IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 21, 2019, accepted April 9, 2019, date of publication April 18, 2019, date of current version April 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911409

System Capacity Analysis for Ultra-Dense Multi-Tier Future Cellular Networks

SYED WAQAS HAIDER SHAH[1], (Student Member, IEEE), ADNAN NOOR MIAN[1,2], (Member, IEEE), SHAHID MUMTAZ[3], (Senior Member, IEEE), AND JON CROWCROFT[2], (Fellow, IEEE)

[1]Department of Electrical Engineering, Information Technology University, Lahore 45000, Pakistan
[2]Computer Laboratory, University of Cambridge, Cambridge CB3 0FD, U.K.
[3]DETI, Instituto de Telecomunicações, University of Aveiro, 4554 Aveiro, Portugal

Corresponding author: Syed Waqas Haider Shah (waqas.haider@itu.edu.pk)

This work was supported by the University of Cambridge.

ABSTRACT Ultra-dense multi-tier cellular networks have recently drawn the attention of researchers due to their potential efficiency in dealing with high-data rate demands in upcoming 5G cellular networks. These networks consist of multi-tier base stations including micro base stations with very high-system capacity and short inter-site distances, overlooked by central macro base stations. In this way, network densification is achieved in the same area as that of traditional mobile networks, which offers much higher system capacity and bandwidth reuse. This paper utilizes a well-known analytical tool, stochastic geometry for modeling and analyzing interference in ultra-dense multi-tier cellular networks. Primarily, we have studied different factors affecting the system capacity including the network densification, cell load, and multi-tier interference. The role of the ergodic channel capacity is also discussed. Moreover, the effects of channel interference, system bandwidth, and the network densification on the spectral and energy efficiencies of the network are observed. Finally, the results show that the network densification and the cell load have a profound impact on system performance as well as spectral and energy efficiencies of the networks.

INDEX TERMS System capacity, ultra-dense multi-tier networks (UDMN), spectral efficiency, energy efficiency, stochastic geometry, 5G.

I. INTRODUCTION

The trend towards a digitizing world is gaining popularity in recent times as billions of new devices and users are being connected to the global Internet. In order to provide seamless connectivity to this massive number of new users and devices, cellular networks can be an appropriate solution. Advantages of cellular networks like mobility, roaming support, ubiquitous coverage, and reliable data delivery distinguish them from other wireless networks [1]. In the last decade, the cellular industry has emerged as one of the leading industries in providing seamless connectivity to various sectors of the society. The telecommunication sector has a mature ecosystem and it is governed by the 3rd Generation Partnership Project (3GPP), which ensures industry-academia partnership for future developments. The 3GPP has issued various releases based upon exponentially increasing demands of high data rate, low latency, and better quality-of-service (QoS). In its recent Release 14 and beyond for long term evolution (LTE) and LTE advanced (LTE-A), the 3GPP has also ensured low power communication for machine-type devices which enables it as ''one solution fits all'' [2]. Furthermore, to provide high data rate connectivity to the massive number of cellular users, 3GPP has announced fifth generation (5G) cellular network which will be commercially deployed as early as by 2020 [3]. One of the key enabling technologies for 5G is ultra-dense multi-tier networks (UDMN). In UDMN, multiple micro-cells (mCs) are deployed within a coverage area of a macro-cell (MC) to provide high data rate connectivity. This multi-tier network approach can provide access to a massive number of users as well as deliver significant capacity gains [4]. Generally, an MC operates in low-frequency bands (legacy LTE frequency bands of 1800 MHZ & 2100 MHz) and mCs are

The associate editor coordinating the review of this manuscript and approving it for publication was Fuhui Zhou.

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

assigned high-frequency bands (above 6GHz). A cellular user always keeps track of both the mC and MC for high data rate connectivity and to ensure high-speed mobility, respectively. UDMN is one of the most efficient ways of providing high data rate connectivity to the exponentially increasing number of devices which are expected to reach 50 billion by 2020 [5].

The channel capacity, being one of the key performance indicators (KPIs) of a wireless communication system, has been analyzed previously in many different scenarios [6]–[9]. Sharif and Hassibi [6] presented the transmission capacity for multiple-antenna broadcast channels in which the authors proposed a novel scheme that constructs multiple random beams and those beams transmit information to the users with the highest signal-to-interference-plus-noise-ratio (SINR). The authors assumed that channel state information (CSI) is available at the transmitter with very little feedback. With the emergence of relay-assisted networks, the traditional concept of transmission capacity has been changed. In a relay network, the overall end-to-end transmission capacity is equal to the transmission capacity of the link which has the minimum link capacity among all the links. The authors in [7] derived analytical expressions of the ergodic capacity and maximum achievable throughput of a decode-and-forward (DF) relaying network. Moreover, the authors have used the energy harvesting technique for powering up the relay node which can significantly increase the energy efficiency of the network. The upcoming 5G cellular networks should support ultra-reliability, massive connectivity, and very high system capacity in order to provide connectivity to the diverse nature of future applications. The authors in [9] presented a joint scheduling scheme for ultra-reliable and low-latency communication (URLLC) to achieve high data rate connectivity. Theoretical QoS guarantees and upper bound on delay probability for URLLC traffic is also presented.

On the other hand, realistic network modeling is very essential to calculate the effective network capacity. For network modeling, researchers are more inclined towards stochastic geometry approach as compared to other network models such as grid-based models. In stochastic geometry, cell deployment is modeled using poison point process (PPP). Cell deployment modeling in any network is of utmost importance as it defines various network properties including user association, mobility, traffic patterns, and most importantly the network capacity. Stochastic geometry provides more realistic network models and their properties such as coverage/outage probability, SINR etc [10]. The authors in [8] presented transmission capacity analysis for device-to-device (D2D) communication. Various D2D communication modes (underlay/overlay) has been studied with and without relay assistance. The authors have used stochastic geometry tool to identify the success and outage probability as well. They have proved that D2D transmission capacity can be enhanced using relay transmission. There are various other studies as well [11], [12] which explain this KPI for various wireless networks such as Ad-hoc networks and cloud-empowered heterogeneous networks, respectively. *However, to the best of authors' knowledge, there does not exist any work related to the capacity and outage probability analysis for UDMN in 5G cellular networks.*

TABLE 1. Mathematical notations.

