Interference Alignment in Two-Tier Randomly Distributed Heterogeneous Wireless Networks Using Stochastic Geometry Approach

Y. Luo, *Student Member, IEEE*, T. Ratnarajah, *Senior Member, IEEE*, J. Xue *Member, IEEE* and F. A. Khan, *Member, IEEE*

***Abstract*—With the massive increase in wireless data traffic in recent years, multi-tier wireless networks have been deployed to provide much higher capacities and coverage. However, heterogeneity of wireless networks bring new challenges for interference analysis and coordination due to spatial randomly distributed transmitters. In this paper, we present a distance-dependent interference alignment (IA) approach for a generic 2-tier heterogeneous wireless network, where transmitters in the first and second tiers are distributed as Poisson Point Process (PPP) and Poisson Cluster Process (PCP) respectively. The feasibility condition of the IA approach is used to find upper bound of the number of interference streams that can be aligned. The proposed IA scheme maximizes the second-tier throughput by using the trade-off between signal-to-interference ratio and multiplexing gain. It is shown that acquiring accurate knowledge of the distance between the receiver in the second-tier and the nearest cross-tier transmitter only brings insignificant throughput gain compared to statistical knowledge of distance. Furthermore, the remaining cross-tier and inter-cluster interferences are modeled and analyzed using stochastic geometry technique. Numerical results validate the derived expressions of success probabilities and throughput, and show that the distance-dependent IA scheme significantly outperforms the traditional IA scheme in the presence of path-loss effect.**

***Index Terms*—Heterogeneous network, interference alignment, interference analysis, stochastic geometry, throughput.**

I. Introduction

The explosive growth of wireless services and applications in recent years has motivated strong interest to find the throughput capacity of large-scale random wireless networks. Starting from the pioneering work of [1], the throughput capacity of random wireless networks has attracted significant attention in the last decade and many key results have been obtained [2], [3]. However, all the above works only considered homogeneous wireless networks where the nodes are assumed to be identical. On the other hand, in practice, the ever-growing demand for wireless connectivity has spurred network densification by deploying additional access nodes which are overlaid on the existing homogeneous network infrastructure and differ in their size and power characteristics, leading to heterogeneous networks (HetNets).

Inter-cell interference cancellation is one of the key factors to achieve the desired performance in the HetNets. Most of the existing works on interference cancellation in HetNets focus only on the fading effect of the channel whereas path-loss effect is largely ignored. However, with the increasing network densification, the effect of path-loss can no longer be ignored. Stochastic geometry has recently been used as a powerful tool to model and analyze randomly distributed wireless networks [4]–[6]. Using fundamental tools from stochastic geometry, a mathematical model was introduced in [7] for coexistence in networks composed of both narrowband (NB) and ultrawideband (UWB) wireless nodes. The theory and techniques of cooperation for multi-cell multiple-input-multiple-output (MIMO) wireless networks was discussed in [8] which shows that multi-cell cooperation can dramatically improve the system performance. In [9], the K-tier downlink heterogeneous cellular network was modeled and analyzed and it shows that the probability of coverage or outage would not be affected when adding more nodes in the network for interference limited open access networks when all the tiers had same target Signal-to-Interference-plus-Noise-Ratio (SINR). The author provided a tractable model to analyze the effects of offloading in a K-tier wireless HetNets in [10], where the distribution of rate over the entire network was then derived for a weighted association strategy. Furthermore, a load aware model was introduced and studied for carrier aggregation enabled multi-band HetNets in [11]. Ordering results were derived for coverage probability and per user rate to compare different transmission techniques, such as space division multiple access (SDMA), single user beamforming (SUBF), and baseline single-input single-output (SISO) transmission of downlink multi-antenna heterogeneous cellular HetNets [12].

Meanwhile, two-tier heterogeneous wireless networks have drawn significant attention in the recent years, especially for single-antenna cellular networks. In [13], a heterogeneous cellular network was considered wherein the macro base stations (BSs) followed PPP while the femto BSs followed

Y. Luo was with the School of Engineering, Institute for Digital Communications, The University of Edinburgh, Edinburgh, EH9 3JL, U.K. He is now with the with school of electrical engineering, Tsinghua University, China.

T. Ratnarajah and J. Xue are with the School of Engineering, Institute for Digital Communications, The University of Edinburgh, Edinburgh, EH9 3JL, U.K.

F. A. Khan was with the School of Engineering, Institute for Digital Communications, The University of Edinburgh, Edinburgh, EH9 3JL, U.K. He is now with the with School of Computing and Engineering, University of Huddersfield.

The work was supported by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission under FET-Open grant number HIATUS-265578. This work was also supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/L025299/1.

Dr. J. Xue is the correspondence author.

PPP or PCP. The trade-off between the performance of the subscriber and non-subscriber femto BSs' users was studied based on multi-carrier techniques. In [14], a heterogeneous cellular network was considered with different tiers, namely the macro BS tier and the femto BS tier which followed PPP, and another tier of the femto BSs' users following Neyman-Scott process. The optimal spectrum allocation policy was derived and aggregated network throughput was maximized with constraint on the quality of service (QoS). However, most of the analysis is restricted to simple random arrangement of nodes such as PPP and single-antenna setting, for reasons of tractability. This motivated us to apply more generalized and realistic stochastic geometry tools for analyzing and improving the QoS by using interference alignment (IA) technique.

IA is a promising approach to increase the degrees-of-freedom (DoF) of HetNets by aligning the interference from different transmitters into the same lower-dimensional interference subspace at each receiver [15]–[18]. IA concept originated from an elegant coding scheme for 2-user MIMO X channel [15]. In [19], the full potential of IA was realized by achieving half of the total DoF at each receiver in K-user interference channel. [16] and [17] studied the performance of wireless networks with Poisson clustered process (PCP) based on the probability generating function (PGF) where perfect intra-cluster channel state information (CSI) was available and the intra-cluster interference was completely eliminated by using the IA approach. Moreover, the case of imperfect CSI with bounded errors was considered in [18], and a capacity lower bound of the channel by using IA was derived. These traditional methods to implement IA may not be optimal to achieve the highest spectrum efficiency in the presence of the path-loss effect. In terms of the proposed two-layer distribution, tier-A transmitters can be thought of as macro BSs, tier-B transmitters can be considered as femto BSs and tier-B receivers represent the user terminals.

In this paper, we consider a generic randomly distributed 2-tier multi-antenna heterogeneous wireless network. Transmitters in tier-A follow a PPP with large coverage range, while transmitters in tier-B are Poisson clustered distributed with much smaller coverage range [1]. Four different kinds of interferences are considered at each tier-B receiver: the nearest cross-tier interference, the remaining cross-tier interference, the inter-cluster interference and the intra-cluster interference. The main contributions of the paper are as follows.

- The intra-cluster interference at each tier-B receiver is always assumed to be completely eliminated using IA technique, and the throughput of tier-B receivers in each cluster is derived and maximized.
- In order to derive the throughput, the nearest cross-tier interference is mitigated at the tier-B receivers by using a distance-dependent IA approach. Specifically, transmitters in tier-B partially align their intra-cluster interference to the nearest cross-tier interference, which is overheard by the receivers in the same cluster, in order to increase the multiplexing gain of their own data links.
- The nearest cross-tier interference is modeled as shot noise when the distance from the nearest cross-tier interferer to the tier-B receiver is large enough. Otherwise, IA approach is used to align this interference. With the distance information, the throughput can be maximized by the trade-off between SIR at each data stream and the number of data streams. The gain of the accurate distance information compared to the statistical distance information is then studied and validated by the simulation results.
- Apart from the nearest cross-tier interference and the intra-cluster interference, stochastic geometry based approach is utilized to model the remaining interference, which consists of the cross-tier and the inter-cluster interferences.
- The inter-cluster interference will affect the outage probability independently after considering the expectation of remaining cross-tier interference power. Moreover, outage probability affected by the inter-cluster interference is derived as well as its closed-form approximation. All of the derived expressions of the success probabilities are finally validated by Monte Carlo simulations.

The remainder of this paper is organized as follows. Section II introduces the system model, including the distribution of the transmitters, CSI availability, interference categorization and performance metrics. Section III presents the implementation of the distance-dependent IA approach to mitigate the nearest cross-tier interference. The success probabilities affected by the remaining cross-tier interference and the inter-cluster interference are investigated in Section IV which is followed by the simulation results in Section V, where we have presented all numerical results. Finally, conclusions are presented in Section VI.

