Hybrid MU-MIMO and Non-orthogonal Multiple Access Design in Wireless Heterogeneous Networks

Yiran Xu, Haijian Sun, Rose Qingyang Hu
Department of Electrical and Computer Engineering
Utah State University, Logan, UT, USA
Emails: {yiran.xu, h.j.sun, rosehu}@ieee.org

***Abstract*—In order to meet the ever increasing mobile application proliferation and data traffic growth, next-generation wireless networks or 5G networks are under a revolutionary technology innovation path towards these goals. In this paper, we introduce a hybrid MU-MIMO and NOMA design scheme in wireless heterogeneous networks to improve the system throughput and also to increase multi-user diversity gains by exploiting the heterogeneous nature of the supporting wireless networks. The best user cluster is formed in a NOMA group and then a precoding based MU-MIMO scheme is applied to NOMA composite signals. The problem is further formulated as a resource scheduling optimization problem to achieve the proportional fairness. Aiming to ensure the global optimality, a brute-force search algorithm is used to solve the problem. Simulations results show that the proposed scheme can improve the overall system performance notably.**

I. Introduction

The widespread popularity of smart phones and tablets is triggering explosive mobile application proliferation and data traffic growth. Based on the forecast data, global mobile traffic grew 69% in 2014, which was nearly 30 times the size of the entire global Internet in 2000, and it will increase nearly 10-fold by 2019. In contrast, the average data speed will only increase 19% annually in the next five years [1]. Clearly there exists a huge gap between the growth rate from air interface technologies and the growth rate of customer needs. To maintain mobile service profitability, and narrow the gap between increasing demands and scarce network resources, it is necessary to explore the potential benefits of new network architecture and wireless technologies simultaneously.

In traditional cellular networks, a base station (BS) consumes a significant amount of resources to support the activities of user equipments (UEs), especially cell edge UEs. Emerging high-density, heterogeneous wireless networks introduce a hierarchical infrastructure, where high power BSs provide a blanket coverage and seamless mobility while low power nodes, such as femto and pico BSs, are usually deployed at coverage holes or capacity-demanding hotspots and can greatly extend the wireless service coverage range and expand the cell capacity [2]- [3]. For densely deployed small cells in a heterogeneous network, cell-edge users might suffer from severe interference from neighboring BSs. Thereby, precoding based multi-user multiple-input and multiple-output (MU-MIMO) is applied in heterogeneous network to mitigate the interference and improve the throughput. In [4], authors applied intra-cell cooperation in a heterogeneous network to improve the system throughput and cell-edge performance. Furthermore, [5] combined precoding technique with intra-cell cooperation to increase cell-edge user's achievable data rates, so that the overall system performance was improved considerably.

Non-orthogonal multiple access (NOMA) is considered as a future radio access technology candidate [6] in 5G networks. NOMA can explore power domain for multiple access [7] with advanced transmission/reception technique such as success interference cancellation. Particularly, NOMA allocates the same spectrum to different users, where different users are served with different power levels. At the receiving side, based on the received signal strengths, UEs employ successive interference cancellation to remove the signals intended for other UEs before decoding their own. Since more information can be delivered by sharing the same spectrum resource among different users, NOMA can potentially achieve a high spectral efficiency and increase the total system throughput. In [8], a system-level study on downlink NOMA was conducted and compared with traditional orthogonal multiple access. Simulation results demonstrated the performance gain of NOMA over traditional orthogonal multiple access techniques. [9] explored NOMA in a heterogeneous network and proposed a cooperative NOMA scheme so that the system can achieve diversity gains at the receiver side. UEs with cooperative NOMA can receive information from both macro-node and pico-node simultaneously, thereby, the total system throughput was increased. There exist some limitations in NOMA [7] [8]. If the received signals among different UEs have relative large power disparity, NOMA is able to achieve a good performance gain on overall throughput. When the received power difference becomes small among the received signals, NOMA gain diminishes. Alternatively, MU-MIMO can work well under this situation given that there is enough channel diversity. Thus, in this paper, we introduce a hybrid MU-MIMO and NOMA design scheme to improve the system throughput and to increase multi-user diversity gains by exploiting the heterogeneous nature of the wireless HetNets. The best user cluster is formed in a NOMA group and then a precoded MU-MIMO scheme is applied to the superposed signals. The problem is further formulated as a resource scheduling optimization problem with proportional fairness purpose. Aiming to ensure the global optimality, a brute-force search algorithm is used

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

to solve it.

