IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 11, 2018, accepted May 27, 2018, date of publication June 8, 2018, date of current version June 29, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2845690

An Optimization Design of Ultra Dense Networks Balancing Mobility and Densification

JINKANG ZHU[1], **(Life Member, IEEE), MING ZHAO**[1], **(Member, IEEE), AND SHENGLI ZHOU**[2], **(Fellow, IEEE)**
[1]Key Laboratory of Wireless-Optical Communications, School of Information Science and Technology, Chinese Academy of Sciences, University of Science and Technology of China, Hefei 230027, China
[2]Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269-4157, USA

Corresponding author: Ming Zhao (zhaoming@ustc.edu.cn)

This work was supported by the State Key Program of the National Natural Science Foundation of China under Grant 61631018.

ABSTRACT The ultra-dense network (UDN) has been identified as an appealing solution to address the huge service demands in future 5G and beyond. In this paper, we propose an optimization design of UDN balancing user mobility and network densification, where the massive users are divided into different groups according to their moving speeds, and different groups are served by different subnets. This way, the UDN can adjust subnet parameters to meet the service demands using minimum resource. Specifically, we first present the objective of the UDN optimization, then look at the area spectral efficiency of each individual subnet if selected to serve users within a speed range, and finally carry out the joint optimization across multiple subnets. Numerical examples demonstrate that the proposed approach is effective and practical, meeting the service demands with the least bandwidth resource consumption.

INDEX TERMS Ultra dense network, UDN optimization, user mobility, cell radius, area spectral efficiency, balancing mobility and densification, future 5G.

I. INTRODUCTION

As is known to all, 4G networks and services have been deployed all over the world. Now, facing the massive user access and the huge service requirements, especially as the proliferation of internet-connected mobile devices will continue to drive the data traffic growth dramatically in an exponential fashion, 4G network will not be adequate. Many companies have already began to consider deployment of 5G, and international organizations are stepping up their efforts to promote 5G standardization. The future 5G needs to achieve very comprehensive networking performance, such as higher spectral efficiency, higher area capacity, higher connectivity, and more stable user throughput, including to support very high mobile speeds [1]. To meet these demands, more efficient transmission technologies, more available spectrum strategies, and dense network architectures are expected as three important research paradigms [2], [3]. Indeed, wireless evolution has come a long way over the last 30 years, shifting the focus of innovation from transceiver technologies (e.g., MIMO, mMIMO) to network architecture and resource allocation, finally, to network densification.

Heterogeneous network (HetNet) has been introduced in 3G and 4G that is a typical progress of network densification, which is adding smaller footprint cells within macro cell deployments as a means of offloading traffic in hotspot localized areas. The basic features of HetNets are large cell/small cell non-overlapping random coverage based on random geometry, and the regular cellular coverage based on fractional frequency reuse [4], [5], where the interference cancellation and management in the cell edge with a few pico eNB deployments have been deeply discussed. The fractional frequency reuse (FFR)/soft frequency reuse (SFR) schemes as inter-cell interference coordination (ICIC) techniques of HetNets have been used in 4G LTE, the performance improvements for adding heterogeneous elements to the existing network have been demonstrated [6].

Along with increased spectrum efficiency (SE), spectrum bandwidth expansion and traffic offloading through WiFi, small cell deployment is regarded as one of the most promising ways to meet the use of broadband mobile service in both outdoor and indoor scenarios. In 2015, the European Commission has proposed DECADE project for deploying high capacity dense small cell heterogeneous networks (dense HetNets) to study multi-tier heterogeneous networks, in which a mix of access node types, such as macrocell, femtocell and relay, will co-exist to achieve successfully small cell and HetNet deployment [7], [8]. The energy efficiency (EE) and spectral efficiency improvement of uniform and non-uniform user equipment and base stations distributions has been studied, using an ON/OFF switching algorithm in a dense HetNet environment [9], [10].

A framework for joint user, power and subcarrier allocation on the downlink of a multi-tier HetNet is proposed in [11]. Based on different link distances characterized by multi-slope path loss models, the proposed framework ensures the users quality of service requirements, namely, user minimum rate and the base stations maximum transmission power [11]. In consisting multi-tier densely deployed base stations of dense HetNet, the downlink coverage performance of the HetNet with non-cooperative and cooperative schemes is studied, where the cooperative NOMA can significantly improve the coverage performance [12]. These advances show that heterogeneous networks are moving towards dense HetNets, whose main feature is the overlapping coverage of multi-tier networks.

The ultra dense network (UDN) has recently emerged as a prominent solution to meet the challenges of extremely high capacity density requirements of 5G, which is a network with much higher density of wireless spectral resources than that in current networks [13], [14]. Heterogeneous, overlapping and efficient deployment will be the important coverage features of ultra-dense networks with a large number of access points. The different UDN deployment strategies have been evaluated from the spectrum efficiency to the area capacity for 5G [15], [16]. In particular, how to enable a wide range of mobility support is a great challenge [17], [18].

A. RELATED WORKS

Towards realizing the goals set for 5G, UDN has been viewed as the new wireless frontier that can be made a reality only through re-thinking of current cellular practice [19]. A successful UDN deployment is not an easy task. Prominent solutions on UDN have started being proposed from different views. Refrence [13] has given an overview of the current state of UDN development and discussed that base station (BS) sleeping effectively improve the average SE and EE.

The insights on fundamental issues related to UDN deployment is provided in [20], where mobile user devices acting as "infrastructure prosumers" are used as access nodes. Two efficient localized mobility management schemes considering small cell deployments and backhaul topology are proposed in [21]. The first one centralizes mobility management control from small cell access points into a local server, another one allows individual small cell access points to handle mobility events. In [22], the backhaul network capacity and the backhaul energy efficiency of ultra dense cellular networks are investigated to answer how much densification can be deployed for 5G ultra dense cellular networks. The mm-Wave bands used in UDN is discussed [20], due to the availability of vast amount of frequency resources compared to traditional bands, and hence one can use different spectral resources among cells with different sizes. The hypergraph theory is recognized as a useful mathematical tool to model the complex relations among multiple entities, for solving the problem of resource allocation in UDN [23]. After presenting a potential network architecture for the UDN, a generalized orthogonal/non-orthogonal random access scheme is proposed to improve the network efficiency while reducing the signaling overhead, based on a Markov chain model [24].

