CHALMERS
UNIVERSITY OF TECHNOLOGY

MSE minimized joint transmission in coordinated multipoint systems with sparse feedback and constrained backhaul requirements

Nezafati *et al. J Wireless Com Network* (2021) 2021:103
https://doi.org/10.1186/s13638-021-01979-3

EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

MSE minimized joint transmission in coordinated multipoint systems with sparse feedback and constrained backhaul requirements

Mohammad Bagher Nezafati[1], Mehrdad Taki[2*] and Tommy Svensson[3]

*Correspondence: mehrdadtaki@gmail.com
[2] Department of Electrical Engineering, University of Qom, Qom, Iran
Full list of author information is available at the end of the article

Abstract

In a joint transmission coordinated multipoint (JT-CoMP) system, a shared spectrum is utilized by all neighbor cells. In the downlink, a group of base stations (BSs) coordinately transmit the users' data to avoid serious interference at the users in the boundary of the cells, thus substantially improving area fairness. However, this comes at the cost of high feedback and backhaul load; In a frequency division duplex system, all users at the cell boundaries have to collect and send feedback of the downlink channel state information (CSI). In centralized JT-CoMP, although with capabilities for perfect coordination, a central coordination node have to send the computed precoding weights and corresponding data to all cells which can overwhelm the backhaul resources. In this paper, we design a JT-CoMP scheme, by which the sum of the mean square error (MSE) at the boundary users is minimized, while feedback and backhaul loads are constrained and the load is balanced between BSs. Our design is based on the singular value decomposition of CSI matrix and optimization of a binary link selection matrix to provide sparse feedback—constrained backhaul link. For comparison, we adopt the previously presented schemes for feedback and backhaul reduction in the physical layer. Extensive numerical evaluations show that the proposed scheme can reduce the MSE with at least 25%, compared to the adopted and existing schemes.

Keywords: JT-CoMP, Sparse feedback, Constrained backhaul, MSE minimization, SVD decomposition

1 Introduction

Mobile communication systems are becoming an essential part of social networks, interactive media (e.g. augmented and virtual reality), internet of things (IoT), and a facilitator for the digital economy. This drives the fifth-generation and beyond (B5G) of mobile communication systems to scale mainly in three dimensions; (a) Rate is scaled to 10 Gbps which is ten-times of peak data rate in the fourth generation (4G) of long term evolution (LTE), release 10, and end-to-end latency decreases to 1 ms which is one-fifth of latency compared to 4G. (b) A massive scale in the number of connected devices in

IoT is needed. (c) Higher system reliability and quicker round trip times are to be available in transportation systems and industrial process control. To fulfill these requirements, a combination of various new techniques such as massive multiple input multiple output (MMIMO), dense small cells, cooperative communications such as device-to-device (D2D) and coordinated multipoint (CoMP), advanced air interface, additional spectrum at higher frequencies (mm-wave), and integrated access and backhaul (IAB) are needed [1–5].

One of the most promising concepts to cover all the above requirements is ultra-dense networks (UDN) with frequency reuse (close to) one, where more small base stations (BSs) are deployed within the service area. However, using the same frequency for all BSs, exposes the users to severe inter-cell-interference (ICI), especially at the cell edges. Recently, to reap the benefits of UDN, cloud radio access network (CRAN) architecture has been proposed, where ICI can be effectively mitigated by employing the CoMP technique [6, 7]. Initially, CoMP was introduced for LTE-A by the third generation partnership project (3GPP) to mitigate ICI in cell-edge users [8]. Qualcomm has implemented a fifth-generation (5G) CoMP testbed and showed a fourfold increase in system capacity [9]. Currently, CoMP is considered as one of the potential technologies for 5G cellular networks. Accordingly, 5G enhanced some aspects of CoMP such as control signaling and channel feedback [10, 11].

There are three main categories for downlink CoMP, i.e. joint transmission (JT), dynamic point selection (DPS), and coordinated scheduling/beam-forming (CS/CB). In JT-CoMP, the data related to a user is available at all serving BSs and is transmitted simultaneously by each BS. This transmission can be coherent or non-coherent. In coherent transmission (also known as multiple input multiple output (MIMO) network), the signal strength is enhanced by precoding the data to exploit the phase and amplitude information of each channel. We consider coherent JT-CoMP, since it generally outperforms the laters in system performance [12].

In JT-CoMP, a group of BSs forms virtual antenna array distributed across multiple cells. In the downlink, two or more geographically separated BSs cooperate to jointly and coordinately transmit to cell-edge users, where the improvement is most needed and exploit the interference as a useful signal. Superposition of signals at the user position is performed in a way to maximize the desired signal (constructive) and at the same time minimize the ICI (destructive). This requires that accurate channel state information (CSI) is available at the transmitter side. In time division duplex (TDD) transmission, CSI is acquired from the reciprocity of uplink and downlink channels, while in frequency division duplex (FDD) transmission, users need to feedback the received CSI from all serving BSs to the BS with the strongest link. In the centralized approach of FDD JT-CoMP, the CSI for all users is aggregated in the central coordination node (CCN) to calculate the precoding weights for the subsequent downlink transmission. Through backhaul, the precoding weights with the users' data are to be sent from CCN to the corresponding BSs. Finally, each BS transmits a weighted combination of all users' data to the users.

The throughput of downlink JT-CoMP heavily relies on the quality of the CSI at the transmitters. Feedback latency and reliability of the feedback channel degrade the system performance [13, 14]. Other impairments such as imperfect carrier and sampling

frequencies among BSs cause a mismatch between the precoder and the actual channel, which limits the potential gains of JT-CoMP [15, 16]. In practical implementations of JT-CoMP, feedback and backhaul loads are two main challenges that need to be addressed properly. As system performance depends on the CSI quality, CSI has to be sent back at very low latency to avoid it to be outdated before being used for precoding. Therefore a large amount of CSI is required, which poses a considerable feedback load. The users' data needs to be available at all serving BSs. Moreover, the precoding weights are to be sent to the corresponding BSs, which pose a heavy burden on the backhaul traffic. Also, increasing the number of cooperating BSs to improve the spectral efficiency, increases the backhaul load [17–19].

Although CoMP is one of the main solutions to mitigate ICI and has been considered from 4G to B5G, feedback and backhaul loads have been identified as two of the key challenges for its practical implementation and prevent the CoMP from real take off. It is expected in the 5G era, a powerful fiber backhaul is available at least for macro-cells in CRAN architecture. Such deployment can inherently provide a low-latency and high-capacity backhaul needed for JT-CoMP, while connections of small-cells might be ranging from fiber to various relaying and IAB approaches, which have lower capacity but might be more cost-effective [3, 5].

It is highly desirable to reduce the feedback and backhaul loads by routing users' data to a limited number of BSs. This reduction can be done in the medium access control (MAC) layer [20–24] or physical (PHY) layer [7, 25–27]. Furthermore, recently by considering the cache mechanism to BSs, the transmission latency and the total backhaul bandwidth consumption is reduced which is based on upper layers of the network [28]. In this paper, we aim to reduce feedback and backhaul loads simultaneously in JT-CoMP using the PHY layer schemes.

In the PHY layer based schemes, limited feedback and backhaul precoders are designed with respect to sum-rate maximization [25–27] or maximizing the number of users admitted to the network [7]. In [25], absolute and relative thresholding was proposed for feedback load reduction in the FDD downlink. In absolute thresholding, only CSI with a corresponding signal to noise ratio (SNR) exceeding a predefined threshold is sent back, whereas in relative thresholding, the threshold is set based on the strongest channel. It has been shown that the latter technique provides a good trade-off between sum-rate performance and feedback overhead. An absolute threshold is used for selective feedback in [26], where for limiting the user data exchange through the backhaul, two schemes including scheduling and precoding techniques are employed. The proposed framework has a good sum-rate performance with limited system overhead. The idea of relative thresholding is followed by [27], where the precoder is designed using a successive second order cone programming (SSOCP) to maximize the sum-rate for all cell-edge users.

Recently, the 3GPP initiated a standardization activity to employ codebook-based precoding at BSs with an aim to decrease CSI feedback overhead to satisfy the spectral efficiency requirement of future cellular systems. In 3GPP LTE, codebook type I was introduced [29] and for more accurate CSI feedback to better support the transmission in new radio (NR), codebook type II was introduced in 3GPP Release 15 [30]. This throughput gain comes at the expense of a significant increase in feedback overhead. To

this end, Release 16 introduces enhanced CSI feedback by compressing the CSI report in the frequency domain and extending the codebook type II to support MIMO channels with rank larger than two [31]. These enhancements increase throughput and reduce CSI feedback overhead [32].

In this paper, we design a novel JT-CoMP transmission scheme based on singular value decomposition (SVD) to minimize the sum mean square error (MSE) at boundary users. As in practical networks especially mobile systems, implementation of continuous rate adaptation is impossible and rate is selected from a limited discrete set [33–35], we consider the MSE criteria for system performance evaluation. We optimize a binary link selection matrix, in which each element corresponds to the link between an antenna of a BS and an antenna of a user. If an element is one, the corresponding antenna serves the user, otherwise, it is not involved in the transmission. We propose a two-layer recursive optimization method; In the inner layer, the SVD of the CSI matrix is utilized to design a precoder fulfilling a sum-power constraint. In the outer layer, the link selection matrix is designed, providing required feedback and backhaul load reductions and load balancing between BSs. To obtain a further reduction of the feedback load, we consider a CSI codebook based limited feedback strategy, where each user selects a codeword from a CSI codebook and feeds back its index to the serving BS. The CCN collects all the codeword indexes and calculates the precoding matrix. Random vector and uniformly distributed quantizations are employed respectively for quantizing the channel direction information and phase ambiguity.

We compare our scheme with two recent works in this area [26, 27], having a close target performance goal for our design. Moreover, we adopt two existing precoders, zero forcing (ZF) and Wiener, in our two layer optimization scheme to provide sparsity constraints on feedback and backhaul. As shown, our scheme outperforms [26, 27] with at least 30%, and adopted ZF and Wiener precoders with 25% from the MSE aspect. The key contributions of this paper are summarized as follows:

- In the previously presented approaches, just the average backhaul or feedback loads are controlled, while the hardware must be available for the worst-case scenario. In our scheme, the feedback and backhaul loads are strictly constrained.
- As the load balancing has a key role in radio resource optimization, we consider a constraint for the number of users that are served by a specific BS. This association constraint between BSs and users, reduces the maximum load in each BS.
- The proposed scheme is not sensitive to the type of receiver and has the same performance in receivers with different receive filters, while the performance of adopted ZF and Wiener precoders depend on the type of receive filter. Therefore, an advantage of the proposed scheme is that it has good performance in simplified receivers such as receivers using no receive filter.
- We employ a CSI codebook based feedback strategy to further reduce the feedback load. In this regard, the random vector and uniform distribution are used for quantizing the channel direction and phase information. It is showed that by employing only 6 bits for quantization, a performance near to the full CSI feedback is attainable. In this technique, users employ different CSI codebooks to independently quantize their CSI.

Table 1 List of main variables

$\boldsymbol{H}_u$	Channel matrix for the u-th user
$\boldsymbol{W}_u$	Precoding matrix for the u-th user
$\bar{\boldsymbol{H}}_u$	Effective channel matrix for the u-th user
$\tilde{\boldsymbol{H}}_u$	Composite channel matrix for the u-th user
$\hat{\boldsymbol{W}}$	Sparse precoding matrix
$\boldsymbol{S}$	Link selection matrix
$\boldsymbol{S}^b$	Link selection sub-matrix for the b-th BS
$\boldsymbol{S}_u$	Link selection sub-matrix for the u-th user
N_b	Number of BSs
N_t	Number of BS's transmission antennas
N_u	Number of users
N_r	Number of user's transmission antennas

- The proposed scheme has good convergence properties. It converges after transmitting 5 subframes, in the worst case.

The remainder of the paper is organized as follows, in Sect. 2, preliminaries including notation and system model are presented. Section 3 is devoted to designing the new proposed JT-CoMP transmission scheme, in Sect. 4, benchmark schemes are explained, and Sect. 5 numerically evaluates the proposed scheme and its efficiency in comparison with the benchmarks. The paper conclusions are made in Sect. 6.

2 Preliminaries

2.1 Notation

In this paper, scalar variables are denoted by small italic letters e.g. x and vector variables by small italic bold letters e.g. $\boldsymbol{x}$. Sets are denoted by calligraphic letters e.g. $\mathcal{W}$. The absolute value of scalar variables or the number of members in a set or matrix is shown by $|.|$, the maximum integer lower than x is shown with $\lfloor x \rfloor$. The Euclidean and Frobenius norm of vector $\boldsymbol{x}$ are denoted by $\|\boldsymbol{x}\|_2$ and $\|\boldsymbol{x}\|_F$. The transpose, conjugate, and conjugate transpose (Hermitian) of matrix $\boldsymbol{H}$ are shown respectively by $\boldsymbol{H}^T$, $\boldsymbol{H}^*$ and $\boldsymbol{H}^H$. Indeed $\boldsymbol{H}(i,:)$ and $\boldsymbol{H}(:,j)$ denote the i-th row vector and the j-th column vector of the matrix $\boldsymbol{H}$, respectively, while $\boldsymbol{H}(i,j)$ is the i-th row and the j-th column element of the matrix $\boldsymbol{H}$. The vectors $\boldsymbol{d}(\boldsymbol{A})$ and $\boldsymbol{\lambda}(\boldsymbol{A})$ contain diagonal and eigenvalues of the square matrix $\boldsymbol{A}$. The operator $Tr(\boldsymbol{A}) = \sum_i \boldsymbol{A}(i,i)$ and $Tr(\boldsymbol{d}) = \sum_i \boldsymbol{d}(i)$ denote trace of a matrix or vector. The maximum eigenvalue of a Hermitian matrix $\boldsymbol{A}$ is represented by $\lambda_{\max}(\boldsymbol{A})$. The number of nonzero elements in matrix $\boldsymbol{A}$ is shown by $nnze(\boldsymbol{A})$, while the number of nonzero rows and columns are shown by $nnzr(\boldsymbol{A})$ and $nnzc(\boldsymbol{A})$, respectively. Inner matrix product of $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ and $\boldsymbol{B} \in \mathbb{C}^{M\times N}$ is denoted with $\boldsymbol{C} = \boldsymbol{A} \cdot \boldsymbol{B}$, where $\boldsymbol{C}(i,j) = \boldsymbol{A}(i,j)\boldsymbol{B}(i,j), 1 \le i \le M, 1 \le j \le N$. Identity matrix of size $N \times N$ is represented by $\mathbf{I}_N$ and the all one matrix of size $M \times N$ is shown by $\mathbf{1}_{M\times N}$. The list of main variables in the paper is presented in Table 1.

2.2 System model and network structure

Figure 1 shows a schematic form of the considered network, including 3 neighboring cells and 3 users at the common boundary of the cells. There is a cluster area in the

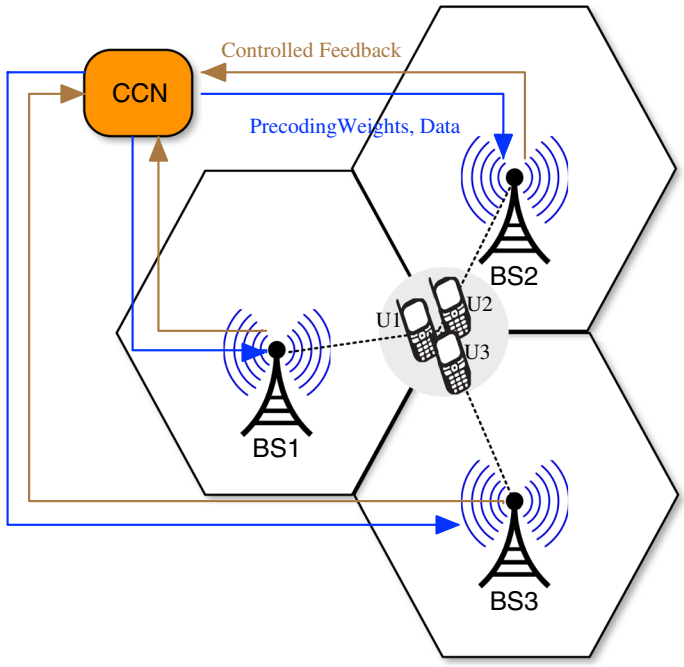

Fig. 1 A simple schematic model for the considered network

middle of the cells, in which there would be high interference, in case of no media division or coordination between cells. This is shown with a gray area in Fig. 1.

