# Layered Downlink Precoding for C-RAN Systems With Full Dimensional MIMO

Jinkyu Kang, Osvaldo Simeone, *Fellow, IEEE*, Joonhyuk Kang, *Member, IEEE*, and Shlomo Shamai, *Fellow, IEEE*

***Abstract*—The implementation of a cloud radio access network (C-RAN) with full dimensional (FD) multiple-input multiple-output (MIMO) is faced with the challenge of controlling the fronthaul overhead for the transmission of baseband signals as the number of horizontal and vertical antennas grows larger. This paper proposes to leverage the special low-rank structure of the FD-MIMO channel, which is characterized by a time-invariant elevation component and a time-varying azimuth component, by means of a layered precoding approach, to reduce the fronthaul overhead. According to this scheme, separate precoding matrices are applied for the azimuth and elevation channel components, with different rates of adaptation to the channel variations and correspondingly different impacts on the fronthaul capacity. Moreover, we consider two different central unit (CU)–radio unit (RU) functional splits at the physical layer, namely, the conventional C-RAN implementation and an alternative one in which coding and precoding are performed at the RUs. Via numerical results, it is shown that the layered schemes significantly outperform conventional nonlayered schemes, particularly in the regime of low fronthaul capacity and a large number of vertical antennas.**

***Index Terms*—Cloud radio access networks (C-RAN), fronthaul compression, full-dimensional multiple-input multiple-output (FD-MIMO), layered precoding.**

## I. Introduction

CLOUD radio access network (C-RAN) architecture consists of multiple radio units (RUs) connected via fronthaul links to a central unit (CU) that implements the protocol stack of the RUs, including baseband processing [1], [2]. C-RAN enables a significant reduction in capital and operating expenses, as well as an enhanced spectral efficiency by means of joint interference management at the physical layer across all connected RUs. Nevertheless, it is well recognized that the performance of this architecture is limited by the capacity and latency constraints of the fronthaul network connecting the RUs and the CU [1]–[4].

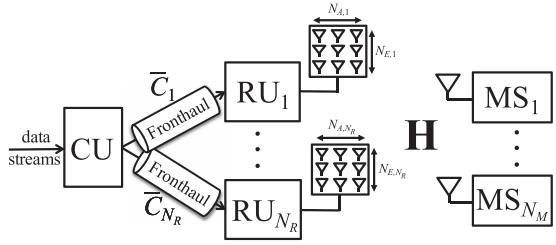


Fig. 1. Downlink of a C-RAN system with FD-MIMO.

In a standard C-RAN implementation, the fronthaul links carry digitized baseband signals. Hence, the bit rate required for a fronthaul link is determined by the quantization and compression operations applied to the baseband signals prior to transmission on the fronthaul links. As such, the fronthaul rate is proportional to the signal bandwidth, to the oversampling factor, to the resolution of the quantizer/compressor, and to the number of antennas [5]. The fronthaul bit rate can be reduced by implementing alternative functional splits between CU and RU, whereby some baseband functionalities are implemented at the RU [6]–[8].

As a concurrent trend in the evolution of wireless networks, in the Third-Generation Partnership Project Long-Term Evolution (LTE) Release-13, 3-D MIMO, where base stations are equipped with 2-D rectangular antenna arrays, has been intensely discussed as a promising tool to boost spectral efficiency [9]–[11]. Three-dimensional MIMO technology is classified into three categories, namely vertical sectorization (VS), elevation beamforming (EB), and full-dimensional MIMO (FD-MIMO), in order of complexity. The VS scheme splits a sector of cellular coverage into multiple sectors by means of different electrical downtilt angles. With the EB approach, instead, users are supported by predetermined or adaptive beams in the elevation direction. Finally, in FD-MIMO, the spatial diversity provided by vertical and horizontal antennas is leveraged jointly to serve multiple users using multiuser-MIMO techniques.

Endowing RUs with 2-D arrays in a C-RAN system (see Fig. 1), while promising from a spectral efficiency perspective, creates significant challenges in terms of fronthaul overhead

Manuscript received November 26, 2015; revised March 30, 2016; accepted May 4, 2016. Date of publication May 24, 2016; date of current version March 10, 2017. This work was supported in part by the Institution for Information and Communications Technology Promotion of the Ministry of Science, ICT, and Future Planning of South Korea through the ICT R&D Program under Grant B0101-16-1372 (Development of Mobile Multimode Transmission Technology based on Spatial Spreading). The work of O. Simeone was supported in part by the U.S. National Science Foundation under Grant 1525629. The work of S. Shamai was supported in part by the Israel Science Foundation and in part by the European Research Council under Advanced Grant 694630. The review of this paper was coordinated by Prof. R. Dinis.

J. Kang was with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, South Korea. He is now with the John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: jkkang@g.harvard.edu).

O. Simeone is with the Center for Wireless Communications and Signal Processing Research, Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: osvaldo.simeone@njit.edu).

J. Kang is with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, South Korea (e-mail: jhkang@ee.kaist.ac.kr).

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: sshlomo@ee.technion.ac.il).

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

Digital Object Identifier 10.1109/TVT.2016.2572199

as the number of antennas grows larger [12]. In this paper, we focus on the design of downlink precoding for C-RANs with FD-MIMO RUs by accounting for the impact of fronthaul capacity limitations. Previous works [4], [13]–[16] on precoding design for the downlink of C-RAN systems either assume fixed channel matrices with full channel state information (see [4] and [13]–[15]) or consider ergodic channels with generic correlation structure and possibly imperfect CSI [16]. Importantly, these works do not account for the special features of FD channel models [17]–[19] and hence do not bring insights into the feasibility of a C-RAN deployment based on FD-MIMO.

The starting point of this paper is the recent characterization of the FD-MIMO channel, which has been demonstrated to be well described by a Kronecker structure for the elevation and azimuth components of the transmit-side correlation matrix, and reported to exhibit time variability at different timescales for elevation and azimuth components [17]. More specifically, the elevation component was seen to change significantly more slowly than the rate of change of the more conventional azimuth component. This can be ascribed to less-significant users' mobility in the elevation coordinate. The results in [17] are obtained by means of comparison of the ergodic capacity of a point-to-point channel under the Kronecker model with that under a standard geometry-based ray-tracing channel model (see also [20] for standard geometry-based models). In order to enhance the performance of C-RAN system with FD-MIMO, this paper puts forth the following contributions.

- A *layered precoding* scheme is proposed, whereby separate precoding matrices are applied for the azimuth and elevation channel components with a different rate of adaptation to the channel variations. Specifically, a single precoding matrix is designed for the elevation channel across all coherence times based on stochastic CSI, whereas precoding matrices are optimized for the azimuth channel by adapting instantaneous CSI. This layered approach, considered in [18] and [19] for a conventional cellular architecture and included in LTE Release-13 [21], has the *unique* advantage in a C-RAN of potentially reducing the fronthaul transmission rate, due to the opportunity to amortize the overhead related to the elevation channel component across multiple coherence times. Due to this property, which stems from the layered structure, the layered coding strategies proposed here for the first time in the context of C-RAN have the potential of strongly improving over the unconstrained precoding techniques typically used in C-RAN [16].
- We study layered precoding in a C-RAN system by considering two different CU-RU functional splits at the physical layer, namely the conventional C-RAN implementation, which is referred to as compress-after-precoding (CAP) as in [4] and [13]–[16], whereby all baseband processing is done at the CU, and an alternative split, known as compress-before-precoding (CBP) [16], [22], in which channel encoding and precoding are instead performed at the RUs.
- We develop a novel optimization strategy based on stochastic successive upper bound minimization (SSUM) and difference-of-convex (DC) methods; this strategy is tailored to the optimization of layered precoding and specifically to the design of long- and short-term parameters of the precoding matrices, namely, the elevation components and the azimuth components.
- We carry out a performance comparison between standard nonlayered precoding strategies and layered precoding for C-RAN systems with FD-MIMO under different functional splits as a function of system parameters such as the fronthaul capacity and the duration of the coherence period.

The remainder of this paper is organized as follows. We describe the system model in Section II. In Section III, we review the conventional nonlayered precoding schemes corresponding to the mentioned functional splits, namely CAP and CBP [16]. Then, we propose and optimize the layered precoding strategy for fronthaul compression in Section IV. In Section V, numerical results are presented. Concluding remarks are summarized in Section VI.

*Notation:* $E[\cdot]$ and $\mathrm{tr}(\cdot)$ denote the expectation and trace of the argument matrix, respectively. We use the standard notation for mutual information [23]. $\boldsymbol{\nu}_{\max}(\mathbf{A})$ is the eigenvector corresponding to the largest eigenvalue of the semi-positive definite matrix $\mathbf{A}$. We reserve the superscript $\mathbf{A}^T$ for the transpose of $\mathbf{A}$, $\mathbf{A}^\dagger$ for the conjugate transpose of $\mathbf{A}$, and $\mathbf{A}^{-1} = (\mathbf{A}^\dagger\mathbf{A})^{-1}\mathbf{A}^\dagger$, which reduces to the usual inverse if the number of columns and rows are same. $[\mathbf{A}]_{j,i}$ denotes the $(j,i)$ entry of the matrix $\mathbf{A}$. The identity matrix is denoted $\mathbf{I}$. $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$.

## II. System Model

We consider the downlink of a C-RAN in which a cluster of $N_R$ RUs provides wireless service to $N_M$ mobile stations (MSs), as shown in Fig. 1. Each RU $i$ has a FD (or 2-D) antenna array of $N_{A,i}$ horizontal antennas by $N_{E,i}$ vertical antennas, and each MS has a single antenna. RU $i$ is connected to the CU via fronthaul link of capacity $\bar{C}_i$ bit per downlink symbol, where the downlink symbol rate equals the baud rate, i.e., no oversampling is performed.

### A. Signal Model

Each coded transmission block spans multiple coherence periods, e.g., multiple distinct resource blocks in an LTE system, of the downlink channel that contain $T$ symbols each. The $T \times 1$ signal $\mathbf{y}_j$ received by the MS $j$ in a given coherence interval is given as follows:

$$\mathbf{y}_j = \mathbf{X}^T\mathbf{h}_j + \mathbf{z}_j \tag{1}$$

where $\mathbf{z}_j$ is the $T \times 1$ noise vector with independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ components; $\mathbf{h}_j = [\mathbf{h}_{j1}^T, \ldots, \mathbf{h}_{jN_R}^T]^T$ denotes the $\sum_{i=1}^{N_R} N_{A,i}N_{E,i} \times 1$ channel vector for MS $j$, with $\mathbf{h}_{ji}$ being the $N_{A,i}N_{E,i} \times 1$ channel vector from the $i$th RU to the MS $j$, as further discussed in the following; and $\mathbf{X}$ is an $\sum_{i=1}^{N_R} N_{A,i}N_{E,i} \times T$ matrix that stacks the signals

transmitted by all the RUs, i.e., $\mathbf{X} = [\mathbf{X}_1^T, \ldots, \mathbf{X}_{N_R}^T]^T$, with $\mathbf{X}_i$ being a $N_{A,i}N_{E,i} \times T$ complex baseband signal matrix transmitted by the $i$th RU with each channel coherence period of duration $T$ channel uses. Note that each column of the signal matrix $\mathbf{X}_i$ corresponds to the signal transmitted from the $N_{A,i}N_{E,i}$ antennas in a channel use and that all transmitted signals consist of $T$ channel uses. The transmit signal $\mathbf{X}_i$ has a power constraint given as $E[|\mathbf{X}_i|^2] = T\bar{P}_i$.

The channel vector $\mathbf{h}_j$ is assumed constant during each channel coherence block and to change according to a stationary ergodic process from block to block. We assume that the CU has perfect instantaneous information about the channel matrix $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_{N_M}]$ and MSs have full CSI about their respective channel matrices.

### B. FD Channel Model

As in, for example, [17] and [18], we assume that each RU is equipped with a uniform rectangular array (URA). Furthermore, the channel vector $\mathbf{h}_{ji}$ from RU $i$ to MS $j$ is modeled by means of a Kronecker product spatial correlation model [17], [18]. This was shown to provide a good modeling choice under the condition that the MS is sufficiently far away from the RUs [17], According to this model, the covariance of the 3-D channel $\mathbf{h}_{ji}$, which is defined as $\mathbf{R}_{ji} = E[\mathbf{h}_{ji}\mathbf{h}_{ji}^\dagger]$, is written as

$$\mathbf{R}_{ji} = \mathbf{R}_{ji}^A \otimes \mathbf{R}_{ji}^E \tag{2}$$

where $\mathbf{R}_{ji}^A$ and $\mathbf{R}_{ji}^E$ represent the covariance matrices in the azimuth and elevation directions, respectively. Since the elevation direction is typically subject to negligible scattering [24], [25], the elevation covariance matrix $\mathbf{R}_{ji}^E$ may be assumed to be a rank-1 matrix, i.e., $\mathbf{R}_{ji}^E = \mathbf{u}_{ji}^E \mathbf{u}_{ji}^{E\,\dagger}$, where $\mathbf{u}_{ji}^E$ is a $N_{E,i} \times 1$ unit-norm vector [18]. Under this assumption, the channel vector $\mathbf{h}_{ji}$ can be written as

$$\mathbf{h}_{ji} = \sqrt{\alpha_{ji}}\mathbf{h}_{ji}^A \otimes \mathbf{u}_{ji}^E \tag{3}$$

where $\alpha_{ji}$ denotes the path-loss coefficient between MS $j$ and RU $i$ as

$$\alpha_{ji} = \frac{1}{1 + \left(\frac{d_{ji}}{d_0}\right)^\eta} \tag{4}$$

with $d_{ji}$ being the distance between the $j$th MS and the $i$th RU, $d_0$ being a reference distance, and $\eta$ being the path-loss exponent; and $\mathbf{h}_{ji}^A \sim \mathcal{CN}(0, \mathbf{R}_{ji}^A)$, with $\mathbf{R}_{ji}^A$ having diagonal elements equal to one. This model entails that the elevation components $\mathbf{h}_{ji}^E$ remains constant over coherence interval, whereas the azimuth component changes independent across coherence interval as $\mathbf{h}_{ji}^A \sim \mathcal{CN}(0, \mathbf{R}_{ji}^A)$, as shown in Fig. 2.