In short, UDN has a multi-tier networks architecture building upon macro cell, micro cell, mmwave cell, and so on, to decompose into two-tier, three-tier, four-tier and even more multi-tier networks. Since various types of cell structures clustering users to meet different requirements is designed with more complexity of infrastructure and higher diversity of associated devices and resources. How to establish an effective design method and optimization framework for UDN, balancing user mobility and cell densification, is an important issue that must be addressed.

B. CONTRIBUTIONS

In this paper, we propose a novel optimization design for UDN by user grouping based on the mobility. Specifically, the massive users are divided into different user groups, the corresponding cell radius and suitable subnets are selected, and the subnet parameters are jointly adjusted to meet the service demand with minimum resource cost. We prove optimization theorems for a UDN with multi-tier subnets to meet the service demands, establish the mathematical relationship between the speed range and the cell radius of subnets, and analyze the performance for adjusting various subnets. Numerical examples illustrate the effectiveness of the proposed optimization design and show that the total bandwidth is minimized while serving the demands in a given area.

The rest of this paper is organized as follows. Section II describes the problem formulation and the main idea of mobility-based user grouping. Section III focuses on the optimization of each individual subnet while Section IV presents the joint optimization across subnets. Section V provides numerical examples and Section VI offers concluding remarks.

II. PROBLEM FORMULATION

Heterogeneous ultra dense networks are envisioned to accommodate fast growing user demands in future wireless systems. At any given space and time, user traffics can go through different networks depending on the system scheduling. How to coordinate the network resources of co-existing heterogeneous networks to best serve the user demands in a given area is the research goal of this paper, where we propose a novel optimization framework via mobility-based user grouping and network access.

Consider a typical scenario with M users, $m = 1, 2, \ldots\ldots, M$, located in an area of A_{req} km^2. Each user, say the mth user, has a different moving velocity v_m, where the user moving velocity may also change over time. The mth user has a service request $S_m(t)$, which is a time-dependent function (e.g., daytime and night time traffics are different). The total service demands from M users at time t is denoted as

$$S_{\text{total}}(t) = \sum_{m=1}^{M} S_m(t). \quad (1)$$

If $M \gg 1$, the total demand $S_{\text{total}}(t)$ will be very large, and the user moving speeds will be diverse.

The aim here is to efficiently build a UDN, structured by multi-tier subnets, to meet the service demands. Assume that the number of available subnets is K and the transmitted capacity of the kth subnet is $C_k(t)$, the achieved capacity of the UDN is then

$$C_{\text{total}}(t) = \sum_{k=1}^{K} C_k(t), \quad (2)$$

where $C_{\text{total}}(t)$ (Gbps/km^2) is also called as the area throughput of the UDN. To meet the service demand, we need:

$$C_{\text{total}}(t) \geq S_{\text{total}}(t). \quad (3)$$

The question is on how to achieve this goal while spending the least system resources. We next present a novel optimization framework based on partition of users into different speed groups.

A. USER GROUPING BASED ON MOBILITY

Due to the time-varying nature of the service demand and the user mobility patterns, the resource allocation of the UDN will be time-dependent. For notational brevity, we focus on one period of time and drop the dependence on time t, e.g., expressing $S_{\text{total}}(t)$, $C_{\text{total}}(t)$, $S_m(t)$ and $C_k(t)$ as S_{total}, C_{total}, S_m and C_k, respectively.

The key innovation of this paper is network optimization through the participation of users based on their mobility. We profile the service demand based on the user speed as follows. Assume that the total number of users is large: $M \gg 1$, and let $\rho_S(v)$ denote the service demand density[1] at speed v. In other words, the quantity

$$\int_{V_1}^{V_2} \rho_S(v) dv \quad (4)$$

represents the aggregated service demands from all users with moving speeds in the range of $[V_1, V_2)$. For the overall system, we have

$$S_{\text{total}} = \int_0^{V_{\max}} \rho_S(v) dv \quad (5)$$

where the user speed is distributed in the range of $[0, V_{\max}]$.

Assume that each subset is used to support one group of users. Corresponding to the available K subnets, the users are now split into at most K different groups. For the kth group, the speed set is defined as $\mathcal{V}_k$, which could be one contiguous interval, or consists of multiple intervals. If $\mathcal{V}_k$ is empty, it means that the kth subnet is not actually used.

The transmission capacity C_k of the kth subnet depends greatly on the resources that can be scheduled and used by each subnet. Let B_k be the bandwidth of the kth subnet and assume that there are other Q resources types (such as sectors, MIMO beamforming, cell density, interference management schemes) that determine the capacity, which are collected in the parameter set $\mathcal{Q}_k = (x_1, \ldots, x_q, \ldots, x_Q)$. We explicitly express C_k as a function of the bandwidth B_k, the users that it supports $\mathcal{V}_k$, and other parameters in $\mathcal{Q}_k$ as

$$C_k = C_k(B_k, \mathcal{V}_k, \mathcal{Q}_k). \quad (6)$$

B. OPTIMIZATION FRAMEWORK

The goal of the network optimization is to coordinate K suitable subnets of UDN to meet the huge service demands from all users with the minimum resource. The proposed optimization framework can be suitable to various cost functions. Specifically, one can define the cost for the kth subnet as $cost(B_k, \mathcal{V}_k, \mathcal{Q}_k)$, and define an optimization problem as

$$\min \sum_{k=1}^{K} \text{cost}(B_k, \mathcal{V}_k, \mathcal{Q}_k), \quad (7)$$

$$\text{s.t. } C_k(B_k, \mathcal{V}_k, , \mathcal{Q}_k) \geq \int_{\mathcal{V}_k} \rho_S(v) dv, \quad (8)$$

$$\bigcup_{k=1}^{K} \mathcal{V}_k = [0, V_{\max}]. \quad (9)$$

Constraint (8) means that the kth subset needs to support all the users allocated to it, and constraint (9) makes sure that all the users in the system are supported. The unique feature of the proposed framework in (7) is network optimization through mobility-based user grouping.

