Downlink Channel Sounding For Multiuser Massive MIMO With Limited Feedback

Petteri Pulkkinen, *Member, IEEE*, David J. Love, *Fellow, IEEE*, and Visa Koivunen, *Fellow, IEEE*

***Abstract*—Massive multiple-input multiple-output (MIMO) technology is pivotal for next-generation wireless networks, offering higher data rates and seamless connectivity. Despite its promise, channel estimation in feedback-based massive MIMO presents unique challenges that must be addressed to realize massive MIMO for frequency-division duplexing (FDD) and other communications systems without channel reciprocity. This paper proposes a novel base station (BS) centric closed-loop channel sounding method for multiuser massive MIMO systems, aiming to reduce the pilot overhead and improve the data rates. It also reduces the computational load on the UEs and facilitates accurate channel estimation at the BS. The proposed approach utilizes beamspace processing in the pilot optimization and the feedback scheme. Furthermore, we show that the proposed method can be employed with hybrid analog-digital architecture in BSs. An efficient algorithm for designing the pilots to maximize the approximate mutual information (MI) between the channel coefficients and received feedback signals is developed, along with an efficient hybrid zero-forcing (HZF) precoding algorithm. The simulation results show that this approach significantly improves channel estimation accuracy and sum rate performance in multiuser massive MIMO systems.**

***Index Terms*—Beam alignment, millimeter wave, frequency division duplexing, convex optimization, mutual information.**

I. Introduction

MASSIVE MIMO technology is pivotal for 5G and emerging 6G wireless networks to enable higher data rates and seamless connectivity [1]. The multiplexing gains and spatial diversity enable higher spectral efficiencies for wireless users. Furthermore, the array gains facilitate larger network cell sizes and enable tackling high propagation losses in millimeter and terahertz frequencies [2]. Higher frequencies above sub-6 GHz band allow for using larger antenna array apertures with smaller physical size and employing broader bandwidths. This facilitates new innovative use cases and alleviating the spectrum congestion.

Despite the promise of massive MIMO technology, it presents unique challenges that must be addressed to realize its full potential. Previously, massive MIMO has been mainly studied in sub-6 Ghz wireless communication systems obeying fully digital and relatively small antenna arrays and time-division duplexing (TDD) scheme [1], [3]. However, research and industry interest is directed towards large antenna arrays. To reduce costs and power consumption, algorithms for low-bit analog-to-digital converters (ADCs) [4] and hybrid analog-digital architectures [5], [6] are developed to tackle hardware-related problems when using large antenna arrays [7].

This paper considers multiuser channel sounding and precoding in limited feedback multiuser massive MIMO systems [8], [9], [10]. Awareness of channel states and interference levels plays a significant role in reaching the achievable communications capacity [10]. In particular, the transmitter requires accurate knowledge of the communication channel to achieve optimal beamforming gains and to effectively mitigate interference between multiple spatial streams. Obtaining channel state information at the trasmitter (CSIT) in mobile massive MIMO systems poses a significant challenge. It is particularly difficult in frequency-division duplexing (FDD) scheme with limited feedback and under high frequencies with short coherence times and hardware constraints. Therefore, this paper focuses on downlink channel estimation in FDD multiuser massive MIMO systems and proposes a channel sounding method tailored for such systems with hybrid digital-analog hardware.

Acquiring CSIT for downlink transmissions in FDD massive MIMO systems can be broadly categorized into user equipment (UE)-centric [11], [12], [13], [14], [15], [16], [17], [18], [19], [20] and BS-centric [21], [22] schemes. In UE-centric scheme, UEs estimate the channels from downlink pilots and then provide channel state information (CSI) feedback to the BS. Conversely, BS-centric schemes rely on UEs to feed back the received pilot signals to the BS, which then performs the downlink channel estimation. To the authors' best knowledge, the UE-centric scheme has been widely adopted in the literature with digital feedback, while BS-centric scheme have been

Received 24 February 2025; revised 18 July 2025 and 6 September 2025; accepted 14 September 2025. Date of publication 1 October 2025; date of current version 4 November 2025. This work was supported in part by the National Science Foundation under Grant CNS2225578 and Grant CNS2212565. The associate editor coordinating the review of this article and approving it for publication was Foad Sohrabi. *(Corresponding author: Petteri Pulkkinen.)*

Petteri Pulkkinen is with the Department of Information and Communications Engineering, Aalto University, FI-02150, Espoo, Finland and also with Saab Finland Oy FI-00100 Helsinki, Finland (e-mail: petteri.pulkkinen@aalto.fi).

David J. Love is with the Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: djlove@purdue.edu).

Visa Koivunen is with the Department of Information and Communications Engineering, Aalto University, FI-02150 Espoo, Finland (e-mail: visa.koivunen@aalto.fi).

Digital Object Identifier 10.1109/TSP.2025.3615853

considered only in conjunction with analog feedback [23]. Most communications systems with limited digital feedback use spatial domain codebooks to alleviate CSI quantization. For example, the third generation partnership project (3GPP) standards (e.g., release 17) are based on the discrete Fourier transform (DFT) codebooks [24] where the sparsity in spatial [25] or delay domain [26] is utilized. However, there are also many other codebook designs, such as large dimension beamforming vector quantization based on trellis-coding [27], [28], [29], and codebooks tailored for millimeter wave communications [30], [31], [32], [33].

Downlink channel sounding frameworks within the UE-centric scheme are generally divided into *open-loop* [11], [12], [13], [14] and *closed-loop* [15], [16], [17], [18], [19] methods. The open-loop indicates that there is no feedback from UEs to BS regarding the preferred pilot sequences. In contrast, in closed-loop methods [15], [16], [17], [18], [19], UEs exploit channel awareness and associated uncertainties to design downlink pilots that minimize predicted channel estimation errors and subsequently transmit the pilot preferences to the BS via a feedback link. Open-loop and closed-loop sounding methods may integrate information from multiple received subsequent pilots using sequential estimation, for example, Kalman filtering (KF) [34], [35], [36], [37], [38]. In contrast, open-loop methods, where each UE estimates the channel using only the most recent pilot signal, are referred to as *single-shot* methods [11], [12], [13], [39]. Closed-loop methods require additional feedback from UEs to BSs for determining the pilots to be used. Methods that aim to minimize this feedback overhead have been proposed, for example, in [18], [19], where UEs control the downlink pilot lengths via 1-bit feedback, reducing the feedback overhead but significantly limiting the degrees of freedom in the pilot adaptation.

Related to channel sounding methods, the process of iteratively probing transmit and receive beams to find the best beamformer from the codebook is called beam alignment (BA) [40], [41], [42], especially in the context of millimeter wave systems. These approaches are conventionally developed for analog transmit-receive beamforming, and the methods are based on exhaustive [41] or hierarchical [32], [43], [44], [45] search. BA approaches for hybrid architectures are typically based on compressive sensing (CS) where BS transmit pseudo-random pilots, and UEs estimate the strongest beam pairs and reports them back to the BS [46]. Another BA approach is to leverage auxiliary data, like GPS [47], [48] or sensor inputs [49], to inform the BA process. Also, sub-6 GHz frequency bands can be utilized for initial BA, as shown in [50], [51], [52] since lower frequencies facilitate more favorable propagation characteristics than millimeter waves. A large body of BA literature considers closed-loop beam alignment [47], [53], [54], [55], [56], [57], [58], [59], [60], [61]. Many of them are based on the multi-armed bandit (MAB) algorithms [47], [55], [56], [57], [58], [59] and partially observable Markov decision processes (POMDPs) [60], [61]. These algorithms are motivated by the optimal tradeoff between exploration (gaining more knowledge about unknown beam pairs) and exploitation (picking the best beam pair).

This paper proposes a BS-centric channel sounding algorithm that sequentially optimizes pilot signals, leveraging the proposed pilot and feedback scheme to estimate downlink channels at BS centrally. In contrast to closed-loop UE-centric approaches [15], [16], [17], [18], [19], [20], the proposed BS-centric method enables centralized pilot designs, as well as facilitates the pilot adaptation without requiring additional feedback from the UEs. Compared to existing BS-centric methods in [21], [22], which rely on analog feedback, the proposed approach can be used with limited digital feedback. Furthermore, unlike [21], [22], the proposed method considers hybrid digital-analog architectures and is formulated within a Bayesian framework. Therefore, the method can leverage the memory of past channel observations and optimize pilots sequentially. The proposed algorithm accommodates limited digital feedback via optimizing pilot structures based on beamspace codebooks and approximate mutual information (MI) criterion [62]. Due to the use of beam codebooks for probing the channel, it is also related to the BA methods. The method is tailored for hybrid digital-analog architectures and can be implemented with various antenna configurations. For example, the proposed method can be implemented with a discrete lens array (DLA) [63], effectively reducing hardware costs.

Summarizing the contributions in this paper compared to the existing literature:

1) A novel BS-centric channel sounding algorithm is proposed for FDD multiuser massive MIMO systems, and is compatible with hybrid transceiver architectures.
2) The BS-centric approach enables the use of centralized channel awareness for designing pilot signals sequentially, as well as avoids additional feedback from the UEs required by closed-loop UE-centric channel sounding methods. An optimization method is proposed for designing pilots that minimizes the approximate MI criterion.
3) In contrast to prior BS-centric approaches that rely on analog feedback, the proposed beamspace domain pilot structure facilitates limited digital feedback. Additionally, a Bayesian formulation is employed to exploit channel memory, enabling sequential pilot optimization.
4) A zero-forcing precoder design for downlink data transmissions using the acquired CSIT is proposed for hybrid systems, referred to as hybrid zero-forcing (HZF).

The performance of the proposed algorithm is evaluated in diverse simulation settings to verify the analytical results and compare against both UE-centric methods that use DFT codebooks for CSI quantization [24], and simplified variants of the proposed BS-centric channel sounding algorithm with reduced pilot design complexity.

II. Notation

We use $\mathbb{N}$ to denote set of natural numbers $\{1, 2, \dots\}$, excluding 0. The shorthand notation $[N] = \{1, \dots, N\}$ denotes the set of natural numbers from 1 to N. The probability distribution $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a scalar or multivariate complex Gaussian with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Specifically, for $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

$\mathbf{x} - \boldsymbol{\mu}$ obeys a zero-mean circular complex Gaussian distribution. Matrix $\mathbf{I}_N$ is an identity matrix of size N. Operator $\mathcal{C}(\mathcal{X}, K)$ returns a set that comprises all subsets of the set $\mathcal{X}$ with K distinct elements.

III. Problem Description

This paper considers downlink (DL) communications and channel estimation with limited feedback in multiuser massive MIMO systems. We do not assume channel reciprocity such that the proposed method applies to FDD and other communications systems that do not rely on reciprocity. We assume a limited feedback channel, where the number of feedback bits per one time block is finite. Furthermore, we consider a BS-centric system model where the channel estimation, pilot design, and precoder design are implemented at the BS side. This distinguishes the proposed method from better-known UE-centric methods [11], [12], [13], [14], [15], [16], [17], [18], [19], [20] that estimate the channel at the UE side, and feed the quantized CSI back to the BS to be used in designing the precoder for downlink data transmissions. Moreover, closed-loop UE-centric methods [15], [16], [17], [18], [19], that inform BS about the preferred pilot signals, are difficult to extend to the multiuser case, as the optimal design requires centralized optimization.

The system model comprises a BS with M antennas and N single antenna UEs.[1] Each downlink channel obeys a widely used temporally correlated block fading model that evolves according to a first-order Gauss-Markov process [64], [65]. Coherence blocks of the channel are referred to with indices $k \in \mathbb{N}$. The DL channel to the user $n \in [N]$ is denoted as $\mathbf{h}_k^{(n)} \in \mathbb{C}^{M\times 1}$. The Gauss-Markov process is a linear transition model that relates two subsequent channel states as

$$\mathbf{h}_{k+1}^{(n)} = \eta_n \, \mathbf{h}_k^{(n)} + \sqrt{1-\eta_n^2}\, \mathbf{q}_k^{(n)} \tag{1}$$

where $\eta_n \in [0,1)$ is the temporal-correlation coefficient and $\mathbf{q}_k^{(n)} \sim \mathcal{N}_{\mathbb{C}}(0, \boldsymbol{\Phi}_n)$ is the state process noise. The matrix $\boldsymbol{\Phi}_n \in \mathbb{C}^{M\times M}$ is the covariance matrix of the process noise $\mathbf{q}_k$, but also exhibits the covariance matrix of the channel in the stationary regime, i.e., $\boldsymbol{\Phi}_n = \lim_{k\to\infty} \mathrm{cov}(\mathbf{h}_k^{(n)} | \mathbf{h}_0^{(n)})$, where $\mathrm{cov}(\cdot|\cdot)$ denotes the conditional covariance. Therefore, it is also referred to as the channel covariance matrix.