II. System Model

We consider *two* tiers of transmitters, namely tier-A transmitters and tier-B transmitters, on an infinite plane. The tier-A transmitter set, denoted as $\Phi_A = \{N_i | i \in \mathbb{N}\}$, follows a PPP with intensity λ_A and the tier-B transmitters, denoted as $\Phi_B = \{T_i | i \in \mathbb{N}\}$, follows a Matérn process on the plane with the intensity of the parent process being λ_B. In each cluster of Φ_B, there are K transmitters uniformly distributed in a ball centred at the parent point with radius r_a. Meanwhile, there are K intended receivers in each cluster in tier-B, denoted as R_i, at distance l_{ii} from each transmitter, where l_{ii} is constant. In order to simplify the following descriptions, we use the term *tier-B receivers* to denote the intended receivers of the related tier-B transmitters. The linear density function of the transmitters in each tier-B cluster is given by,

$$f_B(x) = \begin{cases} \frac{2\|x\|}{r_a^2}, & \|x\| \le r_a \\ 0, & \|x\| > r_a \end{cases}, \tag{1}$$

[1] In realistic scenario, tier-A transmitters can be thought of as macro BSs which can be approached as homogeneous distributed, while tier-B transmitters can be considered as femto BSs and tier-B receivers represent the user terminals. Roughly speaking, a set of points are in clusters if they lack spatial homogeneity, i.e., one observes points forming groups which are well spaced out. Because of this, it is more sensible to assume that macro BSs follow PPP as in [20] and femto BSs follow PCP because more services are required in crowded places such as parks, shopping malls and stations.

where we assume r_a to be small enough so that all the transmitters and receivers in a cluster have the same nearest cross-tier transmitter.

It is worth to notice that various multi-tier wireless network structures can be represented using our system model, such as Mobile Ad-Hoc Networks (MANETs) and the Long-Term Evolution-Advanced (LTE-A) cellular network. In the case of LTE-A cellular networks, which consist of macro BS, pico BSs, femto BSs and relays, each layer of devices has unique transmit power constraint, distribution, number of antennas, etc. Macro BSs typically transmit with a high power level (5W - 40W) and cover an area with a radius of up to $2\mathrm{km}$, while femto BSs typically have a low transmit power level ($100\mathrm{mW}$ - 2W) and cover an area of up to tens of meters [21]. Therefore, in the presence of path-loss effect, unmitigated interference from tier-A transmitters, especially the nearest one, and the intra-cluster interference may reduce the system performance of tier-B.

Each transmitter in tier-A is assumed to transmit d_A data streams with N_A antennas and each transmitter in tier-B is assumed to transmit d_B data streams with N_B antennas. Tier-B receivers are also equipped with N_B antennas. We consider a receiver R_0 which is in the Voronoi cell of N_0 and its intended transmitter is in ϕ_0, where ϕ_i is defined as the i-th cluster in tier-B. The signal received at R_0 is given as

$$\mathbf{y}_0 = l_{00}^{-\alpha/2}\mathbf{H}_{00}\mathbf{V}_0\mathbf{s}_0 + \underbrace{\sum_{N_i\in\Phi_A} l_{0N_i}^{-\alpha/2}\mathbf{H}_{0N_i}\mathbf{V}_{N_i}\mathbf{s}_{N_i}}_{\text{Interference from tier-}A} + \underbrace{\sum_{T_i\in\phi_0} l_{0i}^{-\alpha/2}\mathbf{H}_{0i}\mathbf{V}_i\mathbf{s}_i}_{\text{Intra-cluster interference}} + \underbrace{\sum_{T_i\in\Phi_B/\phi_0} l_{0i}^{-\alpha/2}\mathbf{H}_{0i}\mathbf{V}_i\mathbf{s}_i}_{\text{Inter-cluster interference}} + \mathbf{n}_0,$$

where l_{iN_j} is the distance between tier-A transmitter to R_i and l_{ij} represents the distance between R_i and the tier-B nodes, T_j. Symbols $\mathbf{H}_{iN_j} \in \mathbb{C}^{N_B\times N_A}$ and $\mathbf{H}_{ij} \in \mathbb{C}^{N_B\times N_B}$ stand for the channel matrices from transmitters in tier-A and tier-B to R_i, respectively. We assume that all entries of channel matrices are independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$. Beamforming matrices $\mathbf{V}_{N_i} \in \mathbb{C}^{N_A\times d_A}$ and $\mathbf{V}_i \in \mathbb{C}^{N_B\times d_B}$ are designed by nodes in tier-A and tier-B, respectively. Symbol vectors $\mathbf{s}_{N_j} \in \mathbb{C}^{d_A\times 1}$ and $\mathbf{s}_j \in \mathbb{C}^{d_B\times 1}$ are transmitted by nodes in tier-A and tier-B, respectively. We assume that all symbols in each symbol vector $\mathbf{s}_i = [s_i(1)\ s_i(2)\ \dots]^T$, where $i = 1, 2, \dots$ or $N_1, N_2, \dots$ share the same power allocation, i.e., $\rho_A \triangleq P_A/d_A$ and $\rho_B \triangleq P_B/d_B$ are for tier-A transmitters and tier-B transmitters. α denotes the path-loss exponent and vector $\mathbf{n}_0 \in \mathbb{C}^{N_B\times 1}$ denotes the additive white noise.

It is reasonable to assume that the transmitter has access to high quality CSI when the receiver is nearby. Hence, we assume that perfect CSI is shared between transmitter-receiver pairs in a tier-B cluster due to small size of the cluster. As the interference from the nearest transmitter is dominant [22] and each BS in tier-A only serves receivers in its Voronoi cell [13], we assume the tier-B transmitters and receivers only have access to CSI of the nearest tier-A transmitter. Moreover, as each tier-A transmitter covers a large area, it can be reasonably assumed that the CSI of the nearest tier-A transmitter at tier-B transmitters and receivers is subject to estimation error. The relationship between the real channel and the corresponding estimate at R_0 can be written as

$$\mathbf{H}_{iN_0} = \sqrt{1-\varepsilon^2}\hat{\mathbf{H}}_{iN_0} + \varepsilon\tilde{\mathbf{H}}_{iN_0}, \tag{2}$$

where $\varepsilon \in [0,1]$ is the standard error of the channel estimation error. Symbol $\hat{\mathbf{H}}_{0N_0}$ is the estimated channel matrix and $\tilde{\mathbf{H}}_{0N_0}$ is the normalized channel estimation error matrix. The system model with interference and typical tier-B cluster is shown by Fig. 1.

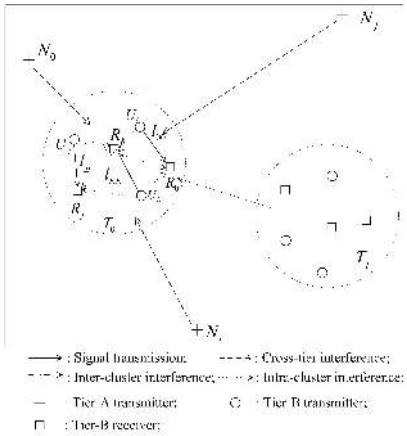


Fig. 1: System model with different interference and tier-B typical cluster.

The desired signal is d_B-dimensional and the interference signals are confined in the remaining $(N_B - d_B)$-dimensional subspace at the tier-B receivers. When it comes to the role of IA within this paper, our main aim is to show that it is indeed possible to use IA to cancel out the nearest cross-tier interference and the intra-cluster interference. Thus we only provide details on the feasibility conditions; this keeps the IA parts brief and to point, so as not to distract from the main focus of the paper which is the throughput analysis. Looking closely at our proposed system it can be noticed that the tier B nodes on their own essentially form a standard interference channel (IC), however our IA scheme also needs to handle an additional interference component from the closest tier A interferer. Given this system structure, we can adapt well-established IA schemes proposed for the IC such as Max-SINR, Min-WLI [23] and MMSE based IA [24] to our system. Such algorithms have already been shown to be applicable to networks beyond the standard IC, with extensions to the interference broadcast channels (IBC) reported in [25]. Because all of the intra-cluster interference is eliminated, therefore, the post-processing signal at R_0 can be expressed as,

$$\hat{\mathbf{y}}_0 = l_{00}^{-\alpha/2}\mathbf{U}_0^H\mathbf{H}_{00}\mathbf{V}_0\mathbf{s}_0 + \mathbf{U}_0^H\sum_{N_i\in\Phi_A} l_{0N_i}^{-\alpha/2}\mathbf{H}_{0N_i}\mathbf{V}_{N_i}\mathbf{s}_{N_i} + \mathbf{U}_0^H\sum_{T_i\in\Phi_B/\phi_0} l_{0i}^{-\alpha/2}\mathbf{H}_{0i}\mathbf{V}_i\mathbf{s}_i + \mathbf{n}_0, \tag{3}$$

where $\mathbf{U}_0 \in \mathbb{C}^{N_B\times d_B}$ is a decorrelator (also known as interference nulling or zero-forcing (ZF) receiver) for IA at R_0 and the intra-cluster interference is eliminated with the help of perfect CSIT. At the q-th data stream of R_0, the desired signal power, denoted as $s_{0,q}$, the cross-tier interference power, denoted as I_c, and the inter-cluster interference power, denoted

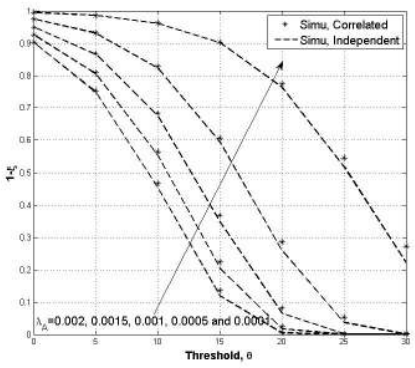


Fig. 2: Independence assumption of Tier-A interference.

as I_{inter}, are given by the following expressions:

$$s_{0,q} = l_{00}^{-\alpha} \left\| \mathbf{w}_q^H \mathbf{H}_{00} \mathbf{V}_0 \mathbf{s}_0 \right\|^2, \tag{4}$$

$$I_c = \left\| \mathbf{w}_q^H \sum_{N_i \in \Phi_A} l_{0N_i}^{-\alpha/2} \mathbf{H}_{0N_i} \mathbf{V}_{N_i} \mathbf{s}_{N_i} \right\|^2, \tag{5}$$

$$I_{inter} = \left\| \mathbf{w}_q^H \sum_{T_i \in \Phi_B / \phi_0} l_{0i}^{-\alpha/2} \mathbf{H}_{0i} \mathbf{V}_i \mathbf{s}_i \right\|^2, \tag{6}$$

where $\mathbf{w}_q$ is the decorrelator of the q-th data stream and is also the q-th column of $\mathbf{U}_0$.

