-1]$ as in (3)
  Node 2 computes $\mathbf{z}[k]$ as in (29)
  Node 1 receives $\mathbf{y}_e[k-1]$ as in (4)
  Node 1 computes $\mathbf{f}[k]$ as in (27)
end for

For further analysis of the proposed algorithm, it is useful to define the state-space model of the combined system state: $\mathbf{s}[k] = \left[\mathbf{s}_e^{\mathsf{T}}[k]\ \mathbf{s}_o^{\mathsf{T}}[k]\right]^{\mathsf{T}}$. A straightforward simplification of $\mathbf{s}[k]$ shows that

$$\mathbf{s}[k] = \begin{bmatrix}\mathbf{s}_e[k]\\ \mathbf{s}_o[k]\end{bmatrix} \tag{33}$$

$$= \begin{bmatrix}\mathbf{I} & \sqrt{\rho_e}\beta_{k-1}\mathbf{H}^*\\ \sqrt{\rho_o}\alpha_{k-1}\mathbf{H} & \mathbf{I}\end{bmatrix}\mathbf{s}[k-1] + \mathbf{n}[k] \tag{34}$$

$$= \prod_{i=0}^{k-1}\begin{bmatrix}\mathbf{I} & \sqrt{\rho_e}\beta_{k-1-i}\mathbf{H}^*\\ \sqrt{\rho_o}\alpha_{k-1-i}\mathbf{H} & \mathbf{I}\end{bmatrix}\mathbf{s}[0] + \sum_{\ell=1}^{k}\prod_{j=\ell}^{k-1}\begin{bmatrix}\mathbf{I} & \sqrt{\rho_e}\beta_{k-1+\ell-j}\mathbf{H}^*\\ \sqrt{\rho_o}\alpha_{k-1+\ell-j}\mathbf{H} & \mathbf{I}\end{bmatrix}\mathbf{n}[\ell] \tag{35}$$

where

$$\mathbf{n}[k] = \begin{bmatrix}\overline{\mathbf{n}}_e[k]\\ \mathbf{n}_o[k]\end{bmatrix}. \tag{36}$$

Without loss in generality, we can transform an $M_r \times M_t$ channel matrix to an $M \times M$ channel matrix by appending zero columns/rows where $M = \max(M_r, M_t)$. Thus, we restrict attention to square channel matrices. We can also assume that $\rho_e = \rho_o = \rho$ without loss in generality to simplify the convergence studies. While establishing a convergence result under the general Rayleigh fading model appears difficult, we now establish this under certain restrictions. Nevertheless, numerical studies in Section VI show that convergence of the summed power method holds true even for general channel matrix settings. These assumptions (listed as Hypotheses 1-3) are as follows:

- **Hypothesis 1:** Since convergence studies make more sense in the high-SNR regime, we assume that $\rho \gg 1$.
- **Hypothesis 2:** Let $\mathbf{f}[i] = [f_{i,1}, \cdots, f_{i,M}]^{\mathsf{T}}$ and $\mathbf{z}[i] = [z_{i,1}, \cdots, z_{i,M}]^{\mathsf{T}}$. We make the assumptions that as k increases, $\sum_{i=0}^{k} f_{i,n} \approx \mathsf{C}_k$ for all n and $\sum_{i=0}^{k} z_{i,m} \approx \mathsf{C}_k$ for all m. In other words, the statistics of the beamformers remain invariant to the antenna indices at either node as k increases.
- **Hypothesis 3:** We consider real-valued, diagonal channel matrices $\mathbf{H} = \mathrm{diag}\left([h_1, \cdots, h_M]\right)$ with diagonal elements ordered in non-increasing order. These assumptions can be viewed as restricting all the signal processing to happen within the bases corresponding to the left- and right-singular vectors of $\mathbf{H}$. Also, assume that $h_1 > h_2$ implying a singular dominant eigen-mode for $\mathbf{H}$.

We now discuss the behavior of the summed power method as k (the number of iterations) increases under the above assumptions. Under Hypothesis 3, it can be seen that the optimal beamformers reduce to a scaled version of the first column of the $M \times M$-dimensional identity matrix, denoted as $\mathbf{e}_1$. Thus, the desired state vector is $\mathbf{s}_{\mathsf{opt}} = \left[\alpha \mathbf{e}_1^{\mathsf{T}} \ \beta \mathbf{e}_1^{\mathsf{T}}\right]^{\mathsf{T}} = \begin{bmatrix} \alpha & 0 & \cdots & 0 & \beta & 0 & \cdots & 0 \end{bmatrix}^{\mathsf{T}}$ for some α and β. The impreciseness in the choice of α and β is because the beamforming vector is defined only up to a point on the Grassmann manifold [18,19,22].

Convergence of the summed power method is equivalent to the limiting behavior/convergence of $\mathbf{s}[k]$ from (33) to $\mathbf{s}_{\mathsf{opt}}$. Lemma 2 provides a preliminary result needed to establish this convergence result.

Lemma 2. *Under Hypothesis 3, the state transition matrix from (35) is diagonalized by*

$$\mathbf{U}_{k-1} = \begin{bmatrix} \sqrt{\frac{\beta_{k-1}}{\alpha_{k-1}+\beta_{k-1}}}\mathbf{I} & \sqrt{\frac{\beta_{k-1}}{\alpha_{k-1}+\beta_{k-1}}}\mathbf{I} \\ \sqrt{\frac{\alpha_{k-1}}{\alpha_{k-1}+\beta_{k-1}}}\mathbf{I} & -\sqrt{\frac{\alpha_{k-1}}{\alpha_{k-1}+\beta_{k-1}}}\mathbf{I} \end{bmatrix}. \quad (37)$$

For the proof, see Appendix D.

Note that $\mathbf{U}_{k-1}$ is not unitary for general α_{k-1} and β_{k-1}. However, we have the following additional result that simplifies $\mathbf{U}_{k-1}$.

Lemma 3. *Under Hypotheses 1-3, we can assume that $\alpha_k \approx \beta_k$ for large k. Thus, as k increases, $\mathbf{U}_{k-1}$ converges to*

$$\widetilde{\mathbf{U}} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix}. \quad (38)$$

For the proof, see Appendix E.

We now have the following main result.

Theorem 4. *Under Hypotheses 1-3, we have that $\mathbf{s}[k] \to \mathbf{s}_{\mathsf{opt}}$ as k increases.*

For the proof, see Appendix F.

The results of Section VI will show that these results hold for more general channel models and are not restricted to satisfaction of Hypotheses 1-3. In addition, Section V presents two modifications to the summed power method which aim to improve performance over a wider range of SNRs while maintaining low computational complexity.

V. Least-Squares Initialized Summed Power Method

We now consider a refinement that trades off the advantages of both the approaches in Sections III and IV in terms of complexity, feedback and performance. The main motivation behind this approach is the observation that the performance of a beam alignment algorithm critically depends on how $\mathbf{f}[0]$ (or $\mathbf{z}[0]$) is initialized. When $\mathbf{f}[0]$ is initialized as a complex random unit-norm vector, we rely on multiple iterations over the channel to re-align this choice towards the singular vectors of the channel. Depending on the approach used for alignment as well as the SNR on the downlink and uplink, the beam alignment algorithm could take a substantial number of iterations to improve the effective channel gain.

In this context, we note that the (sequential/batch) least-squares approach from Section III achieves good performance in the high-SNR regime by optimally estimating the channel matrix over every iteration and re-aligning the alignment problem at every step. However, this gain comes at the cost of complexity and feedback overhead of the algorithm. On the other hand, at low-SNR, averaging over the noise results in significant performance improvement with the summed power method from Section IV, which is a low-complexity/feedback overhead scheme.

These observations suggest that the two approaches can be married together, which is the focus of the *least-squares initialized summed power (LISP) method*. In this method, both nodes "prime" their beamformers using either the batch/sequential least-squares method for the first k_{switch} iterations, after which they switch to the summed power method. In particular, we have the following description in Algorithm 4 for the proposed technique with the sequential least-squares initialization. The switching point k_{switch} can be chosen in multiple ways. Specific choices for k_{switch} include $\min(M_r, M_t)$, $\max(M_r, M_t)$ or via some optimality studies as in Sec. VI.

VI. Numerical Studies

In this section, we present performance comparisons of the proposed schemes obtained via Monte Carlo experiments. We first present results on the convergence properties of the different techniques under varying conditions. We then present the impact of an increase in M_t on the performance of these schemes.

Algorithm 4 Least-squares Initialized Summed Power Method

Initialize $\mathbf{f}[0]$ as a complex random unit-norm vector and obtain $\widehat{\mathbf{H}}_{o,0}$, $\mathbf{z}[0]$, $\widehat{\mathbf{H}}_{e,1}$ and $\mathbf{f}[1]$ as in (23)-(26).
Initialize $\mathbf{C}_{o,0} = \mathbf{C}_{e,1} = \alpha\mathbf{I}$ for an appropriate α.
for all $k = 1, \ldots, k_{\mathsf{max}}$ **do**
  if $k \leq k_{\mathsf{switch}}$ **then**
    Node 2 receives $\mathbf{y}_o[k]$ as in (3) and gets $\mathbf{f}[k]$ from Node 1
    Node 2 estimates $\widehat{\mathbf{H}}_{o,k}$ as in (20)
    Node 2 computes $\mathbf{z}[k]$ as in (14)
  else if $k > k_{\mathsf{switch}}$ **then**
    Node 2 receives $\mathbf{y}_o[k-1]$ as in (3)
    Node 2 computes $\mathbf{z}[k]$ as in (29)
  end if

  if $k \leq k_{\mathsf{switch}} - 1$ **then**
    Node 1 receives $\mathbf{y}_e[k]$ as in (4) and gets $\mathbf{z}[k]$ from Node 2
    Node 1 estimates $\widehat{\mathbf{H}}_{e,k+1}$ as in (17)
    Node 1 computes $\mathbf{f}[k+1]$ as in (13)
  else if $k > k_{\mathsf{switch}} - 1$ **then**
    Node 1 receives $\mathbf{y}_e[k]$ as in (4)
    Node 1 computes $\mathbf{f}[k+1]$ as in (27)
  end if
end for

A. Convergence Studies

We study two variants of the proposed sequential least-squares technique from Section III: "SLS (Optimal)" and "SLS (Suboptimal)." The first variant computes the batch least-squares estimator for the first M_r (or M_t) iterations before switching to the sequential version after that. The second variant relies on the result from Theorem 3 to be computationally efficient and to avoid having to compute the batch estimator. It is initialized with $\alpha = 1000$ and uses the sequential estimator starting at the first iteration. We also study the performance of the iterative solutions based on the summed power method from Section IV and the least-squares initialized summed power method with $k_{\mathsf{switch}} = \max(M_r, M_t)$ from Section V. These approaches are denoted as "Summed Power" and "LISP" in the plots, respectively.

