Low-Complexity Design of Decode-Forward Relaying in Massive MIMO Heterogeneous Networks

Ahmad Abu Al Haija*, Min Dong‡, Ben Liang*, Gary Boudreau†
* University of Toronto, Toronto, ON, Canada, †Ericsson Canada, Ottawa, ON, Canada
‡ University of Ontario Institute of Technology, Oshawa, ON, Canada

***Abstract*—We investigate the impact of massive MIMO on the uplink transmission design for a heterogeneous network (HetNet) where multiple users communicate with a macro-cell base station (MCBS) through multiple small-cell BSs (SCBSs). We develop a new scheme in which the SCBSs deploy decode-forward (DF) relaying, multi-layer binning and time division transmission where the number of binning layers (resp. time slots) is equal to the number of SCBSs (resp. users). The MCBS separately and sequentially decodes the binning indices and each user's message that belongs to those indices. The proposed scheme is simpler than the scheme with common transmission of all users' messages by each SCBS and joint decoding at the MCBS. Specifically, in the proposed scheme, 1) the codebook size and the decoding complexity increase linearly with the number of users instead of exponentially, 2) every transmission-decoding step is similar to the conventional point-to-point communication, and 3) the same set of time slot durations at all SCBSs is sufficient to achieve the maximum rate. Despite its simplicity, the proposed scheme is efficient since it achieves the same rate performance of the more complex schemes with joint decoding.**

I. Introduction

The development of the fifth-generation (5G) cellular networks aims to drastically improve the spectral efficiency and data rate of current networks to serve a large number of connected devices. Hence, some key enabling technologies have been proposed, such as massive MIMO systems, HetNet with wireless backhaul, and full-duplex transmissions [1].

Consider the uplink transmission in a massive MIMO HetNet in Fig. 1, where the MCBS and the SCBSs are equipped with a large number of antennas, and the UEs in two close-by small cells intend to transmit data to the MCBS. Instead of deploying interference coordination as in the current LTE-A standard [2] and having each UE served by one SCBS or the MCBS [3], [4], both SCBSs can help relay the data from both UEs to the MCBS. Moreover, in a dense network [5], the MCBS may also receive signals that a UE transmits to its SCBS. Such channel setting is theoretically modeled as a multiple-access multiple-relay channel (MAMRC) where the SCBSs (resp. MCBS) resemble the relays (resp. destination).

For the MAMRC, several relaying schemes were proposed based on quantize-forward [6], amplify-forward [7], and decode-forward (DF) [8], [9] relaying. In DF relaying, the MCBS may jointly decode all UEs' messages. This joint decoding (JD) leads to the largest rate regions but has high computational complexity [10]. To reduce the decoding complexity, simpler decoding with the facilitation of binning techniques [9], [10] at each SCBS can be considered, such that the MCBS sequentially decodes the binning indices and then the UEs' messages [10]. However, this may result in rate loss as compared with that of JD [10].

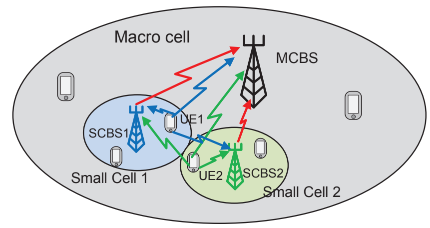


Fig. 1. Uplink transmission in a HetNet with two UEs and two SCBSs.

For a massive MIMO HetNet, receive antenna processing at the BS may be explored to further simplify the DF relaying design. Massive MIMO systems [11] are able to 1) neglect the small scale fading through channel hardening [12], 2) orthogonalize different user transmissions through beamforming [13], and 3) approach the optimal performance with simple linear receivers, e.g, zero-forcing (ZF) receiver [13]. Given these properties, here, we investigate how massive MIMO systems can simplify the uplink transmission in the HetNet with DF relaying at each SCBS, and develop a low-complexity and efficient DF scheme, while maximizing the rate performance.

The contributions of this paper are as follows:

- We propose a new uplink transmission scheme for a massive MIMO HetNet. In the proposed scheme, the ZF receiver is used at each SCBS and the MCBS; each UE performs conventional transmission; each SCBS performs DF relaying for all UEs' messages with multi-layer binning (MB) [9] and time-division (TD) transmission over multiple phases; and the MCBS performs separate and sequential decoding for each bin index in each layer and each UE's message.
- We derive the rate region and show that it is irrelevant to the SCBSs' order (i.e. which SCBS transmits each layer of binning) and allocating the same set of phase durations at all SCBSs is sufficient to maximize the rate region.
- We analyze the complexity of the proposed scheme in terms of the codebook size and decoding complexity. We show that the complexity increases linearly with the number of UEs, and ordering the SCBSs based on their link qualities to the MCBS minimizes the codebook size.
- We compare our scheme with other more sophisticated schemes with exponential complexity where the MCBS either jointly decodes all UEs' messages [8] or the binning indices within each different layer and then UEs' messages [9]. We show that our scheme with linear complexity achieves the same rate regions of the other schemes.

This work has been funded in part by Ericsson Canada and by the Natural Sciences and Engineering Research Council of Canada.

