Hybrid beamforming designs for 5G new radio with fronthaul compression and functional splits

Jing Li¹, Dianwu Yue¹, Ha H. Nguyen²

¹College of Information Science and Technology, Dalian Maritime University, Dalian 116026, People’s Republic of China
²Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada

Abstract: In this study, the authors investigate the intra-physical functional splits of 5G new radio protocol stack proposed by different groups. Based on the location of the digital beamforming block, the radio units (RUs) are divided into two categories: Category A and Category B. Two implementation modes of hybrid beamforming at the physical layer are considered, in which digital beamforming is performed either at the distributed unit (DU), as in the Category A RU based hybrid beamforming (HBF-A) scheme, or at the RUs, as in the Category B RU based hybrid beamforming (HBF-B) scheme. To maximise the weighted sum rate, the authors formulate the problems of jointly designing hybrid beamforming, analogue combining and fronthaul compression strategies for both HBF-A and HBF-B. The formulated problems are simplified by adopting the codebook-based design, and further tackled by leveraging the majorisation–minimisation algorithm. Finally, numerical results confirm that the HBF-A scheme outperforms the HBF-B scheme in the large power regime. Compared with the HBF-A method, the HBF-B method is more sensitive to changes in system parameters, such as the compression noise and the number of receive antennas, in the large power regime, while it is less sensitive in the small power regime.

1 Introduction

The cloud radio access network (C-RAN) architecture has been adopted in 4G wireless networks by operators in many countries and continues to be considered for 5G wireless networks. In the traditional C-RAN, a centralised baseband unit (BBU) pool is connected to remote radio units (RRUs) via highly reliable transport links, known as fronthaul. The Common Public Radio Interface (CPRI) specifications define how to transport the digitised radio frequency (RF) signals between the BBU and RRUs. In the 4G C-RAN networks, all the baseband processing functions of layers 1–3 reside within the BBU and the RRUs are confined to RF functions only. However, the traditional C-RAN architecture has a major challenge: this functional split leads to an extremely high demand on the capacity of the fronthaul link, which may become a bottleneck of network performance. Moreover, it has been pointed that wider frequency bandwidths available in both sub-6 GHz and millimetre wave (mmWave) bands in the 5G era as well as much larger antenna counts [thanks to the promising massive multiple-input–multiple-output (MIMO) technology] will put higher and higher demand for fronthaul bandwidth, eventually making it excessively large [1].

The initial focus of C-RAN is centralisation of BBU physical (PHY) elements, but the end game is cloud, or virtualisation, and 5G rollout is the primary driver to promote the evolution from the centralised RAN to a virtualised cloud RAN. Different from the traditional C-RAN architecture deployed in 4G, the 5G new radio (NR) RAN architecture consisting of interconnected radio base stations includes three functional entities: the centralised unit (CU), the distributed unit (DU) and the radio unit (RU), which can be deployed in multiple combinations. These splits create two transport segments: fronthaul (between DU and RU), and midhaul (between CU and DU). As for the BBU in the traditional C-RAN architecture, some PHY functions are migrated to the RU, while the real-time functions are deployed in the DU and the non-real-time functions are hosted centrally in the CU.

Recently, there have been many research works focusing on the fronthaul capacity problem in the C-RAN architecture. On the one hand, an evolved network, referred to as fog RAN (F-RAN), has been proposed with the goal of taking full advantages of both fog computing and C-RAN by equipping the edge devices with storing capabilities [2]. Unlike the centralised processing architecture of C-RAN, part of network functions in F-RAN are distributed to the network edge, which not only can reduce the transmission burden of fronthaul links, but also avoid large-scale radio signal processing in the BBU pool. On the other hand, flexible functional splits solve the excessive fronthaul bandwidth problem by placing in the RUs some PHY-layer functions traditionally located in the DUs. In order to alleviate the fronthaul capacity burden, the ideas of F-RAN and flexible functional split coincide with each other as both of them attempt to execute more functions locally. It is worth noting that some computing and control functions are even implemented at the intelligent user equipment (UE) rather than in the RU under the F-RAN architecture. By contrast, functional splits do not involve the processing in the UEs and address the issue from the perspective of the fundamental RAN architecture. Furthermore, there are other ways of solving the fronthaul capacity problem, but they are not related to the main focus of this paper.

The Third Generation Partnership Project (3GPP) has finalised the 5G NR specifications for making connections via multiple radio technologies in Release 15 [3]. It has defined two different frequency ranges: frequency bands from 450 MHz to 6 GHz (commonly known as sub-6 GHz bands) and frequency bands from 24.25 to 52.6 GHz (also known as mmWave bands). Compared to sub-6 GHz bands, signals in mmWave bands undergo more propagation loss and thus have smaller coverage. To solve this issue, C-RAN architecture can be utilised to deploy RUs close to the UEs. Meanwhile, benefiting from the short wavelength of mmWave signals, massive MIMO technology can be exploited to provide both spatial multiplexing and beamforming (BF) gains by packing large-scale antennas into small dimensions. In massive MIMO systems, conventional analogue BF can be conveniently realised by using phase shifters, which have constant amplitude constraints and may lead to a poor performance. By contrast, although fully-digital BF can control the phase and amplitude of the signal to achieve an optimal performance, it requires higher hardware cost and power consumption.

To balance the tradeoff between system performance and hardware cost/complexity, hybrid BF that combines analogue BF in the RF domain and digital BF in the baseband domain has been
proposed and extensively investigated [4–7]. The digital beamformer is designed to perform baseband processing and facilitate multi-stream processing, whereas the analogue beamformer is used to perform RF processing and obtain power gains via phase shifters. Determined by the mapping mode of RF chains and transmit antennas, three hybrid BF structures are commonly adopted: fully-connected structure as in [4], partially-connected structure as in [5, 6] and hybridly-connected structure as in [7]. Each RF chain is connected to all transmit antennas in the fully-connected structure, while each sub-array is connected to a single RF chain in the partially-connected structure. In the hybridly-connected structure, each sub-array is connected to multiple RF chains, and each RF chain is connected to all antennas with the sub-array in question. Compared with the fully-connected and partially-connected structure, the hybridly-connected structure is an attractive solution to balance the tradeoff between spectral efficiency and implementation cost.

In this paper, taking into account hybrid BF implementation and focusing on the downlink direction, we investigate the intra-PHY functional splits of 5G NR protocol stack proposed by different groups, and discuss the characteristics of diverse split options in detail. As the RUs can be classified into two categories depending on the location of digital BF, we propose two implementation modes of hybrid BF on the basis of the RU classification, and present simulation results to compare the performances of the proposed schemes. The main contributions of this paper are summarised as follows.