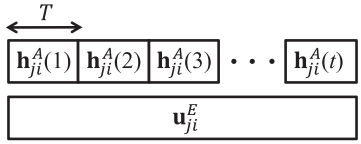


Fig. 2. Time variability of the azimuth component $\{\mathbf{h}_{ji}^A(t)\}$ and of the elevation component $\mathbf{u}_{ji}^E$ in the FD channel model (3). The notation $\mathbf{h}_{ji}^A(t)$ emphasizes the dependence on the coherence block $t$ of the azimuth component of the channel.

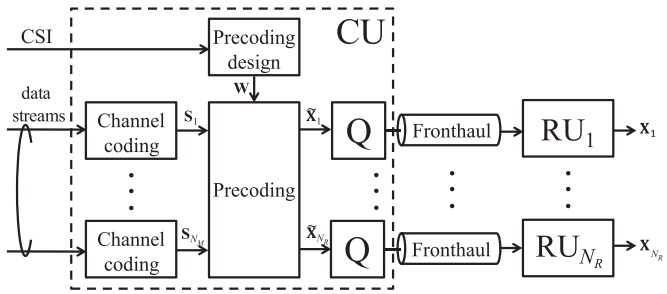


Fig. 3. Block diagram of the (nonlayered) CAP scheme ("Q" represents fronthaul compression).

## III. Background

Here, we briefly recall in an informal fashion two baseline strategies for downlink transmission in the C-RAN system introduced earlier. The strategies correspond to two different functional splits at the physical layer between the CU and RUs [5], [6], as detailed in [16]. We note that these schemes were previously proposed and studied without specific reference to FD-MIMO and, hence, do not leverage the special structure of the channel model (3).

### A. Standard C-RAN Processing: Precoding at the CU

In the standard C-RAN approach, all baseband processing is done at the CU. Specifically, as shown in Fig. 3, the CU performs channel coding and precoding, and then compresses the resulting baseband signals so that they can be forwarded on the fronthaul links to the corresponding RUs. The RUs upconvert the received quantized baseband signal prior to transmission on the wireless channel. Following [16], we refer to this strategy as CAP. Analysis and optimization of the CAP strategy can be found in [16]. Note that the optimization of precoding and quantization noise is carried out in [15] based on the DC approach [26].

### B. Alternative Functional Split: Precoding at the RUs

As an alternative to the standard C-RAN approach just described, one can instead implement channel encoding and precoding at the RUs. This is referred to as CBP in [16] and [22], According to this solution, as shown in Fig. 4, the CU calculates the precoding matrices based on the available CSI but does not perform precoding. Instead, it uses the fronthaul links to communicate the downlink information streams to each RU, along with the compressed precoding matrix. Each RU can then encode and precode the messages of the MSs based on the information received from the fronthaul link. We emphasize that this approach does not preclude cooperative interference management techniques such as interference alignment [27] or interference cancellation [28], but it generally makes it more difficult to implement cooperative transmission strategies such

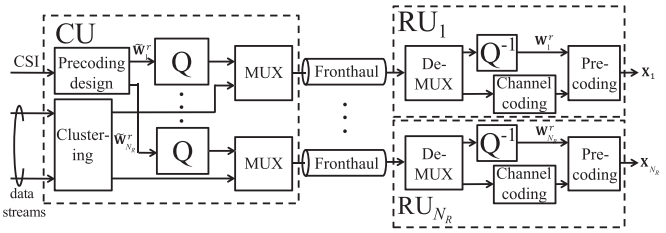


Fig. 4. Block diagram of the (nonlayered) CBP scheme ("Q" represents fronthaul compression).

as cooperative beamforming due to the need to convey the MSs' data streams to all the participating RUs. As a result, as elaborated on in [16], this alternative functional split is generally advantageous when the number of MSs is not too large and when the coherence period $T$ is large enough. This is because, when the number of MSs is small, a lower fronthaul overhead is needed to communicate the smaller number of data streams of the MSs on the fronthaul link. Furthermore, when the coherence period is large, the compressed precoding information can be amortized over a longer period $T$, hence reducing the fronthaul rate. As for CAP, the optimization of precoding and quantization noise is performed in [15] based on the DC approach [26].

## IV. Layered Precoding for Reduced Fronthaul Overhead

The baseline state-of-the-art fronthaul transmission strategies mentioned earlier do not make any provision to exploit the special structure of the FD channel model (3) and can hence be inefficient if the number of vertical antennas is large. Here, we propose a layered precoding that instead leverages the different dynamic characteristic of the elevation and azimuth channels as per channel model (3). We recall that, according to this model, the elevation channel has a constant direction across the coherence periods in its elevation component due to the rank-1 covariance matrix, whereas its azimuth component changes in each coherence period due to the generally larger rank of its covariance matrix (see Fig. 2).

To exploit this channel decomposition, we propose that the CU designs separate precoding matrices for the elevation and azimuth channels following a layered precoding approach. The key idea is that of designing a single precoding matrix for the elevation channel across all coherence times based on long-term CSI, while adapting only the azimuth precoding matrix to the instantaneous channel conditions. This allows the CU to accurately describe the elevation precoding matrix through the fronthaul links via quantization with negligible overhead given that the latter is amortized across all coherence periods. Precoding on the azimuth channel can instead be handled via either a CAP or CBP-like scheme, as detailed in the following.

In the following, we first describe the layered precoding approach in Section IV-A; then, we introduce the precoding and fronthaul compression strategy based on CAP in Section IV-B; finally, we introduce CBP-based fronthaul compression and layered precoding design in Section IV-C.

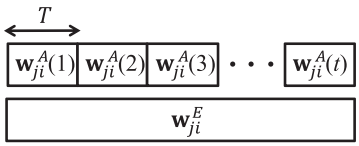


Fig. 5. Time variability of the azimuth and elevation components of beamforming in the layered precoding scheme (5).

### A. Layered Precoding

Leveraging the channel decomposition resulting from the Kronecker channel model (3), we propose to factorize the $N_{A,i}N_{E,i} \times 1$ precoding vector $\mathbf{w}_{ji}$ for RU $i$ toward MS $j$ as

$$\mathbf{w}_{ji} = \mathbf{w}_{ji}^A \otimes \mathbf{w}_{ji}^E \tag{5}$$

where $\mathbf{w}_{ji}^A$ denotes the $N_{A,i} \times 1$ azimuth component, and $\mathbf{w}_{ji}^E$ is the $N_{E,i} \times 1$ elevation component of the precoding vector for MS $j$ and RU $i$ designed based on the elevation channels. A similar model was proposed in [18] for colocated antenna arrays. The corresponding $N_{A,i} \times N_M$ azimuth precoding matrix $\mathbf{W}_i^A$ and the $N_{E,i} \times N_M$ elevation precoding matrix $\mathbf{W}_i^E$ for RU $i$ are defined as $\mathbf{W}_i^A = [\mathbf{w}_{1i}^A, \ldots, \mathbf{w}_{N_M i}^A]$ and $\mathbf{W}_i^E = [\mathbf{w}_{1i}^E, \ldots, \mathbf{w}_{N_M i}^E]$, respectively. In the proposed solutions, each elevation component $\mathbf{w}_{ji}^E$ is quantized by the CU and sent to the $j$th RU via the corresponding fronthaul links. Since this vector is to be used for all coherence times, as shown in Fig. 5, its fronthaul overhead can be amortized across multiple coherence interval. As a result, it can be assumed known accurately at the RUs. Moreover, the corresponding fronthaul overhead for the transfer of elevation precoding information on the fronthaul links can be assumed negligible. For the azimuth components, we may adopt either a CAP or CBP approach, as discussed in the following.

### B. CAP-Based Fronthaul Compression for Layered Precoding

In the proposed CAP-based solution, the CU applies precoding only for the azimuth component. Accordingly, the azimuth-precoded baseband signals, as well as the precoding matrix for the elevation component, are separately compressed at the CU and forwarded over the fronthaul links to each RU. In order to perform precoding over both elevation and azimuth channels, each RU finally performs the Kronecker product of the compressed baseband signal $\mathbf{X}_{ji}^A$ and the precoding vector $\mathbf{w}_{ji}^E$ for elevation channel. A block diagram can be found in Fig. 6 and details are provided in the following.

*1) Details and Analysis:* Let $\widetilde{\mathbf{X}}_{ji}^A$ be the $N_{A,i} \times T$ precoded signal only for the azimuth channel between RU $i$ and MS $j$ in a given coherence period. This is defined as $\widetilde{\mathbf{X}}_{ji}^A = \mathbf{w}_{ji}^A \mathbf{s}_j^T$, where $\mathbf{s}_j$ is the $T \times 1$ vector containing the encoded data stream for MS $j$ in the given coherence period. Note that all the entries of vector $\mathbf{s}_j$ are assumed to have i.i.d. $\mathcal{CN}(0,1)$ from standard random coding arguments. Adopting a CAP-like approach, the CU quantizes each sequence of baseband signals $\{\widetilde{\mathbf{X}}_{ji}^A\}$, for all $j \in \mathcal{N}_M$, across all coherence periods intended for RU $i$ for transfer on $i$th fronthaul. The compressed signal $\mathbf{X}_{ji}^A$ is modeled as

$$\mathbf{X}_{ji}^A = \widetilde{\mathbf{X}}_{ji}^A + \mathbf{Q}_{x,ji}^A = \mathbf{w}_{ji}^A \mathbf{s}_j^T + \mathbf{Q}_{x,ji}^A \tag{6}$$

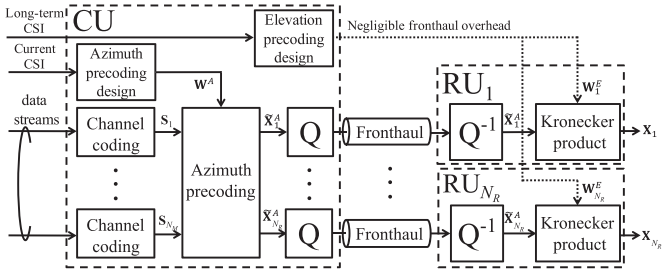


Fig. 6. Block diagram of the layered CAP scheme ("Q" represents fronthaul compression).

where $\mathbf{Q}^A_{x,ji}$ is the quantization noise matrix. From standard rate–distortion arguments [23], [29], the required rate for transfer of the precoded data signals $\{\widetilde{\mathbf{X}}^A_{ji}\}_{j\in\mathcal{N}_M}$ on fronthaul link between the CU and RU $i$ is given as $\sum_{j=1}^{N_M} I(\mathbf{X}^A_{ji};\widetilde{\mathbf{X}}^A_{ji})$ [23, Ch. 3]. Choosing the quantization noise matrix to have i.i.d. $\mathcal{CN}(0,\sigma^2_{x,ji})$ entries, we obtain

$$\begin{aligned} C_{x,i}\left(\mathbf{W}^A_i,\boldsymbol{\sigma}^2_{x,i}\right) &= \sum_{j=1}^{N_M} I\left(\mathbf{X}^A_{ji};\widetilde{\mathbf{X}}^A_{ji}\right) \\ &= \sum_{j=1}^{N_M}\left\{\log\left(\left\|\mathbf{w}^A_{ji}\right\|^2+\sigma^2_{x,ji}\right)-\log\sigma^2_{x,ji}\right\} \end{aligned} \tag{7}$$

where we have defined $\boldsymbol{\sigma}^2_{x,i}=[\sigma^2_{x,1i},\ldots,\sigma^2_{x,N_Mi}]^T$ and we have used the assumption that the data signals $\mathbf{X}^A_{ji}$ are independent across the MS index $j$ due to the independence of the data streams for different users. We note that the assumption of Gaussian quantization noise can be justified by fact that large-block lattice quantization codes, such as trellis-coded quantization, are able to approximate a Gaussian quantization noise distribution and to achieve the rate–distortion tradeoff implied by (7) [30]. Note that, unlike the standard CAP scheme, here the signals for different MSs are separately compressed as per (6).

