Users Association in Small Cell Networks with Massive MIMO

Radwa Sultan[†], Lingyang Song[*], Karim G. Seddik [‡], and Zhu Han[†]
[†] Electrical and Computer Engineering Department, University of Houston, TX, USA
[*] School of Electrical Engineering and Computer Science, Peking University, Beijing, China
[‡] Electronics and Communications Engineering Department, American University in Cairo, Egypt

***Abstract*—In this paper, an investigation of the effect of deploying massive MIMO in two-tier cellular networks is presented. In our model, the macrocell base station (BS) as well as femtocell access points (FAPs) are equipped with a very large number of antennas. Each mobile user will attempt to connect to the BS or FAPs. However, it is assumed that the users may be biased to connect to FAPs rather than the BS. A resource allocation problem is formulated to find the optimal bias that maximizes the total system capacity while keeping the transmitted power from the BS and the FAPs within a certain limit. An algorithm for solving the optimization problem is proposed and numerical results are presented to illustrate how deploying massive MIMO can affect the optimal bias value and the total capacity. In addition, the performance of our proposed scheme is evaluated when using different precoding schemes.**

I. INTRODUCTION

Increasing modern communication networks' capacity has become a continuous demand. Accordingly, many new methodologies have been recently proposed. Two of these solutions are the very large scale multiple-input multiple-output or "massive MIMO" which is proposed as a solution that is capable of enhancing the communication link efficiency, and the small/femto cell networks (SCNs) which are proposed to increase the network capacity. Consequently, utilizing both massive MIMO and SCNs is expected to enhance the communication network's performance. Massive MIMO [1], [2], [3] has been proposed as a technique that is able to achieve network densification by largely increasing the number of active antennas. It is particularly fascinated that, with an infinite number of antennas, the simplest form of user detection and beamforming become optimal.

Massive MIMO challenges have been studied in recent literature. In [4], the effect of a finite number of antennas on the sum rate of the time-shifted method is analyzed. In [5], a new algorithm which is based on outer multi-cellular precoding is used to eliminate inter-cell interference in TDD large scale antenna systems. In [6], an energy efficient power allocation scheme has been proposed for maximum ratio combining massive MIMO system.

On the other track, SCNs have also been proposed as a solution of high data rate demands by reducing the size of the network's cells. In [7], a hierarchical resource allocation scheme for the downlink of a large-scale small-cell network is proposed. In [8], [9], different resource allocation problems in full-duplex heterogeneous network are proposed.

Actually, jointly utilizing massive MIMO and SCNs can further improve the network's QoS. In [10], a comparison between massive MIMO and SCN is done. In [11], the potential benefits of incorporating a massive MIMO BS in heterogeneous network are investigated. In [12], a TDD two-tier network architecture is presented. In this framework, the BS with massive MIMO designs its precoding vectors such that orthogonality to the subspace spanned by the strongest interference directions is achieved.

However, the effect of deploying massive MIMO on resource allocation problems in SCNs has not been fully investigated in literature. In this paper, we study the effect of deploying massive MIMO on users' association in SCNs. Accordingly, we consider the downlink transmission of a two-tier network in which both the BS and the FAPs are equipped with massive MIMO. There is no definition of macro-users and femto-users, since each user can be served by the BS or one of the FAPs. However, it must be noticed that the users are biased to connect to the femto-tier rather than the macro-tier. Based on these assumptions, a resource allocation problem will be presented whose objective is to find the optimal biasing value that maximizes the network's capacity while keeping the power constraints of the BS and the FAPs satisfied. Our contribution in this work can be stated as follows:

1) Studying the effects of deploying massive MIMO on users' association to both macrocell and femtocell networks.
2) Proposing an optimization problem that determines the optimum biasing value that maximizes the network's sum capacity.
3) Presenting numerical results evaluating the system performance when using different precoding schemes.

The remainder of this paper is organised as follows. In Section II, the system model is presented. In Section III, the proposed optimization problem is formulated. In Section IV,

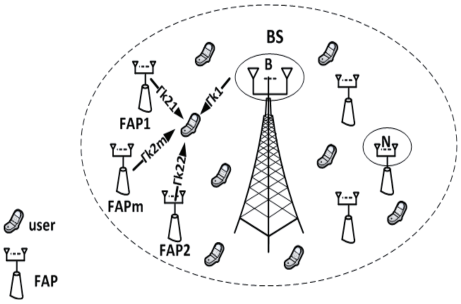

Fig. 1. System Model.

the numerical results is presented. Finally, in Section V, the paper is concluded.