II. System Model

We consider the uplink transmission in a dense HetNet of K UEs communicating with a MCBS via L SCBSs. Each UE has a single antenna while the MCBS (resp. each SCBS) has M (resp. N) antennas, with $N \gg K$ and $M \gg LN$. For a dense HetNet, a SCBS may receive strong signals from UEs in other small cells. Thus, each SCBS relays the received data of its serving UEs and UEs in other small cells.

For simplicity, we first consider $K = 2$ and $L = 2$ as in Fig. 2 for MAMRC where Relay 1, Relay 2, and the destination respectively resemble SCBS1, SCBS2, and the MCBS. Generalization to any K and L is given in Remark 4.

We assume a block fading channel model where the channel over each link remains constant in each transmission block and changes independently between blocks. Consider B transmission blocks, where $B \gg 1$. Let $\mathbf{h}_{rl,i,j}$ denote the $N \times 1$ complex Gaussian channel vector from UE_i to the SCBS l in block j, for $i \in \{1, 2\}$, $l \in \{1, 2\}$, $j \in \{1, \ldots, B\}$. Each $\mathbf{h}_{rl,i,j}$ is a complex Gaussian random vector with zero mean and covariance $\sigma^2_{h,rl}\mathbf{I}$. The variance $\sigma^2_{h,rl}$ is determined based on a pathloss model as $\sigma^2_{h,rl} = d^{-\alpha}_{rl,i}$, with $d_{rl,i}$ being the distance between UE_i and the SCBS l, and α is the pathloss exponent. A Similar definition holds for each element of the $M \times 1$ channel vector $\mathbf{h}_{di,j}$ from UE_i to the MCBS and the $M \times N$ channel matrix $\mathbf{H}_{drl,j}$ from SCBS l to the MCBS with d_{di} (resp. d_{drl}) being the distance from UE_i (resp. SCBS l) to the MCBS. We assume all channel coefficients are independent of each other.

For each transmission block $j \in \{1, \ldots, B\}$, the respective received signal vectors $\mathbf{y}_{rl,j}$ and $\mathbf{y}_{d,j}$ at each SCBS and the MCBS, are given by

$$\begin{aligned}\mathbf{y}_{rl,j} &= \mathbf{h}_{rl,1,j}x_{1,j} + \mathbf{h}_{rl,2,j}x_{2,j} + \mathbf{z}_{rl,j},\ l \in \{1, 2\}\\ \mathbf{y}_{d,j} &= \mathbf{h}_{d1,j}x_{1,j} + \mathbf{h}_{d2,j}x_{2,j}\\ &\quad + \mathbf{H}_{dr1,j}\mathbf{x}_{r1,j} + \mathbf{H}_{dr2,j}\mathbf{x}_{r2,j} + \mathbf{z}_{d,j},\end{aligned} \tag{1}$$

where $x_{i,j}$ is the transmit signal by UEi and $\mathbf{x}_{rl,j}$ is the $N \times 1$ transmit signal vector from SCBS l; $\mathbf{z}_{rl,j}$ and $\mathbf{z}_{d,j}$ are independent complex Gaussian noise vectors with zero mean and covariance $\mathbf{I}_N$ and $\mathbf{I}_M$, respectively.

We assume perfect channel information at the respective receivers, i.e. SCBS l knows $\mathbf{h}_{rl,1,j}$ and $\mathbf{h}_{rl,2,j}$ and the MCBS knows $\mathbf{h}_{d1}$, $\mathbf{h}_{d2}$, $\mathbf{H}_{dr1}$, and $\mathbf{H}_{dr2}$. We consider full-duplex relaying at each SCBS. Although full-duplex relaying suffers from self-interference, it can be substantially alleviated by analog and digital cancellation techniques [14]. Hence, we assume perfect cancellation at the SCBSs in our design.

In massive MIMO systems, linear detectors like ZF or maximum ratio combining (MRC) can diminish the inter-user interference and achieve high data rate [13]. In this work, we choose the ZF detector for simplicity; but similar analysis is applicable to other detectors. The SCBSs and MCSB receivers can perform ZF detection to separate the data streams from different origins. Specifically, SCBS1 (resp. SCBS2) processes its received signal vector $\mathbf{y}_{r1,j}$ (resp. $\mathbf{y}_{r2,j}$) by ZF matrix $\mathbf{A}_{r1,j}$ (resp. $\mathbf{A}_{r2,j}$), and the MCBS processes $\mathbf{y}_{d,j}$ by $\mathbf{A}_{d,j}$ where

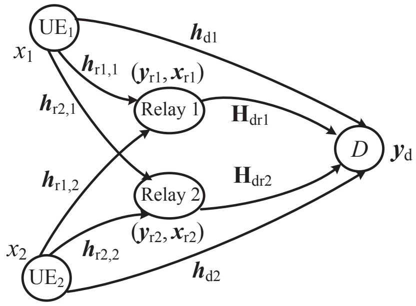


Fig. 2. The channel model of the MAMRC with two users and two relays.