- Though there have been many research works on functional splits until now, the split options proposed by different groups are varying, and there are only a few split options taking hybrid BF into consideration. This paper summarises the intra-PHY split options proposed by 3GPP, the evolved CPRI (eCPRI) group and Open RAN (O-RAN) Alliance, of which the suggested functional splits affect how hybrid BF is implemented.
- In the traditional C-RAN using fully-digital BF, the RUs can be divided into two categories based on the location of the BF block. Similarly, depending on whether the digital BF function is included, we classify the RUs of the 5G NR RANs into two categories, namely Category A RU and Category B RU, and discuss their characteristics, advantages and disadvantages.
- Based on the classification of RUs, we propose two implementation modes of the hybrid BF: Category A RU based hybrid BF (HBF-A) scheme and Category B RU based hybrid BF (HBF-B) scheme. Then, for the two proposed schemes, we formulate the optimisation of jointly designing hybrid BF, analogue combining and fronthaul compression strategies to maximise the weighted sum rate (WSR). Finally, the formulated problems are simplified by codebook-based analogue BF/combiner design and solved with the majorisation-minimisation (MM) algorithm.
- Numerical results are provided to assess the system performance of the proposed HBF-A and HBF-B schemes. It is shown that the HBF-A scheme outperforms the HBF-B scheme in the large power regime. Moreover, in the latency-sensitive and dense-traffic networks with large power, compared with the HBF-A scheme, the HBF-B scheme can better balance the tradeoff among fronthaul bandwidth, transmit latency and system performance.
- To the authors’ best knowledge, although [8] is the first work exploring the classification of RU based on fully-digital BF and there are other works about hybrid BF in C-RAN architecture, this paper is the first to discuss hybrid BF implementations using different types of RUs in 5G NR RANs.

The remainder of this paper is organised as follows. Section 2 presents the locations of low layer functional splits in 5G NR RANs and introduces two types of RUs. In Section 3, corresponding to different types of RUs, two implementation modes of hybrid BF are studied. Numerical results are presented in Section 4. Section 5 concludes the paper.

Notation: Boldface uppercase and boldface lowercase indicate matrices and vectors, respectively. The notations \( A^\top \), \( A^\dagger \), \( | \cdot | \), \( \det ( \cdot ) \), \( \text{tr} ( \cdot ) \) and \( \| \cdot \| \) denote transpose, conjugate transpose, expectation, determinant, trace and the Euclidean norm, respectively. We adopt the standard information-theoretic definition for mutual information \( I(X; Y) \) between random variables \( X \) and \( Y \). A circularly symmetric complex Gaussian distribution with mean \( \mu \) and covariance matrix \( R \) is denoted by \( C \mathcal{N}(\mu, R) \). By \( \left[ \cdot \right]_i \) gives the \( i \)th entry of the matrix. The set of all \( M \times N \) complex matrices is defined as \( \mathbb{C}^{M \times N} \). \( I_d \) and \( \text{diag}(a_1, \ldots, a_d) \) stand for the identity matrix of size \( d \) and diagonal matrix with diagonal elements \( \{a_1, \ldots, a_d\} \). \( \otimes \) stands for the Kronecker product.

### 2 Low layer functional splits

#### 2.1 Different functional split options

3GPP has initially identified eight functional split options and several suboptions [9]. Meanwhile, the cCPRI group has proposed five possible functional splits with different balances between DU and RU to support different traffic patterns and business cases [10].

The functional split options can be broadly classified as either a high layer split between the packet data convergence protocol layer and radio link control layer, or a low layer split between the medium access control (MAC) layer and PHY layer. Here, we focus on the low layer split options, which may affect how hybrid BF is implemented.

Fig. 1 illustrates the locations of functional split options recommended by 3GPP, eCPRI group and O-RAN Alliance in the PHY layer for 5G NR RANs regarding the downlink direction, in which the lines present different split options. The functions below the line will be accomplished in the RU while the functions above the line will be performed in the DU. The more functions located in the RU, the lower the bit rate on the fronthaul link. Concerning these functional blocks, Fig. 1 just gives one possible implementation in the PHY layer which is considered solely for the purpose of study, and the actual implementation may be different due to various reasons. Moreover, Fig. 1 only focuses on the data channels.

It can be seen from Fig. 1 that, in the downlink process, the functions in the PHY layer will transform the transport blocks received from the MAC layer into in-phase and quadrature (IQ) signals ready for the RF block [11]. To be specific, the users' bit sequences received from the MAC layer go through coding, scrambling, modulation, layer mapping, precoding and resource element mapping, resulting in IQ sample sequences of orthogonal frequency division multiplexing (OFDM) signals in the frequency domain. Then, the IQ sample sequences undergo inverse fast
Fourier transform (IFFT) where they are converted to OFDM symbols in the time domain. The cyclic prefix (CP) is added to distinguish the frames from another. Finally, the symbols are converted to analogue signals in the RF block. Digital BF is accomplished before IFFT processing while analogue BF is performed after analogue signal conversion. Digital BF can be implemented by a single block along the function chain, or it may be spread across different functions of the PHY layer [12]. Here, we adopt the former. The preceding block in Fig. 1 is only designed for antenna port mapping as discussed in [13], and the digital and analogue BF blocks are truly responsible for modifying the amplitudes and phases of antenna arrays to offer directional BF.

Different functional split options in the PHY/RF layer are highlighted in Fig. 1. The characteristics of each split option are summarised below.

- **Option 8:** 3GPP split option 8 (also called split E by eCPRI group) is what has already been introduced as a traditional BBU-RU split in the C-RAN architecture. It has been known for several years and supported in current long term evolution (LTE) networks. All the baseband processing functions reside in the DU and only the RF sampler and up-converter reside in the RU, which results in the highly centralised DU and minimally functional RU. The separation of RF and PHY layers makes the reuse of RF components to serve the PHY layer of different radio access technologies (e.g. Global System for Mobile, 3G and LTE) possible. As mentioned in Section 1, this split option could be bandwidth-intensive.

- **Option 7-1:** This option is specified by 3GPP. IFFT, CP addition and RF functions are located in the RU, and the rest of PHY functions are left in the DU. Similar to options 7-2x and 7-2, this split can reduce the fronthaul requirement in terms of throughput and realise centralised scheduling and joint processing (both transmit and receive).