II. SYSTEM MODEL

We consider a single channel a single cell network that includes a single BS with B antennas and M FAPs each with N antennas. The cell has K users that are randomly located in the cell. It must be noticed that, in order to employ massive MIMO, we must have $B \gg K$, and $N \gg K$ [1]. The system model is shown in Fig. 1.

In our model, a downlink transmission is considered in which users may be served by BS or any FAP. Therefore, in order to guarantee the best QoS for each user, each user needs to be assigned to the best serving node. Therefore, each user compares the received powers from all nodes and chooses the node with the highest power. In addition, it is assumed that users may be biased to be served by the femto-tier rather than the macro-tier. In other words, the users may choose to connect to one of the FAPs even if the actual received power from the chosen FAP is less than that received from the BS.[2] Thus, users assignments will depend on the actual received signal-to-interference-noise ratio (SINR) at the BS and the biased received SINR from the FAPs. Therefore, in our proposed model, the actual received SINR at the k^{th} user from the BS is given by,

$$\Gamma_{k_1} = \frac{D_{1_k}^{-\alpha} |\mathbf{h}_{k,1}\mathbf{w}_{k,1}|^2}{\sigma^2 + \sum\limits_{\substack{i \in MU \\ i \neq k}} D_{1_k}^{-\alpha} |\mathbf{h}_{k,1}\mathbf{w}_{i,1}|^2 + \sum\limits_{j \in FU} \sum\limits_{m=1}^{M} D_{2_{mk}}^{-\alpha} |\mathbf{h}_{mk,2}\mathbf{w}_{mj,2}|^2}, \tag{1}$$

where $D_{1_k}^{-\alpha}$ represents the large scale propagation loss between the BS and the k^{th} user with path loss exponent α and distance D_{1_k}, $\mathbf{w}_{k,1} \in \mathbb{C}^{B \times 1}$ represents the BS precoding vector for the k^{th} user data. However, if the k^{th} user is not connected to the BS, the value of $\mathbf{w}_{k,1}$ will be zero. $\mathbf{h}_{k,1} \in \mathbb{C}^{1 \times B}$ is the channel vector between the k^{th} user and the BS, all the channel coefficients are assumed to be an i.i.d. zero mean complex Gaussian random variables with unit variance, i.e, Rayleigh fading, σ^2 represents the additive white Gaussian noise (AWGN) variance, the first summation in the denominator denotes the interference power introduced by other transmissions to users connected to the BS, where MU denotes the set of users that are connected to the BS. Similarly, the second summation in the denominator represents the interference power introduced by other transmissions to users connected to the FAPs, where FU denotes the set of users that are connected to the FAPs, $\mathbf{w}_{mj,2} \in \mathbb{C}^{N \times 1}$ denotes the m^{th} FAP precoding vector for the j^{th} user data, $\mathbf{h}_{mk,2} \in \mathbb{C}^{1 \times N}$ denotes the channel vector between the k^{th} user and the m^{th} FAP.

On the other hand, the actual received SINR at the k^{th} user from the m^{th} FAP is given by,

$$\Gamma_{k_{2m}} = \frac{D_{2_{mk}}^{-\alpha} |\mathbf{h}_{mk,2}\mathbf{w}_{mk,2}|^2}{\sigma^2 + \sum\limits_{i \in MU} D_{1_k}^{-\alpha} |\mathbf{h}_{k,1}\mathbf{w}_{i,1}|^2 + \sum\limits_{\substack{j \in FU \\ j \neq k}} \sum\limits_{m=1}^{M} D_{2_{mk}}^{-\alpha} |\mathbf{h}_{mk,2}\mathbf{w}_{mj,2}|^2}. \tag{2}$$

Furthermore, according to [14], the biased received SINR is a scaled version from the actual received SINR. Accordingly, the biased received SINR at the k^{th} user from the m^{th} FAP is given by,

$$\Gamma_{k_{2m}}^{*} = \beta \times \Gamma_{k_{2m}}, \tag{3}$$

where β represents the biasing factor of the users to connect to the FAPs. It must be mentioned that from equations (1) and (3), two important points needs to be mentioned,