Consider a BS with a hybrid analog-digital multiantenna architecture and $M_{\mathrm{rf}} \leq M$ radio frequency (RF) chains. Transmit symbols in each coherence block can be divided into pilot and data symbols, as illustrated in Fig. 1. The symbols in each coherence block are indexed as $\ell \in [L]$, where L is the block length in symbols. The indices $\ell \in \mathcal{L}_{\mathrm{P}} \subset [L]$ correspond to pilot symbols, while indices $\ell \in [L] \backslash \mathcal{L}_{\mathrm{P}}$ correspond to the data indices. The number of pilot symbols is $L_{\mathrm{P}} = |\mathcal{L}_{\mathrm{P}}|$. The transmit signal $\mathbf{x}_k[\ell] \in \mathbb{C}^{M\times 1}$ for all $\ell \in [L]$ can be written as

$$\mathbf{x}_k[\ell] = \sqrt{\rho_{\mathrm{tx}}} \cdot \begin{cases} \mathbf{A}_k^{\mathrm{P}} \mathbf{d}_k^{\mathrm{P}}[\ell], & \text{if } \ell \in \mathcal{L}_{\mathrm{P}} \\ \mathbf{A}_k^{\mathrm{D}} \mathbf{D}_k^{\mathrm{D}} \mathbf{c}_k[\ell], & \text{else} \end{cases} \tag{2}$$

where ρ_{tx} is the total RF chain power, $\mathbf{A}_k^{\mathrm{P}} \in \mathbb{C}^{M\times M_{\mathrm{rf}}}$ is the analog part of the pilot and $\mathbf{d}_k^{\mathrm{P}}[\ell] \in \mathbb{C}^{M_{\mathrm{rf}}\times 1}$ is the digital part of the pilot. Furthermore, for the DL data symbols $\mathbf{c}_k[\ell] \in \mathbb{C}^{N\times 1}$, $\mathbf{A}_k^{\mathrm{D}} \in \mathbb{C}^{M\times M_{\mathrm{rf}}}$ is the analog precoder and $\mathbf{D}_k^{\mathrm{D}} \in \mathbb{C}^{M_{\mathrm{rf}}\times N}$ is the digital precoder. Define matrix $\mathbf{D}_k^{\mathrm{P}} \in \mathbb{C}^{M_{\mathrm{rf}}\times L_{\mathrm{P}}}$ which columns comprise the set of vectors $\{\mathbf{d}_k^{\mathrm{P}}[\ell]\}_{\ell\in\mathcal{L}_{\mathrm{P}}}$. Due to the average power constraint of the RF chains $\mathrm{tr}\left\{(\mathbf{D}_k^{\mathrm{P}})^{\mathrm{H}}\mathbf{D}_k^{\mathrm{P}}\right\} \leq L_{\mathrm{P}}$ and $\mathrm{tr}\left\{(\mathbf{D}_k^{\mathrm{D}})^{\mathrm{H}}\mathbf{D}_k^{\mathrm{D}}\right\} \leq 1$. The matrices $\mathbf{A}_k^{\mathrm{D}}$ and $\mathbf{A}_k^{\mathrm{P}}$ corresponding to the analog beamforming control the phases of the phase-shifter network.

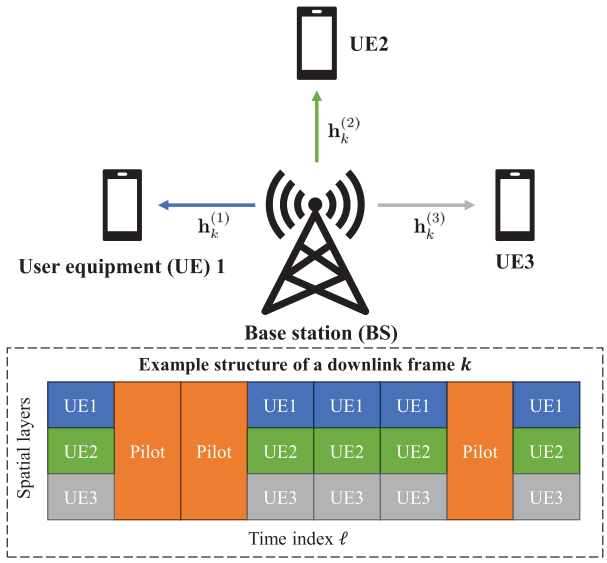

Fig. 1. BS communicating to three UEs. Some time indices within the frame are allocated for the optimized pilot sent simultaneously to all UEs.

The received signal at the UE can be written as

$$r_k^{(n)}[\ell] = \mathbf{x}_k^{\mathrm{T}}[\ell]\, \mathbf{h}_k^{(n)} + v_k[\ell] \tag{3}$$

where $v_k^{(n)}[\ell] \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_n^2)$ is zero-mean and circular complex Gaussian noise with variance σ_n^2. Thus, the vector of received pilot symbols $(\{r_k^{(n)}[\ell]\}_{\ell\in\mathcal{L}_{\mathrm{P}}})$ can be written as

$$\mathbf{r}_{\mathrm{P},k}^{(n)} = \sqrt{\rho_{\mathrm{tx}}}(\mathbf{A}_k^{\mathrm{P}}\mathbf{D}_k^{\mathrm{P}})^{\mathrm{T}}\mathbf{h}_k^{(n)} + \mathbf{v}_{\mathrm{P},k}^{(n)} \tag{4}$$

where the noise $\mathbf{v}_{\mathrm{P},k}^{(n)} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_n^2 \mathbf{I}_{L\mathrm{P}})$ obeys zero-mean circular complex Gaussian distribution with covariance $\sigma_n^2 \mathbf{I}_{L\mathrm{P}}$. We define matrix $\mathbf{X}_k^{\mathrm{P}} \in \mathbb{C}^{L_{\mathrm{P}}\times M}$ as

$$\mathbf{X}_k^{\mathrm{P}} = \sqrt{\rho_{\mathrm{tx}}}(\mathbf{A}_k^{\mathrm{P}}\mathbf{D}_k^{\mathrm{P}})^{\mathrm{T}} \tag{5}$$

which rows comprise the pilots $\{\mathbf{x}_k[\ell]\}_{\ell\in\mathcal{L}_{\mathrm{P}}}$. The matrix $\mathbf{X}_k^{\mathrm{P}}$ is later referred to as the pilot matrix.

IV. Channel Estimation With Limited Feedback and Memory

In a traditional UE-centric channel estimation scheme, UEs need to know the pilot matrix $\mathbf{X}_k^{\mathrm{P}}$ in (4) to estimate the channel. If UEs know the pilot matrix, they can estimate the channels and feedback the quantized channels or precoders from a codebook. This means that in an *open-loop* setting, there is a set of pilot matrices $\mathbf{X}_k^{\mathrm{P}}$ shared among the UEs [14]. Then, the pilot matrices are scheduled in a round-robin manner, allowing UEs to

[1] The proposed approach can be extended to multiantenna UEs in future research.

determine which pilot was received. In UE-centric *closed-loop* setting, UEs themselves design the pilot matrices and provide additional feedback about the index of the desired pilot matrix to the BS [15]. This UE-centric closed-loop scheme has three major drawbacks: (i) additional feedback bits are required, (ii) pilots among users may cause mutual interference, and (iii) pilots are not centrally and jointly optimized for multiuser channel estimation. This section introduces a BS-centric feedback scheme that facilitates accurate channel estimation at the BS, enabling *centralized* and *sequential* design of the pilot matrices. The BS-centric approach also reduces the computation required at the UE side.

A. BS Centric Feedback Model

A naïve BS-centric approach would design the pilot matrix $\mathbf{X}_k^{\mathrm{P}}$ in (5) to optimize some performance metric, and each UE would feedback the received pilot signal $\mathbf{r}_{\mathrm{P},k}^{(n)}$ back to the BS as in [21]. However, providing feedback on the whole observation vector introduces too much overhead, especially if L_{P} is large. Introducing structure to the pilot matrix can reduce this feedback overhead while still allowing for *closed-loop* optimization.

The proposed method is based on estimating the channels using beamspace transformation. Define the beamspace transformation of the communications channel as $\bar{\mathbf{h}}_k^{(n)} = \mathbf{\Gamma}\mathbf{h}_k^{(n)}$ where rows of the codebook matrix $\mathbf{\Gamma} \in \mathbb{C}^{M_{\mathrm{cb}} \times M}$ correspond to different beams. The rank of matrix $\mathbf{\Gamma}$ needs to be M to facilitate estimating $\mathbf{h}_k^{(k)}$ when using the transformation. Furthermore, the codebook matrix $\mathbf{\Gamma}$ should be designed in such a way that it can resolve distinct and approximately independent propagation paths. Thus, the beamspace domain channel $\bar{\mathbf{h}}_k^{(n)}$ is typically a sparse vector, especially in higher frequencies. For a uniform linear array (ULA), the codebook matrix can be, for example, the DFT matrix with $M_{\mathrm{cb}} = M$. If a DFT matrix is considered, the proposed method can also be implemented with reduced complexity hardware, such as using a DLA [63].

Consider any codebook matrix $\mathbf{\Gamma}$ that can be implemented with only phase-shift elements. Therefore, we can set $(\mathbf{A}_k^{\mathrm{P}})^{\mathrm{T}} = \mathbf{S}_k^{\mathrm{P}}\mathbf{\Gamma}$ where $\mathbf{S}_k^{\mathrm{P}} \in \{0,1\}^{M_{\mathrm{rf}} \times M_{\mathrm{cb}}}$ is the beam selection matrix. In addition, the digital part of the pilot matrix can be set $(\mathbf{D}_k^{\mathrm{P}})^{\mathrm{T}} = \mathbf{U}\mathbf{\Lambda}_k^{\mathrm{P}}$ where $\mathbf{\Lambda}_k^{\mathrm{P}} \in \mathbb{R}_+^{L_{\mathrm{P}} \times M_{\mathrm{rf}}}$ is a rectangular diagonal matrix and $\mathbf{U} \in \mathbb{C}^{L_{\mathrm{P}} \times L_{\mathrm{P}}}$ is any unitary matrix. A total power constraint $\mathrm{tr}\left\{(\mathbf{\Lambda}_k^{\mathrm{P}})^{\mathrm{H}}\mathbf{\Lambda}_k^{\mathrm{P}}\right\} \leq L_{\mathrm{P}}$ is imposed. Define variable $M_{\mathrm{nz}} \leq \min(M_{\mathrm{rf}}, L_{\mathrm{P}})$ that is the maximum number of beams allowed to be probed simultaneously, later shown to control the amount of feedback. The diagonal elements $[\mathbf{\Lambda}_k^{\mathrm{P}}]_{m,m} \geq 0$ for all $m \in [M_{\mathrm{nz}}]$, and other elements are strictly zero.