The remainder of this paper is organized as follows. Section II introduces a downlink wireless heterogeneous model. Specifically, we present the details on the hybrid design of MU-MIMO and NOMA. A resource scheduling optimization problem is formulated in Section III. To properly form MONA pairs at each scheduling circle, a brute-force search algorithm is implemented in Section IV. In Section V, simulations results are presented. The paper is concluded in Section VI.

II. System Model

We consider downlink communications in a wireless heterogeneous network in Fig. 1, where each macro-cell is equipped with one high-power macro BS (mBS) and several overlaid lower-power pico BSs (pBSs). We denote N_m as the total number of mBSs in the system and N_p as the number of pBSs per macro-cell. N_u UEs are uniformly distributed in the network, so that each UE can be served by either an mBS or a pBS, or both, depending on the location and service requirement of the UE. All the UEs and BSs are equipped with single antenna. Compared to an mBS, a pBS typically has a much lower transmit power and a smaller coverage range. We denote the transmit powers for an mBS and a pBS as P_m and P_p, respectively. The overlaid pico-cells reuse the same spectrum of the macro-cells and aim to provide services at hotspots and coverage holes, such that the overall system spectrum efficiency, energy efficiency and coverage are greatly improved.

As mentioned in [7] [8], NOMA usually achieves great performance gains if there exist relatively large disparities between the received signals among a cluster of users. For users with small difference in received signal strengths, NOMA might not provide any performance gain. As illustrated in Fig. 1, we categories overall service areas into three ranges, namely mBS MU-MIMO+NOMA range, mBS+pBS MU-MIMO range, and pBS MU-MIMO+NOMA range, based on the association scheme and the received power levels from mBS and pBS. For UEs located in the MU-MIMO range, UEs receive relatively equal signal powers from both mBS and pBS and thus MU-MIMO is favorable. For UEs located in MU-MIMO+NOMA ranges, signals received from mBS and pBS are largely different, making hybrid MU-MIMO+NOMA as a more spectrum efficient transmission mechanism than either NOMA alone or MU-MIMO alone. In a wireless heterogeneous network, due to the high disparity on the powers from different BSs, a high percentage of downlink UEs locate in the regions where interference power is even stronger than the intended signal power. So using a hybrid MU-MIMO+NOMA scheme in these regions is highly desirable since it turns a destructive interference issue into a constructive contributor. In this paper, for the sake of clarity on presentation, we assume all the UEs and BSs have only one antenna. However, the algorithm is applied to the general multi-antenna case as well.

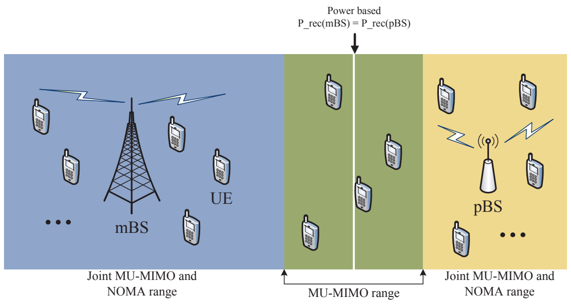

Fig. 1. Wireless Heterogeneous Network

A. MU-MIMO

We first address how MU-MIMO works in the MU-MIMO only range. Without loss of generality, we formulate the channel matrix for two users that form an MU-MIMO pair as $\mathbf{H}_{1,2}$ as:

$$\mathbf{H}_{1,2} = \begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \end{bmatrix}, \tag{1}$$

where channel gain $h_{i,j}$ considers both path-loss and Rayleigh fading. We can simply use $\mathbf{H}$ to represent $\mathbf{H}_{i,j}$. The received downlink signal vector at two UEs can be expressed as $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$, where $\mathbf{z} = [z_1 \quad z_2]^T$ represents the noise vector at receiver side. In order to mitigate inter-user interference, we assume perfect channel state information at both mBS and pBS. Then we apply dirty paper coding (DPC) [10] by designing the precoding matrix as $\mathbf{W} = \mathbf{Q}^H\mathbf{G}$, where $\mathbf{Q}^H$ is the Hermitian matrix of $\mathbf{Q}$ and can be obtained by proceeding LQ decomposition to $\mathbf{H}$:

$$\mathbf{H} = \begin{bmatrix} l_{1,1} & 0 \\ l_{2,1} & l_{2,2} \end{bmatrix} \begin{bmatrix} \mathbf{q_1} \\ \mathbf{q_2} \end{bmatrix} = \mathbf{LQ}. \tag{2}$$