$$\begin{aligned}\mathbf{A}_{rl,j} &\triangleq (\mathbf{G}^H_{rl,j}\mathbf{G}_{rl,j})^{-1}\mathbf{G}^H_{rl,j},\ \mathbf{G}_{rl,j} \triangleq [\mathbf{h}_{rl,1,j}\ \mathbf{h}_{rl,2,j}],\\ \mathbf{A}_{d,j} &\triangleq (\mathbf{G}^H_{d,j}\mathbf{G}_{d,j})^{-1}\mathbf{G}^H_{d,j},\ \mathbf{G}_{d,j} \triangleq [\mathbf{h}_{d1,j}\ \mathbf{h}_{d2,j}\ \mathbf{H}_{dr,j}],\end{aligned} \tag{2}$$

for $i \in \{1, 2\}$ and $l \in \{1, 2\}$. Denote $\mathbf{A}_{r1,j} = [\mathbf{a}_{r1,1,j}, \mathbf{a}_{r1,2,j}]^H$, $\mathbf{A}_{r2,j} = [\mathbf{a}_{r2,1,j}, \mathbf{a}_{r2,2,j}]^H$, and $\mathbf{A}_{dj} = [\mathbf{a}_{d1,j}, \mathbf{a}_{d2,j}, \mathbf{A}_{dr1,j}, \mathbf{A}_{dr2,j}]^H$, where $\mathbf{a}_{rl,i,j}$ is an $N \times 1$ vector, $\mathbf{a}_{di,j}$ is an $M \times 1$ vector, and $\mathbf{A}_{drl,j}$ is an $M \times N$ matrix. After applying ZF detection in (2) to $y_{rl,j}$ and $y_{d,j}$ in (1), we obtain the 2×1 received vector $\tilde{\mathbf{y}}_{rl,j}$ at SCBS l and $(2 + 2N) \times 1$ received vector $\tilde{\mathbf{y}}_{d,j}$ at the MCBS as follows

$$\begin{aligned}&\tilde{\mathbf{y}}_{r1,j} = [\tilde{y}_{r1,1,j}\ \tilde{y}_{r1,2,j}]^T,\ \tilde{\mathbf{y}}_{r2,j} = [\tilde{y}_{r2,1,j}\ \tilde{y}_{r2,2,j}]^T,\\ &\tilde{\mathbf{y}}_{d,j} = [\tilde{y}_{d1,j}\ \tilde{y}_{d2,j}\ \tilde{\mathbf{y}}_{dr1,j}\ \tilde{\mathbf{y}}_{dr2,j}]^T,\\ &\tilde{y}_{rl,i,j} = x_{i,j} + \mathbf{a}^H_{rl,i,j}\mathbf{z}_{rl,j},\ i \in \{1, 2\},\ l \in \{1, 2\},\\ &\tilde{y}_{di,j} = x_{i,j} + \mathbf{a}^H_{di,j}\mathbf{z}_{d,j},\ \tilde{\mathbf{y}}_{drl,j} = \mathbf{x}_{rl,j} + \mathbf{A}^H_{drl,j}\mathbf{z}_{d,j}.\end{aligned} \tag{3}$$

where $\tilde{y}_{rl,i,j}$ (resp. $\tilde{y}_{di,j}$) is the data received at the SCBS l (resp. MCBS) from UE_i, and $\tilde{\mathbf{y}}_{drl,j}$ is the $N \times 1$ data vector received at the MCBS from SCBS l.

III. DF-MB-TD Scheme for Massive MIMO HetNets

We propose a simple yet efficient scheme which consists of ZF detection at each BS, DF relaying at the SCBSs with MB and TD transmission, and sliding window decoding at the MCBS with separate and sequential decoding. First, we explain some of the techniques used:

- To apply MB [9], the number of layers is equal to the number of SCBSs, i.e., two. In layer 1, each UE's message set is partitioned into equal-size groups and a bin index is given to each group. In layer 2, the set of binning indices in layer 1 are further partitioned in equal-size groups and a bin index (second layer) is given to each group. Consider the binning of UE_1 message set. Let n be the length of the transmitted codewords, while the transmission rates for the UE_1 message, the bin index of layer 1, and layer 2 are R_1, $R^{(1)}_{b1}$, and $R^{(2)}_{b1}$, respectively, where $0 < R^{(2)}_{b1} \le R^{(1)}_{b1} \le R_1$. Therefore, their corresponding set sizes are 2^{nR_1}, $2^{nR^{(1)}_{b1}}$, and $2^{nR^{(2)}_{b1}}$, respectively. In layer 1, the UE_1 message set is partitioned into $2^{nR^{(1)}_{b1}}$ groups, where each group is given a bin index and represents $2^{n(R_1 - R^{(1)}_{b1})}$ elements. Similarly, in layer 2, the set of layer 1 bin indices is partitioned into $2^{nR^{(2)}_{b1}}$ groups and each group has $2^{n(R^{(1)}_{b1} - R^{(2)}_{b1})}$ elements.
- In TD, different messages (codewords) are separately transmitted over different time slots (phases). The number of

phases is equal to the number of UEs. Hence, each SCBS transmits one layer of binning indices over two phases.

Next, we describe the scheme in detail.

