Performance Optimization for Multicast MmWave MIMO Networks with Mobile Users

Songling Zhang*, Mingzhe Chen†, Wenjing Zhang*, Zhaohui Yang‡, Danpu Liu*,
Zhilong Zhang*, and Kai-Kit Wong§
*Beijing Key Laboratory of Network System Architecture and Convergence,
Beijing University of Posts and Telecommunications, Beijing, China
†Department of Electrical and Computer Engineering and Institute for Data Science and Computing,
University of Miami, Coral Gables, FL, 33146, USA
‡Zhejiang University, Hangzhou 310027, China, and with Zhejiang Lab, Hangzhou 31121, China
§Department of Electronic and Electrical Engineering, University College London, London, WC1E 6BT, UK
Emails:{slzhang, zhangwenjing, dpliu, zhangzhilong}@bupt.edu.cn, mingzhe.chen@miami.edu, yang_zhaohui@zju.edu.cn,
kai-kit.wong@ucl.ac.uk

***Abstract*—In this paper, the problem of maximizing the sum rate of mobile users in a multi-base station (BS) cooperative millimeter-wave (mmWave) multicast communication system is studied. In the considered model, due to the real-time mobility of users, the users being served by a given BS and beamforming of BSs and users are dynamic. Multiple BSs must cooperate to serve dynamic requests of multiple mobile users. This problem is posed as an optimization framework whose goal is to maximize the sum rate of all mobile users by jointly optimizing the number of users served by all BSs and beamforming matrices of both BSs and users. To solve this non-convex optimization problem, we first introduce a value decomposition based reinforcement learning (VD-RL) algorithm to determine the users to be served by each BS. Then, we use the block diagonalization method to obtain the fully digital transmit beamforming matrices of all BSs as well as the receive beamforming matrices of the users. Finally, a fast optimization algorithm is used to optimize the hybrid beamforming matrices of both BSs and users. Simulation results show that, the proposed algorithm can achieve up to 51% gain in terms of the sum rate of all mobile users compared to baseline multi-agent algorithms.**

I. Introduction

Millimeter wave (mmWave) communication is a key enabling technology for wireless mobile communication, as the abundant spectrum of mmWave band dramatically improves data rates [1]. However, mmWave communication with higher carrier frequency experiences higher path loss compared to lower carrier frequency communication [2]. Massive multiple-input multiple-output (MIMO) can be employed to accurately align the beamforming of the transceiver antennas to overcome serious path loss and improve both spectral and energy efficiency [3]. However, the number of radio frequency (RF) chains required for traditional fully digital beamforming increases with the number of antennas, resulting in high energy consumption and hardware costs. Therefore, in mmWave systems, the widespread adoption of effective hybrid beamforming contributes to significantly reducing energy consumption and hardware costs, whilst maintaining a satisfactory signal-to-noise ratio (SNR) in the communication link. However, optimizing hybrid beamforming in mmWave massive MIMO systems faces several challenges such as the design of the constant-amplitude RF beamforming matrices, and the joint optimization of the coupled digital beamforming matrices and RF beamforming matrices.

A number of works have studied the design of hybrid beamforming for base station (BS) and users in mmWave unicast networks. The work in [4] considered hybrid beamforming for mmWave MIMO communication system. The authors in [5] jointly optimized the hybrid analog-digital beamforming and the reconfigurable intelligent surface (RIS) reflection matrix to minimize the sum-mean-square-error in the RIS assisted mmWave multi-user MIMO system. In [6], a proficient beam and channel tracking strategy was established within a reconfigurable hybrid beamforming structure for broadband mmWave MIMO systems. Furthermore, in [7], a multi-user hybrid architecture was proposed for mmWave MIMO communication systems. In [8], a novel two-timescale analog-digital hybrid beamforming scheme was proposed for full-duplex (FD) mmWave MIMO multiple-relay systems. In [9], the authors studied the user association, hybrid beamforming, and fronthaul compression strategies for cell-free mmWave MIMO systems. In [10], the authors focused on the creation of beam squint-aware channel covariance-based hybrid beamformers for mmWave massive MIMO-orthogonal frequency division multiplexing (OFDM) systems.

Several existing studies [11]–[15] have considered beamforming in mmWave multicast systems. The authors in [11] investigated the robust hybrid beamforming architecture for multigroup mmWave multicast transmission. Besides, in [12], the authors studied the hybrid beamforming design for covert multicast mmWave communications. In [13], a low-complexity multicast beamforming design was proposed for mmWave communications. The authors in [14] studied energy efficient analog beamforming in mmWave single-group multicast communication systems. Considering the

This work is supported in part by the National Natural Science Foundation of China under Grant No. 62271065, U22B2001, the National Key R & D Program of China under Grant No. 2023YFB2904804, the National Natural Science Foundation of China (NSFC) under Grants 62394292, 62394290.

Fig. 1: The considered multigroup multicast MIMO communication system.

device-to-device (D2D) communication [15], an efficient multicast scheduling scheme was proposed for mmWave systems, where D2D communications in close proximity, concurrent transmissions, and the multi-level antenna codebook are exploited to improve multicast efficiency. However, these works [11]–[15] only consider static users. Considering user mobility in mmWave multicast systems will introduce several new challenges such as dynamic BS-user associations and user clustering, and real-time optimization of hybrid beamforming.