The total interference of the q-th data stream at R_0, denoted as I_q, is the sum of the two different interferences: the inter-cluster interference I_{inter} and the cross-tier interference I_c. By further splitting the cross-tier interference into two sub-interferences: the nearest cross-tier interference I_n and the remaining cross-tier interference I_A, the total interference of the q-th data stream at R_0 is given by,

$$I_q = I_{inter} + I_n + I_A, \quad \text{denote} \quad I_n + I_A = I_c. \tag{7}$$

If the channel entries are i.i.d. $\mathcal{CN}(0,1)$ and R_0 performs single-stream decoding as in [26], the signal power is exponentially distributed with unit mean, i.e., $s_{0,q} \sim \text{Exp}(1)$. Therefore, the outage probability, which is the probability that the SIR is smaller than the pre-specified threshold θ, can be expressed as

$$\xi = \Pr\left(\frac{s_{0,q}}{I_q} < \theta \right) = 1 - \mathbb{E}_{I_q}\left[\exp\left(-\frac{\rho_A \theta l_{00}^{\alpha} I_q}{\rho_B} \right) \right]. \tag{8}$$

Note that as N_0 is the nearest Tier-A transmitter to R_0, and we assume the location l_{iN_0} is independent with the other locations l_{iN_i}, where $N_i \neq N_0$, leading to independence between the nearest cross-tier interference and the remaining cross-tier interference and can be approximated as

$$\begin{aligned} &\mathbb{E}\left[\exp\left(-\frac{\rho_A \theta l_{00}^{\alpha} \left(I_n + I_A \right)}{\rho_B} \right) \right] \\ &\approx \mathbb{E}\left[\exp\left(-\frac{\rho_A \theta l_{00}^{\alpha} I_n}{\rho_B} \right) \right] \mathbb{E}\left[\exp\left(-\frac{\rho_A \theta l_{00}^{\alpha} I_A}{\rho_B} \right) \right]. \end{aligned} \tag{9}$$

This approximation is reasonable for Tier-A BSs and can be verified by Fig. 2, where we assume $R = 100$, $N_A = N_B = 2$, $l_{00} = 1$, $\alpha = 2$ and $\rho_A = \rho_B = 1$. The simulation shows that the approximation provides tight lower bound when we ignore the correlation between the interference of the nearest and the other Tier-A notes in our system.

In order to evaluate the performance of the IA approach in the proposed HetNet, throughput is selected as the performance metric, which is defined as the rate of successful message delivery over a communication system as in [26]. It is observed that throughput not only considers the success probability that is affected by interference, but also includes the multiplexing gain that can be enhanced by the IA approach. Therefore, it is an ideal metric to evaluate the Tier-B performance in our system. With properties of the exponential function, the throughput denoted as $\mathcal{T}$, can be expressed as

$$\begin{aligned} \mathcal{T} \approx & KRd_B \underbrace{\mathbb{E}\left[\exp\left(-\frac{\rho_A \theta l_{00}^{\alpha} I_n}{\rho_B} \right) \right]}_{p_s^n} \underbrace{\mathbb{E}\left[\exp\left(-\frac{\rho_A \theta l_{00}^{\alpha} I_A}{\rho_B} \right) \right]}_{p_s^A} \\ & \times \underbrace{\mathbb{E}\left[\exp\left(-\frac{\rho_A \theta l_{00}^{\alpha} I_{inter}}{\rho_B} \right) \right]}_{p_s^{inter}}, \end{aligned} \tag{10}$$

where p_s denotes the overall success probability which is equal to $1 - \xi$ and R is the data rate at each data stream. Symbols p_s^n, p_s^A and p_s^{inter} denote the success probabilities affected by the nearest cross-tier interference, the remaining cross-tier interference and the inter-cluster interference, respectively. Note that the success probability affected by the inter-cluster interference has been well-studied in [27]. In the scenario when all the transmitters have the same power constraint and number of data streams, p_s^{inter} can be expressed by the following expression:

$$p_s^{inter} = \exp\left(-\lambda_B \left(\theta l_{00}^{\alpha} \right)^{\frac{2}{\alpha}} \frac{\Gamma(Kd_B + \frac{2}{\alpha})\Gamma(1 - \frac{2}{\alpha})}{\Gamma(Kd_B)} \right), \tag{11}$$

where $\Gamma(\cdot)$ is the Gamma function.

III. Interference Mitigation

In this section, the nearest cross-tier interference, i.e., the interference from N_0, is decomposed at first, which leads to the feasibility condition of the IA approach. We then design the distance-dependent IA approach using transmitters in the cluster ϕ_0 to maximize the throughput where the receivers utilize the decorrelators to obtain the desired signals.

A. Interference decomposition

The interference signal from the nearest cross-tier transmitter N_0 received at R_0 is $l_{0N_0}^{-\alpha/2} \mathbf{U}_0^H \mathbf{H}_{0N_0} \mathbf{V}_{N_0} \mathbf{s}_{N_0}$, where $\mathbf{s}_{N_0} = [s_{N_0}^{(1)}\ s_{N_0}^{(2)}\ \ldots\ s_{N_0}^{(d_A)}]^T$ and the matrix $\mathbf{H}_{0N_0}\mathbf{V}_{N_0}$ can be rewritten as $[\mathbf{v}_1\ \mathbf{v}_2\ \ldots\ \mathbf{v}_{d_A}]$, where $\mathbf{v}_i \in \mathbb{C}^{N_B \times 1}$. Intuitively, $\mathbf{v}_i$ is the beamforming vector for the i-th symbol in $\mathbf{s}_{N_0}$. Recall that each column of $\mathbf{U}_0$ is the decorrelator for each data stream, i.e., $\mathbf{U}_0 = [\mathbf{w}_1\ \mathbf{w}_2\ \ldots\ \mathbf{w}_{d_B}]$.

The squared of the norm of each entry $\mathbf{w}_i^H \mathbf{v}_j$ is Chi-squared distributed with *two* degree of freedom (DoF) for following reasons. The vectors $\mathbf{v}_i, i = 1, 2, \ldots, d_A$ include the complex Gaussian distributed channel entries, the squared sum of the real and imaginary parts of it leading to Chi-squared distribution. Moreover, because all the columns of

$\mathbf{U}_0$ and $\mathbf{V}_{N_0}$ have unit norm, they have no impact on the signal distribution. Therefore, multiplying the effective channel $\mathbf{H}_{0N_0}\mathbf{V}_{N_0}$ with $\mathbf{w}_q^H$, the interference received at the q-th data stream of R_0 is obtained, with the power having Chi-squared distributed with $2d_A$ DoF. If the interference from the j-th data stream from N_0 is eliminated at the i-th data stream of R_0, then $\mathbf{w}_i^H\mathbf{v}_j = 0, i = 1, 2, \ldots, d_B$. Because all of the channel entries and symbols are i.i.d., it is possible to partially eliminate interference from N_0 at R_0. If d_a out of d_A data streams are treated as interference, the partially interference estimation can be expressed as follows:

$$l_{0N_0}^{-\alpha/2}\mathbf{U}_0^H\hat{\mathbf{H}}_{iN_0}\mathbf{V}_{N_0}\mathbf{\Sigma}_{d_a} = \mathbf{0}, \tag{12}$$

where $\mathbf{\Sigma}_{d_a}$ contains arbitrary d_a columns of a $d_A \times d_A$ identity matrix, denoted as $\mathbf{I}^{(d_A)}$.

Intuitively, when the distance from R_0 to N_0 is large, the interference power from N_0 overheard by R_0 is modeled as shot noise without sacrificing much SIR at R_0. However, as the distance decreases, the interference cannot be regarded as noise and it is reasonable to partially eliminate the interference. Hence, there exists a trade-off between treating interference from N_0 as noise and treating it as interference, i.e., a trade-off between increasing the SIR of each data stream of R_0 and reducing the number of data streams from T_0, depending on the distance information (i.e. implementing IA approach in a distance-dependent way).