C. FOCUS OF THIS PAPER

In this article, we focus on one specific example where the bandwidth consumption is used as the objective function, formulated as

$$\min B_{\text{total}} = \sum_{k=1}^{K} B_k \quad (10)$$

$$\text{s.t. } C_k(B_k, \mathcal{V}_k, , \mathcal{Q}_k) \geq \int_{\mathcal{V}_k} \rho_S(v) dv, \quad (11)$$

$$\bigcup_{k=1}^{K} \mathcal{V}_k = [0, V_{\max}]. \quad (12)$$

[1] Here we clarify how the service demand density can be estimated empirically based on the communication profiles of network operations. Consider a scenario of interest, e.g., a specific day-time interval over a given area, where the historical data of all service demands can be obtained, along with the user moving speeds collected when the service requests were made. One can write the service demand as a function of the user index m and the moving speed v_m, as $\tilde{\rho}(m, v_m)$. Now project the service demands into a speed interval $[v, v + \Delta v]$ as

$$\rho_S(v) = \sum_{\forall m \mid v_m \in [v, \Delta v)} \tilde{\rho}(m, v_m),$$

where Δv is the step size. This way, the service demand density $\rho_S(v)$ can be obtained empirically. Due to the large number of users, the individual user behaviours, e.g., not all users with common velocity may require identical data demands or one user might have varying velocity profiles, might not affect the accuracy of the profiling on the aggregated rates over the dense networks.

The reason of using bandwidth consumption as the optimization objective is two-fold. First, wireless spectrum is one of the most valuable resources for wireless systems as the spectrum is limited while the user demands are growing in a fast pace.[2] Second, the optimization problem in (10) is one concrete example of the general framework of (7), whose solutions can be developed as shown in Sections III and IV.

We next solve the specific optimization problem in (10), minimizing the spectral resource consumption while serving the total service demands. We first look at individual optimization of each subnet, after which we carry out the joint optimization across the subnets.

III. OPTIMIZATION OF INDIVIDUAL SUBNET

Different types of subnets might be deployed in a given area to support the high user demand. In this paper, we adopt four example subnets as illustrated in Fig. 1. Subnet 1 is a typical macro cell cellular subnet. Subnet 2 is a micro cell subnet with fractional frequency reuse (FFR) [5], which can also adopt cooperative or antenna active system (AAS) techniques. Subnet 3 has a small cell structure for outdoor coverage, and subnet 4 deploys pico/femto cells for indoor applications. Note that in our optimization framework subnets could be homogeneous or heterogeneous. In the extreme case, all the subnets could be of the same type but designed with different parameters and operating at different frequency bands.

For each subnet, the area capacity is proportional to the bandwidth B_k as

$$C_k = B_k \sigma_k, \tag{13}$$

where σ_k denotes the area capacity subject to a unit bandwidth (also called as the area spectral efficiency) for the kth subnet. Therefore, maximization of σ_k is a must to reduce the bandwidth consumption. We next analyze the area spectral efficiency for different types of networks.

A. AREA SPECTRAL EFFICIENCY EVALUATION

In the literature, analysis of regular networks and random networks have been carried out. Here, we first define the area spectral efficiency (ASE) for a general network, and then evaluate the ASE of regular and random networks to validate the theoretical results.

Define an area of A_{req} square meters, where there are N_{cell} base stations with each base station in one cell. Once the positions of the base stations are known, the cell boundary can be drawn if we assume that the average SNR is used for a user to select the base stations. Now consider the ith cell $\mathcal{C}_i$. Consider a downlink scenario where each base station is in communication with one mobile user at a given

[2]Normally a communication transceiver would be designed with a particular bandwidth in the first place, and operates on one assigned subchannel when a user is admitted into the network. A communication network can be allocated with multiple subchannels when there are more bandwidth resources become available. If the subchannels of all the subnets have equal bandwidth, the bandwidth optimization problem can be viewed as a subchannel assignment problem where the least amount of subchannels are allocated to different subnets to support the total service demand.

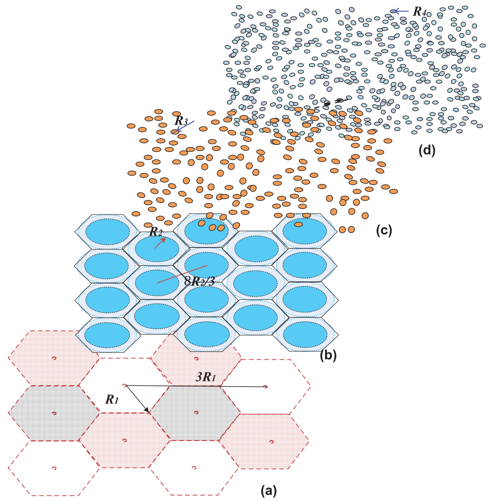

FIGURE 1. Four types of selected subnets. (a) Subnet 1. (b) Subnet 2. (c) Subnet 3. (d) Subnet 4.

unit bandwidth.[3] The user is randomly located within the cell. The mobile unit is located at (x, y), and (x_j, y_j) is the position of the jth interfering base station. Let P_j denote the transmission power of base station j, h_j is the normalized channel from the mobile to base station j, and $\|X_j\|$ denote the distance from the user to base station j. The signal to noise plus ratio (SINR) for the mobile user is then

$$\text{SINR}_i = \frac{P_i |h_i|^2 \|X_i\|^{-\alpha}}{\sum_{j \in \Phi_i} P_j |h_j|^2 \|X_j\|^{-\alpha} + \sigma_i^2}, \tag{14}$$

where the interference is from the set of interfering base stations (denoted by Φ_i) which use the same frequency. At a given location (x_0, y_0) within $\mathcal{C}_i$, the average data rate of the user is

$$\bar{I}_i|_{(x_0,y_0)} = E_{h_i,\{h_j\}} \left[\log_2(1 + \text{SINR}_i)\right], \tag{15}$$

where the expectation is over the channel realizations. Further assume that the user is uniformly distributed in the cell, the average data rate for a user in the cell is

$$\bar{I}_i = E_{(x_0,y_0) \in \mathcal{C}_i} \left[\bar{I}_i|_{(x_0,y_0)}\right], \tag{16}$$

where the expectation is taken over user locations.

The area spectral efficiency for the network is defined as

$$\text{ASE} = \frac{\sum_{i=1}^{N_{\text{cell}}} \bar{I}_i}{A_{\text{req}} \Delta_{\text{FR}}}, \tag{17}$$

where Δ_{FR} is the frequency reuse factor as a particular frequency might not be allowed to use in all the cells. Now we look at the special cases of regular and random networks.

[3]In this paper, we assume FDMA or OFDMA within each cell and frequency reuse across cells. A full loaded scenario is assumed where each base station is in communication with one mobile user on each available frequency point. Therefore, the SINR performance does not change as the number of users is increasing. More frequency resources are demanded from the network to accommodate all the users when more users make communication requests.