With the pilot model above, the pilot matrix in (5) can be rewritten as

$$\mathbf{X}_k^{\mathrm{P}} = \sqrt{\rho_{\mathrm{tx}}}\,\mathbf{U}\mathbf{\Lambda}_k^{\mathrm{P}}\mathbf{S}_k^{\mathrm{P}}\mathbf{\Gamma}. \tag{6}$$

By plugging (6) to (4) we obtain

$$\mathbf{r}_{\mathrm{P},k}^{(n)} = \sqrt{\rho_{\mathrm{tx}}}\,\mathbf{U}\mathbf{\Lambda}_k^{\mathrm{P}}\tilde{\mathbf{h}}_k^{(n)} + \mathbf{v}_{\mathrm{P},k}^{(n)} \tag{7}$$

where the vector $\tilde{\mathbf{h}}_k^{(n)} = \mathbf{S}_k^{\mathrm{P}}\bar{\mathbf{h}}_k^{(n)} \in \mathbb{C}^{M_{\mathrm{rf}} \times 1}$ comprises M_{rf} elements of the beamspace channel $\bar{\mathbf{h}}_k^{(n)} = \mathbf{\Gamma}\mathbf{h}_k^{(n)}$. Assume that BS can share the matrix $\mathbf{U}$ with the UEs.[2] Thus, the UEs can pre-multiply the received pilot signal in (4) by $(\sqrt{\rho_{\mathrm{tx}}})^{-1}\mathbf{U}^{\mathrm{H}}$. Furthermore, after the pre-multiplication, we know that at most M_{nz} first elements of the resulting signal of interest $\mathbf{\Lambda}_k^{\mathrm{P}}\tilde{\mathbf{h}}_k^{(n)}$ can be non-zero, and the rest of the elements are strictly zero due to the structure of $\mathbf{\Lambda}_k^{\mathrm{P}}$ imposed previously. The elements that are not strictly zero and contain information about the channel can be captured by $\mathbf{\Omega} \in \{0,1\}^{M_{\mathrm{nz}} \times L}$ which is a rectangular diagonal binary matrix where the first M_{nz} diagonal elements are 1 and other components are 0. First, multiplying $\mathbf{r}_{\mathrm{P},k}^{(n)}$ by $(\sqrt{\rho_{\mathrm{tx}}})^{-1}\mathbf{U}^{\mathrm{H}}$ and then by $\mathbf{\Omega}$, we get the following feedback model

$$\mathbf{y}_k^{(n)} = \mathbf{\Omega}\mathbf{\Lambda}_k^{\mathrm{P}}\mathbf{S}_k^{\mathrm{P}}\mathbf{\Gamma}\mathbf{h}_k^{(n)} + \boldsymbol{\epsilon}_k^{(n)} \tag{8}$$

where $\boldsymbol{\epsilon}_k^{(n)} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_n^2\rho_{\mathrm{tx}}^{-1}\mathbf{I}_{M_{\mathrm{nz}}})$. Finally, we quantize each complex valued element in $\mathbf{y}_k^{(n)}$ using Q_{bc} bits such that the total number of feedback bits in the BS-centric scheme is $B_{\mathrm{bc}} = Q_{\mathrm{bc}}M_{\mathrm{nz}}$.

B. Channel Estimation With Memory

As the channels are time-dependent, we can integrate information among the coherence blocks using KF [[66], Ch. 13]. We assume that Q_{bc} is large enough, so the quantization noise is not explicitly considered in the feedback model (8).[3] Linear KF is employed to filter and predict the states since the state transition model in (1) and the feedback model in (8) form a linear state-space model. The filtered channel state obeys conditional distribution $\mathbf{h}_k^{(n)}|\mathbf{y}_1^{(n)}, \ldots, \mathbf{y}_k^{(n)} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}_{k|k}^{(n)}, \mathbf{\Sigma}_{k|k}^{(n)})$ where $\boldsymbol{\mu}_{k|k}^{(n)}$ and $\mathbf{\Sigma}_{k|k}^{(n)}$ are the mean and covariance of the filtered estimate, respectively. Similarly, the prior distribution can be written as $\mathbf{h}_k^{(n)}|\mathbf{y}_1^{(n)}, \ldots, \mathbf{y}_{k-1}^{(n)} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}_{k|k-1}^{(n)}, \mathbf{\Sigma}_{k|k-1}^{(n)})$.

Using the above definitions for the state-space model, it is possible to write the linear KF prediction and update steps in the element or beamspace domain, as shown in Appendix A. The process covariance matrices $\mathbf{\Phi}_n$ and the temporal correlation coefficients η_n are assumed to be known for all $n \in [N]$. The KF should be initialized by $\boldsymbol{\mu}_{0|0}^{(n)} = \mathbf{0}$ and $\mathbf{\Sigma}_{0|0}^{(n)} = \mathbf{\Phi}_n$, corresponding to the stationary distribution of the channel in (1). However, methods extending to cases where these parameters are estimated can be considered in future work, for example, using sparse Bayesian learning (SBL) as in [56]. Although these extensions would affect the channel estimation part of this work, the pilot and precoder designs proposed in this paper require no changes.

V. Downlink Pilot Optimization Problem

Suppose that the BS has acquired the prior distribution about the beamspace domain channel vectors $\bar{\mathbf{h}}_k^{(n)} \sim \mathcal{N}_{\mathbb{C}}(\bar{\boldsymbol{\mu}}_{k|k-1}^{(n)}, \bar{\mathbf{\Sigma}}_{k|k-1}^{(n)})$, where $\bar{\boldsymbol{\mu}}_{k|k-1}^{(n)} = \mathbf{\Gamma}\boldsymbol{\mu}_{k|k-1}^{(n)}$ and $\bar{\mathbf{\Sigma}}_{k|k-1}^{(n)} =$

[2] It may occur only once (e.g., the index of the matrix from a codebook) as it does not vary between coherence blocks.

[3] In our simulations, we evaluate the impact of this assumption by comparing the performance of quantized feedback signals against non-quantized feedback signals.

$\mathbf{\Gamma}\mathbf{\Sigma}^{(n)}_{k|k-1}\mathbf{\Gamma}^{\mathrm{H}}$ are the beamspace domain mean and covariance, respectively. Since the codebook is designed to resolve distinct and approximately independent paths, the beamspace channel covariance matrix $\bar{\mathbf{\Phi}}_n = \mathbf{\Gamma}\mathbf{\Phi}_n\mathbf{\Gamma}^{\mathrm{H}}$ is nearly diagonal. Consequently, also $\bar{\mathbf{\Sigma}}^{(n)}_{k|k-1}$ for all $k \in \mathbb{N}$ can be regarded as approximately diagonal, as detailed in Appendix A. This structure allows us to approximate the MI criterion for the channel estimation problem, proved in Appendix B, by

$$\mathcal{I}\left(\mathbf{y}^{(n)}_k; \mathbf{h}^{(n)}_k\right) \approx \sum_{i=1}^{M_{\mathrm{nz}}} \log\left(1 + \frac{\rho_{\mathrm{tx}} L_{\mathrm{P}}}{\sigma_n^2}\, \delta^{(n)}_{k,\alpha_{k,i}} p_{k,i}\right) \tag{9}$$

where $\alpha_{k,i} \in [M_{\mathrm{cb}}]\ \forall\, i \in [M_{\mathrm{rf}}]$ refers to the beam index in the codebook assigned to the specific RF chain i (i.e., $[\mathbf{S}^{\mathrm{P}}_k]_{i,\alpha_{k,i}} = 1$), $\delta^{(n)}_{k,i}$ is the ith diagonal element of $\bar{\mathbf{\Sigma}}^{(n)}_{k|k-1}$, and $p_{k,i} = L_{\mathrm{P}}^{-1}[\mathbf{\Lambda}^{\mathrm{P}}_k]^2_{i,i}$ is the power allocated to ith RF chain. The power allocations $p_{k,i}\ \forall\, i \in [M_{\mathrm{rf}}]$ must sum up to 1, and comprise with the constraint $p_{k,i} \geq 0\ \forall\, i \in [M_{\mathrm{nz}}]$, while $p_{k,i} = 0\ \forall\, i = M_{\mathrm{nz}}+1, \ldots, M_{\mathrm{rf}}$, because only the first M_{nz} RF chains are selected by matrix $\mathbf{S}^{\mathrm{P}}_k$.

Define vector $\mathbf{p}_k = [p_{k,1}, \ldots, p_{k,M_{\mathrm{nz}}}]^{\mathrm{T}}$ and set $\mathcal{A}_k = \{\alpha_{k,1}, \ldots, \alpha_{k,M_{\mathrm{nz}}}\} \in \mathcal{C}\left([M_{\mathrm{cb}}], M_{\mathrm{nz}}\right)$ where $\mathcal{C}\left([M_{\mathrm{cb}}], M_{\mathrm{nz}}\right)$ is a collection of sets that comprises all subsets of the beam index set $[M_{\mathrm{cb}}]$ with M_{nz} distinct elements. Therefore, the pilot design problem can be written as

$$\max_{\mathcal{A}_k, \mathbf{p}_k} \sum_{n=1}^{N} \sum_{i=1}^{M_{\mathrm{nz}}} \log\left(1 + \frac{\rho_{\mathrm{tx}} L_{\mathrm{P}}}{\sigma_n^2}\, \delta^{(n)}_{k,\alpha_{k,i}} p_{k,i}\right) \tag{10a}$$

$$\text{s.t. } \mathcal{A}_k \in \mathcal{C}\left([M_{\mathrm{cb}}], M_{\mathrm{nz}}\right) \tag{10b}$$

$$\mathbf{1}^{\mathrm{T}}\mathbf{p}_k = 1 \text{ and } \mathbf{p}_k \geq 0. \tag{10c}$$

where (10a) is the approximate MI between the communications channels and feedback signals, (10b) is the constraint for the selected beams, and (10c) ensures that the total power constraint is satisfied. Solving the optimization problem in (10) aims to find the M_{nz} best beams and their power allocations, such that they maximize the sum of approximate MIs given the predicted UE channel statistics. In particular, this approach allows for designing downlink pilots sequentially to estimate multiple channels most effectively in terms of the approximate MI.

Given the index set $\mathcal{A}_k$, the problem in (10) can be solved efficiently using the well-known water-filling algorithm [67]. When there is just a *single user* ($N = 1$), solving for the optimal $\mathcal{A}_k$, i.e., the selected beams, is trivial because the logarithm appearing in equation (10a) is a monotonically increasing function. Therefore, the optimal beam set $\mathcal{A}_k$ is obtained by selecting M_{nz} beams with the largest values of $\delta^{(1)}_{k,\alpha}$, where $\alpha \in [M_{\mathrm{cb}}]$. When there are *multiple users* ($N > 1$), the optimization must account for user-specific uncertainties $\delta^{(n)}_{k,\alpha}$ and noise levels σ_n^2. Therefore, the optimal beam selection problem cannot be decoupled from the power allocation problem. Motivated by the optimal solution in the single user case, we propose an efficient suboptimal approach, where a score $\sum_{n=1}^{N} \delta^{(n)}_{k,\alpha}/\sigma_n^2$ is defined for each beam index $\alpha \in [M_{\mathrm{cb}}]$. The set $\mathcal{A}_k$ is then formed by selecting the M_{nz} beams with the highest scores. However, different score expressions can also be used, such as the maximum normalized uncertainty among users $\max_{n\in[N]} \delta^{(n)}_{k,\alpha}/\sigma_n^2$.

VI. Downlink Precoding

Denote the precoding matrix for data transmission $\ell \in [L]\backslash\mathcal{L}_{\mathrm{P}}$ as $\mathbf{W}_k = \mathbf{A}^{\mathrm{D}}_k\mathbf{D}^{\mathrm{D}}_k = [\mathbf{w}^{(1)}_k, \ldots, \mathbf{w}^{(N)}_k]$ where $\mathbf{w}^{(n)}_k \in \mathbb{C}^{M\times 1}$ is the beamforming vector of user $n \in [N]$. Due to the feedback delay, $\mathbf{W}_k$ is designed during coherence block $k-1$ using the prior with parameters $\boldsymbol{\mu}^{(n)}_{k|k-1}$ and $\mathbf{\Sigma}^{(n)}_{k|k-1}$. The beamforming gain of UE $n \in [N]$ is $|(\mathbf{h}^{(n)}_k)^{\mathrm{T}}\mathbf{w}^{(n)}_k|^2$ and the unintentional interference caused to UE n from the symbols dedicated to UE $n' \in [N] \setminus \{n\}$ is $|(\mathbf{h}^{(n)}_k)^{\mathrm{T}}\mathbf{w}^{(n')}_k|^2$. Thus, the expected beamforming gain ($n = n'$) or unintentional interference ($n \neq n'$) can be written as

$$\mathbb{E}\left[|(\mathbf{h}^{(n)}_k)^{\mathrm{T}}\mathbf{w}^{(n')}_k|^2 \,\middle|\, \boldsymbol{\mu}^{(n)}_{k|k-1}, \mathbf{\Sigma}^{(n)}_{k|k-1}\right] = (\mathbf{w}^{(n')}_k)^{\mathrm{H}}\left(\mathbf{\Psi}^{(n)}_{k|k-1}\right)^{*}\mathbf{w}^{(n')}_k \tag{11}$$

where

$$\mathbf{\Psi}^{(n)}_{k|k-1} = \boldsymbol{\mu}^{(n)}_{k|k-1}(\boldsymbol{\mu}^{(n)}_{k|k-1})^{\mathrm{H}} + \mathbf{\Sigma}^{(n)}_{k|k-1} \tag{12}$$

is the second moment of the prior distribution of the nth UE channel. Therefore, the maximum expected beamforming gain, or interference power, is obtained by transmitting towards the maximum conjugate eigenvector of the matrix $\mathbf{\Psi}_n^{(k|k-1)}$[4]. We can approximate any standard precoding scheme (e.g., maximum-ratio transmission (MRT), zero-forcing (ZF), minimum mean square error (MMSE)) by setting

$$\widehat{\mathbf{h}}^{(n)}_k = e_{\max}\left(\mathbf{\Psi}^{(n)}_{k|k-1}\right) \tag{13}$$

where $\widehat{\mathbf{h}}^{(n)}_k$ is the channel estimate used for the precoding, and $e_{\max}(\cdot)$ takes the eigenvector corresponding to the maximum eigenvalue of the input matrix. This results in the exact precoding schemes in the special case when $\mathbf{\Sigma}^{(n)}_{k|k-1} = \mathbf{0}$.