1) Transmission Scheme: The proposed scheme is carried over B independent blocks and each UE sends $B-1$ messages through B transmission blocks. In block $j \in \{1, 2, ..., B\}$:

Each UE transmits a new message to both SCBSs and the MCBS, i.e., UE$_i$ sends its new message $w_{i,j}$ as follows.

$$x_{i,j} = \sqrt{P_i} U_i(w_{i,j}),\ i \in \{1, 2\}, \tag{4}$$

where P_i is the UE$_i$ transmit power and U_i is a Gaussian signal with zero mean and unit variance that conveys the $w_{i,j}$ codeword.

Each SCBS deploys ZF detection as in (3) and then separately decodes each UE message. Specifically, SCBS l utilizes $\tilde{y}_{rl,i,j}$ to decode (estimate) $\hat{w}_{i,j}$. For reliable decoding, R_i should satisfy the following constraint [13]:

$$R_i \leq \mathcal{C}\big(P_i N d_{rl,i}^{-\alpha}\big) \triangleq I_{l,i},\ i \in \{1, 2\},\ l \in \{1, 2\}, \tag{5}$$

where $\mathcal{C}(x) \triangleq \log(1 + x)$. Next,

1. SCBS1 finds the binning indices in layer 1 $(b_{1,j}^{(1)}, b_{2,j}^{(1)})$ that include $\hat{w}_{1,j}$ and $\hat{w}_{2,j}$, i.e., $\hat{w}_{1,j} \in b_{1,j}^{(1)}$ and $\hat{w}_{2,j} \in b_{2,j}^{(1)}$.

2. SCBS1 generates a separate codeword for each bin index in layer 1 (i.e., $\mathbf{U}_{r1,1}(b_{1,j}^{(1)})$ and $\mathbf{U}_{r1,2}(b_{2,j}^{(1)})$) and transmits them in block $j + 1$ over two phases of different durations $\beta_1^{(1)}$ and $\beta_2^{(1)}$, respectively where $\beta_1^{(1)} + \beta_2^{(1)} = 1$. Hence, for $i \in \{1, 2\}$, SCBS1 transmits the following signal in phase i:

$$\text{Phase } i\text{: } \mathbf{x}_{r1,i,j+1} = \sqrt{\rho_{r1,i}/(\beta_i^{(1)} N)}\,\mathbf{U}_{r1,i}(b_{i,j}^{(1)}), \tag{6}$$

where $\rho_{r1,1} + \rho_{r1,2} = P_{r1}$, P_{r1} is the transmit power of SCBS1, and $\rho_{r1,i}$ is the power allocated by SCBS1 to transmit $b_{i,j}^{(1)}$. Note that SCBS1 also deploys power control as it transmits with power $(\rho_{r1,i}/\beta_i^{(1)})$ in phase i.

3. SCBS2 finds the binning indices in layer 2 $(b_{1,j}^{(2)}, b_{2,j}^{(2)})$ such that $\hat{w}_{i,j} \in b_{i,j}^{(1)}$ and $b_{i,j}^{(1)} \in b_{i,j}^{(2)}$. Then, similar to SCBS1, SCBS2 transmits these binning indices over two phases of durations $\beta_1^{(2)}$ and $\beta_2^{(2)}$ (for $\beta_1^{(2)} + \beta_2^{(2)} = 1$) as follows.

$$\text{Phase } i\text{: } \mathbf{x}_{r2,i,j+1} = \sqrt{\rho_{r2,i}/(\beta_i^{(2)} N)}\,\mathbf{U}_{r2,i}(b_{i,j}^{(2)}), \tag{7}$$

where $\rho_{r2,1} + \rho_{r2,2} = P_{r2}$, P_{r2} is the transmit power of SCBS2, and $\rho_{r2,i}$ is the power allocated to transmit $b_{i,j}^{(2)}$.

The MCBS deploys ZF detection as in (3) but receives different SCBSs' signals in different phases, i.e., the received signal from SCBS l in phase i and block $j + 1$ is given as

$$\tilde{\mathbf{y}}_{drl,i,j+1} = \mathbf{x}_{rl,i,j+1} + \mathbf{A}_{drl,j+1}^H \mathbf{z}_{d,j+1}. \tag{8}$$

Then, the MCBS utilizes the signals from the SCBSs in block $j + 1$ and the UEs in block j to separately and sequentially decode each bin index of layer 2, 1 and then the UE message. For the UE$_1$ message, the MCBS sequentially decodes:

1. The bin index of layer 2 (i.e., $b_{1,j}^{(2)}$) using the received signal from SCBS2 in phase 1 (i.e., $\tilde{\mathbf{y}}_{dr2,1,j+1}$). Reliable decoding is ensured if $R_{b1}^{(2)}$ satisfies

$$R_{b1}^{(2)} \leq \beta_1^{(2)} N \mathcal{C}\Big(\rho_{r2,1}(M - 2N)/(N\beta_1^{(2)} d_{dr2}^{\alpha})\Big) \triangleq I_{dr2,1}. \tag{9}$$

2. The bin index of layer 1 (i.e., $b_{1,j}^{(1)}$) given that $b_{1,j}^{(1)} \in b_{1,j}^{(2)}$ using $\tilde{\mathbf{y}}_{dr1,1,j+1}$. Reliable decoding is ensured if

$$R_{b1}^{(1)} - R_{b1}^{(2)} \leq \beta_1^{(1)} N \mathcal{C}\Big(\frac{\rho_{r1,1}(M - 2N)}{N\beta_1^{(1)} d_{dr1}^{\alpha}}\Big) \triangleq I_{dr1,1}. \tag{10}$$

The constraint is on $R_{b1}^{(1)} - R_{b1}^{(2)}$ instead of $R_{b1}^{(1)}$ as the MCBS only looks for $b_{1,j}^{(1)}$ such that $b_{1,j}^{(1)} \in b_{1,j}^{(2)}$.