Notation	Description
γ_u	SINR experienced by the user
γ_u^i	SINR of user u when connected to the cell i
d	Network densification factor
$\hat{R}_u$	Desired rate of a user u
ζ_u^i	Minimum amount of resources assigned to a user u in a cell i
ζ_i	Cell load of a cell i
h_x	Fast fading coefficients at point x
$l(\parallel x - y \parallel)$	Path-loss function
γ_{req}	Desired threshold (function of noise and interference)
ϕ	Intensity of nodes in a specific region
λ	Density of PPP
$\Lambda(A)$	Intensity measure in region 'A'
α	Path-loss exponent
I	Interference caused by neighboring BS
$P_c(\gamma_{req})$	Transmission success probability
C_{ecc}^i	Ergodic channel capacity of the fading channel
ϕ^*	PPP of the distances
η_s	Spectral efficiency of the network
η_E	Energy efficiency of the network

In this work, we utilize a well known analytical tool, stochastic geometry for modeling and analyzing interference in UDMN. Primarily, we have studied different factors affecting the system capacity including network densification, cell load, and multi-tier interference. The role of the ergodic channel capacity is also discussed. Furthermore, the effects of channel interference, system bandwidth, and network densification on the spectral and energy efficiencies are observed. Finally, the results show that the network densification and the cell load have a profound impact on system performance as well as spectral and energy efficiencies.

The rest of the paper is organized as follows. Section II introduces the system model and highlights various features which we have considered in this paper. Section III provides the channel capacity and its underlying factors with interference modeling. More specifically the stochastic geometry has been used in this section for interference modeling. Section IV provides the ergodic channel capacity as well as the spectral and energy efficiencies. Section V provides the numerical results followed by some discussions. Finally, Section VI concludes the paper. For the readers' facilitation, TABLE 1 shows all the mathematical notations used in this paper for convenient referencing.

II. SYSTEM MODEL

In this work, we consider a UDMN consisting a MC of radius D with multiple mCs each with a radius R as shown in FIGURE 1. MC is operating on lower frequency channels with higher bandwidth to provide low rate connectivity to a large number of cellular users within a large coverage area, whereas mCs are operating on higher frequency bands with smaller bandwidth as compared to MC in the small

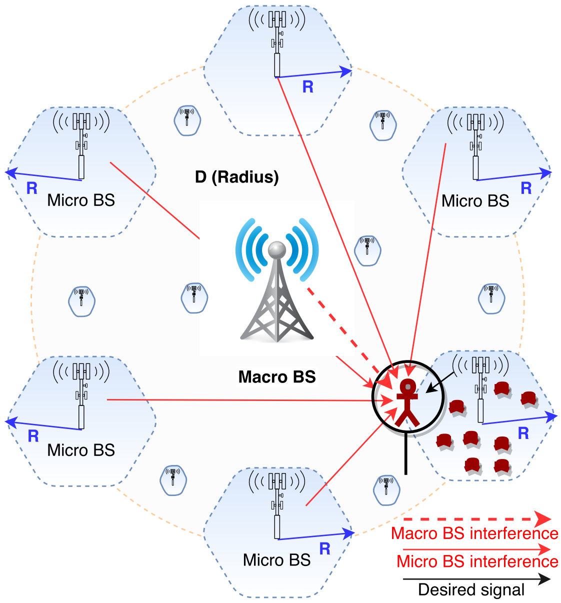


FIGURE 1. **The system model: a possible architecture for ultra-dense multi-tier cellular network. The dotted red arrow represent the interference caused by MC, while the solid red arrows represent the interference caused by mCs. The black arrow represent the desired signal. The MC has a large coverage area of radius *D*, while the mCs have a radius *R*.**

coverage area. Cellular users connected to mCs can achieve higher data rate. We assume that no mCs have overlapping regions, which means that in order to perform handovers (HO), a mobile user needs to connect to the MC. The users are spread across the coverage area in a random fashion and they can experience interference from both the MC and mCs. The transmission channel is considered to be a Rayleigh fading channel. Moreover, we assume that CSI at the transmitter (CSIT) is not available, thus; the transmitter schedules the source data at a constant rate.

III. CHANNEL CAPACITY AND UNDERLYING FACTORS

According to Shannon channel capacity, maximum achievable capacity of a channel in traditional cellular network when connected to a cell i is,

$$C^{i}_{channel} = d(\frac{B}{\zeta_i})\log_2(1+\frac{P_t}{N_o+I_u}) \tag{1}$$

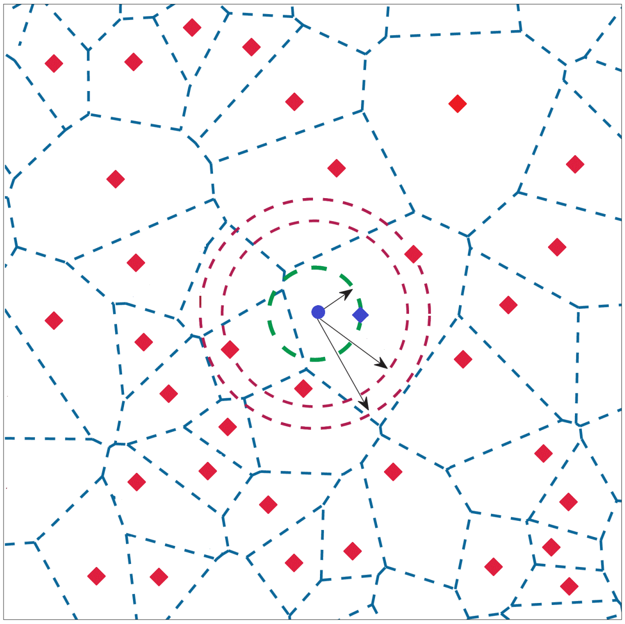
where P_t is the transmission power of the transmitter, I_u is the average power of the interfering base stations (BS), N_o is the thermal noise power, B is the available channel bandwidth, ζ_i is the cell load of a cell i, and d is the network densification factor. In order to achieve higher channel capacity and accommodate exponentially increasing number of users, there is a need to either look for alternate frequency spectrums or to employ the network densification by deploying the massive numbers of smaller BS. Although, the network densification can bring in many benefits including the massive number of users accommodation with the higher achievable data rate, with the massive deployment of smaller BS in a network, interference among users as well as between BS would increase drastically and would deteriorate the channel capacity. Hence, the SINR associated with each user must be calculated beforehand, which can be written as $\frac{P_t}{N_o+I_u}$.