Here, $\mathbf{L}$ is a lower triangle matrix. $\mathbf{q_1}$ and $\mathbf{q_2}$ are row vectors, respectively. In order to form an interference free transmission channel, $\mathbf{G}$ is given as

$$\mathbf{G} = \begin{bmatrix} 1 & 0 \\ -\frac{l_{1,2}}{l_{2,2}} & 1 \end{bmatrix}. \tag{3}$$

At each mBS and each pBS, the signal vector $\mathbf{x} = [x_1 \quad x_2]^T$ is precoded to $\hat{\mathbf{x}} = \mathbf{W}\mathbf{x} = [\hat{x}_1 \quad \hat{x}_2]^T$ before transmission, where $\hat{x}_1$ is the precode signal transmitted from mBS and $\hat{x}_2$ is from pBS. Thereby, the received signal $\mathbf{y} = [y_1 \quad y_2]^T$ is expressed as

$$\begin{aligned} \mathbf{y} &= \mathbf{H}\hat{\mathbf{x}} + \mathbf{z} = \mathbf{HW}\mathbf{x} + \mathbf{z} \\ &= \begin{bmatrix} l_{1,1} & 0 \\ 0 & l_{2,2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}. \end{aligned} \tag{4}$$

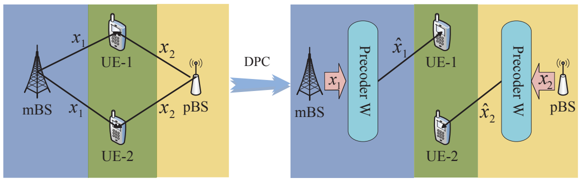

Fig. 2. Transmission Model for MU-MIMO only

It is observed that the inter-user interference can be canceled out in the ideal case, which is illustrated in Fig. 2. Then the total achievable data rate from MU-MIMO is R^c:

$$R^c = R^c_{k_1} + R^c_{k_2}, \tag{5}$$

$$R^c_{k_1} = W \log_2 \left(1 + \frac{|l_{1,1}|^2 P_m}{N_0}\right), \tag{6}$$

$$R^c_{k_2} = W \log_2 \left(1 + \frac{|l_{2,2}|^2 P_p}{N_0}\right). \tag{7}$$

Here, $l_{m,m}$, $m = 1, 2$, represents the equivalent channel gain between UE 1 (or UE 2) and mBS (or pBS). W is denoted as the bandwidth of one resource block (RB).

B. Hybrid MU-MIMO and NOMA

In the MU-MIMO+NOMA range, an MU-MIMO+NOMA pair consists of 2 UEs, one from mBS and one from pBS. Each BS can transmit two different signals to 2 UEs, with one close to itself and the other one further away, thus forming a desirable downlink NOMA pair. As shown in Fig. 3, x_1 and x_4 are intended to UE 1, and x_2 and x_3 are intended to UE 2, so that in total 4 signals are transmitted to two UEs by using hybrid MU-MIMO+NOMA, compared with only 2 signals to 2 UEs in the MU-MIMO only case. In order to transmit two different signals from a single antenna, power disparity between transmitted signals needs to be imposed in order for the receiving side to achieve a notable NOMA gain. Therefore, we introduce a power allocation parameter $\theta \in (0, 1)$ to partition the transmit power at each BS.