We particularly consider ZF precoding within the context of a hybrid digital-analog architecture, which we refer to as HZF. Typically, HZF algorithms [68], [69], [70], [71] are based on a formula

$$\mathbf{D}^{\mathrm{D}}_k = \widetilde{\mathbf{H}}^{\mathrm{H}}_k(\widetilde{\mathbf{H}}_k\widetilde{\mathbf{H}}^{\mathrm{H}}_k)^{-1}\mathbf{\Lambda}^{\mathrm{D}}_k \tag{14}$$

where $\widetilde{\mathbf{H}}_k = \widehat{\mathbf{H}}_k\mathbf{A}^{\mathrm{D}}_k$ with the channel estimate matrix $\widehat{\mathbf{H}}_k = [\widehat{\mathbf{h}}^{(1)}_k, \ldots, \widehat{\mathbf{h}}^{(N)}_k]^{\mathrm{T}}$, and $\mathbf{\Lambda}^{\mathrm{D}}_k$ is a diagonal matrix implementing column normalization to satisfy the total power constraint. The matrix $\mathbf{\Lambda}^{\mathrm{D}}_k$ may be obtained by the water-filling algorithm or by normalizing each column to have equal power, for example.[5]

The proposed algorithm uses (13) to obtain the channel estimates. Furthermore, our low-complexity precoding method differs from the algorithms in [68], [69], [70], [71] in how the analog precoder matrix $\mathbf{A}^{\mathrm{D}}_k$ is designed. We use the same analog

[4]The conjugate appears in (11) as the signal model uses transpose instead of conjugate transpose.

[5]In simulations, we normalize each column to have equal power.

precoder structure $\mathbf{A}_k^{\mathrm{D}} = (\mathbf{S}_k^{\mathrm{D}}\mathbf{\Gamma})^{\mathrm{T}}$ as used for the pilots.[6] Therefore, the analog precoders can be optimized by determining the selected beams, i.e., optimizing the matrix $\mathbf{S}_k^{\mathrm{D}}$. We propose to select the beams by first computing the precoder for a fully digital array and selecting the beams associated with the most power.

VII. Numerical Examples

In the numerical examples, various downlink channel sounding strategies are evaluated under a limited digital feedback. In particular, we compare the performance gains achieved for downlink data transmission in terms of the sum of achievable rates, based on the Shannon–Hartley theorem (capacity expression), hereafter referred to as the *sum rate*. The sum rate at each block k is defined as

$$\zeta_k = \left(1 - \frac{L_{\mathrm{P}}}{L}\right) \Delta_f \sum_{n=1}^{N} \log_2\left(1 + \gamma_k^{(n)}\right) \tag{15}$$

where $(1 - L_{\mathrm{P}}/L)$ accounts for the pilot overhead, Δ_f is the bandwidth, and

$$\gamma_k^{(n)} = \frac{|(\mathbf{h}_k^{(n)})^{\mathrm{T}}\mathbf{w}_k^{(n)}|^2}{\sum_{n' \neq n} |(\mathbf{h}_k^{(n)})^{\mathrm{T}}\mathbf{w}_k^{(n')}|^2 + \frac{\sigma_n^2}{\rho_{\mathrm{tx}}}} \tag{16}$$

is the signal-to-interference-plus-noise ratio (SINR) of user n in a linear scale. The pilot length affects the sum rate in (15) both explicitly via the pilot overhead and implicitly via the channel estimation capability. To ensure fair comparison among channel sounding algorithms, the pilot length is fixed to $L_{\mathrm{P}} = 16$ unless otherwise noted. Also, the sensitivity to the choice of L_{P} is studied, as shown in Fig. 4. The linear KF and the HZF precoder are used in conjunction with all channel sounding algorithms considered to enable *apples-to-apples* comparison. The simulation model and parameters are described in Section VII-C and summarized in Table I.

A. Baseline Algorithms for Channel Sounding

The baseline algorithms are motivated by methods applied in current wireless communications systems and available in the open literature. However, there is a lack of existing methods for optimizing downlink pilots for multiuser scenarios and systems employing hybrid digital-analog architectures. The considered baseline channel sounding algorithms are briefly explained below.

- **Perfect CSI**: The channel sounding method is not considered; instead, the precoder uses perfect CSI. In other words, it is assumed that the downlink channels $\mathbf{h}_k^{(n)}$ are known for all $k \in \mathbb{N}$ and $n \in [N]$. This method presents the upper bound for the sum rate performance.
- **Only CDI**: The channel sounding method is not considered; instead, the precoder uses only the channel distribution information (CDI). In other words, it assumed that the downlink channel covariance matrices $\mathbf{\Phi}_n$ are known for all $n \in [N]$. The HZF precoding is implemented by using $\mathbf{\Psi}_{k|k-1}^{(n)} = \mathbf{\Phi}_n$ in (13).
- **Random**: UE-centric method, where pseudo-random pilots are transmitted in a round-robin manner, and the pilots are assumed to be known by the UE. The analog pilot matrices $\mathbf{A}_k^{\mathrm{P}}$ are generated by sampling random phase shifts from a uniform distribution. The digital pilot vectors $\mathbf{d}_k^{\mathrm{P}}[\ell] \forall\, \ell \in \mathcal{L}_{\mathrm{P}}$ (i.e., columns of the matrix $\mathbf{D}_k^{\mathrm{P}}$) are sampled from $\mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \frac{1}{M_{\mathrm{rf}}}\mathbf{I}_{M_{\mathrm{rf}}}\right)$.
- **Non-Optimized Proposed**: Reduced complexity variation of the proposed method. In particular, the vector $\mathbf{p}_k$ allocating powers to the beams in (10) is sampled from a uniform distribution defined in interval $[0, 1]$ and normalized to sum up to 1, and $\mathbf{S}_k^{\mathrm{P}}$ is sampled such that each beam has an equal probability of being selected.
- **Exhaustive**: The pilot is comprised of probing one beam at a time, corresponding to something similar to the exhaustive search method in BA literature [41]. This can be presented using the pilot and feedback model developed in this paper when the maximum number of active beams M_{nz} is restricted to one.
- **WMSE [72]**: A method proposed in [72] that optimize the weighted sum of channel estimation mean square errors (MSEs) across all UEs using CDI. In this approach, the pilot is designed at the BS side while the channel is estimated at the UE side, hence it is based on the UE-centric feedback scheme. The channel covariance matrix $\mathbf{\Phi}_n$ for all $n \in [N]$ is assumed to be known by the BS, enabling centralized optimization of the pilots. The optimized pilot is then communicated back to the UEs because UE requires it for estimating the channel. In practice, there is overhead associated with obtaining the matrices $\mathbf{\Phi}_n$ and distributing the optimized pilots, but this is not accounted for in the overhead computations. Additionally, the design does not account for the constraints of a hybrid analog-digital architecture. This gives the method a slight performance advantage over the proposed approach and other considered alternatives.

B. Feedback Quantization

In this subsection, we describe the feedback quantization methods used for the UE-centric baselines and the proposed BS-centric scheme.

The UE-centric baselines use a similar channel quantization methodology defined for type-II codebook feedback in 3GPP standard (release 17) [24]. In particular, the quantization is done for the beamspace channel $\bar{\mathbf{h}}_k^{(n)}$. This quantization is done in two stages:

1) The method finds the element of $\bar{\mathbf{h}}_k^{(n)}$ with the largest amplitude and normalizes the beamspace channel with the corresponding element.
2) The method quantizes the phase and amplitude of M_{fb} elements of the normalized beamspace channel, with M_{fb} representing the number of beamspace coefficients selected for feedback. The amplitude is quantized with $Q_{\mathrm{uc}}^{\mathrm{A}}$ bits with 3dB intervals into

[6] The beam codebook $\mathbf{\Gamma}$ can be different for the data transmission. For simplicity, we do not define different variable for it.

decibel bins $\{0, -3, \ldots, -3(2^{Q_{\rm uc}^{\rm A}} - 1)\}$. The phase is quantized with $Q_{\rm uc}^{\rm P}$ bits uniformly into bins $\{0, 2\pi/2^{Q_{\rm uc}^{\rm P}}, \ldots, 2\pi(2^{Q_{\rm uc}^{\rm P}} - 1)/2^{Q_{\rm uc}^{\rm P}}\}$.

For the UE-centric scheme, we use $Q_{\rm uc}^{\rm A} = 3$ and $Q_{\rm uc}^{\rm P} = 4$ bits that coincide with the 3GPP standard [24]. There are diminishing returns in increasing $Q_{\rm uc}^{\rm A}$ and $Q_{\rm uc}^{\rm P}$. In addition, the UE needs to inform which beams are returned. Therefore, as derived in Appendix C, the number of feedback bits is

$$B_{\rm uc} = \log_2\left(\frac{M!}{M_{\rm fb}!(M - M_{\rm fb})!}\right) + Q_{\rm uc} M_{\rm fb} \tag{17}$$

where $Q_{\rm uc} = Q_{\rm uc}^{\rm A} + Q_{\rm uc}^{\rm P} = 7$ is the number of bits per feedback element. Thus, the feedback codebook performance most significantly depends on $M_{\rm fb}$. For example, $M_{\rm fb}$ in $\{1, 4, 7\}$ correspond to a number of feedback bits $B_{\rm uc} \in \{14, 40, 86\}$, respectively.

The proposed BS-centric approach quantizes the feedback elements in (8) using a similar approach as with the UE-centric approach. However, there is no normalization stage since the proposed approach requires scale information. Therefore, the amplitude is quantized with uniform interval $A_{\rm I} = 1.5$ in dB scale between the maximum value $A_{\max} = -100$ dB and minimum value $A_{\min} = A_{\max} - A_{\rm I}(2^{Q_{\rm bc}^{\rm A}} - 1) = -194.5$dB where $Q_{\rm bc}^{\rm A} = 6$ is the number of bits. The BS-centric method also uses $Q_{\rm bc}^{\rm P} = 4$ bits for phase quantization. Further increasing $Q_{\rm bc}^{\rm A}$ and $Q_{\rm bc}^{\rm P}$ leads to diminishing returns. For example, $M_{\rm nz}$ in 1,4,8 correspond to $B_{\rm bc} \in \{10, 40, 80\}$ feedback bits.

C. Simulation Environment

The BS has a critically sampled ULA with the steering vector $\mathbf{a}(\theta)$ with the element $m \in [M]$ defined as

$$[\mathbf{a}(\theta)]_m = \frac{1}{\sqrt{M}} e^{-j(m-1)\pi \sin(\theta)} \tag{18}$$

where θ is the direction of the steering vector. The codebook matrix $\mathbf{\Gamma} \in \mathbb{C}^{M \times M}$ is a unitary DFT matrix. We use a first-order Markov model for the channel in (1) where $\eta_n = \eta_{n'}$ for all $n \in [N]$ and $n' \in [N]$. The channel covariance matrix $\mathbf{\Phi}_n$ is computed using the Saleh-Valenzuela channel model [73]. Therefore, $\mathbf{\Phi}_n$ is modeled as follows

$$\mathbf{\Phi}_n = \sum_{i=1}^{N_{\rm p}} \xi_i^{(n)} \mathbf{a}(\theta_i^{(n)}) \mathbf{a}^{\rm H}(\theta_i^{(n)}) \tag{19}$$

where $N_{\rm p}$ is the number of propagation paths, and $\xi_i^{(n)}$ is the average path gain with angle of departure (AoD) $\theta_i^{(n)}$. The path gains $\xi_i^{(n)}$ are computed using the log-distance path loss model [74]

$$10\log_{10}(\xi_i^{(n)}) = -10\log_{10}\left(\left(4\pi\frac{f_c}{v_c}\right)^2 d_{\rm ue}^{\beta}\right) + X_i^{(n)} \tag{20}$$

where f_c is the carrier frequency, $d_{\rm ue}$ is the distance (the same distance is used for all users), v_c is the speed of light, β is the path-loss exponent, and $X_i^{(n)} \sim \mathcal{N}(0, \sigma_X^2)$ is the shadowing loss in dB scale with a standard deviation of σ_X. In addition, the path AoDs are generated by first sampling the UE directions $\bar{\theta}_n \; \forall \, n \in [N]$ from the uniform distribution $\mathcal{U}(-\pi/2, \pi/2)$, and then sampling $\theta_i^{(n)} - \bar{\theta}_n \sim \mathcal{U}(-\sigma_{\rm aod}/2, \sigma_{\rm aod}/2)$. The simulation parameters are shown in Table I. All results are averaged over 100 Monte Carlo runs, each simulation consisting of 100 time blocks.