A. NETWORK DENSIFICATION

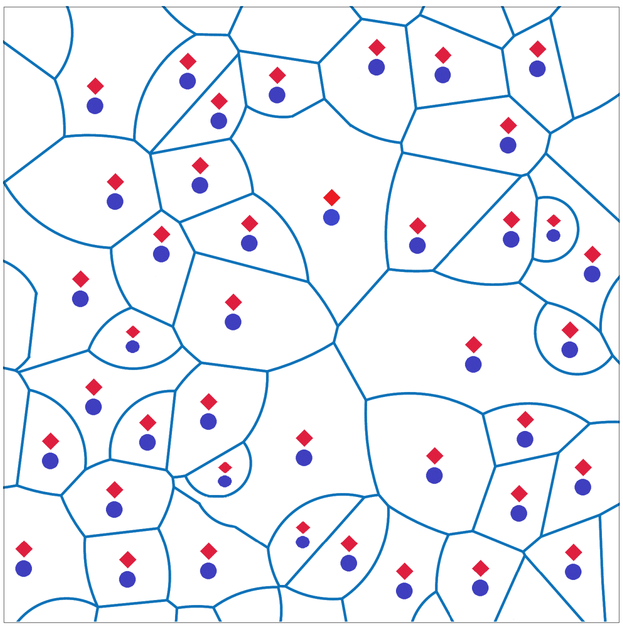
System capacity must be enhanced for efficient management of exponentially growing data traffic in 5G cellular networks. The term network densification covers all the aspects which are aimed at enhancing the system capacity for 5G [13]. Multi-dimensional solutions can be incorporated in the 5G cellular networks to commensurate the system capacity with the demand. These solutions include techniques related to physical layer enhancement, which can increase the network capacity by 3 to 5 times, such as coordinated multipoint transmission (CoMP) [14], Massive MIMO [14], etc. Incorporating new spectrum to enhance the bandwidth such as mm-Wave transmission [15], can lead to 10× better network capacity. But most importantly, spatial densification i.e. increasing the number of small cells in the same area alone can account for a minimum of 40× increase in the capacity gain of the system [16], especially in the traffic congested areas. It also poses several challenges including, increase in system energy consumption and a higher number of HO events. Therefore, densification must be accompanied by an efficient network management strategy such as self-organizing networks (SON), which can dynamically increase or decrease the number of active cells according to the traffic congestion [17]. As evident from (1), densification serves the purpose of enhancing system capacity well by lowering the cell load (ζ), while enhancing the transmission power (P_t).

1) CELL LOAD

As already mentioned, network densification leads to a reduction in the cell load factor i.e. the number of active users associated with the BS, which in turn increases the overall network capacity. Therefore, an efficient network management scheme must take into account the reduction of the cell load, while also meeting the challenges associated with the densification such as HO and blocking probabilities. One of the efficient network management strategies is to introduce smart mCs (cloud-cells), which may turn on or off according to the user traffic. The MC decides the activation of these cloud-cells depending upon certain predefined parameters such as number of active users, user throughput and delay demands, user priority, and system performance level. The MC can also offload existing active cloud-cells with newly activated mCs in the given coverage area [18]. A possible mCs deployment in a cellular network using PPP is shown in FIGURE 2(a). The service area or the footprint of each mC is geometrically represented by using Voronoi-tessellation [19]. Generally, in order to reduce the cell load, the multi-tier deployment of cells is more efficient. FIGURE 2(b) shows a multi-tier cell deployment in a cellular network where both the tiers are mutually independent and the BS location also follows the independent PPP model. In this way, not only the



(a)



(b)

FIGURE 2. (a) The blue diamond represents the desired mC and the red diamonds represents interfering mCs. Blue dotted lines represents the mCs' footprints. Green dotted line shows the area where no horizontal handoff occurs and red dotted circles shows the area of maximum interference. (b) The mCs and MCs are deployed on the same location using independent PPP. Red diamonds and blue circles represents the BS of the mCs and MCs, respectively. (a) The mCs deployment in a cellular network using PPP. (b) The ultra-dense multi-tier cellular network deployment using PPP.

network load is balanced which ultimately leads to higher system capacity, but also lesser power is consumed along with lower HO probability due to dynamic (rather than static) operation of cloud-cells. This is specifically true for places where the probability of user traffic is diverse and time-dependent, such as shopping malls and offices. Moreover, an interesting factor regarding the effect of network densification on the transmission capacity is that it is not always a monotonic function. It is due to the fact that with an increase in the number of mCs in a coverage area of an MC (increase in mCs' density), the transmission capacity does not always increase as shown in the FIGURE 3. The transmission capacity increases in the beginning as mCs density increases because more mCs can bring capacity enhancement to the entire system. Whereas, if the mCs density increases continuously, the interference caused by the neighboring mCs will become more critical. This interference will ultimately lead to a reduction in the users' transmission capacity on the whole.

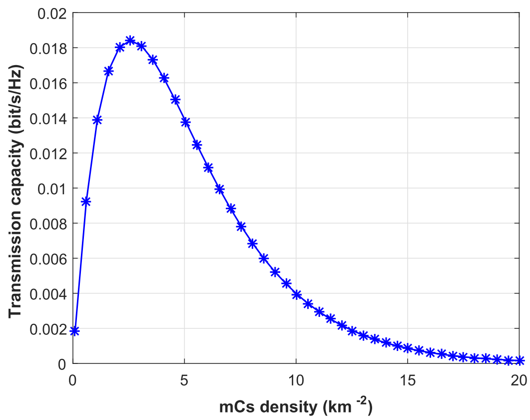


FIGURE 3. User transmission capacity versus the micro cells (mCs) density.

The cell load can also be reduced using adaptive user association, in which the mCs with higher load factor can be offloaded by shifting some of their active users (or incoming users) to adjacent mCs, which although provide comparatively lower SINR than the serving mC, but are comparatively lightly loaded and can, therefore, boost up the users and network capacities as unused resources are put to use [20]. Maximum capacity a user can actually achieve is the instantaneous rate multiplied by the allowed fraction of resources (PRB). We as cellular users must have experienced in our daily life a considerable drop in rate (throughput) at peak hours or in crowded public places regardless of the signal quality (SINR). This is because of the high number of users associated with that cell (generally known as the cell load). A saturated cell (fully or 80-90% loaded) can provide less throughput to a user as compared to a less or partially loaded cell. In order to define the cell load, first, we need to calculate the minimum amount of resources assigned to a user,

$$\zeta_u^i = \frac{1}{B_r} \frac{\hat{R}_u}{f(\gamma_u^i)} \tag{2}$$

where B_r represents the bandwidth associated to one physical resource block (PRB), $\hat{R}_u$ is the desired rate of a user u and γ_u^i is the SINR of user u when connected to the cell i. $f(\gamma_u^i)$ defines the spectral efficiency for a given SINR, thus $f(\gamma_u^i) = \log_2(1+\gamma_u^i)$ [1]. Now taking into account resource allocated to a user to find the total cell load [21].

$$\zeta_i = \frac{1}{B_i}(\frac{1}{B_r}\sum_{U_i}\frac{\hat{R_u}}{\log_2(1+\gamma_u^i)}) \tag{3}$$

where B_i is total bandwidth allocated to cell i and U_i represents a set of all users (active) connected with the cell i. Previously, achievable capacity in a cell is considered to be only as a function of distance or location. As the effect of the cell load on the throughout is clear now, ζ_i will define actual achievable capacity by a cell i on a specific location z as a function of location as well as the cell load [22].

$$C_i(z,\zeta) = \min\{B_i.\log_2(1+\gamma_i(z,\zeta)), c_{max}\} \tag{4}$$

where c_{max} is the maximum channel capacity achievable when CSI is available at the transmitting end. The SINR experienced at location z by a user u connected to the cell i is,

$$\gamma_i(z,\zeta) = \frac{P_i(z)}{\sum_{j\neq i}(\zeta_j.P_j(z)+N_o)} \tag{5}$$

where ζ_j defines the neighboring cell load. It is worth noting that SINR in one cell also depends on the resources consumed in neighboring cells.