B. Feasibility condition

The feasibility condition is the main factor that limits the effectiveness of MIMO IA. As proved in [28], IA is feasible only when the number of variables is no less than the number of equations that are required by the IA beamforming scheme. The feasibility condition of this HetNet has been studied in [29]. Considering the scenario that $K \geq 3$, which is typical when IA is implemented, the constraint of d_a can be equivalently written as,

$$d_a \leq \min\left\{\frac{N_B(K-1)}{K}, d_A\right\}. \tag{13}$$

This inequality shows the upper bound of the number of data streams from N_0 that can be feasibly cancelled at R_0.

C. Distance-Dependent Interference Alignment

As shown in the previous subsection, d_a out of d_A data streams from N_0 are placed in the interference subspace (treated as interference) at R_0 using the IA approach and the remaining $d_A - d_a$ data streams are modeled as shot noise. The optimal value of d_a is investigated so that the throughput can be maximized. Considering the receiving matrix, the IA equations that show the interference from ϕ_0 is fully eliminated and interference from N_0 is partially eliminated are given by

$$\mathbf{U}_0^H\hat{\mathbf{H}}_{0N_0}\mathbf{V}_{N_0}\mathbf{\Sigma}_{d_a} = \mathbf{0}, \tag{14}$$

$$\mathbf{U}_0\mathbf{H}_{0i}\mathbf{V}_i = \mathbf{0}, \quad \forall T_i \in \phi_0, i \neq 0 \tag{15}$$

$$\mathrm{rank}(\mathbf{U}_0\mathbf{H}_{00}\mathbf{V}_0) = d_B. \tag{16}$$

Because of the imperfect CSI (with the variance of estimation error being ε^2), each of the d_a data streams that is treated as interference, still causes interference with power reduced by ε^2 compared to the interference caused by each of the other $d_A - d_a$ data streams. The interference power at the q-th data stream of R_0 is, therefore, represented as,

$$I_n = \frac{\rho_A}{l_{0N_0}^{\alpha}}\left(\varepsilon^2\|\mathbf{w}_q^H\tilde{\mathbf{H}}_{0N_0}\mathbf{V}_{N_0}\mathbf{\Sigma}_{d_a}\|^2 + \|\mathbf{w}_q^H\mathbf{H}_{0N_0}\mathbf{V}_{N_0}\tilde{\mathbf{\Sigma}}_{d_a}\|^2\right) \tag{17}$$

$$= I_n^{(1)} + I_n^{(2)}, \tag{18}$$

where $\tilde{\mathbf{\Sigma}}_{d_a}$ is the complementary $d_A - d_a$ columns of $\mathbf{\Sigma}_{d_a}$ in the identity matrix. As the channel is complex Gaussian distributed, $I_n^{(1)}$ and $I_n^{(2)}$ are Chi-squared distributed with $2d_a$ and $2(d_A - d_a)$ DoF respectively.

Theorem 1: The success probability at each data stream of receiver R_0 affected by the nearest cross-tier interferer N_0 under Rayleigh fading is given by,

$$p_s^n = \left(\frac{\rho_A}{\rho_B}l_{0N_0}^{-\alpha}\theta l_{00}^{\alpha} + 1\right)^{-d}, \tag{19}$$

where N_0 is assumed to transmit d data streams.

Proof: See Appendix A.

The interference from N_0 here consists of two additive interference sources: the d_a data streams that are treated as interference and remaining data streams that are modeled as shot noise. Hence, the success probability affected by interference from N_0 is the product of the success probabilities affected by the two interference sources. Based on Theorem 1, the success probability affected by N_0 is considered as a product of the two success probabilities, one is affected by the interference from d_a data streams with power $\varepsilon^2\rho_A$ and the other affected by the interference from $d_A - d_a$ data streams with symbol power ρ_A. Thus, considering the success probability of each data stream of R_0 affected by N_0 and $d_B = \frac{2N_B - d_a}{K+1}$, the corresponding throughput is obtained as,

$$\mathcal{T} = \frac{KRp_s^A p_s^{inter}\left(\frac{2N_B-d_a}{K+1}\right)\left(\frac{\rho_A\theta l_{00}^{\alpha}}{\rho_B l_{0N_0}^{\alpha}} + 1\right)^{d_a - d_A}}{\left(\frac{\rho_A}{\rho_B}l_{0N_0}^{-\alpha}\varepsilon^2\theta l_{00}^{\alpha} + 1\right)^{d_a}}. \tag{20}$$

From this equation, it is observed that the success probability is affected by the distance between N_0 and R_0 which may or may not be known in practice. Therefore, the value of d_a that maximizes the throughput is found in two scenarios: (i) l_{0N_0} is only known statistically and (ii) l_{0N_0} is accurately known. Generally speaking, a large d_a leaves a lower dimensional subspace for the desired signals because $d_B = \frac{2N_B - d_a}{K+1}$, but the SIR of each data stream increases. For small d_a, there will be a higher dimensional subspace for the desired signals, but the SIR at each data stream will be lower. It is worth to notice that the expected throughput is maximized independently with p_s^A. The reason behind this is that transmitter-receiver pairs in a cluster is maximizing the throughput wherein the Tier-A transmitters are not part of the scheme.

1) Receivers have statistical distance knowledge: In this scenario, we assume that each Tier-B receiver only has access to the intensity of the Tier-A transmitters which means that

$$\begin{aligned}\mathbb{E}[\mathcal{T}] =&2\lambda_A\pi K R p_s^A \left(\frac{2N_B-d_a}{K+1}\right)\exp\left(-\lambda_B\left(\theta l_{00}^{\alpha}\right)^{\frac{2}{\alpha}}\frac{\Gamma\left(K\left(\frac{2N_B-d_a}{K+1}\right)+\frac{2}{\alpha}\right)\Gamma\left(1-\frac{2}{\alpha}\right)}{\Gamma\left(K\left(\frac{2N_B-d_a}{K+1}\right)\right)}\right)\\ &\times\int_0^{\infty}\left(\frac{\rho_A(2N_B-d_a)}{P_B(K+1)}x^{-\alpha}\varepsilon^2\theta l_{00}^{\alpha}+1\right)^{-d_a}\left(\frac{\rho_A(2N_B-d_a)}{P_B(K+1)}x^{-\alpha}\theta l_{00}^{\alpha}+1\right)^{-d_A+d_a}xe^{-\lambda_A\pi x^2}\mathrm{d}x,\end{aligned} \tag{21}$$

it only knows the probability density function (PDF) of its distance to the nearest cross-tier transmitter. In this way, all the Tier-B receivers will have the same expected throughput and will choose the same optimal d_a, so that the overall expected throughput is maximized, where the optimal value of d_a is denoted as $\hat{d}_a$. Considering $d_B = \frac{2N_B-d_a}{K+1}$, the expectation of the throughput is given by (21), where (21) is obtained by taking (11) into consideration. As a result, the value of d_a that maximizes the throughput, i.e., $\hat{d}_a$, is found by the following expressions,

$$\begin{aligned}\arg\max_{d_a}\quad &\mathbb{E}[\mathcal{T}],\\ \text{s.t.}\quad &d_a \le \min\left\{\frac{N_B(K-1)}{K}, d_A\right\}.\end{aligned} \tag{22}$$

After obtaining $\hat{d}_a$, the closed-form expression of (21) is not available for general α. However, it is available when the quality of the channel between R_0 and N_0 is high, i.e., $\varepsilon = 0$ and α is fixed. For example, when $\varepsilon = 0$ and $\alpha = 4$, the integral part of (21) becomes,

$$\begin{aligned}&\int_0^{\infty}\left(\frac{\rho_A}{\rho_B}l^{-4}\theta l_{00}^{\alpha}+1\right)^{-d_A+\hat{d}_a} le^{-\lambda_A\pi l^2}dl\\ &=\frac{\sqrt{\theta l_{00}^{\alpha}\frac{\rho_A}{\rho_B}}\mathcal{G}_{1,3}^{3,1}\left(\frac{(\pi\lambda_A)^2\theta l_{00}^{\alpha}\rho_A}{4\rho_B}\middle|\begin{matrix}\hat{d}_a-d_A+\frac{1}{2}\\ \\ -\frac{1}{2},0,\frac{1}{2}\end{matrix}\right)}{4\sqrt{\pi}\Gamma(d_A-d_a)},\end{aligned} \tag{23}$$

where $\mathcal{G}(\cdot)$ is the Meijer-G function. The closed-form expression of an upper bound of (21) is available. Specifically, considering the PDF of $f_{N_0}(x)$, the integral part can be rewritten as

$$\mathbb{E}_x\left[\left(\frac{\rho_A}{\rho_B}x^{-\alpha}\varepsilon^2\theta l_{00}^{\alpha}+1\right)^{-d_a}\left(\frac{\rho_A}{\rho_B}x^{-\alpha}\theta l_{00}^{\alpha}+1\right)^{-d_A+d_a}\right], \tag{24}$$

where the inside term is a concave function with respect to x. Using Jensen's inequality, a closed-form expression of upper bound of throughput is given by (25).