In terms of performance benchmarking, we consider the one-dimensional versions of the techniques proposed in [27] and [28]. The algorithm from [27] is called *Blind Iterative MIMO Algorithm (BIMA)* by the authors and is denoted as "BIMA" in the plots here. The algorithm from [28] is called *Best Singular Mode (BSM) estimation* by the authors and is denoted as "BSM" in the plots here. The value of the design parameter μ for the BSM algorithm from [28] is set to $1.5k$ where k is the time-index.

In Figs. 2-4, we compare the performance of these six schemes at different SNR values with $M_r = 4$ and $M_t = 32$ (corresponding to a downlink channel matrix $\mathbf{H}$ of dimensions 4×32). The channel matrix $\mathbf{H}$ has i.i.d. entries. In particular, Figs. 2a and 2b show the results for uplink and downlink SNR values of -10 dB, whereas, Figs. 3a and 3b, and Figs. 4a and 4b provide similar plots for an SNR of 0 dB and 20 dB, respectively. These SNR values are expected to be typical of low-, medium- and high-SNR regimes, respectively.

Practical mmWave channels are expected to be sparser [32, 36] than i.i.d. channels. In this context, Fig. 5 illustrates the performance of the same set of six schemes in a sparse MIMO channel model with $\lambda/2$ spaced uniform linear arrays (ULAs) at both ends corresponding to $M_r = 4$ and $M_r = 32$. Both downlink and uplink SNRs are assumed to be -10 dB and $f_{\mathsf{c}} = 28$ GHz is used. The channel is made of $K = 3$ dominant clusters with one path per cluster (hence the channel matrix $\mathbf{H}$ is rank-deficient). The angles of arrival and departure are assumed to be in the azimuth plane and uniformly distributed in a $120°$ angular spread at both ends. Rayleigh fading is assumed for the path gains. Such a model is commonly used in mmWave system analysis (see [36] and references therein for details).

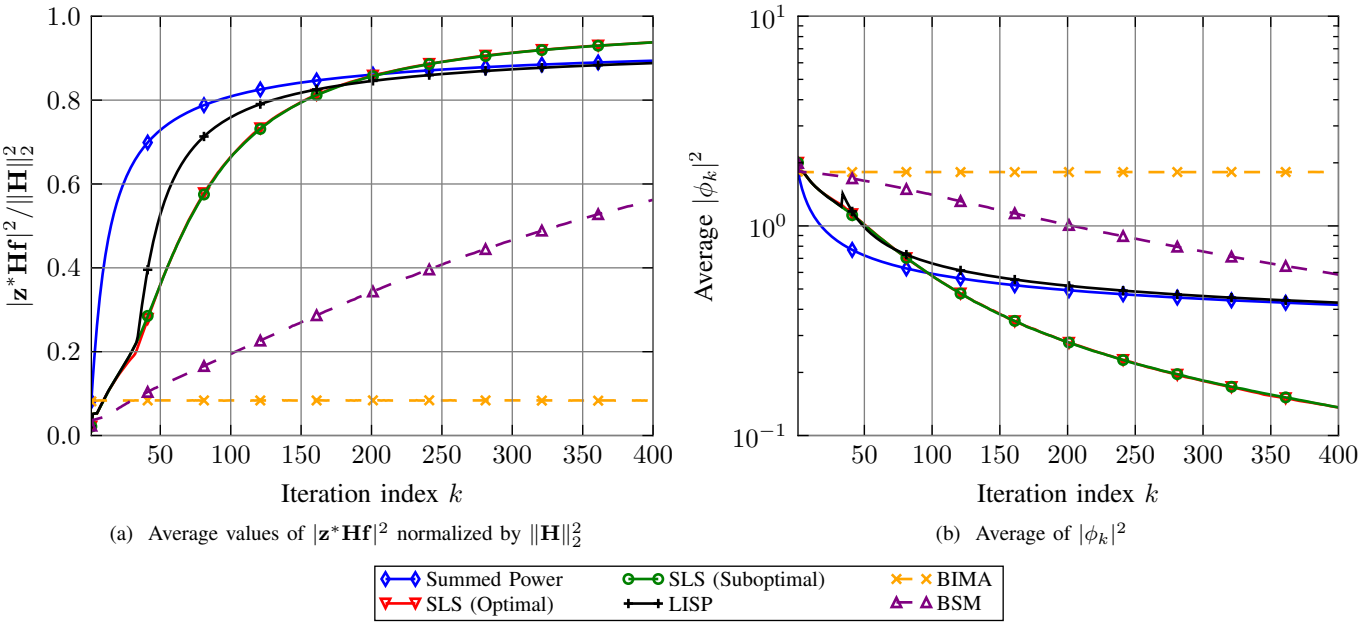
We study two metrics capturing the performance of these six schemes: i) the instantaneous effective channel gain $|\mathbf{z}^*[k]\mathbf{H}\mathbf{f}[k]|^2$ at time-index k, and ii) the angle between the true singular vector $\mathbf{f}_{\mathsf{opt}}$ and its estimate $\mathbf{f}[k]$, given as

$$\phi_k = \cos^{-1}\left(|\mathbf{f}_{\mathsf{opt}}^*\mathbf{f}[k]|\right), \tag{39}$$

and measured in radians. Note that ϕ_k equivalently captures the chordal distance between $\mathbf{f}_{\mathsf{opt}}$ and $\mathbf{f}[k]$. In order to average results over different channel realizations, we normalize the effective channel gain by $\|\mathbf{H}\|_2^2 = \lambda_{\mathsf{max}}(\mathbf{H}^*\mathbf{H})$. Fast convergence of the algorithm is then equivalent to fast convergence of the normalized instantaneous effective channel gain to 1.

From Figs. 2-4, we make the following remarks:

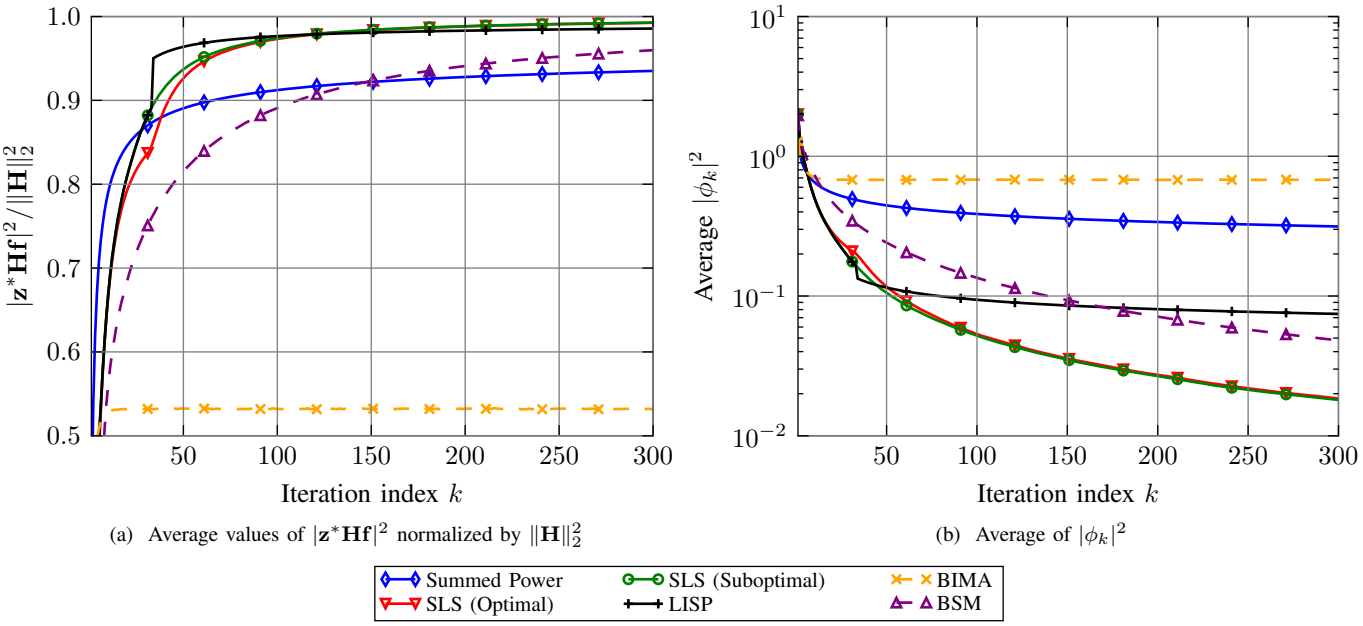
1) There is a minor performance gap (both in terms of gains and angles) between the optimal and suboptimal variants of the SLS estimator across all the three SNRs, even though there is a significant complexity reduction with the suboptimal variant. Thus, this study motivates the use of the suboptimal variant of the SLS estimator over the optimal variant.
2) In the low-SNR regime typical of mmWave settings, the summed power method significantly outperforms all the methods for small k values, whereas the additional channel estimation step of the SLS estimator contributes to its utility for large k values. While the method from [28] is better in performance than the one from [27], neither method produces a performance comparable to the schemes proposed in this work.
3) The performance of the schemes in [28] and [27] improve with SNR. In the high-SNR regime, both methods become comparable to the SLS estimator. However, the summed power method is significantly inferior in this regime as it cannot suppress the effect of noise from the beamformer estimates for large k values.
4) The switching between the SLS part and the summed power part means that the LISP method shows a switch in terms of performance at $k = k_{\mathsf{switch}} = \max(M_r, M_t) = 32$. But more importantly, in the low-SNR regime, the LISP method approaches the performance of the summed power method for large k and in the high-SNR regime, it approaches the performance of the SLS estimator (even for small k values) without