- **Option 7-2x:** In this split, IFFT, CP addition and digital BF functions are located in the RU, and the rest of PHY functions are placed in the DU. Here, ‘x’ refers to xRAN, which is an industry organisation pursuing standardising and promoting a software-based extensible RAN [14]. In February 2018, the xRAN Forum was merged with C-RAN Alliance into the O-RAN Alliance by operators to evolve RANs to be more open and smarter than previous generations [15, 16]. As the fronthaul specifications of xRAN are inherited and maintained by O-RAN Alliance now, option 7-2x is proposed to be the functional split option for O-RAN architecture [17]. Though this split is not specified by 3GPP, it attracts interest from 5G-PICTURE Project [18], where it is called 7-2a, and the industry as a whole.

- **Split IIp:** The eCPRI specification recommends this split to separate eCPRI radio equipment control and eCPRI radio equipment, which are known as DU and RU in 3GPP terminology, respectively. Since the eCPRI interface operates in the frequency domain for this split, it is more bandwidth-efficient than the CPRI protocol using the traditional split (i.e. option 8) working in the time domain.

- **Option 7-2:** This option is specified by 3GPP. Compared with option 7-2x, it further moves the precoding and resource element mapping blocks to the RU, which indicates that more hardware resources essential for baseband processing are required at the RUs.

- **Option 7-3:** For 3GPP split option 7-3 (also called split I3P by the eCPRI group), only the encoder resides in the DU and the rest of PHY functions reside in the RU. The payload for this option is encoded data, which leads to a significant lower bit rate on the fronthaul link. Unlike the split IIp, in which the data is IQ symbol oriented, the data in this split is bit oriented [10].

When designing the functional split options for practical usage scenarios, we should consider the tradeoff between the degree of centralisation and the fronthaul bandwidth/latency. Since part, or all of the PHY processing is centralised in the DU, low layer splits (somewhere in the RAN protocol stack PHY layer) can simplify the RU and obtain significant pooling gains, though they also introduce long latency, high power consumption and stringent capacity constraints on the transport network. In other words, low layer splits (e.g. split options 8, 7-1 and 7-2x) can bring more flexible tradeoffs than high layer splits.

### 2.2 Two categories of RUs

Though the 5G White Paper published by the Next Generation Mobile Networks Alliance [8] has discussed the possibility of moving baseband processing to RU, it adopts fully-digital BF instead of the hybrid scheme. It is expected that hybrid BF will be incorporated into 5G NR base stations, which integrate mmWave bands, large-scale antenna arrays and C-RAN architecture, to further improve the system performance. Some research works on the new generation RAN have studied hybrid solutions and described feasible locations of digital and analogue BF blocks in the implementations of PHY layer [14, 17, 19, 20]. The digital BF and analogue BF can be performed in different physical units corresponding to the functional splits indicated in Fig. 1. For options 7-1 and 8, digital BF and analogue BF are separately executed in the DU and RU, respectively. Nevertheless, for option 7-2x, split IIp and other higher layer splits, both digital BF and analogue BF are implemented in the RU. Therefore, depending on whether the digital BF function is included, the RUs can be divided into two categories, called Category A and Category B.

#### 2.2.1 Category A

As the digital BF block resides in the DU, Category A devices do not have the digital BF function, which means that the implementation of RU is simple. The inter-RU interference can be alleviated by optimising the digital beamformer. The previous works on hybrid BF algorithms of the traditional C-RAN architecture as in [21–25] (where option 8 is the default functional split point) still apply to the 5G NR networks using Category A RU. When the C-RAN system employs fully-digital BF, the fronthaul interface suffers high traffic as the fronthaul bandwidth requirement scales on the number of transmit antennas, no matter whether the split option 8 or 7-1 is chosen. However, when Category A RUs are deployed in the system, this relationship no longer holds since the number of transmit antennas is not equal to the number of RF chains.

#### 2.2.2 Category B

Category B devices include both digital and analogue BF functions. The benefit is that the fronthaul bandwidth requirement should be relatively low since it carries layers or bits, which allows using a large number of antennas without requiring high transport bandwidth. The challenge is that the RU design is more complex compared with Category A RU. If each RU calculates BF matrices separately, there will be inter-RU interferences. However, if the DU takes the responsibility for the design of BF matrices and the RUs perform BF with the information provided by the DU, the inter-RU interferences can also be eliminated [26]. In this way, the RUs do not need to handle the computational work, and hence, the RUs can be as simple as possible. But since part of the baseband processing is migrated to the RUs, the cost of Category B RU is still higher than that of Category A RU. The Category B RU has a high scalability as it can support different BF techniques (e.g. digital, analogue and hybrid) as well as different BF algorithms [27]. Moreover, as both the digital BF and analogue BF reside in the Category B RU, no matter which split option is used (e.g. option 7-2x, split IIp), there is no need to think about the compression loss and bandwidth requirement in the design of the beamformers.

Two types of RUs are suitable for different use cases, which have specified requirements on latency and data transmit rate. Category A RU fits low frequency and latency-insensitive services, while Category B RU applies to high frequency and dense traffic networks. To meet the service demands, both categories of RUs are likely to coexist in the system. A survey recently conducted by Heavy Reading [28] reveals that many operators around the world have clear interests in moving towards open interoperability. For PHY layer implementation in the RAN, this requires not only the open fronthaul interfaces to enable multi-vendor DU-RU interoperability, but also the scalability of DU to support both...
categories of RUs from different vendors. Moreover, though all of the aforementioned split options are feasible depending on the particular radio configuration, converging on one or two split options is necessary to facilitate the implementation of interoperability that the O-RAN Alliance promotes [16].

3 Hybrid BF implementation in 5G NR RANs

In this section, based on the classification of RUs discussed in Section 2, we propose two implementation modes of hybrid BF and formulate the corresponding problems of jointly designing hybrid BF, analogue combining and fronthaul compression strategies to maximise the WSR. The formulated problems are then simplified and solved with existing algorithms. At the end, we compare the proposed schemes of this paper with other existing works.