Considering also the elevation component, the resulting signal $\mathbf{X}_i$ computed and transmitted by RU $i$ is obtained as $\mathbf{X}_i=\sum_{j=1}^{N_M}\mathbf{X}_{ji}$, with

$$\begin{aligned} \mathbf{X}_{ji}=\mathbf{X}^A_{ji}\otimes\mathbf{w}^E_{ji} &= \left(\mathbf{w}^A_{ji}\mathbf{s}^T_j+\mathbf{Q}^A_{x,ji}\right)\otimes\mathbf{w}^E_{ji} \\ &= \left(\mathbf{w}^A_{ji}\otimes\mathbf{w}^E_{ji}\right)\mathbf{s}^T_j+\mathbf{Q}^A_{x,ji}\otimes\mathbf{w}^E_{ji}. \end{aligned} \tag{8}$$

The power transmitted at RU $i$ is then computed as

$$\begin{aligned} &P_i\left(\mathbf{W}^A_i,\mathbf{W}^E_i,\boldsymbol{\sigma}^2_{x,i}\right) \\ &\quad=\mathrm{tr}\left(\mathbf{X}_i\mathbf{X}^\dagger_i\right) \\ &\quad=\mathrm{tr}\left(\sum_{j=1}^{N_M}\left(\left(\mathbf{w}^A_{ji}\mathbf{s}^T_j+\mathbf{Q}^A_{x,ji}\right)\otimes\mathbf{w}^E_{ji}\right)\right. \\ &\qquad\left.\times\left(\left(\mathbf{w}^A_{ji}\mathbf{s}^T_j+\mathbf{Q}^A_{x,ji}\right)\otimes\mathbf{w}^E_{ji}\right)^\dagger\right) \\ &\quad=\sum_{j=1}^{N_M}\left(\left\|\mathbf{w}^A_{ji}\right\|^2\left\|\mathbf{w}^E_{ji}\right\|^2+N_{A,i}\sigma^2_{x,ji}\left\|\mathbf{w}^E_{ji}\right\|^2\right) \end{aligned} \tag{9}$$

where we have used the property of the Kronecker product that $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D})=(\mathbf{AC}\otimes\mathbf{BD})$ and $\mathrm{tr}(\mathbf{A}\otimes\mathbf{B})=\mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})$ [31].

The ergodic achievable rate for MS $j$ is evaluated as $E[R_j(\mathbf{H},\mathbf{W}^A,\mathbf{W}^E,\boldsymbol{\sigma}^2_x)]$, with $R_j(\mathbf{H},\mathbf{W}^A,\mathbf{W}^E,\boldsymbol{\sigma}^2_x)=I_{\mathbf{H}}(\mathbf{s}_j;\mathbf{y}_j)/T$, where $I_{\mathbf{H}}(\mathbf{s}_j;\mathbf{y}_j)$ is the mutual information conditioned on the value of channel matrix $\mathbf{H}$, the expectation is taken with respect to $\mathbf{H}$, and

$$\begin{aligned} &R_j\left(\mathbf{h},\mathbf{w}^A,\mathbf{w}^E,\boldsymbol{\sigma}^2_x\right) \\ &=\log\left(1+\sum_{k=1}^{N_M}\sum_{i=1}^{N_R}\lambda^E_{ji}\left|\mathbf{u}^E_{ji}\mathbf{w}^E_{ki}\right|^2\left(\left|\mathbf{w}^{A\dagger}_{ki}\mathbf{h}^A_{ji}\right|^2+\sigma^2_{x,ki}\left\|\mathbf{h}^A_{ji}\right\|^2\right)\right) \\ &\quad-\log\left(1+\sum_{k=1,k\neq j}^{N_M}\sum_{i=1}^{N_R}\lambda^E_{ji}\left|\mathbf{u}^E_{ji}\mathbf{w}^E_{ki}\right|^2\right. \\ &\qquad\left.\times\left(\left|\mathbf{w}^{A\dagger}_{ki}\mathbf{h}^A_{ji}\right|^2+\sigma^2_{x,ki}\left\|\mathbf{h}^A_{ji}\right\|^2\right)\right) \end{aligned} \tag{10}$$

where $\mathbf{W}^A=[(\mathbf{W}^A_1)^T,\ldots,(\mathbf{W}^A_{N_R})^T]^T$, $\mathbf{W}^E=[(\mathbf{W}^E_1)^T,\ldots,(\mathbf{W}^E_{N_R})^T]^T$, and $\boldsymbol{\sigma}^2_x=[\boldsymbol{\sigma}^2_{x,1},\ldots,\boldsymbol{\sigma}^2_{x,N_R}]$.

**Algorithm 1** CAP-Based Fronthaul Compression and Layered Precoding Design

**1) Long-term Optimization of Elevation Precoding**
**Input:** Long-term statistics of the channel
**Output:** Elevation precoding $\mathbf{W}^{E*}$
**Initialization (outer loop)**: Initialize the covariance matrix $\mathbf{V}^{E\,(n)}\succeq 0$ subject to $\mathrm{tr}(\mathbf{V}^{E\,(n)})=1$ and set $n=0$.
**Repeat**
$n\leftarrow n+1$
Generate a channel matrix realization $\mathbf{H}^{(n)}$ using the available stochastic CSI.
**Inner loop**: Obtain $\mathbf{V}^{A(n)}(\mathbf{H}^{(n)})$ and $\boldsymbol{\sigma}^{2(n)}_x(\mathbf{H}^{(n)})$ with $\mathbf{V}^E\leftarrow\mathbf{V}^{E(n-1)}$ using Algorithm 2.
Update $\mathbf{V}^{E\,(n)}$ by solving problem (25), which de-pends on $\mathbf{V}^{A(m)}(\mathbf{H}^{(m)})$ and $\boldsymbol{\sigma}^{2(m)}_x(\mathbf{H}^{(m)})$ for all $m\leq n$.
**Until** a convergence criterion is satisfied.
Set $\mathbf{V}^E\leftarrow\mathbf{V}^{E\,(n)}$.
**Calculation of $\mathbf{W}^{E*}$**: Calculate the precoding matrix $\mathbf{W}^{E*}$ for elevation channel from the covariance matrix $\mathbf{V}^E$ via rank reduction as $\mathbf{w}^{E*}_{ji}=\nu_{\max}(\mathbf{V}^E_{ji})$ for all $j\in\mathcal{N}_M$ and $i\in\mathcal{N}_R$.
**2) Short-term Optimization of Azimuth Precoding and Quantization Noise**
**Input:** Channel $\mathbf{H}$ and elevation precoding $\mathbf{W}^{E*}$
**Output:** Azimuth precoding $\mathbf{W}^{A*}(\mathbf{H})$ and quantization noise vector $\boldsymbol{\sigma}^{2*}_x(\mathbf{H})$
Obtain $\mathbf{V}^A(\mathbf{H})$ and $\boldsymbol{\sigma}^2_x(\mathbf{H})$ with $\mathbf{W}^E\leftarrow\mathbf{W}^{E*}$ using Algorithm 2.
**Calculation of $\mathbf{W}^{A*}(\mathbf{H})$**: Calculate the precoding matrix $\mathbf{W}^{A*}(\mathbf{H})$ for the azimuth channel from the covariance matrix $\mathbf{V}^A(\mathbf{H})$ via rank reduction as $\mathbf{w}^{A*}_{ji}(\mathbf{H})=\beta_{ji}\nu_{\max}(\mathbf{V}^A_{ji}(\mathbf{H}))$ for all $j\in\mathcal{N}_M$ and $i\in\mathcal{N}_R$, where $\beta_{ji}$ is obtained by imposing $P_i(\mathbf{W}^{A*}_i(\mathbf{H}),\mathbf{W}^{E*}_i,\boldsymbol{\sigma}^{2,*}_{x,i}(\mathbf{H}))=\bar{P}_i$ using (9).

**Algorithm 2** DC Algorithm for Optimization of $\mathbf{V}^A(\mathbf{H})$ and $\boldsymbol{\sigma}_x^2(\mathbf{H})$

**Input:** Channel $\mathbf{H}$ and elevation precoding $\mathbf{V}^E$.
**Output:** $\mathbf{V}^A(\mathbf{H})$ and $\boldsymbol{\sigma}_x^2(\mathbf{H})$
**Initialization**: Initialize $\mathbf{V}^{A\,(0)}(\mathbf{H}) \succeq 0$ and $\boldsymbol{\sigma}_x^{2\,(0)}(\mathbf{H}) \in \mathbb{R}^+$, and set $l = 0$.
**Repeat**
$l \leftarrow l + 1$
Update $\mathbf{V}^{A(l)}(\mathbf{H})$ and $\boldsymbol{\sigma}_x^{2(l)}(\mathbf{H}$ by solving problem (22).
**Until** a convergence criterion is satisfied.
Set $\mathbf{V}^A(\mathbf{H}) \leftarrow \mathbf{V}^{A\,(l)}(\mathbf{H})$ and $\boldsymbol{\sigma}_x^2(\mathbf{H}) \leftarrow \boldsymbol{\sigma}_x^{2\,(l)}(\mathbf{H})$.

*2) Problem Formulation:* The ergodic achievable sum rate (10) can be optimized over the precoding matrices $\mathbf{W}^A$ and $\mathbf{W}^E$, and over the quantization noise variance vector $\boldsymbol{\sigma}_x^2$ under fronthaul capacity and power constraints. Since the design of the precoding matrix $\mathbf{W}^A$ for azimuth channel and of the compression noise variance $\boldsymbol{\sigma}_x^2$ is adapted to the channel realization $\mathbf{H}$ for each coherence block, we use the notations $\mathbf{W}^A(\mathbf{H})$ and $\boldsymbol{\sigma}_x^2(\mathbf{H})$. The problem of maximizing the achievable rate is then formulated as follows:

$$\underset{\mathbf{w}^A(\mathbf{h}),\mathbf{w}^E,\boldsymbol{\sigma}_x^2(\mathbf{h})}{\text{maximize}} \quad \sum_{j\in\mathcal{N}_M} E\left[R_j\left(\mathbf{h},\mathbf{w}^A(\mathbf{h}),\mathbf{w}^E,\boldsymbol{\sigma}_x^2(\mathbf{h})\right)\right] \tag{11a}$$

$$\text{s.t.} \quad C_{x,i}\left(\mathbf{w}_i^A(\mathbf{h}),\boldsymbol{\sigma}_{x,i}^2(\mathbf{h})\right) \le \bar{C}_i \tag{11b}$$

$$P_i\left(\mathbf{w}_i^A(\mathbf{h}),\mathbf{w}_i^E,\boldsymbol{\sigma}_{x,i}^2(\mathbf{h})\right) \le \bar{P}_i \tag{11c}$$

$\forall\, i \in \mathcal{N}_R$, where the constraints apply for all channel realizations $\mathbf{H}$, and we recall that the capacity constraint on $i$th fronthaul link is $\bar{C}_i$ and the power constraint for RU $i$ is $\bar{P}_i$.

*3) Optimization Algorithm:* In problem (11), the objective function (11a) and constraint (11b) are nonconvex in terms of $\mathbf{W}^A(\mathbf{H})$, $\mathbf{W}^E$, and $\boldsymbol{\sigma}_x^2(\mathbf{H})$. Furthermore, as discussed earlier, $\mathbf{W}^E$ is designed based on stochastic CSI (long-term CSI), whereas $\mathbf{W}^A(\mathbf{H})$ and $\boldsymbol{\sigma}_x^2(\mathbf{H})$ are adapted to instantaneous CSI (short-term CSI). To tackle this problem, we propose an algorithm that optimizes separately the long-term and short-term variables $\mathbf{W}^E$ and $(\mathbf{W}^A(\mathbf{H}),\boldsymbol{\sigma}_x^2(\mathbf{H}))$, respectively. For the former optimization, we adopt a stochastic-optimization-approach-based empirical approximation of the ensemble averages in (11a) following the SSUM method [32]. For the latter, we instead invoke the DC method [26], [33] by leveraging the rank relaxation in obtained by reformulating the optimization problem in terms of the covariance matrices $\mathbf{V}_{ji}^A(\mathbf{H}) = \mathbf{w}_{ji}^A(\mathbf{H})\mathbf{w}_{ji}^{A\dagger}(\mathbf{H})$ and $\mathbf{V}_{ji}^E = \mathbf{w}_{ji}^E\mathbf{w}_{ji}^{E\dagger}$ for all $j \in \mathcal{N}_M$ and $i \in \mathcal{N}_R$. The resulting algorithm is detailed in Algorithm 1 and Appendix A. Note that, in Algorithm 1, long-term optimization has two nested loops in which the inner loop requires at each iteration the solution of a convex problem, whose complexity is polynomial in the problem size [34].

### C. CBP-Based Fronthaul Compression for Layered Precoding

In the proposed CBP-based strategy, as shown in Fig. 7, the CU designs the precoding matrices for both azimuth and elevation components, which are transferred, along with a given subset of downlink information messages, over the fronthaul link to the each RU. As discussed, since the design of the elevation precoding is done based on long-term CSI and hence entails the use of a negligible portion of the fronthaul capacity, the fronthaul overhead depends only on the azimuth precoding matrices, which are adapted to current CSI, and on the information messages. As in [16], the subset of information messages sent to each RU is determined by a preliminary clustering step at the CU, whereby each RU is assigned to serve a subset of the MSs. At each RU, the precoding matrix for FD-MIMO is computed via the Kronecker product between the precoding matrices for the azimuth and elevation channels. Based on the calculated precoding matrix, each RU can then encode and precode the received messages of the assigned MSs. Details are provided in the following.

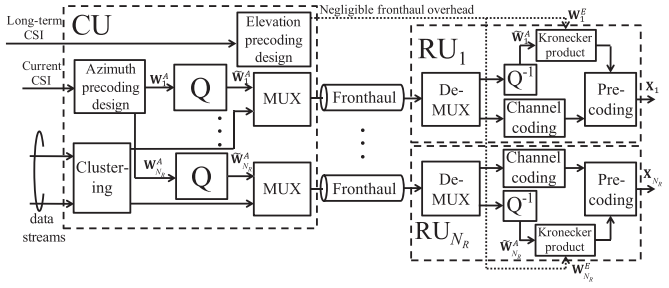


Fig. 7. Block diagram of the Layered CBP scheme ("Q" represents fronthaul compression).