B. INTERFERENCE MODELLING

The cumulated interference experienced by a user at any point $z\epsilon R^d$ in the network is

$$I(z) = \sum_{y\epsilon\tau} P_y h_y l(\parallel z-y \parallel) \tag{6}$$

where $I_y = T_y.f(d).h$, d is pathloss, and $\tau\epsilon R^d$ represents set of all transmitting nodes. h_y and $l(\parallel z-y \parallel)$ are random processes representing fast fading coefficients and pathloss function, respectively. It is assumed that the pathloss function l depends on difference of distances only. Generally, l is modeled as exponential ($l \parallel z-y \parallel = c_o e^{-\gamma\parallel z-y\parallel}$) or power law distributed ($l \parallel z-y \parallel = c_o \parallel z-y \parallel^{-a}$) [23]. Fast fading in this case is the consequence of shadowing. We can write transmission success probability, also known as coverage probability, as,

$$\begin{aligned} P_c(\gamma_{\text{req}}) &= P(\gamma_u > \gamma_{\text{req}}) = P(\gamma_u > \gamma_{\text{req}}(N_o + I_u)) \\ P_c(\gamma_{\text{req}}) &= P(\gamma_{\text{req}}N_o).P(\gamma_{\text{req}}I_u) \end{aligned} \tag{7}$$

where γ_{req} is any desired threshold and is a function of noise and interference powers. It is shown from (7) that coverage probability is the product of two independent factors, noise and interference. This allows us to find the probability distributions of noise and interference independently in order to find the coverage probability. Considering the fact that the spectrum sharing systems are not noise-limited but only interference-limited, one could ignore the effect of thermal noise. Therefore, coverage probability is only a factor of interference. Now in order to find the probability distribution of interference, first we need to find the user distribution in the cell. The spread of users in a cell can be modeled using PPP as it ensures independence among users' existence. Now we present Lemma 1 which gives us the probability of k^{th} user present outside the coverage area of a cell.

Lemma 1: Probability that k^{th} user is outside the coverage region of a cell (circular) having area $A = \pi R^2$ where R is radius of the circle can be calculated using PPP with the following simplified result,

$$P_{R_k}(R) = R^{2k-1}(\lambda\pi)^k \frac{2}{\Gamma(k)} e^{(\lambda\pi R^2)}$$

Proof: Proof is shown in Appendix A.

Lemma 1 also provides the total number of users present in a coverage area of a cell. Now in order to find the intensity measure, intensity of users present in a specific region A can be calculated using Lemma 2.

Lemma 2: (Intensity Measure). Intensity measure of users present in a coverage region having area A can be represented as the expectation of the counting measure of the users present in that area,

$$\Lambda(A) = \mathbb{E}[\phi(A)]$$

Proof: Proof is shown in Appendix B.

As calculations in two or higher dimensions grow complex, the solution is to map the model (circle) into one dimension (real line with varying density). More specifically, let f be a function for mapping $R^2 \rightarrow R$ and $\phi = \{X_1, X_2, X_3, \ldots, X_{k-1}\}$ then $\phi^* = \{f(x_1), f(x_2), f(x_3), \ldots, f(x_{k-1})\}$, where ϕ^* is a poison process with $\Lambda^*(A) = \Lambda(f^{-1}(A))$. In our case of distances, ϕ is a stationary PPP with intensity λ and $f(x) = ||x||$. For $A = [0, r], f^{-1}(A) = (0, R)$, the cell of radius R at origin has a mean measure given as $\Lambda^*(A) = \Lambda(a(0,R)) = \lambda\pi R^2$. Since the dimensions are reduced, the intensity would change, which is measured by taking derivative of $\Lambda^*(A)$ w.r.t R which gives us $\lambda^*(R) = 2\lambda\pi R$. Where $R \geq 0$. It is evident that the distance of the points of a PPP, that is homogeneous on the plane form a non-homogeneous PPP on R^* with linearly increasing density. However, when the squared distances are taken as $\phi^* = \{||x_1||^2, ||x_2||^2, ||x_3||^2 \ldots, ||x_{k-1}||^2\}$, they again form homogeneous PPP with intensity $\lambda^* = \lambda\pi$.

Furthermore, the interference experienced by a user at the center of the cell, provided that all nodes transmit at unit power and pathloss ($l = R^{-a}$) can be written as $I_u^o = \sum_{x\epsilon\phi} 1h_x||x||^{-a}$. The transmitting nodes form a stationary PPP ϕ of intensity λ in R^2. Due to this stationarity, interference is considered same everywhere (across the cell) and therefore, can be written as $I_u = \sum_{R\epsilon\phi^*} h_R R^{-a}$. Where ϕ^* in this case is $\{||x_1||, ||x_2||, ||x_3|| \ldots, ||x_{k-1}||\}\epsilon R^+$ the PPP of the distances. Therefore, mean of interference I_u is given as $\mathbb{E}[I_u] = \mathbb{E}[\sum_{R\epsilon\phi^*} h_R R^{-a}]$. Moreover, since the pathloss and shadowing are considered two independent random variables so, they can be dealt separately $\mathbb{E}[I_u] = \mathbb{E}[\sum_{R\epsilon\phi^*} h_R].\mathbb{E}[\sum_{R\epsilon\phi^*} R^{-a}]$. Since $\mathbb{E}[h] = 1$,

[1]Only possible when various gains such as channel coding, pre-coding, scheduling and MIMO scheme gains are considered. Moreover, $f(\gamma_u^i) = X\log_2(1+Y\gamma_u^i)$, owing to simplicity we consider X = Y = 1

then $\mathbb{E}[I_u] = \mathbb{E}[\sum_{R\epsilon\phi^*} R^{-a}]$. Now assuming f to be a non-negative function, the Campbell's theorem for general PPP states that,

$$\mathbb{E}[I_u] = \int_{R^+} R^{-a}\lambda 2\pi R.dR = \frac{\lambda 2\pi}{2-a}.R^{2-a}|_0^\alpha \tag{8}$$

However, (8) does not provide convergence for all values of α!, therefore $\mathbb{E}[I_u] = \alpha$ for all stationary PPP. In terms of α, the following two cases arise:
$\alpha \leq 2$: In this case, the upper bound of the integration causes a problem, because there will be too much interference from all the far nodes.
$\alpha \geq 2$: In this case, the lower bound causes a problem, since the nodes near the origin make $\mathbb{E}[I_u]$ diverge because R^{-a} grows very fast as R is reduced if $\alpha > 2$. This problem can be solved by introducing bounded pathloss model for $\alpha > 2$, then $\mathbb{E}[I_u]$ remains finite, provided that no node moves to the origin, which implies that lower bound be change to $\rho > 0$ and therefore, $\mathbb{E}[I_u] = \frac{\lambda 2\pi}{\alpha-2}\rho^{2-\alpha}|_{\rho>0}^{\alpha}$.