3. The UE$_1$ message $w_{1,j}$ given that $w_{1,j} \in b_{1,j}^{(1)}$ using the signal from UE$_1$ (i.e., $\tilde{y}_{d1,j}$). Reliable decoding is ensured if

$$R_1 - R_{b1}^{(1)} \leq \mathcal{C}\left(P_1(M - 2N) d_{d1}^{-\alpha}\right) \triangleq I_{d1}. \tag{11}$$

As in (10), the constraint is on $R_1 - R_{b1}^{(1)}$ since $w_{1.j} \in b_{1,j}^{(1)}$.

4. Similar decoding holds for the UE$_2$ message where the MCBS decodes $b_{2,j}^{(2)}$ from $\tilde{\mathbf{y}}_{dr2,2,j+1}$, $b_{2,j}^{(1)}$ from $\tilde{\mathbf{y}}_{dr1,2,j+1}$, and then $w_{2,j}$ from $\tilde{y}_{d2,j}$ at the following rates:

$$R_{b2}^{(2)} \leq I_{dr2,2},\ R_{b2}^{(1)} - R_{b2}^{(2)} \leq I_{dr1,2},\ R_2 - R_{b2}^{(1)} \leq I_{d2}. \tag{12}$$

$I_{dr2,2}$, $I_{dr1,2}$, and I_{d2} are similar to $I_{dr2,1}$, $I_{dr1,1}$, and I_{d1} in (9)–(11) but replacing $\rho_{r2,1}$, $\rho_{r1,1}$, $\beta_1^{(2)}$, $\beta_1^{(1)}$, P_1, and d_{d1} by $\rho_{r2,2}$, $\rho_{r1,2}$, $\beta_2^{(2)}$, $\beta_2^{(1)}$, P_2, and d_{d2}, respectively.

2) Achievable Rate Region: The rate constraints that ensure reliable decoding determine the rate region follows.

Theorem 1. *For the considered massive MIMO HetNet, the DF-MB-TD scheme achieves a rate region that consists of all rate pairs (R_1, R_2) satisfying*

$$R_1 \leq \min\{I_{1,1}, I_{2,1}, J_{d1}\},\ R_2 \leq \min\{I_{1,2}, I_{2,2}, J_{d2}\}, \tag{13}$$

with power allocation and phase durations satisfying $\rho_{rl,1} + \rho_{rl,2} = P_{rl}$ and $\beta_1^{(i)} + \beta_2^{(i)} = 1$. Here, $J_{di} = I_{di} + I_{dr1,i} + I_{dr2,i}$ and $I_{l,i}$, I_{d1}, $I_{dr1,1}$, $I_{dr2,1}$, and $(I_{d1}, I_{dr1,2}, I_{dr2,2})$ are given in (5), (11), (10), (9), and (12), respectively.

Proof: $I_{l,i}$ is obtained from (5), and J_{d1} (resp J_{d2}) is obtained by combining (9), (10), and (11) (resp. (12)). ■

Remark 1. Each separate and sequential decoding step at the MCBS is similar to the point-to-point communication.

Remark 2. The transmission rate of the binning indices are obtained from (9) and (10) as follows:

$$R_{bi}^{(2)} \leq I_{dr2,i},\ R_{bi}^{(1)} \leq I_{dr2,i} + I_{dr1,i},\ i \in \{1, 2\}. \tag{14}$$

Remark 3. Any order of the SCBSs leads to the same rate region. However, the transmission rates of the binning indices differ. Specifically, by switching the order of the SCBSs, i.e., SCBS1 (resp. SCBS2) transmits layer 2 (resp. layer 1) of the binning indices, the MCBS sequentially decodes the binning indices of layer 2 (resp. layer 1) using the received signal from SCBS1 (resp. SCBS2), and then each UE message. This decoding leads to the rate region in Theorem 1 but the binning index transmission rates in (14) become as follows:

$$R_{bi}^{(2)} \leq I_{dr1,i},\ R_{bi}^{(1)} \leq I_{dr2,i} + I_{dr1,i},\ i \in \{1, 2\}, \tag{15}$$

where $R_{bi}^{(2)}$ is different in (14) and (15) while $R_{bi}^{(1)}$ is the same.