- In both (1) and (2), the effect of small scale fading in the received signals and the interference signals will be averaged to its mean value which is equal to one. This is actually one of the key features of the deploying massive MIMO, in which deploying a very large number of antennas will cause channel hardening [15], which offers high immunity to Rayleigh fading. From experiments, it is shown that channel hardening assumption is valid, if the number of transmitting antennas is in the order of 20 antennas.
- In (3), the value of β determines whether the users are biased to connect to the BS or the FAPs. If the value of $\beta > 1$, the biased SINR will be larger than the actual SINR. Hence, the user will be more attracted to connect to FAPs. This will result in virtually expanding the coverage area of femtocells. However, if $0 < \beta < 1$, then the biased SINR will be less than the actual SINR. Therefore, the user will be less attracted to FAPs. Accordingly, Both MU and FU sets' members depend on the value of β.

---

[1] It must be noticed that deploying Massive MIMO in FAPs will have many implementation challenges like space and power limitations. However, some recent work [13] shows that, it is possible to implement massive MIMO even in relatively small base stations within a reasonable form factor. In this paper, we show the results of different numbers of antennas in FAPs so as to show the effects.

[2] Generally speaking, biasing is considered as an assigning algorithm that helps distribute the network's traffic between the BS and the FAPs. In this assigning algorithm, the user compares between the true received power from the BS and biased power from the FAPs, and picks the node with the highest received power.

Algorithm 1: Finding Optimum Biasing Value

```
Data: $D_{1k}$ , $D_{2_{mk}}$ $\forall\, m,k$ , $P_{max1}$ , $P_{max2}$ , $\beta_s$ , all
      CSI infromation, $\varepsilon$.
Result: Find $\beta$ maximizing $C_T(\beta)$

Intially: $n = 1$, ;
1. Find $\chi = \frac{\delta C_T}{\delta \beta}$ ;
2. Find $a_n \in \beta_s$ such that $\chi(a_n) > 0$;
3. Find $b_n \in \beta_s$ such that $\chi(b_n) < 0$;
4. Compute $p_n = \frac{a_n + b_n}{2}$;
5. if $\chi(p_n) = 0$ or $b_n - a_n \le \varepsilon$ then
    $\beta = p_n$
    end
else
    if $\chi(p_n) > 0$ then
        $a_{n+1} = p_n$, $b_{n+1} = b_n$, $n = n + 1$
        Return to step 4.
    else
        $b_{n+1} = p_n$, $a_{n+1} = a_n$, $n = n + 1$
        Return to step 4.
    end
end
```

Since, the value of β controls the number of users in both MU and FU, then the macrocell and the femtocell channel capacities will be dependent on the value of β. Accordingly, the total network capacity per unit bandwidth is given by,

$$C_T(\beta) = W_1\Big(\sum_{k\in MU} \log_2(1+\Gamma_{k_1})\Big) + W_2\Big(\sum_{k\in FU} \log_2(1+\Gamma_{k_{2m}})\Big), \tag{4}$$

where, the first summation calculates the macrocell capacity, the second summation computes the femtocell capacity, W_1, and W_2 are arbitrarily weight factors to balance the offloading of the BS.

III. PROPOSED SCHEME

In this section, we aim at finding the optimal value of β that maximizes C_T, while keeping the transmission power constraints imposed on the BS and the FAPs. Therefore, the optimization problem is given by

$$\begin{aligned} \max_{\beta} \quad & C_T(\beta) \\ \text{subject to} \quad & \sum_{k=1}^{K} |\mathbf{w}_{k,1}|^2 \le P_{max_1}, \\ & \sum_{k=1}^{K} |\mathbf{w}_{mk,2}|^2 \le P_{max_2}, \text{for } m \in [1, M], \end{aligned} \tag{5}$$

where P_{max_1} and P_{max_2} denote the maximum power consumption for the BS and each FAP, respectively. The summations $\sum_{k=1}^{K} |\mathbf{w}_{k,1}|^2$ and $\sum_{k=1}^{K} |\mathbf{w}_{mk,2}|^2$ represent the total power transmitted by the BS and the FAPs, respectively [3]. It must be noticed that in the proposed scheme, the main purpose of biasing is to maximize the total system capacity, and not to offload the users from the macro-tier to the femto-tier.