2) Receivers have accurate distance knowledge: In this scenario, each Tier-B receiver has accurate knowledge of its distance to the nearest cross-tier transmitter, instead of statistical knowledge. We assume that the average optimal d_a chosen by all the Tier-B transmitters and receivers is $\bar{d}_a$ and the number of data streams from N_0 that are treated as interference at R_0 is denoted as $d_a^{(0)}$, the optimal value of that is $\hat{d}_a^{(0)}$. Unlike the case with statistical distance knowledge where all the Tier-B clusters have the same expected throughput, the challenge in this case is that the receivers in different clusters have different optimal solutions that affect $\bar{d}_a$. Specifically, the choice of $\hat{d}_a^{(0)}$ also affects $\bar{d}_a$, which in turn, affect the choice of $\hat{d}_a^{(0)}$. Therefore, R_0 has to estimate $\bar{d}_a$ first before solving $\hat{d}_a^{(0)}$.

Step 1: Finding $\bar{d}_a$. Defining all the terms in (20) that contain $d_a^{(0)}$ as $\mathcal{Q}$ and considering ρ_B is the Tier-B transmit power per data stream, the $\mathcal{Q}$ can be expressed as (26). It can be easily proved that $\mathcal{Q}$ is a concave function with respect to $d_a^{(0)}$, which implies that throughput is also concave with respect to $d_a^{(0)}$. Hence, the optimal value $\hat{d}_a^{(0)}$ is the value of $d_a^{(0)}$ that maximizes $\mathcal{Q}$, which is obtained by the following expressions:

$$\begin{aligned}\arg\max_{d_a^{(0)}}\quad &\mathcal{Q},\\ \text{s.t.}\quad &d_a^{(0)} \le \min\left\{\frac{N_B(K-1)}{K}, d_A\right\}.\end{aligned} \tag{27}$$

The average number of data streams that are treated as interference by Tier-B transmitters and receivers is obtained by solving the equation, $\int_0^{\infty} d_a^{(0)}(x) f_{N_0}(x)\mathrm{d}x = \bar{d}_a$. The solution of this equation is the $\bar{d}_a$ so that the expected value of $d_a^{(0)}$ is exactly equal to $\bar{d}_a$. Therefore, the Tier-B transmitters transmit $\frac{2N_B-\bar{d}_a}{K+1}$ data streams on average.

Step 2: Finding $\hat{d}_a^{(0)}$. After obtaining $\bar{d}_a$, the optimal number of data streams from N_0 that are treated as interference by R_0 is found by using (27) with a known $\bar{d}_a$. By maximizing the throughput for each specific l_{0N_0} realization, the expected throughput achieved by the K transmitter-receiver pairs in ϕ_0 is given by (28).

IV. Success probability and throughput

In this section, the success probabilities that are affected by the remaining interference from Tier-A transmitters as well as the inter-cluster interference from Tier-B transmitters are derived using stochastic geometry technique.

A. Cross-tier interference from Tier-A nodes except for N_0.

The remaining cross-tier interference is the interference caused by the Tier-A transmitters except N_0, which is essentially a PPP without the nearest point. As an extension of the analysis in [29], for Tier-A transmitters with intensity λ_A and $v(l) = \exp\left(-\frac{\theta l_{00}^{\alpha} h\rho_A l^{-\alpha}}{2\rho_B}\right)$, the success probability affected by all Tier-A transmitters except N_0 is expressed as,

$$\begin{aligned}p_s^A =&\exp\left(-2\pi\lambda_A\mathbb{E}_h\left[\int_0^{\infty}\int_x^{\infty}\left(1-\exp\left(-\frac{\theta l_{00}^{\alpha}h\rho_A}{2\rho_B l^{\alpha}}\right)\right)\right.\right.\\ &\left.\left.\times x f_{N_0}(l)\mathrm{d}l\mathrm{d}x\right]\right).\end{aligned} \tag{29}$$

Remark: Symbol h denotes the effective fading coefficient and is Chi-squared distributed with $2d_A$ DoF. In Tier-A, all of the transmitters are independently located which means that all of the transmitters except N_0 itself are distributed within radius $r_0 < r_\xi < \infty$. This property of PPP is used to calculate the average value of interference power from $N_i, \forall i \neq 0$. Specifically, as seen in the inner integral, the integration

$$\begin{aligned}\mathbb{E}[\mathcal{T}] \leq & KRp_s^A \left(\frac{2N_B-\hat{d}_a}{K+1}\right) \exp\left(-\lambda_B\left(\theta l_{00}^{\alpha}\right)^{\frac{2}{\alpha}} \frac{\Gamma\left(K\left(\frac{2N_B-\hat{d}_a}{K+1}\right)+\frac{2}{\alpha}\right)\Gamma\left(1-\frac{2}{\alpha}\right)}{\Gamma\left(K\left(\frac{2N_B-\hat{d}_a}{K+1}\right)\right)}\right) \\ & \times \left(\frac{\rho_A(2N_B-d_a)}{P_B(K+1)} 2^{\alpha}\lambda_A^{\alpha/2}\varepsilon^2\theta l_{00}^{\alpha}+1\right)^{-\hat{d}_a} \left(\frac{\rho_A(2N_B-d_a)}{P_B(K+1)} 2^{\alpha}\lambda_A^{\alpha/2}\theta l_{00}^{\alpha}+1\right)^{-d_A+\hat{d}_a}\end{aligned} \tag{25}$$

$$\begin{aligned}\mathcal{Q} = & (2N_B-d_a^{(0)}) \exp\left(-\lambda_B\left(\theta l_{00}^{\alpha}\right)^{\frac{2}{\alpha}} \frac{\left(2N_B-d_a^{(0)}\right)\Gamma\left(K\left(\frac{2N_B-\bar{d}_a}{K+1}\right)+\frac{2}{\alpha}\right)\Gamma\left(1-\frac{2}{\alpha}\right)}{\left(2N_B-\bar{d}_a\right)\Gamma\left(K\left(\frac{2N_B-\bar{d}_a}{K+1}\right)\right)}\right) \\ & \times \left(\frac{\rho_A(2N_B-d_a^{(0)})}{P_B(K+1)} l_{0N_0}^{-\alpha}\varepsilon^2\theta l_{00}^{\alpha}+1\right)^{-d_a^{(0)}} \left(\frac{\rho_A(2N_B-d_a^{(0)})}{P_B(K+1)} l_{0N_0}^{-\alpha}\theta l_{00}^{\alpha}+1\right)^{-d_A+d_a^{(0)}}.\end{aligned} \tag{26}$$

$$\begin{aligned}\mathbb{E}[\mathcal{T}] = & 2\lambda_A\pi kRp_s^A \int_0^{\infty} \left(\frac{2N_B-\hat{d}_a^{(0)}}{K+1}\right) \exp\left(-\lambda_B\left(\theta l_{00}^{\alpha}\right)^{\frac{2}{\alpha}} \frac{\left(2N_B-\hat{d}_a^{(0)}\right)\Gamma\left(K\left(\frac{2N_B-\bar{d}_a}{K+1}\right)+\frac{2}{\alpha}\right)\Gamma\left(1-\frac{2}{\alpha}\right)}{\left(2N_B-\bar{d}_a\right)\Gamma\left(K\left(\frac{2N_B-\bar{d}_a}{K+1}\right)\right)}\right) \\ & \times \left(\frac{\rho_A(2N_B-\hat{d}_a^{(0)})}{P_B(K+1)} x^{-\alpha}\varepsilon^2\theta l_{00}^{\alpha}+1\right)^{-\hat{d}_a^{(0)}} \left(\frac{\rho_A(2N_B-\hat{d}_a^{(0)})}{P_B(K+1)} x^{-\alpha}\theta l_{00}^{\alpha}+1\right)^{-d_A+\hat{d}_a^{(0)}} \mathrm{d}x.\end{aligned} \tag{28}$$

of the interference is from x, which is the integral of all $r_\xi, \xi \neq 0$. By further considering the stochastic properties of N_0, the analytical expression of the expectation of interference power from the remaining cross-tier interferers is obtained. Therefore, the success probability affected by the cross-tier interference is divided into two components: nearest cross-tier interference and all the other remaining cross-tier interference.

Theorem 2: If interferers are Poisson distributed and the nearest interference is ignored, the success probability achieved at the intended receiver is given by

$$\begin{aligned}p_s = & \exp\left(1+\frac{2\pi\lambda\left(\frac{\rho_A}{\rho_B}\beta\right)^{\frac{2}{\alpha}} b^2\Gamma(d+\frac{2}{\alpha})\Gamma(-\frac{2}{\alpha})}{\alpha\Gamma(d)}\right. \\ & \left. -\pi\lambda\int_0^{\infty} e^{-\pi\lambda t}\left(1+\frac{\rho_A}{\rho_B}\theta b^{\alpha}t^{-\alpha/2}\right)^{-d} dt\right),\end{aligned} \tag{30}$$

where λ is the intensity of the PPP and d is the number of data streams transmited by N_0.