3.1 System model

We consider a downlink 5G NR RAN system in which $N_R$ RUs provide services for $N_U$ UEs, as shown in Fig. 2. The RUs are controlled by a DU connected with each RU via a limited-capacity fronthaul link. The $i$th RU has $N_{RF,i}$ transmit antennas and $N_{RF,i}$ RF chains satisfying $N_{RF,i} \leq N_{T,i}$. The $j$th UE has $N_{R,j}$ receive antennas and $N_{URF,j}$ RF chains. For convenience, we define the sets $N_R \triangleq \{1, 2, \ldots, N_R\}$, $N_U \triangleq \{1, 2, \ldots, N_U\}$, $N_{RF,i} \triangleq \{1, 2, \ldots, N_{RF,i}\}$, and $N_{URF,j} \triangleq \{1, 2, \ldots, N_{URF,j}\}$. The total numbers of transmit antennas, receive antennas and RF chains are denoted as $N_T = \sum_{i=1}^{N_R} N_{T,i}$, $N_R = \sum_{j=1}^{N_U} N_{R,j}$ and $N_{RF} = \sum_{i=1}^{N_R} N_{RF,i}$, respectively. We assume perfect channel state information (CSI) at the DU and $N_{T,j} = N_T$ for $j \in N_R$.

The signal received by the $j$th UE is given by

$$x_j = H_j x + z_j, \quad (1)$$

where $z_j \sim \mathcal{CN}(0, \sigma_z^2 I)$ is the complex additive white Gaussian noise vector; $H_j = [h_{1,j}, h_{2,j}, \ldots, h_{N_R,j}]$ denotes the $N_{R,j} \times N_T$ channel matrix for the $j$th UE, where $h_{i,j} \in \mathbb{C}^{N_{R,j} \times N_T}$ represents the channel submatrix from the $i$th RU to the $j$th UE; and $x = [x_1^T, x_2^T, \ldots, x_{N_U}^T]^T$ is the aggregate transmit signal vector, in which the transmit signal $x_i$ has a power constraint given as $\mathbb{E}[\|x_i\|^2] \leq P_i$. The processed baseband signal can be obtained as $y_j = W_j^H x_j$, where $W_j \in \mathbb{C}^{N_{RF,j} \times N_{URF,j}}$ is the analogue combiner of the $j$th UE. The overall analogue combiner in the UE side can be expressed as $\tilde{W} = \text{diag}\{W_1, W_2, \ldots, W_{N_U}\}$. We assume that $W_j$ is implemented using analogue phase shifters, and thus its entries are of constant modulus, i.e. $\|W_j\|_\infty = 1$ for $j \in N_U$, $p \in N_{R,j}$ and $q \in N_{URF,j}$.

3.2 Hybrid BF based on category A RU

Fig. 3 illustrates the block diagram of Category A RU based hybrid BF scheme, in which the DU computes both the digital and analogue BF matrices and performs digital BF. The analogue BF matrices and resulting baseband signals are separately compressed before they are forwarded via the fronthaul links to the corresponding RUs. Finally, the information bits received at each RU are decompressed and precoded with the analogue BF matrix.

We represent the data stream as $s = [s_1^H, s_2^H, \ldots, s_{N_{RF}}^H]^H$, where $s_i \in \mathbb{C}^{N_{R,j} \times 1}$ is the signal intended for the $j$th UE. Without loss of generality, the signals are assumed to be i.i.d. Gaussian random variables with zero mean and unit variance. The total number of data streams is $N_S = \sum_{i=1}^{N_{RF}} N_{U,i}$, which satisfies $N_{U,j} \leq N_{R,j}$ and $N_S \leq N_{RF}$. The signal after performing digital BF by the DU can be expressed as $\tilde{y} = [\tilde{y}_1^H, \tilde{y}_2^H, \ldots, \tilde{y}_{N_{RF}}^H]^H$, where $\tilde{y}_i$ corresponds to the $i$th RU for $i \in N_R$. We have $\tilde{y} = F_{\text{BB}} s$, where $F_{\text{BB}} \in \mathbb{C}^{N_{RF} \times N_S}$ is the digital BF matrix. Specifically, the signal $\tilde{y}_i$ for the $i$th RU can be defined as $\tilde{y}_i = F_{\text{BB},i}^\text{row} s_i$, where the digital BF submatrix for the $i$th RU $F_{\text{BB},i}^\text{row} \in \mathbb{C}^{N_{RF} \times N_S}$ is obtained by properly selecting the rows of matrix $F_{\text{BB}}$. Specifically, the matrix $F_{\text{BB},i}^\text{row}$ is given by $F_{\text{BB},i}^\text{row} = D_i^H F_{\text{BB}}$, with the $N_{RF} \times N_{RF,i}$ matrix $D_i$, having all zero entries except for the rows from $\sum_{k=1}^{i-1} N_{RF,k} + 1$ to $\sum_{k=1}^{i} N_{RF,k}$, which contain an $N_{RF,i} \times N_{RF,i}$ identity matrix. The DU quantises each sequence of the processed baseband signal $\tilde{y}_i$ before transmitting the quantised signal to the $i$th RU on the $i$th fronthaul link. The quantised baseband signal $y_i$ can be written as

$$y_i = \tilde{y}_i + q_i, \quad (2)$$
where the quantisation noise vector $q_{RF,i}$ is assumed to have i.i.d. $\mathcal{CN}(0, \sigma_{RF,i}^2)$ entries, which are independent across the RU index $i$. The covariance of the quantisation noise can be expressed as $\Omega_q = \mathbb{E}[q_{RF,i}q_{RF,i}^H] = \text{diag}[\sigma_{RF,1}^2, \ldots, \sigma_{RF,N_{RF}}^2]$. The analogue BF matrix $\tilde{F}_{RF}$ is block-diagonal, i.e. $\tilde{F}_{RF} = \text{diag}(\tilde{F}_{RF,1}, \ldots, \tilde{F}_{RF,N_{RF}})$, where $\tilde{F}_{RF,i} \in \mathbb{C}^{N_T \times N_{RF,i}}$ is the analogue BF submatrix intended for the $i$th RU. It has a similar constraint as with $W_j$, i.e. $\left|\tilde{F}_{RF,i}^{[n,m]}\right|^2 = 1$ for $m \in N_T,i$ and $n \in N_{RF,i}$. The matrix $\tilde{F}_{RF,i}$ is compressed by the DU and the compressed matrix is forwarded to the $i$th RU over the fronthaul link. The compressed matrix can be represented as $F_{RF,i} = \tilde{F}_{RF,i} + Q_{RF,i}$, where the quantisation noise matrix $Q_{RF,i}$ is assumed to have i.i.d. $\mathcal{CN}(0, \sigma_{RF,i}^2)$ entries and they are independent across the index $i$. The covariance of the analogue BF matrix is $\Omega_{RF} = \text{diag}[\sigma_{RF,1}^2, \ldots, \sigma_{RF,N_{RF}}^2]$. Overall, the compressed analogue BF matrix $F_{RF}$ can be represented as $F_{RF} = \tilde{F}_{RF} + Q_{RF}$. To facilitate future analysis, we define $\Omega_{rf} = \text{diag}[\sigma_{RF,1}^2, \ldots, \sigma_{RF,N_{RF}}^2]$. It follows that the design of fronthaul compression reduces to the optimisation of the quantisation noise variances $\sigma_{RF,i}$ for $i \in N_{RF}$.