In downlink transmission of the JT-CoMP scheme, BSs are coordinated and jointly transmit as a single virtual multi-antenna transmitter with distributed antennas. To this end, data for the users in the cluster center is sent to all BSs via a backhaul link. Each BS transmits a linear combination of users' data with a proper precoding weight. Precoding weights are calculated in the CCN, based on the CSI of all links between BSs and cluster centered users.

In TDD transmission, CSI could be implicitly estimated at the BSs based on channel reciprocity. But in FDD, CSI is estimated by the users and are sent back to the BSs. All CSI is sent from the BSs to the CCN to calculate the precoding weights.

Figure 2 shows the downlink of the mentioned system. In general, there are N_b BSs each of them with N_t antennas that are serving N_u users, each of them with N_r antennas. Thus, there are in total $\mathrm{N}_B = \mathrm{N}_b\mathrm{N}_t$ transmit and $\mathrm{N}_U = \mathrm{N}_u\mathrm{N}_r$ receive antennas. The channel gain at the u-th user, $1 \le u \le \mathrm{N}_u$ is defined by $\boldsymbol{H}_u \in \mathbb{C}^{\mathrm{N}_r \times \mathrm{N}_B}$ in which $\boldsymbol{H}_u(i,j)$, $1 \le i \le \mathrm{N}_r$, $1 \le j \le \mathrm{N}_B$ is the channel gain between the i-th antenna of the u-th user and the j-th transmit antenna. The mentioned transmit antenna belongs to the b-th BS, where $b = \lfloor j/\mathrm{N}_b \rfloor$.

The aggregated data symbols are denoted by $\boldsymbol{x} = \left[\boldsymbol{x}_1^T \cdots \boldsymbol{x}_{\mathrm{N}_u}^T\right]^T \in \mathbb{C}^{\mathrm{N}_U \times 1}$ in which $\boldsymbol{x}_u \in \mathbb{C}^{\mathrm{N}_r \times 1}$ is the data symbol for the u-th user, where $E\{\boldsymbol{x}\boldsymbol{x}^H\} = \sigma_x^2 \mathbf{I}_{\mathrm{N}_U}$. Each BS transmits a linear combination of all precoded data symbols. The precoding matrix corresponding to the u-th user, which belongs to different BSs, is denoted by $\boldsymbol{W}_u \in \mathbb{C}^{\mathrm{N}_B \times \mathrm{N}_r}$, $1 \le u \le \mathrm{N}_u$ and the aggregated precoding matrix is denoted by $\boldsymbol{W} = \left[\boldsymbol{W}_1 \cdots \boldsymbol{W}_{\mathrm{N}_u}\right] \in \mathbb{C}^{\mathrm{N}_B \times \mathrm{N}_U}$. Thus,

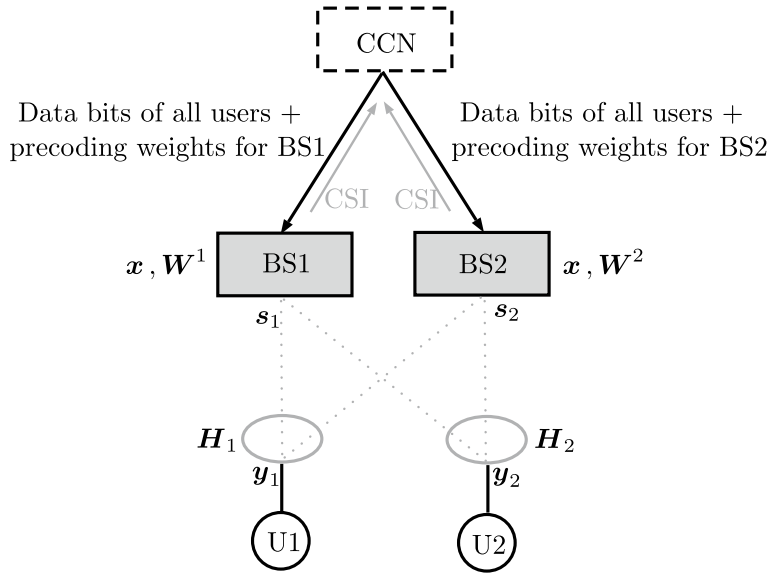

Fig. 2 Downlink transmission setup in a JT-CoMP System

the transmitted signal by all BSs is $\boldsymbol{W}_1\boldsymbol{x}_1 + \cdots + \boldsymbol{W}_{\mathrm{N}_u}\boldsymbol{x}_{\mathrm{N}_u}$ and the received signal at the u-th user is

$$\boldsymbol{y}_u = \boldsymbol{H}_u(\boldsymbol{W}_1\boldsymbol{x}_1 + \cdots + \boldsymbol{W}_{\mathrm{N}_u}\boldsymbol{x}_{\mathrm{N}_u}) + \boldsymbol{n}_u, \tag{1}$$

where $\boldsymbol{n}_u$ is a complex Gaussian random variable with zero mean and variance of σ_n^2. The signal to interference and noise ratio (SINR) at the u-th user is computed as

$$SINR_u = \frac{\|\boldsymbol{H}_u\boldsymbol{W}_u\|_2^2}{\sum_{i=1,i\neq u}^{\mathrm{N}_u}\|\boldsymbol{H}_i\boldsymbol{W}_i\|_2^2 + \sigma_n^2}. \tag{2}$$

By considering the aggregated channel gain as $\boldsymbol{H} = \left[\boldsymbol{H}_1^T \cdots \boldsymbol{H}_{\mathrm{N}_u}^T\right]^T \in \mathbb{C}^{\mathrm{N}_U \times \mathrm{N}_B}$, the aggregated received signal $\boldsymbol{y} = \left[\boldsymbol{y}_1^T \cdots \boldsymbol{y}_{\mathrm{N}_u}^T\right]^T \in \mathbb{C}^{\mathrm{N}_U \times 1}$ is computed as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{W}\boldsymbol{x} + \boldsymbol{n}, \tag{3}$$

where $\boldsymbol{n} = \left[\boldsymbol{n}_1^T \cdots \boldsymbol{n}_{\mathrm{N}_u}^T\right]^T \in \mathbb{C}^{\mathrm{N}_U \times 1}$ is the noise vector. The receive filter at the u-th user is denoted by $\boldsymbol{g}_u \in \mathbb{C}^{\mathrm{N}_r \times \mathrm{N}_r}$ and the detected signal at the u-th receiver is $\widetilde{\boldsymbol{x}}_u = \boldsymbol{g}_u\boldsymbol{y}_u$. If the receive filters are aggregated as $\boldsymbol{G} = diag\left(\boldsymbol{g}_1 \cdots \boldsymbol{g}_{\mathrm{N}_u}\right) \in \mathbb{C}^{\mathrm{N}_U \times \mathrm{N}_U}$, the detected symbols are computed as

$$\widetilde{\boldsymbol{x}} = \boldsymbol{G}\boldsymbol{y}. \tag{4}$$

This paper aims to minimize the weighted sum MSE at all users which can be calculated as

$$\mathrm{MSE} = E\left\{\left\|\boldsymbol{a}\left(\boldsymbol{x} - \boldsymbol{\alpha}\widetilde{\boldsymbol{x}}\right)\right\|_2^2\right\}, \tag{5}$$

where $\boldsymbol{a} = diag(a_1\mathbf{I}_{\mathrm{N}_r} \cdots a_{\mathrm{N}_u}\mathbf{I}_{\mathrm{N}_r}) \in \mathbb{R}_{+}^{\mathrm{N}_U \times \mathrm{N}_U}$, and a_u is the non-negative user weight. To make the MSE calculation to be meaningful, $\boldsymbol{\alpha} = diag(\alpha_1\mathbf{I}_{\mathrm{N}_r} \cdots \alpha_{\mathrm{N}_u}\mathbf{I}_{\mathrm{N}_r}) \in \mathbb{R}_{+}^{\mathrm{N}_U \times \mathrm{N}_U}$ is considered, where α_u is a scalar factor which can be considered as an automatic gain control (AGC) gain in the receiver [36–39]. In Sect. 3.4, the traditional receive filters, and the AGC scalar factor are stated for the proposed system.

3 Design of a novel JT-CoMP transmission scheme with sparse feedback and constrained backhaul

This section aims to design a novel scheme for downlink transmission in a centralized JT-CoMP system. The goal of the design is to minimize the MSE in (5). As stated before, the drawback of JT-CoMP is its feedback and backhaul loads. It is interesting to design a transmission scheme in which the CSI requirement and backhaul load are constrained. Towards this aim, we define a binary link selection matrix $\boldsymbol{S} \in \mathcal{B}^{\mathrm{N}_B \times \mathrm{N}_U}$, where $\boldsymbol{S}(i,j), 1 \le i \le \mathrm{N}_B, 1 \le j \le \mathrm{N}_U$ is 1 when the link between the i-th transmit antenna and the j-th receive antenna is active and it is 0 when the mentioned link is idle. The column-wise sub-matrix of the link selection matrix related to the b-th BS is defined as $\boldsymbol{S}^b = \boldsymbol{S}((b-1)\mathrm{N}_t + 1 : b\mathrm{N}_t, :) \in \mathcal{B}^{\mathrm{N}_t \times \mathrm{N}_U}$, and the row-wise sub-matrix related to the u-th user is denoted as $\boldsymbol{S}_u = \boldsymbol{S}(:, (u-1)\mathrm{N}_r + 1 : u\mathrm{N}_r) \in \mathcal{B}^{\mathrm{N}_B \times \mathrm{N}_r}$. The sparse precoding matrix is defined as

$$\hat{\boldsymbol{W}} = \boldsymbol{S} \cdot \boldsymbol{W}, \tag{6}$$

where the backhaul load reduction is proportional to the cardinality of the set $\mathcal{S}_{BH} = \left\{ \hat{\boldsymbol{W}}(i,j) = 0,\ 1 \le i \le \mathrm{N}_B, 1 \le j \le \mathrm{N}_U \right\}$, and is defined as

$$r_{bl} \triangleq 1 - \frac{nnze(\boldsymbol{S})}{|\boldsymbol{S}|}. \tag{7}$$

Similarly, the feedback load reduction is proportional to the number of zeros in the sparse aggregated channel matrix $\hat{\boldsymbol{H}}$, i.e. the cardinality of $\mathcal{S}_{FB} = \left\{ \hat{\boldsymbol{H}}(i,j) = 0,\ 1 \le i \le \mathrm{N}_U, 1 \le j \le \mathrm{N}_B \right\}$. We assume the equivalent feedback of $\hat{\boldsymbol{W}}$ as $\hat{\boldsymbol{H}} = \boldsymbol{H} \cdot \boldsymbol{S}^T$ and this results in a feedback load reduction as

$$r_{fl} \triangleq 1 - \frac{nnze(\boldsymbol{S}^T)}{|\boldsymbol{S}^T|}. \tag{8}$$

Note that in our proposed scheme, the selection matrices for feedback and precoding are transpose of each other, and the feedback and backhaul load reduction ratios are the same, i.e. $r_{fl} = r_{bl}$. Indeed $nnzc(\boldsymbol{S}^b)$ shows the number of users which are served by the b-th BS. The load of BSs may be balanced, by considering the following constraint [40]

$$\max_{1 \le b \le \mathrm{N}_b} \left\{ nnzc(\boldsymbol{S}^b) \right\} - \min_{1 \le b \le \mathrm{N}_b} \left\{ nnzc(\boldsymbol{S}^b) \right\} \le 1. \tag{9}$$

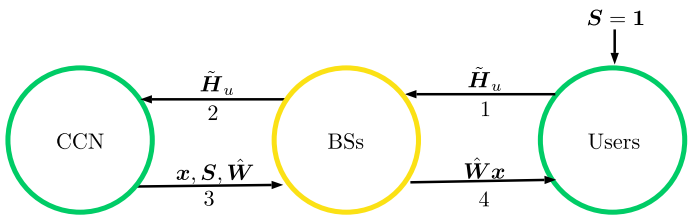

Fig. 3 Representation of data flow in the proposed JT-CoMP system with sparse feedback and constrained backhaul

In the following, a general optimization problem to find $\boldsymbol{W}$ and $\boldsymbol{S}$ is set up and solved. The transmission scheme is briefly shown in Fig. 3. At the beginning of the transmission, $\boldsymbol{S} = \mathbf{1}_{\mathrm{N}_B \times \mathrm{N}_U}$, i.e. we start with a non-sparse case. Substituting (3) and (4) in (5) and considering $\sigma_x^2 = 1$ and $E\{\boldsymbol{n}\} = 0$, the MSE in (5) is computed as

$$\begin{aligned} MSE =& E\left\{ \left\| \boldsymbol{a}\left(\boldsymbol{x} - \boldsymbol{\alpha}\boldsymbol{G}\left(\boldsymbol{H}\hat{\boldsymbol{W}}\boldsymbol{x} + \boldsymbol{n}\right)\right) \right\|_2^2 \right\} = E\left\{ \left\| \boldsymbol{a}\left(\boldsymbol{I} - \boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}\hat{\boldsymbol{W}}\right)\boldsymbol{x} - \boldsymbol{a}\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{n} \right\|_2^2 \right\} \\ =& \left\| \boldsymbol{a}\left(\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}\hat{\boldsymbol{W}} - \boldsymbol{I}\right) \right\|_F^2 + \sigma_n^2 \|\boldsymbol{a}\boldsymbol{\alpha}\boldsymbol{G}\|_F^2, \end{aligned} \tag{10}$$

where σ_n^2 is the noise variance. The detailed steps of MSE computation are described in "Appendix 1". The goal is to minimize the MSE provided that the total transmission power is constrained to P_t, feedback and, backhaul loads are constrained and load is balanced between BSs. Thus the optimization problem is set up as follows

$$\min_{\boldsymbol{S},\boldsymbol{W}} \underbrace{\|\boldsymbol{a}(\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}(\boldsymbol{S}.\boldsymbol{W}) - \mathbf{I})\|_F^2}_{M_1} + \underbrace{\sigma_n^2 \|\boldsymbol{a}\boldsymbol{\alpha}G\|_F^2}_{M_2} \tag{11}$$

subject to:

$$C(1) \ : \ Tr\left((\boldsymbol{S}.\boldsymbol{W})^H(\boldsymbol{S}.\boldsymbol{W})\right) \leq P_t \tag{12}$$

$$C(2) \ : \ nnze(\boldsymbol{S}) \leq \mathrm{Q} \in \{1, 2, \ldots, \mathrm{N}_B\mathrm{N}_U\} \tag{13}$$

$$C(3) \ : \ \max_{1 \leq b \leq \mathrm{N}_b} \left\{ nnzc(\boldsymbol{S}^b) \right\} - \min_{1 \leq b \leq \mathrm{N}_b} \left\{ nnzc(\boldsymbol{S}^b) \right\} \leq 1 \tag{14}$$

Remark 1 Note that $\boldsymbol{S}$ and $\boldsymbol{W}$ are calculated at the CCN, which is not necessarily aware of the receiver and its filter, thus the CCN can only manage to minimize the part M_1 in (11).

Remark 2 Users estimate channel $\boldsymbol{H}$ and find $\boldsymbol{S}$ from the received data, and feed back a composite and sparse version of the CSI, which also contains the information about the receive filter. I.e., $\boldsymbol{H}_f = (\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}) \cdot \boldsymbol{S}^T$ is sent back and aggregated in the CCN. The CCN receives the sparse composite CSI and it estimates the full composite CSI as

$$\tilde{\boldsymbol{H}} = \boldsymbol{H}_f + (\mathbf{1}_{\mathrm{N}_U \times \mathrm{N}_B} - \boldsymbol{S}^T).\boldsymbol{\mathcal{H}} \in \mathbb{C}^{\mathrm{N}_U \times \mathrm{N}_B}, \tag{15}$$

where $\boldsymbol{\mathcal{H}}$ may be an old version of $\boldsymbol{H}_f$ or long term channel statistics (e.g. received signal strength indicator (RSSI)).