1) REGULAR NETWORK

For a regular network, a cell is often assumed to have a radius R, and each cell covers a circular area πR^2. And hence, the number of cells is approximately

$$N_{\mathrm{cell}} \approx \frac{A_{\mathrm{req}}}{\pi R^2}, \tag{18}$$

For a regular network, all the cells are uniform, and $\bar{I}_i = \bar{I}$, and we have

$$\mathrm{ASE} = \frac{\bar{I}}{\pi R^2 \Delta_{\mathrm{FR}}}. \tag{19}$$

A unified analytical framework to evaluate the ASE in (19) is developed in [26]. Note that the channel gain $|h_j|$ can capture both the large-scale shadowing effect and the small-scale multipath fading effect. Explicit expressions are obtained for the ASEs in the presence of composite fading channels, where the large-scale fading follows the log-normal distribution and the fast multipath fading follows the Nakagami-m distribution [26]. In ultra-dense networks where line-of-sight (LoS) paths could exist, Rician fading could be considered for fast multipath fading. Performance analysis with Rician fading can be carried out following the same steps in [26] or can be approximated by Nakagami fading models. Note that other physical layer effects such as Doppler compensation and channel estimation are not included when evaluating the ASEs; this is typical for performance analysis of wireless networks [25]–[27].

Example 1: For the illustration purpose, assume that all the transmission powers are equal $P_j = P_0$, and all the channels follow Rayleigh fading. Based on [26, eq. (13)] and that the moment generation function for an exponential random variable γ is $\mathcal{M}(z) = \mathrm{E}\left[e^{-z\gamma}\right] = 1/(1 + z\mathrm{E}[\gamma])$, one can readily obtain an explicit expression for numerical expression. We also carry out a straight-forward simulation study where users are randomly located within their cells and the channels are randomly generated for performance evaluation. Adopting a typical value $\Delta_{\mathrm{FR}} = 3$, Fig. 2 shows that the theoretical and the simulation curves agree with each other very well.

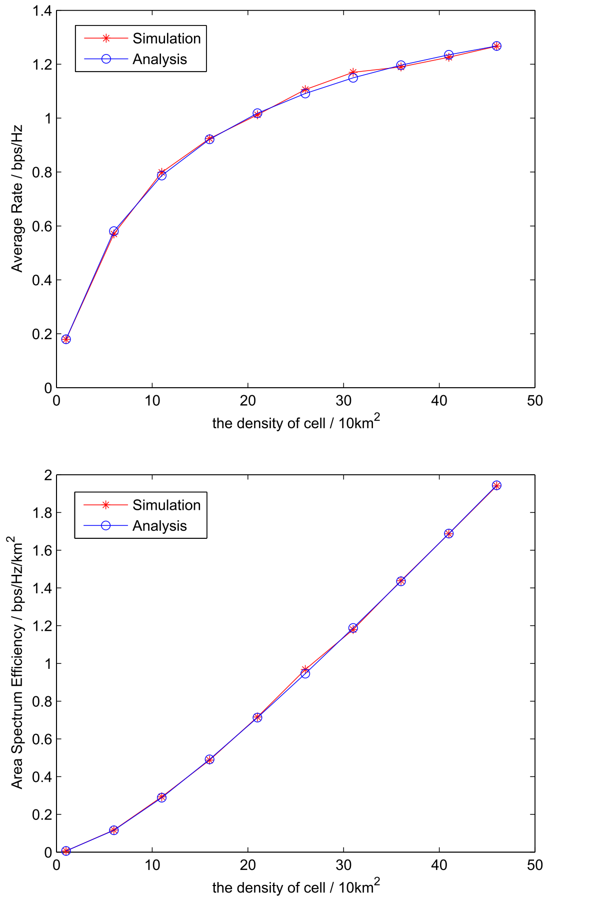

FIGURE 2. Comparison of the simulated and analytic results [26] on area spectral efficiency of a regular network ($\Delta_{\mathrm{FR}} = 3$).

2) RANDOM NETWORK

Most of the recent literature on dense networks and heterogeneous networks assume random networks. While macro cell networks can be carefully planned, small cell or pico/femto cell are often deployed in an ad-hoc manner, and the number of base stations is a random variable. A spatial Poisson point process (PPP) is often assumed with base station density λ, as shown in [27], and the average number of base stations is

$$E\left[N_{\mathrm{cell}}\right] = \lambda A_{\mathrm{req}}. \tag{20}$$

For a random network, there is often no frequency reuse, $\Delta_{\mathrm{FR}} = 1$, and the average spectral efficiency over the Poisson distribution is

$$\overline{\mathrm{ASE}} = E[\sigma] = \lambda E[\bar{I}_i]. \tag{21}$$

Note that in a random network, there could be multiple tiers, with different sets of antennas having different transmission powers P_i, which is often termed as a heterogeneous network (HetNet).

Example 2: For a random HetNet, the average data rate for a user is derived in [27, eq. (5)]. Using only one tier, the simulation result is compared against the theoretical result in Fig. 3. There are some small gaps between the simulated and theoretical curves. Note that the cell association in this simulation is based on the average SNR, and hence the cell boundaries can be drawn once the base stations positions are known, however, in [27] the cell selection is based on the instantaneous SNR. Nonetheless, the ASE curves are pretty close for a validation.

3) NETWORKS WITH VARIOUS ENHANCEMENT TECHNIQUES

In the aforementioned analysis of regular and random networks, only a basic single-antenna setup is examined. Other techniques such as sectioning, beamforming, MIMO, and soft frequency reuse can certainly increase the spectral efficiency of a network. Once a particular technique is adopted, the area spectral efficiency can be evaluated mostly via simulations.

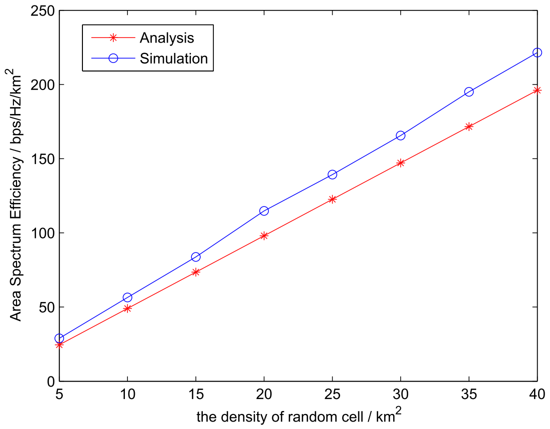

FIGURE 3. Comparison of the simulated and analytic results [27] (threshold SNR set at 0 dB) on area spectral efficiency of a random network.

The main idea of this paper is not on how to compute the area spectral efficiency of different networks, but rather optimize the system resources allocated to different networks once their characteristics are determined.