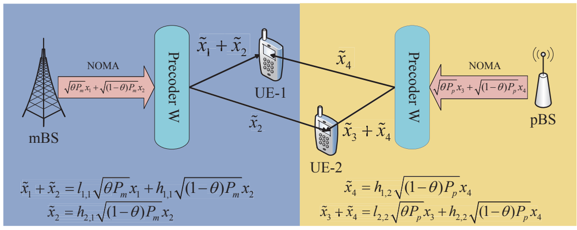
NOMA
$\sqrt{\theta P_m} x_1 + \sqrt{(1-\theta) P_m} x_2$
mBS
Precoder W
$\tilde{x}_1 + \tilde{x}_2$
UE-1
$\tilde{x}_2$
$\tilde{x}_4$
UE-2
$\tilde{x}_3 + \tilde{x}_4$
Precoder W
NOMA
$\sqrt{\theta P_p} x_3 + \sqrt{(1-\theta) P_p} x_4$
pBS
$\tilde{x}_1 + \tilde{x}_2 = l_{1,1}\sqrt{\theta P_m} x_1 + h_{1,1}\sqrt{(1-\theta) P_m} x_2$
$\tilde{x}_2 = h_{2,1}\sqrt{(1-\theta) P_m} x_2$
$\tilde{x}_4 = h_{1,2}\sqrt{(1-\theta) P_p} x_4$
$\tilde{x}_3 + \tilde{x}_4 = l_{2,2}\sqrt{\theta P_p} x_3 + h_{2,2}\sqrt{(1-\theta) P_p} x_4$

Fig. 3. Transmission Model for Hybrid MU-MIMO+NOMA

On the transmitting side, each BS first uses NOMA to superimpose two signals together and then precodes the superimposed signal before sending it out. Without precoding, UE 1 will receive x_1 and x_4 as intended signals while receive x_2 and x_3 as interference. With precoding, UE 1 will still receive x_1 and x_4 as intended signals. But its interference signal can be reduced to x_2 only. The same applies to UE 2 as well.

After the precoding operation, the received signal vector $\mathbf{y}$ is expressed in (8). Here, we assume the first row is the received signal at UE 1, and the second row is the received signal at UE 2. DPC is used to reduce the inter-user interference so that one of the interfering signals is canceled out at each UE. Thus each UE receives a composite signal consisting of three signals, two of which are intended signals. On the receiving side, we apply SIC to retrieve the signals sequentially by following the descending order of the received signal strength. We denote A as the sum power of the received signals that have power lower than x_1, B as the sum power of the received signals that have power lower than x_2, C as the sum power of the received signals that have power lower than x_3, and D as the sum power of the received signals that have power lower than x_4. Then the total achievable data rate R^n can be expressed as

$$R^n = R^n_{k_1} + R^n_{k_2}, \tag{9}$$

$$\begin{aligned} R^n_{k_1} &= W \log_2 \left(1 + \frac{l^2_{1,1} \theta P_m}{A + N_o}\right) \\ &+ W \log_2 \left(1 + \frac{h^2_{1,2}(1-\theta) P_p}{D + N_o}\right), \end{aligned} \tag{10}$$

$$\begin{aligned} R^n_{k_2} &= W \log_2 \left(1 + \frac{h^2_{2,1}(1-\theta) P_m}{B + N_o}\right) \\ &+ W \log_2 \left(1 + \frac{l^2_{2,2} \theta P_p}{C + N_o}\right). \end{aligned} \tag{11}$$

III. Problem Formulation

The objective is to design a dynamic transmission mechanism that can maximize the overall system throughput and deliver satisfactory user experience. Towards that end, at each scheduling cycle, the scheme needs to: 1) decide MU-MIMO with/without NOMA group pair; 2) adjust the transmit power allocation factor θ to maximize MU-MIMO+NOMA performance gain; 3) allocate RBs to UE pairs.

The algorithm will first need to select the transmission mode, i.e., either MU-MIMO only or hybrid MU-MIMO and NOMA for each RB of each BS among all the candidate UEs in the system. Once the BS pair (mBS i, pBS (i, j)) is determined, the rest pBSs (i, j'), $\forall j' \neq j$, will switch to muting mode so that there is no intra-cell interference. Here mBS i represents the mBS in cell i and pBS (i, j) represents the jth pBS in cell i. Each cell can have multiple pBSs. The following scheduling variables are defined. $x^c_{i,0,k_1}(f, t) = 1$ if UE k_1 is served by mBS i on f as the 1st UE in an MU-MIMO pair at t and $x^c_{i,0,k_1}(f, t) = 0$ otherwise. $x^c_{i,j,k_2}(f, t)$ indicates the connection status with the 2nd UE in the MU-MIMO pair; Similarly, $x^n_{i,0,k_1}(f, t) = 1$ if UE k_1 is served by mBS i on f as the 1st UE in a hybrid MU-MIMO and NOMA pair at t and 0 otherwise. $x^n_{i,j,k_2}(f, t)$ is the 2nd UE association status with the hybrid scheme.