TABLE I
SIMULATION PARAMETERS

Description	Symbol	Value
Path loss exponent	β	1.98
Shadowing standard deviation	σ_X	3.1
AoD spread	$\sigma_{\rm aod}$	0.26 rad
Total power	$\rho_{\rm tx}$	$\{0.1, 1, 10\}$ W
Bandwidth	Δ_f	1 MHz
Carrier frequency	f_c	28 GHz
Noise power	$\sigma_n^2 \forall n$	-113.98 dBm
# of UEs	N	3
# of RF chains	$M_{\rm rf}$	16
# of antenna elements	M	128
# of paths	$N_{\rm p}$	6
Pilot length	$L_{\rm P}$	16
Monte Carlo iterations	-	100
Reference distance	$d_{\rm ue}$	8 km
Temporal correlation	$\eta_n \forall n$	$\{0.9983, 0.9589, 0.8515\}$
Block length	τ	100 μs

We compare the performance at three different power levels $\rho_{\rm tx} \in \{0.1, 1.0, 10.0\}$ and using correlation coefficients $\eta_n \in \{0.9983, 0.9589, 0.8515\}$. The varied power levels result in different signal-to-noise ratios (SNRs). In particular, the maximum average SNR given by

$$\bar{\gamma}_k^{(n)} = \max_{\mathbf{w}_k^{(n)}} \frac{(\mathbf{w}_k^{(n)})^{\rm H} \mathbf{\Phi}_n^* \mathbf{w}_k^{(n)}}{\sigma_n^2/\rho_{\rm tx}} \text{ s.t. } \|\mathbf{w}_{\rm k}^{(n)}\|_2^2 \leq 1 \tag{21}$$

is used to characterize the scenarios. With the different power levels, SNRs are in $[-1.9, 2.7]$ dB, $[8.1, 12.7]$ dB, and $[18.1, 22.7]$ dB, respectively. We refer to these three SNR regimes by their average SNRs among the users: 0 dB (low SNR regime), 10 dB (medium SNR regime), and 20 dB (high SNR regime). Specific maximum speeds or Doppler spread can be related to the temporal-correlation coefficient in (1) using Jake's model $\eta_n = J_0(2\pi f_d \tau)$ where τ is the block length in seconds, f_d is the maximum Doppler shift, and $J_0(\cdot)$ is the 0-th order Bessel function of the first kind. With the simulation parameters shown in Table I, the correlation coefficients $\eta_n \in \{0.9983, 0.9589, 0.8515\}$ correspond to maximum speeds of 5 km/h, 25 km/h, and 50 km/h, respectively.[7]

D. Performance Under Limited Feedback Rates

This section investigates the performance of the proposed method in comparison to the baselines when the feedback rate is limited. Fig. 2 shows the achievable sum rate over time blocks for the methods considered under varying SNR and temporal correlation conditions. The proposed, random, and WMSE methods each use 40-bit CSI feedback, while the exhaustive search baseline only requires 10-bit feedback, as only

[7]These maximum speeds depend on the carrier frequency and the block length. With smaller frequencies and longer block lengths, higher speeds are possible with similar α values, or vice versa.

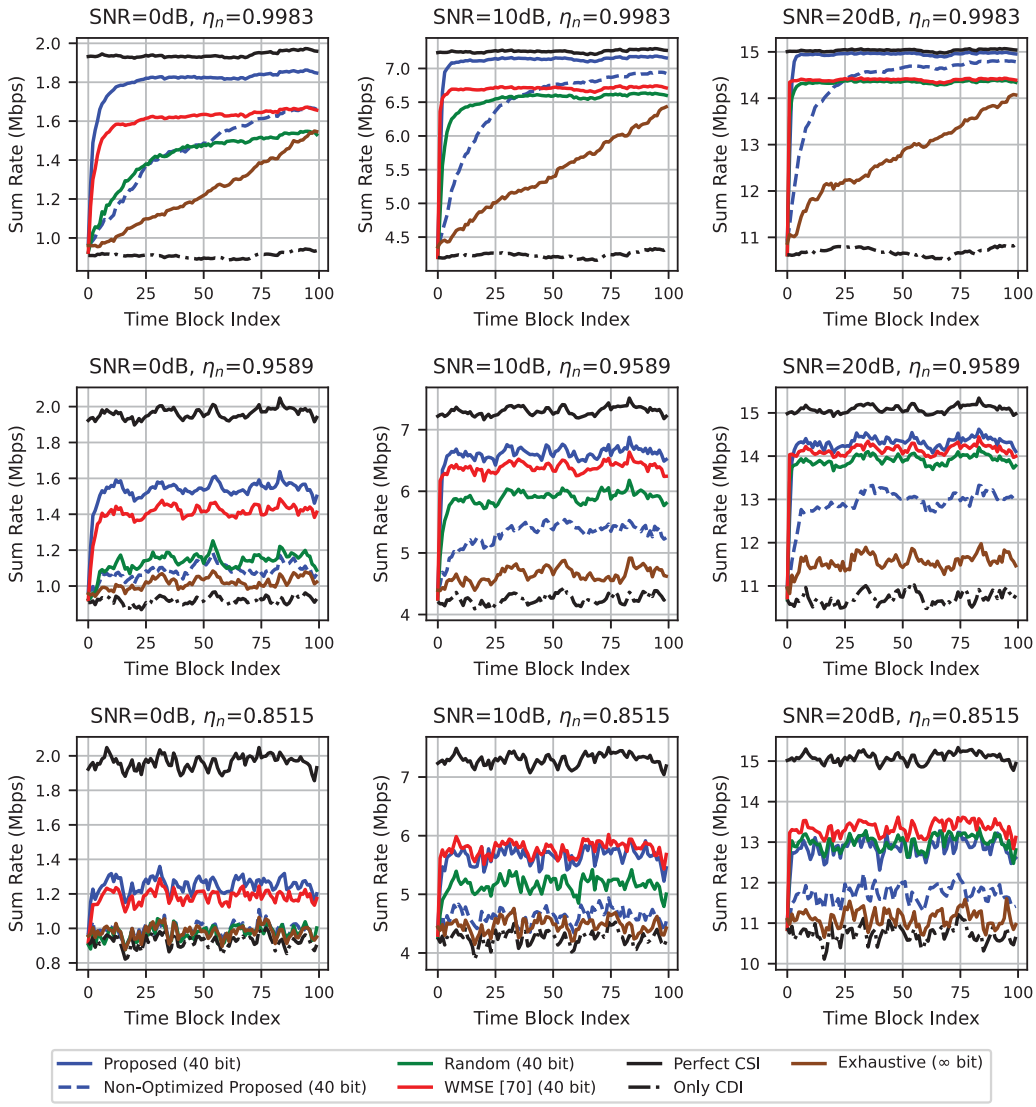

Fig. 2. Achievable sum rate performance over time under different SNR and temporal correlation settings. The proposed method outperforms the baselines, particularly in scenarios with low SNR and high correlation. Performance gains diminish in lower correlation regimes, where all methods achieve lower sum rates.

one beam is sounded at a time. It can be seen that the proposed method outperforms the baselines, especially at low SNR and high correlation regimes. In particular, the results indicate that the proposed method acquires the CSI more quickly than the baselines during the initial time blocks and can track the channel more accurately, as evidenced by the sum rate stabilizing at higher sum rate values.

Regarding the different SNR and temporal correlation conditions, as expected, all methods perform better under higher SNR. Furthermore, under larger temporal correlations, higher sum rates are achievable, approaching those of a perfect CSI precoder. Conversely, under smaller temporal correlations, the performance is only marginally better than with CDI only precoding. In the high correlation scenario with $\eta_n = 0.9983$ and SNR $= 20$ dB, the proposed method is the only one that nearly matches the performance of the perfect CSI precoder. Although the random and WMSE baselines also converge quickly, their performance is constrained by the limited CSI feedback. The proposed method mitigates this limitation by estimating the channel at the BS side, which is enabled by the proposed pilot and feedback structure. This approach seems to be more robust to the feedback rate constraints. However, the effectiveness of the proposed scheme heavily depends on the design of the pilots. The performance degrades noticeably when powers are allocated to the beams randomly or when beams are scanned exhaustively, as demonstrated by the non-optimized variant of the proposed method and the exhaustive search baseline, respectively. When the temporal coefficient is lower $\eta_n \in \{0.9589, 0.8515\}$, the performance improvement compared to the WMSE baseline is not as significant as with $\eta_n = 0.9983$, especially under the high SNR regime. In fact, when $\eta_n = 0.8515$ and SNR$= 20$ dB, the WMSE method slightly outperforms the proposed method. Under these conditions, even the random baseline performs

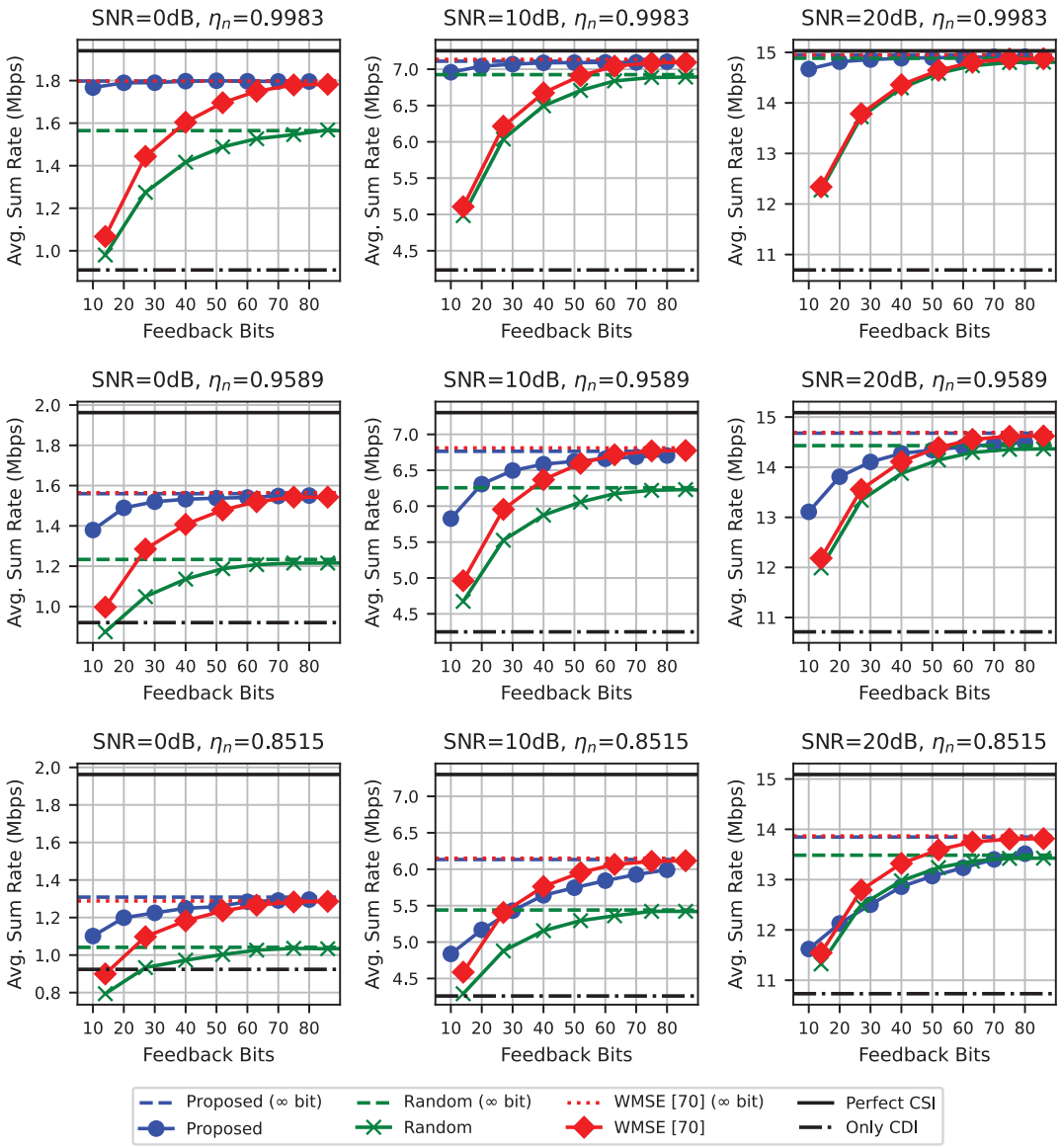

Fig. 3. Average achievable sum rate as a function of the number of feedback bits. The proposed method achieves significantly higher sum rates under low feedback conditions and matches WMSE performance under unlimited feedback.

well, achieving a similar level of performance to the proposed method.