(a) Average values of $|\mathbf{z}^*\mathbf{H}\mathbf{f}|^2$ normalized by $\|\mathbf{H}\|_2^2$

(b) Average of $|\phi_k|^2$

Fig. 2. Results for the i.i.d. channel model at $\rho_e = \rho_o = -10$ dB with $M_r = 4, M_t = 32$



(a) Average values of $|\mathbf{z}^*\mathbf{H}\mathbf{f}|^2$ normalized by $\|\mathbf{H}\|_2^2$

(b) Average of $|\phi_k|^2$

Fig. 3. Results for the i.i.d. channel model at $\rho_e = \rho_o = 0$ dB with $M_r = 4, M_t = 32$

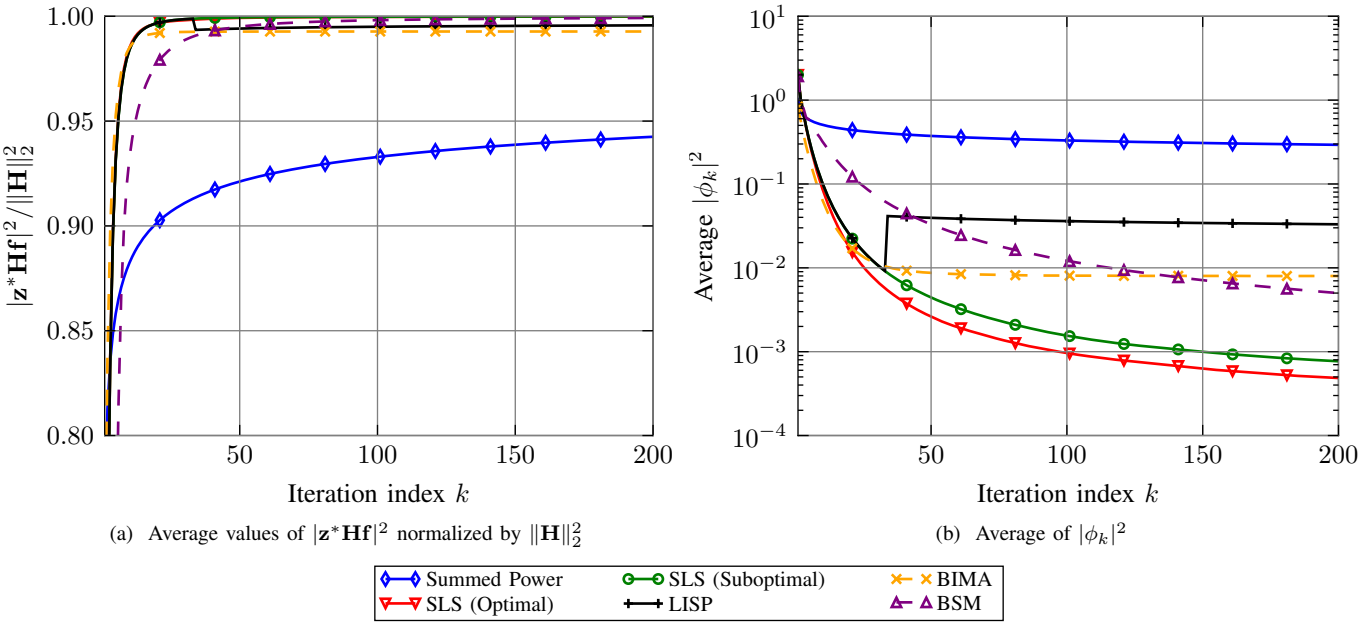
the additional complexity overhead of these methods. Thus, this method may be a suitable low-complexity alternative to the SLS estimator in the medium- to high-SNR regime.

5) In the sparse mmWave setting with low SNR, the summed power method outperforms all the methods over all the values of k considered here. The LISP method quickly approaches the performance of the summed power method after $k = k_{\mathsf{switch}}$.

Summarizing the above statements, we have the following conclusions: i) In the low-SNR regime, the summed power method is advantageous for small k and the SLS estimator is advantageous for large k. If computational complexity is an important issue for large k, the LISP method can be a useful alternative. ii) In the high-SNR regime, the LISP method or the method proposed in [27] are advantageous for all k. iii) These broad conclusions appear to be true for both i.i.d. as well as sparse mmWave channel models.

B. Impact of Antenna Dimensions and k_{switch}

Fig. 6 studies the impact of M_t (as M_t increases from 6 to 64) on the effective channel gain after $k = 100$ iterations with the different beam alignment techniques. The low-SNR regime corresponding to $\rho_e = \rho_o = -10$ dB and $M_r = 4$ is considered in this study. Figs. 6a and 6b present the results



(a) Average values of $|\mathbf{z}^*\mathbf{H}\mathbf{f}|^2$ normalized by $\|\mathbf{H}\|_2^2$

(b) Average of $|\phi_k|^2$

Fig. 4. Results for the i.i.d. channel model at $\rho_e = \rho_o = 20$ dB with $M_r = 4, M_t = 32$

for the i.i.d. channel model and the sparse mmWave channel model introduced earlier.

This study reinforces the advantages of the summed power and least-squares initialized summed power methods relative to other methods. In particular, the performance of the summed power method remains approximately invariant in the i.i.d. case as M_t increases. On the other hand, the smaller rank of the channel matrix in the sparse case improves the fraction of power in the dominant eigen-mode, which is reflected in improving performance as M_t increases. But more importantly, the performance of all other schemes depreciate with M_t suggesting their sensitivity to larger antenna dimensions. Nevertheless, the LISP method appears closest to the summed power method in performance at low-SNR and is also superior at high-SNRs. From these results, we conclude that the proposed beam alignment techniques and in particular, the LISP method can deliver substantial performance improvement as M_t increases with low complexity and feedback overheads making them viable candidates for practical large/massive MIMO systems.

Figs. 7 and 8 study the choice of k_{switch} to be used in the LISP method with $M_r = 4$, $M_t = 32$ and $\rho_e = \rho_o = 0$ dB and $\rho_e = \rho_o = -10$ dB, respectively. From Fig. 7, we note that there exists an optimal k_{switch} that maximizes the effective channel gain for both the i.i.d. and sparse mmWave channel models. The optimal k_{switch} value is typically small in the case of sparse mmWave channels for both SNR settings. In fact, for $\rho_e = \rho_0 = -10$ dB, the optimal k_{switch} in the sparse setting is 1 implying that the summed power method starting at $k = 1$ is better than a noisy initialization based on the SLS estimator. While the optimal k_{switch} can be high in the i.i.d. setting, constraining it to be a small number does not result in a significantly poorer performance relative to the optimal k_{switch} value. Thus, Figs. 7 and 8 suggest that, in the moderate- to high-SNR regime and depending on the level of richness/sparsity structure of the channel, a small k_{switch} may be a better choice than the use of summed power method ($k_{\mathsf{switch}} = 1$). Thus, an improved performance can be ensured with the LISP method at the cost of a small feedback and complexity overhead.

C. Comparison with a Pilot-Based Channel Estimation Scheme

We are now interested in comparing the performance of the proposed beam alignment schemes with a traditional pilot-based channel estimation scheme. In order to simplify the structure of the pilot-based scheme, we assume that the channel matrices are i.i.d. Rayleigh fading. In order to fairly compare the iterative schemes with the batch-oriented pilot-based scheme, we impose a constraint on the total energy used during the beam alignment/channel estimation phase. Let k_{max} be the number of time slots allocated for this phase. With the iterative schemes considered in this work, the total energy used by nodes 1 and 2 reduces to $\rho_o \cdot k_{\mathsf{max}}$ and $\rho_e \cdot k_{\mathsf{max}}$, respectively. With the pilot-based scheme, it is well understood [23,24,51] that the quality of the channel estimate *only* depends on the energy in the training matrices (denoted as $\mathbf{P}_o$ and $\mathbf{P}_e$ for the downlink and uplink, respectively) as long as the number of pilot symbols exceeds the transmit antenna dimensions. Thus, we can assume that $\mathbf{P}_o$ and $\mathbf{P}_e$ are $M_t \times M_t$ and $M_r \times M_r$ scaled-unitary matrices meeting the energy constraint, respectively. With the energy scaling, we have the following system equations:

$$\mathbf{Y}_o = \sqrt{\rho_o \cdot k_{\mathsf{max}}/M_t}\,\mathbf{H}\mathbf{P}_o + \mathbf{N}_o \tag{40}$$

for the downlink, and

$$\mathbf{Y}_e = \sqrt{\rho_e \cdot k_{\mathsf{max}}/M_r}\,\mathbf{H}^{\mathsf{T}}\mathbf{P}_e + \mathbf{N}_e \tag{41}$$

for the uplink.