Fig. 4 shows the analogue BF structure of RU $i$ in the HBF-A scheme. It follows from the above discussion that the compressed analogue BF matrix $F_{RF}$ for the HBF-A scheme is also block-diagonal. Although the RF chains are connected to all antennas in a RU, when referring to the BF structure, we should also take the digital part into consideration. Hence, when all RUs are taken into account, from the system point of view, the BF structure of HBF-A is hybridly-connected, instead of fully-connected.

Since the signal transmitted by the $i$th RU can be expressed as $x_i = F_{RF,i}y$, the transmit power of the $i$th RU can be computed as

$$P(F_{BB}, \tilde{F}_{RF,i}, \sigma_{RF,i}, \sigma_{RF}) = \mathbb{E}[\|x_i\|^2] = \text{tr}(\hat{F}_{RF,i}^H F_{BB} F_{BB}^H D \tilde{F}_{RF,i} + \sigma_{RF,i}^2 \tilde{F}_{RF,i}^H \tilde{F}_{RF,i}) + \sigma_{RF,i}^2 \tilde{F}_{RF,i}^H \tilde{F}_{RF,i}$$

The above expression indicates that the transmit power of the $i$th RU depends on the BF matrices $F_{BB}$ and $F_{RF,i}$, as well as noise variances $\sigma_{RF,i}^2$ and $\sigma_{RF}^2$.

The $i$th fronthaul link with capacity $C_i$ should have a limited data rate. The rate required to transmit the analogue BF matrix on the $i$th fronthaul link is equal to

$$C_i(F_{RF,i}, \sigma_{RF,i}) = \log \det(\tilde{F}_{RF,i}^H F_{RF,i} + \sigma_{RF,i}^2 I) - N_T \log(\sigma_{RF})$$  

As such, the rate required to transmit the precoded baseband signal on the $i$th fronthaul link is determined as

$$C_i(F_{BB}, \sigma_{RF,i}) = \log \det(D_i^H F_{BB} F_{BB}^H D_i + \sigma_{RF,i}^2 I) - N_T \log(\sigma_{RF})$$  

We assume that $\sigma_{RF} = \text{diag}([\sigma_{RF,1}, \ldots, \sigma_{RF,N_{RF}}] \otimes I_{N_T})$ and $\sigma_{BB} = \text{diag}([\sigma_{BB,1}, \ldots, \sigma_{BB,N_{RF}}]^T \otimes I_{N_T})$. The achievable ergodic rate for the $i$th UE can be written as $\mathbb{E}[R_i^A(W_j, F_{BB}, \tilde{F}_{RF,i}, \sigma_{RF}, \sigma_{BB})]$, where $R_i^A(W_j, F_{BB}, \tilde{F}_{RF,i}, \sigma_{RF}, \sigma_{BB}) = I(s_j; s_i')$ is expressed as (7) where

$$\Xi_i^A = \text{diag}([D_i^H F_{BB} F_{BB}^H D_i], \ldots, D_i^H F_{BB} F_{BB}^H D_i)$$  

Instead of considering the instantaneous data rate, we formulate the design problem by considering the ergodic WSR, which is an important metric to evaluate the network performance. The ergodic sum rate can be obtained by averaging over all channel statistics and it has been widely used in the researches on C-RAN systems [22, 23, 29–31]. In this paper, we are interested in maximising the ergodic WSR $R_{sum}^A = \sum_{i \in N_i} \mathbb{E}[R_i^A]$ subject to constraints on per-RU transmit power, fronthaul capacity and constant modulus of analogue BF matrices and combiners. Here $\mu_i > 0$ is a weight expressing the priority for the $i$th UE. In general, introducing the weighting factor into the ergodic sum rate allows us to prioritise different UEs in the system and optimise resource allocation according to different fairness principles, e.g. max–min fairness and proportional fairness [32]. Therefore, the optimisation problem can be formulated as

$$\max_{(W_j), F_{BB}, \tilde{F}_{RF,i}, \sigma_{RF,i}, \sigma_{BB}} \sum_{i \in N_i} \mathbb{E}[R_i^A]$$

s.t.

$$C_i(F_{RF,i}, \sigma_{RF,i}) + C_i(F_{BB}, \sigma_{BB}) \leq \hat{C}_i, i \in N_{RF},$$

$$P_i(F_{BB}, \tilde{F}_{RF,i}, \sigma_{RF,i}, \sigma_{BB}) \leq P_i, i \in N_{RF},$$

$$\|\tilde{F}_{RF,i}\|_F^2 = 1, i \in N_{RF}, m \in N_T, n \in N_{RF,i},$$

$$\|W_j\|_F^2 = 1, j \in N_j, p \in N_{RF,j}, q \in N_{URF,j}.$$  

Conditions (10b), (10c), (10d) and (10e) correspond to the fronthaul capacity constraint on account of multivariate compression, per-RU transmit power constraint and constant modulus constraints of analogue BF matrix and analogue combiner, respectively. The optimisation problem in (10a)–(10e) is quite challenging due to the non-convexity of objective function and constraints. Moreover, the existence of the Kronecker product in (7) makes it even more complex.

Due to practical constraints on RF hardware, the RF phase shifters only have quantised angles. Hence, the analogue BF and combining vectors can only take on certain values, i.e. these vectors need to be selected from finite-size codebooks. To date,
there have been many research works on codebook-based analogue BF combining design [33–35], which inspires us to leverage codebooks to simplify the design problem in (10a)–(10e). When \( N_{i} = N_{j} \) for \( i \in \mathcal{N}_R \) and \( N_{i} = N_{i,j} \) for \( j \in \mathcal{N}_U \), if we neglect the effect of compression noise for the analogue BF matrix, i.e. use the codebook-based analogue beamformer design, the problem (10a)–(10e) is equivalent to joint optimisation of digital and analogue BF matrices and the fronthaul compression strategies as discussed in [21]. With such an approach, the problem is then solved by a block coordinate descent method proposed based on the weighted minimum-mean-square-error approach to relax the constant modulus constraint of the analogue beamformer.