Furthermore, in order to determine whether a RB should be assigned to an MU-MIMO pair or a hybrid MU-MIMO and NOMA pair, we introduce the following proportional fairness

$$\begin{aligned}\mathbf{y} &= \mathbf{H}\hat{\mathbf{x}}_{1,3} + \mathbf{H}\mathbf{x}_{2,4} + \mathbf{z} = \mathbf{H}\mathbf{W}\mathbf{x}_{1,3} + \mathbf{H}\mathbf{x}_{2,4} + \mathbf{z} = \mathbf{L}\mathbf{Q}\mathbf{Q}^H\mathbf{G}\mathbf{x}_{1,3} + \mathbf{H}\mathbf{x}_{2,4} + \mathbf{z} \\ &= \begin{bmatrix} l_{1,1}\sqrt{\theta P_m}x_1 + h_{1,1}\sqrt{(1-\theta)P_m}x_2 + h_{1,2}\sqrt{(1-\theta)P_p}x_4 \\ l_{2,2}\sqrt{\theta P_p}x_3 + h_{2,1}\sqrt{(1-\theta)P_m}x_2 + h_{2,2}\sqrt{(1-\theta)P_p}x_4 \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}. \qquad (8)\end{aligned}$$

(PF) function:

$$U_k(t) = \frac{R_k^\alpha(t)}{T_k^\beta(t)}, \quad T_k(t) = \frac{1}{T_c}\sum_{\tau=t-T_c+1}^{t} R_k(\tau). \qquad (12)$$

Here, $R_k(t)$ is denoted as the instantaneous data rate of UE pair k_1 and k_2 at time t, and $T_k(t)$ is denoted as the window based moving average throughput for UE pair at time t. T_c is the moving average window size. α and β tune the "fairness" of the scheduler. From (12), if a UE gets a low throughout in the past, its PF function value $U_k(t)$ will be elevated so that its priority to be served increases. Thereby, by properly adjusting α and β, we can ensure that when maximizing spectrum efficiency, UEs in different regions can still be served fairly. According to (5) and (9), the achievable data rate at time t on RB f is expressed as:

$$\begin{aligned} R_k(t) &= x^c_{i,0,k_1}(f,t)(1 - x^n_{i,0,k_1}(f,t))R^c_{k_1} \\ &+ x^c_{i,j,k_2}(f,t)(1 - x^n_{i,j,k_2}(f,t))R^c_{k_2} \\ &+ x^n_{i,0,k_1}(f,t)(1 - x^c_{i,0,k_1}(f,t))R^n_{k_1} \\ &+ x^n_{i,j,k_2}(f,t)(1 - x^c_{i,j,k_2}(f,t))R^n_{k_2}. \qquad (13)\end{aligned}$$

The objective function of the scheduling problem is thus formulated as

$$[\mathbf{P}_1] \ \max_{\mathbf{x}(t)} \sum_k U_k(t) \qquad (14)$$

subject to

$$\sum_{k=1}^{N_u} x^c_{i,0,k_1}(f,t) + x^n_{i,0,k_1}(f,t) \ \leq \ 1, \ \forall i, f \qquad (15)$$

$$\sum_{k=1}^{N_u} x^c_{i,j,k_2}(f,t) + x^n_{i,j,k_2}(f,t) \ \leq \ 1, \ \forall i, j, f \qquad (16)$$

Constraints (15) and (16) ensure that at each time slot, each RB can be assigned to only one pair of UEs, either an MU-MIMO+NOMA pair or an MU-MIMO pair.

IV. Brute-force Search Algorithm

As a preliminary study on the hybrid MU-MIMO and NOMA framework, we first consider a system with a low number of BSs and UEs. Therefore, it is possible to apply a brute-force search algorithm with reasonable computational complexity to solve the aforementioned problem. Specifically, in each resource scheduling circle, we search all the UE pairs and form them as either MU-MIMO pairs or MU-MIMO+NOMA pairs, based on which we can compute their objective function values, and choose the UE pair with highest value as the solution. The brute-force search algorithm in summarized in **Algorithm** 1.