Remark 3 As seen, problem (11) is too complicated to be solved directly, especially for its boolean parameters. Thus, we use a sub-optimum solution by converting it to a two-layer optimization procedure. In the inner layer, $\boldsymbol{W}$ is calculated considering a fixed $\boldsymbol{S}$ and subject to constraint $C(1)$. In the outer layer, $\boldsymbol{S}$ is found subject to constraints $C(2)$ and $C(3)$, i.e. the problem is converted to

$$\min_{\boldsymbol{S}} \left[\min_{\boldsymbol{W}} \|\boldsymbol{a}(\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}(\boldsymbol{S} \cdot \boldsymbol{W}) - \mathbf{I})\|_F^2 + \sigma_n^2 \|\boldsymbol{a}\boldsymbol{\alpha}G\|_F^2 \text{ s.to } C(1) \right] \text{ s.to } C(2) \text{ and } C(3). \tag{16}$$

In the following, we first explain the two-layer optimization scheme, and next considerations about traditional receive filters are explained.

3.1 Inner layer optimization: precoder design

In this section, we aim to design a robust precoder by minimizing the part M_1 in (11) for different types of receive filter that also has good performance in a receiver with no filter. We design the precoder weights based on the composite channel gains, $\tilde{\boldsymbol{H}}$. If $\mathcal{W} = \left\{ \hat{\boldsymbol{W}} | Tr\left(\hat{\boldsymbol{W}}^H \hat{\boldsymbol{W}} \right) \leq P_t \right\}$ is the set of all possible weights that is satisfying the total power constraint $C(1)$, by considering fixed $\boldsymbol{S}$ and $\boldsymbol{a} = \mathbf{I}$, the problem (11) can be written as

$$\min_{\hat{\boldsymbol{W}} \in \mathcal{W}} \left\| \tilde{\boldsymbol{H}} \hat{\boldsymbol{W}} - \mathbf{I} \right\|_F^2. \tag{17}$$

By substituting $\hat{\boldsymbol{W}} = \boldsymbol{S}.\boldsymbol{W} = \boldsymbol{W} - \left(\mathbf{1}_{\mathrm{N}_B \times \mathrm{N}_U} - \boldsymbol{S}\right).\boldsymbol{W}$ in (17) and applying the triangle inequality, an upper bound for the objective function is acquired as

$$\begin{aligned} \left\| \tilde{\boldsymbol{H}} \hat{\boldsymbol{W}} - \mathbf{I} \right\|_F^2 &= \left\| \tilde{\boldsymbol{H}} \boldsymbol{W} - \mathbf{I} - \tilde{\boldsymbol{H}} \left((\mathbf{1}_{\mathrm{N}_B \times \mathrm{N}_U} - \boldsymbol{S}).\boldsymbol{W} \right) \right\|_F^2 \\ &\leq \underbrace{\left\| \tilde{\boldsymbol{H}} \boldsymbol{W} - \mathbf{I} \right\|_F^2}_{F1} + \underbrace{\left\| \tilde{\boldsymbol{H}} \left((\mathbf{1}_{\mathrm{N}_B \times \mathrm{N}_U} - \boldsymbol{S}).\boldsymbol{W} \right) \right\|_F^2}_{F2}. \end{aligned} \tag{18}$$

Now, instead of optimization of the objective function in (17), we try to optimize its upper bound. First, we consider the section $F1$ of (18) and adapt the Theorem 2 from [41] to minimize it as follows

$$\min_{\boldsymbol{S}.\boldsymbol{W} \in \mathcal{W}} \left\| \tilde{\boldsymbol{H}} \boldsymbol{W} - \mathbf{I} \right\|_F^2. \tag{19}$$

Theorem 1 *Let $\mathcal{W}$ denotes a nonempty convex set, then $\boldsymbol{U}_W = \boldsymbol{V}_{\tilde{H}}$ and $\boldsymbol{V}_W = \boldsymbol{U}_{\tilde{H}}$ are optimal for the problem (19), where $\boldsymbol{U}_{\tilde{H}}, \boldsymbol{V}_{\tilde{H}}$ are unitary matrices from the SVD of $\tilde{\boldsymbol{H}}$ and $\boldsymbol{U}_W, \boldsymbol{V}_W$ are obtained from the SVD of $\boldsymbol{W}$.*

The proof of Theorem 1 is similar to the proof of Theorem 2 in [41]. However, in [41], the scalar factor in MSE computation is omitted and the Theorem is proved for perfect and statistical CSI.

Consider the SVD of the composite channel as $\tilde{\boldsymbol{H}} = \boldsymbol{U}_{\tilde{H}} \boldsymbol{\Sigma}_{\tilde{H}} \boldsymbol{V}_{\tilde{H}}^H$ with $\boldsymbol{\Sigma}_{\tilde{H}} = \left[\boldsymbol{\Lambda}_{\tilde{H}} \, \mathbf{0}\right] \in \mathbb{C}^{\mathrm{N}_U \times \mathrm{N}_B}$ and $\boldsymbol{\Lambda}_{\tilde{H}} = diag(\lambda_{\tilde{H}}(1), \ldots, \lambda_{\tilde{H}}(\mathrm{N}_U)) \in \mathbb{C}^{\mathrm{N}_U \times \mathrm{N}_U}$ containing the singular values of the composite channel in decreasing order. Unitary matrices $\boldsymbol{U}_{\tilde{H}} \in \mathbb{C}^{\mathrm{N}_U \times \mathrm{N}_U}$ and $\boldsymbol{V}_{\tilde{H}} \in \mathbb{C}^{\mathrm{N}_B \times \mathrm{N}_B}$ are scaling and rotation matrices such that $\boldsymbol{U}_{\tilde{H}} \boldsymbol{U}_{\tilde{H}}^H = \mathbf{I}_{\mathrm{N}_U}$ and $\boldsymbol{V}_{\tilde{H}} \boldsymbol{V}_{\tilde{H}}^H = \mathbf{I}_{\mathrm{N}_B}$. Denote the SVD of $\boldsymbol{W} = \boldsymbol{U}_W \boldsymbol{\Sigma}_W \boldsymbol{V}_W^H$ with $\boldsymbol{\Sigma}_W = [\boldsymbol{\Lambda}_W \, \mathbf{0}]^T \in \mathbb{C}^{\mathrm{N}_B \times \mathrm{N}_U}$ and $\boldsymbol{\Lambda}_W = diag(\lambda_W(1), \ldots, \lambda_W(\mathrm{N}_U)) \in \mathbb{C}^{\mathrm{N}_U \times \mathrm{N}_U}$ containing the singular values of weights. Based on Theorem 1, the left and right singular vectors for the optimal precoder are equal to the right and left singular vectors of the composite channel which simplifies the norm in problem (19) as

$$\left\| \tilde{\boldsymbol{H}} \boldsymbol{W} - \mathbf{I} \right\|_F^2 = \left\| \boldsymbol{U}_{\tilde{H}} \boldsymbol{\Lambda}_{\tilde{H}} \boldsymbol{\Lambda}_W \boldsymbol{U}_{\tilde{H}}^H - \mathbf{I} \right\|_F^2. \tag{20}$$

Since the Frobenius norm is invariant with respect to unitary transformation [42], it is equivalent to

$$\left\| \boldsymbol{U}_{\tilde{H}} \boldsymbol{\Lambda}_{\tilde{H}} \boldsymbol{\Lambda}_W \boldsymbol{U}_{\tilde{H}}^H - \mathbf{I} \right\|_F^2 = \left\| \boldsymbol{\Lambda}_{\tilde{H}} \boldsymbol{\Lambda}_W - \mathbf{I} \right\|_F^2 = \left\| \boldsymbol{\lambda}_{\tilde{H}} . \boldsymbol{\lambda}_W - \mathbf{1}_{\mathrm{N}_U \times 1} \right\|_2^2, \tag{21}$$

where $\boldsymbol{\lambda}_{\tilde{H}} = \left[\lambda_{\tilde{H}}(1), \ldots, \lambda_{\tilde{H}}(\mathrm{N}_U)\right] \in \mathbb{C}^{\mathrm{N}_U \times 1}$ and $\boldsymbol{\lambda}_W = [\lambda_W(1), \ldots, \lambda_W(\mathrm{N}_U)] \in \mathbb{C}^{\mathrm{N}_U \times 1}$. Therefore, the section $F1$ is simplified to

$$\min_{\boldsymbol{\lambda}_W} \left\| \boldsymbol{\lambda}_{\tilde{H}} \cdot \boldsymbol{\lambda}_W - \mathbf{1}_{\mathrm{N}_U \times 1} \right\|_2^2. \tag{22}$$

Second, we try to minimize the section $F2$ of (18). In this regard, an auxiliary vector $\boldsymbol{\Phi}_{i,j}$, which is constructed by stacking the columns of $\boldsymbol{V}_{\tilde{H}} . \boldsymbol{U}_{\tilde{H}}^*$ one below the other, is defined, and the elements of $\boldsymbol{W}$ which are corresponding to idle links (zero elements of $\boldsymbol{S}$) are denoted as $\boldsymbol{Z} = \left(\mathbf{1}_{\mathrm{N}_B \times \mathrm{N}_U} - \boldsymbol{S}\right) . \boldsymbol{W}$, where

$$\boldsymbol{Z}(i,j) = \begin{cases} \boldsymbol{\Phi}_{i,j} \boldsymbol{\lambda}_W & \boldsymbol{S}(i,j) = 0, \; 1 \le i \le \mathrm{N}_B, 1 \le j \le \mathrm{N}_U \\ 0 & \text{otherwise} \end{cases}, \tag{23}$$

$$\boldsymbol{\Phi}_{i,j} = \left[V_{\tilde{H}}(i,1), \ldots, V_{\tilde{H}}(i,\mathrm{N}_U)\right] . \left[U_{\tilde{H}}^*(j,1), \ldots, U_{\tilde{H}}^*(j,\mathrm{N}_U)\right] \in \mathbb{C}^{1 \times \mathrm{N}_U}. \tag{24}$$

Using these definitions, section $F2$ of (18) is rewritten as

$$\min_{\boldsymbol{\lambda}_W} \left\| \tilde{\boldsymbol{H}} \boldsymbol{Z} \right\|_F^2. \tag{25}$$

To compute power constraint, $Tr\left(\hat{\boldsymbol{W}}^H \hat{\boldsymbol{W}}\right)$ is computed as

$$Tr\left(\hat{\boldsymbol{W}}^H \hat{\boldsymbol{W}}\right) = \left\|\hat{\boldsymbol{W}}\right\|_F^2 = \|\boldsymbol{W} - \boldsymbol{Z}\|_F^2 \leq P_t. \tag{26}$$

As $\hat{\boldsymbol{W}}$ and $\boldsymbol{Z}$ are disjoint elements of $\boldsymbol{W}$, the total transmission power, $\left\|\hat{\boldsymbol{W}}\right\|_F^2$ is computed as

$$\begin{aligned} &\left\|\hat{\boldsymbol{W}}\right\|_F^2 = \|\boldsymbol{S}.\boldsymbol{W} + (\boldsymbol{1} - \boldsymbol{S}).\boldsymbol{W}\|_F^2 = \left\|\hat{\boldsymbol{W}} + \boldsymbol{Z}\right\|_F^2 = \left\|\hat{\boldsymbol{W}}\right\|_F^2 + \|\boldsymbol{Z}\|_F^2 \Rightarrow \\ &\left\|\hat{\boldsymbol{W}}\right\|_F^2 = \|\boldsymbol{W}\|_F^2 - \|\boldsymbol{Z}\|_F^2. \end{aligned} \tag{27}$$

By considering (23) and (26), the total power constraint $C(1)$ can be rewritten as

$$Tr\left(\hat{\boldsymbol{W}}^H \hat{\boldsymbol{W}}\right) = \left\|\hat{\boldsymbol{W}}\right\|_F^2 = \|\boldsymbol{W}\|_F^2 - \|\boldsymbol{Z}\|_F^2 = \sum_{u=1}^{\mathrm{N}_U} \lambda_W^2(u) - \sum_{i,j|\boldsymbol{S}(i,j)=0} \left|\boldsymbol{\Phi}_{i,j}\boldsymbol{\lambda}_W\right|^2 \leq P_t. \tag{28}$$

Finally, by adding (25) to problem (22), the sparse precoding design problem is summarized as

$$\min_{\boldsymbol{\lambda}_W} \left\|\boldsymbol{\lambda}_{\tilde{H}} \cdot \boldsymbol{\lambda}_W - \boldsymbol{1}_{\mathrm{N}_U \times 1}\right\|_2^2 + \left\|\tilde{\boldsymbol{H}}\boldsymbol{Z}\right\|_F^2 \tag{29}$$

subject to:

$$\sum_{u=1}^{\mathrm{N}_U} \lambda_W^2(u) - \sum_{i,j|\boldsymbol{S}(i,j)=0} \left|\boldsymbol{\Phi}_{i,j}\boldsymbol{\lambda}_W\right|^2 \leq P_t. \tag{30}$$

In order to solve the above problem, the sequential quadratic programming (SQP) optimization method [43, 44] can be used. In this method, the search direction denoted as $\boldsymbol{d}$ is obtained by solving the following sub-problem

$$\min \nabla f(\boldsymbol{\lambda}_W)^T + \frac{1}{2}\boldsymbol{d}^T\boldsymbol{\phi}\boldsymbol{d} \tag{31}$$

subject to:

$$g(\boldsymbol{\lambda}_W) + \nabla g(\boldsymbol{\lambda}_W)^T\boldsymbol{d} \leq 0, \tag{32}$$

where $f(\boldsymbol{\lambda}_W)$ and $g(\boldsymbol{\lambda}_W)$ are the objective and constraint functions and $\boldsymbol{\phi}$ is a symmetric positive definitive matrix. By a starting point $\boldsymbol{\lambda}_W$, the new point is generated as $\overline{\boldsymbol{\lambda}_W} = \boldsymbol{\lambda}_W + \varepsilon\boldsymbol{d}$, where ε is a scalar non-negative step size.

3.2 Outer layer optimization: selection matrix design

Finding the optimal solution of the outer layer of the problem (11) might be too computationally cumbersome, as it involves Boolean constraints. One naive way to find $\boldsymbol{S}$ is to use exhaustively search among all possible combinations of the selection matrix for the one that gives the best MSE. Although the exhaustive search might be the only mechanism for a truly optimum selection of users to be served by each BS under load balancing constraint, its computational complexity grows quickly and it becomes impractical.

An alternative is to use the convex optimization techniques. Since the elements of $\boldsymbol{S}$ are binary variables, it makes the optimization problem *NP*-hard. To solve this optimization problem, it's possible to use the concept of linear programming relaxation [45], where the constraint that each element must be a binary variable is relaxed to a weaker constraint that each is a real number in the interval $[0, 1]$. The performance of convex optimization with relaxation is very close to the optimal one based on exhaustive search. Although its complexity is not as prohibitive as the exhaustive search, it is still high, being approximately in the order of $\mathcal{O}(n^3)$, where $\mathcal{O}$ presents the complexity and $n = \mathrm{N}_B\mathrm{N}_U$ [46].