In the codebook-based BF design, the DU and RUs are supposed to have some predefined common codebooks, and each element of the codebook can be indexed as a BF vector index. The DU sends BF vector indices to RUs, which perform analogue BF using the corresponding BF vectors. In contrast to transmitting BF matrices, the overhead of transmitting BF vector indices to RUs is very small and negligible. As such, we can get \( C_1(F_{BB}, \sigma_{BB}, \sigma_{FR}) = 0 \) and \( P_1(F_{BB}, \sigma_{BB}, \sigma_{FR}) = tr(F_{BB}D_{BB}^{HH} + \sigma_{BB})P_{BB}^{-1} \). We define \( \mathcal{F} \) and \( \mathcal{W} \) as the analogue BF vector codebook and the analogue combiner vector codebook, respectively. The achievable rate for the \( j \)th UE in (7) can be simplified as (see (11))

Finally, the problem (10a)–(10e) can be rewritten as

\[
\begin{align*}
\max_{\{W_{q}\}, F_{BB}, \sigma_{FR}, \sigma_{BB}} & \sum_{j \in \mathcal{N}_U} \mu_{j} \mathbb{E}[R_{j}^{B}(W_{j}, F_{BB}, F_{RF}, \sigma_{j})] \\
\text{s.t.} \quad & C_{1}(F_{BB}, \sigma_{BB}) \leq \bar{C}_{i}, \quad i \in \mathcal{N}_R, \quad (12a) \\
& P_{1}(F_{BB}, \sigma_{BB}) \leq \bar{P}_{i}, \quad i \in \mathcal{N}_R, \quad (12b) \\
& [F_{RF}]_{i,n} \in \mathcal{F}, \quad i \in \mathcal{N}_R, \quad n \in \mathcal{N}_{RF}, \quad (12c) \\
& [W_{j}]_{q} \in \mathcal{W}, \quad j \in \mathcal{N}_U, \quad q \in \mathcal{N}_{URF}, \quad (12d)
\end{align*}
\]

Since the sizes of analogue BF and combiner codebooks are finite, the problem (12a)–(12e) can be solved by fixing the analogue BF matrix \( F_{BB} \) and the analogue combiner \( W \) to optimise digital processing strategies \( [F_{RF}, \sigma_{BB}] \). When the RF beamformer and the analogue combiner are fixed, the functions \( R_{j}^{B}(W_{j}, F_{BB}, F_{RF}, \sigma_{j}) \) and \( C_{1}(F_{BB}, \sigma_{BB}) \) can be seen as the difference of convex functions of the covariance matrices \( F_{BB}F_{BB}^{HH} \) for \( j \in \mathcal{N}_U \) and variance \( \sigma_{BB}^2 \). The problem (12a)–(12e) can be further simplified as

\[
\begin{align*}
\max_{\{W_{q}\}, F_{BB}, \sigma_{FR}, \sigma_{BB}} & \sum_{j \in \mathcal{N}_U} \mu_{j} \mathbb{E}[R_{j}^{B}(W_{j}, F_{BB}, F_{RF}, \sigma_{j})] \\
\text{s.t.} \quad & C_{1}(F_{BB}, \sigma_{BB}) \leq \bar{C}_{i}, \quad i \in \mathcal{N}_R, \quad (13a) \\
& P_{1}(F_{BB}, \sigma_{BB}) \leq \bar{P}_{i}, \quad i \in \mathcal{N}_R, \quad (13b) \\
& [F_{RF}]_{i,n} \in \mathcal{F}, \quad i \in \mathcal{N}_R, \quad n \in \mathcal{N}_{RF}, \quad (13c)
\end{align*}
\]

Recognising that the problem (13a)–(13c) becomes very similar to the problem (9) in [29], the digital BF matrix \( F_{BB} \) and fronthaul compression strategy \( \sigma_{BB}^2 \) can be obtained by using the MM algorithm to obtain a feasible solution. However, the solution is only optimal for the problem (13a)–(13c) rather than for the problem (12a)–(12e), since it is obtained by supposing that the analogue BF matrix \( F_{BB} \) and analogue combiner \( W \) are known. This means that we need to try all possible implementations of the analogue beamformer and analogue combiner in order to select the globally optimal solution of the problem (12a)–(12e).

### 3.3 Hybrid BF based on Category B RU

This section describes the hybrid BF scheme based on Category B RU. Fig. 5 depicts the implementation block diagram, in which the DU combines both the analogue and digital BF matrices and utilises fronthaul links to communicate the messages of a given subset of UEs to each RU, along with the corresponding compressed BF matrices. Each RU decompresses the BF matrices received over the fronthaul link and precodes the messages of the given UEs. The preliminary clustering step aims to assign each RU with a subset of UEs. This step is performed in the DU, and the results are sent to the RUs via fronthaul.

According to the structure of the digital BF matrix \( \tilde{F}_{BB} \), the set of UEs served by the \( k \)th RU can be denoted as \( \mathcal{U}_k \subseteq \mathcal{N}_U \) for \( i \in \mathcal{N}_R \), which indicates that the \( k \)th RU only processes the signals intended for UEs in the set \( \mathcal{U}_k \). We use the notation \( \mathcal{U}_k(k) \) to denote the \( k \)th UE in the set \( \mathcal{U}_k \). We define the total number of data streams for the \( k \)th RU as \( N_{k} \), and assume that the set of UEs assigned to the \( k \)th RU is given and not subjected to optimisation.

The digital BF matrix \( F_{BB} \) is constrained to have zeros in the positions corresponding to RU–UE pairs such that the UE is not served by the given RU. This constraint can be formulated as

\[
\tilde{F}_{BB} = [F_{BB}^{row_1}, \ldots, F_{BB}^{row_{N_{k}}}]^{H},
\]

where

\[
\begin{align*}
R_{j}^{B}(W_{j}, F_{BB}, F_{RF}, \sigma_{j}) &= \log \det \left( I + W_{j}^{HH}F_{RF} \left( F_{BB}F_{BB}^{HH} + \Omega_{j} \right) F_{BB}^{HH}W_{j} \right) \\
& \quad - \log \det \left( I + W_{j}^{HH}F_{RF} \left( \sum_{i \in \mathcal{N}_U \setminus j} F_{BB,i}F_{BB,i}^{HH} + \Omega_{j} \right) F_{BB}^{HH}W_{j} \right).
\end{align*}
\]
where $F_{BB, i}^{row}$ is the $N_{RF} \times N_u$ digital BF submatrix intended for the $i$th RU, and the $N_u \times N_e$ constant matrix $E_{RF}^{row}$ (having only 0 or 1 entries) defines the association between the data streams and RF chains. The matrix $E_{RF}^{row}$ can be given by $E_{RF}^{row} = [D_1^T, D_2^T, \ldots, D_n^T]^T$, where the $N_u \times N_e$ matrix $D_i$ has all zero elements except for the columns from $\sum_{n=1}^{i-1} N_u + 1$ to $\sum_{n=1}^{i} N_u$ containing an $N_u \times N_e$ identity matrix. The sequence of digital BF matrix $F_{BB, i}^{row}$ is compressed by the DU and the compressed matrix is forwarded over the fronthaul link to the $i$th RU. The compressed matrix $F_{BB, i}^{row}$ for the $i$th RU is then given by