V. Simulation Results and Performance Analysis

The simulation was set up based on 3GPP case 1 configurations specified in [11]. A single cell network structure is divided into three sectors by 120 degree equally. Each sector represents a macrocell, in which one mBS is located in the center and 4 pBSs are equally-distanced deployed within each macrocell, forming a two-tier heterogeneous network. UEs are uniformly distributed in the network. Small scale fading is generated based on the Rayleigh fading channel model [12]. Other parameter settings are shown in Table I.

In Fig. 4, we investigate the MU-MIMO+NOMA performance with different power allocation factors. It is observed that as θ decreases, UEs have a relatively higher average data rate. For example, compared to $\theta = 0.6$, about 80% of the total UEs at $\theta = 0.2$ have an increase of 5000 kbps in average data rate. This is because a small θ reflects a relatively large power disparity within a UE pair. Thereby, implementation of hybrid MU-MIMO+NOMA can deliver additional information to UEs and improve the system throughput considerably. In contrast, with the increase of θ, the received power disparity is not distinct. Then MU-MIMO+NOMA no longer contributes notable performance gains. Hence, more UEs are formed as MU-MIMO pairs. When $\theta = 1$, the system evolves into a pure MU-MIMO system. Specifically, Fig. 5 compares the performance between of MU-MIMO users with the hybrid MU-MIMO+NOMA users. It is shown that hybrid MU-MIMO+NOMA users have relatively higher average data rates than MU-MIMO users. This is because the existence of NOMA introduces the diversity gains, and additional information can be transmitted to UEs with the sharing spectrum resources. Therefore, it leads to a leap on the system performance in terms of users' data rates.

TABLE I
Simulation Parameter Settings

Parameter	Settings
mBS	1
pBS	4 per macro-cell
UE	200 per cell
Transmitting Antenna	1 per BS
Receiving Antenna	1 per UE
Transmit Power	P_m = 30Watt, P_p = 1Watt
System Bandwidth	10 MHz
Number of RBs	$F = 50$
Bandwidth of RB	W = 180kHz
Size of Time Window	T_c = 100 seconds
Fast Fading Model	Rayleigh Fading Channel [12]
Shadowing	8dB, log-normal std. deviation
Noise Model and density	AWGN, -174dBm/Hz

Algorithm 1 Brute-force Search Algorithm

```
Initialization: Given total number of UEs $N_u$, generate all possible UE pairs. Denote the set of total pairs as $\mathcal{P}_{N_u}$
Convergence = false.
for $t = t_0$ to $T$ do
  for $f = 1$ to $F$ do
    for $p = 1$ to $|\mathcal{P}_{N_u}|$ do
      Identify UE pair indexes $(k_1, k_2)_p \in \mathcal{P}_{N_u}$
      Assume a MU-MIMO pair
      Calculate the objective function value:
          $U_p^c = U_{k_1} + U_{k_2}$    (17)
      if $U_p^c \geq U_{p-1}^c$ then
        $U_{p^*}^c = U_p^c$, $(k_1^c, k_2^c) = (k_1, k_2)_p$
      else
        $U_{p^*}^c = U_{p-1}^c$, $(k_1^c, k_2^c) = (k_1, k_2)_{p-1}$
      end if
      Assume a MU-MIMO+NOMA pair
      Calculate the objective function value:
          $U_p^n = U_{k_1} + U_{k_2}$    (18)
      if $U_p^n \geq U_{p-1}^n$ then
        $U_{p^*}^n = U_p^n$, $(k_1^n, k_2^n) = (k_1, k_2)_p$
      else
        $U_{p^*}^n = U_{p-1}^n$, $(k_1^n, k_2^n) = (k_1, k_2)_{p-1}$
      end if
    end for
    Compare $U_{p^*}^c$ and $U_{p^*}^n$
    Determine the transmission mode:
    $(k_1^{t,f}, k_2^{t,f}) = \arg_{k_1,k_2} \left\{ U_{p^*}^c(k_1^c, k_2^c), U_{p^*}^n(k_1^n, k_2^n) \right\}$    (19)
    Assign RB $f$ to UE pair $(k_1^{t,f}, k_2^{t,f})$
    Update average data rate $T_k(t)$
  end for
  Update average data rate $T_k(t)$
end for
Output UEs' average data rates, UEs' transmission modes
```