The main contribution of this paper is to design a new framework for multiple BSs to collaboratively serve multiple mobile users. In the considered model, due to the real-time mobility of users, the users being served by a given BS and beamforming of BSs and users are dynamic. Multiple BSs must cooperate to serve dynamic requests of multiple mobile users. This problem is posed as an optimization framework whose goal is to maximize the sum rate of all mobile users by jointly optimizing the number of users served by all BSs and beamforming matrices of both BSs and users. To solve this problem, we first introduce a value decomposition based reinforcement learning (VD-RL) algorithm to determine the users that all BSs need to serve. Then, we use the block diagonalization (BD) method to obtain the fully digital transmit beamforming matrices of all BSs as well as the receive beamforming matrices of the users. Finally, we introduce a fast optimization algorithm to optimize the obtained transceiver beamforming matrices.

II. System Model And Problem Formulation

A. System Model

We consider a multigroup multicast MIMO communication system where a set $\mathcal{L}$ of L BSs serving K mobile users, as shown in Fig. 1. Each user moves with a certain probability and requests the same content at different time slots. In particular, we assume that each user will only request one content o_k from the set $\mathcal{O}$ of O contents and the probability of user k requesting content o_k at time t is p_k^{C}, while the probability that user k does not request content o_k is $1-p_k^{\mathrm{C}}$. We can easily extend our model to a scenario where each user requests different content at different time slots. At time t, all users are grouped into several clusters according to their content requests and locations. Each user has N^{U} antennas and S^{U} RF chains for receiving ζ data streams from the BS. Each BS is equipped with N^{B} antennas and $O\zeta$ RF chains to serve users.

1) Mobility Model: The user's mobility is depicted by a random walk model. Specifically, users can either remain at their current location or move in one of the four directions: up, down, left, or right. A vector $\boldsymbol{p}_{kt} = [p_{kt,0}, p_{kt,1}, p_{kt,2}, p_{kt,3}, p_{kt,4}]$ represents the possibility of the movement of each user at time t, where $p_{kt,0}$ represents the possibility of user k remaining at present location at time t, $p_{kt,1}$, $p_{kt,2}$, $p_{kt,3}$, and $p_{kt,4}$ respectively represent the possibility of user k moving up, down, left, or right at time t. It is assumed that user k moves at a constant speed v_k. The location coordinate of user k in BS l's coverage area at time t is $\boldsymbol{\phi}_{lk}(t) = [\phi_{lkt,1}, \phi_{lkt,2}]$. With the duration of each time slot t assumed to be Δt, user k's location in the coverage area of BS l at time $t+1$ can be represented by

$$\boldsymbol{\phi}_{lk}(t+1) = \begin{cases} [\phi_{lkt,1}, \phi_{lkt,2}], & \text{probability } p_{kt+1,0}, \\ [\phi_{lkt,1}, \phi_{lkt,2} + v_k\Delta t], & \text{probability } p_{kt+1,1}, \\ [\phi_{lkt,1}, \phi_{lkt,2} - v_k\Delta t], & \text{probability } p_{kt+1,2}, \\ [\phi_{lkt,1} - v_k\Delta t, \phi_{lkt,2}], & \text{probability } p_{kt+1,3}, \\ [\phi_{lkt,1} + v_k\Delta t, \phi_{lkt,2}], & \text{probability } p_{kt+1,4}. \end{cases} \tag{1}$$

2) Data Rate Model: We assume that the BS-user connection vector of BS l at time t is $\boldsymbol{a}_l(t) = [a_{l1}(t), \ldots, a_{lK}(t)]$ where $a_{lk}(t) \in \{0,1\}$ is the index of the connection between BS l and user k with $a_{lk}(t) = 1$ representing user k is connected to BS l, otherwise, we have $a_{lk}(t) = 0$. Here, each user can only connect to one BS. Hence, we have $\sum_{l=1}^{L} a_{lk}(t) \leq 1$. We also assume that the content-user connection vector of content o at time t is $\boldsymbol{y}_o(t) = [y_{o1}(t), \ldots, y_{oK}(t)]^{\mathrm{T}}$ where $y_{ok}(t)) \in \{0,1\}$ is the index indicating whether user k associated with BS l requests content o at time t. $y_{ok}(t) = 1$ implies that user k associated with BS l requests content o at time t, otherwise, we have $y_{ok}(t) = 0$.

Since each user can request one content from O contents in the set $\mathcal{O}$, each BS l can cluster the users into at most O groups, which depends on users' content requests at each time slot. Let $z_{lo}(t) \in \{0,1\}$ be the index indicating whether there are users associated with BS l requesting content o at time t. $z_{lo}(t) = 0$ implies that no users associated with BS l request content o at time t, otherwise, we have $z_{lo}(t) = 1$. The relationship between user connection index $\boldsymbol{a}_l(t)$ and $z_{lo}(t)$ can be given by

$$z_{lo}(t) = \mathcal{X}_{\boldsymbol{a}_l(t)\boldsymbol{y}_o(t)\neq 0}, \tag{2}$$

where $\mathcal{X}$ represents the characteristic function. Given $\boldsymbol{a}_l(t)$,

the detected data of user k of BS l at time t is

$$\begin{aligned}&\hat{\boldsymbol{s}}_{kl}(t)\\&=\sum_{b=1}^{L}\sum_{e=1}^{O}z_{be}(t)\left(\boldsymbol{W}_{kl}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_{kl}^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_{b,k}\left(\boldsymbol{\phi}_{lk}(t)\right)\boldsymbol{F}_{b,e}^{\mathrm{R}}\boldsymbol{F}_{b,e}^{\mathrm{B}}\boldsymbol{s}_{e}\\&+\left(\boldsymbol{W}_{kl}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_{kl}^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{n}_{kl}(t)\end{aligned}\tag{3}$$