Since, averages alone can be misleading we also find the probability distribution of the interference is required. Furthermore, the ultimate goal in network provider perspective is to achieve more capacity and less interference. In order to achieve higher capacity, network densification is required whereas, sparse networks should be used for less interference. This leads to a tradeoff in which such a network should be designed which can effectively serve both the requirements. To address this issue, we need to take our analysis beyond the mean and to find the distribution of the interference. To make our analysis simpler, we take leverage from Laplace transform; thus, it turns out to be

$$L_{I_u}(s) = \exp(-\lambda\pi\ \mathbb{E}(h^\delta)\Gamma(1-\delta)s^\delta) \tag{9}$$

Here, the interference is mapped using a stable exponential distribution where $\lambda\pi\ \mathbb{E}(h^\delta)\Gamma(1-\delta)$ is the dispersion factor and δ is defined as characteristic exponent with $0 < \delta < 1$ bounds. If upper-bound of δ increases from 1 or in other words, value of α is taken less than 2, then we have $L_{I_u}(s) < 0$ for all $s > 0$ and I_u approaches to infinity ($I_u \rightarrow \infty$). So, in order to have finite interference we must set lower limit of $\alpha > 2$. Closed-form expression for probability distribution of the interference can be written as follows,

$$\begin{aligned} L_{I_u}(s) &= \exp(-\lambda\sqrt{s}\frac{\pi}{2}^2) \\ f_{I_u}(x) &= \frac{\pi}{2}\lambda e^{\frac{-\pi^3\lambda^2}{4x}} x^{\frac{-3}{2}} \end{aligned} \tag{10}$$

I_u is the product of two random variables having different probability distributions, shadowing (poison distribution) and fast fading (Rayleigh distribution), so the corresponding distribution is Lèvy distribution [2].

[2]It is from the family of stable distribution and also considered as a special case of the inverse gamma distribution.

IV. ERGODIC CHANNEL CAPACITY

According to the channel capacity given in (1), the cell load given in (3), and the distribution of interference given in (10), we can have

$$C^i_{channel} = d(\frac{B}{\zeta_i})\log_2(1+\frac{P_t}{\frac{\pi}{2}\lambda e^{\frac{-\pi^3\lambda^2}{4x}} x^{\frac{-3}{2}}}) \tag{11}$$

Since, CSIT is not available, the transmitted data will be deteriorated because of the channel fading and the effective channel capacity will be significantly reduced. In this case, ergodic capacity will be a good measure as it is the expected value of the instantaneous channel capacity. The ergodic capacity of a fading channel associated with a cell i (C^i_{ecc}) for an average transmit power $\bar{P}_t$ with no CSIT is given by,

$$\begin{aligned} C^i_{ecc} &= \mathbb{E}[d(\frac{B_i}{\zeta_i})\log_2(1+\gamma_u)] \\ &= \int_0^\infty d(\frac{B_i}{\zeta_i})\log_2(1+\gamma_u)p(\gamma_u)d\gamma \end{aligned} \tag{12}$$

According to Jensons' inequality [24] [3], the condition on the ergodic channel capacity (C^i_{ecc}) is,

$$C^i_{ecc} \leq d(\frac{B_i}{\zeta_i})\log_2(1+\frac{\bar{P}_t}{N+\frac{\lambda 2\pi}{\alpha-2}\rho^{2-\alpha}}) \tag{13}$$

We can conclude from here that the ergodic channel capacity will increase linearly by densification of the network (increasing d) or by allocating more bandwidth to the cell. Whereas, increase in the cell load (more number of active users) can reduce the ergodic channel capacity associated to a user. One could further compute the spectral efficiency and the energy efficiency according to the following subsections.

A. SPECTRAL EFFICIENCY

The link spectral efficiency is a measure of how well the bandwidth resources are exploited in a communication system. As mentioned previously, the spectral efficiency is measured in bits/s/Hz, which distributes the total achievable throughput over the available bandwidth and breaks down these parameters to lowest resolution i.e. what maximum throughput (bits/s) is supported by each hertz of the available bandwidth. The spectral efficiency has a direct relationship with the SINR as shown below,

$$\begin{aligned} \eta_s &= (\frac{d}{\zeta_i})\log_2(1+\frac{\gamma_u}{B}) \\ &= (\frac{d}{\zeta_i})\log_2(1+\frac{P_t}{\frac{\pi}{2}\lambda e^{\frac{-\pi^3\lambda^2}{4x}} x^{\frac{-3}{2}} B}) \end{aligned} \tag{14}$$

In fact, delving further into the case reveals that increasing SINR also allows the system to use higher modulation schemes (such as 16-QAM, 64-QAM), which reciprocates to increase in the spectral efficiency by allowing to send more number of bits per time using the same bandwidth resource.

[3]In general, convex transformation of a mean is always less than or equal to the mean of a convex transformation $\mathbb{E}[B\log_2(1+\gamma_u)] \leq B\log_2(1+\mathbb{E}[\gamma_u])$

On the other hand, when the bandwidth is increased, that efficiency tends to decrease naturally, which is why it has an inverse relation with the spectral efficiency. Another way to increase the spectral efficiency would be to increase the network densification, resulting in the reuse of same bandwidth manyfold.