In order to solve the proposed optimization problem, the bisection search technique is adopted [2]. In the beginning, a set of biasing values β_s, within which the optimal β will be found, is defined. Additionally, a tolerance ε is defined, and the derivative χ of the C_T is calculated with respect to β, then the algorithm tries to find a point $a_n \in \beta_s$ at which χ is positive. Then, the algorithm tries to find a point $b_n \in \beta_s$ at which χ is negative. Subsequently, the average p_n of points a_n and b_n is calculated. Then, χ is calculated at the point p_n. If the value $\chi(p_n)$ is zero, or the difference between the a_n and b_n coefficient is smaller than ε, then the algorithm sets the optimal β to be equal p_n. However, if the value of the derivative $\chi(p_n)$ is positive, then the value of a_n is updated to be p_n. Nevertheless, if the value of $\chi(p_n)$ is negative, then the value of b_n is updated to be p_n. Then, the algorithm repeats the previous procedures till finding the optimal β. The proposed solution is discussed in Algorithm 1. It is worth mentioning that, for the bisection search method, for tolerance value of ε, the number of iterations needed for convergence is $\log_2 \frac{b_1 - a_1}{\varepsilon}$, where a_1, and b_1 are the values of a_n, and b_n, respectively, for $n = 1$.

IV. NUMERICAL RESULTS

In this section, we are going to study the performance of our proposed model. In the beginning, we are going to study the variation of C_T with β. Afterwards, the effects of varying the number of antennas of the BS and the FAPs on the optimal β and the C_T are studied. We consider a square grid with one BS in the center and four FAPs. In addition, we study the performance of both Eigen beamforming (EB) and regularized zero-forcing (ZF) precoding [16]. The EB and the ZF precoding matrices are given, respectively, by

$$\mathbf{W_{EB}} = \mathbf{H}^{\dagger}, \tag{6}$$

$$\mathbf{W_{ZF}} = (\mathbf{H}^{\dagger}\mathbf{H} + \mathbf{Z} + n\varphi I_n)^{-1}\mathbf{H}^{\dagger}, \tag{7}$$

where $\mathbf{H}$ is the channel matrix between the transmitting node and users served by that node, $(.)^{\dagger}$ denotes the Hermitian transpose, n is the number of transmitting antennas, $\mathbf{Z} \in \mathbb{C}^{n\times n}$ is an arbitrary Hermitian non-negative definite matrix, and $\varphi > 0$ is an arbitrary regularization parameter. It must be

---

[3] In this work, we mainly focus on studying the effect of deploying massive MIMO β, and consequently on users' association. However, it must be pointed out that, it is possible to maximize the total network capacity jointly over β and the precoding vectors.

[2] We can use more sophisticated nonlinear optimization solutions such as simulated annealing and interior point methods. However, it will be shown later from the simulation results that due to the shapes of the curves showing the relation between the total achieved channel capacity and β, the bisection search is sufficient enough.

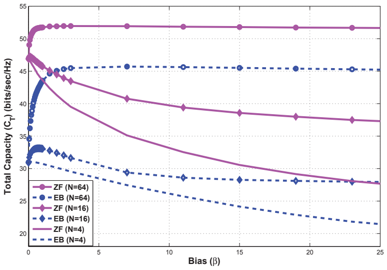

Fig. 2. Variation of Total Capacity with Biasing. { $B = 64$, $K = 10$, $P_{max_1} = 3$ W, $P_{max_2} = 0.9$W. }

noticed that the users precoding vectors are the rows of the precoding matrices in (6), and (7). It must be mentioned that introducing massive MIMO along with channel precoding will make the received power approximately grows linearly with the number of installed antennas, given that the antenna array is scaled with increasing the number of antennas [16]. In addition, the main advantage of ZF over EB is its ability to decrease the interference on other users by transmitting in their null space. consequently, ZF becomes more robust as the number of installed antennas at the transmitting nodes increases. The simulation parameters used are $\alpha = 2.7$, $\mathbf{Z} = 0$, $\varphi = 1$, and $W_1 = W_2 = 1$.

A. The Variation of Total Capacity with Biasing.

Fig. 2 shows the effect of changing β on C_T in case of installing both limited and relatively large N. In the beginning, for small values of β, increasing β increases the value of C_T, which is very intuitive, because if we consider the case of $\beta = 0$, all users will be connected to the BS even if the FAPs can offer better transmission chances. Therefore, a drop in the total capacity occurs. However, by increasing the value of β, users will begin offloading from the BS to the FAPs. Therefore, better transmission chances are exploited. Accordingly, an increase in C_T occurs. However, further increase in β will make all users tend to connect to the FAPs which results in missing better transmission opportunities offered by the BS. Therefore, C_T will decrease. From the results obtained in Fig. 2, two important points must be mentioned. First, introducing massive MIMO offered a noticeable performance gain for both EB and ZF. It can be noticed that the gain achieved from increasing N from 16 to 64 is much higher than that achieved from increasing N from 4 to 16. Second, the relation curve's shape shown in Fig. 2 reassures that there is a global optimum value for β that maximizes the total capacity. Therefore, using the bisection search is sufficiently enough for our proposed problem.