where $\boldsymbol{s}_e=[s_{e,1},\ldots,s_{e,\zeta}]^{\mathrm{T}}\in\mathbb{C}^{\zeta\times1}$ represents the data streams of content e and follows $\mathbb{E}\left[\boldsymbol{s}_e\boldsymbol{s}_e^{\mathrm{H}}\right]=\boldsymbol{I}$, $\boldsymbol{F}_{b,e}^{\mathrm{B}}$ is baseband transmit beamforming matrix of group e under BS b, $\boldsymbol{F}_{b,e}^{\mathrm{R}}$ is RF transmit beamforming matrix of group e under BS b, $\boldsymbol{H}_{b,k}\left(\boldsymbol{\phi}_{lk}(t)\right)$ represents the effective channel between BS b and user k, $\boldsymbol{W}_{kl}^{\mathrm{R}}$ represents the RF receive beamforming matrix of user k located in the coverage area of BS l, $\boldsymbol{W}_{kl}^{\mathrm{B}}$ is baseband receive beamfoming matrix of user k located in the coverage area of BS l, $\boldsymbol{n}_{kl}(t)\in\mathbb{C}^{N^{\mathrm{U}}\times1}$ represents an additive white Gaussian noise vector of user k. Each element of $\boldsymbol{n}_{kl}(t)$ abides by an independent and identically distributed complex Gaussian distribution, which has a zero mean value and variance σ^2. The available data stream i of user k from BS l at the time t is

$$\begin{aligned}&\hat{s}_{kl,i}(t)\\&=\left(\boldsymbol{w}_{kl,i}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_{kl,i}^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_{l,k}\left(\boldsymbol{\phi}_{lk}(t)\right)\boldsymbol{F}_{l,o_k,i}^{\mathrm{R}}\bar{\boldsymbol{f}}_{l,o_k,i}^{\mathrm{B}}s_{o_k,i}\\&+\sum_{j=1,j\neq i}^{\zeta}\left(\boldsymbol{w}_{kl,i}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_{kl,i}^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_{l,k}\left(\boldsymbol{\phi}_{lk}(t)\right)\boldsymbol{F}_{l,o_k,j}^{\mathrm{R}}\\&\quad\bar{\boldsymbol{f}}_{l,o_k,j}^{\mathrm{B}}s_{o_k,j}\\&+\sum_{b=1}^{L}\sum_{e=1}^{O}z_{be}(t)\left(\boldsymbol{W}_{kl}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_{kl}^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_{b,k}\left(\boldsymbol{\phi}_{lk}(t)\right)\boldsymbol{F}_{b,e}^{\mathrm{R}}\boldsymbol{F}_{b,e}^{\mathrm{B}}\boldsymbol{s}_{e}\\&-\left(\boldsymbol{W}_{kl}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_{kl}^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_{l,k}\left(\boldsymbol{\phi}_{lk}(t)\right)\boldsymbol{F}_{l,o_k}^{\mathrm{R}}\boldsymbol{F}_{l,o_k}^{\mathrm{B}}\boldsymbol{s}_{o_k}\\&+\left(\boldsymbol{W}_{kl,i}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_{kl,i}^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{n}_{kl,i}(t),\end{aligned}\tag{4}$$

where $\boldsymbol{w}_{kl,i}^{\mathrm{B}}$ represents the i-th row of matrix $\boldsymbol{W}_{kl}^{\mathrm{B}}$, and $\bar{\boldsymbol{f}}_{l,o_k,i}^{\mathrm{B}}$ represents to the i-th column of matrix $\boldsymbol{F}_{l,o_k}^{\mathrm{B}}$. As detailed in equation (4), the first term denotes the desired signal, the second term represents the interference from different streams, the third and fourth terms represent the interference from other groups, and the fifth term denotes the noise. The signal-to-interference-plus-noise ratio (SINR) for the user k at BS l while receiving data stream i at time t is determined as follows

$$\xi_{kl,i}(t)=\frac{\left|\left(\boldsymbol{w}_{kl,i}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_{kl,i}^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_{l,k}\left(\boldsymbol{\phi}_{lk}(t)\right)\boldsymbol{F}_{l,o_k,i}^{\mathrm{R}}\bar{\boldsymbol{f}}_{l,o_k,i}^{\mathrm{B}}\right|^2}{I_{kl,i}(t)+J_{kl,i}(t)+\sigma^2},\tag{5}$$

where $I_{kl,i}(t)=\sum_{j=1,j\neq i}^{\zeta}\left|\left(\boldsymbol{w}_{kl,i}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_{kl,i}^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_{l,k}\left(\boldsymbol{\phi}_{lk}(t)\right)\boldsymbol{F}_{l,o_k,j}^{\mathrm{R}}\bar{\boldsymbol{f}}_{l,o_k,j}^{\mathrm{B}}\right|^2$ is the interference of other streams, and $J_{kl,i}(t)=\sum_{b=1}^{L}\sum_{e=1}^{O}\left|z_{be}(t)\left(\boldsymbol{W}_{kl}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_{kl}^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_{b,k}\left(\boldsymbol{\phi}_{lk}(t)\right)\boldsymbol{F}_{b,e}^{\mathrm{R}}\boldsymbol{F}_{b,e}^{\mathrm{B}}\right|^2-\left|\left(\boldsymbol{W}_{kl}^{\mathrm{B}}\right)^{\mathrm{H}}\left(\boldsymbol{W}_{kl}^{\mathrm{R}}\right)^{\mathrm{H}}\boldsymbol{H}_{l,k}\left(\boldsymbol{\phi}_{lk}(t)\right)\boldsymbol{F}_{l,o_k}^{\mathrm{R}}\boldsymbol{F}_{l,o_k}^{\mathrm{B}}\right|^2$ is the interference of other groups. The achievable data rate of user k of BS l at time t is given by