To deal with this issue, a Greedy algorithm that attempts to approximate the optimal solution can be implemented. The Greedy algorithms have been widely applied in the framework of wireless communication, particularly in scheduling for CoMP systems, where the objective is to select the set of users that maximizes a certain metric function [47]. The Greedy algorithm which is guaranteed to converge has two main advantages. First, it allows a considerable reduction in complexity, requiring roughly $\mathcal{O}(n^2)$ operations. Second, it can be applied to a wide range of metrics of interest [48]. We use the following Greedy algorithm to find a local optimum for our problem

$$\min_{\boldsymbol{S}} f(\tilde{\boldsymbol{H}}, \boldsymbol{S}) \tag{33}$$

subject to $C(2)$ and $C(3)$, where the metric function $f(\tilde{\boldsymbol{H}}, \boldsymbol{S})$ is defined as

$$f(\tilde{\boldsymbol{H}}, \boldsymbol{S}) = \min_{\boldsymbol{\lambda}_W} \left\| \boldsymbol{\lambda}_{\tilde{H}} \cdots \boldsymbol{\lambda}_W - \mathbf{1}_{\mathrm{N}_U \times 1} \right\|_2^2 + \left\| \tilde{\boldsymbol{H}} \boldsymbol{Z} \right\|_F^2, \text{s.to (30)}. \tag{34}$$

The principle of the proposed Greedy algorithm is as follows. We start with an initial selection matrix obtained by randomly selecting $Q = \lfloor \mathrm{N}_B\mathrm{N}_U(1 - r_{fl}) \rfloor$ elements of $\boldsymbol{S}$ to be one and the remaining to be zero, providing the conditions $C(2)$ and $C(3)$ are fulfilled. Then, we select the first element among the zero-value elements and find the one-value element that when replaced with the selected zero-value leads to a reduction in the metric function. When this occurs, $\boldsymbol{S}$ is updated by replacing the zero-value element with the one-value that presents the largest reduction in the metric function. This process is repeated for other zero-value elements and therefore the selection matrix is designed in a way that the MSE is minimized. The algorithm is summarized in Algorithm 1.

Algorithm 1 Greedy algorithm for design the link selection matrix

Initialize $\boldsymbol{S} = random(\mathrm{N}_B, \mathrm{N}_U, Q)$: Randomly generate binary matrix of size $\mathrm{N}_B \times \mathrm{N}_U$ with Q ones that meets conditions $C(2)$ and $C(3)$
Compute SVD of composite channel, $\tilde{\boldsymbol{H}} = \boldsymbol{U}_{\tilde{H}} diag(\boldsymbol{\lambda}_{\tilde{H}}) \boldsymbol{V}_{\tilde{H}}$
Evaluate $\boldsymbol{\Phi}_{i,j} = \boldsymbol{V}_{\tilde{H}}(i,:).\boldsymbol{U}^*_{\tilde{H}}(:,j)$, **Solve** $f(\tilde{\boldsymbol{H}}, \boldsymbol{S})$ **and calculate** $\boldsymbol{\lambda}_W$
Evaluate metric $MSE = f(\tilde{\boldsymbol{H}}, \boldsymbol{S})$
for $i = 1 \rightarrow \mathrm{MAXRETRIES}$ **do**
    $\mathcal{S} = \{(i,j) | \boldsymbol{S}(i,j) = 0\}$
    $\bar{\mathcal{S}} = \{(i,j) | \boldsymbol{S}(i,j) = 1\}$
    for $j = 1 \rightarrow Q$ **do**
        $\boldsymbol{q} \leftarrow \boldsymbol{S}$
        for $k = 1 \rightarrow \mathrm{N}_B\mathrm{N}_U - Q$ **do**
            $\boldsymbol{q}(\mathcal{S}(k)) \leftarrow 1$
            $\boldsymbol{q}(\bar{\mathcal{S}}(k)) \leftarrow 0$
            solve $f(\tilde{\boldsymbol{H}}, \boldsymbol{q})$ **and calculate new** $\boldsymbol{\lambda}_W$
            $table(k) = f(\tilde{\boldsymbol{H}}, \boldsymbol{q})$
            $\boldsymbol{q} \leftarrow \boldsymbol{S}$
        end for
        if $\min(table) < MSE$ **then**
            $MSE = \min(table)$
            $\boldsymbol{S}(\mathcal{S}(\arg\min(table))) = 1$, $\boldsymbol{S}(\bar{\mathcal{S}}(j)) = 0$, $\mathcal{S}(\arg\min(table)) \longleftrightarrow \bar{\mathcal{S}}(j)$
        end if
    end for
    $\boldsymbol{S} = random(\mathrm{N}_B, \mathrm{N}_U, Q)$
end for

The computational complexity of the pseudo code in Algorithm 1 is computed in "Appendix 2". The overall complexity of the Algorithm 1 is $\mathcal{O}\left(C\mathrm{N}_B^2\mathrm{N}_U^3\right)$, where C is a constant. Then, the complexity is a function of the number of coordinated BSs, number of users and number of transmit and receive antennas. It increases with the square of the number of BSs and transmit antennas and with the third power of the number of users and receive antennas.

3.3 Limited feedback using CSI codebook

For a further reduction of the feedback load, we consider a CSI codebook based limited feedback strategy, where each user selects a codeword from a pre-designed CSI codebook and feeds back its index to the serving BS [7, 49]. The CCN collects all the codeword indexes sent from different BSs and calculates the precoding matrix. Different CSI codebooks are employed by the users, so they can independently quantize their per-BS channel direction information (CDI). The per-BS CDI for the u-th user can be expressed as $\tilde{\boldsymbol{h}}_{u,b} = \boldsymbol{h}_{u,b} / \left\|\boldsymbol{h}_{u,b}\right\|$, where $\boldsymbol{h}_{u,b} \in \mathbb{C}^{\mathrm{N}_r \times \mathrm{N}_t}$ is the CSI of the links spanning from the b-th BS to the u-th user, i.e. $\boldsymbol{h}_{u,b}(i,j) = \boldsymbol{H}_u(i,j)$, $1 \le i \le \mathrm{N}_r$, $(b-1)\mathrm{N}_t + 1 \le j \le b\mathrm{N}_t$. Random vector quantization (RVQ) is considered for quantizing the per-BS CDIs, where the quantized version of the CDI is given by [49] as

$$\hat{\boldsymbol{h}}_{u,b} = \arg \max_{\boldsymbol{c}_n \in \mathcal{C}_{u,b}} \left| \tilde{\boldsymbol{h}}_{u,b} \boldsymbol{c}_n^H \right|, \tag{35}$$

where $\mathcal{C}_{u,b}$ is the CSI codebook used by the u-th user to quantize the CDI of the b-th BS, which consists of $2^{\mathrm{B_{CDI}}}$ codewords. The codeword $\boldsymbol{c}_n \in \mathbb{C}^{\mathrm{N}_r \times \mathrm{N}_t}$, is a random vector with unit norm, and $\mathrm{B_{CDI}}$ denotes the number of bits used for quantizing the CDI.

We assume that the per-BS channel norm, $\left\|\boldsymbol{h}_{u,b}\right\|$ is perfectly known at the CCN, which is not included in the feedback information. The knowledge of these scales can be obtained at each BS by averaging the per-BS channels [50]. After aggregating of all the CDI indexes in the CCN, the CSI is reconstructed as

$$\hat{\boldsymbol{H}}_u = \left[\left\|\boldsymbol{h}_{u,1}\right\| \hat{\boldsymbol{h}}_{u,1}, \ldots, \left\|\boldsymbol{h}_{u,\mathrm{N}_b}\right\| \hat{\boldsymbol{h}}_{u,\mathrm{N}_b} \right]. \tag{36}$$

The global CDI quantization error is computed as

$$\varepsilon_u = 1 - \left| \tilde{\boldsymbol{H}}_u \hat{\boldsymbol{H}}_u^H \right|^2 = 1 - \left| \sum_{b=1}^{\mathrm{N}_b} \gamma_{u,b}^2 \mu_{u,b} e^{j\varphi_{u,b}} \right|^2, \tag{37}$$

where $\tilde{\boldsymbol{H}}_u = \left[\tilde{\boldsymbol{h}}_{u,1}, \ldots, \tilde{\boldsymbol{h}}_{u,\mathrm{N}_b} \right]$ is the global CDI vector, $\gamma_{u,b} = \frac{\|\boldsymbol{h}_{u,b}\|}{\sqrt{\sum_{b=1}^{\mathrm{N}_b} \|\boldsymbol{h}_{u,b}\|}}$ is the normalized per-BS channel norm and $\mu_{u,b} = \left| \tilde{\boldsymbol{h}}_{u,b} \hat{\boldsymbol{h}}_{u,b}^H \right|$ is the normalized per-BS quantization gain. The angle between the per-BS CDI and its codeword is denoted as $\varphi_{u,b}$, i.e. $e^{j\varphi_{u,b}} = \tilde{\boldsymbol{h}}_{u,b} \hat{\boldsymbol{h}}_{u,b}^H \Big/ \left\| \tilde{\boldsymbol{h}}_{u,b} \hat{\boldsymbol{h}}_{u,b}^H \right\|$, which is named as phase ambiguity (PA). As expected, in the perfectly quantized CDI condition, $\varepsilon_u = 0$.

In contrast to a single point transmission system, where the PA does not affect the CDI quantization performance, in coherent transmission, the PA affects the co-phasing of the system and degrades the feedback scheme performance. This is owing to the fact that the codeword selection in (35) only maximizes the magnitude and ignores its phase. The performance degradation due to the PA is more severe for cell-edge users [51].

The PA is uniformly distributed in $[-\pi, \pi]$ and employing a uniform quantizer for PA quantization is optimal [7]. In this regard, the PA can be fed back with aid of a few bits by using a scalar uniform quantizer. By considering $\mathrm{B_{PA}}$ bits to quantize the PA, the quantized PA is given by

$$\hat{\varphi}_{u,b} = \arg \min_{\phi_n} \left| \varphi_{u,b} - \phi_n \right|, \ \phi_n = n \frac{2\pi}{2^{\mathrm{B_{PA}}}} - \pi, \ n = 0, \ldots, 2^{\mathrm{B_{PA}}} - 1.$$

3.4 Receiver considerations

In this section, traditional receive filters are adopted for the transmission scheme. Receivers use stream specific pilots to estimate the effective channel, which includes the precoding weights and feedback (similar to implicit channel in the 3GPP standard). The effective channel for the u-th user is defined as

$$\bar{\boldsymbol{H}}_u = \boldsymbol{H}_u \hat{\boldsymbol{W}}_u \in \mathbb{C}^{\mathrm{N}_r \times \mathrm{N}_r}. \tag{38}$$

Common linear filters such as matched filter (MF) or minimum mean square error (MMSE) filter may be used in receivers to combat channel distortion and noise. Although motivated by aiming for a low-complexity receiver, we consider a receiver to use no filter at all, i.e. the receive filter at the u-th user is $\boldsymbol{g}_u^{No} = \boldsymbol{I}_{\mathrm{N}_r}$. In MF, the filter is designed to maximize the signal portion of the received signal, and as the signal to interference ratio is not minimized, it is useful in low-noise conditions [36]. In our sparse system, the MF receive filter for the u-th user is

$$\boldsymbol{g}_u^{MF} = \bar{\boldsymbol{H}}_u^H. \tag{39}$$

The MMSE receive filter is designed to minimize the MSE and finds a good tradeoff between the signal portion and the interference [52]. To compute the MMSE filter in the sparse system, by setting $\alpha_u = 1, u = 1, \ldots, \mathrm{N}_u$ in (5), the MSE for the u-th user can be expressed as

$$\mathrm{MSE}_u = E\Big\{ \|\boldsymbol{E}_u\|_2^2 \Big\} = E\big\{ tr\big(\boldsymbol{E}_u^H \boldsymbol{E}_u\big) \big\} = tr\Big(E\Big\{ \big(\boldsymbol{x}_u - \boldsymbol{g}_u \boldsymbol{y}_u\big)^H \big(\boldsymbol{x}_u - \boldsymbol{g}_u \boldsymbol{y}_u\big) \Big\} \Big). \tag{40}$$

To minimize the MSE, we can apply the trick of taking the conjugate complex derivative [53] w.r.t. $\boldsymbol{g}_u^H$ and set to zero as

$$\frac{\partial \mathrm{MSE}_u}{\partial \boldsymbol{g}_u^H} = \boldsymbol{g}_u E\big\{ \boldsymbol{y}_u \boldsymbol{y}_u^H \big\} - E\big\{ \boldsymbol{x}_u \boldsymbol{y}_u^H \big\} = 0, \tag{41}$$

where the expectation in the first term is the variance of the received signal by the u-th user and the second term is the cross-correlation of the data symbol with the received signal. Noting (1), these expectations are computed as follows

$$\boldsymbol{R}_{y_u} = E\big\{ \boldsymbol{y}_u \boldsymbol{y}_u^H \big\} = \underbrace{\bar{\boldsymbol{H}}_u \boldsymbol{R}_{x_u} \bar{\boldsymbol{H}}_u^H}_{\text{desired}} + \underbrace{\boldsymbol{H}_u \sum_{i \neq u} \Big(\hat{\boldsymbol{W}}_i \boldsymbol{R}_{x_i} \hat{\boldsymbol{W}}_i^H \Big) \boldsymbol{H}_u^H}_{\text{interference}} + \underbrace{\boldsymbol{R}_{n_u}}_{\text{noise}}, \tag{42}$$

$$E\big\{ \boldsymbol{x}_u \boldsymbol{y}_u^H \big\} = \boldsymbol{R}_{x_u} \bar{\boldsymbol{H}}_u^H, \tag{43}$$

where $\boldsymbol{R}_{x_u} = \sigma_x^2 \mathbf{I}_{\mathrm{N}_r}$ and $\boldsymbol{R}_{n_u} = \sigma_n^2 \mathbf{I}_{\mathrm{N}_r}$ are the variance of the data symbols and noise for the u-th user. Setting the derivative (41) to zero gives the MMSE receive filter in the sparse system as

$$\boldsymbol{g}_u^{\mathrm{MMSE}} = \boldsymbol{R}_{x_u} \bar{\boldsymbol{H}}_u^H \boldsymbol{R}_{y_u}^{-1}. \tag{44}$$

Note that the received signal variance according to (42) has three parts, including desired signal, interference and noise. The desired part can be computed from the effective channel directly, while the interference needs explicit channel estimation and some information about the precoding weights. However, it is possible to compute the received signal variance directly from the two-dimensional (frequency-time) received signal vector (similar to reference signals in LTE Release 14 [54, Section 6.10]).

Finally, scalar factor α in (5) is considered as the AGC gain and is used to adapt the input signal with the dynamic range of the analog to digital converter (ADC) [36]. Similar to

MMSE receive filter computation, the optimum value of the scalar factor can be calculated with minimization of MSE as

$$\alpha_u = \frac{tr\left(\boldsymbol{R}_{x_u}\tilde{\boldsymbol{H}}_u^H \boldsymbol{g}_u^H\right)}{tr\left(\boldsymbol{g}_u \boldsymbol{R}_{y_u} \boldsymbol{g}_u^H\right)}. \tag{45}$$

4 Schemes for comparison

4.1 Two-layer optimization with inner ZF precoder

In this section, we adopt the conventional ZF precoder as the inner precoder for the proposed two-layer optimization scheme. The ZF precoder is designed to remove the interference completely and has good performance in high SNRs [36]. Using the composite channel in (15), the ZF precoder can be designed as

$$\boldsymbol{W}^{\mathrm{ZF}} = \beta_{ZF}\tilde{\boldsymbol{H}}\left(\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^H\right),^{-1} \tag{46}$$

where β_{ZF} is used for power control, and in the sum-power constraint of P_t it is computed as

$$Tr\left(\left(\boldsymbol{W}^{\mathrm{ZF}}\right)^H \boldsymbol{W}^{\mathrm{ZF}}\right) \leq P_t \Rightarrow \beta_{ZF} = \sqrt{\frac{P_t}{tr\left(\left(\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^H\right)^{-1}\boldsymbol{R}_x\right)}}. \tag{47}$$

In the outer layer of the proposed two-layer optimization, we compute the metric function for the Greedy algorithm based on the inner ZF precoder and design the selection matrix in a way that the metric function is minimized. By substituting the closed form precoding matrix from (46) in (11), a new metric function for the Greedy algorithm is computed as

$$f(\tilde{\boldsymbol{H}}, \boldsymbol{S}) = \left\| \beta_{ZF}\tilde{\boldsymbol{H}}\left[\boldsymbol{S}\cdot\left(\tilde{\boldsymbol{H}}\left(\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^H\right)^{-1}\right)\right] - \boldsymbol{I}\right\|_F^2. \tag{48}$$

The following steps of the Greedy algorithm are similar to Algorithm 1, except substitution of the metric function with (48) and removing steps 2 and 3.