B. CELL RADIUS LOWER BOUND

In the proposed system, each subnet serves one category of users where the cell radius of subnet is determined by the user velocity and user density. References [28] and [29] have proposed a dynamic handoff model in a regular network. Assume that the cell radius is R, there are M_{cell} users uniformly distributed within the cell coverage area. Assume that the moving speed of user m is v_m, the maximum speed is $v^{\max}$, and the handover time is t_{ho}. The maximum active radius of a user is defined as

$$r_{\text{ho}} = v^{\max} t_{\text{ho}} \tag{22}$$

at the handover period, and the max active area of a user is πr_{ho}^2. The number of mobile users that can be accommodated in the area of the radius R is then [29, eq.(24)]

$$M_{\text{cell}} = \gamma \frac{\pi R^2}{\pi r_{\text{ho}}^2} = \gamma \left(\frac{R}{r_{\text{ho}}}\right)^2, \tag{23}$$

where γ is the overlapping coefficient of the active area of users. Thus, the cell radius for supporting the user handoff is lowered bounded by

$$R \geq \underbrace{v^{\max} t_{\text{ho}} \sqrt{\frac{M_{\text{cell}}}{\gamma}}}_{:=\Gamma(v^{\max})} \tag{24}$$

In general, $\gamma = 1$ and $t_{\text{ho}} = 1.0$-1.5s. If a user with moving speed 50 km/h or 120 km/h, the minimum cell radius is 80.2m or 192.5m. In practice, the cell radius is often selected to be slightly larger than the lower bound to add a performance margin. In the numerical evaluation, the cell radius is selected to be 1.5 times of the lower bound for a regular network, and is selected to be 4.5 times of the lower bound for a random network. As one increases the ratio of

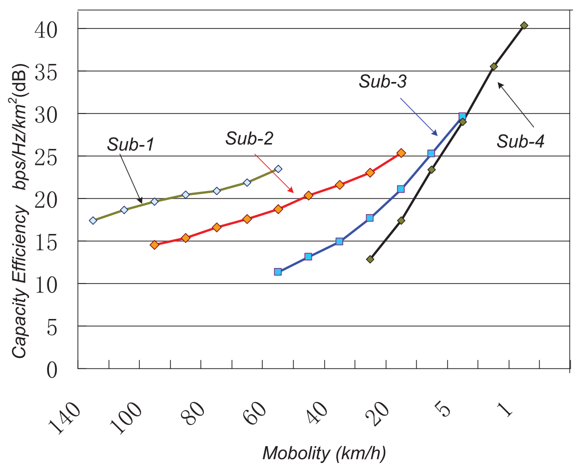

FIGURE 4. Achievable area spectral efficiency as a function of the maximum speed.

the selected cell radius over the lower bound, lower outage would be achieved during the handoff, while the network area spectrum efficiency would decrease. Hence, the ratio could be slightly modified balancing the network throughput and the performance margin; however, the trends of the performance curves with slightly different values would be similar.

C. CAPACITY OPTIMIZATION BY ADJUSTING THE CELL RADIUS

Now consider the k subnet, which support users with speed in $\mathcal{V}_k$, and the area spectral efficiency for a given radius is σ_k. Define the maximum speed as

$$v_k^{\max} = \max\{v | v \in \mathcal{V}_k\}. \tag{25}$$

By adjusting the cell radius while being able to support the users with mobility in $\mathcal{V}_k$, one can maximize the area spectral efficiency through

$$\max_R \; \sigma_k \quad \text{subject to } R \geq \Gamma_k(v_k^{\max}) \tag{26}$$

This way, we can solve the achievable area spectral efficiency of different subnets depending on the up limit value of speed interval, as shown in Fig. 4 for four example subnets, where the Y axis is in the dB scale for illustration convenience.

There is one important property that will be used in Section IV: The area spectral efficiency is non-increasing with the maximum speed. Specifically, if $v_1 \leq v_2$, then

$$\max_{R \geq \Gamma_k(v_1)} \sigma_k \geq \max_{R \geq \Gamma_k(v_2)} \sigma_k \tag{27}$$

IV. JOINT OPTIMIZATION ACROSS SUBNETS

After individual subset optimization, the optimization across multiple subnets can be carried out to meet the service demand with minimal bandwidth consumption. From Section III, we find that the area spectral efficiency only depends on the maximum speed of the users, and hence denoted as $\sigma_k(v_k^{\max})$ in this section. Once the set $\mathcal{V}_k$ is given, the bandwidth needed by the kth subnet is

$$B_k \geq \frac{\int_{\mathcal{V}_k} \rho_S(v)dv}{\sigma_k(v_k^{\max})} \tag{28}$$

to satisfy the service constraint. The challenge is on how to select the subnets and how to assign the users to different subnets. Note that there are K subnets and not all of them might be used. For those subnets that are not used, we have $\mathcal{V}_k = \emptyset$.

We focus on the scenario where each set $\mathcal{V}_k$ is one interval. The scenario where one set might have multiple non-overlapping intervals will not show up if any two area spectral efficiency curves intersect at most once, which is the case for the curves in Fig. 4. Assume that the user speed is divided into J non-empty consecutive intervals as

$$\mathcal{V}_1 = [V_1, V_2) \tag{29}$$
$$\mathcal{V}_2 = [V_2, V_3) \tag{30}$$
$$\vdots$$
$$\mathcal{V}_J = [V_J, 0] \tag{31}$$

where

$$V_1 > V_2 > \cdots > V_J > 0. \tag{32}$$

Out of the K available subnets, the subnet selected to support $\mathcal{V}_j$ is then termed as subnet j. This way, we have

$$v_j^{\max} = V_j. \tag{33}$$

The maximum speed needs to be equal $V_1 = V_{\max}$ to cover all the users. Now, the optimization problem in (10) becomes

$$\min_{V_1>V_2>\cdots>V_J>0} \sum_{j=1}^{J} \frac{\int_{V_{j+1}}^{V_j} \rho_S(v)dv}{\sigma_k(V_j)} \tag{34}$$

where $V_{J+1} = 0$. The solution to (34) will be discussed in Section IV-A.

Taking J subnets out of K subnets in an ordered fashion, there are $K!/J!$ possibilities. Further, J can vary between 1 and K. In total, there are $\sum_{J=1}^{K} K!/J!$ scenarios to be examined, and the one with the smallest bandwidth efficiency will be the final solution. However, many scenarios can be easily discarded based on simple reasoning and the properties of different subnets, which will be discussed in Section IV-B.