$$c_{kl}(t)=W\sum_{i=1}^{\zeta}\log_2\left(1+\xi_{kl,i}(t)\right),\tag{6}$$

where W represents the bandwidth.

Due to the characteristics of the multicast mechanism, the data rate for the group that includes user k at BS l is dictated by the user with the lowest data rate, which can be given by

$$c_{lo}(t)=\min_{k\in\mathcal{U}_{l,o}}\left\{c_{kl}(t)\right\}.\tag{7}$$

B. Problem Formulation

Next, we describe our optimization problem. The objective is to maximize the sum rate for all multicasting groups by simultaneously optimizing the transmit beamforming matrices $\boldsymbol{F}_{l}^{\mathrm{R}}(t)$, $\boldsymbol{F}_{l,o}^{\mathrm{B}}(t)$, the receive beamforming matrices $\boldsymbol{W}_{kl}^{\mathrm{R}}(t)$, $\boldsymbol{W}_{kl}^{\mathrm{B}}(t)$, and the indicator variable $a_{lk}(t)$ at time t. The optimization problem can be described as follows:

$$\max_{\boldsymbol{F}_{l}^{\mathrm{R}}(t),\boldsymbol{F}_{l,o}^{\mathrm{B}}(t),\boldsymbol{W}_{kl}^{\mathrm{R}}(t),\boldsymbol{W}_{kl}^{\mathrm{B}}(t),a_{lk}(t)}\mathbb{E}\left(\sum_{l=1}^{L}\sum_{o=1}^{O}z_{lo}(t)\,c_{lo}(t)\right)\tag{8}$$

$$\text{s.t.}\quad\sum_{l=1}^{L}a_{lk}(t)\leq1,\forall k,\tag{8a}$$

$$\sum_{o=1}^{O}z_{lo}(t)\left\|\boldsymbol{F}_{l}^{\mathrm{R}}(t)\,\boldsymbol{F}_{l,o}^{\mathrm{B}}(t)\right\|_F^2\leq D,\forall l,\tag{8b}$$

$$\left|\boldsymbol{F}_{l}^{\mathrm{R}}(i,j)(t)\right|=\left|\boldsymbol{W}_{kl}^{\mathrm{R}}(i,j)(t)\right|=1,\forall i,j,\tag{8c}$$

where D is the maximum transmit power of each BS. Constraint (8a) represents each user can only connect to one BS. The maximum transmit power limitations for all BSs are depicted in (8b). Constraint (8c) describes the amplitude limitations for the RF beamforming matrices of the BSs and users. Due to the non-convex objective function (8) and non-convex limitations (8a)-(8c), problem (8) is non-convex, which makes it difficult to solve.

III. Proposed Scheme

To effectively solve problem (8), we first introduce a VD-RL algorithm to determine the users that all BSs need to serve. Then, we use the BD method to obtain the fully digital transmit beamforming matrices of all BSs as well as the receive beamforming matrices of the users. Finally, we introduce a fast optimization algorithm to optimize the hybrid transmit beamforming matrices $\boldsymbol{F}_{l}^{\mathrm{R}}(t)$, $\boldsymbol{F}_{l,o}^{\mathrm{B}}(t)$ and the receive beamforming matrices $\boldsymbol{W}_{kl}^{\mathrm{R}}(t)$, $\boldsymbol{W}_{kl}^{\mathrm{B}}(t)$ at time t.

A. VD-RL Algorithm.

Next, we first present the components that constitute the VD-RL scheme. Then, we illustrate the process of utilizing the proposed scheme to solve the problem in (8).

1) VD-RL Components: The VD-RL scheme is composed of six essential components.

- Agent: The agents in VD-RL are the BSs that determine user association.
- States: The state of BS l at time t can be depicted as a vector $\boldsymbol{\xi}_{lt} = [\boldsymbol{\xi}_{l1t}, ..., \boldsymbol{\xi}_{lkt}, ..., \boldsymbol{\xi}_{lE_l t}]$, where E_l is the maximum user index within the coverage range of BS l. $\boldsymbol{\xi}_{lkt} = [\boldsymbol{\phi}_{lk}(t), o_k]$ represents the state of user k in the coverage range of BS l at time t with $\boldsymbol{\phi}_{lk}(t) = [\phi_{lkt,1}, \phi_{lkt,2}]$ being the location coordinate of user k in the coverage of BS l at time t and o_k being the content request of user k. Here, we note that, each BS l can only observe the state $\boldsymbol{\xi}_{lt}$ within its own coverage area at time t.
- Actions: The action of each agent is to determine which users to serve. Hence, an action of BS l at time slot t can be expressed as $\boldsymbol{a}_l(t) = [a_{l1}(t), \ldots, a_{lK}(t)]$, where $a_{lk}(t) \in \{0, 1\}$ is the index of the connection between BS l and user k with $a_{lk}(t) = 1$ representing user k is connected to BS l, otherwise, we have $a_{lk}(t) = 0$. Similarly, the actions of all BSs at time t is $\boldsymbol{a}(t) = [\boldsymbol{a}_1(t), \ldots, \boldsymbol{a}_L(t)]$.
- Reward: The reward of each BS is used to capture the benefit of a selected action. The reward of group o associated with BS l at time t is $c_{lo}(t)$, where $c_{lo}(t)$ is the data rate of group o associated with BS l when the BS takes action $\boldsymbol{a}_l(t)$. To calculate the data rate $c_{lo}(t)$, we first need to solve problem (8). When the BS takes action $\boldsymbol{a}_l(t)$, the optimization problem in (8) can be simplified as