Upon reception of $\mathbf{Y}_o$ and $\mathbf{Y}_e$, each node computes a minimum mean-squared error (MMSE) channel estimate as follows:

$$\widehat{\mathbf{H}}_o = \frac{\sqrt{\rho_o \cdot k_{\mathsf{max}}/M_t}}{1+\rho_o \cdot k_{\mathsf{max}}/M_t} \cdot \mathbf{Y}_o \mathbf{P}_o^* \tag{42}$$

$$\widehat{\mathbf{H}}_e = \frac{\sqrt{\rho_e \cdot k_{\mathsf{max}}/M_r}}{1+\rho_e \cdot k_{\mathsf{max}}/M_r} \cdot \mathbf{Y}_e \mathbf{P}_e^*. \tag{43}$$

The beamformers are estimated using the SVD of the channel estimates. In our study, we use scaled discrete Fourier transform (DFT) matrices for $\mathbf{P}_o$ and $\mathbf{P}_e$ over the i.i.d. channel. With $k_{\mathsf{max}} = 100$, the normalized channel gain across different SNR values is plotted for the different schemes in Fig. 9. These results show that in addition to outperforming iterative schemes from prior works in the low-SNR regime, the proposed methods also compare favorably to the pilot-based channel estimation scheme. The pilot-based scheme requires a substantial pre-beamforming SNR (over 5-10 dB) for improved performance which may not be feasible in practical mmWave systems. Further, it also requires a computational overhead in computing the SVD of the channel estimate.

VII. Concluding Remarks

This paper studied the problem of estimating the dominant singular vectors of a MIMO channel matrix in a TDD system. Such a task is of importance in realizing the full analog beamforming gains in practical mmWave systems, typically impaired with low SNR. We presented multiple iterative approaches based on the power method to address this problem. These approaches included batch and sequential least-squares estimation, summed power method, and least-squares initialized summed power method. Numerical studies and analysis established that the proposed approaches enjoy several advantages over competing approaches from the literature. These advantages include improved convergence and/or performance (beamforming gain) at low- as well as high-SNR at a low-complexity and feedback overhead.

That said, this paper has only scratched the surface of the noisy beam alignment problem. Further studies on developing an analytical/manifold optimization-based framework for the rate of convergence of the proposed algorithms as a function of the SNR, antenna dimensions, mmWave channel eigen-mode/sparsity structure, etc. are important. Such a step could also be of independent interest in problems in machine learning, principal component analysis, and linear algebra. Other problems of interest include understanding the impact of an imperfect (e.g., finite-rate, noisy, etc.) feedback link on the performance of the sequential least-squares estimation scheme, performance comparison with other directional learning approaches [36], impact of temporal variation in the channel and wideband aspects on the performance of the proposed schemes, extending the proposed analog beamforming schemes to a hybrid architectural set-up or multi-user settings [37], intuitive understanding of k_{switch} and further optimization of the beam alignment parameters given an asymmetrical antenna setting in the single-user case, etc.

Appendix

A. Proof of Theorem 1

The derivation of (16) mirrors (15) and thus it suffices to establish (15). Transposing (6) at time $k-1$, we get

$$\mathbf{Y}_{e,k-1}^{\mathsf{T}} = \sqrt{\rho_e}\, \mathbf{Z}_{k-1}^* \mathbf{H} + \mathbf{N}_{e,k-1}^{\mathsf{T}}. \tag{44}$$

Since the columns of $\mathbf{H}$ are i.i.d. complex Gaussian random vectors, we focus on the first column without loss in generality. Let this first column of $\mathbf{H}$ be denoted as $\mathbf{h}_1$ and let its estimator be $\widehat{\mathbf{h}}_1$. With $\widetilde{\mathbf{y}}$ denoting the first column of $\mathbf{Y}_{e,k-1}^{\mathsf{T}}$, we have

$$\widetilde{\mathbf{y}} = \sqrt{\rho_e}\, \mathbf{Z}_{k-1}^* \mathbf{h}_1 + \widetilde{\mathbf{n}}, \tag{45}$$

where $\widetilde{\mathbf{n}}$ is the first column of $\mathbf{N}_{e,k-1}^{\mathsf{T}}$ with i.i.d. complex Gaussian entries. The estimator of $\mathbf{h}_1$ is given as

$$\widehat{\mathbf{h}}_1 = \frac{\left(\mathbf{Z}_{k-1}^*\right)^\dagger \widetilde{\mathbf{y}}}{\sqrt{\rho_e}} \tag{46}$$

with error covariance matrix $\mathbf{C}_{e,k}$, defined as, $\mathbf{C}_{e,k} \triangleq \mathrm{E}\left\{\left(\mathbf{h}_1 - \widehat{\mathbf{h}}_1\right)\left(\mathbf{h}_1 - \widehat{\mathbf{h}}_1\right)^*\right\}$.

It can be seen that

$$\mathbf{h}_1 - \widehat{\mathbf{h}}_1 = \mathbf{h}_1 - \frac{\left(\mathbf{Z}_{k-1}^*\right)^\dagger}{\sqrt{\rho_e}} \left(\sqrt{\rho}_e \mathbf{Z}_{k-1}^* \mathbf{h}_1 + \widetilde{\mathbf{n}}\right) \tag{47}$$

$$= \frac{-\left(\mathbf{Z}_{k-1}^*\right)^\dagger}{\sqrt{\rho_e}} \widetilde{\mathbf{n}} \tag{48}$$

and

$$\mathbf{C}_{e,k} = \frac{\left(\mathbf{Z}_{k-1}^*\right)^\dagger}{\sqrt{\rho_e}} \mathrm{E}\left\{\widetilde{\mathbf{n}}\widetilde{\mathbf{n}}^*\right\} \frac{\left(\mathbf{Z}_{k-1}\right)^\dagger}{\sqrt{\rho_e}} \tag{49}$$

$$= \frac{1}{\rho_e} \left(\mathbf{Z}_{k-1}\mathbf{Z}_{k-1}^*\right)^{-1} \mathbf{Z}_{k-1}\mathbf{Z}_{k-1}^* \left(\mathbf{Z}_{k-1}\mathbf{Z}_{k-1}^*\right)^{-1} \tag{50}$$

$$= \frac{1}{\rho_e} \left(\mathbf{Z}_{k-1}\mathbf{Z}_{k-1}^*\right)^{-1}. \tag{51}$$

Note that the above equation holds only under the i.i.d. $\mathbf{H}$ assumption and for $k \geq M_r$. ∎

B. Proof of Theorem 2

Without loss in generality, we can assume that $\rho_e = \rho_o = 1$. From (10), we have

$$\widehat{\mathbf{H}}_{o,k}^{\mathrm{Batch}}$$

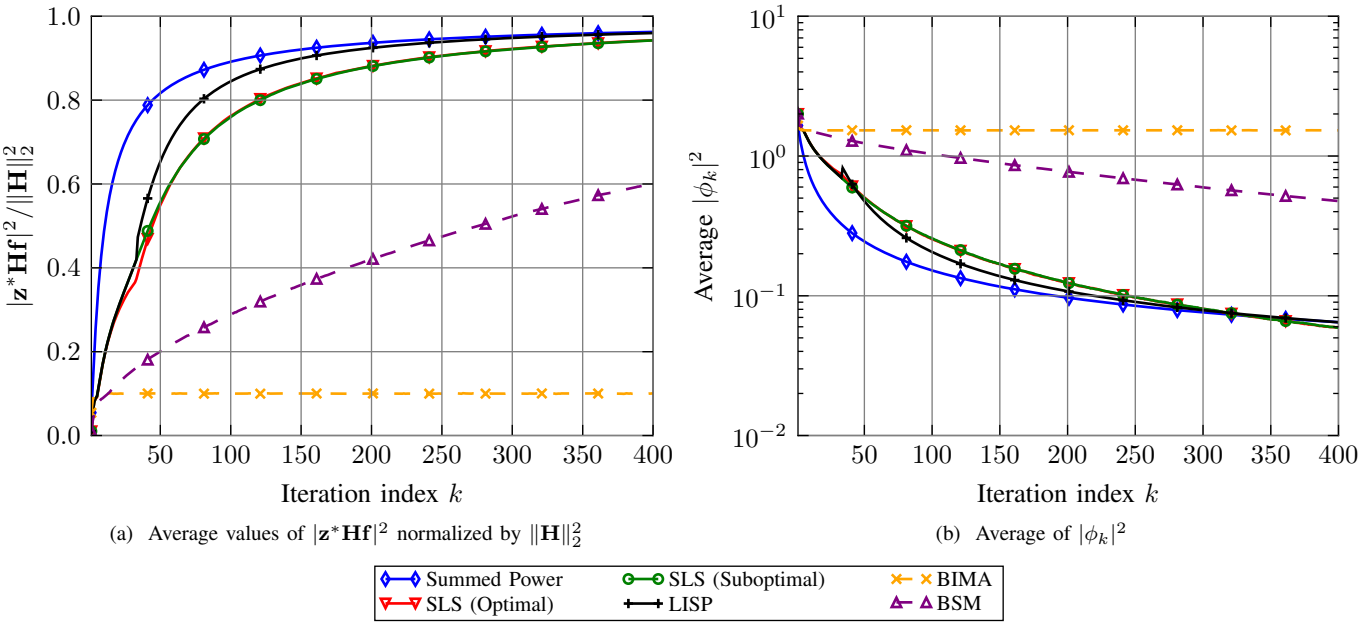
$$= \mathbf{Y}_{o,k}\mathbf{F}_k^*\mathbf{C}_{o,k} = \mathbf{Y}_{o,k}\mathbf{F}_k^*\left(\mathbf{F}_k\mathbf{F}_k^*\right)^{-1} \tag{52}$$

$$= \begin{bmatrix}\mathbf{Y}_{o,k-1} & \mathbf{y}_o[k]\end{bmatrix} \begin{bmatrix}\mathbf{F}_{k-1}^* \\ \mathbf{f}^*[k]\end{bmatrix} \left(\begin{bmatrix}\mathbf{F}_{k-1} & \mathbf{f}[k]\end{bmatrix} \begin{bmatrix}\mathbf{F}_{k-1}^* \\ \mathbf{f}^*[k]\end{bmatrix}\right)^{-1} \tag{53}$$

$$= \left[\mathbf{Y}_{o,k-1}\mathbf{F}_{k-1}^* + \mathbf{y}_o[k]\mathbf{f}^*[k]\right] \left(\mathbf{F}_{k-1}\mathbf{F}_{k-1}^* + \mathbf{f}[k]\mathbf{f}^*[k]\right)^{-1}. \tag{54}$$