Remark 4. (**Extension to general K UEs and L SCBSs**) Applying the scheme to a more general scenario with L SCBSs and K UEs, SCBS l transmits layer l of the binning indices for all UEs over K phases of durations $\beta_i^{(1)}, \beta_i^{(2)}, \ldots,$ and $\beta_i^{(L)}$. Then, for the UE$_i$ message $w_{i,j}$, the MCBS sequentially decodes the bin index in layers $L,\ L-1,\ \ldots, 1$ using the signals from SCBS L, $L-1$, ..., and 1 during phase i in block $j+1$ and then $w_{i,j}$ using the signal from UE$_i$. Hence, the transmission rate for UE$_i$ is given as follows:

$$R_1 \leq \min\{I_{1,i}, I_{2,i}, \ldots, I_{L,i}, J_{di}\}, \tag{16}$$

where $I_{l,i}$ is as given in (5) but with $i \in \{1, 2, \ldots, K\}$ and $l \in \{1, 2, \ldots, L\}$ while $J_{di} = I_{di} + \sum_{l=1}^{l=L} I_{drl,i}$.

3) Complexity: The complexity of a scheme is determined by its codebooksize, decoding computational complexity, and the encoding (trasnmission) design. Although the SCBSs' transmissons require optimization for the phase durations and power allocations, Remark 9 in Section IV shows that they are easily optimized and hence incur negligible computation.

Codebook size at the SCBSs: As the codeword generation at each UE is the same for the proposed DF-MB-TD scheme and the other schemes in Section IV, we only focus on the codebook size for the signals transmitted by the SCBSs. In the proposed DF-MB-TD, a separate codeword is generated for each bin index in layer l, $l \in \{1, 2\}$. Hence, the total number of codewords is $2^{2R_{b1}^{(1)}} + 2^{2R_{b1}^{(2)}} + 2^{2R_{b2}^{(1)}} + 2^{2R_{b2}^{(2)}}$.

Remark 5. While any SCBSs' order leads to the same rate region (Remark 3), from the binning rates in (14) and (15), the codebook size is reduced when ordering the SCBSs based on their link qualities (SNRs) to the MCBS, i.e., the SCBS with the highest SNR transmits the first binning layer.

Decoding complexity: At the MCBS, considering ML decoding, we define the decoding complexity as the number of likelihoods calculated at the MCBS to estimate the transmitted messages. Hence, the MCBS sequentially and separately decodes 1) $b_1^{(2)}$ and $b_2^{(2)}$ by calculating $2^{nR_{b1}^{(2)}} + 2^{nR_{b2}^{(2)}}$ likelihoods, 2) $b_1^{(1)} \in b_1^{(2)}$ and $b_2^{(2)} \in b_2^{(2)}$ by calculating $2^{n(R_{b1}^{(1)} - R_{b1}^{(2)})} + 2^{n(R_{b2}^{(1)} - R_{b2}^{(2)})}$ likelihoods, and 3) each UE's message by calculating $2^{n(R_1 - R_{b1}^{(1)})} + 2^{n(R_2 - R_{b1}^{(1)})}$ likelihoods. Therefore, the total calculations are $2^{n(R_1 - R_{b1}^{(1)})} + 2^{n(R_2 - R_{b2}^{(1)})} + 2^{n(R_{b1}^{(1)} - R_{b1}^{(2)})} + 2^{n(R_{b2}^{(1)} - R_{b2}^{(2)})} + 2^{nR_{b1}^{(2)}} + 2^{nR_{b2}^{(2)}}$.

Remark 6. The codebook size and the decoding complexity increase linearly with the number of UEs because of the separate codeword generation and separate decoding.

IV. Comparison with existing DF schemes

To illustrate the advantages of our scheme, we compare it with other DF schemes that are based on JD [8] or MB [9]. These two schemes in [8] and [9] can be modified for the considered massive MIMO HetNets as follows.

1) DF-JD scheme: In this scheme, each UE's transmission is identical to the DF-MB-TD scheme. However, in block j, each SCBS decodes both UEs' messages and then sends a common codeword for both messages to the MCBS during the whole block $j+1$. The MCBS simultaneously utilizes the received signals from both UEs in block j and both SCBSs in block $j+1$ to jointly decode both UEs' messages.

Theorem 2. *For massive MIMO HetNets, the rate region of the DF-JD scheme consists of all rate pairs (R_1, R_2) satisfying*

$$\begin{aligned} R_i &\leq \min\{I_{1,i}, I_{2,i}, I_{di} + \eta_1 + \eta_2\},\ i \in \{1, 2\} \\ R_1 + R_2 &\leq I_{d1} + I_{d2} + \eta_1 + \eta_2, \end{aligned} \tag{17}$$

where $I_{l,i}$ is as in (5), I_{d1} (resp. I_{d2}) is given in (11) (resp. (12)) and $\eta_l = N\mathcal{C}\left(P_{rl}(M-2N)/(Nd_{drl}^{\alpha})\right)$.

Proof: $I_{l,i}$ is given in Theorem 1, $I_{di} + \eta_1 + \eta_2$ (resp. $I_{d1} + I_{d2} + \eta_1 + \eta_2$) ensures reliable decoding of the UE$_i$ message (resp. both UEs' messages) at the MCBS. The sum rate constraint appears because of the JD at the MCBS. ∎

Codebook size: Each SCBS generates $2^{n(R_1+R_2)}$ codewords to represent each message pair and the two SCBSs sends independent codewords, i.e., $2^{n(R_1+R_2)+1}$ codewords in total.