$$\max_{\boldsymbol{F}_l^{\mathrm{R}}(t), \boldsymbol{F}_{l,o}^{\mathrm{B}}(t), \boldsymbol{W}_{kl}^{\mathrm{R}}(t), \boldsymbol{W}_{kl}^{\mathrm{B}}(t)} \mathbb{E}\left(\sum_{l=1}^{L}\sum_{o=1}^{O} c_{lo}(t)\right) \quad (9)$$

$$\text{s.t.} \quad \sum_{o=1}^{O} \left\|\boldsymbol{F}_l^{\mathrm{R}}(t)\, \boldsymbol{F}_{l,o}^{\mathrm{B}}(t)\right\|_F^2 \le D, \forall l, \quad (9a)$$

$$\left|\boldsymbol{F}_l^{\mathrm{R}}(i,j)(t)\right| = \left|\boldsymbol{W}_{kl}^{\mathrm{R}}(i,j)(t)\right| = 1, \forall i, j. \quad (9b)$$

To optimize $\boldsymbol{F}_l^{\mathrm{R}}(t), \boldsymbol{F}_{l,o}^{\mathrm{B}}(t), \boldsymbol{W}_{kl}^{\mathrm{R}}(t)$, and $\boldsymbol{W}_{kl}^{\mathrm{B}}(t)$, we can use the method in our previous work [16]. Next, we introduce the definition of total reward of all BSs. Since each BS l can only observe its partial state $\boldsymbol{\xi}_{lt}$ at time t and each BS is unaware of the connections between other BSs and users, a user may be served by multiple BSs, which is impractical. To let each user to be served by only one BS, we add a penalty $\rho < 0$ to the reward function for penalizing the scenario where multiple BSs serving one user. Hence, the total reward of all BSs is

$$\begin{aligned} & r(\boldsymbol{\xi}_t, \boldsymbol{a}(t)) \\ &= \sum_{l=1}^{L} r_l(\boldsymbol{\xi}_{lt}, \boldsymbol{a}_l(t)) \\ &= \sum_{l=1}^{L}\sum_{o\in\mathcal{O}} \left[\left(1 - \mathbb{1}_{\left\{\sum_{\zeta\neq l, \zeta\in\mathcal{L}} a_{\zeta k}(t)=0\right\}}\right)\rho + c_{lo_k}(t)\right] \end{aligned} \quad (10)$$

where $\sum_{o\in\mathcal{O}}\left[\left(1 - \mathbb{1}_{\left\{\sum_{\zeta\neq l, \zeta\in\mathcal{L}} a_{\zeta k}(t)=0\right\}}\right)\rho + c_{lo_k}(t)\right]$ is the reward of BS l at time t and $\mathbb{1}_{\left\{\sum_{\zeta\neq l, \zeta\in\mathcal{L}} a_{\zeta k}(t)=0\right\}}$ is a function that indicates whether user k is connected to other BSs. In particular, $\mathbb{1}_{\left\{\sum_{\zeta\neq l, \zeta\in\mathcal{L}} a_{\zeta k}(t)=0\right\}} = 0$ indicates that user k is connected to other BSs, and $\mathbb{1}_{\left\{\sum_{\zeta\neq l, \zeta\in\mathcal{L}} a_{\zeta k}(t)=0\right\}} = 1$, otherwise. From (10), we can see that when multiple BSs serve one user, the reward will add ρ.

- Individual Q value function: For each BS l, the individual Q value function $Q_{\boldsymbol{\vartheta}_{lt}}(\boldsymbol{\xi}_{lt}, \boldsymbol{a}_l(t))$ is utilized to represent the expected reward based on a specific partial state $\boldsymbol{\xi}_{lt}$ of BS l and a chosen action $\boldsymbol{a}_l(t)$ at time t. Each BS l uses a deep neural network (DNN) with the parameter $\boldsymbol{\vartheta}_{lt}$ to estimate its individual Q value function at time t. Due to the fact that a BS can only observe users' states within its coverage area, it shares its individual Q value with other BSs [17] to estimate the global Q value function, which will be further clarified in the next bullet.
- Global Q value function: The global Q value function, denoted as $Q_{tot}(\boldsymbol{\xi}_t, \boldsymbol{a}(t))$, is utilized to compute the overall expected reward from all the BSs. Regarding the proposed VD-RL scheme, it is assumed that the global Q value for all BSs is equivalent to the sum of each BS's individual Q value, which can be expressed as [18], [19]

$$Q_{tot}(\boldsymbol{\xi}_t, \boldsymbol{a}(t)) = \sum_{l=1}^{L} Q_{\boldsymbol{\vartheta}_{lt}}(\boldsymbol{\xi}_{lt}, \boldsymbol{a}_l(t)) \quad (11)$$

The objective for all BSs is to cooperate to maximize the overall expected reward. After training, each BS can utilize the Q function to determine ideal BS-user connections so as to maximize the sum rate for all multicasting groups.