Substituting (16) and applying the Woodbury matrix identity [48] to the second term, we get

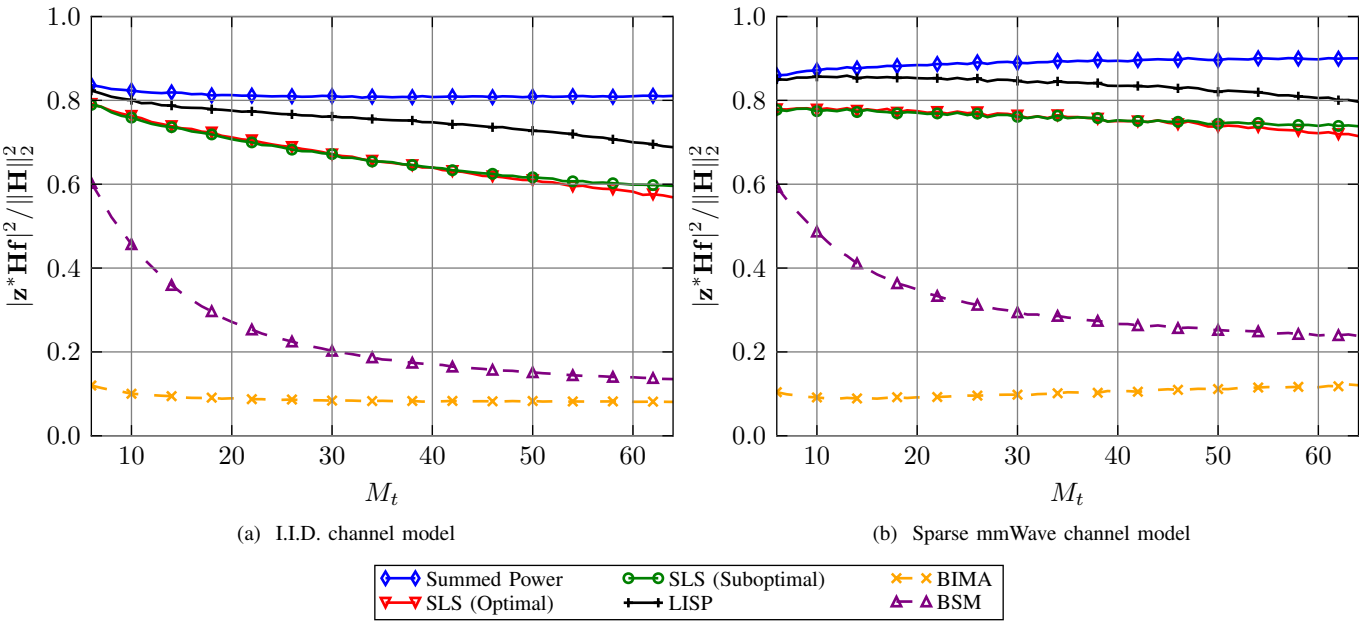
$$\mathbf{C}_{o,k} = \mathbf{C}_{o,k-1} - \frac{\mathbf{C}_{o,k-1}\mathbf{f}[k]\mathbf{f}^*[k]\mathbf{C}_{o,k-1}}{1+\mathbf{f}^*[k]\mathbf{C}_{o,k-1}\mathbf{f}[k]}. \tag{55}$$



(a) Average values of $|\mathbf{z}^*\mathbf{H}\mathbf{f}|^2$ normalized by $\|\mathbf{H}\|_2^2$

(b) Average of $|\phi_k|^2$

Fig. 5. Results for the sparse mmWave channel model at $\rho_e = \rho_o = -10$ dB with $M_r = 4, M_t = 32$



(a) I.I.D. channel model

(b) Sparse mmWave channel model

Fig. 6. Average value of $|\mathbf{z}^*\mathbf{H}\mathbf{f}|^2$ normalized by $\|\mathbf{H}\|_2^2$ at $k = 100$ for $\rho_e = \rho_o = -10$ dB using different channel models with $M_r = 4$ and $M_t \in 6, 8, 10, \cdots, 64$.

We now let

$$\mathbf{K}_{o,k} = \frac{\mathbf{f}^*[k]\mathbf{C}_{o,k-1}}{1 + \mathbf{f}^*[k]\mathbf{C}_{o,k-1}\mathbf{f}[k]} \tag{56}$$

and write

$$\begin{aligned}\widehat{\mathbf{H}}_{o,k}^{\text{Batch}} &= \mathbf{Y}_{o,k-1}\mathbf{F}_{k-1}^*\mathbf{C}_{o,k-1} - \mathbf{Y}_{o,k-1}\mathbf{F}_{k-1}^*\mathbf{C}_{o,k-1}\mathbf{f}[k]\mathbf{K}_{o,k} \\ &\quad + \mathbf{y}_o[k]\mathbf{f}^*[k]\mathbf{C}_{o,k-1} - \mathbf{y}_o[k]\mathbf{f}^*[k]\mathbf{C}_{o,k-1}\mathbf{f}[k]\mathbf{K}_{o,k}.\end{aligned} \tag{57}$$

Now, since

$$\mathbf{y}_o[k]\mathbf{f}^*[k]\mathbf{C}_{o,k-1} = \mathbf{y}_o[k]\left(1 + \mathbf{f}^*[k]\mathbf{C}_{o,k-1}\mathbf{f}[k]\right)\mathbf{K}_{o,k}, \tag{58}$$

we get

$$\widehat{\mathbf{H}}_{o,k}^{\text{Batch}} = \widehat{\mathbf{H}}_{o,k-1}\left(\mathbf{I} - \mathbf{f}[k]\mathbf{K}_{o,k}\right) + \mathbf{y}_o[k]\mathbf{K}_{o,k} \tag{59}$$

$$= \widehat{\mathbf{H}}_{o,k-1} + \left(\mathbf{y}_o[k] - \widehat{\mathbf{H}}_{o,k-1}\mathbf{f}[k]\right)\mathbf{K}_{o,k} \tag{60}$$

$$= \widehat{\mathbf{H}}_{o,k}^{\text{Seq}}. \tag{61}$$

■

C. Proof of Theorem 3

Along the same lines of the proof of Theorem 2, let us assume that $\rho_e = \rho_o = 1$. Suppose that node 2 has access to M_t previous observations at time slot $k = 0$, indexed from

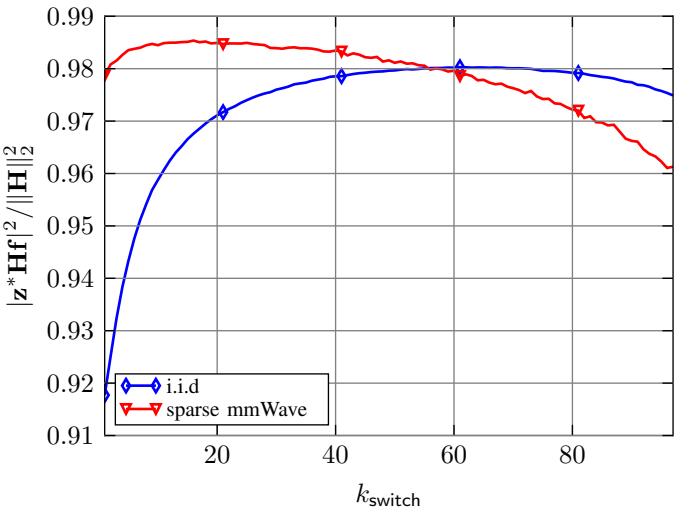


Fig. 7. Normalized channel gain of the LISP method as a function of k_{switch} with $M_r = 4$, $M_t = 32$, $\rho_e = \rho_o = 0$ dB and $k_{\text{max}} = 100$.

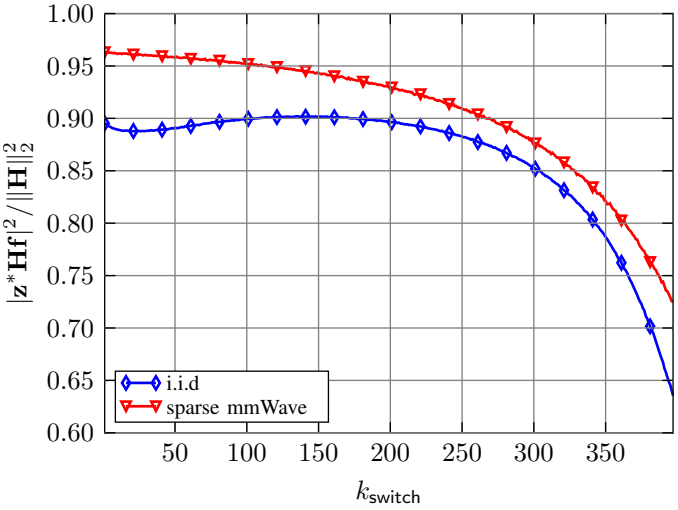


Fig. 8. Normalized channel gain of the LISP method as a function of k_{switch} with $M_r = 4$, $M_t = 32$, $\rho_e = \rho_o = -10$ dB and $k_{\text{max}} = 400$.