4.2 Two-layer optimization with inner Wiener precoder

Similar to the previous section, we adopt the conventional Wiener precoder as inner precoder for the proposed two-layer optimization scheme. The Wiener precoder minimizes the interference and maximizes the signal to interference and noise ratio (SINR). This precoder has better performance in comparison with ZF, especially in low SNRs [36]. The Wiener precoder is derived as

$$\boldsymbol{W}^{WF} = \beta_{\mathrm{WF}}\boldsymbol{F}^{-1}\tilde{\boldsymbol{H}}^H, \tag{49}$$

$$\boldsymbol{F} = \tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}} + \frac{tr\left(\boldsymbol{G}\boldsymbol{R}_n\boldsymbol{G}^H\right)}{P_t}\mathbf{I}_N, \tag{50}$$

where β_{WF} controls the transmitter power, and for the sum-power of P_t it is computed as

$$\beta_{\mathrm{WF}} = \sqrt{\frac{P_t}{tr\left(\boldsymbol{F}^{-2}\tilde{\boldsymbol{H}}^{H}\boldsymbol{R}_x\tilde{\boldsymbol{H}}\right)}}. \tag{51}$$

The metric function for the Greedy algorithm in the outer layer is computed by substituting the closed form precoding matrix (50) in (11) as

$$f(\tilde{\boldsymbol{H}}, \boldsymbol{S}) = \left\| \beta_{\mathrm{WF}}\tilde{\boldsymbol{H}}\left[\boldsymbol{S}\cdot\left(\boldsymbol{F}^{-1}\tilde{\boldsymbol{H}}^{H}\right)\right] - I\right\|_F^2. \tag{52}$$

In Algorithm 1, by substitution of the metric function with (52) and removing steps 2 and 3, the Greedy algorithm for the outer layer is obtained.

Note that for computing the Wiener precoding matrix in (50), it is required that the receive filter is known in the CCN, which is contradictory to designing the precoder using only the composite channel. Therefore, the Wiener precoder cannot be used directly as the inner precoder in the proposed transmission system, and here we consider it only for comparison purpose.

4.3 Selective feedback precoder

In the selective feedback technique [26], users with weak links are prevented from feeding back their CSI to the CCN and each user feeds back at least its strongest CSI. By exploiting a binary feedback index matrix, the coefficients of the channel matrix whose CSI is below a specified threshold are replaced with zeros. This technique can be categorized as an absolute thresholding approach for feedback load reduction.

To overcome the backhaul overhead related to routing users' data to several BSs, two schemes are proposed: one scheme based on MAC layer scheduling, and the other is based on the physical layer precoding. In this paper, we consider the latter, where by vectorization and eliminating of zero elements of the channel matrix, the precoder is designed using the ZF precoding approach.

4.4 SSOCP based relative thresholding precoder

In relative thresholding, users feed back only the CSI of links with channel value within a threshold relative to the strongest BS. In [27], an SSOCP based precoder for maximizing the weighted sum-rate is proposed in which the long term channel statistics are used to model the statistical interference for the unknown CSI. The precoder design problem with per antenna power constraint is considered as

$$\max_{\boldsymbol{W}} \prod_{u} (1+\gamma_u) \tag{53}$$

subject to:

$$\sum_{j=1}^{\mathrm{N}_U} \left|W(i,j)\right|^2 \le P_a, (b-1)\mathrm{N}_t + 1 \le i \le b\mathrm{N}_t, 1 \le b \le \mathrm{N}_b,$$

where γ_u is the SINR for the u-th user and P_a is a per antenna power constraint. For comparison with the sum-power constant of P_t in (12), we consider $P_t = \mathrm{N}_B P_a$.

5 Results and discussion

The numerical evaluation program is developed based on 3GPP time-coherent channel model [55] by MATLAB. The performance of the proposed scheme is evaluated by Monte-Carlo simulation. The inner optimization is performed using the SQP method and the outer layer is based on the Greedy algorithm. The simulation parameters are summarized in Table 2.

5.1 Channel model

Consider a JT-CoMP scenario where a set of $\mathrm{N_u} = 3$ single antenna users at the cluster center are being served by $\mathrm{N}_b = 3$ cooperating BSs with each $\mathrm{N}_t = 1,\ 2$ antennas. The cell radius is $R = 500$ m and the cell-edge SNR is variable. According to an example in Fig. 4, users are uniformly dropped at the cluster center, along an ellipse with semi-major and semi-minor axis of length $\frac{R}{16}$ and $\frac{h/2}{16}$, respectively where $h = \frac{\sqrt{3}}{2}R$ is the height of the hexagon of the cluster area. We consider the 3GPP channel model [55, 56]. The fading channel model includes the path-loss component $\gamma_{PL} = 128.1 + 37.6\log(R)$ (R is in km), $\gamma_{SF} = \mathcal{N}(0,\ 8\ \mathrm{dB})$ shadowing fading and a Rayleigh fast fading component Γ which is simulated as a circularly symmetric complex Gaussian random variable as $\mathcal{CN}(0,\ 1)$. The i.i.d channel between the BSs and the users is calculated as

$$\mathbf{H}_{iid} = \Gamma \mathbf{C}^{\frac{1}{2}} \sqrt{\mathrm{G}\,\gamma_{\mathrm{PL}}\,\gamma_{\mathrm{SF}}}, \tag{54}$$

where $\mathrm{G} = 1$ is the gain of the antennas at the BSs and $\mathbf{C} \in \mathbb{R}^{\mathrm{N_T} \times \mathrm{N_T}}$ is the correlation matrix of the antennas at the BSs, with the correlation between the antennas being $\rho = 0.5$ for all antenna pairs. We consider a time coherent channel model, where the CSI is varied only due to the effect of user movement, and the channel coefficient of the new CSI is based on Clarke's model [57]. The channel evolves in time as

Table 2 Simulation parameters

Parameters	Values
Number of antennas at BS\User	1, 2, 16, 32\1, 2
Receive filter (equalizer)	No, MF, MMSE
Channel model	3GPP channel model with path-loss $128.1 + 37.6\log(R)$, shadowing $\mathcal{N}(0,\ 8\ \mathrm{dB})$ and fast fading $\mathcal{CN}(0,\ 1)$
Channel realizations	10000
Cell radius	500 m
BS antenna gain\ correlation	1\ 0.5
Propagation delay	1 ms
User velocity	5 km/h
Carrier frequency	2 GHz
System bandwidth	10 MHz
Noise power	4×10^{-14} W
Maximum BS power in edge SNR of 10 dB	22.8 dBm (0.19 W)

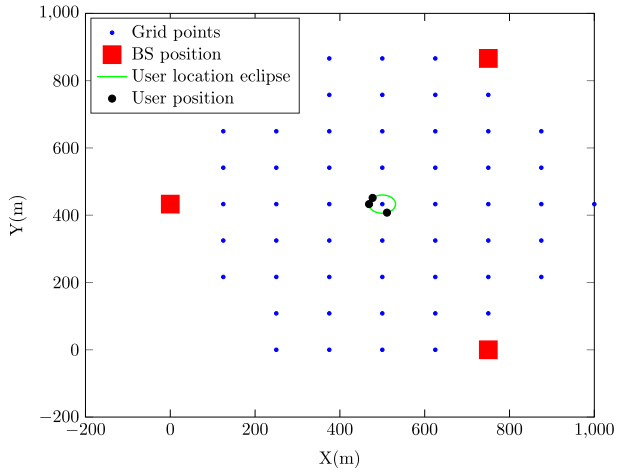

Fig. 4 The hexagon in the middle of the cells denotes the cluster area under consideration where the users are located at the cluster center

$$\mathbf{H}(t+\Delta t) = \sqrt{\rho}\boldsymbol{H}_{iid} + \sqrt{1-\rho}\boldsymbol{H}(t), \tag{55}$$

where $\rho = J_0(2\pi f_d \triangle t)$ is the channel correlation coefficient. Here, $J_0(.)$ is the zero-order Bessel function, the Doppler frequency is $f_d = \frac{vf_c}{c}$ with the velocity of the user being v, the carrier frequency is $f_c = 2$ GHz, c is the velocity of propagation, and $\triangle t$ is the evolved. The value of $\triangle t$ is considered 1 ms as the FDD uplink/downlink frame duration.

The receiver noise power is $\mathrm{N_0} = \mathrm{k_B T_0 B_n}$ Watts, where $\mathrm{k_B}$ is the Boltzmann's constant 1.38×10^{-23} Joules/Kelvin , $\mathrm{T_0} = 290$ Kelvin is the operating temperature, and $\mathrm{B}_n = 10$ MHz is the system bandwidth. The number of channel realization is 10^3 and maximum BS power with cell-edge $SNR = 10$ dB is 22.8 dBm or 0.19 W.

5.2 Simulation results

This section is devoted to the numerical evaluation of the performance of the designed JT-CoMP scheme. The general form of the network structure is depicted in Fig. 1. Time invariant and variant channel models are adapted from (54) and (55). To comprehensively evaluate the proposed scheme, we consider three stages. In the first stage, the proposed precoder is compared with the adopted ZF and Wiener precoders in Figs. 5 and 6, and comparison with selective feedback [26] and SSOCP based relative thresholding [27] precoders are performed in Fig. 9. In the second stage, performance of the proposed scheme is widely analyzed in Figs. 10, 11, and 12, where the effect of load reduction, probability distribution of MSE, and time convergence of the algorithm are investigated. In the third stage, the limited feedback effect on the proposed scheme is analyzed in Figs. 14 and 15.

5.2.1 Comparison to other schemes

In Fig. 5, performance of the adopted ZF, Wiener, and the proposed scheme are compared in a wide range of edge SNR in a receiver without receive filter. Note, in throughout the simulations, SNR is defined before any receive filter. The system is in

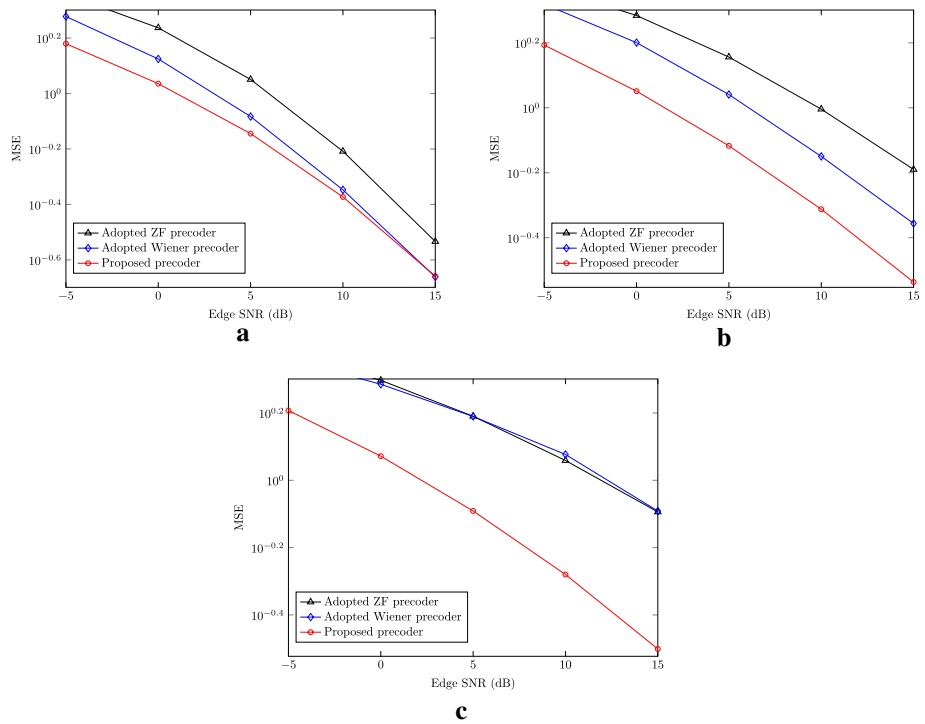

Fig. 5 Performance of the adopted ZF, Wiener and proposed precoders in terms of MSE in a receiver **a** without receive filter, **b** with MF receive filter, and **c** with MMSE receive filter in FFB-FBH configuration with $N_t = 1$. The channel is time-invariant

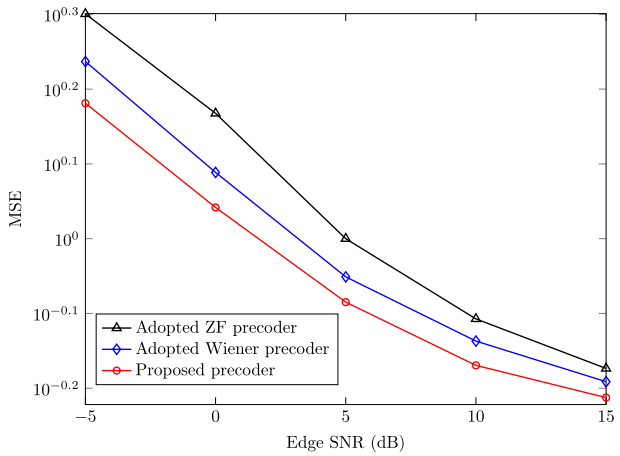

Fig. 6 Performance of the adopted ZF, adopted Wiener and the proposed precoders in terms of MSE in SFB-CBH configuration with $r_{fl} = r_{bl} = 0.33$, $N_t = 1$. The channel is time-variant with $\Delta t = 1$ ms and $v = 5$ km/h

full feedback (FFB)—full backhaul (FBH) configuration and the channel is time-invariant. We consider three types of receivers: receiver without receive filter, MF receiver, and MMSE receiver. In all receivers, the proposed scheme has better performance in comparison to adopted ZF and Wiener precoders. Performance of the proposed

scheme is not sensitive to the type of receiver, while performance of adopted ZF and Wiener precoders depends on the receiver filter. Therefore, a privilege of the proposed scheme is that, it has good performance with simplified receivers, such as receivers using no receive filter.

In Fig. 6, the performance of the proposed scheme is compared with the adopted ZF and Wiener precoders in sparse feedback (SFB) - constrained backhaul (CBH) configuration. We consider load reduction as $r_{fl} = r_{bl} = 0.33$ and the receiver is without receive filter. The channel is time-variant with $\triangle t = 1$ ms and $v = 5$ km/h. The achievable MSE is depicted for a wide range of edge SNRs. From the results shown in this figure, the proposed precoder outperforms adopted ZF and Wiener precoders for at least 25% from the MSE aspect. The superior performance of the proposed scheme is valid for the MF and MMSE receive filters, but due to the space limitation, the simulation results of other common filters are not shown.

By substitution of $\hat{\boldsymbol{W}}_u$ instead of $\boldsymbol{W}_u$ in (2), the SINR for the SFB-CBH configuration is computed. In Fig. 7, the cumulative density function (CDF) of the SINR of the proposed scheme is compared with the adopted ZF and Wiener precoders in edge SNRs of 5 dB and 10 dB. The load reduction is $r_{fl} = r_{bl} = 0.11$ and channel parameters are similar to Fig. 6. From the results shown in this figure, in edge SNR of 5 dB, the proposed precoder outperforms the adopted ZF and Wiener precoders with 6.49 dB and 3.85 dB on 80% point, respectively. In edge SNR of 10 dB, the superiority of the proposed precoder on others is 6.36 dB and 4.73 dB, respectively.