*1) Details and Analysis:* To elaborate, let us denote the set of MSs assigned to RU $i$ as $\mathcal{M}_i \subseteq \mathcal{N}_M$, for all $i \in \mathcal{N}_R$. We also use $\mathcal{M}_i[k]$ to denote the $k$th MS in the set $\mathcal{M}_i$. Note that we assume that the assignment of MSs is given and not subject to optimization. The azimuth precoding vectors $\widetilde{\mathbf{W}}_i^A$ intended for RU $i$ are compressed by the CU and forwarded over the fronthaul link to RU $i$. The compressed azimuth precoding $\mathbf{W}_i^A$ for RU $i$ at the CU is then given by

$$\mathbf{w}_i^A = \widetilde{\mathbf{w}}_i^A + \mathbf{Q}_{w,i} \tag{12}$$

where the quantization noise matrix $\mathbf{Q}_{w,i}$ is assumed to have zero-mean i.i.d. $\mathcal{CN}(0,\sigma_{w,i}^2)$ entries. The required rate for the transfer of the azimuth precoding on fronthaul link is given, similar to (7), as

$$\begin{aligned} C_{w,i}&\left(\widetilde{\mathbf{w}}_i^A,\sigma_{w,i}^2\right) \\ &= \frac{1}{T} I\left(\mathbf{w}_i^A;\widetilde{\mathbf{w}}_i^A\right) \\ &= \frac{1}{T}\left\{\log\det\left(\widetilde{\mathbf{w}}_i^A\widetilde{\mathbf{w}}_i^{A\,\dagger} + \sigma_{w,i}^2\mathbf{I}\right) - \log\det\left(\sigma_{w,i}^2\mathbf{I}\right)\right\} \end{aligned} \tag{13}$$

where $\widetilde{\mathbf{W}}_i^A = [\widetilde{\mathbf{w}}_{\mathcal{M}_i[1]i}^A,\ldots,\widetilde{\mathbf{w}}_{\mathcal{M}_i[|\mathcal{M}_i|]i}^A]$. The remaining fronthaul capacity is used to convey information messages, whose total rate is $\sum_{j\in\mathcal{M}_i} R_j$ with $R_j$ being the user rate for MS $j$. At each RU $i$, the precoding matrix for FD-MIMO is obtained via the Kronecker product of the elevation and azimuth components, yielding the transmitted signal $\mathbf{X}_i = \sum_{j\in\mathcal{M}_i}\mathbf{X}_{ji}$, with

$$\mathbf{X}_{ji} = \left(\mathbf{w}_{ji}^A \otimes \mathbf{w}_{ji}^E\right)\mathbf{s}_j^T = \left(\widetilde{\mathbf{w}}_{ji}^A \otimes \mathbf{w}_{ji}^E\right)\mathbf{s}_j^T + \mathbf{Q}_{w,ji}^A\mathbf{s}_j^T \otimes \mathbf{w}_{ji}^E \tag{14}$$

where we have used the property of the Kronecker product that $(\mathbf{A}\otimes\mathbf{B})\mathbf{C}=(\mathbf{AC}\otimes\mathbf{B})$ if $\mathbf{A}$ and $\mathbf{C}$ are matrices of such sizes that one can form the matrix product $\mathbf{AC}$ (see, e.g., [31]). The power transmitted at RU $i$ is then calculated as

$$P_i\left(\widetilde{\mathbf{w}}_i^A,\mathbf{w}_i^E,\sigma_{w,i}^2\right)=\sum_{j\in\mathcal{M}_i}\left(\left\|\mathbf{w}_{ji}^A\right\|^2\left\|\mathbf{w}_{ji}^E\right\|^2+N_{A,i}\sigma_{w,i}^2\left\|\mathbf{w}_{ji}^E\right\|^2\right). \tag{15}$$

The ergodic achievable rate for MS $j$ is calculated as $E[\bar{R}_j(\mathbf{H},\widetilde{\mathbf{W}}^A,\mathbf{W}^E,\boldsymbol{\sigma}_w^2)]$ with

$$\begin{aligned}&\bar{R}_j\left(\mathbf{h},\widetilde{\mathbf{w}}^A,\mathbf{w}^E,\boldsymbol{\sigma}_w^2\right)\\&=\log\left(1+\sum_{i=1}^{N_R}\sum_{k\in\mathcal{M}_i}\lambda_{ji}^E\left|\mathbf{u}_{ji}^E\mathbf{w}_{ki}^E\right|^2\left(\left|\widetilde{\mathbf{w}}_{ki}^{A\dagger}\mathbf{h}_{ji}^A\right|^2+\sigma_{w,i}^2\left\|\mathbf{h}_{ji}^A\right\|^2\right)\right)\\&\quad-\log\left(1+\sum_{i=1}^{N_R}\sum_{k\in\mathcal{M}_i\setminus j}\lambda_{ji}^E\left|\mathbf{u}_{ji}^E\mathbf{w}_{ki}^E\right|^2\right.\\&\qquad\left.\times\left(\left|\widetilde{\mathbf{w}}_{ki}^{A\dagger}\mathbf{h}_{ji}^A\right|^2+\sigma_{w,i}^2\left\|\mathbf{h}_{ji}^A\right\|^2\right)\right)\end{aligned} \tag{16}$$

where $\widetilde{\mathbf{W}}^A=[\widetilde{\mathbf{W}}_1^{A\,T},\ldots,\widetilde{\mathbf{W}}_{N_R}^{A\,T}]^T$ and $\boldsymbol{\sigma}_w^2=[\sigma_{w,1}^2,\ldots,\sigma_{w,N_R}^2]$.

**Algorithm 3** CBP-Based Fronthaul Compression and Layered Precoding Design

**1) Long-term Optimization of Elevation Precoding and User Rates**

**Input**: Long-term statistics of the channel and clustering $\{\mathcal{M}_i\}$

**Output:** Elevation precoding $\mathbf{W}^{E\,*}$ and MSs' rates $\{R_j\}$

**Initialization (outer loop)**: Initialize the covariance matrix $\mathbf{V}^{E\,(n)}\succeq 0$ subject to $\mathrm{tr}(\mathbf{V}^{E\,(n)})=1$ and $\{R_j^{(n)}\}\in\mathbb{R}^+$, and set $n=0$.

**Repeat**

$n\leftarrow n+1$

Generate a channel matrix realization $\mathbf{H}^{(n)}$ using the available stochastic CSI.

**Inner loop**: Obtain $\widetilde{\mathbf{V}}^{A(n)}(\mathbf{H}^{(n)})$ and $\boldsymbol{\sigma}_w^{2(n)}(\mathbf{H}^{(n)})$ with $\mathbf{V}^E\leftarrow\mathbf{V}^{E(n-1)}$ using Algorithm 4.

Update $\mathbf{V}^{E\,(n)}$ and $\{R_j^{(n)}\}$ by solving problem (31), which depends on $\widetilde{\mathbf{V}}^{A\,(m)}(\mathbf{H}^{(m)})$ and $\boldsymbol{\sigma}_w^{2\,(m)}(\mathbf{H}^{(m)})$ for all $m\le n$.

**Until** a convergence criterion is satisfied.

Set $\mathbf{V}^E\leftarrow\mathbf{V}^{E\,(n)}$ and $\{R_j\}\leftarrow\{R_j^{(n)}\}$.

**Calculation of $\mathbf{W}^{E*}$**: Calculate the precoding matrix $\mathbf{W}^{E*}$ for elevation channel from the covariance matrix $\mathbf{V}^E$ via rank reduction as $\mathbf{w}_{ji}^{E*}=\nu_{\max}(\mathbf{V}_{ji}^E)$ for all $j\in\mathcal{N}_M$ and $i\in\mathcal{N}_R$.

**2) Short-term Optimization of Azimuth Precoding and Quantization Noise**

**Input:** Channel $\mathbf{H}$ and elevation precoding $\mathbf{W}^{E*}$

**Output:** Azimuth precoding $\widetilde{\mathbf{W}}^{A\,*}(\mathbf{H})$ and quantization noise vector $\boldsymbol{\sigma}_w^{2\,*}(\mathbf{H})$

Obtain $\widetilde{\mathbf{V}}^A(\mathbf{H})$ and $\boldsymbol{\sigma}_w^2(\mathbf{H})$ with $\mathbf{W}^E\leftarrow\mathbf{W}^{E*}$ using Algorithm 4.

**Calculation of $\widetilde{\mathbf{W}}^{A\,*}(\mathbf{H})$}**: Calculate the precoding matrix $\widetilde{\mathbf{W}}^{A\,*}(\mathbf{H})$ for the azimuth channel from the covariance matrix $\widetilde{\mathbf{V}}^A(\mathbf{H})$ via rank reduction as $\widetilde{\mathbf{w}}_{ji}^{A\,*}(\mathbf{H})=\beta_{ji}\nu_{\max}(\widetilde{\mathbf{V}}_{ji}^A(\mathbf{H}))$ for all $j\in\mathcal{N}_M$ and $i\in\mathcal{N}_R$, where $\beta_{ji}$ is obtained by imposing $P_i(\widetilde{\mathbf{W}}_i^{A\,*}(\mathbf{H}),\mathbf{W}_i^E,\sigma_{w,i}^{2\,*}(\mathbf{H}))=\bar{P}_i$ using (15).

*2) Problem Formulation:* As discussed in Section IV-B, the azimuth precoding $\widetilde{\mathbf{W}}^A(\mathbf{H})$ and the compression noise variance $\boldsymbol{\sigma}_w^2(\mathbf{H})$ can be adapted to the current channel realization at each coherence block. Accordingly, the optimization problem of interest can be formulated as follows:

$$\underset{\widetilde{\mathbf{w}}^A(\mathbf{h}),\mathbf{w}^E,\{R_j\},\boldsymbol{\sigma}_w^2(\mathbf{h})}{\text{maximize}}\quad\sum_{j\in\mathcal{N}_M}R_j \tag{17a}$$

$$\text{s.t.}\quad R_j\le E\left[\bar{R}_j\left(\mathbf{h},\widetilde{\mathbf{w}}^A(\mathbf{h}),\mathbf{w}^E,\boldsymbol{\sigma}_w^2(\mathbf{h})\right)\right] \tag{17b}$$

$$C_{w,i}\left(\widetilde{\mathbf{w}}_i^A(\mathbf{h}),\sigma_{w,i}^2(\mathbf{h})\right)\le\bar{C}_i-\sum_{j\in\mathcal{M}_i}R_j \tag{17c}$$

$$P_i\left(\widetilde{\mathbf{w}}_i^A(\mathbf{h}),\mathbf{w}_i^E,\sigma_{w,i}^2(\mathbf{h})\right)\le\bar{P}_i \tag{17d}$$

$\forall\, j\in\mathcal{N}_M$ and $\forall\, i\in\mathcal{N}_R$, where the constraints apply to every channel realization $\mathbf{H}$.

**Algorithm 4** DC Algorithm for Optimization of $\widetilde{\mathbf{V}}^A(\mathbf{H})$ and $\boldsymbol{\sigma}_w^2(\mathbf{H})$

**Input:** Channel $\mathbf{H}$ and elevation precoding $\mathbf{V}^E$.

**Output:** $\widetilde{\mathbf{V}}^A(\mathbf{H})$ and $\boldsymbol{\sigma}_w^2(\mathbf{H})$

**Initialization**: Initialize $\widetilde{\mathbf{V}}^{A\,(0)}(\mathbf{H})\succeq 0$ and $\boldsymbol{\sigma}_w^{2\,(0)}(\mathbf{H})\in\mathbb{R}^+$, and set $l=0$.

**Repeat**

$l\leftarrow l+1$

Update $\widetilde{\mathbf{V}}^{A(l)}(\mathbf{H})$ and $\boldsymbol{\sigma}_w^{2(l)}(\mathbf{H})$ by solving problem (28).

**Until** a convergence criterion is satisfied.

Set $\widetilde{\mathbf{V}}^A(\mathbf{H})\leftarrow\widetilde{\mathbf{V}}^{A\,(l)}(\mathbf{H})$ and $\boldsymbol{\sigma}_w^2(\mathbf{H})\leftarrow\boldsymbol{\sigma}_w^{2\,(l)}(\mathbf{H})$.

*3) Optimization Algorithm:* Similar to Section IV-B, the nonconvex functions $\bar{R}_j(\mathbf{H},\widetilde{\mathbf{W}}^A(\mathbf{H}),\mathbf{W}^E,\boldsymbol{\sigma}_w^2(\mathbf{H}))$ and $C_{w,i}(\widetilde{\mathbf{W}}_i^A(\mathbf{H}),\sigma_{w,i}^2(\mathbf{H}))$ can be seen to be DC functions of the covariance matrices $\widetilde{\mathbf{V}}_{ji}^A(\mathbf{H})=\widetilde{\mathbf{w}}_{ji}^A(\mathbf{H})\widetilde{\mathbf{w}}_{ji}^{A\,\dagger}(\mathbf{H})$ and $\mathbf{V}_{ji}^E=\mathbf{w}_{ji}^E\mathbf{w}_{ji}^{E\,\dagger}$ for all $j\in\mathcal{N}_M$ and $i\in\mathcal{N}_R$. Moreover, the optimization problem can be divided into long-term and short-term optimizations, which can be tackled via the SSUM and DC methods, respectively, as summarized in Algorithm 3 and detailed in Appendix B. Moreover, as in Algorithm 1, it is required to solve one convex problem, which has polynomial complexity [34], at each inner iteration.