$k = -(M_t - 1)$ to $k = 0$. Using this data, node 2 could thus compute the batch estimate at time slot $k = 0$, given as

$$\widehat{\mathbf{H}}_{o,0}^{\text{Batch}} = \mathbf{Y}_{o,0}\mathbf{F}_0^{\dagger} = \mathbf{Y}_{o,0}\mathbf{F}_0^*\left(\mathbf{F}_0\mathbf{F}_0^*\right)^{-1}, \tag{62}$$

where $\mathbf{F}_0 = \begin{bmatrix}\mathbf{f}[-(M_t-1)] & \mathbf{f}[-M_t] & \dots & \mathbf{f}[0]\end{bmatrix}$ and $\mathbf{Y}_{o,0} = \begin{bmatrix}\mathbf{y}_o[-(M_t-1)] & \mathbf{y}_o[-M_t] & \dots & \mathbf{y}_o[0]\end{bmatrix}$. Using Theorem 1, we note that the covariance matrix of each column of this estimated matrix is given as

$$\mathbf{C}_{o,0} = \left(\mathbf{F}_0\mathbf{F}_0^*\right)^{-1}. \tag{63}$$

Applying the result of Theorem 2, we note that for any $k > 0$, a sequential least-squares estimator would be identical to the batch estimator using all of the data from $k = -(M_t - 1)$ up

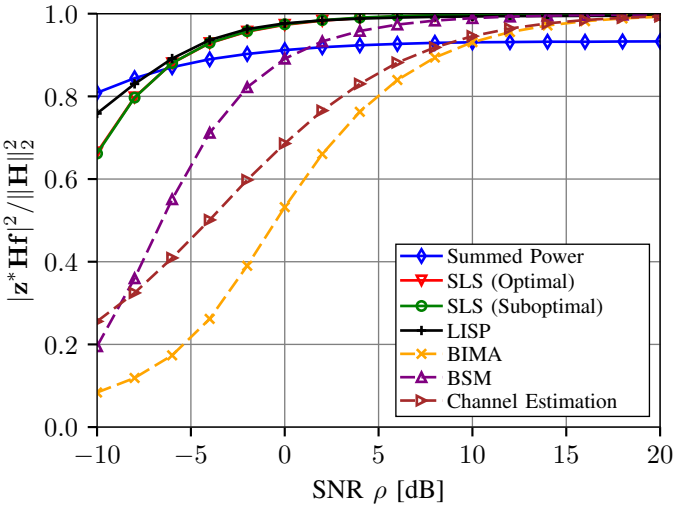


Fig. 9. Normalized channel gain with $k_{\text{max}} = 100$ for varying values of $\rho = \rho_o = \rho_e$ in the i.i.d. Rayleigh fading channel case.

to k. We can thus write

$$\widehat{\mathbf{H}}_{o,k}^{\text{Seq}} = \widehat{\mathbf{H}}_{o,k}^{\text{Batch}} = \mathbf{Y}_{o,k}\mathbf{F}_k^{\dagger} \tag{64}$$

$$= \left(\sum_{n=-(M_t-1)}^{k}\mathbf{y}_o[n]\mathbf{f}^*[n]\right)\left(\sum_{n=-(M_t-1)}^{k}\mathbf{f}[n]\mathbf{f}^*[n]\right)^{-1}, \tag{65}$$

where we have rewritten $\widehat{\mathbf{H}}_{o,k}^{\text{Batch}}$ in terms of individual vector outer products. After separating the hypothetical data from $k = -(M_t - 1)$ to $k = 0$ from the data starting at $k = 1$, we have for the sequential estimator

$$\widehat{\mathbf{H}}_{o,k}^{\text{Seq}} = \left(\sum_{n=-(M_t-1)}^{0}\mathbf{y}_o[n]\mathbf{f}^*[n] + \sum_{n=1}^{k}\mathbf{y}_o[n]\mathbf{f}^*[n]\right) \cdot \left(\sum_{n=-(M_t-1)}^{0}\mathbf{f}[n]\mathbf{f}^*[n] + \sum_{n=1}^{k}\mathbf{f}[n]\mathbf{f}^*[n]\right)^{-1} \tag{66}$$

$$= \left(\mathbf{Y}_{o,0}\mathbf{F}_0^* + \mathbf{Y}_{o,k}\mathbf{F}_k^*\right)\left(\mathbf{F}_0\mathbf{F}_0^* + \mathbf{F}_k\mathbf{F}_k^*\right)^{-1} \tag{67}$$

$$= \left(\widehat{\mathbf{H}}_{o,0}\mathbf{C}_{o,0}^{-1} + \mathbf{Y}_{o,k}\mathbf{F}_k^*\right)\left(\mathbf{C}_{o,0}^{-1} + \mathbf{F}_k\mathbf{F}_k^*\right)^{-1}. \tag{68}$$

Upon further inspection of (66), we observe that for any $k > M_t$, the product $\mathbf{F}_k\mathbf{F}_k^*$ is invertible and we can let $\mathbf{C}_{o,0}^{-1}$ arbitrarily approach the matrix of all zeros. This can be accomplished by choosing $\mathbf{C}_{o,0} = \alpha\mathbf{I}$. If $\mathbf{C}_{o,0}$ is indeed chosen this way, (66) loses its dependence on the previous data from the supposition and we can start the sequential iteration at $k = 1$. For sufficiently large α, the sequential estimator will approach the batch estimator for any $k > M_t$. ■

D. Proof of Lemma 2

Under Hypothesis 3, the state transition matrix for a real, diagonal channel matrix $\mathbf{H}$ is given as

$$\mathbf{S}_{k-1} = \begin{bmatrix}\mathbf{I} & \sqrt{\rho}\beta_{k-1}\cdot\mathbf{H} \\ \sqrt{\rho}\alpha_{k-1}\cdot\mathbf{H} & \mathbf{I}\end{bmatrix} \tag{69}$$

since $\mathbf{H} = \mathbf{H}^* = \text{diag}\left([h_1, \cdots, h_M]\right)$. Note that the size of $\mathbf{S}_{k-1}$ is $2M \times 2M$. The characteristic equation of $\mathbf{S}_{k-1}$ is given as

$$\chi(\mathbf{S}_{k-1}, \lambda) = \det\left(\mathbf{S}_{k-1} - \lambda\mathbf{I}\right) \tag{70}$$

$$= \det\left(\begin{bmatrix} \mathbf{I} - \lambda\mathbf{I} & \sqrt{\rho}\beta_{k-1} \cdot \mathbf{H} \\ \sqrt{\rho}\alpha_{k-1} \cdot \mathbf{H} & \mathbf{I} - \lambda\mathbf{I} \end{bmatrix}\right). \tag{71}$$

By using the Schur complement lemma [48], this equation can be written as

$$\chi(\mathbf{S}_{k-1}, \lambda) = \det\Big(\mathbf{I} - \lambda\mathbf{I}\Big) \cdot \det\left(\mathbf{I} - \lambda\mathbf{I} - \rho\alpha_{k-1}\beta_{k-1} \cdot \mathbf{H}\left(\mathbf{I} - \lambda\mathbf{I}\right)^{-1}\mathbf{H}\right). \tag{72}$$

Since all of the matrices involved are diagonal, we can write the determinants as the product of the diagonal elements, resulting in

$$\chi(\mathbf{S}_{k-1}, \lambda) = (1-\lambda)^{2M} \prod_{i=1}^{2M} \left\{(1-\lambda) - \frac{\rho\alpha_{k-1}\beta_{k-1} \cdot h_i^2}{1-\lambda}\right\} \tag{73}$$

$$= \prod_{i=1}^{2M} \left\{(1-\lambda)^2 - \rho\alpha_{k-1}\beta_{k-1} \cdot h_i^2\right\}, \tag{74}$$

which has $2M$ roots (denoted as $\lambda_1, \cdots, \lambda_{2M}$) of the form $1 \pm \sqrt{\rho\alpha_{k-1}\beta_{k-1}} \cdot h_i$ for $i = 1, \ldots, M$. We can thus write the eigenvalue matrix $\mathbf{\Lambda}_{k-1}$ as

$$\mathbf{\Lambda}_{k-1} = \text{diag}\left(\begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_{2M} \end{bmatrix}\right) \tag{75}$$

$$= \text{diag}\left(\begin{bmatrix} 1 + \sqrt{\rho\alpha_{k-1}\beta_{k-1}} \cdot h_1 \\ \vdots \\ 1 + \sqrt{\rho\alpha_{k-1}\beta_{k-1}} \cdot h_M \\ 1 - \sqrt{\rho\alpha_{k-1}\beta_{k-1}} \cdot h_1 \\ \vdots \\ 1 - \sqrt{\rho\alpha_{k-1}\beta_{k-1}} \cdot h_M \end{bmatrix}\right). \tag{76}$$

Solving for the $2M$ eigenvectors (i.e. solving $\mathbf{S}_{k-1}\mathbf{u}_i = \lambda_i\mathbf{u}_i$ for $i = 1, \ldots, 2M$) and normalizing each column to unit-norm finally results in the following eigenvector matrix:

$$\mathbf{U}_{k-1} = \begin{bmatrix} \sqrt{\frac{\beta_{k-1}}{\alpha_{k-1}+\beta_{k-1}}}\mathbf{I} & \sqrt{\frac{\beta_{k-1}}{\alpha_{k-1}+\beta_{k-1}}}\mathbf{I} \\ \sqrt{\frac{\alpha_{k-1}}{\alpha_{k-1}+\beta_{k-1}}}\mathbf{I} & -\sqrt{\frac{\alpha_{k-1}}{\alpha_{k-1}+\beta_{k-1}}}\mathbf{I} \end{bmatrix}. \tag{77}$$