A. PROPERTIES OF OPTIMAL SOLUTION TO (34)

We have the following necessary conditions on $\{V_j\}_{j=2}^{J}$ for the optimal solution.

Theorem 1: The optimal speed ranges shall satisfy the following conditions

$$\frac{\rho_S(V_{j+1})}{\sigma_{j+1}(V_{j+1})} - \frac{\rho_S(V_{j+1})}{\sigma_j(V_j)} - \frac{\int_{V_{j+2}}^{V_{j+1}} \rho_S(v)dv}{[\sigma_{j+1}(V_{j+1})]^2}\dot{\sigma}_{j+1}(V_{j+1}) = 0 \quad \forall\, j = 1, 2, \ldots, J-1 \tag{35}$$

where $\dot{\sigma}_j(v)$ is the derivative of $\sigma_j(v)$.

Proof: Based on (34), the objective function may have an extreme point when its derivative on $V_{j+1}, j = 1, 2, \ldots, J-1$, goes to zero:

$$\sum_{k=1}^{J} \frac{d\left(\frac{\int_{V_{k+1}}^{V_k} \rho_S(v)dv}{\sigma_k(V_k)}\right)}{d(V_{j+1})} = 0. \tag{36}$$

For each j, there are two terms from the sum that are relevant to V_{j+1}, where $k = j$ and $k = j+1$. This means

$$\frac{d\left(\frac{\int_{V_{j+1}}^{V_j} \rho_S(v)dv}{\sigma_j(V_j)}\right)}{d(V_{j+1})} + \frac{d\left(\frac{\int_{V_{j+2}}^{V_{j+1}} \rho_S(v)dv}{\sigma_{j+1}(V_{j+1})}\right)}{d(V_{j+1})} = 0. \tag{37}$$

Carrying out the differentiation by the chain rule, we obtain (35).

Theorem 2: With the property in (27)*, the optimal solution of* $(V_1, \ldots, V_J)$ *should satisfy*

$$\sigma_1(V_1) \leq \sigma_2(V_2) \leq \ldots \leq \sigma_J(V_J). \tag{38}$$

Proof: The property in (27) means that $\sigma_{j+1}(v)$ is nonincreasing as the speed upperlimit increases from low to high: hence $\dot{\sigma}_{j+1}(v) \leq 0$. Using this property in (35), we have speed up-limit increases from low to high, obtain $\dot{\sigma}_{k+1}(v) \leq 0$.

$$\frac{\rho_S(V_{j+1})}{\sigma_{j+1}(V_{j+1})} - \frac{\rho_S(V_{j+1})}{\sigma_j(V_j)} \leq 0, \tag{39}$$

which leads to

$$\sigma_j(V_j) \leq \sigma_{j+1}(V_{j+1}). \tag{40}$$

Putting the inequalities corresponding to $j = 1, \ldots, J-1$ together, we have (38).

We can see that the optimal V_{j+1} depends on the values of V_j and V_{j+2}. Therefore, a joint search of $\{V_2, \ldots, V_J\}$ is needed to find the optimal solution to (34). However, Theorem 2 can help to reduce the search space. For each tentative V_j, the starting point of V_{j+1} should be that $\sigma_{j+1}(V_{j+1}) \geq \sigma_j(V_j)$. If we cannot find $\sigma_{j+1}(V_{j+1}) \geq \sigma_j(V_j)$, where $V_{j+1} < V_j$, this means that subnet $j+1$ should not be used, and the number of active subsets is reduced by one.

B. DISCUSSIONS ON IMPLEMENTATION COMPLEXITY

There are $\sum_{J=1}^{K} K!/J!$ cases to be considered. When there are J active subnets, there are $J-1$ variables $\{V_2, \ldots, V_J\}$ to be jointly optimized. The optimization problem is implemented based on numerical search where the speed variables are taken from a discrete grid. Note that there are certain properties which might help reduce the number of cases. For example, one would prefer the macro cell subnet as the first subnet to support high-mobility users, as done in the numerical section.

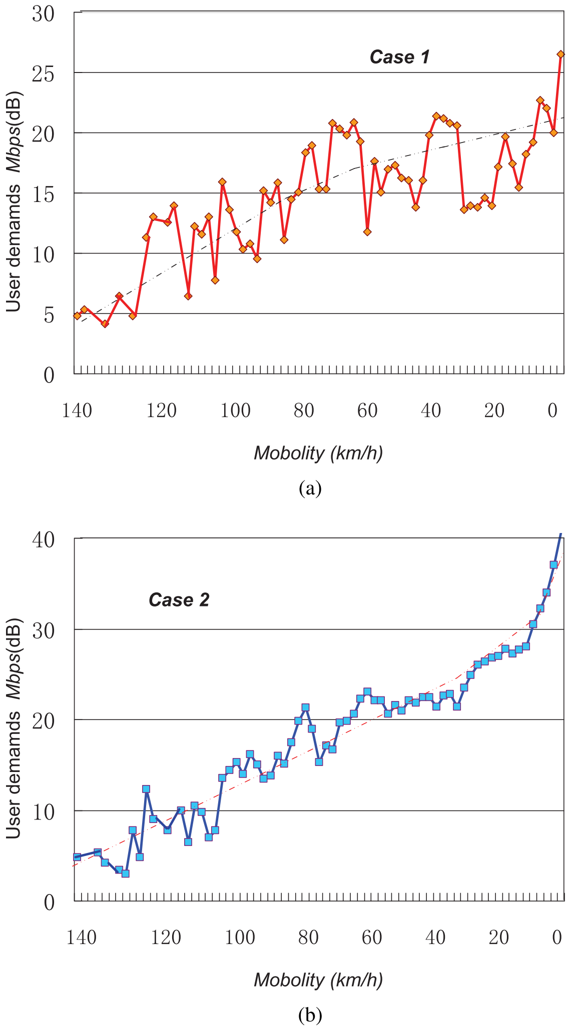

FIGURE 5. User demands of different moving speeds. (a) Case 1. (b) Case 2.

TABLE 1. Two fixed allocation of user mobility.

# of subnets	Labels	Speed Sets
$J = 2$	A	(140,60], (60,0]
	B	(140, 40], (40,0]
$J = 3$	A	(140,80], (80,40], (40,0]
	B	(140,62], (62,22], (22,0]
$J = 4$	A	(140,90], (90,60], (60-30], (30-0]
	B	(140,90], (90, 46], (46, 16], (16,0]

TABLE 2. The optimal user grouping based on mobility for cases 1 and 2.