B. ENERGY EFFICIENCY

The energy efficiency measures the cost of each transmission bit sent over the link, in terms of consumed energy in joules usually defined as *bits/joule*. On the network level, it can be measured as the total amount of throughput divided by the network power consumption. In wireless communication and especially in upcoming 5G networks, the energy efficiency is an important metric due to energy-constraint nature of the devices, such as wireless sensor nodes and machine-type communication devices. Apart from this, the researchers are also working on reducing power consumption and providing greener 5G solutions at network level due to a global increase in CO_2 footprint by employing smarter networks such as SON, as discussed previously. 80% of the total energy consumption in cellular communication networks is hogged by BS operations [25]. This can also be concluded by looking at the energy efficiency relation as shown below,

$$\eta_E = \frac{\text{area spectral efficiency}}{\text{average network power consumption}}$$

where area spectral efficiency is $\eta_s B$ and the average network power consumption includes the BS transmit power (P_T) and the power consumed in other functions of the BS (P_{BS}).

$$\eta_E = \frac{\eta_s B}{\frac{P_T}{\rho} + P_{BS}} \tag{15}$$

by putting the value of spectral efficiency (η_s) from (14), the energy efficiency at the network level would be,

$$\eta_E = \frac{(\frac{d}{\zeta_i}) B \log_2\left(1 + \frac{P_t}{\frac{\pi}{2}\lambda e^{\frac{-\pi^3\lambda^2}{4x}} x^{\frac{-3}{2}} B}\right)}{\frac{P_T}{\rho} + P_{BS}} \tag{16}$$

The energy efficiency has an outright inverse relation with the BS transmit power and power consumed in other BS functions. While the discussion is focused on UDMN, the role of the network densification on the energy efficiency should also be questioned. It has been established that the network densification leads to better spectral efficiency, which implies higher achievable bit rate for the given bandwidth and energy and therefore the causality between the energy efficiency and the spectral efficiency. In fact, this can also be corroborated by the fact that increased SINR in a link indicates better channel conditions and lower interference, leading to better achievable bit rate (bits/s) at lower consumed energy.

V. NUMERICAL ANALYSIS

In this section, we first describe our simulation setup, and then present the simulation results related to ergodic channel capacity, spectral efficiency, and energy efficiency. These simulation results will highlight the effect of network densification, cell load, and SINR on the ergodic channel capacity as well as the spectral and energy efficiencies of the network.

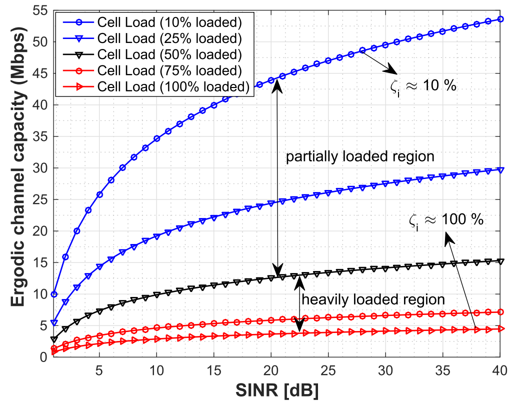


FIGURE 4. Ergodic channel capacity vs SINR for different cell loads.

A. SIMULATION SETUP

The simulations was done in MATLAB. We consider an MC of radius 1000 m with multiple number of mCs in its coverage area (placed randomly using PPP model) each with a radius of 100 m. We randomly place cellular users inside the coverage region of the MC. Reason for choosing PPP model is because of its advantages such as analytical tractability and ensuring maximum entropy. It is also more realistic tool for user and cell deployment as compared to traditional grid-based network models. Channel bandwidths of 100 kHz and 60 kHz are associated with MC and each mC, respectively. We used the following pathloss models for transmission: $PL(distance) = 128.1 + 37.6\ log_{10}(distance)$ and $PL(distance) = 140.7 + 36.7\ log_{10}(distance)$ for MC and mCs, respectively [26]. Average transmit power for MC is set to be 47 dBm and for mCs is 30 dBm. Range of γ_u is set between 0-40 dB.

B. SIMULATION RESULTS

In FIGURE 4, we plot the ergodic channel capacity as a function of SINR. It is evident from the figure that as SINR of the channel increases, the ergodic channel capacity will also increase. One could also see the effect of the cell load on service capacity as well; the ergodic channel capacity for the high cell load (large number of active users associated with BS) is lower when compared to the low cell load. Additionally, when BS is operating at 100% cell load, then the ergodic channel capacity will not change with increase in SINR. Moreover, change in the ergodic channel capacity in partially loaded region is very high when compared to the change in ergodic channel capacity in heavily loaded region.

FIGURE 5 also plots the ergodic channel capacity as a function of SINR but with the additional effect of the network densification. One could see that the ergodic channel capacity for a denser network is larger than a sparser network. Densification allows more number of mCs deployed in the coverage

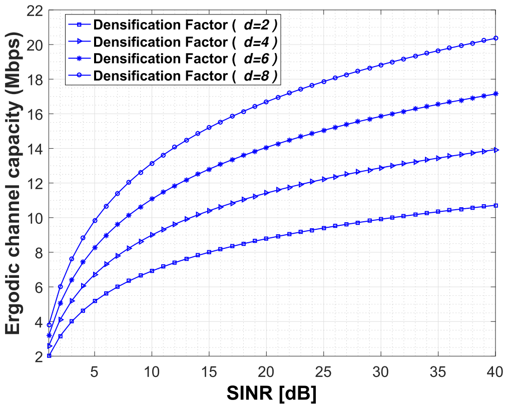


FIGURE 5. Ergodic channel capacity vs SINR and the effect of densification.

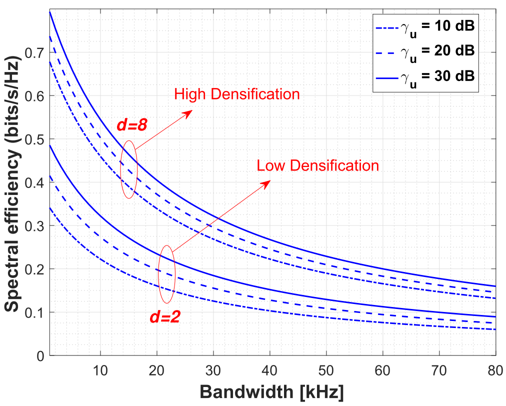


FIGURE 6. Spectral efficiency vs network bandwidth: for low (sparse) and high denser networks.

region of one MC which leads to more frequency reuse; thus, allows users to achieve higher transmission capacity.

In FIGURE 6, we plot the spectral efficiency as a function of network bandwidth. It is evident that as the bandwidth of the network increases, spectral efficiency of the system reduces. The rationale for this inverse relationship lies in the fact that more bandwidth accommodates the same number of bits transmitted per second. It is also evident from the curves that with an increase in the channel SINR, better spectral efficiency can be achieved for a given bandwidth. The curves have been drawn for three different cases of channel SINR. Finally, higher network densification also yields more efficient use of the given spectrum, since it can accommodate higher number of mCs within the same coverage region.