Decoding complexity: The MCBS jointly decodes both UEs' messages, i.e., it calculates $2^{n(R_1+R_2)}$ likelihoods.

Remark 7. The codebook size and the decoding complexity of the DF-JD scheme increase exponentially with number of UEs because of the common codeword generation and the JD.

2) DF-MB scheme: It is similar to the DF-MB-TD scheme but without TD. Specifically, instead of separate transmission, SCBS l transmits a common codeword for both binning indices in layer l. The MCBS jointly decodes both binning indices in layer l using the signal from SCBS l. Finally, it separately decodes each UE's message using the signals from the UEs.

Theorem 3. *The rate region of the DF-MB scheme is identical to that of the DF-JD scheme in Theorem 2.*

Proof: $I_{l,i}$ is given in Theorem 1. Next, the MCBS decodes $b_1^{(2)}$ and $b_2^{(2)}$ reliably if $R_{b1}^{(2)} + R_{b2}^{(2)} \leq \eta_2$, $b_1^{(1)}$ and $b_2^{(1)}$ reliably if $R_{b1}^{(1)} - R_{b1}^{(2)} + R_{b2}^{(1)} - R_{b2}^{(2)} \leq \eta_1$, and then each UE message as in the DF-MB-TD scheme. By combining these constraints, we obtain the rate region in Theorem 2. ∎

Codebook size: SCBS l generates $2^{n(R_{b1}^{(l)} + R_{b2}^{(l)})}$ common codewords to represent each bin index pair in layer l. Hence, the total number of codewords is $2^{n(R_{b1}^{(l)} + R_{b2}^{(1)})} + 2^{n(R_{b1}^{(2)} + R_{b2}^{(2)})}$.

Decoding complexity: The MCBS decodes in three steps that require the following calculations as the MCBS 1) jointly decodes $b_1^{(2)}$ and $b_2^{(2)}$ by calculating $2^{n(R_{b1}^{(2)} + R_{b2}^{(2)})}$ likelihoods; 2) jointly decodes $b_1^{(1)} \in b_1^{(2)}$ and $b_2^{(2)} \in b_2^{(2)}$ by calculating $2^{n((R_{b1}^{(1)} - R_{b1}^{(2)}) + (R_{b2}^{(1)} - R_{b2}^{(2)}))}$ likelihoods; and 3) separately decodes each UE's message by calculating $2^{n(R_1 - R_{b1}^{(1)})} + 2^{n(R_2 - R_{b2}^{(1)})}$ likelihoods. Hence, the total calculations: $2^{n(R_1 - R_{b1}^{(1)})} + 2^{n(R_2 - R_{b2}^{(1)})} + 2^{n((R_{b1}^{(1)} - R_{b1}^{(2)}) + (R_{b2}^{(1)} - R_{b2}^{(2)}))} + 2^{n(R_{b1}^{(2)} + R_{b2}^{(2)})}$.

Remark 8. The complexity of the DF-MB scheme grows exponentially with UEs' number due to the common codeword generation and JD for all binning indices in each layer.

3) Comparison with the DF-MB-TD scheme: We compare the three schemes with the following result.

Theorem 4. *The DF-MB-TD scheme achieves the same rate region of DF-JD and DF-MB schemes with linear complexity.*

Proof: See Remarks 6—8 for complexity comparison. For rate region comparison, by considering (13) with (17), $I_{l,i}$ for

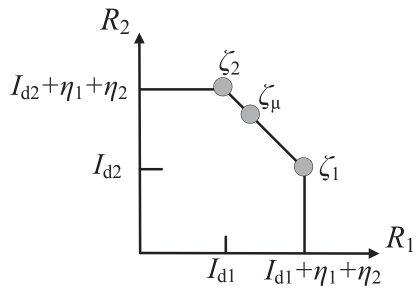


Fig. 3. Rate Region 2 in (18).

$i \in \{1,2\}$ and $l \in \{1,2\}$ are identical. Then, these regions are equivalent if the following regions are equivalent:

$$\text{Region 1 (DF-MB-TD): } R_1 \leq J_{d1},\ R_2 \leq J_{d2} \tag{18}$$
$$\text{Region 2 (DF-JD): } R_1 \leq I_{d1} + \eta_1 + \eta_2,\ R_2 \leq I_{d2} + \eta_1 + \eta_2,$$
$$R_1 + R_2 \leq I_{d1} + I_{d2} + \eta_1 + \eta_2.$$

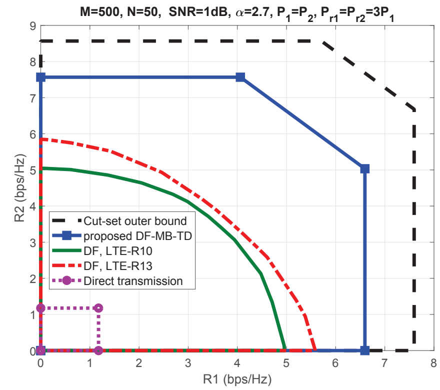
In (18), Region 2 is a pentagon (see Fig. 3) and Region 1 can achieve all points in Fig. 3 by setting $\rho_{rl,i} = \beta_i^{(l)} P_{ri}$ while varying $\beta_i^{(l)}$, i.e., the point ζ_1 (resp. ζ_2) is achieved with $\beta_1^{(1)} = \beta_1^{(2)} = 1$ (resp. 0) and $\beta_2^{(1)} = \beta_2^{(2)} = 0$ (resp. 1) while any point (ζ_μ) for $\mu \in [0,1]$ is achieved by time sharing [10] ($\beta_1^{(1)} = \beta_1^{(2)} = \mu$ and $\beta_2^{(1)} = \beta_2^{(2)} = 1 - \mu$). ■