To better understand the implications of limited feedback, we examine the tradeoff between feedback rates and achievable sum rates, averaged over time blocks. The feedback rate of the proposed BS-centric scheme is varied by adjusting the parameter $M_{\mathrm{nz}} = 1, \ldots, 8$, while for the UE-centric scheme $M_{\mathrm{fb}} = 1, \ldots, 7$. Fig. 3 illustrates the average sum rate as a function of the number of feedback bits for the proposed, random, and WMSE methods. It can be seen that the proposed method delivers significantly better performance, especially with low feedback rates, compared to the other methods tested. Furthermore, the proposed method incurs only a slight performance loss when decreasing the feedback rate under high temporal correlation coefficients. The performance of the random and WMSE baselines is significantly dependent on the feedback rates, with performance being very weak at low rates. Especially, under the low temporal correlation and high SNR regime, the performance degrades noticeably as the feedback rate decreases. This is because lower temporal correlations indicate that less memory of the previous channel instances can be exploited. Therefore, channels can be estimated and tracked more accurately with higher feedback rates that allow more beams to be probed simultaneously.

With an unlimited feedback rate shown in Fig. 3, the performance of the proposed method is similar to the WMSE method. Furthermore, both methods are distinctively better than the random baseline, except in the high SNR regime, where the random baseline also works well. The similarity in the performance between the proposed and WMSE may be explained due to the well-established connection between MI and MSE objectives [75]. Although the proposed method uses the approximate MI

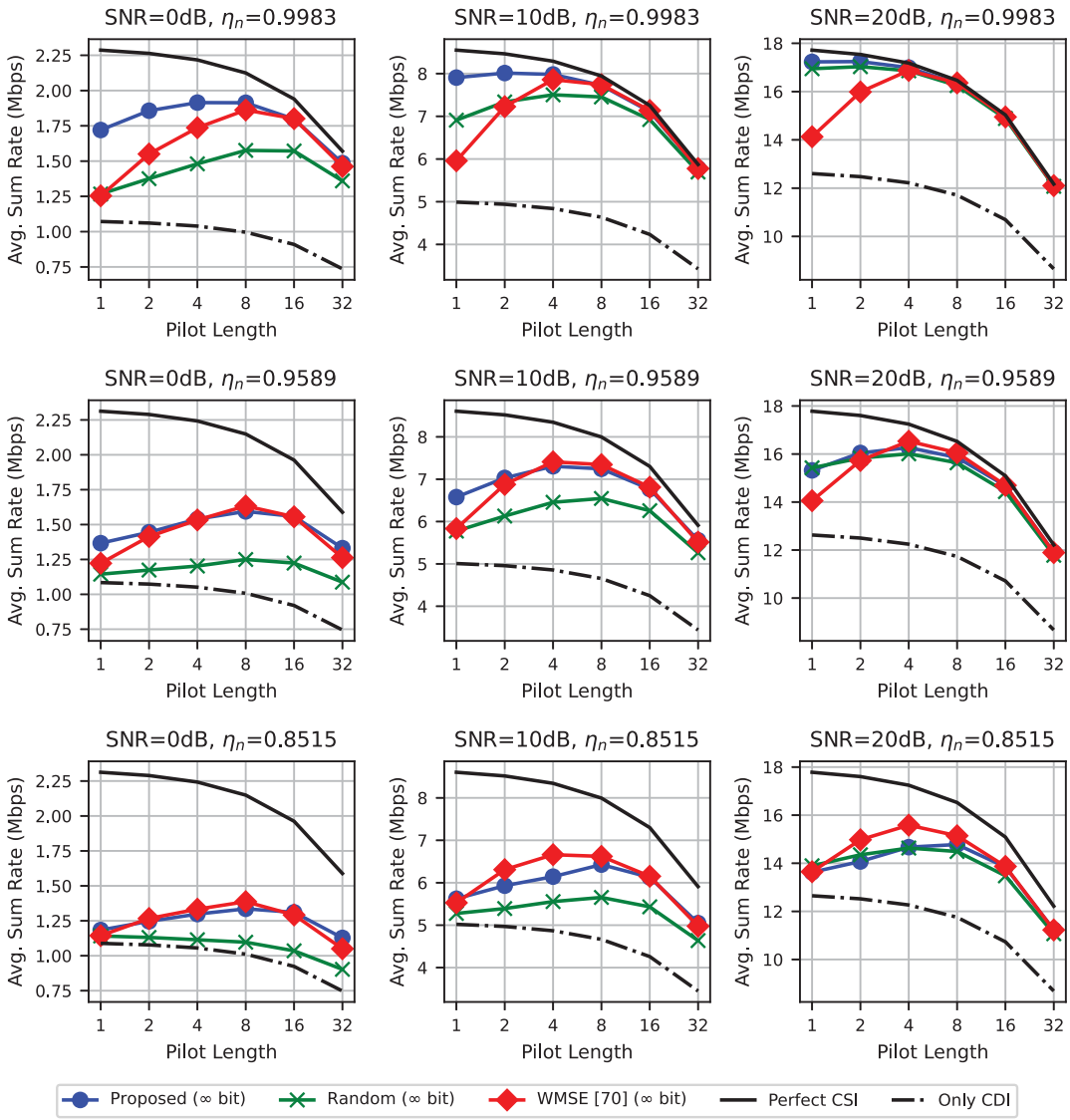

Fig. 4. Average achievable sum rate as a function of pilot length under unlimited feedback. A tradeoff between pilot diversity and pilot overhead is optimally achieved with 4 to 8 pilots, depending on the SNR and temporal correlation coefficient. Especially under high temporal correlation, the proposed method outperforms the WMSE baseline at shorter pilot lengths.

optimization criterion instead of the exact MI, the results remain competitive. It is also worth noting that, unlike the proposed method, the WMSE is not directly applicable to hybrid digital-analog hardware. Moreover, the practical implementation of WMSE requires extra signaling and coordination for sharing the channel covariance matrices and the downlink pilots.

E. Sensitivity to Pilot Lengths

In this section, we evaluate the sensitivity of the proposed and baseline approaches to the pilot length parameter. For simplicity, we consider unlimited feedback for each method. Fig. 4 shows the average achievable sum rates as a function of the pilot length under different SNR and temporal correlation conditions. It shows that there is a tradeoff between increasing the pilot length to gain more diverse probing signals and the overhead introduced by the pilot. In other words, larger pilot lengths improve the channel estimation capability, but decrease the scaling factor $1 - L_{\mathrm{P}}/L$ in (15). The figure indicates that the optimal pilot length lies between 4 and 8 symbols, depending on the SNR and temporal correlation conditions. There is no significant difference in the optimal pilot length among the considered methods. The optimal pilot being close to the number of channel paths $L_{\mathrm{P}} = 6$ is expected, because L_{P} is also the rank of the channel covariance matrices. However, the proposed method is more robust to the variations in the pilot length compared to the WMSE baseline. This is especially visible under the high temporal correlation coefficient $\eta_n = 0.9983$ scenario, where the memory of the KF can be exploited. In particular, the proposed method maintains strong performance even with shorter pilot sequences, whereas the WMSE baseline suffers significant degradation in such cases.

VIII. Conclusion

We proposed a novel BS-centric channel sounding method for multiuser massive MIMO systems. This method is specifically developed for feedback-based communications systems that work when channel reciprocity cannot be exploited, for example, when using FDD. This approach differs from traditional UE-centric methods, particularly developed to address multiuser settings with spatial multiplexing. The UE-centric approach has several drawbacks, including additional feedback bits, the potential for mutual interference between users, and the pilots not being centrally optimized for multiuser channel estimation. Our BS-centric method addresses these issues using a beamspace-based pilot model, reducing the computational load on the UEs and facilitating accurate channel estimation at the BS.

The proposed method is specifically designed for hybrid RF architectures and applicable commonly used multiantenna configurations where beamspace processing may be employed. DLAs are particularly interesting in this context because they lend themselves to low-complexity processing. We developed an efficient pilot optimization algorithm to maximize the approximate MI between the received feedback signal and channel coefficients. We also proposed an efficient ZF precoding algorithm designed for hybrid RF chains.

Comparative analysis showed that our proposed method outperforms the considered baseline algorithms. The improvement is significant in scenarios with low feedback rates and at the lower SNR regime. However, there are areas for further exploration and improvement in future research. For example, extension from single antenna to multi-antenna UEs and wideband communications systems. Also, it is worth investigating the case of a priori unknown channel covariance matrices and very dynamic settings where the channel correlation coefficient is small. In particular, utilizing integrated sensing and communications (ISAC) [76], [77], [78] to obtain prior information on the channel is a promising avenue for future research.

Appendix

A. Kalman Filtering Equations

Using the state-space model definitions in Section IV-B, we can write linear KF prediction and update steps in the beamspace domain, where $\bar{\mathbf{h}}_k^{(n)} = \boldsymbol{\Gamma}\mathbf{h}_k^{(n)}$ and $\bar{\boldsymbol{\Phi}}_n = \boldsymbol{\Gamma}\boldsymbol{\Phi}_n\boldsymbol{\Gamma}^{\mathrm{H}}$. According to [[66], Ch. 13], the equations required for the prediction step can be written as

$$\bar{\boldsymbol{\mu}}_{k|k-1}^{(n)} = \eta_n\, \bar{\boldsymbol{\mu}}_{k-1|k-1}^{(n)}, \tag{22a}$$

$$\bar{\boldsymbol{\Sigma}}_{k|k-1}^{(n)} = \eta_n^2\, \bar{\boldsymbol{\Sigma}}_{k-1|k-1}^{(n)} + (1-\eta_n^2)\, \bar{\boldsymbol{\Phi}}_n. \tag{22b}$$

Define matrix $\mathbf{F}_k = \boldsymbol{\Omega}\boldsymbol{\Lambda}_k^{\mathrm{P}}\mathbf{S}_k^{\mathrm{P}}$, which is a diagonal measurement matrix corresponding to the proposed pilot and feedback model in (8). Following [[66], Ch. 13], we can write the equations corresponding to the update step of the KF as

$$\mathbf{C}_k^{(n)} = \mathbf{F}_k\bar{\boldsymbol{\Sigma}}_{k|k-1}^{(n)}\mathbf{F}_k^{\mathrm{H}} + \sigma_n^2\rho_{\mathrm{tx}}^{-1}\mathbf{I}_{M_{\mathrm{nz}}}, \tag{23a}$$

$$\mathbf{G}_k^{(n)} = \bar{\boldsymbol{\Sigma}}_{k|k-1}^{(n)}\mathbf{F}_k^{\mathrm{H}}(\mathbf{C}_k^{(n)})^{-1}, \tag{23b}$$

$$\bar{\boldsymbol{\mu}}_{k|k}^{(n)} = \bar{\boldsymbol{\mu}}_{k|k-1}^{(n)} + \mathbf{G}_k^{(n)}\left(\mathbf{y}_k^{(n)} - \mathbf{F}_k\bar{\boldsymbol{\mu}}_{k|k-1}^{(n)}\right), \tag{23c}$$

$$\bar{\boldsymbol{\Sigma}}_{k|k}^{(n)} = \bar{\boldsymbol{\Sigma}}_{k|k-1}^{(n)} - \mathbf{G}_k^{(n)}\mathbf{F}_k^{(n)}\bar{\boldsymbol{\Sigma}}_{k|k-1}^{(n)}. \tag{23d}$$

As the beamspace codebook is designed to resolve distinct paths in different directions, the process noise covariance matrix $\bar{\Phi}_n = \lim_{k\to\infty}\mathrm{cov}(\bar{\mathbf{h}}_k|\bar{\mathbf{h}}_0^{(n)})$ approximately obeys a diagonal structure. For example, in the case of DFT codebooks for ULA, if all channel paths are at the centers of the codebook beams, there is no leakage to the other beams. However, in practice, the same communication path may be illuminated by multiple adjacent beams, and hence, correlation would be introduced. Additionally, some of this correlation is introduced due to sidelobes. However, the off-diagonal values are still typically significantly smaller compared to the diagonal values. It is well justified to set the initial covariance matrix as $\bar{\boldsymbol{\Sigma}}_{0|0} = \bar{\Phi}_n$ to represent the stationary distribution. As the prediction and update equations in (22) and (23) involve only operations with diagonal or approximately diagonal matrices, the resulting matrices $\bar{\boldsymbol{\Sigma}}_{k|k-1}^{(n)}$ and $\bar{\boldsymbol{\Sigma}}_{k|k}^{(n)}$ also convey nearly diagonal form for any $k \in [K]$.