$$F_{BB, i}^{row} = \tilde{F}_{BB, i}^{row} + Q_{BB, i},$$

where the $N_{RF} \times N_u$ quantisation noise matrix $Q_{BB, i}$ is assumed to have i.i.d. $\mathcal{C}(0, \sigma_{BB}^2)$ entries, which are independent across index $i$. For the following analysis, we define $\Omega_{BB} = \text{diag}\{\sigma_{BB}^2, \sigma_{BB}^2, \ldots, \sigma_{BB}^2, \sigma_{BB}^2, I\}$. We also have $\Omega_{BB} = E[\text{vec}(Q_{BB})\text{vec}(Q_{BB})^H] = \text{diag}\{\sigma_{BB}^2, \ldots, \sigma_{BB}^2, I\}$.

For the HBF-B scheme, the analogue BF structure is the same as that of the HBF-A scheme, as shown in Fig. 4. It can be verified that the compressed analogue BF matrix $F_{RF}$ is also block-diagonal. Since the digital BF is performed in a RU, from the viewpoint of a specific RU, the BF structure of the HBF-B scheme can be seen as fully connected, which is different from that of the HBF-A scheme.

The power transmitted by the $i$th RU can be calculated as

$$P_i(\tilde{F}_{BB, i}^{row}, \tilde{F}_{RF, i}^{row}, \sigma_{BB, i}^2, \sigma_{RF, i}^2) = \text{tr}(\tilde{F}_{RF, i}^{row}\tilde{F}_{BB, i}^{row}\tilde{F}_{RF, i}^{row}\tilde{F}_{BB, i}^{row} + \sigma_{BB, i}^2\tilde{F}_{RF, i}^{row}\tilde{F}_{RF, i}^{row} + \sigma_{RF, i}^2\tilde{F}_{BB, i}^{row}\tilde{F}_{BB, i}^{row} + \sigma_{BB, i}^2\sigma_{RF, i}^2I).$$

It is assumed that the $N_e \times N_u$ matrix $V_j$ has all zero elements except for the rows from $\sum_{i=1}^{j-1} N_e + 1$ to $\sum_{i=1}^{j} N_e$, which contain an $N_e \times N_u$ identity matrix. The achievable ergodic rate for the $j$th UE can be written as $E[R_{BB}^B(W_j, \tilde{F}_{BB}, \tilde{F}_{RF}, \sigma_{BB, i}^2, \sigma_{RF, i}^2)]$, where $\sigma_{BB}^2 = [\sigma_{BB, 1}^2, \ldots, \sigma_{BB, N_u}^2]^T$ and $R_{BB}^B(W_j, \tilde{F}_{BB}, \tilde{F}_{RF}, \sigma_{BB, i}^2, \sigma_{RF, i}^2) = I(s_j, s_j')$ is calculated as (see (18)), where (see (19))

$$R_{BB}^B(W_j, \tilde{F}_{BB}, \tilde{F}_{RF}, \sigma_{BB, i}^2, \sigma_{RF, i}^2) = \log \det \left( I + W_j^H H \left( \tilde{F}_{RF, i}^{row} \tilde{F}_{BB, i}^{row} + \Omega_{BB} \right) \tilde{F}_{RF, i}^{row} + \Omega_{RF} \left( \Omega_{BB} \otimes I_{N_e} \right) \right) W_j \right) \text{and}
- \log \det \left( I + W_j^H H \left( \tilde{F}_{RF, i}^{row} \sum_{k \in \mathcal{A}, j'} \tilde{F}_{BB, k}^{row} \tilde{F}_{BB, j'}^{row} + \Omega_{BB} \right) \tilde{F}_{RF, i}^{row} + \Omega_{RF} \left( \Omega_{BB} \otimes I_{N_e} \right) \right) W_j \right).$$

$$E_{BB}^B = \text{diag} \{ \text{tr} \left[ F_{BB, \eta_1}^{row} E_{BB, \eta_1}^{row} V_{BB, \eta_1}^{row} H_{BB, \eta_1}^{row} \right], \ldots, \text{tr} \left[ F_{BB, \eta_N}^{row} E_{BB, \eta_N}^{row} V_{BB, \eta_N}^{row} H_{BB, \eta_N}^{row} \right] \}.$$
by fixing the analogue BF matrix and analogue combiner to optimise digital processing strategies with the MM algorithm.

Before closing this section, it is pointed out that there is another way of implementing the HBF-B scheme, in which the DU is responsible for computing a composite BF matrix, separately compressing the submatrices for RUs and transmitting the compressed submatrices via the fronthaul link. Finally, in each RU, the analogue and digital beamformers are calculated by leveraging the received composite submatrix. The composite BF matrix can be designed using the CBP method, and the analogue and digital beamformers can be further obtained through a matrix factorisation based on the near-optimal design in [7]. Unfortunately, such an approach may cause a certain performance loss in the process of generating the analogue and digital beamformers, which will lead to a performance gap between HBF-B and CBP.

### 3.4 Comparison with existing works

The HBF-A approach is similar to the compress-after-preceding (CAP) strategy envisioned for C-RANs in [29, 36], while the HBF-B approach may cause a certain performance loss in the process of compressing the submatrices for RUs and transmitting the compressed submatrices via the fronthaul link. Unfortunately, such an approach may cause a certain performance loss in the process of generating the analogue and digital beamformers, which will lead to a performance gap between HBF-B and CBP.

![Fig 6 Setup under consideration for the numerical results in Section 4, where the RUs and UEs are independently and uniformly distributed in a circular region](image)

**Fig. 6** Setup under consideration for the numerical results in Section 4, where the RUs and UEs are independently and uniformly distributed in a circular region.