*Remark 1:* For additional insight regarding the fronthaul overhead for the four considered strategies, we report the number of quantized complex numbers and the number of data streams that are transmitted over the $i$th fronthaul link during a coherence period $T$ in Table I.

TABLE I
FRONTHAUL OVERHEAD FOR THE $i$TH FRONTHAUL LINK (Q.C.N. = QUANTIZED COMPLEX NUMBER)

| Strategy | Fronthaul overhead |
|---|---|
| CAP | $N_{A,i}N_{E,i}T$ q.c.n. |
| CBP | $N_{A,i}N_{E,i}N_C$ q.c.n. + $N_C$ data streams |
| Layered CAP | $N_{A,i}T$ q.c.n. |
| Layered CBP | $N_{A,i}N_C$ q.c.n. + $N_C$ data streams |

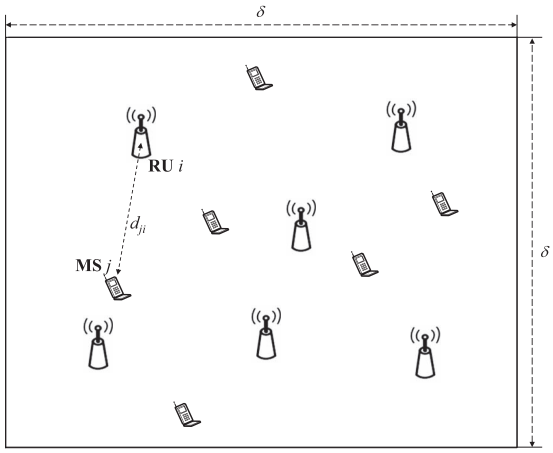


Fig. 8. Simulation environment for the numerical results.

## V. NUMERICAL RESULTS

Here, we compare the performance of the strategies with layered precoding, namely, layered CAP and CBP schemes, and the conventional strategies, namely, CAP and CBP schemes, for FD-MIMO systems. To this end, we consider a setup simulation environment where the RUs and MSs are randomly located in a square area with side $\delta = 500$ m as in Fig. 8. In the path loss formula (4), we set the reference distance to $d_0 = 50$ m and the path-loss exponent to $\eta = 3$ with $d_{ji}$ being the Euclidean distance between the $i$th RU and the $j$th MS. For all numerical results, the Kronecker channel model is, as in [17], such that the $(k, p)$ entry of the elevation covariance matrix is given as

$$\left[\mathbf{R}_{ji}^{E}\right]_{k,p} = e^{j\frac{2\pi d_1}{\lambda}(p-k)\cos\theta} e^{-\frac{1}{2}\left(\xi\frac{2\pi d_1}{\lambda}\right)^2 (p-k)^2 \sin^2\theta} \quad (18)$$

where the elevation angle-of-departure (AoD) $\theta = 3\pi/8$, the variance of elevation angular perturbation $\xi = \pi/36$, and $d_1 = \lambda/2$, with $\lambda = c/f_c$ and carrier frequency $f_c = 2.4$ GHz, whereas the $(l, q)$ entry of the azimuth correlation matrix is given as

$$\left[\mathbf{R}_{ji}^{A}\right]_{l,q} = \frac{1}{\sqrt{D_5}} e^{-\frac{D_3^2\cos^2\phi_{ji}}{2-D_5}} e^{j\frac{D_2\cos\phi_{ji}}{D_5}} e^{-\frac{1}{2}\frac{(D_2\tilde{\sigma}_{ji})^2}{D_5}} \quad (19)$$

where the azimuth AoD $\phi_{ji}$ is calculated based on the location of RU $i$ and MS $j$, $D_2 = 2\pi d_2/\lambda(q-l)\sin\theta$, $D_3 = \xi(2\pi d_2/\lambda)(q-l)\cos\theta$, $D_5 = D_3^2\tilde{\sigma}^2 + 1$, $d_2 = \lambda/2$, and $\tilde{\sigma}_{ji} = (\sin\phi_{ji})\sigma$, with the variance of azimuth angular perturbation $\sigma = \pi/12$. In the following, we assume that every RU is subject to the same fronthaul capacity $\bar{C}$ and has the same power constraint $\bar{P}$, namely $\bar{C}_i = \bar{C}$ and $\bar{P}_i = \bar{P}$ for $i \in \mathcal{N}_R$. We also consider CBP strategies in which each RU serves $N_C$ MSs that have the largest instantaneous channel norms. Note that this assignment is done for each coherence block.

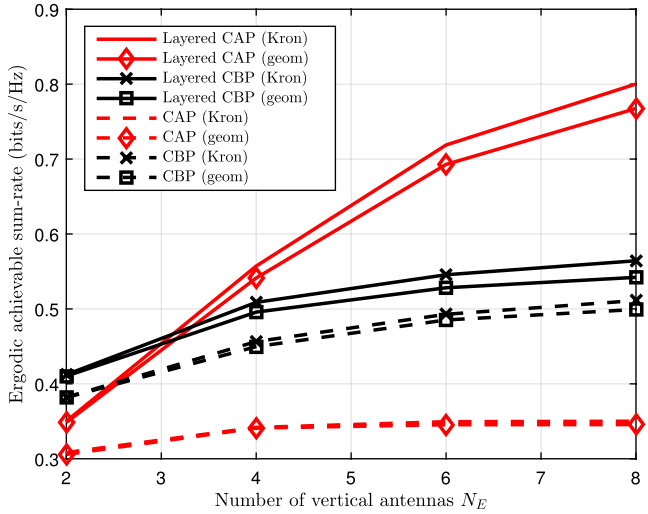


Fig. 9. Ergodic achievable sum rate versus the number of vertical antennas $N_E$ ($N_R = N_M = 2$, $N_{A,i} = 2$, $\bar{C} = 1$ bit/s/Hz, $\bar{P} = 0$ dB, and $T = 20$).

Fig. 9 shows the ergodic achievable sum rate as function of the number of vertical antennas $N_E$, where the number of RUs and MSs is $N_R = N_M = 2$, the number of horizontal antennas is $N_{A,i} = 2$ for all $i \in \mathcal{N}_R$, the number of MSs served by each RU is $N_C = N_M$, the fronthaul capacity is $\bar{C} = 1$ bit/s/Hz, the transmit power is $\bar{P} = 0$ dB, and the coherence time is $T = 20$. Note that we set $N_E \geq 2$ since for $N_E = 1$ layered precoding is not applicable. In Fig. 9, we compare the performance achievable under the Kronecker model with that under a standard geometry-based channel model (see [17, Eq. (6)]), with same angular spread parameters and 20 multipath components. It is observed that, although the proposed scheme is justified by a Kronecker channel model, the approach yields comparable gains also under a geometry-based model. These results provide a validation of the adoption of the Kronecker model.

Moreover, we observe in Fig. 9 that the layered precoding schemes provide increasingly large gains as $N_E$ grows larger. This is because, in the conventional strategies, the fronthaul overhead for the transfer of elevation precoding information increases with the number of vertical antennas. This gain is less pronounced here for layered CBP strategies, whose achievable rate is limited here by the relatively small coherence interval, as further discussed in the following (see also Section III-B). Moreover, it is observed that the conventional CAP strategy outperforms the layered CAP strategy for small values of $N_E$. This is caused by the fact that, with the layered CAP strategy, the azimuth precoded signals for the MSs are separately compressed, hence entailing an inefficient use of the fronthaul when $N_E$ is small enough.

Fig. 10 shows the effect of the number of MSs $N_M$ on the ergodic achievable sum rate with $N_R = 2$, $N_{A,i} = 2$, and $N_{E,i} = 4$ for all $i \in \mathcal{N}_R$, $N_C = N_M$, $\bar{C} = 3$ bit/s/Hz, $\bar{P} = 5$ dB, and $T = 20$. The CBP methods show the known poor performance as the number of MSs increases, due to the need for the transmission of the messages of all MSs on all fronthaul links [16]. Moreover, in keeping with the discussion earlier, we observe that the conventional CAP method is to be preferred in the regime of a large number of MSs. This is due to the separate

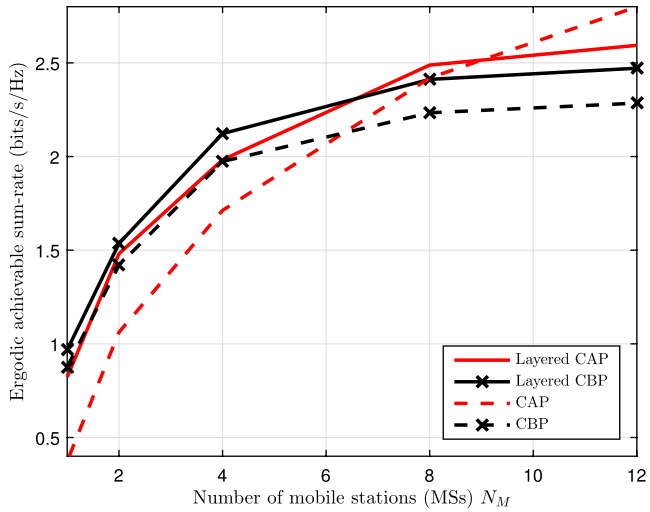


Fig. 10. Ergodic achievable sum rate versus the number of MSs $N_M$ ($N_R = 2$, $N_{A,i} = 2$, $N_{E,i} = 4$, $\bar{C} = 3$ bit/s/Hz, $\bar{P} = 5$ dB, and $T = 20$).

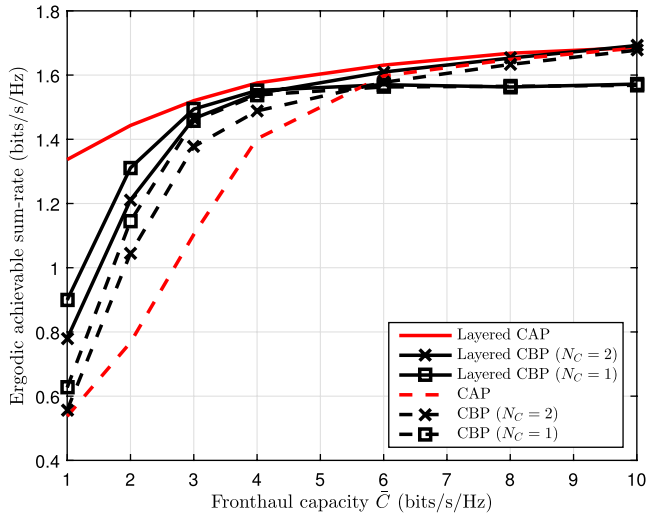


Fig. 11. Ergodic achievable sum rate versus the fronthaul capacity $\bar{C}$ ($N_R = 2$, $N_M = 2$, $N_{A,i} = 2$, $N_{E,i} = 4$, $\bar{P} = 5$ dB, and $T = 10$).

compression of the azimuth precoded signals of layered CAP, which entails a fronthaul overhead proportional to the number of MSs.

In Fig. 11, the ergodic achievable sum rate is plotted versus the fronthaul capacity $\bar{C}$ for $N_R = N_M = 2$, $N_{A,i} = 2$, and $N_{E,i} = 4$ for all $i \in \mathcal{N}_R$, $\bar{P} = 5$ dB, and $T = 10$. We first remark that the performance gain of the layered strategies is observed at low-to-moderate fronthaul capacities, while, for large fronthaul capacities, the performance of the conventional strategies approach that of the layered strategies. As a general rule, the conventional CAP strategy is uniformly better than conventional CBP, as long as the fronthaul capacity is sufficiently large, due to the enhanced interference mitigation capabilities of CAP [16]. Instead, the layered CAP strategy is advantageous here across all values of fronthaul capacity. Moreover, it is shown that, for small values of the fronthaul capacity $\bar{C}$, CBP schemes with smaller $N_C$ tend to have better performance. This is because the fronthaul overhead required for supplying a smaller number of data streams to the RUs is reduced. Instead, for large values of $N_C$, it is preferable to

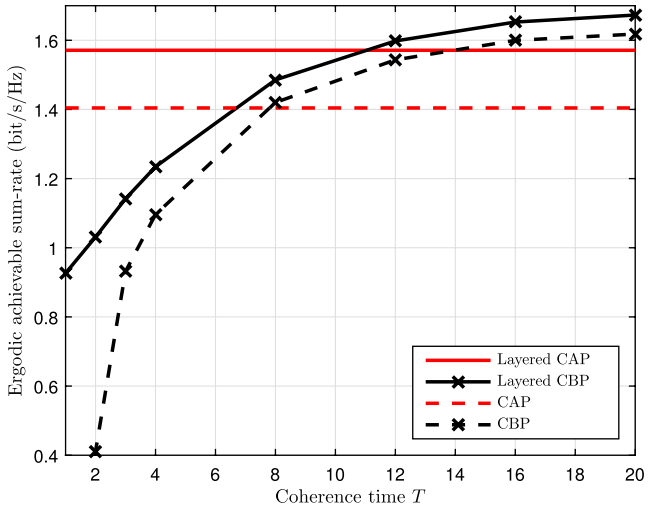


Fig. 12. Ergodic achievable sum rate versus the coherence time $T$ ($N_R = N_M = 2$, $N_{A,i} = 2$, $N_{E,i} = 4$, $\bar{C} = 4$ bit/s/Hz, and $\bar{P} = 5$ dB).

choose a large $N_C$ in order to benefit from the cooperation gains due to joint RU precoding.