Proof: See Appendix B

Although Theorem 2 is valid for any $\alpha > 2$, the integral part cannot be simplified directly. However, for any particular value of α, closed-form expressions are available. For example, when $\alpha = 4$, we have,

$$\begin{aligned}& \int_0^{\infty} e^{-\pi\lambda_A t}\left(1+\frac{\rho_A}{\rho_B}\theta l_{00}^{\alpha}t^{-2}\right)^{-d_A} \mathrm{d}t \\ & = \frac{\sqrt{\frac{\rho_A}{\rho_B}\theta l_{00}^{\alpha}}\mathcal{G}_{1,3}^{3,1}\left(\frac{(\pi\lambda_A)^2\rho_A\theta l_{00}^{\alpha}}{4\rho_B}\middle|\begin{matrix}\frac{1}{2}-d_A\\ -\frac{1}{2},0,\frac{1}{2}\end{matrix}\right)}{2\sqrt{\pi}\Gamma(d_A)}.\end{aligned} \tag{31}$$

Although closed-form expression is not available for general α, the upper bound can be found by regarding $\pi\lambda_A\exp(-\pi\lambda_A t)$ as the PDF of t. Therefore, the integral part in the exponential function of Theorem 2 is interpreted as $\mathbb{E}\left[(1+\frac{\rho_A}{\rho_B}\theta l_{00}^{\alpha}t^{-\alpha/2})^{-d_A}\right]$, which is concave in terms of t. Furthermore, using Jensen's inequality, an upper bound of the success probability affected by the remaining cross-tier interference is written as,

$$\begin{aligned}p_s^A \leq & \exp\left(1+\frac{2\pi\lambda_A\left(\frac{\rho_A}{\rho_B}\theta\right)^{\frac{2}{\alpha}} l_{00}^2\Gamma(d_A+\frac{2}{\alpha})\Gamma(-\frac{2}{\alpha})}{\alpha\Gamma(d_A)}\right. \\ & \left. -\left(1+\frac{\rho_A}{\rho_B}\theta l_{00}^{\alpha}(\pi\lambda_A)^{\alpha/2}\right)^{-d_A}\right).\end{aligned} \tag{32}$$

V. Numerical Results

In this section, we present numerical results based on the system model given in Section II. The Monte Carlo simulations are conducted by first generating transmitters on the plane, where at least 200 transmitters in each tier are included. Unless stated otherwise, all the simulations have the pre-specified SIR threshold as $\theta = 10$, path-loss exponent $\alpha = 4$, distance between the intended transceiver is $l_{00} = 1$ and the number of transceiver pairs in each cluster is $K = 3$. Each point is averaged over 3000 channel and distribution realizations.

Fig. 3 shows the optimal number of data streams from N_0 that should be treated as interference at R_0, i.e., $\hat{d}_a$, versus the Tier-A transmitter intensity λ_A, which is assumed to be statistically known by Tier-B receivers. According to (13), $\hat{d}_a$ is constrained to $\frac{N_B(K-1)}{K} = 8$. It is observed that $\hat{d}_a$ increases with Tier-A transmitter intensity. The reason behind this is that the nearest cross-tier interference becomes more dominant when λ_A increases, thus, it is reasonable to cancel more of its interference. Meanwhile, $\bar{d}_a$ increases with better channel estimation quality, i.e., smaller variance of channel estimation error ξ. This is because with smaller ξ, there will be significant difference between the interference generated by the data streams that are treated as interference and that are modeled as shot noise. Thus, canceling one data stream from N_0 brings more SIR gain. Another important observation from this figure is that $\bar{d}_a$ is larger than *zero* even when $\xi = 1$, i.e., there is no difference between treating one data stream as interference or as shot noise. One reasonable explanation is that d_B will be large when $\bar{d}_a$ is small when we assume

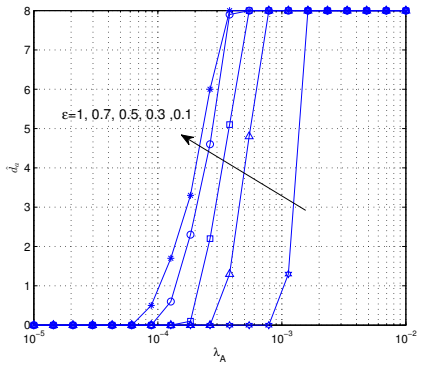


Fig. 3: Optimal number of data streams from N_0 that need to be put into interference subspace at R_0 versus intensity λ_A with different ξ for $\lambda_B = 0.005$, $\rho_A = 10$ W, $P_B = 1$ W, $\alpha = 3$ and $N_B = d_A = 12$.

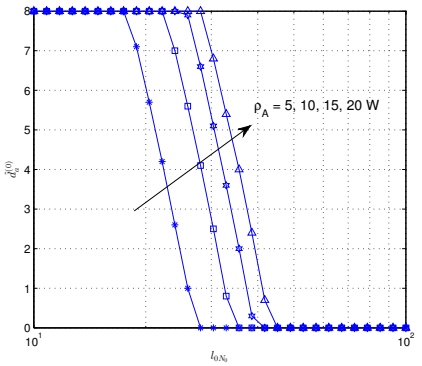


Fig. 4: Optimal number of data streams from N_0 that need to be put into interference subspace at R_0 versus distance between R_i and N_0 with different Tier-A transmit power for $\lambda_B = 0.005$, $\xi = 0.1$, $P_B = 1$ W, $\alpha = 3$ and $N_B = d_A = 12$.

$d_B = \frac{2N_B - \bar{d}_a}{K+1}$. Thus the decrease in the power per data stream offsets the gain brought by the multiplexing gain.

Fig. 4 shows $\hat{d}_a^{(0)}$ versus the distance to the nearest Tier-A transmitter, i.e., l_{0,N_0}, which is assumed to be statistically known by Tier-B receivers. Similarly, $\hat{d}_a^{(0)}$ is constrained to 8 which is obtained from (13). It is observed that with larger ρ_A or smaller l_{0N_0}, interference power from each data stream of N_0 to R_0 becomes so critical that it will considerably reduce the SIR at R_0 if the interference is modeled as shot noise. Therefore, more data streams from N_0 are put into the interference subspace. One important observation is that $\hat{d}_a$ will decrease to *zero* as l_{0N_0} increases. Intuitively, as there is always d_B-dimensional interference subspace at each Tier-B receiver, $\hat{d}_a$ should be no smaller than d_B. However, Bezout's theorem states that the essence of the IA scheme is a number of equations and variables. Aligning every single additional data stream is an overhead to the beamforming or receiving matrices. Finally, when $\hat{d}_a = 0$, all nearest cross-tier interferences are modeled as shot noise, which is how IA is implemented in [16], [17]. Clearly, ignoring the nearest cross-tier interference is only optimal when l_{0N_0} is large.

Fig. 5 illustrates the throughput versus transmit power P_A where $N_A = d_A = N_B = 8$, $\alpha = 3$ and $\lambda_A = 0.01$. In order to focus on the nearest cross-tier interference, we assume

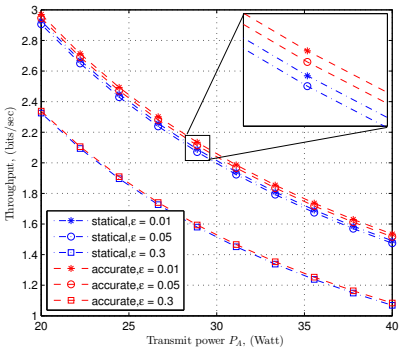


Fig. 5: Throughput affected by the nearest cross-tier interferer versus Tier-A transmit power with different channel estimation error. Throughputs with and without the knowledge to the nearest cross-tier interferer are considered for $\alpha = 3$, $\lambda_A = 0.01$ and $N_A = d_A = N_B = 8$.

$p_s^A = 1$. Throughput achieved with the accurate knowledge of the distance l_{0N_0} outperforms the curve with only statistical knowledge. The reason is that, with accurate knowledge, a proper value of $\hat{d}_a$ can be selected to maximize the throughput for each transmitter distribution realisation, while with statistical knowledge of l_{0N_0}, $\hat{d}_a$ can only be designed to maximize the expected throughput. Interestingly, the gap between the curves with accurate knowledge and statistical knowledge increases as the quality of CSI increases, the reason being that with larger ε, convexity of (20) in terms of $\bar{d}_a$ increases. However, even when ε decreases to 0.01, the throughput gap between the curves with and without the accurate distance knowledge is less than 0.1 bits/s. Conclusively speaking, obtaining accurate distance knowledge to the nearest cross-tier transmitter cannot significantly improve the performance. Another observation from this figure is that decreasing ε from 0.05 to 0.01 will not achieve significant throughput gain. This observation is contrary to that in [22]. The reason behind this is that the proposed distance-dependent IA approach also takes the variance of CSI error into consideration, i.e., when the variance is large, less data streams are expected to be eliminated. In other words, the trade-off between increasing the multiplexing gain and SIR enhancement is effective.

Fig. 6 demonstrates the success probabilities at R_0, affected only by the Tier-A transmitters without the nearest one, i.e., p_s^A in Theorem 2, together with the corresponding Monte Carlo simulations. In this figure, $\alpha = 3.5$ and 4.5 are considered, respectively. It is observed that the success probability increases with α, which will cause more severe path-loss through the interference signal propagation. The gap between two groups of lines increases with Tier-A transmit power until a *zero* success probability is enforced. The reason is that success probability is an exponential function of interference power, i.e., $\exp(-\theta I_A)$, while the interference power increases only linearly with Tier-A transmit power. Moreover, this figure also shows that Theorem 2 accurately modeled and successfully considered the interference from the remaining cross-tier interferers.