4 Performance analysis

There is a simpler approach to design the BF, combining and compression strategies [37]. First, the BF matrices and combiners are designed using a modified or simplified version of the HBF-A approach. The HBF-A strategy in [29] and the data-sharing (DS) strategy in [36]. If we compare the performances of the proposed HBF-A and HBF-B approaches in the setup under the mmWave channels in [7]. To match the characteristics of mmWave propagation, we model the RUs and UEs as two independent stationary Poisson point processes with densities \( \lambda_{RU} \) and \( \lambda_{UE} \), respectively. The RUs and UEs are independently and uniformly distributed in a circular region with a radius \( R = 200 \) m as shown in Fig. 6. Both the RUs and UEs are equipped with uniform linear arrays, and the angles of arrival and departure are assumed to be uniformly distributed in \([0, 2\pi]\).

To facilitate the analysis of simulation results, we assume that each RU has a single data stream and \( N_{RF} = 64 \) receive antennas for \( j \in \mathcal{M}_U \), while each RU has \( N_{RF} = 2 \) RF chains and \( N_{RF} = 64 \) transmit antennas for \( i \in \mathcal{M}_R \). For simplicity, the weight of \( U_j \) is assumed to be \( \mu_j = 1 \) for all \( j \in \mathcal{U}_j \).

The noise power is given by \( \sigma^2 = N/\sigma_0 \), where \( B = 500 \) MHz is the signal bandwidth, \( \sigma_0 = 3 \text{ dB} \) is the receiver's noise figure and \( N_0 = -174 \text{ dBm/Hz} \) is the noise power spectral density.

**Fig. 7** Shows the ergodic achievable sum rate versus the UE density under different transmit power constraints and the same compression noise. As shown in Fig. 7, the HBF-A scheme is preferred in the large power regime, while the HBF-B scheme is advantageous in the lower power condition. When the compression noise is fixed, increasing \( P \) weakens the relative impact of the quantisation noise on the system performance, which contributes to a larger ergodic achievable sum rate. For the HBF-A method, since the baseband signal is quantised as in (2), the performance loss caused by the compression noise is not affected by the transmit power, which explains its better performance. In contrast, for the HBF-B method, since the digital BF matrix instead of the processed baseband signal is quantised as in (15), the performance loss caused by compression increases with the transmit power. Thus, in the large power regime, the performance improvement of the HBF-B scheme is very limited, which results in a poor performance compared to the HBF-A scheme.

The HBF-B scheme still achieves performance very close to that achieved with the HBF-A scheme. Together with the discussion on Category B RU in Section 2.2, it can be seen that the HBF-B scheme not only can alleviate the burden of fronthaul bandwidth, but also shorten the transmit delay. Although it increases the computational complexity of RU, from another point of view, it also reduces the workload of DU, which is vital for 5G networks requiring massive information processing. Thus, in the latency-sensitive and dense traffic networks with large power, compared with HBF-A, the HBF-B scheme appears to be an attractive...
solution to balance the tradeoff among fronthaul bandwidth, transmit latency and system performance.

The ergodic achievable sum rate versus the UE density under different compression noise variances and transmit powers is investigated in Fig. 8. It can be observed that, the compression noise could largely degrade the system performance, which emphasises the importance of properly selecting the quantiser during the system design. With the same transmit power and compression noise, the ergodic achievable sum rate versus the UE density under different numbers of receive antennas and transmit powers is presented in Fig. 9. It is shown in Fig. 9 that, benefiting from large BF gains, increasing the number of receive antennas improves the performances of both the HBF-A and HBF-B schemes. Moreover, as illustrated in Figs. 8 and 9, the performance gap between the HBF-A and HBF-B schemes becomes more pronounced with the increase of compression noise or the number of receive antennas, and for both large and small power cases. This is because, as the compression noise or the number of receive antennas increases, a large power exacerbates the compression loss, which leads to a much worse performance of the HBF-B scheme, while a small power alleviates the compression loss, which makes HBF-B achieve a much better performance than HBF-A.

Therefore, compared with the HBF-A method, the HBF-B method is more sensitive to changes in system parameters, such as the compression noise and the number of receive antennas, in the large power regime, while it is less sensitive in the small power regime.

Considering that it is difficult to obtain perfect CSI in practice, in order to evaluate the effect of imperfect CSI on the system performance, the estimated channel model \( \tilde{H} = \xi H + \sqrt{1 - \xi^2} \Theta \) is used [7]. In this model, \( \tilde{H} \) and \( H \) represent the channel matrices with imperfect CSI and perfect CSI, respectively, \( \Theta \) is the error matrix having i.i.d. \( \mathcal{CN}(0, 1) \) entries, and \( \xi \) is the accuracy of channel estimation, which satisfies \( 0 \leq \xi \leq 1 \). Fig. 10 illustrates the ergodic achievable sum rate versus the UE density for perfect CSI and imperfect CSI with \( \xi = 0.7 \), \( \lambda_{RU} = (10/\pi R^2) \), \( \sigma^2 = -60 \text{ dBm} \), and \( \lambda_{RU} = (10/\pi R^2) \), \( \sigma^2 = -60 \text{ dBm} \).

5 Conclusion

In this paper, we have investigated the intra-PHY functional splits of 5G NR protocol stack and studied the features and merits of various split options. We have introduced two categories of RUs and discussed two implementation modes of hybrid BF at the PHY layer, in which the digital BF is performed either at the DU (as in the HBF-A scheme), or at the RU (as in the HBF-B scheme). Then, while taking into account the constant modulus constraints, transmit power and fronthaul capacity constraints, we have formulated the problem of jointly designing hybrid BF, analogue combining and fronthaul compression strategies for both HBF-A
and HBF-B schemes with the goal of maximising the WSR. The formulated problems were tackled by leveraging the MM algorithm when the codebook-based solution was adopted. Numerical results have confirmed that the HBF-A scheme outperforms the HBF-B scheme in the large power regime. Compared with the HBF-A method, the HBF-B method is more sensitive to changes in system parameters, such as the compression noise and the number of receive antennas, in the large power regime while it is less sensitive in the small power regime. In the future, we plan to solve the formulated problem using different algorithms and compare their performance differences. The issue of robustness against channel estimation error is another relevant and interesting research topic.

References


[3] 3GPP TR 38.912 v15.0.0: ‘Study on new radio (NR) access technology issues and challenges’, 3GPP, 2018


[9] 3GPP TR 38.801 v14.0.0: ‘Study on new radio access technology: radio access interface and technologies’, 3GPP, 2017


[18] Li, J., Yue, D., Sun, Y.: ‘Performance analysis of millimeter wave massive MIMO systems in centralized and distributed schemes’, IEEE Access, 2018, 6, pp. 75482–75494