2) VD-RL Solution: At first, each agent collects local data which contains partial states $\boldsymbol{\xi}_{lt}$ and actions $\boldsymbol{a}_l(t)$ at time t. Then, each agent shares its local data with other agents [20] for the purpose of determining its reward $r_l(\boldsymbol{\xi}_{lt}, \boldsymbol{a}_l(t))$ and the overall reward $r(\boldsymbol{\xi}_t, \boldsymbol{a}(t)) = \sum_{l=1}^{L} r_l(\boldsymbol{\xi}_{lt}, \boldsymbol{a}_l(t))$ for all BSs. Finally, each agent updates its individual Q value function based on the global Q value function. The definition for the loss function of the global Q value function $Q_{tot}(\boldsymbol{\xi}_t, \boldsymbol{a}(t))$ is as follows:

$$\begin{aligned} A(\boldsymbol{\vartheta}_{1t}, \ldots, \boldsymbol{\vartheta}_{Lt}) =& \mathbb{E}\left[Q_{tot}(\boldsymbol{\xi}_t, \boldsymbol{a}(t)) - r(\boldsymbol{\xi}_t, \boldsymbol{a}(t))\right. \\ & \left. - \max_{\boldsymbol{a}(t+1)'} Q_{tot}\left(\boldsymbol{\xi}_{t+1}, \boldsymbol{a}(t+1)'\right)\right]^2 \end{aligned} \quad (12)$$

where $\max_{\boldsymbol{a}(t+1)'} Q_{tot}\left(\boldsymbol{\xi}_{t+1}, \boldsymbol{a}(t+1)'\right)$ is defined as the maximum global Q value for the state $\boldsymbol{\xi}_{t+1}$. As each individual Q value consistently rises, the global Q value exhibits a steady growth, i.e., an action taken by an agent having a high individual Q value also holds significant value for the entire wireless network. Therefore, the purpose of each BS training its individual Q value function is to optimize

TABLE I: Simulation Parameters

Parameters	Values	Parameters	Values
L	5	K	50
N^{U}	64	S^{U}	4
O	5	ζ	4
ρ	2	D	50 dBm
v_k	2 m/s	σ^2	-90 dBm

the global Q value. Each BS's individual Q value function $Q_{\boldsymbol{\vartheta}_{lt}}\left(\boldsymbol{\xi}_{lt}, \boldsymbol{a}_l\left(t\right)\right)$ can be updated by applying a gradient descent approach, as described below:

$$\boldsymbol{\vartheta}_{lt} \leftarrow \boldsymbol{\vartheta}_{lt} - \lambda_{\boldsymbol{\vartheta}_{lt}} \nabla_{\boldsymbol{\vartheta}_{lt}} A\left(\boldsymbol{\vartheta}_{1t}, \ldots, \boldsymbol{\vartheta}_{Lt}\right), \tag{13}$$

where $\lambda_{\boldsymbol{\vartheta}_{lt}}$ is the updating rate and $\nabla_{\boldsymbol{\vartheta}_{lt}} A\left(\boldsymbol{\vartheta}_{1t}, \ldots, \boldsymbol{\vartheta}_{Lt}\right)$ is the gradient of the global Q value function, which can be described as follows:

$$\begin{aligned} &\nabla_{\boldsymbol{\vartheta}_{lt}} A\left(\boldsymbol{\vartheta}_{1t}, \ldots, \boldsymbol{\vartheta}_{Lt}\right) \\ &= \nabla_{\boldsymbol{\vartheta}_{lt}} \left[Q_{tot}\left(\boldsymbol{\xi}_t, \boldsymbol{a}\left(t\right)\right) - r\left(\boldsymbol{\xi}_t, \boldsymbol{a}\left(t\right)\right)\right. \\ &\quad \left. - \max_{\boldsymbol{a}(t+1)'} Q_{tot}\left(\boldsymbol{\xi}_{t+1}, \boldsymbol{a}\left(t+1\right)'\right)\right]^2 \\ &= 2\Delta Q_{tot} \nabla_{\boldsymbol{\vartheta}_{lt}} Q_{\boldsymbol{\vartheta}_{lt}}\left(\boldsymbol{\xi}_{lt}, \boldsymbol{a}_l\left(t\right)\right), \end{aligned} \tag{14}$$

where $\Delta Q_{tot} = Q_{tot}\left(\boldsymbol{\xi}_t, \boldsymbol{a}\left(t\right)\right) - r\left(\boldsymbol{\xi}_t, \boldsymbol{a}\left(t\right)\right) - \max_{\boldsymbol{a}(t+1)'} Q_{tot}\left(\boldsymbol{\xi}_{t+1}, \boldsymbol{a}\left(t+1\right)'\right)$.