We can further extend the discussion to the effect of channel SINR on the energy efficiency of the network. As shown in FIGURE 7, the energy efficiency (bits/joule) of the network increases with channel SINR. This is because the increase in SINR is linked to a decrease in the interference experienced by the user, which therefore allows the network to transmit the same number of bits with lower energy expended per bit. The effect of densification is also closely linked to the energy efficiency: a higher density of mCs

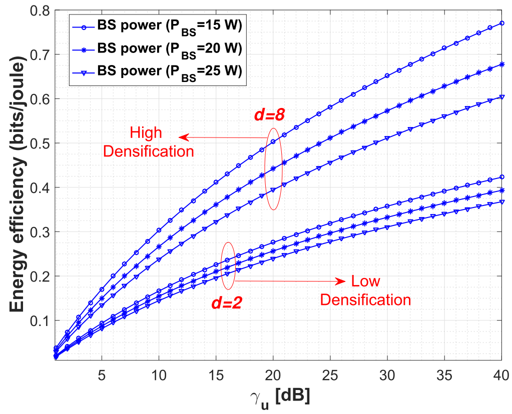


FIGURE 7. Energy efficiency vs SINR: for low (sparse) and high denser networks.

increases the system capacity and enables a higher number of transmitted bits at a given transmission power.

In fact, the transmission power also plays an important role in the energy efficiency; this can be observed in the curves that higher transmission power intuitively reduces the energy efficiency of the system. The curves have been drawn for three different levels of BS transmission power for both cases of densification. It is also visible that lower transmission power has a significant effect on the energy efficiency when the system has a higher density of mCs for a given SINR. This is again linked to higher capacity of the system in denser networks, which can accommodate more bits and therefore, reduce energy consumption per bit.

VI. CONCLUSION AND FUTURE RESEARCH

This paper is a discourse on the enhancement of system capacity in ultra-dense multi-tier cellular networks. These networks are the future of cellular communication due to their potential high data rates and higher spectral efficiency when compared to traditional networks. They allow the network densification through the use of a higher number of small cells and therefore, increasing the resource reuse. We have discussed several factors which define and affect the system capacity, most importantly the network densification and its underlying optimization techniques. We have also discussed the role of interference in restricting the system capacity by modeling it as both stationary and non-stationary PPP. The effective channel capacity is being modeled as the ergodic channel capacity as well as spectral and energy efficiencies are also computed. The results show the heavy dependence of the ergodic channel capacity and spectral/energy efficiencies on channel SINR, network densification, and cell load.

As a future research direction, this work can be extended to analyze the system capacity for ultra-dense multi-tier cellular networks using the technique of cooperative relaying. This technique can provide enhanced energy efficiency and throughput specifically in D2D communication. Another direction could be the analysis of system capacity for joint multi-user beamforming (JMB) using MegaMIMO. Such an analysis can help gain insights in not only reducing the neighboring cells' interference but also provide transmission capacity enhancement.

APPENDIX A
PROOF OF LEMMA 1

Using poison distribution probability that k number of users are present in a coverage area A is given by,

$$P(\phi(A)=k)=e^{(-\lambda|A|)}.\frac{(\lambda|A|)^k}{k!}$$

where $|.|$ is the coverage area, $\phi(A)$ is the counting measure and λ is the density of PPP. For simplicity, the coverage area of a cell is considered to be a circle, thus $A=\pi R^2$

$$P(\phi(\pi.R^2)=k)=e^{(-\lambda\pi R^2)}.\frac{(\lambda\pi R^2)^k}{k!}$$

as mCs have no overlapping coverage area so, $\phi(A)$ for any $A\epsilon R^2$ is an independent random variable. Now in order to find all users in the coverage area, we need to take sum of $k-1$ users,

$$\begin{aligned}P(\phi(\pi.R^2)) &= \sum_{m=0}^{k-1}[e^{(-\lambda\pi R^2)}\frac{(\lambda\pi R^2)^m}{m!}]\\ &= e^{(-\lambda\pi R^2)}\sum_{m=0}^{k-1}[\frac{(\lambda\pi R^2)^m}{m!}]\end{aligned}$$

Since this is assumed that there are $k-1$ users residing inside the coverage region of an mC, it means that a k^{th} user will reside outside the region A and that might be connected to the MC. A complementary cumulative distribution function (CCDF) is given above, now in order to find the probability that k^{th} user reside outside the region A or in other words probability that $k-1$ users resides inside the region we need to take derivative with respect to the radius of the circle.

$$\begin{aligned}P_{R_k}(R) &= \frac{-d}{dr}(P(\phi(\pi.R^2)>R))\\ &= \frac{-d}{dr}(e^{(-\lambda\pi R^2)}\sum_{m=0}^{k-1}[\frac{(\lambda\pi R^2)^m}{m!}])\end{aligned}$$

for k is a positive real number $k\epsilon R$ and $R\geq 0$, it simplifies to the following result,

$$P_{R_k}(R)=R^{2k-1}(\lambda\pi)^k\frac{2e^{(\lambda\pi R^2)}}{(k-1)!}$$

where $(k-1)!$ can be written as $\Gamma(k)$. ■

APPENDIX B
PROOF OF LEMMA 2

For a PPP $\phi=\{X_1,X_2\ldots X_{k-1}\}$ the number of users present in a specific coverage region which has an area of A and $A\subset R^2$ can be written as [27],

$$\phi(A)=|\phi\cap A|=\sum_{x\epsilon\phi}1(x\epsilon A)$$

and the intensity measure Λ is the expected number of users available in the coverage region A.

Case.1 (Stationary/Homogenous PPP): When user distribution is stationary and or homogenous PPP, intensity is location independent,

$$\Lambda(A)=\lambda|A|$$

Case.2 (Non-stationary/Non-homogenous PPP): When user distribution is non-stationary and or non-homogenous PPP, intensity is location dependent thus intensity measure becomes the integration of intensity function $\lambda(x)$ over the coverage area A,