The effect of the coherence time $T$ is investigated in Fig. 12, with $N_R = N_M = 2$, $N_{A,i} = 2$, and $N_{E,i} = 4$ for all $i \in \mathcal{N}_R$, $N_C = N_M$, $\bar{C} = 4$ bit/s/Hz, and $\bar{P} = 5$ dB. The CBP schemes benefit from a larger coherence time $T$ since the fronthaul overhead required to transmit precoding information is amortized over a larger period. In contrast, such overhead in layered CAP and CAP schemes scales proportionally to the coherence time $T$; hence, the layered CAP and CAP schemes are not affected by the coherence time.

## VI. Concluding Remarks

In this paper, we have studied the design of downlink C-RAN systems in which the RUs are equipped with FD-MIMO arrays. We proposed to leverage the special low-rank structure of the FD-MIMO channel, which exhibits different rates of variability in the elevation and azimuth components, by means of a novel layered precoding strategy coupled with an adaptive fronthaul compression scheme. Specifically, in the layered strategy, a single precoding matrix is optimized for the elevation channel across all coherence times based on long-term CSI, although azimuth precoding matrices are optimized across independent coherence interval by adapting to instantaneous CSI. This proposed layered approach has the unique advantage in a C-RAN of potentially reducing the fronthaul overhead, due to the opportunity to amortize the overhead related to the elevation channel component across multiple coherence times. Via numerical results, it is shown that the layered strategies significantly outperform standard nonlayered schemes, particularly in the regime of low fronthaul capacity and a large number of vertical antennas.

We have also considered two different functional splits for both layered and nonlayered precoding, namely the conventional C-RAN implementation, also known as the CAP scheme, and an alternative split, referred to as CBP, whereby channel coding and precoding are performed at the RUs. Layered precoding is seen to work better under a CAP implementation

when the coherence interval is not too large and when the number of vertical antennas is sufficiently large, whereas the CBP approach benefits from a longer coherence interval due to its capability to amortize the fronthaul overhead for transfer of azimuth precoding information. Interesting open issues include the investigation of a scenario with multiple interfering clusters of RUs controlled by distinct CUs (see [35]), and the analysis of the performance in the presence of more general FD-MIMO channel models (see, e.g., [12] and [20]) and in more advanced system models (e.g., with stochastically coupled azimuth and elevation components).

## APPENDIX A
## OPTIMIZATION ALGORITHM FOR THE LAYERED COMPRESS-AFTER-PRECODING STRATEGY

Here, we detail the derivation of Algorithm 1 for the optimization of the layered CAP strategy. We first discuss the optimization problem for the short-term variables, namely the covariance matrix $\mathbf{V}^A(\mathbf{H})$ for azimuth precoding and the quantization noise variance $\boldsymbol{\sigma}_x^2(\mathbf{H})$, which are adapted to the channel realization $\mathbf{H}$, for given the elevation covariance matrix $\mathbf{V}^E$. We then consider the optimization of the long-term variable, namely the covariance matrix $\mathbf{V}^E$ for elevation precoding, with the given covariance matrices $\mathbf{V}^A(\mathbf{H})$ for azimuth precoding and quantization noise vectors $\boldsymbol{\sigma}_x^2(\mathbf{H})$.

After obtaining the elevation covariance matrix $\mathbf{V}^{E*}$, using the approach in Algorithm 1, the precoding matrix $\mathbf{W}^{E*}$ for the elevation channel is calculated via the principal eigenvector approximation [36] of the obtained solution $\mathbf{V}^{E*}$ as $\mathbf{w}_{ji}^{E*} = \nu_{\max}(\mathbf{V}_{ji}^{E*})$ for all $j \in \mathcal{N}_M$ and $i \in \mathcal{N}_R$. In a similar fashion, the algorithm obtains the precoding matrix $\mathbf{W}^{A*}(\mathbf{H})$ for the azimuth channel via the standard rank-reduction approach [36] from the obtained solution $\mathbf{V}^A(\mathbf{H})^*$ as $\mathbf{w}_{ji}^{A*}(\mathbf{H}) = \beta_{ji}\nu_{\max}(\mathbf{V}_{ji}^A(\mathbf{H})^*)$ with the normalization factors $\beta_{ji}$ selected to satisfy the power constraint with equality, namely $P_i(\mathbf{W}_i^{A*}(\mathbf{H}), \mathbf{W}_i^{E*}, \boldsymbol{\sigma}_{x,i}^{2\,*}(\mathbf{H})) = \bar{P}_i$.

*1) Optimization Over $\mathbf{V}^A(\mathbf{H})$ and $\boldsymbol{\sigma}_x^2(\mathbf{H})$ With Given $\mathbf{V}^E$:* Here, we tackle the problem (11) based on the DC algorithm [26] given the elevation precoding covariance matrix $\mathbf{V}^E$ over the azimuth covariance matrix $\mathbf{V}^A(\mathbf{H})$ and the quantization noise variance $\boldsymbol{\sigma}_x^2(\mathbf{H})$. To this end, the objective function $R_j(\mathbf{H}, \mathbf{W}^A(\mathbf{H}), \mathbf{W}^E, \boldsymbol{\sigma}_x^2(\mathbf{H}))$ is approximated by a locally tight lower bound $\widetilde{R}_j(\mathbf{H}, \mathbf{V}^A(\mathbf{H}), \boldsymbol{\sigma}_x^2(\mathbf{H}) | \mathbf{V}^{A(l-1)}(\mathbf{H}), \boldsymbol{\sigma}_x^{2\,(l-1)}(\mathbf{H}), \mathbf{V}^E)$ around solutions $\mathbf{V}^{A\,(l-1)}(\mathbf{H})$ and $\boldsymbol{\sigma}_x^{2\,(l-1)}(\mathbf{H})$ obtained at $(l-1)$th inner iteration with

$$\begin{aligned}
&\widetilde{R}_j\left(\mathbf{h}, \mathbf{V}^A(\mathbf{h}), \boldsymbol{\sigma}_x^2(\mathbf{h}) | \mathbf{V}^{A\,(l-1)}(\mathbf{h}), \boldsymbol{\sigma}_x^{2\,(l-1)}(\mathbf{h}), \mathbf{V}^E\right) \\
&= \log\left(1 + \sum_{k=1}^{N_M}\sum_{i=1}^{N_R} \rho_{ji}\left(\mathbf{V}_{ki}^A(\mathbf{h}), \mathbf{V}_{ki}^E, \sigma_{x,ki}^2(\mathbf{h})\right)\right) \\
&\quad - f\left(\Gamma_j\left(\mathbf{V}^{A(l-1)}(\mathbf{h}), \mathbf{V}^E, \boldsymbol{\sigma}_x^{2(l-1)}(\mathbf{h})\right)\right. \\
&\qquad \left.\Gamma_j\left(\mathbf{V}^A(\mathbf{h}), \mathbf{V}^E, \boldsymbol{\sigma}_x^2(\mathbf{h})\right)\right) \qquad (20)
\end{aligned}$$

where $\Gamma_j(\mathbf{V}^A, \mathbf{V}^E, \boldsymbol{\sigma}_x^2) = 1 + \sum_{k=1, k\neq j}^{N_M}\sum_{i=1}^{N_R} \rho_{ji}(\mathbf{V}_{ki}^A, \mathbf{V}_{ki}^E, \sigma_{x,ki}^2)$ and $\rho_{ji}(\mathbf{V}_{ki}^A, \mathbf{V}_{ki}^E, \sigma_{x,ki}^2) = \lambda_{ji}^E \mathbf{u}_{ji}^E \mathbf{V}_{ki}^E \mathbf{u}_{ji}^\dagger (\mathbf{h}_{ji}^A \mathbf{V}_{ki}^A \mathbf{h}_{ji}^{A\,\dagger} + \sigma_{x,ki}^2 \|\mathbf{h}_{ji}^A\|^2)$, and the linearized function $f(a,b)$ is obtained from the first-order Taylor expansion of the log function as $f(a,b) = \log(a) + (b-a)/a$. Since the fronthaul constraint (11b) is a DC constraint, the left-hand side of the constraint (11b) is approximated by applying successive locally tight convex lower bounds as

$$\begin{aligned}
&\widetilde{C}_{x,i}\left(\mathbf{V}_i^A(\mathbf{h}), \boldsymbol{\sigma}_{x,i}^2(\mathbf{h}) \middle| \mathbf{V}_i^{A\,(l-1)}(\mathbf{h}), \boldsymbol{\sigma}_{x,i}^{2\,(l-1)}(\mathbf{h})\right) \qquad (21) \\
&\quad \triangleq \sum_{j=1}^{N_M}\left\{ f\left(\mathrm{tr}\left(\mathbf{V}_{ji}^{A(l-1)}(\mathbf{h})\right) + \sigma_{x,ji}^{2(l-1)}(\mathbf{h}), \mathrm{tr}\left(\mathbf{V}_{ji}^A(\mathbf{h})\right)\right. \right. \\
&\qquad \left. \left. + \sigma_{x,ji}^2(\mathbf{h})\right) - \log\sigma_{x,ji}^2(\mathbf{h})\right\}.
\end{aligned}$$

At the $l$th inner loop, the following convex optimization problem, for given $\mathbf{V}^{A\,(l-1)}(\mathbf{H})$, $\boldsymbol{\sigma}_x^{2\,(l-1)}(\mathbf{H})$, and $\mathbf{V}^E$, is solved for obtaining new iterates $\mathbf{V}^{A\,(l)}(\mathbf{H})$ and $\boldsymbol{\sigma}_x^{2\,(l)}(\mathbf{H})$ as

$$\begin{aligned}
\max \quad & \sum_{j\in\mathcal{N}_M} \widetilde{R}_j\Big(\mathbf{h}, \mathbf{V}^A(\mathbf{h}), \boldsymbol{\sigma}_x^2(\mathbf{h}) \Big| \mathbf{V}^{A(l-1)}(\mathbf{h}), \boldsymbol{\sigma}_x^{2(l-1)}(\mathbf{h}), \mathbf{V}^E\Big) \\
\text{s.t.} \quad & \widetilde{C}_{x,i}\left(\mathbf{V}_i^A(\mathbf{h}), \boldsymbol{\sigma}_{x,i}^2(\mathbf{h}) \middle| \mathbf{V}_i^{A(l-1)}(\mathbf{h}), \boldsymbol{\sigma}_{x,i}^{2(l-1)}(\mathbf{h})\right) \le \bar{C}_i \\
& P_i\left(\mathbf{V}_i^A(\mathbf{h}), \mathbf{V}_i^E, \boldsymbol{\sigma}_{x,i}^2(\mathbf{h})\right) \le \bar{P}_i \qquad (22)
\end{aligned}$$

$\forall\, i \in \mathcal{N}_R$. The DC method obtains the solutions $\mathbf{V}^A(\mathbf{H})$ and $\boldsymbol{\sigma}_x^2(\mathbf{H})$ by solving the problem (22) iteratively over $l$ until a convergence criterion is satisfied and the resulting algorithm is summarized in Algorithm 2.

*2) Optimization over $\mathbf{V}^E$:* In this part, the covariance matrix $\mathbf{V}^E$ for elevation precoding is designed for given azimuth precoding covariance matrices $\mathbf{V}^{A\,(m)} = \mathbf{V}^{A\,(m)}(\mathbf{H}^{(m)})$ and quantization noise vectors $\boldsymbol{\sigma}_x^{2\,(m)} = \boldsymbol{\sigma}_x^{2\,(m)}(\mathbf{H}^{(m)})$ for all $m = 1, \ldots, n$. Since the elevation covariance matrix $\mathbf{V}^{E\,(n)}$ is not adapted to the channel realization $\mathbf{H}$ and the objective function (11) is nonconvex with respect to $\mathbf{V}^{E\,(n)}$, in this optimization, we use the SSUM algorithm [32]. To this end, at each step, a stochastic lower bound of the objective function is maximized around the current iterate. Following the SSUM method, at $n$th outer loop, the objective function with given $\mathbf{V}^{A\,(m)}$ and $\boldsymbol{\sigma}_x^{2\,(m)}$, for all $m = 1, \ldots, n$, is reformulated as the empirical average, i.e.,

$$\frac{1}{n}\sum_{m=1}^{n} \widetilde{R}_j\left(\mathbf{h}^{(m)}, \mathbf{V}^E \middle| \mathbf{V}^{E\,(m-1)}, \mathbf{V}^{A\,(m)}, \boldsymbol{\sigma}_x^{2\,(m)}\right) \qquad (23)$$

where $\widetilde{R}_j(\mathbf{H}^{(m)}, \mathbf{V}^E | \mathbf{V}^{E\,(m-1)}, \mathbf{V}^A, \boldsymbol{\sigma}_x^2)$ is a locally tight convex lower bound around the previous iterate $\mathbf{V}^{E\,(m-1)}$, when the channel realization is $\mathbf{H}^{(m)}$, and is calculated as