B. Approximate MI Criterion

Consider the pilot matrix $\mathbf{X}_k^{\mathrm{P}}$ with the structure in (6) and the feedback model in (8). For convenience, re-write the feedback model as

$$\mathbf{y} = \boldsymbol{\Omega}\boldsymbol{\Lambda}\mathbf{S}\boldsymbol{\Gamma}\mathbf{h} + \boldsymbol{\epsilon} \tag{24}$$

where we omit the superscript $\cdot^{(n)}$ about the user index and subscript $\cdot_k$ about the time block index for clarity, and $\boldsymbol{\epsilon} \sim \mathcal{N}_{\mathbb{C}}(0, \frac{\sigma^2}{\rho_{\mathrm{tx}}}\mathbf{I}_{M_{\mathrm{nz}}})$. We assume that the prior or posterior distribution of the channel coefficient is $\mathbf{h} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is the mean and $\boldsymbol{\Sigma}$ is the covariance, computed, for example, via Kalman filtering in Section IV-B. Thus, the marginal distribution of the feedback signal is

$$\mathbb{P}(\mathbf{y}) = \mathcal{N}_{\mathbb{C}}\left(\boldsymbol{\Omega}\boldsymbol{\Lambda}\mathbf{S}\boldsymbol{\Gamma}\boldsymbol{\mu}, \boldsymbol{\Omega}\boldsymbol{\Lambda}\mathbf{S}\bar{\boldsymbol{\Sigma}}\mathbf{S}^{\mathrm{H}}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{\Omega}^{\mathrm{H}} + \frac{\sigma^2}{\rho_{\mathrm{tx}}}\mathbf{I}_{M_{\mathrm{nz}}}\right) \tag{25}$$

where $\bar{\boldsymbol{\Sigma}} = \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{\mathrm{H}}$ is the covariance matrix of the channel in the beamspace domain. Furthermore, the conditional distribution is

$$\mathbb{P}(\mathbf{y}|\mathbf{h}) = \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \frac{\sigma^2}{\rho_{\mathrm{tx}}}\mathbf{I}_{M_{\mathrm{nz}}}\right). \tag{26}$$

Therefore, the MI $\mathcal{I}(\mathbf{y};\mathbf{h})$ can be written as

$$\mathcal{I}(\mathbf{y};\mathbf{h}) = \mathcal{H}(\mathbf{y}) - \mathcal{H}(\mathbf{y}|\mathbf{h}) \tag{27}$$

$$= \log\left(\frac{\det\left(\boldsymbol{\Omega}\boldsymbol{\Lambda}\mathbf{S}\bar{\boldsymbol{\Sigma}}\mathbf{S}^{\mathrm{H}}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{\Omega}^{\mathrm{H}} + \frac{\sigma^2}{\rho_{\mathrm{tx}}}\mathbf{I}_{M_{\mathrm{nz}}}\right)}{\det\left(\frac{\sigma^2}{\rho_{\mathrm{tx}}}\mathbf{I}_{M_{\mathrm{nz}}}\right)}\right) \tag{28}$$

$$= \log\det\left(\frac{\rho_{\mathrm{tx}}}{\sigma^2}\boldsymbol{\Omega}\boldsymbol{\Lambda}\mathbf{S}\bar{\boldsymbol{\Sigma}}\mathbf{S}^{\mathrm{H}}\boldsymbol{\Lambda}^{\mathrm{H}}\boldsymbol{\Omega}^{\mathrm{H}} + \mathbf{I}_{M_{\mathrm{nz}}}\right). \tag{29}$$

The beamspace transform matrix $\boldsymbol{\Gamma}$ is designed to resolve distinct and approximately independent paths of the channel.

According to Appendix A, this means that $\bar{\mathbf{\Sigma}}$ is approximately a diagonal matrix. If we approximate $\bar{\mathbf{\Sigma}}$ as a diagonal, the matrix $\mathbf{G} = \frac{\rho_{\mathrm{tx}}}{\sigma^2}\mathbf{\Omega}\mathbf{\Lambda}\mathbf{S}\bar{\mathbf{\Sigma}}\mathbf{S}^{\mathrm{H}}\mathbf{\Lambda}^{\mathrm{H}}\mathbf{\Omega}^{\mathrm{H}} + \mathbf{I}_{M_{\mathrm{nz}}}$ in (29) is also diagonal due to the simple row and column selection and scaling. As follows, we can write an approximation of the MI criterion as

$$\mathcal{I}(\mathbf{y};\mathbf{h}) \approx \sum_{i=1}^{M_{\mathrm{nz}}} \log\left(1 + \frac{\rho_{\mathrm{tx}}}{\sigma^2}\left[\bar{\mathbf{\Sigma}}\right]_{\alpha_i,\alpha_i}\left[\mathbf{\Lambda}\right]_{i,i}^2\right), \tag{30}$$

where the variable $\alpha_i \,\forall\, i \in [M_{\mathrm{rf}}]$ that refers to the beam index in the codebook assigned to the specific chain RF i (i.e., $[\mathbf{S}]_{i,\alpha_i} = 1$). Even when the matrix G is not exactly diagonal, the determinant can still be approximated by the product of its diagonal elements. For example, this follows from the Geršgorin disc theorem [80, Ch. 8], written in terms of log-determinant as:

$$\log\det(\mathbf{G}) \geq \sum_{i=1}^{M}{}_{\mathrm{nz}} \log\left([\mathbf{G}]_{i,i} - \sum_{j=1,j\neq i}^{M}{}_{\mathrm{nz}}|[\mathbf{G}]_{i,j}|\right), \tag{31a}$$

$$\log\det(\mathbf{G}) \leq \sum_{i=1}^{M}{}_{\mathrm{nz}} \log\left([\mathbf{G}]_{i,i} + \sum_{j=1,j\neq i}^{M}{}_{\mathrm{nz}}|[\mathbf{G}]_{i,j}|\right). \tag{31b}$$

The inequalities in (31) hold with equality if and only if $\mathbf{G}$ is a diagonal matrix. However, (31) indicate that the determinant of the matrix $\mathbf{G}$ can be well-approximated by the product of diagonal elements if $[\mathbf{G}]_{i,i} \gg \sum_{j=i+1}^{M}{}_{\mathrm{nz}}|[\mathbf{G}]_{i,j}|$ for all $i \in [M_{\mathrm{nz}}]$.

C. Derivation of Feedback Bits for UE-Centric Scheme

In the considered UE-centric feedback scheme described in Section IV-A, M_{fb} elements are selected to be fed back to the BS from the total of M beamspace elements. The number of distinct combinations is given by the binomial coefficient

$$\binom{M}{M_{\mathrm{fb}}} = \frac{M!}{M_{\mathrm{fb}}!(M - M_{\mathrm{fb}})!}.$$

Therefore, the number of bits needed to indicate which elements are included in the feedback is

$$\log_2\left(\binom{M}{M_{\mathrm{fb}}}\right) = \log_2\left(\frac{M!}{M_{\mathrm{fb}}!(M - M_{\mathrm{fb}})!}\right). \tag{32}$$

Each of the M_{fb} selected elements is quantized using $Q_{\mathrm{uc}}^{\mathrm{A}}$ bits for amplitude and $Q_{\mathrm{uc}}^{\mathrm{P}}$ bits for phase, resulting in a total of $Q_{\mathrm{uc}} = Q_{\mathrm{uc}}^{\mathrm{A}} + Q_{\mathrm{uc}}^{\mathrm{P}}$ bits per element. Therefore, the total number of bits required to quantize all selected elements is $Q_{\mathrm{uc}}M_{\mathrm{fb}}$. Consequently, the total number of feedback bits B_{uc} in (17) is the sum of two terms: the bits needed to indicate the selected beam indices and the bits required to quantize the corresponding complex coefficients.

ACKNOWLEDGMENT

We acknowledge the computational resources provided by the Aalto Science-IT project.

REFERENCES

[1] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive MIMO in sub-6 Ghz and mmWave: Physical, practical, and use-case differences," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 100–108, Apr. 2019.

[2] X. Wang et al., "Millimeter wave communication: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1616–1653, 3rd Quart. 2018.

[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] J. Liu, Z. Luo, and X. Xiong, "Low-resolution ADCs for wireless communication: A comprehensive survey," *IEEE Access*, vol. 7, pp. 91291–91324, 2019.

[5] J. Liu and E. Bentley, "Hybrid-beamforming-based millimeter-wave cellular network optimization," in *Proc. 15th Int. Symp. Model. Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt),* 2017, pp. 1–8.

[6] Y. J. Cho, G.-Y. Suk, B. Kim, D. K. Kim, and C.-B. Chae, "RF lens-embedded antenna array for mmWave MIMO: Design and performance," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 42–48, Jul. 2018.

[7] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[8] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[9] P. Ding, D. J. Love, and M. D. Zoltowski, "Multiple antenna broadcast channels with shape feedback and limited feedback," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3417–3428, Jul. 2007.

[10] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

[11] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[12] J. Kotecha and A. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[13] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated RICIAN MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.

[14] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, Oct. 2014.

[15] 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 J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[16] B. Liao, Z. Zhou, and S. Zhang, "An MAP method for closed-loop channel training in massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 5534–5539, May 2022.

[17] Y. Han, J. Lee, and D. J. Love, "Compressed sensing-aided downlink channel training for FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 2852–2862, Jul. 2017.

[18] V. K. N. Lau, S. Cai, and A. Liu, "Closed-loop compressive CSIT estimation in FDD massive MIMO systems with 1 bit feedback," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2146–2155, Apr. 2016.

[19] A. Liu, F. Zhu, and V. K. N. Lau, "Closed-loop autonomous pilot and compressive CSIT feedback resource adaptation in multi-user FDD massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 1, pp. 173–183, Jan. 2017.

[20] J. Jee and H. Park, "Deep learning-based joint optimization of closed-loop FDD mmWave massive MIMO: Pilot adaptation, CSI feedback, and beamforming," *IEEE Trans. Veh. Technol.*, vol. 73, no. 3, pp. 4019–4034, Mar. 2024.

[21] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[22] Y. Song, T. Yang, M. B. Khalilsarai, and G. Caire, "Deep-learning aided channel training and precoding in FDD massive MIMO with channel statistics knowledge," in *Proc. IEEE Int. Conf. Commun. (ICC),* 2023, pp. 2791–2797.

[23] Y. Song, T. Yang, M. B. Khalilsarai, and G. Caire, "Joint vs. separate source-channel coding in CSI feedback for massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2024, pp. 5129–5134.

[24] X. Fu et al., "A tutorial on downlink precoder selection strategies for 3GPP MIMO codebooks," *IEEE Access*, vol. 11, pp. 138897–138922, 2023.

[25] W. Shen, L. Dai, B. Shim, Z. Wang, and R. W. Heath, "Channel feedback based on AoD-adaptive subspace codebook in FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5235–5248, Nov. 2018.