IV. Simulation Results And Analysis

For our simulations, we consider a scenario with five BSs serving 50 mobile users. The coordinates of the BSs are (50m, 50m, 10m), (200m, 800m, 10m), (400m, 500m, 10m), (750m, 50m, 10m) and (1000, 600m, 10m), respectively. The moving speed of each user is v_k = 2 m/s and the time duration of a time slot is Δt = 1 s. Other parameters are listed in Table I. For comparison purposes, we consider three baselines, which are given as follows:

- a) VDRL-WMMSE is an algorithm in which the beamforming matrices of the BSs and the users are determined based on the weighted minimum mean-square error (WMMSE) criterion [21], and the connections between BSs and users are determined by VDRL.
- b) VDRL-Heuristic is an algorithm in which the beamforming matrices of the BS and the users are determined based on a heuristic algorithm, and the connections between BSs and users are determined by VDRL.
- c) IDRL-BD is an algorithm in which the beamforming matrices of the BSs and the users are optimized by a BD method, and the connections between BSs and users are determined by independent distributed reinforcement learning (IDRL).

Fig. 2 illustrates the convergence of the proposed scheme. The figure indicates that around 200 iterations are needed for the proposed algorithm to converge, thus improving the convergence speed by up to 39.8%, 44.5%, and 71.1% compared to VDRL-WMMSE, VDRL-Heuristic, and IDRL-BD. The 39.8% and 44.5% gains stem from the fact that VDRL-WMMSE and VDRL-Heuristic do not eliminate the

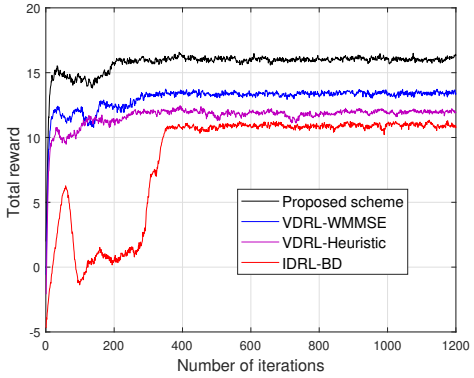


Fig. 2: Value of the total rewards as the number of iterations.

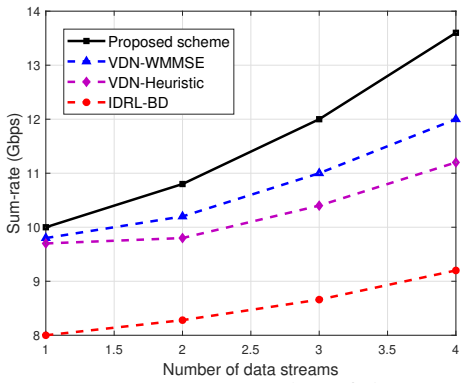
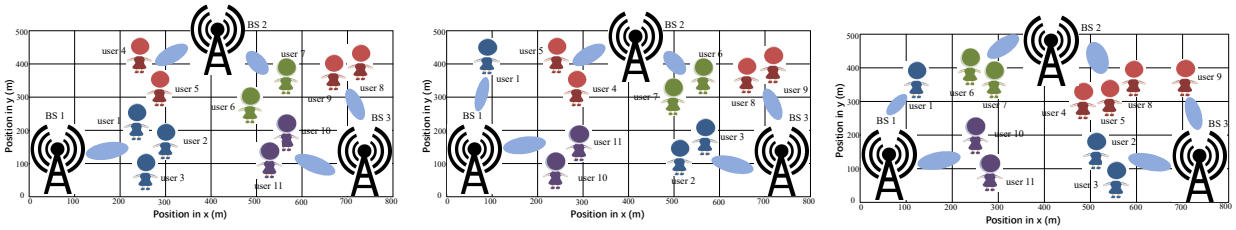


Fig. 3: Sum-rate versus number of data streams.

inter-group interference and the interference among multiple streams of each user. The 71.1% gain stems from the fact that the BSs in our method can make the decisions cooperatively to maximize the sum rate. Fig. 2 also shows that the proposed scheme can achieve 20.7%, 35.3%, and 51% gains in terms of total reward compared to VDRL-WMMSE, VDRL-Heuristic, and IDRL-BD. The performance improvements of 20.7% and 35.3% are attributed to the BD method's capability to mitigate interference both inter-group and each user's multiple streams. The 51% gain stems from the fact that the proposed algorithm optimizes the total reward of all BSs while the IDRL-BD maximizes individual reward of each BS and ignores the actions of other BSs.

Fig. 3 shows the relationship between the sum rate of mobile users and the number of data streams. From Fig. 3, we can see that the difference between the proposed algorithm and VDRL-WMMSE, VDRL-Heuristic becomes larger with the growth in data streams. This is because when only one flow of data is transmitted, only inter-group interference exists between users and the interference among multiple streams of each user does not exist. As the number of data streams increases, the interference among multiple streams of each user also increases. The proposed algorithm eliminates both inter-group and the interference among multiple streams of each user.

Fig. 4 is a visualization of using our proposed scheme to determine user associations. In this figure, the users that



(a) (b) (c)

Fig. 4: Visualization of BS-user connentions and user grouping obtained by the proposed scheme.

request the same content are represented by the same color. From Fig. 4(a), we can see that users who are geographically close and request the same content have a higher probability of being assigned to the same group of the same BS. Fig. 4(b) shows that as the users move, BS-user connections change. From Fig. 4(c), we can also see that users who are geographically close and request the same content may not necessarily be assigned to the same BS. For example, in Fig. 4(c), users 4, 5, 8, and 9 are not served by the same BS. This is because users 4, 5, 8, and 9 have significantly different signal reception directions.