Furthermore, it can be noticed that it is more beneficial to utilize ZF rather that EB. This superiority is due to the fact that the ZF decreases the interference level introduced on the user's transmission, hence the received SINR, and hence the channel capacity, increase.

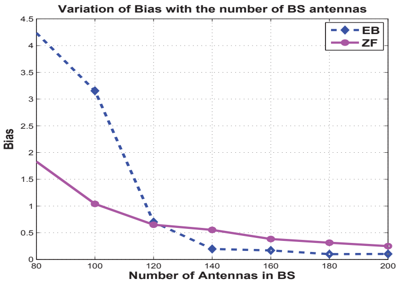

Fig. 3. Variation of Optimal Biasing Value with Number of Antennas at the BS. {$N = 60$, $K = 20$, $P_{max_1} = 7$ W, $P_{max_2} = 0.9$W.}

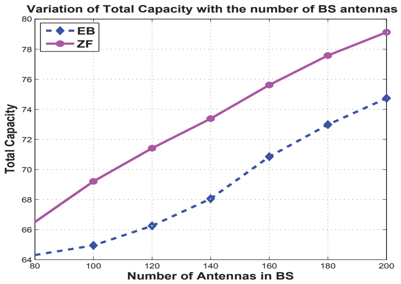

Fig. 4. Variation of The Maximum Total Capacity with Number of Antennas at the BS. { $N = 60$, $K = 20$, $P_{max_1} = 7$ W, $P_{max_2} = 0.9$W.}

B. The variation of The Optimal Biasing Value and System Total Capacity with Number of Antennas in the BS

Fig. 3 and Fig. 4 show the effect of increasing the number of BS antennas B on the optimal β and C_T, respectively. Actually, increasing B is expected to decrease the value of the optimal β. This is explained by the fact that increasing B will increase the antenna array gain offered by the BS. As a result, the BS will offer better transmission chance for the users. Therefore, the value of the optimal β will decrease to make users connect to the BS. This behavior can be verified from Fig. 3, for a relatively small B, the value of the optimal β is high. However, by increasing the B, the value of the optimal β decreases to achieve the highest possible system capacity. The success of the proposed model to achieve the maximum capacity can be verified from Fig. 4 as it can be noticed that increasing B will increase C_T for both precoding algorithms.

C. The variation of The Optimal Biasing Value and System Total Capacity with Number of Antennas in FAPs.

Fig. 5 and Fig. 6 show the effect of increasing the number of FAPs antennas N on the optimal β and C_T, respectively. Actually, increasing N is expected to increase the optimal β. This is due to the fact that increasing N will increase the antenna array gain offered by the FAPs. As a result, the FAPs will offer better transmission chance for the users. Therefore, the optimal β will increase to make

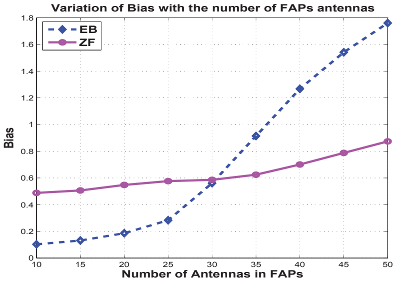

Fig. 5. Variation of Optimal Biasing Value with Number of Antennas at the FAPs.$\{$ $B = 80$, $K = 20$, $P_{max_1} = 7$ W, $P_{max_2} = 0.9$W.$\}$

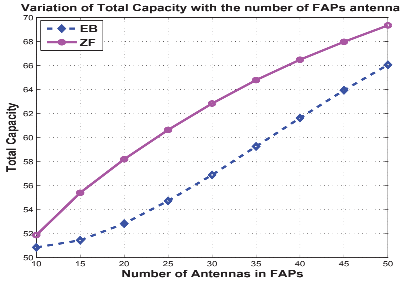

Fig. 6. Variation of The Maximum Total Capacity with Number of Antennas at the FAPs.$\{B = 80$, $K = 20$, $P_{max_1} = 7$ W, $P_{max_2} = 0.9$W.$\}$

users connected to the FAPs. This behavior can be verified from Fig. 5, for a relatively small N, the optimal β is low. However, by increasing tN, the value of the optimal β starts to increase to achieve the highest possible system capacity. The success of the proposed model to achieve the maximum system capacity can be verified from Fig. 6 as it can be easily noticed that increasing N will increase C_T for both precoding algorithms.