To evaluate the performance of the individual users in the proposed scheme, we define the MSE difference as $\triangle\text{MSE} = \text{MSE}_m - \text{MSE}_{\acute{m}}$, where the user m experiences the best MSE and the $\acute{m}$ experiences the worst one in a given channel realization. The computation of MSE at each user is described in "Appendix 1". The CDF of the MSE difference is shown in Fig. 8. Based on this result, the proposed scheme has less variance compared to the ZF precoder in SFB-CBH configuration. Although in this configuration, the CDF of MSE difference in the proposed precoder is slightly better than

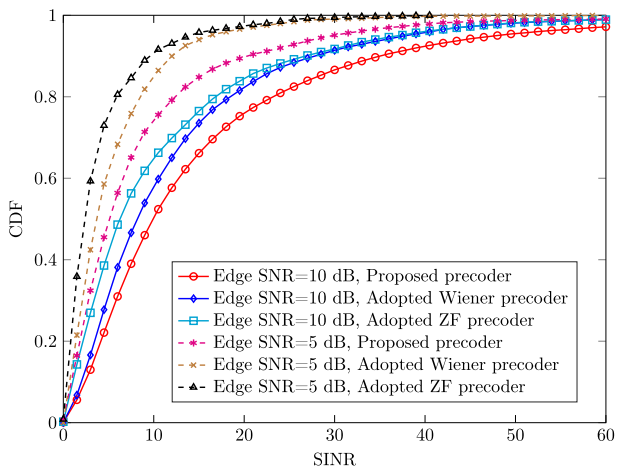

Fig. 7 CDF comparison of the proposed precoder with adopted ZF and Wiener precoders in terms of the users' SINR. The channel is time-variant and $r_{fl} = r_{bl} = 0.11, \mathrm{N}_t = 1, \mathrm{N}_r = 1, \Delta t = 1$ ms and $v = 5$ km/h

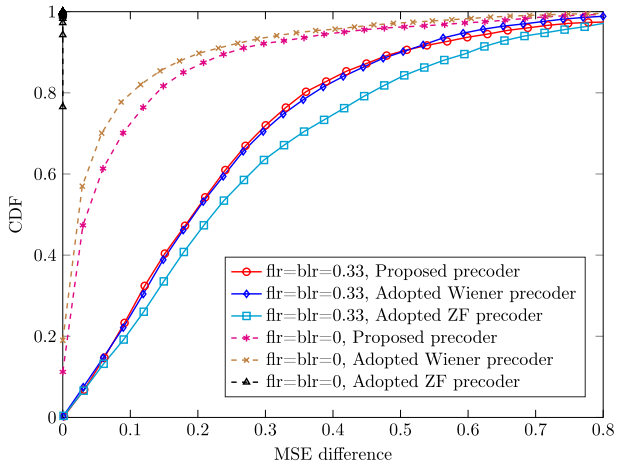

Fig. 8 CDF comparison of the proposed precoder with adopted ZF and Wiener precoders in terms of the users' MSE difference in SFB-CBH ($r_{fl} = r_{bl} = 0.33$) and FFB-FBH configurations and edge SNR of 10 dB. The channel is time-variant and $N_t = 1, N_r = 1, \Delta t = 1$ ms and $v = 5$ km/h

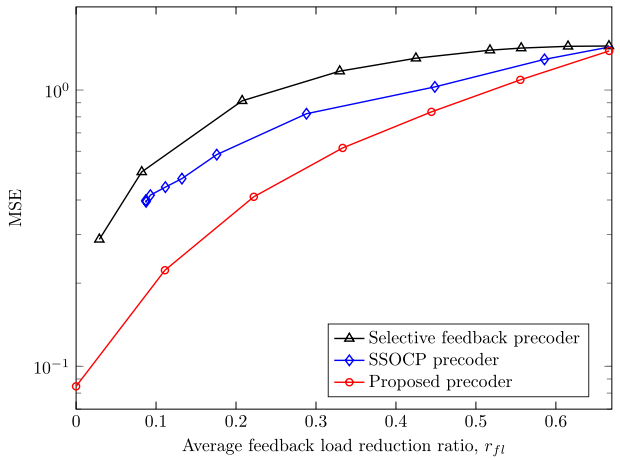

Fig. 9 Performance comparison of the proposed scheme with selective feedback and SSOCP based relative thresholding precoders in edge SNR of 10 dB in time-variant channel

the Wiener precoder. As expected, in the FFB-FBH configuration, the ZF precoder distributes an equal MSE to the users, hence the difference becomes zero.

In Fig. 9, the MSE of the proposed scheme is compared with selective feedback [26] and SSOCP based relative thresholding [27] precoders in a time-variant channel with edge SNR of 10 dB, $N_t = 1$ and SFB-CBH configuration. In the selective feedback precoder, to change r_{fl} from 0 to 60%, it is needed to change the absolute threshold level from −100 to −120 dB, while in the SSOCP precoder, the relative threshold level must be changed from 0 to 11 dB. Note that, in these precoders, only average loads can be controlled by adjusting the threshold value, while in the proposed precoder, the loads can be controlled strictly. As seen, the proposed scheme outperforms the selective feedback and SSOCP precoders for at least 30% from the MSE aspect.

Based on numerical evaluations, we can conclude that the proposed scheme has better MSE performance in comparison to the 3GPP Release 15 codebook type II precoder. These numerical comparisons are omitted here due to limited space. Although, in Release 16 and 17 more advanced and effective CSI reporting is possible.

5.2.2 Performance analysis of the proposed scheme

In Fig. 10, performance of the proposed scheme is compared in various configurations w.r.t to feedback and backhaul load reductions. The receiver is without receive filter and the channel is time-variant. As expected, the proposed precoder has the best performance in FFB-FBH configuration and with increasing the r_{fl} and r_{bl}, the system performance decreases. It is worth noting, for $r_{fl} = r_{bl} = 0.11$, the MSE increases as 40%. To evaluate the effect of backhaul load reduction alone, a sparse feedback-full backhaul (SFB-FBH) configuration is considered where $r_{fl} = 0.11$ and $r_{bl} = 0$. As expected, the proposed precoder has better performance in comparison to SFB-CBH with equal r_{fl} and slightly worse performance in comparison to FFB-FBH configuration.

In Fig. 11, the CDF of the MSE in the proposed scheme is showed for different feedback and backhaul load reduction values in SFB-CBH configuration. Edge SNR is 10 dB and the μ values in the legend show the average value. As seen, the average MSE is increased by increasing the feedback and backhaul load reductions.

Figure 12 depicts the convergence behavior of the proposed scheme. A SFB-CBH configuration with $r_{fl} = 0.11$ and $r_{bl} = 0.11$ is assumed and the MSE is shown for different edge SNRs. As seen, the scheme converges after transmitting an acceptable number of precoded data. In the worst case, the MSE converges after 5 subframes.

Based on numerical evaluations, the SINR performance of the proposed scheme in SFB-CBH configuration is slightly decreases with increasing $\triangle t$ that can be considered as CSI reporting period. These numerical comparisons are omitted here due to limited space.

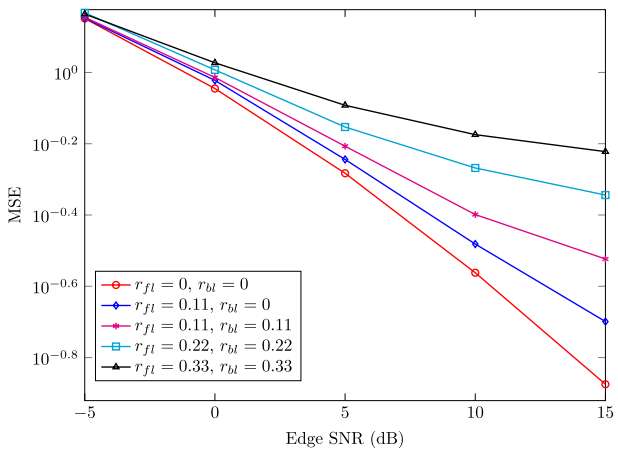

Fig. 10 MSE performance of the proposed scheme as function of edge SNR in terms of feedback load ratios in time-variant channel with $N_t = 1$

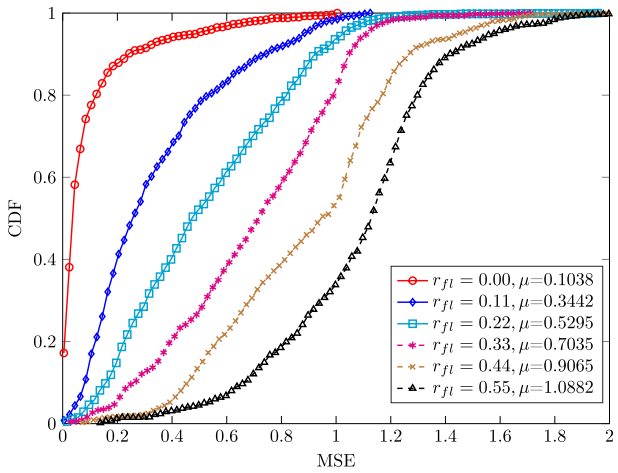

Fig. 11 Comparison of CDF of the SFB-CBH system in terms of MSE in time-variant channel. The cell-edge SNR is 10 dB. The μ values in the legend, shows the average value

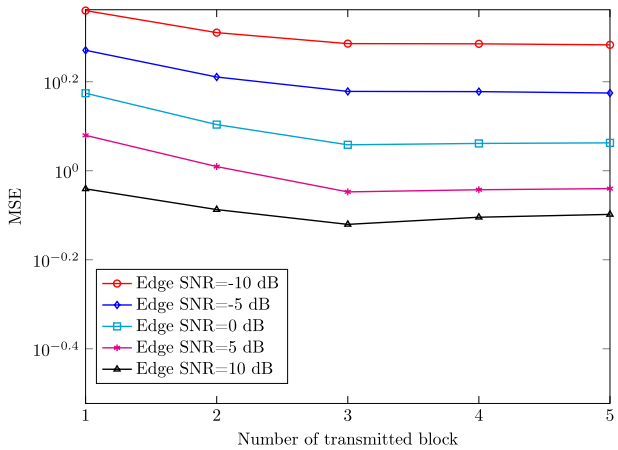

Fig. 12 MSE over time of SFB-CBH with $r_{fl} = r_{bl} = 0.11$in various cell-edge SNRs and time-variant channel

5.2.3 Performance of the proposed scheme in the CRAN network

To evaluate the performance of the proposed scheme in 5G and B5G systems, a scenario in the ultra-dense CRAN is considered, where in a square area of 400m $\times$ 400m, both users and BSs are uniformly distributed. To satisfy seamless coverage, the density of BSs is anticipated to come up to 40−50 BSs/km^2 [7, 58], therefore the number of BSs and users are set to 8 and 14 to have densities of 50 BSs/km^2 and 87 Users/km^2. Because of limitation in the maximum number of users that can be supported by each backhaul link, each user is assumed to be served with its nearest 3 BSs [7]. It is considered each BS has $N_t = 8, 16, 32$ transmit antennas and the users are equipped with $N_r = 2$ receive antennas. Figure 13 compares the performance of the proposed scheme in SFB-CBH ($r_{fl} = r_{bl} = 0.3$) and FFB-FBH configurations. As expected, the proposed precoder has the best performance when $N_t = 32$ and by decreasing the number of transmit antennas, the MSE increases. It is worth to note that when the

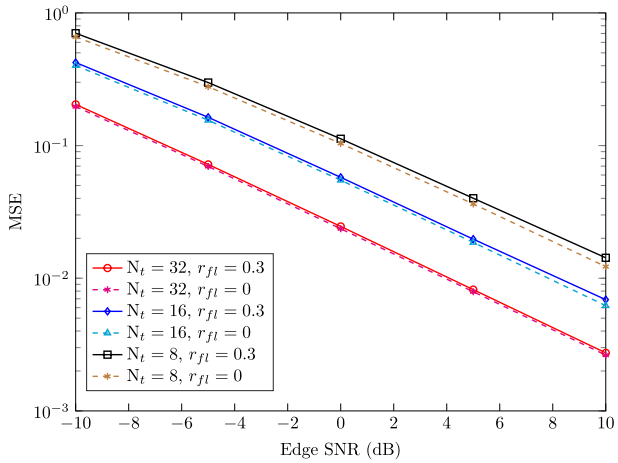

Fig. 13 MSE performance of the proposed precoder as function of edge SNR in various number of transmit antennas in SFB-CBH ($r_{fl} = r_{bl} = 0.3$) and FFB-FBH configurations. The channel is time-variant with $\Delta t = 1$ ms, $v = 5$ km/h and $N_r = 2$

number of transmit antennas is high, the performance degradation arising from feedback load reduction is negligible.

5.2.4 Feedback quantization effect

In Fig. 14, performance of the proposed quantization scheme in the SFB-CBH configuration with $r_{fl} = 0.167$, $N_t = 2$, $N_r = 1$, perfect PA, and varying bit number for CDI quantization is depicted. The achievable MSE is plotted for a wide range of edge SNRs in a receiver without receive filter. We can see a performance gap between the scheme of perfect CDI and of B_{CDI} bits quantization. However, with few bits for CDI quantization, the performance of the CSI codebook based feedback is significantly improved, and with $B_{CDI} = 8$ bits, the performance loss is negligible.

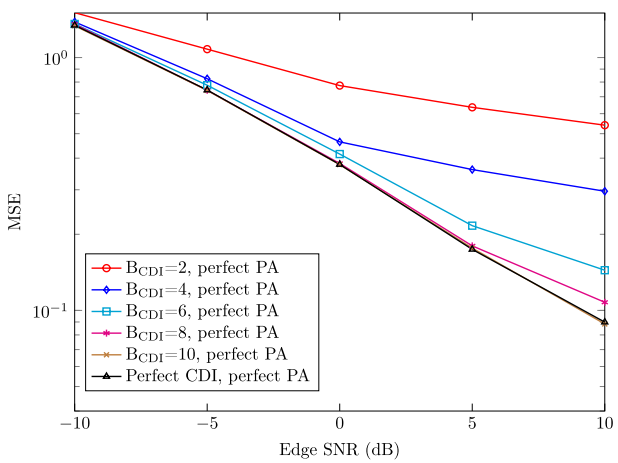

Fig. 14 MSE performance of the CSI codebook based quantization in SFB-CBH proposed scheme with $flr = 0.167$, $N_t = 2$, perfect PA and varying bit number for CDI quantization

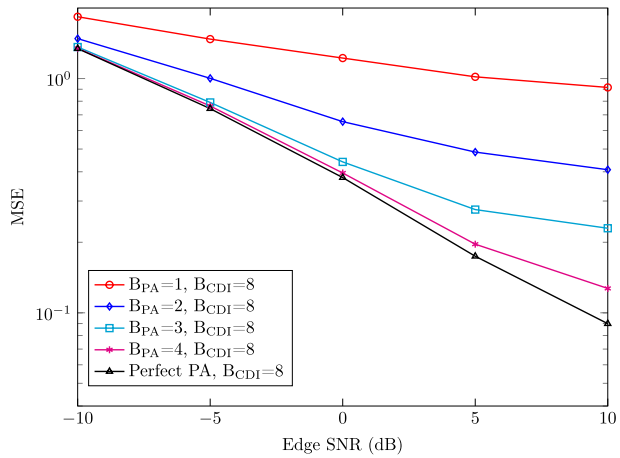

Fig. 15 Impact of bit number of PA quantization in MSE performance of the proposed scheme with $flr = 0.167$, $N_t = 2$ and $B_{CDI} = 8$ bits

Figure 15, shows the performance of the proposed quantization scheme in a similar system configuration stated for Fig. 14, with $B_{CDI} = 8$ bits for CDI quantization and varying number of bits for the PA quantization. From the results shown in this figure, when the B_{PA} increases from 1 to 4, the MSE is considerably decreased and in comparison to perfect PA, when $B_{PA} = 4$ bits, the performance loss is negligible.

The impact of the number of CDI and PA quantization bits on system performance implies that, only a small number of bits is necessary to benefit from a CSI codebook based quantization scheme in the proposed precoder. Especially, by considering the total number of bits for quantization of each link spanning from a BS to a user as $B = \left(\frac{B_{CDI}+B_{PA}}{N_t}\right)$, it is clear that by employing $B = 6$ bits quantization, a performance near to the full CSI feedback is attainable.