V. Conclusions

In this paper, we have studied the problem of maximizing the sum rate of mobile users in a multi-BS cooperative mmWave multicast communication system. In the considered model, due to the real-time mobility of users, multiple BSs must cooperate to serve dynamic requests of multiple mobile users. We have formulated the problem into an optimization framework and have put forth VD-RL algorithm as a solution. Simulation results indicate that the proposed VD-RL algorithm outperforms the traditional multi-agent algorithms in terms of convergence behavior and sum rate.

References

[1] Y. Yang, F. Gao, X. Tao, G. Liu, and C. Pan, "Environment semantics aided wireless communications: A case study of mmWave beam prediction and blockage prediction," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 7, pp. 2025–2040, Jul. 2023.

[2] Y. Niu, Y. Liu, Y. Li, X. Chen, Z. Zhong, and Z. Han, "Device-to-device communications enabled energy efficient multicast scheduling in mmWave small cells," *IEEE Transactions on Communications*, vol. 66, no. 3, pp. 1093–1109, Mar. 2018.

[3] J. An, C. Xu, D. W. K. Ng, G. C. Alexandropoulos, C. Huang, C. Yuen, and L. Hanzo, "Stacked intelligent metasurfaces for efficient holographic MIMO communications in 6G," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2380–2396, Aug. 2023.

[4] H. Yu, H. D. Tuan, E. Dutkiewicz, H. V. Poor, and L. Hanzo, "Regularized zero-forcing aided hybrid beamforming for millimeter-wave multiuser MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 5, pp. 3280–3295, May 2023.

[5] S. Gong, C. Xing, P. Yue, L. Zhao, and T. Q. S. Quek, "Hybrid analog and digital beamforming for RIS-assisted mmWave communications," *IEEE Transactions on Wireless Communications*, vol. 22, no. 3, pp. 1537–1554, Mar. 2023.

[6] S.-H. Wu and G.-Y. Lu, "Compressive beam and channel tracking with reconfigurable hybrid beamforming in mmWave MIMO OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 2, pp. 1145–1160, Feb. 2023.

[7] L. Zhao, M. Li, C. Liu, S. V. Hanly, I. B. Collings, and P. A. Whiting, "Energy efficient hybrid beamforming for multi-user millimeter wave communication with low-resolution A/D at transceivers," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2142–2155, Sept. 2020.

[8] Y. Cai, K. Xu, A. Liu, M. Zhao, B. Champagne, and L. Hanzo, "Two-timescale hybrid analog-digital beamforming for mmWave full-duplex MIMO multiple-relay aided systems," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2086–2103, Sept. 2020.

[9] Z. Wang, M. Li, R. Liu, and Q. Liu, "Joint user association and hybrid beamforming designs for cell-free mmWave MIMO communications," *IEEE Transactions on Communications*, vol. 70, no. 11, pp. 7307–7321, Nov. 2022.

[10] G. Femenias and F. Riera-Palou, "Wideband cell-free mmWave massive MIMO-OFDM: Beam squint-aware channel covariance-based hybrid beamforming," *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 4695–4710, Jul. 2022.

[11] J. Li, Z. Wang, Y. Zhang, P. Zhu, D. Wang, and X. You, "Robust hybrid beamforming for outage-constrained multigroup multicast mmWave transmission with phase shifter impairments," *IEEE Systems Journal*, vol. 17, no. 1, pp. 869–880, Mar. 2023.

[12] W. Ci, C. Qi, G. Y. Li, and S. Mao, "Hybrid beamforming design for covert multicast mmWave massive MIMO communications," in *Proc. IEEE Global Communications Conference*, Madrid, Spain, Dec. 2021, pp. 1–6.

[13] Z. Li, C. Qi, and G. Y. Li, "Low-complexity multicast beamforming for millimeter wave communications," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 12 317–12 320, Oct. 2020.

[14] Z. Wang, Q. Liu, M. Li, and W. Kellerer, "Energy efficient analog beamformer design for mmWave multicast transmission," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 2, pp. 552–564, Jun. 2019.

[15] Y. Niu, L. Yu, Y. Li, Z. Zhong, and B. Ai, "Device-to-device communications enabled multicast scheduling for mmWave small cells using multi-level codebooks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2724–2738, Mar. 2019.

[16] S. Zhang, Z. Yang, M. Chen, D. Liu, K.-K. Wong, and H. V. Poor, "Beamforming design for the performance optimization of intelligent reflecting surface assisted multicast MIMO networks," *IEEE Transactions on Wireless Communications*, vol. 23, no. 3, pp. 2325–2339, Mar. 2024.

[17] W. Xu, Z. Yang, D. W. K. Ng, M. Levorato, Y. C. Eldar, and M. Debbah, "Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 17, no. 1, pp. 9–39, Jan. 2023.

[18] P. Sunehag, G. Lever, A. Gruslys, W. M. Czarnecki, V. Zambaldi, M. Jaderberg, M. Lanctot, N. Sonnerat, J. Z. Leibo, K. Tuyls, and T. Graepel, "Value-decomposition networks for cooperative multi-agent learning," Available online: http://arxiv.org/abs/1706.05296, Jun. 2017.

[19] Y. Hu, M. Chen, W. Saad, H. V. Poor, and S. Cui, "Distributed multi-agent meta learning for trajectory design in wireless drone networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 10, pp. 3177–3192, Oct. 2021.

[20] M. Chen, D. Gündüz, K. Huang, W. Saad, M. Bennis, A. V. Feljan, and H. V. Poor, "Distributed learning in wireless networks: Recent progress and future challenges," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 12, pp. 3579–3605, Dec. 2021.

[21] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.