Note that $\mathbf{U}_{k-1}$ is not unitary in general and $\mathbf{S}_{k-1}$ can be written as $\mathbf{S}_{k-1} = \mathbf{U}_{k-1} \cdot \mathbf{\Lambda}_{k-1} \cdot \left(\mathbf{U}_{k-1}\right)^{-1}$. ■

E. Proof of Lemma 3

Let $\mathbf{H} = \{\mathbf{H}_{m,n}\}$, $\mathbf{f}[i] = [f_{i,1}, \cdots, f_{i,M}]^{\mathsf{T}}$ and $\mathbf{z}[i] = [z_{i,1}, \cdots, z_{i,M}]^{\mathsf{T}}$. Also, let $\mathbf{n}_e[i] = [\mathbf{n}_{e,1}[i], \cdots, \mathbf{n}_{e,M}[i]]^{\mathsf{T}}$ and $\mathbf{n}_o[i] = [\mathbf{n}_{o,1}[i], \cdots, \mathbf{n}_{o,M}[i]]^{\mathsf{T}}$. Now observe that

$$\frac{1}{\alpha_k^2} = \|\mathbf{s}_e[k]\|_2^2 = \left\|\sum_{i=0}^{k} \overline{\mathbf{y}}_e[i]\right\|_2^2 \tag{78}$$

$$= \sum_{n=1}^{M} \left|\sqrt{\rho}\sum_{m=1}^{M} \overline{\mathbf{H}}_{m,n} \sum_{i=0}^{k} z_{i,m} + \sum_{i=0}^{k} \overline{\mathbf{n}}_{e,n}[i]\right|^2. \tag{79}$$

Similarly, we have

$$\frac{1}{\beta_k^2} = \sum_{m=1}^{M} \left|\sqrt{\rho}\sum_{n=1}^{M} \mathbf{H}_{m,n} \sum_{i=0}^{k} f_{i,n} + \sum_{i=0}^{k} \mathbf{n}_{o,m}[i]\right|^2. \tag{80}$$

From (79) and (80), we have the following simplifications

$$\frac{1}{\rho \cdot \alpha_k^2} \overset{(a)}{\approx} \sum_{n=1}^{M} \left|\sum_{m=1}^{M} \overline{\mathbf{H}}_{m,n} \sum_{i=0}^{k} z_{i,m}\right|^2 \tag{81}$$

$$\overset{(b)}{\approx} |\mathsf{C}_k|^2 \cdot \sum_{n=1}^{M} \left|\sum_{m=1}^{M} \overline{\mathbf{H}}_{m,n}\right|^2 \tag{82}$$

$$\overset{(c)}{=} |\mathsf{C}_k|^2 \cdot \mathsf{Tr}(\mathbf{H}\mathbf{H}^*) \tag{83}$$

where (a), (b) and (c) follow from Hypotheses 1-3, respectively. Similarly, we have

$$\frac{1}{\rho \cdot \beta_k^2} \approx |\mathsf{C}_k|^2 \cdot \mathsf{Tr}(\mathbf{H}\mathbf{H}^*). \tag{84}$$

Thus, when Hypotheses 1-3 hold, $\alpha_k \approx \beta_k$ as k increases. ■

F. Proof of Theorem 4

When Hypotheses 1-3 hold, from Lemma 3, we have that $\alpha_k \approx \beta_k$ and $\mathbf{U}_{k-1} \approx \widetilde{\mathbf{U}}$. Thus, the state-space model in (35) can be written as

$$\mathbf{s}[k] = \widetilde{\mathbf{U}} \begin{bmatrix} \prod_{i=0}^{k-1}\left(\mathbf{I} + \sqrt{\rho}\alpha_i\mathbf{H}\right) & \mathbf{0} \\ \mathbf{0} & \prod_{i=0}^{k-1}\left(\mathbf{I} - \sqrt{\rho}\alpha_i\mathbf{H}\right) \end{bmatrix} \widetilde{\mathbf{U}}^*\mathbf{s}[0] + \widetilde{\mathbf{U}} \sum_{\ell=1}^{k} \begin{bmatrix} \prod_{j=\ell}^{k-1}\left(\mathbf{I} + \sqrt{\rho}\alpha_j\mathbf{H}\right) & \mathbf{0} \\ \mathbf{0} & \prod_{j=\ell}^{k-1}\left(\mathbf{I} - \sqrt{\rho}\alpha_j\mathbf{H}\right) \end{bmatrix} \widetilde{\mathbf{U}}^*\mathbf{n}[\ell]. \tag{85}$$

Let us now consider the $2M \times 2M$-dimensional diagonal matrix $\widetilde{\mathbf{\Lambda}}_{k-1,0}$

$$\widetilde{\mathbf{\Lambda}}_{k-1,0} = \begin{bmatrix} \prod_{i=0}^{k-1}\left(\mathbf{I} + \sqrt{\rho}\alpha_i\mathbf{H}\right) & \mathbf{0} \\ \mathbf{0} & \prod_{i=0}^{k-1}\left(\mathbf{I} - \sqrt{\rho}\alpha_i\mathbf{H}\right) \end{bmatrix}. \tag{86}$$

From Hypothesis 3, since the diagonal entries of $\mathbf{H}$ are arranged in non-increasing order and $h_1 > h_2$, we have

$$\frac{\prod_{i=0}^{k-1}(1 + \sqrt{\rho}\alpha_i h_1)}{\prod_{i=0}^{k-1}(1 + \sqrt{\rho}\alpha_i h_\ell)} \approx \left(\frac{h_1}{h_\ell}\right)^k \to \infty \text{ as } k \to \infty \tag{87}$$

for $\ell = 2, \cdots, M$. Similarly, we have

$$\frac{\prod_{i=0}^{k-1}(1 - \sqrt{\rho}\alpha_i h_1)}{\prod_{i=0}^{k-1}(1 - \sqrt{\rho}\alpha_i h_\ell)} \approx \left(\frac{h_1}{h_\ell}\right)^k \to \infty \text{ as } k \to \infty \tag{88}$$

for $\ell = 2, \cdots, M$ and

$$\frac{\prod_{i=0}^{k-1}(1+\sqrt{\rho}\alpha_i h_1)}{\prod_{i=0}^{k-1}(1-\sqrt{\rho}\alpha_i h_1)} \approx (-1)^k \text{ as } k \to \infty. \tag{89}$$

Thus, the diagonal entries of $\widetilde{\mathbf{\Lambda}}_{k-1,0}$ are dominated by (as k increases) the first entry, which is denoted as

$$\lambda_{1,0} \approx \left(\sqrt{\rho}h_1\right)^k \cdot \prod_{i=0}^{k} \alpha_i, \tag{90}$$

and the $(M+1)$th entry, which is denoted as

$$\lambda_{M+1,0} \approx \left(-\sqrt{\rho}h_1\right)^k \cdot \prod_{i=0}^{k} \alpha_i. \tag{91}$$

Similarly, we can consider the diagonal matrices $\widetilde{\mathbf{\Lambda}}_{k-1,\ell}$ for $\ell = 1, \cdots, k-1$:

$$\widetilde{\mathbf{\Lambda}}_{k-1,\ell} = \begin{bmatrix} \prod_{i=\ell}^{k-1}\left(\mathbf{I}+\sqrt{\rho}\alpha_i\mathbf{H}\right) & \mathbf{0} \\ \mathbf{0} & \prod_{i=\ell}^{k-1}\left(\mathbf{I}-\sqrt{\rho}\alpha_i\mathbf{H}\right) \end{bmatrix}. \tag{92}$$

Following the same logic as before, these matrices are also dominated by the first entry, which is denoted as

$$\lambda_{1,\ell} \approx \left(\sqrt{\rho}h_1\right)^{k-\ell} \cdot \prod_{i=\ell}^{k-1} \alpha_i, \tag{93}$$

and the $(M+1)$th entry, which is denoted as

$$\lambda_{M+1,\ell} \approx \left(-\sqrt{\rho}h_1\right)^{k-\ell} \cdot \prod_{i=\ell}^{k-1} \alpha_i. \tag{94}$$

With $\mathbf{s}[0] = [s_1(0), \cdots, s_{2M}(0)]^{\mathsf{T}}$, $\mathbf{s}[k] = [s_1(k), \cdots, s_{2M}(k)]^{\mathsf{T}}$, and $\mathbf{n}[\ell] = [n_1(\ell), \cdots, n_{2M}(\ell)]^{\mathsf{T}}$, it is straightforward to see that as k increases and for $\rho \gg 1$, we have

$$\frac{s_1(k)}{\left(\sqrt{\rho}h_1\right)^k \cdot \prod_{i=0}^{k-1}\alpha_i} \to \begin{cases} s_1(0) & \text{if } k \text{ is even} \\ s_{M+1}(0) & \text{if } k \text{ is odd} \end{cases} \tag{95}$$

$$\frac{s_{M+1}(k)}{\left(\sqrt{\rho}h_1\right)^k \cdot \prod_{i=0}^{k-1}\alpha_i} \to \begin{cases} s_{M+1}(0) & \text{if } k \text{ is even} \\ s_1(0) & \text{if } k \text{ is odd.} \end{cases} \tag{96}$$

And we also have $\frac{s_\ell(k)}{\left(\sqrt{\rho}h_1\right)^k \cdot \prod_{i=0}^{k-1}\alpha_i} \to 0$ for all $\ell \in \{2, \cdots, M, M+2, \cdots, 2M\}$. Thus, $\mathbf{s}[k] \to \mathbf{s}_{\mathsf{opt}}$ as k increases. ■

REFERENCES

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. Zhang, "What will 5G be?," *IEEE Journ. Sel. Areas in Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[2] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Magaz.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] F. Khan and Z. Pi, "An introduction to millimeter wave mobile broadband systems," *IEEE Commun. Magaz.*, vol. 49, no. 6, pp. 101–107, June 2011.