V. CONCLUSION

In this paper, a two-tier network with a single BS and multiple FAPs is presented. Both the BS and the FAPs can be equipped with massive MIMO. Furthermore, a resource allocation problem is formulated to find the optimal biasing that maximizes the system capacity while satisfying the power constraints of the transmitting nodes. In addition, a solution algorithm using the bisection search method is presented. In order to evaluate our system performance, both EB and ZF are considered. Moreover, the effects of varying both the number of antennas employed in the BS and the FAPs on the optimal biasing value and the system capacity are studied. Finally, the effects of deploying massive MIMO on users' association are summarized as follows:

1) If massive MIMO is deployed in the FAPs, users will tend to associate to the femto-tier, hence the optimal biasing value will be large.
2) If massive MIMO is deployed in the BS, users will tend to associate to the macro-tier, hence the optimal biasing value will be small.
3) If massive MIMO is deployed in both BS and FAPs, the optimal biasing value will depend mainly on the imposed power constraints on the transmitting node.

ACKNOWLEDGMENT

Thanks to U.S. NSF CCF-1456921, CNS-1443917, ECCS-1405121, ECCS 1547201, and NSFC 61428101.

REFERENCES

[1] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T.L.Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, pp. 40–60, January 2013.

[2] T. L. Marzetta, G. Caire, M. Debbah, I. Chih-Lin, and S. K. Mohammed, "Special issue on massive mimo," *Journal of Communications and Networks*, vol. 15, pp. 333 – 337, August 2013.

[3] J. Hoydis, K. Hosseini, S. ten Brink, and M. Debbah, "Making smart use of excess antennas: Massive mimo, small cells, and tdd," *Bell Labs Technical Journal*, vol. 18, pp. 5–21, August 2013.

[4] W. Mahyiddin, P. Martin, and P. Smith, "Pilot contamination reduction using time-shifted pilots in finite massive mimo systems," in *IEEE 80th Vehicular Technology Conference (VTC Fall)*, Vancouver, Canada, September 2014.

[5] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, Cambridge, MA, July 2012, pp. 1137 – 1141.

[6] L. Zhao, H. Zhao, F. Hu, K. Zheng, and J. Zhang, "Energy efficient power allocation algorithm for downlink massive mimo with mrt precoding," in *IEEE 78th Vehicular Technology Conference (VTC Fall)*, Las Vegas, NV, September 2013, pp. 1–5.

[7] S. Sadr and R. Adve, "Partially-distributed resource allocation in small-cell networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 12, pp. 6851 – 6862, June 2014.

[8] R. Sultan, L. Song, and Z. Han, "Impact of full duplex on resource allocation for small cell networks," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta, GA, December 2014.

[9] R. Sultan, L. Song, K. G. Seddik, Y. Li, and Z. Han, "Mode selection, user pairing, subcarrier allocation and power control in full-duplex ofdma hetnets," in *IEEE International Conference on Communication Workshop (ICCW)*, London, June 2015.

[10] W. Liu, S. Han, and C. Yang, "Massive mimo or small cell network: Who is more energy efficient?" in *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Shanghai, China, April 2013, pp. 24 – 29.

[11] M. Kountouris and N. Pappas, "Hetnets and massive mimo: Modeling, potential gains, and performance analysis," in *IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC)*, Torino, Italy, September 2013, pp. 1319 – 1322.

[12] K. Hosseini, J. Hoydis, S. ten Brink, and M. Debbah, "Massive mimo and small cells: How to densify heterogeneous networks," in *IEEE International Conference on Communications (ICC)*, Budapest, Hungary, June 2013, pp. 5442 – 5447.

[13] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *Access, IEEE*, vol. 1, pp. 335–349, May 2013.

[14] J. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, "An overview of load balancing in hetnets: old myths and open problems," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 18 – 25, April 2014.

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

[16] J. Hoydis, S. ten Brink, and M. Debbah, "Massive mimo in the ul/dl of cellular networks: How many antennas do we need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 160 – 171, February 2013.