Used subnets	Cases	Speed Sets
(1,2)	Case 1	(140,42], (41,0]
	Case 2	(140,28], (27,0]
(1,2,4)	Case 1	(140,62], (61,12], (11,0]
	Case 2	(140, 60], (59,12], (11,0]
(1,2,3,4)	Case 1	(140,68], (67,26], (25,10], (9,0]
	Case 2	(140,68], (67,26], (25,10], (9,0]

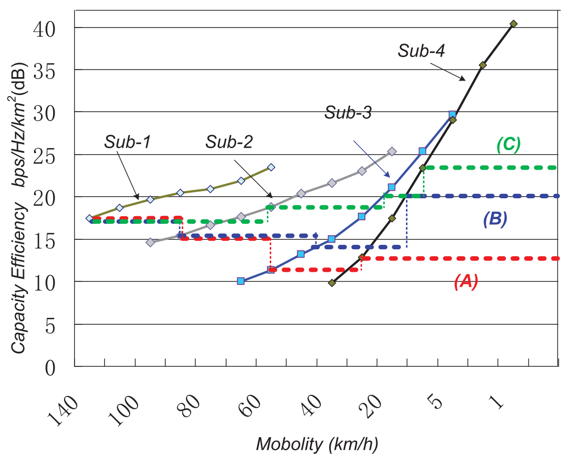

FIGURE 6. Illustration of of the speed sets $\mathcal{V}_1$ to $\mathcal{V}_4$ for (A), (B) and (C).

V. NUMERICAL EXAMPLE

For numerical illustration, we provide two data sets of service requirements, denoted as cases 1 and case 2, which correspond to network operations before and after working hours. As shown in Figs. 5(a) and 5(b), the services demands (Mbps) of 200 users are organized based on the moving speed on a 2km/h of quantization interval. The total service demands are $S_{\text{total}} = 3.54$ Gbps and $S_{\text{total}} = 30.61$ Gbps for cases 1 and case 2, respectively. One might obtain these profiles from data collection campaigns [30], by sorting out the data service demand based on the moving speed.

There is a total of $K = 4$ subnets, and there is flexibility to choose J subnets to serve the service demand, $1 \leq J \leq K$. To compare with the optimal allocation methods, we list two fixed allocation methods in Table 1 for comparison, when $J = 1$ subnet 1 is selected to cover all speed interval (140, 0). In method A, the intervals are equally split between 120 km/h and 0, except the first interval which will support the user speed up to 140 km/h. In method B, the intervals are successively decreasing. We run the optimization algorithm to select the speed sets. The results for selected configurations are shown in Table 2. For the $J = K = 4$ case, Fig. 6 illustrates the subnet selection path for (A), (B), and (C), where (C) is the optimal interval selection, satisfying the property in (38).

The total bandwidth consumption for different cases are summarized in Figs. 7 and 8. When $J = 1$, only the 1st subnet (macro cell) is selected to support the high moving speed. The resource cost (frequency bandwidth) for meeting the total service demands for Case 1 and Case 2 is $B_{\text{total}} = 62.8$ MHz and $B_{\text{total}} = 542.9$ MHz respectively. For $J = 2$, there are three configurations: $(1, 2)$, $(1, 3)$, and $(1, 4)$. For $J = 3$, there are also three configurations: $(1, 2, 3)$, $(1, 2, 4)$, and $(1, 3, 4)$. For $J = 4$, all the four subnets are used. For Case 1, the minimum frequency bandwidth is reduced to $B_{\text{total}} = 22.2$ MHz, and for Case 2 it is $B_{total} = 53.0$ MHz.

Figs. 7 and 8 show that the bandwidth allocation through the optimal splitting of user mobility is better than that based on fixed splittings, and increasing the number of subnets can considerably reduce the bandwidth consumption while serving the total service demand. The percentage of bandwidth reduction by using multiple subnets, relative to a single macro subnet (Subnet 1) is summarized in Fig. 9.

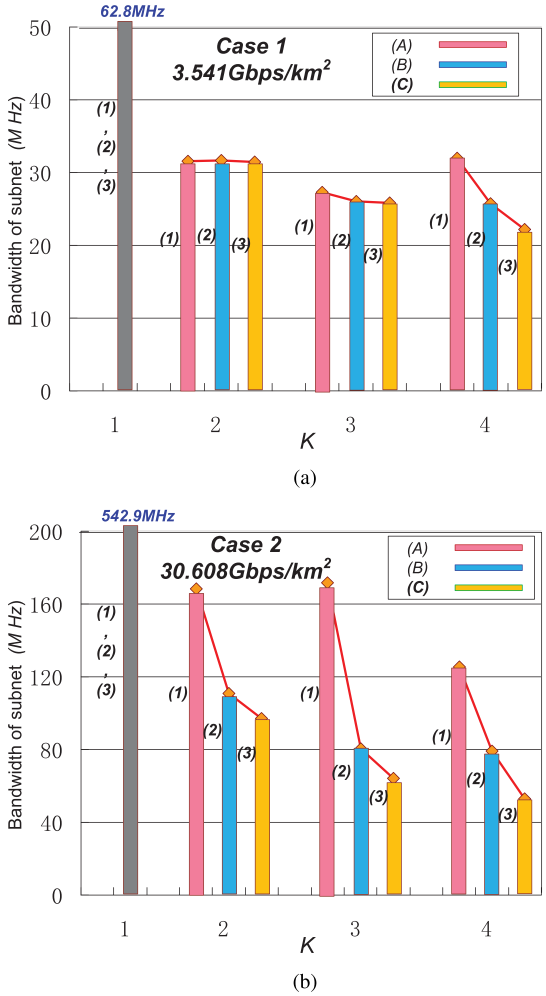

FIGURE 7. The total bandwidth consumption for different strategies. (a) Case 1. (b) Case 2.