Remark 9. There are several implications from the proof in Theorem 4:

- **Same phase durations at both SCBSs**: The proposed DF-MB-TD scheme achieves the rate region of the DF-JD scheme by varying μ between 0 and 1 while setting $\beta_1^{(1)} = \beta_1^{(2)} = \mu$ and $\beta_2^{(1)} = \beta_2^{(2)} = 1 - \mu$. Hence, using the same set of phase durations is sufficient to maximize the rate region.
- **Same transmit power at each phase**: The DF-MB-TD scheme achieves the performance of DF-JD with $\rho_{rl,i} = \beta_i^{(l)} P_{ri}$, being substituted into (6) and (7), each SCBS transmits with power P_r in each phase. Hence, there is no need for different power allocation or power control.

V. Numerical Results

Fig. 4 compares the rate regions of the massive MIMO HetNet for the proposed DF-MB-TD scheme (which is identical to DF-JD and DF-MB) for $K = L = 2$, with LTE-R10 [15], LTE-R13 [16], direct transmission (i.e., without the SCBSs), and the cut-set outer bound [10]. In the LTE standards, 1) each SCBS works in half-duplex mode and deploys DF relaying for its own UE only, 2) two close SCBSs transmit in different bands, and 3) the MCBS uses signals from SCBSs only for decoding in LTE-R10 [15], while it can also use the UEs' signals in LTE-R13 [16]. Our proposed scheme outperforms the LTE schemes because of the concurrent transmission from the UEs and SCBSs, and full-duplex DF relaying at each SCBS for both UEs' messages.



;

Fig. 4. Rate regions of DF-MB-TD (DF-JD or DF-MB), LTE-R10, LTE-R13, and direct transmission, where $d_{r1,i} = 9$, $d_{r2,i} = 7$, $d_{di} = 105$ and $d_{drl} = 100$ in meters for $i \in \{1,2\}$ and $l \in \{1,2\}$, and P_i is related to the SNR as $\text{SNR} = 10\log_{10}(P_1(M-N)/d_{d1}^{\alpha})$.

VI. Conclusion

We have proposed a new uplink transmission scheme for the massive MIMO HetNet. It is based on DF relaying at each SCBS with multiple layers of binning and time division transmission, and separate and sequential decoding at the MCBS. The proposed scheme has a linear complexity w.r.t the number of UEs, and ordering the SCBSs based on their link qualities to the MCBS minizes the codebook size. Furthermore, allocating the same set of phase durations at all SCBSs is sufficient to maximize the rate region. Despite its simplicity, the proposed scheme achieves the same rate region of more advanced schemes with exponential complexity.

References

[1] M. Jaber *et al.*, "5G backhaul challenges and emerging research directions: A survey," *IEEE Access*, vol. 4, pp. 1743–1766, May 2016.

[2] K. Loa *et al.*, "IMT-advanced relay standards [WiMAX/LTE Update]," *IEEE Commun. Mag.*, vol. 48, no. 8, pp. 40–48, Aug. 2010.

[3] H.-H. Yang, G. Geraci, and Q. S. T. Quek, "Energy-efficient design of MIMO heterogeneous networks with wireless backhaul," *IEEE Trans. Wireless. Commun.*, vol. 15, no. 7, pp. 4914–4927, Jul. 2016.

[4] M. Feng, S. Mao, and T. Jiang, "Joint frame design, resource allocation and user association for massive MIMO heterogeneous networks with wireless backhaul," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1937–1950, Mar. 2018.

[5] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, 2016.

[6] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, May 2011.

[7] G. O. Okeke *et al.*, "Novel low-complexity joint user-relay selection and association for multi-user multi-relay MIMO uplink," *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 309–312, Jun. 2015.

[8] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.

[9] M. Aref, "Information flow in relay networks," *Ph.D. dissertation, Stanford Univ.*, Oct. 1980.

[10] R. El Gamal and Y.-H. Kim, *Network Information Theory*, 1st ed. Cambridge University Press, 2011.

[11] E. Bjrnson *et al.*, "Massive MIMO: ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[12] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1893–1909, Sept. 2004.

[13] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[14] D. Bharadia, E. Mcmilin, and S. Katti, "Full duplex radios," *ACM, SIGCOMM*, vol. 43, no. 4, pp. 375–386, Aug. 2013.

[15] J. Korhonen, *Introduction to 4G mobile communications*. Artech House, 2014.

[16] C. Rosa *et al.*, "Dual connectivity for LTE small cell evolution: functionality and performance aspects," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 137–143,, Jun. 2016.