Fig. 7 illustrates the mean throughput affected by all of the interference sources. Intra-cluster interference is perfectly can-

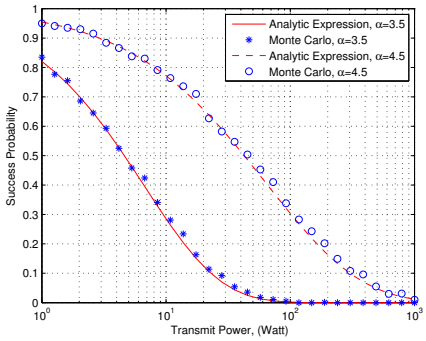
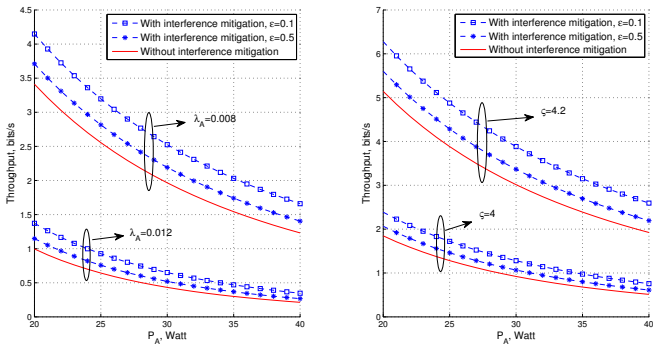


Fig. 6: Monte Carlo and analytic expression of success probability at R_i affected by the Tier-A nodes excluding the nearest one versus different transmit power with different path-loss exponents for $\lambda_A = 0.01$ and $d_A = 4$.



(a) With various λ_A. (b) With various α.

Fig. 7: The overall expected throughput versus Tier-A transmit power.

celled and the interference mitigation in this figure specifically refers to mitigating the interference from the nearest cross-tier transmitter. The distance l_{0N_0} is assumed to be statistically known and we set $\lambda_B = 0.05$ and $N_B = d_A = 8$. In Fig. 7a, we fix $\alpha = 4$ and change λ_A while in Fig. 7b, we fix $\lambda_A = 0.01$ and change α. Generally, in both subfigures, the throughput decreases with Tier-A transmit power and ξ. Specifically, in Fig. 7a, the gain from the line without interference mitigation to the line with interference mitigation ($\xi = 0.1$) is 23% when $\lambda_A = 0.008$, but rises by 40% when $\lambda_A = 0.012$. This is because the expected distance from Tier-A transmitters to R_0 decreases at the same rate when intensity increases, while the interference power from one transmitter decays exponentially with its distance to R_0. Hence, when transmitter intensity increases, the interference power from the closest transmitters becomes more dominant. As the distance-dependent IA approach only deals with the nearest cross-tier interference, it is more effective with larger λ_A. In Fig. 7b, it is observed that the throughput is highly sensitive to the path-loss exponent α. Although increasing α makes the nearest cross-tier interference more dominant, as shown in Fig. 7a, the throughput gain decreases with larger α. In detail, the throughput gain is 20% when $\alpha = 4.2$ and 31% when $\alpha = 4$. The reason behind this is that when α increases, interference power from all of the transmitters decreases, which reduces the effectiveness of interference cancellation. Actually, when $\alpha \to \infty$, the impact of path-loss will overwhelm all of the interferences.

VI. Conclusion

This paper considered a generic two-tier heterogeneous wireless network in which Tier-A transmitters (with large coverage range) were Poisson distributed while Tier-B transmitters (with small coverage range) were Poisson clustered distributed. A distance-dependent IA approach was proposed within the Tier-B clusters in which the intra-cluster interference was aligned to a part of the interference from the nearest cross-tier transmitter. The throughput of this HetNets was compared under various system settings and it was shown that: 1) obtaining the accurate knowledge of the distance to the nearest cross-tier transmitter was not significantly important; 2) small channel estimation error was not harmful to the throughput due to the effective trade-off between increasing multiplexing gain and the SIR enhancement; 3) distance-dependent IA approach was more effective with a smaller path-loss exponent and a higher intensity of Tier-A transmitters. The remaining different interferences were modelled with stochastic geometry tool. The closed-form expression of the success probability, affected by the remaining cross-tier interference, was derived together with a closed-form upper bound. Furthermore, accurate expression of the success probability was derived as well as its closed-form approximation when the cluster size was small. It was shown that the path-loss exponent influenced the throughput more significantly when the power of Tier-A transmitters is large.

References

[1] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[2] S. Toumpis and A. Goldsmith, "Capacity regions for wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 736–748, Jul. 2003.

[3] M. Ebrahimi, M. Maddah-Ali, and A. Khandani, "Throughput scaling laws for wireless networks with fading channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4250–4254, Nov. 2007.

[4] J. Zhang, C. K. Wen, S. Jin, X. Gao, and K. K. Wong, "On capacity of large-scale MIMO multiple access channels with distributed sets of correlated antennas," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 133–148, Feb. 2013.

[5] F. Khan, H. He, J. Xue, and T. Ratnarajah, "Performance analysis of cloud radio access networks with distributed multiple antenna remote radio heads," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4784–4799, Aug. 2015.

[6] S. Biswas, J. Xue, F. Khan, and T. Ratnarajah, "On the capacity of correlated massive MIMO systems using stochastic geometry," Jun. 2015.

[7] P. C. Pinto, A. Giorgetti, M. Z. Win, and M. Chiani, "A stochastic geometry approach to coexistence in heterogeneous wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1268–1282, Sep. 2009.

[8] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[9] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of k-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550–560, Apr. 2012.

[10] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, May 2013.

[11] X. Lin, J. G. Andrews, and A. Ghosh, "Modeling, analysis and design for carrier aggregation in heterogeneous cellular networks," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 4002–4015, Sep. 2013.

[12] H. S. Dhillon, M. Kountouris, and J. G. Andrews, "Downlink MIMO hetnets: Modeling, ordering results and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 5208–5222, Oct. 2013.

[13] Y. Zhong and W. Zhang, "Multi-channel hybrid access femtocells: A stochastic geometric analysis," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 3016–3026, Jul. 2013.

[14] W. Cheung, T. Quek, and M. Kountouris, "Throughput optimization, spectrum allocation, and access control in two-tier femtocell networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 561–574, Apr. 2012.

[15] M. A. Maddad-Ali, A. S. Motahari, and A. K. Khandani, "Signaling over MIMO multi-base systems: combaination of multi-access and broadcast schemes," in *Proc. IEEE Int. Symp. on Information Theory*, 2006, pp. 2104–2108.

[16] R. Tresch and M. Guillaud, "Performance of interference alignment in clustered wireless ad hoc networks," in *Proc. IEEE Int. Symp. on Information Theory*, Jun. 2010, p. 1703.

[17] R. Tresch, G. Alfano, and M. Guillaud, "Interference alignment in clustered ad hoc networks: high reliablity regime and per-cluster ALOHA," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, May 2011, pp. 3348–3351.

[18] R. F. Guiazon, K. K. Wong, and D. Wisely, "Capacity analysis of interference alignment with bounded CSI uncertainty," vol. 3, no. 5, pp. 505–508, Oct. 2014.

[19] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, Aug. 2008.

[20] T. Brown, "Cellular performance bounds via shotgun cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 11, pp. 2443–2455, Nov. 2000.

[21] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1479–1489, Dec. 2010.

[22] S. Weber, J. Andrews, X. Yang, and G. Veciana, "Transmission capacity of wireless ad hoc networks with successive interference cancellation," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2799–2814, Aug. 2007.

[23] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, Jun. 2011.

[24] D. A. Schmidt, S. Changxin, R. A. Berry, M. L. Honig, and W. Utschick, "Minimum mean squared error interference alignment," *in Proc. Asilomar Conf. Signals Sys. Comput. (ACSSC)*, 2009.

[25] P. Aquilina and T. Ratnarajah, "Performance analysis of IA techniques in the MIMO IBC with imperfect CSI," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1259–1270, Apr. 2015.

[26] R. Vaze and R. Heath, "Transmission capacity of ad-hoc networks with multiple antennas using transmit stream adaptation and interference cancellation," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 780–792, Feb. 2012.

[27] B. Nosrat-Makouei, J. Andrews, and R. Heath, "MIMO interference alignment over correlated channels with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2783–2794, 2011.

[28] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4771–4782, Sep. 2010.

[29] Y. Luo and T. Ratnarajah, "Throughput of two-tier heterogeneous wireless networks with interference coordination," *IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pp. 618–622, Sep. 2014.