[4] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[5] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[6] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?," *IEEE Commun. Magaz.*, vol. 52, no. 9, pp. 56–62, Sept. 2014.

[7] T. L. Marzetta, "Non-cooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[8] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[9] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Sig. Proc. Magaz.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[10] Aalto University, AT&T, BUPT, CMCC, Ericsson, Huawei, Intel, KT Corporation, Nokia, NTT DOCOMO, NYU, Qualcomm, Samsung, U. Bristol, and USC, "White paper on '5G Channel Model for bands up to 100 GHz'," *v2.3*, Oct. 2016.

[11] 3GPP TR 38.900 V14.1.0 (2016-09), "Study on channel model for frequency spectrum above 6 GHz (Rel. 14)," Sept. 2016.

[12] D. Gesbert, M. Kountouris, R. W. Heath, Jr., C.-B. Chae, and T. Salzer, "Shifting the MIMO paradigm: From single user to multiuser communications," *IEEE Sig. Proc. Magaz.*, vol. 24, no. 5, pp. 36–46, Oct. 2007.

[13] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Commun. Magaz.*, vol. 42, no. 10, pp. 60–67, Oct. 2004.

[14] V. Venkateswaran and A.-J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Sig. Proc.*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.

[15] E. Torkildson, U. Madhow, and M. Rodwell, "Indoor millimeter wave MIMO: Feasibility and performance," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4150–4160, Dec. 2011.

[16] J. Brady, N. Behdad, and A. M. Sayeed, "Beamspace MIMO for millimeter-wave communications: System architecture, modeling, analysis and measurements," *IEEE Trans. Ant. Propag.*, vol. 61, no. 7, pp. 3814–3827, July 2013.

[17] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Magaz.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[18] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[19] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[20] F. Boccardi, B. Clerckx, A. Ghosh, E. Hardouin, G. Jöngren, K. Kusume, E. Onggosanusi, and Y. Tang, "Multiple antenna techniques in LTE-Advanced," *IEEE Commun. Magaz.*, vol. 50, no. 2, pp. 114–121, Mar. 2012.

[21] C. Lim, T. Yoo, B. Clerckx, B. Lee, and B. Shim, "Recent trend of multiuser MIMO in LTE-Advanced," *IEEE Commun. Magaz.*, vol. 51, no. 3, pp. 127–135, Mar. 2013.

[22] J. J. Choi, V. Raghavan, and D. J. Love, "Limited feedback design for the spatially correlated multi-antenna broadcast channel," *Proc. IEEE Global Telecommun. Conf., Atlanta, GA*, pp. 3481–3486, Dec. 2013.

[23] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 935–946, May 2000.

[24] B. Hassibi and B. Hochwald, "How much training is needed in a multiple antenna wireless link?," *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–964, Apr. 2003.

[25] G. Hariharan, V. Raghavan, and A. M. Sayeed, "Capacity of sparse wideband channels with partial channel feedback," *Trans. Emerging Telecommun. Technologies*, vol. 19, no. 4, pp. 475–493, June 2008.

[26] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. 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.

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

[28] S. Gazor and K. AlSuhaili, "Communications over the best singular mode of a reciprocal MIMO channel," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 1993–2001, July 2010.

[29] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[30] P. Schniter and A. M. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," *Proc. IEEE Asilomar Conf. Signals, Systems and Computers, Pacific Grove, CA*, pp. 273–277, Nov. 2014.

[31] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive tracking with 1000-element arrays: A framework for multi-Gbps mm-wave cellular downlinks," *Proc. Annual Allerton Conf. Commun., Control and Computing, Allerton, IL*, pp. 690–697, Oct. 2012.

[32] V. Raghavan and A. M. Sayeed, "MIMO capacity scaling and saturation in correlated environments," *Proc. IEEE Intern. Conf. Commun.*, vol. 5, pp. 3006–3010, May 2003.

[33] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journ. Sel. Topics in Sig. Proc.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[34] S. Sun, T. S. Rappaport, R. W. Heath, Jr., A. Nix, and S. Rangan, "MIMO for millimeter wave wireless communications: Beamforming, spatial multiplexing, or both?," *IEEE Commun. Magaz.*, vol. 52, no. 12, pp. 110–121, Dec. 2014.

[35] A. Adhikary, E. Al Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, "Joint spatial division and multiplexing for mm-Wave channels," *IEEE Journ. Sel. Areas in Commun.*, vol. 32, no. 6, pp. 1239–1255, June 2014.

[36] V. Raghavan, J. Cezanne, S. Subramanian, A. Sampath, and O. H. Koymen, "Beamforming tradeoffs for initial UE discovery in millimeter-wave MIMO systems," *IEEE Journ. Sel. Topics in Sig. Proc.*, vol. 10, no. 3, pp. 543–559, Apr. 2016.

[37] V. Raghavan, S. Subramanian, J. Cezanne, A. Sampath, O. H. Koymen, and J. Li, "Directional hybrid precoding in millimeter-wave MIMO systems," *Proc. IEEE Global Telecommun. Conf., Washington, DC*, pp. 1–7, Dec. 2016.

[38] Y. Tang, B. Vucetic, and Y. Li, "An iterative singular vectors estimation scheme for beamforming transmission and detection in MIMO systems," *IEEE Commun. Letters*, vol. 9, no. 6, pp. 505–507, June 2005.

[39] S. Mandelli and M. Magarini, "Blind iterative singular vectors estimation and adaptive spatial loading in a reciprocal MIMO channel," *Proc. IEEE Wireless Commun. and Netwk. Conf., Istanbul, Turkey*, pp. 1036–1041, Apr. 2014.

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

[41] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, fourth edition, Dec. 2012.

[42] R. Prasad, B. N. Bharath, and C. R. Murthy, "Joint data detection and dominant singular mode estimation in time varying reciprocal MIMO systems," *Proc. IEEE Intern. Conf. on Acoust., Speech and Sig. Proc., Prague, Czech Rep.*, pp. 3240–3243, May 2011.

[43] H. Ghauch, T. Kim, M. Skoglund, and M. Bengtsson, "Subspace estimation and decomposition in large millimeter-wave MIMO systems," *IEEE Journ. Sel. Topics in Sig. Proc.*, vol. 10, no. 3, pp. 528–542, Apr. 2016.

[44] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE Journ. Sel. Topics in Sig. Proc.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[45] S. Noh, M. D. Zoltowski, and D. J. Love, "Training sequence design for feedback assisted hybrid beamforming in massive MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 187–200, Jan. 2016.

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

[47] C-H. Tse, K-W. Yip, and T-S. Ng, "Performance tradeoffs between maximum ratio transmission and switched-transmit diversity," *Proc. IEEE Intern. Symp. Pers. Indoor and Mob. Radio Commun., London, UK*, vol. 2, pp. 1485–1489, Sept. 2000.

[48] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 2nd edition, Oct. 2012.

[49] S. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall, Englewood Cliffs, N.J, 1st edition, Apr. 1993.

[50] F. Eicker, "Asymptotic normality and consistency of the least squares estimators for families of linear regressions," *Ann. Math. Stat.*, vol. 34, no. 2, pp. 447–456, June 1963.

[51] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Capacity of sparse multipath channels in the ultra-wideband regime," *IEEE Journ. Sel. Topics Sig. Proc.*, vol. 1, no. 3, pp. 357–371, Oct. 2007.

Dennis Ogbe (S'13) received the B.S. degree (with honors) in electrical engineering in 2014 from Tennessee Technological University, Cookeville, TN and is currently working towards the Ph.D. degree at Purdue University, West Lafayette, IN. During the summer of 2016, he was an intern at Nokia Networks. His current research interests are in the design of adaptive multiple antenna wireless systems and software defined radio. Mr. Ogbe is an active member of Eta Kappa Nu.

David J. Love (S'98–M'05–SM'09–F'15) received the B.S. (with highest honors), M.S.E., and Ph.D. degrees in electrical engineering from the University of Texas at Austin in 2000, 2002, and 2004, respectively. Since August 2004, he has been with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, where he is now a Professor. He has served as an Editor for the IEEE Transactions on Communications, an Associate Editor for the IEEE Transactions on Signal Processing, and a guest editor for special issues of the IEEE Journal on Selected Areas in Communications and the EURASIP Journal on Wireless Communications and Networking. His industry experience includes work as a summer co-op and consultant for Texas Instruments. Dr. Love holds 27 issued US patents.

Dr. Love was recognized as a Thomson Reuters Highly Cited Researcher in 2014 and 2015. Along with his co-authors, he has won best paper awards from the IEEE Communications Society (2016 IEEE Communications Society Stephen O. Rice Prize), the IEEE Signal Processing Society (2015 IEEE Signal Processing Society Best Paper Award), and the IEEE Vehicular Technology Society (2009 IEEE Transactions on Vehicular Technology Jack Neubauer Memorial Award). He has received multiple IEEE Global Communications Conference (Globecom) best paper awards.

Vasanthan Raghavan (S'01–M'06–SM'11) received the B.Tech degree in Electrical Engineering from the Indian Institute of Technology at Madras, India in 2001, M.S. and Ph.D. degrees in Electrical and Computer Engineering in 2004 and 2006, respectively, and M.A. degree in Mathematics in 2005, all from the University of Wisconsin, Madison, WI. He is currently with the New Jersey Research Center of Qualcomm, Inc. His research interests span multi-antenna communication techniques,information theory, quickest changepoint detection, and random matrix theory.