[26] G. Kwon and H. Park, "Limited feedback hybrid beamforming for multi-mode transmission in wideband millimeter wave channel," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4008–4022, Jun. 2020.

[27] C. K. Au-Yeung, D. J. Love, and S. Sanayei, "Trellis coded line packing: Large dimensional beamforming vector quantization and feedback transmission," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1844–1853, Jun. 2011.

[28] J. Choi, Z. Chance, D. J. Love, and U. Madhow, "Noncoherent trellis coded quantization: A practical limited feedback technique for massive MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5016–5029, Dec. 2013.

[29] J. Choi, D. J. Love, and T. Kim, "Trellis-extended codebooks and successive phase adjustment: A path from LTE-Advanced to FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2007–2016, Apr. 2015.

[30] Z. Xiao, T. He, P. Xia, and X.-G. Xia, "Hierarchical codebook design for beamforming training in millimeter-wave communication," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3380–3392, May 2016.

[31] Z. Xiao, P. Xia, and X.-G. Xia, "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 141–153, Jan. 2017.

[32] J. Zhang, Y. Huang, Q. Shi, J. Wang, and L. Yang, "Codebook design for beam alignment in millimeter wave communication systems," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4980–4995, Nov. 2017.

[33] J. Mo et al., "Beam codebook design for 5G mmWave terminals," *IEEE Access*, vol. 7, pp. 98387–98404, 2019.

[34] K. J. Kim, J. Yue, R. Iltis, and J. Gibson, "A QRD-M/Kalman filter-based detection and channel estimation algorithm for MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 710–721, Mar. 2005.

[35] K. Huber and S. Haykin, "Improved Bayesian MIMO channel tracking for wireless communications: Incorporating a dynamical model," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2458–2466, Sep. 2006.

[36] C. Min, N. Chang, J. Cha, and J. Kang, "MIMO-OFDM downlink channel prediction for IEEE802.16e systems using Kalman filter," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2007, pp. 942–946.

[37] X. Zhang, P. Xiao, D. Ma, and J. Wei, "Variational-Bayes-assisted joint signal detection, noise covariance estimation, and channel tracking in MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 63, no. 9, pp. 4436–4449, Nov. 2014.

[38] S. G. Larew and D. J. Love, "Adaptive beam tracking with the unscented Kalman filter for millimeter wave communication," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1658–1662, Nov. 2019.

[39] M. B. Khalilsarai, Y. Song, T. Yang, and G. Caire, "FDD massive MIMO channel training: Optimal rate-distortion bounds and the spectral efficiency of 'one-shot' schemes," *IEEE Trans. Wireless Commun.*, vol. 22, no. 9, pp. 6018–6032, Sep. 2023.

[40] M. Giordani, M. Mezzavilla, and M. Zorzi, "Initial access in 5G mmWave cellular networks," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 40–47, Nov. 2016.

[41] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 173–196, 1st Quart. 2019.

[42] W. Attaoui, K. Bouraqia, and E. Sabir, "Initial access & beam alignment for mmWave and terahertz communications," *IEEE Access*, vol. 10, pp. 35363–35397, 2022.

[43] 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.

[44] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. 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.

[45] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5689–5701, Sep. 2017.

[46] X. Song, S. Haghighatshoar, and G. Caire, "A scalable and statistically robust beam alignment technique for millimeter-wave systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4792–4805, Jul. 2018.

[47] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, "Online learning for position-aided millimeter wave beam training," *IEEE Access*, vol. 7, pp. 30507–30526, 2019.

[48] R. Wang, P. V. Klaine, O. Onireti, Y. Sun, M. A. Imran, and L. Zhang, "Deep learning enabled beam tracking for non-line of sight millimeter wave communications," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1710–1720, 2021.

[49] H. Luo, U. Demirhan, and A. Alkhateeb, "Millimeter wave V2V beam tracking using radar: Algorithms and real-world demonstration," in *Proc. 31st Eur. Signal Process. Conf. (EUSIPCO)*, 2023, pp. 740–744.

[50] A. Ali, N. González-Prelcic, and R. W. Heath, "Millimeter wave beam-selection using out-of-band spatial information," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1038–1052, Feb. 2018.

[51] M. S. Sim, Y.-G. Lim, S. H. Park, L. Dai, and C.-B. Chae, "Deep learning-based mmWave beam selection for 5G NR/6G with sub-6 GHz channel information: Algorithms and prototype validation," *IEEE Access*, vol. 8, pp. 51634–51646, 2020.

[52] V. Raj, N. Nayak, and S. Kalyani, "Deep reinforcement learning based blind mmWave MIMO beam alignment," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8772–8785, Oct. 2022.

[53] V. Suresh and D. J. Love, "Error control sounding strategies for millimeter wave beam alignment," in *Proc. Inf. Theory Appl. Workshop (ITA)*, 2018, pp. 1–6.

[54] A. J. Duly, T. Kim, D. J. Love, and J. V. Krogmeier, "Closed-loop beam alignment for massive MIMO channel estimation," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1439–1442, Aug. 2014.

[55] W. Wu, N. Cheng, N. Zhang, P. Yang, W. Zhuang, and X. Shen, "Fast mmWave beam alignment via correlated bandit learning," *IEEE Trans. Wireless Commun.*, vol. 18, no. 12, pp. 5894–5908, Dec. 2019.

[56] M. B. Booth, V. Suresh, N. Michelusi, and D. J. Love, "Multi-armed bandit beam alignment and tracking for mobile millimeter wave communications," *IEEE Commun. Lett.*, vol. 23, no. 7, pp. 1244–1248, Jul. 2019.

[57] I. Chafaa, E. V. Belmega, and M. Debbah, "One-bit feedback exponential learning for beam alignment in mobile mmWave," *IEEE Access*, vol. 8, pp. 194575–194589, 2020.

[58] C. Liu, L. Zhao, M. Li, and L. Yang, "Adaptive beam search for initial beam alignment in millimetre-wave communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6801–6806, Jun. 2022.

[59] Y. Wei, Z. Zhong, and V. Y. F. Tan, "Fast beam alignment via pure exploration in multi-armed bandits," *IEEE Trans. Wireless Commun.*, vol. 22, no. 5, pp. 3264–3279, May 2023.

[60] J. Seo, Y. Sung, G. Lee, and D. Kim, "Training beam sequence design for millimeter-wave MIMO systems: A POMDP framework," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1228–1242, Mar. 2016.

[61] M. Hussain and N. Michelusi, "Learning and adaptation for millimeter-wave beam tracking and training: A dual timescale variational framework," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 37–53, Jan. 2022.

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

[63] P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2223, Jun. 2015.

[64] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[65] Y. Zhao, M. Zhao, L. Xiao, and J. Wang, "Capacity of time-varying rayleigh fading MIMO channels," in *Proc. IEEE 16th Int. Symp. Pers., Indoor Mobile Radio Commun.,* 2005, vol. 1, pp. 547–551.

[66] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[67] P. He, L. Zhao, S. Zhou, and Z. Niu, "Water-filling: A geometric approach and its application to solve generalized radio resource allocation problems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3637–3647, Jul. 2013.

[68] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653–656, Dec. 2014.

[69] D. Ying, F. W. Vook, T. A. Thomas, and D. J. Love, "Hybrid structure in massive MIMO: Achieving large sum rate with fewer RF chains," in *Proc. IEEE Int. Conf. Commun. (ICC),* 2015, pp. 2344–2349.
[70] L. Zhao, D. W. K. Ng, and J. Yuan, "Multi-user precoding and channel estimation for hybrid millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1576–1590, Jul. 2017.
[71] S. Cornelis, N. Noels, and M. Moeneclaey, "Rate-maximizing zero-forcing hybrid precoder for MU-MISO-OFDM," *IEEE Access*, vol. 11, pp. 3275–3290, 2023.
[72] S. Bazzi and W. Xu, "Downlink training sequence design for FDD multiuser massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4732–4744, Sep. 2017.
[73] A. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, Feb. 1987.
[74] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.
[75] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
[76] P. Pulkkinen, M. Esfandiari, and V. Koivunen, "Cognitive beamspace algorithm for integrated sensing and communications," in *Proc. IEEE Radar Conf. (RadarConf)*, 2024, pp. 1–6.
[77] K. Mishra, M. Bhavani Shankar, V. Koivunen, B. Ottersten, and S. A. Vorobyov, "Toward millimeter-wave joint radar communications: A signal processing perspective," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 100–114, Sep. 2019.
[78] F. Liu et al., "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 6, pp. 1728–1767, Jun. 2022.

Petteri Pulkkinen (Member, IEEE) received the B.Sc. degree in electronics and electrical engineering and the M.Sc. degree in signal processing from Aalto University, Espoo, Finland, in 2019 and 2020, respectively. Since 2020, he has been a Research Engineer with Saab Finland Oy and is working toward the Ph.D. degree in signal processing and data analysis with Aalto University. From 2023 to 2024, he was a Visiting Scholar with Purdue University, West Lafayette, IN, USA. His research interests include cognitive radars, integrated sensing and communication systems, statistical signal processing, convex optimization, and machine learning.

David J. Love (Fellow, IEEE) received the B.S. (with highest honors), M.S.E., and Ph.D. degrees in electrical engineering from the University of Texas, Austin, in 2000, 2002, and 2004, respectively. Since 2004, he has been with the Elmore Family School of Electrical and Computer Engineering with Purdue University, where he is now the Nick Trbovich Professor of Electrical and Computer Engineering. He serves as a Senior Editor for IEEE Journal on Selected Areas in Communications (JSAC) and previously served as a Senior Editor for IEEE Signal Processing Magazine, Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, and guest editor for special issues of the JSAC and the EURASIP Journal on Wireless Communications and Networking. He was a member of the Executive Committee for the National Spectrum Consortium. He holds 32 issued U.S. patents. His research interests are in the design and analysis of broadband wireless communication systems, beyond-5G wireless systems, multiple-input multiple-output (MIMO) communications, millimeter wave wireless, software defined radios and wireless networks, coding theory, and MIMO array processing. He is a Fellow of the American Association for the Advancement of Science (AAAS) and the National Academy of Inventors (NAI). He was named a Thomson Reuters Highly Cited Researcher (2014 and 2015). Along with his co-authors, he won best paper awards from IEEE Communications Society (2016 Stephen O. Rice Prize, 2020 Fred W. Ellersick Prize, and 2024 William R. Bennett Prize), the IEEE Signal Processing Society (2015 IEEE Signal Processing Society Best Paper Award), and the IEEE Vehicular Technology Society (2010 Jack Neubauer Memorial Award).

Visa Koivunen (Fellow, IEEE) received the D.Sc. (EE) degree with honors from the Department of Electrical Engineering, University of Oulu. He is a member of Eta Kappa Nu. He was a Visiting Researcher with the University of Pennsylvania, Philadelphia, USA, 1991–1995. Since 1999 he has been a Full Professor of Signal Processing with Aalto University, Finland. He received the Academy Professor position in 2010 and Aalto Distinguished Professor in 2020. Years 2003–2006 he was also Adjunct Full Professor with the University of Pennsylvania, USA. During his sabbatical terms in 2006–2007 and 2013–2014 he was a Visiting Faculty with Princeton University. He has also been a Visiting Fellow with Nokia Research (2006–2012). Since 2010 he has spent mini-sabbaticals with Princeton University each year. On his sabbatical year in 2022–2023, he was a Visiting Professor with EPFL, Lausanne, Switzerland. His research interests include statistical signal processing, wireless comms, radar, multisensor systems, data science and machine learning. He has published about 500 papers in international scientific conferences and journals and holds five patents. He has co-authored multiple papers receiving the best paper award in IEEE and other conferences. He was awarded the IEEE SP Society best paper award for the year 2007 (with J. Eriksson) and 2017 (w Zoubir, Muma and Chakhchouk). He has served in editorial board for The Proceedings of the IEEE, IEEE SP Letters, IEEE TR on SP, and IEEE SP Magazine. He was awarded the 2015 EURASIP (European Association for Signal Processing) Technical Achievement Award for fundamental contributions to statistical signal processing and its applications in wireless communications, radar and related fields. He has served in the IEEE Fourier Award, Kilby Medal and Fellow Evaluation committees, SPS Award Board and as IEEE SP Society Distinguished Lecturer in 2015–2016 and Board of Governors for Asilomar conferences. He is also EURASIP Fellow, AAIA Fellow and a member of the Academia Europaea.