APPENDIX A
PROOF OF THEOREM 1

The interference power from N_0 at the q-th data stream of R_0 is given by

$$\rho_A l_{0N_0}^{-\alpha} \left\| \mathbf{w}_q^H \mathbf{H}_{iN_0} \mathbf{V}_{N_0} \right\|^2, \tag{33}$$

Because each of the channel entry is complex Gaussian distribution with unit-variance and zero-mean, $\left\| \mathbf{w}_q^H \mathbf{H}_{iN_0} \mathbf{V}_{N_0} \right\|^2$ is Chi-squared distributed with $2d$ DoF with the PDF given by

$$f_{\chi^2}(2x, 2d) = \frac{(2x)^{d-1} e^{-x}}{2^d \Gamma(d)}. \tag{34}$$

Therefore, the success probability is expressed as,

$$p_s^n = \left(\frac{\rho_A}{\rho_B} l_{0N_0}^{-\alpha} \theta l_{00}^{\alpha} + 1 \right)^{-d}. \tag{35}$$

APPENDIX B
PROOF OF THEOREM 2

Taking the PDF of the distance to the nearest point in PPP $f_{N_0}(x)$ into (29), we have,

$$\begin{aligned} p_s^A =& \exp\left(-(2\pi\lambda_A)^2 \mathbb{E}_h \left[\int_0^\infty \int_x^\infty \left(1 - \exp\left(-\frac{\theta l_{00}^{\alpha} h \rho_A}{2\rho_B l^{\alpha}} \right) \right) \right.\right. \\ & \left.\left. \times l x \exp(-\pi\lambda_A x^2) \mathrm{d}l \mathrm{d}x \right] \right). \end{aligned} \tag{36}$$

With the following integral,

$$\begin{aligned} & \int_x^\infty \left(1 - \exp\left(-\frac{\theta l_{00}^{\alpha} h \rho_A l^{-\alpha}}{2\rho_B} \right) \right) l \mathrm{d}l \\ &= -\frac{x^2}{2} + \frac{1}{\alpha} \left(\frac{h\theta l_{00}^{\alpha} \rho_A}{2\rho_B} \right)^{2/\alpha} \\ & \times \left(-\Gamma\left(-\frac{2}{\alpha} \right) + \Gamma\left(-\frac{2}{\alpha}, \frac{h\theta l_{00}^{\alpha} \rho_A x^{-\alpha}}{2\rho_B} \right) \right), \end{aligned}$$

and $\int_0^\infty -\frac{x^2}{2} \left[x \exp\left(-\pi\lambda_A x^2 \right) \right] \mathrm{d}x = -\frac{1}{4\pi^2\lambda_A^2}$. The success probability is reduced to the following equation:

$$\begin{aligned} p_s^A =& \exp\left(\mathbb{E}_h \left[1 + \frac{2\pi\lambda_A \Gamma\left(-\frac{2}{\alpha}\right)}{\alpha} \left(\frac{h\theta l_{00}^{\alpha} \rho_A}{2\rho_B} \right)^{2/\alpha} \right.\right. \\ & - \frac{(2\pi\lambda)^2}{\alpha} \left(\frac{h\theta l_{00}^{\alpha} \rho_A}{2\rho_B} \right)^{2/\alpha} \int_0^\infty x \exp(-\pi\lambda_A x^2) \\ & \left.\left. \times \Gamma\left(-\frac{2}{\alpha}, \frac{h\theta l_{00}^{\alpha} \rho_A x^{-\alpha}}{2\rho_B} \right) \mathrm{d}x \right] \right). \end{aligned} \tag{37}$$

The expectation operation in terms of the fading coefficient h can be simply solved for the first two terms. However, the third term needs further derivation, i.e.,

$$\int_0^\infty x \exp(-\pi\lambda_A x^2) \Gamma\left(-\frac{2}{\alpha}, \frac{h\theta l_{00}^{\alpha} \rho_A x^{-\alpha}}{2\rho_B} \right) \mathrm{d}x \tag{38}$$

$$= \frac{1}{2\pi\lambda_A} \int_0^\infty \exp(-\pi\lambda_A x^2) \mathrm{d}\Gamma\left(-\frac{2}{\alpha}, \frac{h\theta l_{00}^{\alpha} \rho_A x^{-\alpha}}{2\rho_B} \right) \tag{39}$$

$$\begin{aligned} &= \frac{\alpha}{4\pi\lambda_A} \left(\frac{h\theta l_{00}^{\alpha} \rho_A x^{-\alpha}}{2\rho_B} \right)^{-\frac{2}{\alpha}} \\ & \times \int_0^\infty \exp\left(-\pi\lambda_A t - \frac{h\theta l_{00}^{\alpha} \rho_A t^{-\alpha/2}}{2\rho_B} \right) \mathrm{d}t. \end{aligned} \tag{40}$$

Because the PDF of the fading coefficient is Chi-squared distributed with $2d_A$ DoF, the third term of (37) is translated using the following equations:

$$\begin{aligned} & \mathbb{E}_h \left[\frac{(2\pi\lambda_A)^2}{\alpha} \left(\frac{h\theta l_{00}^{\alpha} \rho_A}{2\rho_B} \right)^{2/\alpha} \frac{\alpha \left(\frac{h\theta l_{00}^{\alpha} \rho_A x^{-\alpha}}{2\rho_B} \right)^{-\frac{2}{\alpha}}}{4\pi\lambda_A} \right. \\ & \left. \times \int_0^\infty \exp\left(-\pi\lambda_A t - \frac{h\theta l_{00}^{\alpha} \rho_A t^{-\alpha/2}}{2\rho_B} \right) \mathrm{d}t \right] \end{aligned} \tag{41}$$

$$= \pi\lambda_A \int_0^\infty e^{-\pi\lambda_A t} \left(1 + \frac{\rho_A}{\rho_B} \theta l_{00}^{\alpha} t^{-\alpha/2} \right)^{-d_A} \mathrm{d}t. \tag{42}$$

Because the $\frac{2}{\alpha}$-th moment of h can be written as, $\mathbb{E}\left[h^{2/\alpha}\right] = \frac{2^{2/\alpha}\Gamma(d_A+2/\alpha)}{\Gamma(d_A)}$. Hence, the success probability can be given by,

$$p_s^A = \exp\left(1 + \frac{2\pi\lambda_A \left(\frac{\rho_A}{\rho_B}\theta l_{00}^{\alpha}\right)^{\frac{2}{\alpha}} \Gamma(d_A+\frac{2}{\alpha})\Gamma(-\frac{2}{\alpha})}{\alpha\Gamma(d_A)} - \pi\lambda_A \int_0^{\infty} e^{-\pi\lambda_A t}\left(1+\frac{\rho_A}{\rho_B}\theta l_{00}^{\alpha} t^{-\alpha/2}\right)^{-d_A} \mathrm{d}t\right). \quad (43)$$

Yi Luo (S11) received the B.Eng. degree in solid state electronics from the University of Electronic Science and Technology of China, Chengdu, China, in 2010 and PhD degree from the Institute of Digital Communication, The University of Edinburgh, Edinburgh, U.K, in 2015. His current research interests include stochastic geometry-based analysis for wireless communication and interference alignment. He is now a research fellow with school of electrical engineering, Tsinghua University, China.

Tharmalingam Ratnarajah [A96-M05-SM05] is currently with the Institute for Digital Communications, University of Edinburgh, Edinburgh, UK, as a Professor in Digital Communications and Signal Processing and the Head of Institute for Digital Communications. His research interests include signal processing and information theoretic aspects of 5G and beyond wireless networks, full-duplex radio, mmWave communications, random matrices theory, interference alignment, statistical and array signal processing and quantum information theory. He has published over 300 publications in these areas and holds four U.S. patents. He is currently the coordinator of the FP7 project ADEL (3.7M€) in the area of licensed shared access for 5G wireless networks. Previously, he was the coordinator of the FP7 project HARP (3.2M€) in the area of highly distributed MIMO and FP7 Future and Emerging Technologies projects HIATUS (2.7M€) in the area of interference alignment and CROWN (2.3M€) in the area of cognitive radio networks. Dr Ratnarajah is a Fellow of Higher Education Academy (FHEA), U.K., and an associate editor of the IEEE Transactions on Signal Processing.

Jiang Xue [M10] received the B.S. degree in Information and Computing Science from Xi'an Jiaotong University, Xi'an, China, in 2005, the M.S. degrees in Applied Mathematics from Lanzhou University, China and Uppsala University, Sweden, in 2008 and 2009, respectively. Dr. J. Xue received the Ph.D. degree in Electrical and Electronic Engineering from ECIT, the Queen's University of Belfast, U.K., in 2012. He is currently a Research Fellow with IDCoM, the University of Edinburgh, UK. His main interests include information theory, performance analysis of general multiple antenna systems, stochastic geometry, cooperative communications, and cognitive radio.

Faheem A. Khan [M02] received his Ph.D. degree in electrical and electronic engineering from Queens University Belfast, United Kingdom, in 2012. He is currently a lecturer with the School of Computing and Engineering, University of Huddersfield, UK. Previously, he worked as a research associate at the Institute for Digital Communications, University of Edinburgh, UK, where he contributed to research in EU research projects HARP and ADEL. His research interests lie in the field of wireless communications and signal processing with particular focus on cognitive radio, MIMO and millimeter wave communications. He has significant previous teaching and research experience at academic institutions in the UK, Middle East and India. He has authored and co-authored more than 20 papers in refereed journals and conferences.