$$\begin{aligned}
&\widetilde{R}_j\left(\mathbf{h}^{(m)}, \mathbf{V}^E | \mathbf{V}^{E\,(m-1)}, \mathbf{V}^{A\,(m)}, \boldsymbol{\sigma}_x^{2\,(m)}\right) \\
&= \log\left(1 + \sum_{k=1}^{N_M}\sum_{i=1}^{N_R} \rho_{ji}\left(\mathbf{h}^{(m)}, \mathbf{V}_{ki}^{A\,(m)}, \mathbf{V}_{ki}^E, \sigma_{x,ki}^{2\,(m)}\right)\right) \\
&\quad - f\left(\Gamma_j^{(m)}\left(\mathbf{V}^{A(m)}, \mathbf{V}^{E(m-1)}, \boldsymbol{\sigma}_x^{2(m)}\right)\right. \\
&\qquad \left.\times\, \Gamma_j^{(m)}\left(\mathbf{V}^{A(m)}, \mathbf{V}^E, \boldsymbol{\sigma}_x^{2(m)}\right)\right) \qquad (24)
\end{aligned}$$

with $\Gamma_j^{(m)}(\mathbf{V}^A,\mathbf{V}^E,\boldsymbol{\sigma}_x^2)=1+\sum_{k=1,k\neq j}^{N_M}\sum_{i=1}^{N_R}\rho_{ji}(\mathbf{H}^{(m)},\mathbf{V}_{ki}^A,\mathbf{V}_{ki}^E,\sigma_{x,ki}^2)$ and $\rho_{ji}(\mathbf{H}^{(m)},\mathbf{V}_{ki}^A,\mathbf{V}_{ki}^E,\sigma_{x,ki}^2)=\lambda_{ji}^{E\,(m)}\mathbf{u}_{ji}^{E\,(m)}\mathbf{V}_{ki}^E\mathbf{u}_{ji}^{(m)\,\dagger}(\mathbf{h}_{ji}^{A\,(m)}\mathbf{V}_{ki}^A\mathbf{h}_{ji}^{A\,(m)\,\dagger}+\sigma_{x,ki}^2\|\mathbf{h}_{ji}^{A\,(m)}\|^2)$. The $n$th iterate $\mathbf{V}^{E\,(n)}$ is obtained by solving the following convex optimization problem:

$$\begin{aligned}\max\quad & \frac{1}{n}\sum_{m=1}^{n}\sum_{j\in\mathcal{N}_M}\widetilde{R}_j\left(\mathbf{h}^{(m)},\mathbf{V}^E|\mathbf{V}^{E(m-1)},\mathbf{V}^{A(m)},\boldsymbol{\sigma}_x^{2(m)}\right)\\ \text{s.t.}\quad & C_{x,i}\left(\mathbf{V}_i^{A\,(n)},\boldsymbol{\sigma}_{x,i}^{2\,(n)}\right)\le\bar{C}_i\qquad\forall\, i\in\mathcal{N}_R\\ & P_i\left(\mathbf{V}_i^{A\,(n)},\mathbf{V}_i^E,\boldsymbol{\sigma}_{x,i}^{2\,(n)}\right)\le\bar{P}_i\qquad\forall\, i\in\mathcal{N}_R.\end{aligned}\tag{25}$$

As in Appendix A1, the outer loop in Algorithm 1 is repeated until the convergence is achieved.

## Appendix B
## Optimization Algorithm for the Layered Compress-Before-Precoding Strategy

Here, the precoding matrices $\mathbf{W}^{E*}$ and $\widetilde{\mathbf{W}}^{A\,*}$, MSs' rates $\{R_j\}$, and quantization noise vector $\boldsymbol{\sigma}_w^{2\,*}$ are jointly optimized for the CBP-based strategy. The optimization of short-term variables, namely the covariance matrix $\widetilde{\mathbf{V}}^A(\mathbf{H})$ for azimuth precoding and the quantization noise variance $\boldsymbol{\sigma}_w^2(\mathbf{H})$, which are adapted to the channel realization $\mathbf{H}$ for given the elevation covariance matrix $\mathbf{V}^E$, is described first. Then, the optimization over the long-term variables, namely the covariance matrix $\mathbf{V}^E$ for elevation precoding and the user rates $\{R_j\}$, is discussed given covariance matrices $\mathbf{V}^{A\,(m)}(\mathbf{H})$ for azimuth precoding and quantization noise vectors $\boldsymbol{\sigma}_w^{2\,(m)}(\mathbf{H})$, for all $m=1,\ldots,n$, as detailed in Appendix B-B.

As in Appendix A, the elevation precoding matrix $\mathbf{W}^{E*}$ and the azimuth precoding matrix $\widetilde{\mathbf{W}}^{A\,*}$ are calculated via the standard rank-reduction approach [36] with the obtained solutions $\mathbf{V}^{E*}$ and $\widetilde{\mathbf{V}}^{A\,*}$, respectively, as detailed in Algorithm 3.

*1) Optimization Over $\widetilde{\mathbf{V}}^A(\mathbf{H})$ and $\boldsymbol{\sigma}_w^2(\mathbf{H})$ with Given $\mathbf{V}^E$:* Here, we aim at maximizing the objective function (17a) over the azimuth precoding covariance matrix $\widetilde{\mathbf{V}}^A(\mathbf{H})$ and the quantization noise variance $\boldsymbol{\sigma}_w^2(\mathbf{H})$ given the elevation precoding covariance matrix $\mathbf{V}^E$ using the DC method [26]. At the $l$th iteration of the DC method, the nonconvex functions $\bar{R}_j(\mathbf{H},\widetilde{\mathbf{V}}^A(\mathbf{H}),\mathbf{V}^E,\boldsymbol{\sigma}_w^2(\mathbf{H}))$ and $C_{w,i}(\widetilde{\mathbf{V}}_i^A(\mathbf{H}),\sigma_{w,i}^2(\mathbf{H}))$ are, respectively, substituted with a locally tight lower bound $\widetilde{R}_j(\mathbf{H},\widetilde{\mathbf{V}}^A(\mathbf{H}),\boldsymbol{\sigma}_w^2(\mathbf{H})|\widetilde{\mathbf{V}}^{A\,(l-1)}(\mathbf{H}),\boldsymbol{\sigma}_w^{2\,(l-1)}(\mathbf{H}),\mathbf{V}^E)$ and a tight upper bound $\widetilde{C}_{w,i}(\widetilde{\mathbf{V}}_i^A(\mathbf{H}),\sigma_{w,i}^2(\mathbf{H})|\widetilde{\mathbf{V}}_i^{A\,(l-1)}(\mathbf{H}),\sigma_{w,i}^{2\,(l-1)}(\mathbf{H}))$ obtained as in Appendix A. The bounds are given by

$$\begin{aligned}&\widetilde{R}_j\left(\mathbf{h},\widetilde{\mathbf{V}}^A(\mathbf{h}),\boldsymbol{\sigma}_w^2(\mathbf{h})\left|\widetilde{\mathbf{V}}^{A\,(l-1)}(\mathbf{h}),\boldsymbol{\sigma}_w^{2\,(l-1)}(\mathbf{h}),\mathbf{V}^E\right.\right)\\ &\quad=\log\left(1+\sum_{i=1}^{N_R}\sum_{k\in\mathcal{M}_i}\rho_{ji}\left(\widetilde{\mathbf{V}}_{ki}^A(\mathbf{h}),\mathbf{V}_{ki}^E,\sigma_{w,i}^2(\mathbf{h})\right)\right)\\ &\qquad-f\left(\Gamma_j\left(\widetilde{\mathbf{V}}^{A(l-1)}(\mathbf{h}),\mathbf{V}^E,\boldsymbol{\sigma}_w^{2(l-1)}(\mathbf{h})\right)\right.\\ &\qquad\left.\Gamma_j\left(\widetilde{\mathbf{V}}^A(\mathbf{h}),\mathbf{V}^E,\boldsymbol{\sigma}_w^2(\mathbf{h})\right)\right)\end{aligned}\tag{26}$$

$$\begin{aligned}&\widetilde{C}_{w,i}\left(\widetilde{\mathbf{V}}_i^A(\mathbf{h}),\sigma_{w,i}^2(\mathbf{h})\left|\widetilde{\mathbf{V}}_i^{A\,(l-1)}(\mathbf{h}),\sigma_{w,i}^{2\,(l-1)}(\mathbf{h})\right.\right)\\ &\quad\triangleq\frac{1}{T}\left\{f(\widetilde{\mathbf{V}}_i^{A\,(l-1)}(\mathbf{h})+\sigma_{w,i}^{2\,(l-1)}(\mathbf{h})\mathbf{I},\widetilde{\mathbf{V}}_i^A(\mathbf{h})+\sigma_{w,i}^2(\mathbf{h})\mathbf{I})\right.\\ &\qquad\left.-N_{A,i}\log\left(\sigma_{w,i}^2\right)\right\}\end{aligned}\tag{27}$$

where $\Gamma_j(\widetilde{\mathbf{V}}^A,\mathbf{V}^E,\boldsymbol{\sigma}_w^2)=1+\sum_{i=1}^{N_R}\sum_{k\in\mathcal{M}_i\setminus j}\rho_{ji}(\widetilde{\mathbf{V}}_{ki}^A,\mathbf{V}_{ki}^E,\sigma_{w,i}^2)$ and $\rho_{ji}(\widetilde{\mathbf{V}}_{ki}^A,\mathbf{V}_{ki}^E,\sigma_{w,i}^2)=\lambda_{ji}^E\mathbf{u}_{ji}^E\mathbf{V}_{ki}^E\mathbf{u}_{ji}^{\dagger}(\mathbf{h}_{ji}^A\widetilde{\mathbf{V}}_{ki}^A\mathbf{h}_{ji}^{A\,\dagger}+\sigma_{w,i}^2\|\mathbf{h}_{ji}^A\|^2)$, and the linearization function $f(\mathbf{A},\mathbf{B})$ for the matrices is defined as $f(\mathbf{A},\mathbf{B})\triangleq\log\det(\mathbf{A})+\mathrm{tr}(\mathbf{A}^{-1}(\mathbf{B}-\mathbf{A}))$.

At $l$th iteration of the DC method, the following convex optimization problem for given $\widetilde{\mathbf{V}}^{A\,(l-1)}(\mathbf{H})$, $\boldsymbol{\sigma}_w^{2\,(l-1)}(\mathbf{H})$, and $\mathbf{V}^E$ is solved for obtaining new iterates $\widetilde{\mathbf{V}}^{A\,(l)}(\mathbf{H})$ and $\boldsymbol{\sigma}_w^{2\,(l)}(\mathbf{H})$, i.e.,

$$\begin{aligned}&\underset{\widetilde{\mathbf{V}}^A(\mathbf{h}),\boldsymbol{\sigma}_w^2(\mathbf{h}),\{R_j\}}{\arg\max}\quad\sum_{j\in\mathcal{N}_M}R_j\\ &\text{s.t.}\quad R_j\le\widetilde{R}_j\left(\mathbf{h},\widetilde{\mathbf{V}}^A(\mathbf{h}),\boldsymbol{\sigma}_w^2(\mathbf{h})\left|\widetilde{\mathbf{V}}^{A(l-1)}(\mathbf{h}),\boldsymbol{\sigma}_w^{2(l-1)}(\mathbf{h}),\mathbf{V}^E\right.\right)\\ &\widetilde{C}_{w,i}\left(\widetilde{\mathbf{V}}_i^A(\mathbf{h}),\sigma_{w,i}^2(\mathbf{h})|\widetilde{\mathbf{V}}_i^{A(l-1)}(\mathbf{h}),\sigma_{w,i}^{2(l-1)}(\mathbf{h})\right)\le\bar{C}_i-\sum_{j\in\mathcal{M}_i}R_j\\ &P_i\left(\widetilde{\mathbf{V}}_i^A(\mathbf{h}),\mathbf{V}_i^E,\sigma_{w,i}^2(\mathbf{h})\right)\le\bar{P}_i\end{aligned}\tag{28}$$

$\forall\, j\in\mathcal{N}_M$ and $\forall\, i\in\mathcal{N}_R$. Problem (28) is solved iteratively over $l$ until convergence and the resulting algorithm is summarized in Algorithm 4.