6 Conclusion

For a centralized JT-CoMP FDD downlink system, we designed and investigated the performance of a novel sparse feedback and constrained backhaul transmission scheme. To design the precoder matrix by providing feedback and backhaul load reductions, under a total power constraint and load balancing between BSs, a sub-optimum two-layer optimization method was proposed. In the inner layer, SVD of the CSI matrix was utilized to design the precoder matrix fulfilling a sum-power constraint and pre-known idle links. In the outer layer, the Greedy algorithm was exploited to design the link selection matrix, providing required feedback and backhaul load reductions and load balancing between BSs. In addition, sparse feedback and constrained backhaul schemes were introduced with adopting ZF and Wiener precoders. To further reduction of the feedback load, a CSI codebook based limited feedback strategy was considered. Numerical evaluations show a performance gain in terms of MSE of the proposed scheme, when compared to adopted ZF, Wiener, selective feedback and SSOCP based relative thresholding.

Appendix 1

The MSE at the u-th user is computed as

$$\begin{aligned}\mathrm{MSE}_u =& E\left\{\left\|\boldsymbol{x}_u-\alpha_u\widetilde{\boldsymbol{x}}_u\right\|_2^2\right\}=E\left\{\left(\boldsymbol{x}_u-\alpha_u\widetilde{\boldsymbol{x}}_u\right)\left(\boldsymbol{x}_u-\alpha_u\widetilde{\boldsymbol{x}}_u\right)^H\right\}\\ =&\left\{\boldsymbol{x}_u\boldsymbol{x}_u^H\right\}-\alpha_u^H E\left\{\boldsymbol{x}_u\widetilde{\boldsymbol{x}}_u^H\right\}-\alpha_u E\left\{\widetilde{\boldsymbol{x}}_u\boldsymbol{x}_u^H\right\}+\alpha_u\alpha_u^H E\left\{\widetilde{\boldsymbol{x}}_u\widetilde{\boldsymbol{x}}_u^H\right\}.\end{aligned} \tag{56}$$

As the user's data symbols have unit power variance, $E\left\{\boldsymbol{x}_u\boldsymbol{x}_u^H\right\}=\mathbf{I}_{\mathrm{N}_r}$. By considering the detected symbol at the u-th user as $\widetilde{\boldsymbol{x}}_u=\boldsymbol{g}_u\boldsymbol{y}_u$ and using (1), $E\left\{\boldsymbol{x}_u\widetilde{\boldsymbol{x}}_u^H\right\}$ and $E\left\{\widetilde{\boldsymbol{x}}_u\boldsymbol{x}_u^H\right\}$ are computed as

$$\begin{aligned}E\left\{\boldsymbol{x}_u\widetilde{\boldsymbol{x}}_u^H\right\}&=E\left\{\boldsymbol{x}_u\left(\boldsymbol{g}_u\boldsymbol{H}_u\sum_{i=1}^{\mathrm{N}_u}\boldsymbol{W}_i\boldsymbol{x}_i+\boldsymbol{g}_u\boldsymbol{n}_u\right)^H\right\}\\ &=E\left\{\left(\boldsymbol{x}_u\sum_{i=1}^{\mathrm{N}_u}\boldsymbol{x}_i^H\boldsymbol{W}_i^H\right)\boldsymbol{H}_u^H\boldsymbol{g}_u^H+\boldsymbol{x}_u\boldsymbol{n}_u^H\boldsymbol{g}_u^H\right\}\\ &=E\left\{\boldsymbol{x}_u\sum_{i=1}^{\mathrm{N}_u}\boldsymbol{x}_i^H\boldsymbol{W}_i^H\right\}\boldsymbol{H}_u^H\boldsymbol{g}_u^H+\boldsymbol{x}_u E\{\boldsymbol{n}_u^H\}\boldsymbol{g}_u^H=\boldsymbol{W}_u^H\boldsymbol{H}_u^H\boldsymbol{g}_u^H,\end{aligned} \tag{57}$$

$$\begin{aligned}E\{\widetilde{\boldsymbol{x}}_u\boldsymbol{x}_u^H\}&=E\left\{\left(\boldsymbol{g}_u\boldsymbol{H}_u\sum_{i=1}^{\mathrm{N}_u}\boldsymbol{W}_i\boldsymbol{x}_i+\boldsymbol{g}_u\boldsymbol{n}_u\right)\boldsymbol{x}_u^H\right\}\\ &=E\left\{\left(\boldsymbol{g}_u\boldsymbol{H}_u\sum_{i=1}^{\mathrm{N}_u}\boldsymbol{W}_i\boldsymbol{x}_i\boldsymbol{x}_u^H\right)+\boldsymbol{g}_u\boldsymbol{n}_u\boldsymbol{x}_u^H\right\}\\ &=\boldsymbol{g}_u\boldsymbol{H}_u E\left\{\sum_{i=1}^{\mathrm{N}_u}\boldsymbol{W}_i\boldsymbol{x}_i\boldsymbol{x}_u^H\right\}+\boldsymbol{g}_u E\{\boldsymbol{n}_u\}\boldsymbol{x}_u^H=\boldsymbol{g}_u\boldsymbol{H}_u\boldsymbol{W}_u,\end{aligned} \tag{58}$$

where the user data symbols are independent and zero mean. The term $E\left\{\widetilde{\boldsymbol{x}}_u\widetilde{\boldsymbol{x}}_u\right\}$ in (56) is the variance of the received signal which can be evaluated as

$$\begin{aligned}E\{\widetilde{\boldsymbol{x}}_u\widetilde{\boldsymbol{x}}_u\}&=E\left\{\boldsymbol{g}_u\left[\left(\boldsymbol{H}_u\sum_{i=1}^{\mathrm{N}_u}\boldsymbol{W}_i\boldsymbol{x}_i+\boldsymbol{n}_u\right)\left(\sum_{j=1}^{\mathrm{N}_u}\boldsymbol{x}_j^H\boldsymbol{W}_j^H\boldsymbol{H}_u^H+\boldsymbol{n}_u^H\right)\right]\boldsymbol{g}_u^H\right\}\\ &=E\left\{\boldsymbol{g}_u\left[\boldsymbol{H}_u\sum_{i=1}^{\mathrm{N}_u}\boldsymbol{W}_i\boldsymbol{x}_i\sum_{j=1}^{\mathrm{N}_u}\boldsymbol{x}_j^H\boldsymbol{W}_j^H\boldsymbol{H}_u^H+\boldsymbol{H}_u\sum_{i=1}^{\mathrm{N}_u}\boldsymbol{W}_i\boldsymbol{x}_i\boldsymbol{n}_u^H+\boldsymbol{n}_u\sum_{j=1}^{\mathrm{N}_u}\boldsymbol{x}_j^H\boldsymbol{W}_j^H\boldsymbol{H}_u^H+\boldsymbol{n}_u\boldsymbol{n}_u^H\right]\boldsymbol{g}_u^H\right\}\\ &=\boldsymbol{g}_u E\left\{\boldsymbol{H}_u\sum_{i=1}^{\mathrm{N}_u}\boldsymbol{W}_i\boldsymbol{x}_i\sum_{j=1}^{\mathrm{N}_u}\boldsymbol{x}_j^H\boldsymbol{W}_j^H\right\}\boldsymbol{g}_u^H+\boldsymbol{g}_u E\{\boldsymbol{n}_u\boldsymbol{n}_u^H\}\boldsymbol{g}_u^H\\ &=\boldsymbol{g}_u\left(\sum_{i=1}^{\mathrm{N}_u}(\boldsymbol{H}_u\boldsymbol{W}_i)(\boldsymbol{H}_u\boldsymbol{W}_i)^H\right)\boldsymbol{g}_u^H+\sigma_n^2\boldsymbol{g}_u\boldsymbol{g}_u^H.\end{aligned} \tag{59}$$

Finally, the MSE at the u-th user is obtained as

$$\begin{aligned}\mathrm{MSE}_u = \mathbf{I}_{\mathrm{N}_r} - \alpha_u\left(\boldsymbol{g}_u\boldsymbol{H}_u\boldsymbol{W}_u\right)^H - \alpha_u\boldsymbol{g}_u\boldsymbol{H}_u\boldsymbol{W}_u \\ + \alpha_u\alpha_u^H\left[\boldsymbol{g}_u\left(\sum_{i=1}^{\mathrm{N}_u}(\boldsymbol{H}_u\boldsymbol{W}_i)(\boldsymbol{H}_u\boldsymbol{W}_i)^H\right)\boldsymbol{g}_u^H + \sigma_n^2\boldsymbol{g}_u\boldsymbol{g}_u^H\right].\end{aligned} \tag{60}$$

Similarly, the sum MSE at all users is computed as

$$\begin{aligned}\mathrm{MSE} =& E\left\{\left\|\boldsymbol{x}-\boldsymbol{\alpha}\widetilde{\boldsymbol{x}}\right\|_2^2\right\} = E\left\{\left\|\boldsymbol{x}-\boldsymbol{\alpha}\boldsymbol{G}(\boldsymbol{H}\boldsymbol{W}\boldsymbol{x}+\boldsymbol{n})\right\|_2^2\right\} \\ =& E\left\{\left\|(\mathbf{I}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}\boldsymbol{W})\boldsymbol{x}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{n}\right\|_2^2\right\} = E\left\{((\mathbf{I}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}\boldsymbol{W})\boldsymbol{x}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{n})((\mathbf{I}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}\boldsymbol{W})\boldsymbol{x}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{n})^H\right\} \\ =& E\left\{(\mathbf{I}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}\boldsymbol{W})\boldsymbol{x}\boldsymbol{x}^H(\mathbf{I}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}\boldsymbol{W})^H\right\} - E\left\{(\mathbf{I}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}\boldsymbol{W})\boldsymbol{x}(\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{n})^H\right\} \\ & - E\left\{(\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{n})\boldsymbol{x}^H(\mathbf{I}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}\boldsymbol{W})^H\right\} + E\left\{(\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{n})(\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{n})^H\right\} \\ =& (\mathbf{I}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}\boldsymbol{W})E\left\{\boldsymbol{x}\boldsymbol{x}^H\right\}(\mathbf{I}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}\boldsymbol{W})^H + (\boldsymbol{\alpha}\boldsymbol{G})E\left\{\boldsymbol{n}\boldsymbol{n}^H\right\}(\boldsymbol{\alpha}\boldsymbol{G})^H \\ =& \|\mathbf{I}-\boldsymbol{\alpha}\boldsymbol{G}\boldsymbol{H}\boldsymbol{W}\|_F^2 + \sigma_n^2\|\boldsymbol{\alpha}\boldsymbol{G}\|_F^2,\end{aligned} \tag{61}$$

where $E\{\boldsymbol{n}\} = \mathbf{0}_{\mathrm{N}_U}$, $E\left\{\boldsymbol{n}\boldsymbol{n}^H\right\} = \sigma_n^2\mathbf{I}_{\mathrm{N}_U}$ and $E\left\{\boldsymbol{x}\boldsymbol{x}^H\right\} = \mathbf{I}_{\mathrm{N}_U}$.

Appendix 2

To compute the complexity of the SQP based inner loop optimization, the subproblem (31, 32) is considered, where the gradient of the objective function is computed as

$$\begin{aligned}\nabla f(\boldsymbol{\lambda}_w) =& \nabla\left[\left\|\boldsymbol{\lambda}_{\widetilde{H}}.\boldsymbol{\lambda}_W - \mathbf{1}_{\mathrm{N}_U\times 1}\right\|_2^2 + \left\|\widetilde{\boldsymbol{H}}\boldsymbol{Z}\right\|_F^2\right] = \nabla\left\|\boldsymbol{\lambda}_{\widetilde{H}}.\boldsymbol{\lambda}_W - \mathbf{1}_{\mathrm{N}_U\times 1}\right\|_2^2 + \nabla\left\|\widetilde{\boldsymbol{H}}\boldsymbol{Z}\right\|_F^2 \\ =& \nabla f_1(\boldsymbol{\lambda}_W) + \nabla f_2(\boldsymbol{\lambda}_W)\end{aligned} \tag{62}$$

By considering $\nabla f(\boldsymbol{\lambda}_W) = \left[\frac{\partial f(\boldsymbol{\lambda}_W)}{\partial\lambda_1},\ldots,\frac{\partial f(\boldsymbol{\lambda}_W)}{\partial\lambda_{\mathrm{N}_U}}\right]$, the $\nabla f_1(\boldsymbol{\lambda}_W)$ is computed as

$$\frac{\partial f_1(\boldsymbol{\lambda}_W)}{\partial\lambda_u} = \frac{\partial}{\partial\lambda_u}\sum_{i=1}^{\mathrm{N}_U}\left(\lambda_{\widetilde{H}}(i)\lambda_W(i)-1\right)^2 = 2\lambda_{\widetilde{H}}(u)\left(\lambda_{\widetilde{H}}(u)\lambda_W(u)-1\right) \tag{63}$$

$$\nabla f_1(\boldsymbol{\lambda}_W) = 2\boldsymbol{\lambda}_{\widetilde{H}}\cdot\left(\boldsymbol{\lambda}_{\widetilde{H}}.\boldsymbol{\lambda}_W - \mathbf{1}_{\mathrm{N}_U\times 1}\right) \tag{64}$$

By applying the chain rule and using the gradient of norm as $\nabla\|\boldsymbol{A}\boldsymbol{x}\|^2 = 2\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}$, the $\nabla f_2(\boldsymbol{\lambda}_W)$ is computed as

$$\nabla f_2(\boldsymbol{\lambda}_W) = 2\widetilde{\boldsymbol{H}}^H\widetilde{\boldsymbol{H}}\boldsymbol{Z}(\nabla\boldsymbol{Z}) = 2\widetilde{\boldsymbol{H}}^H\widetilde{\boldsymbol{H}}\boldsymbol{Z}\left[\left(\mathbf{1}_{\mathrm{N}_B\times\mathrm{N}_U}-\boldsymbol{S}\right).\left(\boldsymbol{V}_{\widetilde{H}}\boldsymbol{U}_{\widetilde{H}}^H\right)\right] \tag{65}$$

Similarly, the gradient of the constraint is computed as

$$\begin{aligned}\nabla g(\boldsymbol{\lambda}_W) =& \nabla\left[\sum_{u=1}^{\mathrm{N}_U}\lambda_W(u)^2 - \sum_{i,j|\boldsymbol{S}(i,j)=0}\left\|\boldsymbol{\Phi}_{i,j}\boldsymbol{\lambda}_W\right\|^2 - P_t\right] \\ =& 2\boldsymbol{\lambda}_W - \sum_{i,j|\boldsymbol{S}(i,j)=0}2\boldsymbol{\Phi}_{i,j}^H\boldsymbol{\Phi}_{i,j}\boldsymbol{\lambda}_W\end{aligned} \tag{66}$$

The computational complexity of $\nabla f_1(\boldsymbol{\lambda}_W)$ and $\nabla f_2(\boldsymbol{\lambda}_W)$ is $\mathcal{O}(\mathrm{N}_U)$ and $\mathcal{O}\left(\mathrm{N}_B\mathrm{N}_U^2\right)$ respectively, therefore the complexity of $\nabla f(\boldsymbol{\lambda}_W)$ is simplified to $\mathcal{O}\left(\mathrm{N}_B\mathrm{N}_U^2\right)$. The complexity of

$g(\boldsymbol{\lambda}_W)$ and $\nabla g(\boldsymbol{\lambda}_W)$ is $\mathcal{O}(C\mathrm{N}_U)$ and $\mathcal{O}\left(Q\mathrm{N}_U^2\right)$ respectively, where Q is the number of zero elements in the matrix $\boldsymbol{S}$. From the simulations, the SQP sub-problem converges within $C < 100$ iterations with no further improvement. By assuming $Q \ll \mathrm{N}_B\mathrm{N}_U$ overall computational complexity of solving the problem (29) can be simplified to $\mathcal{O}\left(C\mathrm{N}_B\mathrm{N}_U^2\right)$.