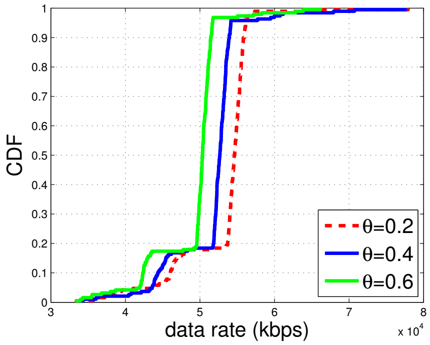

Fig. 4. The CDF of User Average Data Rate With Different Power Allocation Factor θ

VI. Conclusions

In this paper, we investigate a hybrid MU-MIMO+NOMA scheme in a wireless heterogeneous network. A proportional fair resource scheduling problem is formulated to justify the advantage of hybrid scheme. A brute-force search algorithm is applied to solve the problem. Simulation results show that the heterogeneous network can receive considerable benefits from the hybrid MU-MIMO and NOMA implementation. In the future, brute-force search might be inadequate due to its high computational complexity and time consumption. Therefore, it is necessary to explore advanced scheduling and pairing methods. Moreover, it is also necessary to consider the hybrid application of MU-MIMO and NOMA in multi-antenna multi-cell systems with inter-cell interferences.

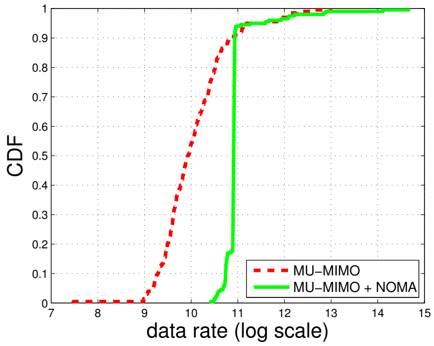

Fig. 5. Comparison Between MU-MIMO and MU-MIMO + NOMA.

Acknowledgment

This work was supported by the National Science Foundation under the grants ECCS-1308006, EARS-1547312.

References

[1] White_paper_c11-520862, “Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014-2019 ,” Feb. 2015.

[2] R. Q. Hu and Y. Qian, *Heterogeneous Cellular Networks*, John Wiley & Sons, Ltd., 2013.

[3] R. Q. Hu, Y. Qian, “An energy efficient and spectrum efficient wireless heterogeneous network framework for 5G systems,” *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 94-101, May 2014.

[4] Y. Xu, and R. Q. Hu, “Optimal intra-cell cooperation in the heterogeneous relay networks,” in *Proc. IEEE GLOBECOM 2012*, pp. 4120-4125, Dec., 2012.

[5] Y. Xu, R. Q. Hu, Q. Li, and Y. Qian, “Optimal intra-cell cooperation with precoding in wireless heterogeneous networks,” in *Proc. IEEE WCNC 2013*, pp. 761-766, Apr. 2013.

[6] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, “Non-orthogonal multiple access (NOMA) for future radio access,” in *Proc. IEEE TVC spring 2013*, Jun. 2013.

[7] Z. Ding, P. Fan and H. V. Poor, “Impact of user pairing on 5G nonorthogonal multiple access,” *arXiv preprint*, arXiv:1412.2799, 2014.

[8] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, “System-level performance evaluation of downlink non-orthogonal multiple access (NOMA),” in *Proc. IEEE PIMRC 2013*, pp. 611-515, Sep. 2013.

[9] Y. Xu, H. Sun, R. Q. Hu, and Y. Qian, “Cooperative Nonorthogonal Multiple Access in Heterogeneous Networks,” in *Proc. IEEE GlOBECOMM 2015*, San Diego, Dec. 2015

[10] M. Costa, “Writing on dirty paper,” *IEEE Trans. Info. Theory*, vol. 29, no. 3, pp. 439-441, May 1983.

[11] 3GPP TR36.814,“ Further Advancements for E-UTRA Physical Layer Aspects,” v9.0.0, Mar. 2010.

[12] Y. R. Zheng, and C. Xiao, “Simulation models with correct statistical properties for Rayleigh fading channels,” *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920-928, Jun. 2003.