*2) Optimization Over $\mathbf{V}^E$ and $\{R_j\}$:* We design the covariance matrix $\mathbf{V}^E$ for elevation precoding and the user rates $\{R_j\}$ for given azimuth precoding covariance matrices $\widetilde{\mathbf{V}}^{A\,(m)}=\widetilde{\mathbf{V}}^{A\,(m)}(\mathbf{H}^{(m)})$ and quantization noise vectors $\boldsymbol{\sigma}_w^{2\,(m)}=\boldsymbol{\sigma}_w^{2\,(m)}(\mathbf{H}^{(m)})$ for all $m=1,\ldots,n$. As in Appendix A, this optimization problem can be tackled via the SSUM method. To this end, the function $E[\bar{R}_j(\mathbf{H},\widetilde{\mathbf{W}}^A(\mathbf{H}),\mathbf{W}^E,\boldsymbol{\sigma}_w^2(\mathbf{H}))]$ in (17b) is approximated with the stochastic upper bound as

$$\frac{1}{n}\sum_{m=1}^{n}\widetilde{R}_j\left(\mathbf{h}^{(m)},\mathbf{V}^E\left|\mathbf{V}^{E\,(m-1)},\widetilde{\mathbf{V}}^{A\,(m)},\boldsymbol{\sigma}_w^{2\,(m)}\right.\right)\tag{29}$$

with

$$\begin{aligned}&\widetilde{R}_j\left(\mathbf{h}^{(m)},\mathbf{V}^E\left|\mathbf{V}^{E\,(m-1)},\widetilde{\mathbf{V}}^{A\,(m)},\boldsymbol{\sigma}_w^{2\,(m)}\right.\right)\\ &\quad=\log\left(1+\sum_{i=1}^{N_R}\sum_{k\in\mathcal{M}_i}\rho_{ji}\left(\mathbf{h}^{(m)},\widetilde{\mathbf{V}}_{ki}^{A\,(m)},\mathbf{V}_{ki}^E,\sigma_{w,i}^{2\,(m)}\right)\right)\\ &\qquad-f\left(\Gamma_j^{(m)}\left(\widetilde{\mathbf{V}}^{A(m)},\mathbf{V}^{E(m-1)},\boldsymbol{\sigma}_w^{2(m)}\right)\right.\\ &\qquad\left.\Gamma_j^{(m)}\left(\widetilde{\mathbf{V}}^{A(m)},\mathbf{V}^E,\boldsymbol{\sigma}_w^{2(m)}\right)\right)\end{aligned}\tag{30}$$

where $\Gamma_j^{(m)}(\widetilde{\mathbf{V}}^A,\mathbf{V}^E,\boldsymbol{\sigma}_w^2)=1+\sum_{i=1}^{N_R}\sum_{k\in\mathcal{M}_i\setminus j}\rho_{ji}(\mathbf{H}^{(m)},\widetilde{\mathbf{V}}_{ki}^A,\mathbf{V}_{ki}^E,\sigma_{w,i}^2)$ and $\rho_{ji}(\mathbf{H}^{(m)},\widetilde{\mathbf{V}}_{ki}^A,\mathbf{V}_{ki}^E,\sigma_{w,i}^2)=\lambda_{ji}^{E\,(m)}\mathbf{u}_{ji}^{E\,(m)}\mathbf{V}_{ki}^E\mathbf{u}_{ji}^{(m)\,\dagger}(\mathbf{h}_{ji}^{A\,(m)}\widetilde{\mathbf{V}}_{ki}^A\mathbf{h}_{ji}^{A\,(m)\,\dagger}+\sigma_{w,i}^2\|\mathbf{h}_{ji}^{A\,(m)}\|^2)$. At the $n$th

iteration, $\mathbf{V}^{E\,(n)}$ and $\{R_j^{(n)}\}$ are obtained by solving the following optimization problem based on the SSUM method:

$$\begin{aligned}
&\underset{\mathbf{V}^E,\{R_j\}}{\arg\max} \sum_{j\in\mathcal{N}_M} R_j \\
&\text{s.t.} \quad R_j \le \frac{1}{n}\sum_{m=1}^{n} \widetilde{R}_j\left(\mathbf{h}^{(m)}, \mathbf{V}^E \,\middle|\, \mathbf{V}^{E(m-1)}, \widetilde{\mathbf{V}}^{A(m)}, \boldsymbol{\sigma}_w^{2\,(m)}\right) \\
&\qquad C_{w,i}\left(\widetilde{\mathbf{V}}_i^A(\mathbf{h}), \sigma_{w,i}^2(\mathbf{h})\right) \le \bar{C}_i - \sum_{j\in\mathcal{M}_i} R_j \\
&\qquad P_i\left(\widetilde{\mathbf{V}}_i^A(\mathbf{h}), \mathbf{V}_i^E, \sigma_{w,i}^2(\mathbf{h})\right) \le \bar{P}_i \qquad (31)
\end{aligned}$$

$\forall\, j \in \mathcal{N}_M$ and $\forall\, i \in \mathcal{N}_R$, until convergence.

## REFERENCES


[1] "C-RAN: the road towards green RAN," in *White Paper, ver. 2.5*, Beijing, China: China Mobile Res. Inst., Oct. 2011, pp. 1–48.

[2] A. Checko *et al.*, "Cloud RAN for mobile networks—a technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, 1st Quart. 2015.

[3] D. Samardzija, J. Pastalan, M. MacDonald, S. Walker, and R. Valenzuela, "Compressed transport of baseband signals in radio access networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3216–3225, Sep. 2012.

[4] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 69–79, Nov. 2014.

[5] U. Dotsch, M. Doll, H.-P. Mayer, F. Schaich, J. Segel, and P. Sehier, "Quantitative analysis of split base station processing and determination of advantageous architectures for LTE," *Bell Labs Tech. J.*, vol. 18, no. 1, pp. 105–128, Jun. 2013.

[6] D. Wubben *et al.*, "Benefits and impact of cloud computing on 5 G signal processing: Flexible centralization through cloud-RAN," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 35–44, Nov. 2014.

[7] J. Bartelt, P. Rost, D. Wubben, J. Lessmann, B. Melis, and G. Fettweis, "Fronthaul and backhaul requirements of flexibly centralized radio access networks," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 105–111, Oct. 2015.

[8] A. D. L. Oliva *et al.*, "Xhaul: Toward an integrated fronthaul/backhaul architecture in 5G networks," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 32–40, Oct. 2015.

[9] P.-H. Kuo, "New physical layer features of 3GPP LTE release-13 [Industry Perspectives]," *IEEE Wireless Commun.*, vol. 22, no. 4, pp. 4–5, Aug. 2015.

[10] Y.-H. Nam *et al.*, "Full-dimension MIMO (FD-MIMO) for next generation cellular technology," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 172–179, Jun. 2013.

[11] H. Murata *et al.*, "R&D activities for 5 G in IEICE technical committee on radio communication systems," in *Proc. IEEE 21st APCC*, Kyoto, Japan, Oct. 2015, pp. 250–256.

[12] G. Xu, Y. Li, Y.-H. Nam, C. Zhang, T. Kim, and J.-Y. Seol, "Full-dimension MIMO: Status and challenges in design and implementation," in *Proc. IEEE CTW*, Curaçcao, May 2014, pp. 1–21.

[13] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai, "Downlink multicell processing with limited-backhaul capacity," *EURASIP J. Adv. Signal Process.*, vol. 2009, Jun. 2009, Art. no. 840814.

[14] P. Marsch and G. Fettweis, "On downlink network MIMO under a constrained backhaul and imperfect channel knowledge," in *Proc. IEEE GLOBECOM*, Honolulu, HI, USA, Nov. 2009, pp. 1–6.

[15] P. Patil and W. Yu, "Hybrid compression and message-sharing strategy for the downlink cloud radio-access network," in *Proc. IEEE Inf. Theory Appl. Workshop*, San Diego, CA, USA, Feb. 2014, pp. 1–6.

[16] J. Kang, O. Simeone, J. Kang, and S. Shamai, "Fronthaul compression and precoding design for C-RANs over ergodic fading channel," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5022–5032, Jul. 2016.

[17] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3-D channel model," in *Proc. IEEE Int. Conf. Commun.*, Sydney, NSW, Australia, Jun. 2014, pp. 5865–5870.

[18] A. Alkhateeb, G. Leus, and R. W. Heath Jr., "Multi-layer precoding for full-dimensional massive MIMO systems," in *Proc. IEEE Asilomar Conf. Signal, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2014, pp. 815–819.

[19] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing-the large-scale array regime," *IEEE Trans. Info. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[20] B. Mondal *et al.*, "3D channel model in 3GPP," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 16–23, Mar. 2015.

[21] P.-H. Kuo, "A glance at FD-MIMO technologies for LTE," *IEEE Wireless Commun.*, vol. 23, no. 1, pp. 2–5, Feb. 2016.

[22] S. Park, C.-B. Chae, and S. Bahk, "Before/after precoded massive MIMO in cloud radio access networks," in *Proc. IEEE Int. Conf. Commun.*, Budapest, Hungary, Jun. 2013, pp. 169–173.

[23] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[24] N. Seifi, J. Zhang, R. W. Heath, Jr., T. Svensson, and M. Coldrey, "Coordinated 3-D beamforming for interference management in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5396–5410, Oct. 2014.

[25] Z. Zhong, X. Yin, X. Li, and X. Li, "Extension of ITU IMT-advanced channel models for elevation domains and line-of-sight scenarios," in *Proc. IEEE 78th VTC—Fall*, Las Vegas, NV, USA, Sep. 2013, pp. 1–5.

[26] A. Beck, and M. Teboulle, "Gradient-based algorithms with applications to signal recovery problems," in *Convex Optimization in Signal Processing and Communications*, Y. Eldar, and D. Palomar, Eds., Cambridge, U.K.: Cambridge Univ. Press, 2010, pp. 42–48.

[27] C. Suh and D. Tse, "Interference alignment for cellular networks," in *Proc. IEEE 46th Annu. Allerton Conf. Commun., Control, Comput.*, Urbana-Champaign, IL, USA, Sep. 2008, pp. 1037–1044.

[28] G. Thiagarajan and C. R. Murthy, "Novel transmit precoding methods for rayleigh fading multiuser TDD-MIMO systems with CSIT and no CSIR," *IEEE Trans. Veh. Technol.*, vol. 64, no. 3, pp. 973–984, Mar. 2015.

[29] T. M. Cover and J. A. Thomas, *Element of Information Theory*. New York, NY, USA: Wiley, 2006.

[30] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, Jul. 1996.

[31] M. Brookes, "The matrix reference manual," Imperial College, London, UK, 2011. [Online]. Available: http://www.ee.imperial.ac.uk/hp/staff/dmb/matrix/intro.html

[32] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks," *Math. Program.*, vol. 157, no. 2, pp. 515–545, Jun. 2016.

[33] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[35] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Inter-cluster design of precoding and fronthaul compression for cloud radio access networks," *IEEE Wireless Commun. Lett.*, vol. 3, no. 4, pp. 369–372, Aug. 2014.

[36] L. Vandenberghe and S. Boyd, "Semidefinite relaxation of quadratic optimization problems," *SIAM Rev.*, vol. 38, no. 1, pp. 49–95, 1996.



**Jinkyu Kang** received the B.Sc. degree in electrical communications engineering, the M.Sc. degree in electrical engineering, and the Ph.D. degree in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2009, 2011, and 2015, respectively.

From 2012 to 2013, he was a Visiting Scholar with New Jersey Institute of Technology, Newark, NJ, USA. From 2015 to 2016, he was a Postdoctoral Research Fellow with the Information and Electronics Research Institute, KAIST. Since then, he has been a Postdoctoral Research Fellow with the John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. His research interests include wireless communication, signal processing for cooperative communication, and information theory.

**Osvaldo Simeone** (F'16) received the M.Sc. degree (with honors) and the Ph.D. degree in information engineering from Politecnico di Milano, Milan, Italy, in 2001 and 2005, respectively.

He is currently with the Center for Wireless Communications and Signal Processing Research, Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ, USA, where he is currently an Associate Professor. His research interests include wireless communications, information theory, optimization, and machine learning.

Dr. Simeone currently serves as an Editor for IEEE TRANSACTIONS ON INFORMATION THEORY. He co-received the Best Paper Awards at the IEEE International Workshop on Signal Processing Advances in Wireless Communications and the IEEE Conference on Wireless Rural and Emergency Communications in 2007 and the IEEE Communication Society Best Tutorial Paper Award in 2015.

**Joonhyuk Kang** (M'15) received the B.S.E. and M.S.E. degrees from Seoul National University, Seoul, South Korea, in 1991 and 1993, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin, Austin, TX, USA, in 2002.

From 1993 to 1998, he was a research staff with Samsung Electronics, Suwon, South Korea, where he was involved in the development of digital-signal-processing-based real-time control systems. In 2000, he was with Cwill Telecommunications, Austin, where he participated in the project for multicarrier code-division multiple-access systems with antenna arrays. From 2008 to 2009, he was a Visiting Scholar with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. He is currently a faculty member with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea. His research interests include signal processing for cognitive radio, cooperative communication, physical-layer security, and wireless localization.

Dr. Kang received the Texas Telecommunication Consortium (TxTEC) Graduate Fellowship for 2000–2002. He is a member of the Korea Information and Communications Society and the Tau Beta Pi Engineering Honor Society.

**Shlomo Shamai (Shitz)** (F'94) received the B.Sc., M.Sc., and Ph.D. degrees from Technion—Israel Institute of Technology, in 1975, 1981, and 1986, respectively, all in electrical engineering.

From 1975 to 1985, he was a Senior Research Engineer with the Communications Research Labs. Since 1986, he has been with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is currently a Technion Distinguished Professor and holds the William Fondiller Chair of Telecommunications. His research interests include a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) served twice on the Board of Governors of the IEEE Information Theory Society. He has served as an Associate Editor for Shannon theory and on the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY. He received the Alon Grant for Distinguished Young Scientists in 1985, the Technion Henry Taub Prize for Excellence in Research in 2000, the van der Pol Gold Medal of the Union RadioScientifique Internationale (URSI) in 1999, the Claude E. Shannon Award in 2011, and the Rothschild Prize in Mathematics/Computer Sciences and Engineering in 2014. He was a corecipient of the IEEE Donald G. Fink Prize Paper Award in 2000, the Joint IEEE Information Technology and Communication Societies Paper Award in 2003 and 2004, the European Commission FP7 Network of Excellence in Wireless Communications (NEWCOM++, NEWCOM#) Best Paper Awards in 2009 and 2015, the Thomson Reuters Award for International Excellence in Scientific Research in 2010, the EURASIP Best Paper Award (for the *EURASIP Journal on Wireless Communications and Networking*) in 2014, and the IEEE Communications Society Best Tutorial Paper Award in 2015. He is a member of the Israeli Academy of Sciences and Humanities and a foreign member of the U.S. National Academy of Engineering.