$$\Lambda(A)=\int_A\lambda(x)dx<\alpha$$

■

REFERENCES

[1] X. Ge, S. Tu, G. Mao, and C. X. Wang, "5G ultra-dense cellular networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 1, pp. 72–79, Feb. 2016.
[2] J. Lee *et al.*, "LTE-advanced in 3GPP rel-13/14: An evolution toward 5G," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 36–42, Mar. 2016.
[3] M. Shafi *et al.*, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.
[4] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 120–128, Feb. 2017.
[5] L. Ericsson, "More than 50 billion connected devices," Ericsson, Stockholm, Sweden, White Paper 14124, 2011.
[6] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.
[7] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, "Throughput and ergodic capacity of wireless energy harvesting based DF relaying network," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 4066–4071.
[8] Y. Yang, Y. Zhang, L. Dai, J. Li, S. Mumtaz, and J. Rodriguez, "Transmission capacity analysis of relay-assisted device-to-device overlay/underlay communication," *IEEE Trans. Ind. Informat.*, vol. 13, no. 1, pp. 380–389, Feb. 2017.
[9] A. Anand, G. De Veciana, and S. Shakkottai, "Joint scheduling of URLLC and eMBB traffic in 5G wireless networks," in *Proc. IEEE INFOCOM*, Apr. 2018, pp. 1970–1978.
[10] R. Li, Z. Zhao, Y. Zhong, C. Qi, and H. Zhang, "The stochastic geometry analyses of cellular networks with α-stable self-similarity," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2487–2503, Mar. 2019.
[11] S. Weber, J. G. Andrews, and N. Jindal, "An overview of the transmission capacity of wireless networks," *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3593–3604, Dec. 2010.
[12] S. A. R. Zaidi, A. Imran, D. C. McLernon, and M. Ghogho, "Characterizing coverage and downlink throughput of cloud empowered HetNets," *IEEE Commun. Lett.*, vol. 19, no. 6, pp. 1013–1016, Jun. 2015.
[13] N. Bhushan *et al.*, "Network densification: The dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb. 2014.
[14] V. Jungnickel *et al.*, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.
[15] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.
[16] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: How to empower SON with big data for enabling 5G," *IEEE Netw.*, vol. 28, no. 6, pp. 27–33, Nov. 2014.
[17] F. Han, S. Zhao, L. Zhang, and J. Wu, "Survey of strategies for switching off base stations in heterogeneous networks for greener 5G systems," *IEEE Access*, vol. 4, pp. 4959–4973, 2016.
[18] T. Alsedairy, Y. Qi, A. Imran, M. A. Imran, and B. Evans, "Self organising cloud cells: A resource efficient network densification strategy," *Trans. Emerg. Telecommun. Technol.*, vol. 26, no. 8, pp. 1096–1107, Apr. 2015.

[19] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Hoboken, NJ, USA: Wiley 2009.

[20] 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. Commun.*, vol. 21, no. 2, pp. 18–25, Apr. 2014.

[21] A. Asghar, H. Farooq, and A. Imran, "A novel load-aware cell association for simultaneous network capacity and user QoS optimization in emerging HetNets," in *Proc. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–7.

[22] A. J. Fehske, H. Klessig, J. Voigt, and G. P. Fettweis, "Concurrent load-aware adjustment of user association and antenna tilts in self-organizing radio networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1974–1988, Jun. 2013.

[23] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1029–1046, Sep. 2009.

[24] M. Kuczma, *An Introduction to the Theory of Functional Equations and Inequalities: Cauchy's Equation and Jensen's Inequality*. New York, NY, USA: Springer, 2009.

[25] A. De Domenico, E. C. Strinati, and A. Capone, "Enabling green cellular networks: A survey and outlook," *Comput. Commun.*, vol. 37, pp. 5–24, Jan. 2014.

[26] S. W. H. Shah, M. M. U. Rahman, A. N. Mian, A. Imran, S. Mumtaz, and O. A. Dobre, "On the impact of mode selection on effective capacity of device-to-device communication," *IEEE Wireless Commun. Lett.*, to be published.

[27] A. Klenke, *Probability Theory: A Comprehensive Course*. Springer, 2013.

SYED WAQAS HAIDER SHAH received the master's degree in electrical engineering from the National University of Science and Technology, Islamabad, Pakistan. He is currently a Ph.D. Research Scholar at the Internet-of-Things Research Laboratory, Information Technology University, Pakistan. He has published various papers in refereed international conferences and journals. His areas of research include wireless communication, future cellular networks, 5G, device-to-device communication, the analytical analysis of mobile networks, and the quality-of-service provisioning. He uses mathematical and system level tools to model and analyze emerging wireless communication architectures, leading to innovative and/or theoretically optimal new communication techniques. He is a Reviewer of various international journals, including the IEEE IoT Journal, the IEEE WCL, *Sensors* (MDPI), and the *Journal of Sensor and Actuator Networks* (MDPI).

ADNAN NOOR MIAN received the Ph.D. degree in computer engineering from the Sapienza University of Rome, Italy, where he was a Postdoctoral Researcher. He is currently a Visiting Research Scholar with the Computer Laboratory, University of Cambridge, U.K. He has published more than 30 papers in refereed conferences and journals. His research interests include wireless sensors and ad hoc networks, the Internet of Things (IoT), distributed algorithms, mobile and distributed systems, and cloud computing. He is currently involved in algorithms for finding efficient virtual LAN topologies, opportunistic networks, heuristics for dial a ride problem, the applications of random walk in large-scale wireless sensor networks, and exploiting EFI vehicle vs sensors using onboard diagnostics. He is a member of the Middleware Laboratory (Midlab), Department of Computer Science and Engineering, Sapienza University of Rome, where he was with Prof. R. Baldoni and Dr. R. Beraldi. In addition to research, he is also assisting the research community by serving in a number of technical program committees of national and international conferences. He is a Reviewer of many international journals and government project proposals.

SHAHID MUMTAZ received the master's degree in electrical and electronic engineering from the Blekinge Institute of Technology, Sweden, and the Ph.D. degree in electrical and electronic engineering from the University of Aveiro, Portugal. He has more than 10 years of wireless industry/academic experience. He is currently a Principal Research Scientist and the Technical Manager with the Instituto de Telecomunicações (IT), University of Aveiro, Portugal. From 2005 to 2006, he was a Researcher with Ericsson and Research Labs, Huawei, Sweden. Since 2014, he has been an Aux Professor with IT. He is also a Visiting Professor with various renowned universities and a Visiting Fellow with Nokia Bell Labs. He has published three books and more than 150 publications in a very high-rank IEEE Transactions, journals, book chapters, and international conferences. He uses mathematical and system level tools to model and analyze emerging wireless communication architectures, leading to innovative and/or theoretically optimal new communication techniques. He becomes the Vice-Chair of the IEEE Research Project on vision of green standardization due to his pioneer work in green communication. He is also actively involved in 3GPP standardization on LTE release 12 onwards, along with major manufacturers (Huawei and Intel). He is serving as the Scientific Expert and the Evaluator for Research Funding Agencies such as the European Commission, COST, and NSF China.

JON CROWCROFT received the degree in physics from Trinity College, University of Cambridge, in 1979, and the M.Sc. degree in computing and the Ph.D. degree from UCL, in 1981 and 1993, respectively. He has been the Marconi Professor of communications systems with the Computer Laboratory, University of Cambridge, since 2001. He was involved in the area of Internet support for multimedia communications for more than 30 years. The three main topics of interest have been scalable multicast routing, practical approaches to traffic management, and the design of deployable end-to-end protocols. His current active research areas include opportunistic communications, social networks, and techniques and algorithms to scale infrastructure-free mobile systems. He is a Fellow of the Royal Society, the ACM, the British Computer Society, the IET, and the Royal Academy of Engineering.