By using the computed complexity for inner loop optimization, the complexity of the Algorithm1 is computed. The complexity of the pseudo code presented in Algorithm 1 is

$$\mathcal{O}(B_1 + MAXRETRIES(B_2 + Q(B_3 + (\mathrm{N}_B\mathrm{N}_U - Q)B_4))) \tag{67}$$

where $MAXRETRIES$ is the number of outer iteration and the terms B_1, B_2, B_3 and B_4 present blocks of the pseudo code in steps $1-4$, $6-7, 22$, $9, 17-20$ and $11-15$, respectively. Based on the simulations, the Greedy algorithm for outer loop converges within $MAXRETRIES \leq 3$ iterations.

The initialization steps of the Algorithm 1 in block B_1 include computation of $\boldsymbol{S}$ with complexity of $\mathcal{O}(Q\mathrm{N}_B\mathrm{N}_U)$, computation of the SVD with complexity of $\mathcal{O}(Q\mathrm{N}_B\mathrm{N}_U \min(\mathrm{N}_B, \mathrm{N}_U))$, solving problem (29) with complexity of $\mathcal{O}\left(C\mathrm{N}_B\mathrm{N}_U^2\right)$ and evaluating the MSE with complexity of $\mathcal{O}\left(\mathrm{N}_B\mathrm{N}_U^2\right)$. By considering the worst-case complexity, the complexity of B_1 is $\mathcal{O}\left(\mathrm{N}_B\mathrm{N}_U^2\right)$. In blocks B_2 and B_3, the time and computational complexity can grow with $\mathrm{N}_B\mathrm{N}_U$, therefore the complexity of these blocks is $\mathcal{O}(\mathrm{N}_B\mathrm{N}_U)$. The complexity of the block B_4 mainly depends on solving the $f\left(\widetilde{\boldsymbol{H}}, q\right)$ in step 13 which is $\mathcal{O}\left(\mathrm{N}_B\mathrm{N}_U^2\right)$. Finally, by ignoring the lower-order terms, the overall computational complexity of the Algorithm 1 can be simplified to $\mathcal{O}\left(MAXRETRIES \times Q(\mathrm{N}_B\mathrm{N}_U - Q)C\mathrm{N}_B\mathrm{N}_U^2\right)$. By assuming $Q \ll \mathrm{N}_B\mathrm{N}_U$ and ignoring the small constants, the overall complexity is simplified to $\mathcal{O}\left(C\mathrm{N}_B\mathrm{N}_U^3\right)$.

Methods/experimental

The purpose of this study was to design a downlink in a centralized JT-CoMP system with sparse feedback and constrained backhaul links. The system consists of neighboring cells and users at the common boundary, or cluster area, in the middle of the cells. The channels are assumed to be time-variant following 3GPP channel model. The throughput of the system in terms of MSE was optimized using a two-layer method including an inner SVD precoder design and an outer Greedy link selection approach. Furthermore, sparse feedback and constrained backhaul schemes based on ZF and Wiener precoders were defined and used as benchmark for the proposed scheme.

Abbreviations

JT-CoMP: Joint transmission coordinated multipoint; BS: Base station; CSI: Channel state information; MSE: Mean square error; SVD: Singular value decomposition; SFB: Sparse feedback; CBH: Constrained backhaul; ZF: Zero forcing; FDD: Frequency division duplex; CCN: Central coordination node; SNR: Signal to noise ratio; TDD: Time division duplex; AGC: Automatic gain control; RSSI: Received signal strength indicator; MF: Matched filter; MMSE: Minimum mean square error; ADC: Analog to digital converter; WF: Wiener filter; SSOCP: Successive second order cone programming; SNIR: Signal to noise and interference ratio; FFB: Full feedback; FBH: Full backhaul; CDF: Cumulative density function.

Authors' contributions

The individual contributions of each authors to the manuscript are the same. All authors read and approved the final manuscript.

Funding

The authors declare that there is no fund of this work.

Availability of data and materials

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Institute for Research in Fundamental Sciences, Tehran, Iran. [2] Department of Electrical Engineering, University of Qom, Qom, Iran. [3] Department of Signals and Systems, Chalmers University of technology, Gothenburg, Sweden.

Received: 9 July 2020 Accepted: 30 March 2021
Published online: 23 April 2021

References
1. K. Manolakis, V. Jungnickel, C. Oberli, T. Wild, V. Braun, N. Vucic, M. Castaneda, Cooperative cellular networks: overcoming the effects of real-world impairments. IEEE Veh. Technol. Mag. **10**(3), 30–40 (2015)
2. V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfröjd, T. Svensson, The role of small cells, coordinated multipoint, and massive MIMO in 5g. IEEE Commun. Mag. **52**(5), 44–51 (2014)
3. A. Osseiran, J. Monserrat, P. Marsch, O. Queseth, H. Tullberg, M. Fallgren, K. Kusume, A. Höglund, H. Droste, I. Silva, P. Rost, M. Boldi, J. Sachs, P. Popovski, D. Gozalvez-Serrano, P. Fertl, Z. Li, F. Moya, G. Fodor, J. Lianghai, *5G Mobile and Wireless Communications Technology*, vol. 06 (Cambridge University Press, Cambridge, 2016)
4. S.M.R. Islam, N. Avazov, O.A. Dobre, K. Kwak, Power-domain non-orthogonal multiple access (noma) in 5g systems: potentials and challenges. IEEE Commun. Surv. Tutor. **19**(2), 721–742 (2017)
5. M. Hashemi, M. Coldrey, M. Johansson, S. Petersson, Integrated access and backhaul in fixed wireless access systems, in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, pp. 1–5 (2017)
6. A. Checko, H.L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M.S. Berger, L. Dittmann, Cloud ran for mobile networks—a technology overview. IEEE Commun. Surv. Tutor. **17**(1), 405–426 (2015)
7. C. Pan, H. Ren, M. Elkashlan, A. Nallanathan, L. Hanzo, Robust beamforming design for ultra-dense user-centric c-ran in the face of realistic pilot contamination and limited feedback. IEEE Trans. Wirel. Commun. **18**(2), 780–795 (2019)
8. 3GPP, Coordinated multi-point operation for lte; 3gpp tr 36.819 v11.0.0, 3GPP TSG RAN WG1, Tech. Rep., Sept. (2011)
9. Qualcomm, How can comp extend 5g nr to high capacity and ultra-reliable communications? Tech. Rep. (2018)
10. 3GPP, Technical specification group services and system aspects; release 15 description; 3gpp tr 21.915, v15.0.0, Tech. Rep., 2019-09
11. 3GPP, Technical specification group radio access network; study on further enhancements to coordinated multi-point (comp) operation for lte; 3gpp tr 36.741, v14.0.0, Tech. Rep., 2017-03
12. D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, K. Sayana, Coordinated multipoint transmission and reception in lte-advanced: deployment scenarios and operational challenges. IEEE Commun. Mag. **50**(2), 148–155 (2012)
13. J. Li, A. Papadogiannis, R. Apelfröjd, T. Svensson, M. Sternad, Performance evaluation of coordinated multi-point transmission schemes with predicted CSI, in *IEEE 23rd International Symposium on Personal. Indoor and Mobile Radio Communications—(PIMRC) 2012*, pp. 1055–1060 (2012)
14. Z. Mayer, J. Li, A. Papadogiannis, T. Svensson, On the impact of control channel reliability on coordinated multi-point transmission. EURASIP J. Wirel. Commun. Netw. **2014**(1), 28 (2014). https://doi.org/10.1186/1687-1499-2014-28
15. K. Manolakis, V. Jungnickel, C. Oberli, T. Wild, V. Braun, Impairments in cooperative mobile networks: Models, impact on performance and mitigation, in *European Wireless 2014; 20th European Wireless Conference*, pp. 1–8 (2014)
16. K. Manolakis, C. Oberli, V. Jungnickel, F. Rosas, Analysis of synchronization impairments for cooperative base stations using OFDM. Int. J. Antennas Propag. **2015**, 1–14 (2015)
17. B. Makki, J. Li, T. Eriksson, T. Svensson, Throughput analysis for multi-point joint transmission with quantized CSI feedback, in *76th IEEE Vehicular Technology Conference, VTC Fall 2012; Quebec City, QC; Canada; 3 September 2012 through 6 September 2012*, pp. 1 – 5 (2012)
18. L. Shi, Z. Hu, T. Zhang, Z. Zeng, Performance analysis of delayed limited feedback based on per-cell codebook in comp systems, in *2015 IEEE Wireless Communications and Networking Conference, WCNC 2015, New Orleans, LA, USA, March 9–12, 2015*, pp. 363–368 (2015)
19. S. Zhou, J. Gong, Z. Niu, Distributed adaptation of quantized feedback for downlink network MIMO systems. IEEE Trans. Wireless Commun. **10**(1), 61–67 (2011)
20. B. Makki, J. Li, T. Eriksson, T. Svensson, Throughput analysis for multi-point joint transmission with quantized CSI feedback. IEEE Veh. Technol. Conf. (VTC Fall) **2012**, 1–5 (2012)
21. B. Makki, J. Li, T. Eriksson, T. Svensson, Coordinated multi-point joint transmission with partial channel information feedback, in *European Wireless 2013; 19th European Wireless Conference*, pp. 1–5 (2013)
22. M. Lossow, S. Jaeckel, V. Jungnickel, V. Braun, Efficient mac protocol for JT comp in small cells. IEEE Int. Conf. Commun. Workshops ICC **2013**, 1166–1171 (2013)
23. T.R. Lakshmana, J. Li, C. Botella, A. Papadogiannis, T. Svensson, Scheduling for backhaul load reduction in comp. IEEE Wirel. Commun. Netw. Conf. WCNC **2013**, 227–232 (2013)
24. D. Marabissi, G. Bartoli, R. Fantacci, M. Pucci, An optimized comp transmission for a heterogeneous network using EICIC approach. IEEE Trans. Veh. Technol. **65**(10), 8230–8239 (2016)
25. A. Papadogiannis, H.J. Bang, D. Gesbert, E. Hardouin, Downlink overhead reduction for multi-cell cooperative processing enabled wireless networks, in *IEEE 19th International Symposium on Personal. Indoor and Mobile Radio Communications*, Sept 2008, 1–5 (2008)
26. A. Papadogiannis, H.J. Bang, D. Gesbert, E. Hardouin, Efficient selective feedback design for multicell cooperative networks. IEEE Trans. Veh. Technol. **60**(1), 196–205 (2011)

27. T.R. Lakshmana, A. Tölli, R. Devassy, T. Svensson, Precoder design with incomplete feedback for joint transmission. IEEE Trans. Wirel. Commun. **15**(3), 1923–1936 (2016)
28. Y. Yu, T. Hsieh, A. Pang, Millimeter-wave backhaul traffic minimization for comp over 5g cellular networks. IEEE Trans. Veh. Technol. **68**(4), 4003–4015 (2019)
29. 3GPP, Technical specification group radio access network; physical channels and modulation, release 10, ts 36.211 v 12.5.0, Tech. Rep., March 2015
30. The 5g evolution: 3gpp releases 16–17, January 2020. https://www.5gamericas.org/5g-evolution-3gpp-releases-16-17, Tech. Rep
31. 3GPP, Nr; physical layer procedures for data, ts 38.214 v15.4.0, Tech. Rep., December 2018
32. R. Ahmed, F. Tosato, M. Maso, "Overhead reduction of NR type ii CSI for NR release 16, in *WSA 2019; 23rd International ITG Workshop on Smart Antennas*, pp. 1–5 (2019)
33. M. Taki, M.B. Nezafati, Delay constrained throughput optimisation with imperfect CSI using discrete adaptive power. Int. J. Electron. **105**(12), 2033–2051 (2018). https://doi.org/10.1080/00207217.2018.1494330
34. M. Taki, M.B. Nezafati, Integrated scheduling and link adaptation for heterogeneous networks: design and performance analysis. Int. J. Electron. (2019). https://doi.org/10.1080/00207217.2019.1703222
35. M. Taki, T. Svensson, M.B. Nezafati, Delay constrained throughput optimization in multi-hop AF relay networks, using limited quantized CSI. EURASIP J. Wirel. Commun. Netw. **2019**(1), 102 (2019). https://doi.org/10.1186/s13638-019-1423-3
36. M. Joham, W. Utschick, J.A. Nossek, Linear transmit processing in MIMO communications systems. IEEE Trans. Signal Process. **53**(8), 2700–2712 (2005)
37. P. Xiao, M. Sellathurai, Improved linear transmit processing for single-user and multi-user MIMO communications systems. IEEE Trans. Signal Process. **58**(3), 1768–1779 (2010)
38. S. Wahls, H. Boche, Linear lIR-MMSE precoding for frequency selective MIMO channels," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3264–3267 (2011)
39. D. Sacristán-Murga, M. Payaró, A. Pascual-Iserte, Robust linear precoding for MSE minimization in MIMO broadcast systems with channel gram matrix feedback, in *2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications*, pp. 341–345 (2011)
40. L. You, L. Lei, D. Yuan, Load balancing via joint transmission in heterogeneous LTE: modeling and computation, in *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1173–1177 (2015)
41. J. Wang, D.P. Palomar, Robust MMSE precoding in MIMO channels with pre-fixed receivers. IEEE Trans. Signal Process. **58**(11), 5802–5818 (2010)
42. F. Zhang, *Matrix Theory Basic Results and Techniques* (Springer, New York, 2011)
43. P. Boggs, J. Tolle, Sequential quadratic programming. Acta Numer. **4**, 1–51 (1995)
44. Z. Zhu, An efficient sequential quadratic programming algorithm for nonlinear programming. J. Comput. Appl. Math. **175**(2), 447–464 (2005)
45. A. Schrijver, *Theory of Linear and Integer Programming* (Wiley, New York, NY, 1998)
46. S. Joshi, S. Boyd, Sensor selection via convex optimization. Trans. Sig. Proc. **57**(2), 451–462 (2009). https://doi.org/10.1109/TSP.2008.2007095
47. S. Sigdel, W. A. Krzymien, M. Al-Shalash, Greedy and progressive user scheduling for comp wireless networks, pp. 4218–4223 (2012)
48. K. Elkhalil, A. Kammoun, T. Y. Al-Naffouri, M. S. Alouini, A blind antenna selection scheme for single-cell uplink massive MIMO, in *2016 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6 (2016)
49. F. Yuan, C. Yang, Bit allocation between per-cell codebook and phase ambiguity quantization for limited feedback coordinated multi-point transmission systems. IEEE Trans. Commun. **60**(9), 2546–2559 (2012)
50. R. Bhagavatula, R.W. Heath, Adaptive limited feedback for sum-rate maximizing beamforming in cooperative multicell systems. IEEE Trans. Signal Process. **59**(2), 800–811 (2011)
51. D. Su, X. Hou, C. Yang, Quantization based on per-cell codebook in cooperative multi-cell systems. IEEE Wirel. Commun. Netw. Conf. **2011**, 1753–1758 (2011)
52. E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, H. Poor, *MIMO Wireless Communications* (Cambridge University Press, Cambridge, 2007)
53. K. B. Petersen, M. S. Pedersen, J. Larsen, K. Strimmer, L. Christiansen, K. Hansen, L. He, L. Thibaut, M. Barão, S. Hattinger, V. Sima, and W. The, "The matrix cookbook,"Tech. Rep. (2006)
54. 3GPP, Evolved universal terrestrial radio access (e-utra); physical channels and modulation, TS 36.211 V14.4.0, 2017-09
55. 3GPP, Evolved universal terrestrial radio access, radio frequency system scenarios, TR 36.942-a20 Release 10 (2015)
56. J. Anderson, A. Svensson, *Coded Modulation Systems*, ser. Information Technology Series. Springer US, 2003. [Online]. https://books.google.com/books?id=oP-xAwgjr5AC
57. T. Rappaport, A.O.M.C. Safari, *Wireless Communications: Principles and Practice*, ser. Prentice Hall communications engineering and emerging technologies series. Prentice Hall PTR, 2002. [Online]. https://books.google.com/books?id=TbgQAQAAMAAJ
58. M. Yu, A. Tang, X. Wang, C. Han, Joint scheduling and power allocation for 6g terahertz mesh networks, in *2020 International Conference on Computing, Networking and Communications (ICNC)*, pp. 631–635 (2020)

Publisher's Note