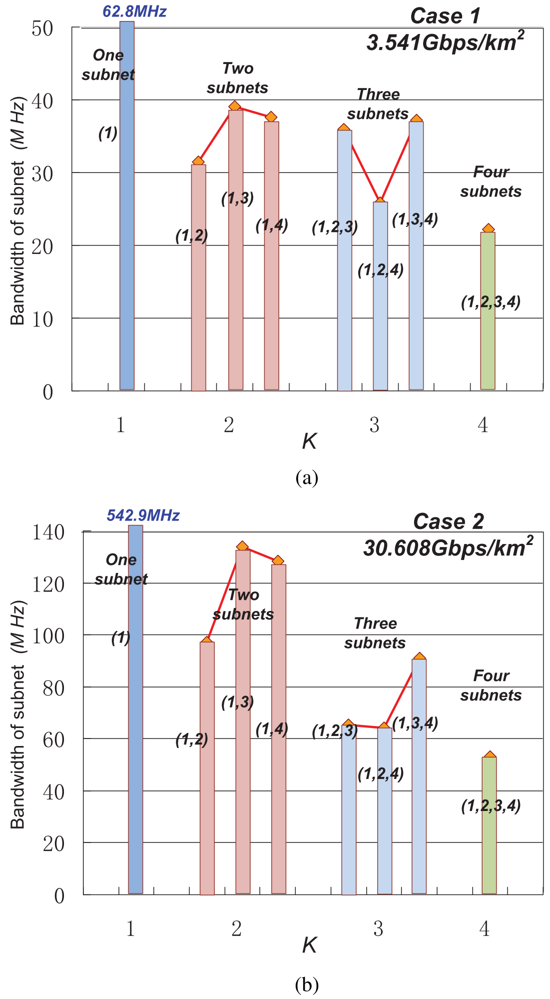

FIGURE 8. Total bandwidth consumption with optimal speed splitting for different setups. (a) Case 1. (b) Case 2.

VI. CONCLUSION

In this paper, we proposed a novel optimization design for ultra dense networks. Specifically, the mobile users are divided into different groups based on their moving speeds, and different subnets serve different user groups, optimizing its system parameters based on the speed range of the allocated users. Based on the profiles of the service demand and user speeds, optimization is carried out across multiple subnets meeting the service demands with minimal bandwidth consumption. The analysis and numerical results illustrate that: (i) The area spectral efficiency of each subnet is of crucial importance for the optimization of UDN; and (ii) The optimization of user speed ranges covered by subnets for different user groups depend on the area spectral efficiency of all the subnets and the variations of the service demands.

The proposed optimization design identified the user moving speed as one reference variable for the UDN design and used the bandwidth as one example objective function for network resource optimization. The proposed optimization design lays out a foundation for network optimization with

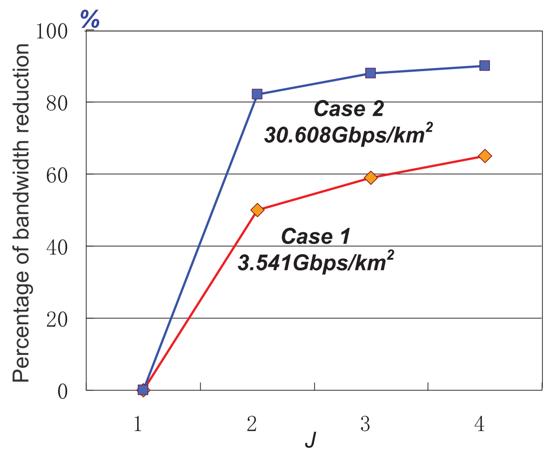

FIGURE 9. The bandwidth reduction of using multiple subnets relative to one macro subnet.

other types of system costs. Further, the users may have different traffic types with different quality of service requirements. The proposed framework can be also expanded to include the quality-of-service (QoS) requirements of different traffic types, which might be impacted by user mobility.

REFERENCES

[1] (Oct. 2014). *4G Americans' Recommendations on 5G Requirements and Solutions*. [Online]. Available: http://www.4gamericas.org/index.cfm?fuseaction=page&sectionid=428

[2] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 Gbps/UE in cellular systems: Understanding ultra-dense small cell deployments," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2078–2101, 4th Quart., 2015.

[3] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and devices for the 5G era," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 90–96, Feb. 2014.

[4] A. Damnjanovic *et al.*, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 10–21, Jun. 2011.

[5] A. Ghosh *et al.*, "Heterogeneous cellular networks: From theory to practice," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 54–64, Jun. 2012.

[6] (Feb. 2009). *3GPP LTE Standard: 3GPP Rel-8 Beyond*. [Online]. Available: http://www.3GAmericas.org

[7] *Deploying High Capacity Dense Small Cell Heterogeneous Networks, DECADE Project*, European Commission, Brussels, Belgium, Aug. 2015.

[8] J. Zhu, N. Deng, and M. Zhao, "A novel frequency reuse scheme for future wireless networks," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Sep. 2015, pp. 1–6.

[9] J. Zhu, N. Deng, M. Zhao, and Y. Chen, "A unified frequency reuse framework for heterogeneous cellular networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–7.

[10] A. Mehbodniya, R. Yoneya, and F. Adachi, "A study of energy- and spectral-efficiency for dense HetNet scenario with non-unifom BS and UE distribution," in *Proc. IEEE VTS Asia–Pacific Wireless Commun. Symp. (APWCS)*, Aug. 2016, pp. 489–492.

[11] H. Munir, S. A. Hassan, H. Pervaiz, Q. Ni, and L. Musavian, "Resource optimization in multi-tier HetNets exploiting multi-slope path loss model," *IEEE Access*, vol. 5, pp. 8714–8726, 2017.

[12] C.-H. Liu, D.-C. Liang, P.-C. Chen, and J.-R. Yang, "Coverage analysis for dense heterogeneous networks with cooperative NOMA," in *Proc. IEEE 85th Veh. Technol. Conf.*, Jun. 2017, pp. 1–6.

[13] J. Liu, W. Xiao, C.-L. I, C. Yang, and A. Soong, "Ultra-dense networks (UDNs) for 5G," *IEEE 5G Tech Focus*, vol. 1, no. 1, pp. 6–10, Mar. 2017.

[14] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2522–2545, 4th Quart., 2016.

[15] C.-X. Wang *et al.*, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[16] S. F. Yunas, M. Valkama, and J. Niemelä, "Spectral and energy efficiency of ultra-dense networks under different deployment strategies," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 90–100, Jan. 2015.

[17] H. Zhang, C. Jiang, M. Bennis, M. Debbah, Z. Han, and V. C. M. Leung, "Heterogeneous ultra-dense networks: Part 1," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 68–69, Dec. 2017.

[18] P. Kela, J. Turkka, and M. Costa, "Borderless mobility in 5G outdoor ultra-dense networks," *IEEE Access*, vol. 3, pp. 1462–1476, 2015.

[19] S. Chen, F. Qin, B. Hu, X. Li, and Z. Chen, "User-centric ultra-dense networks for 5G: Challenges, methodologies, and directions," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 78–85, Apr. 2016.

[20] A. Gotsis, S. Stefanatos, and A. Alexiou, "UltraDense networks: The new wireless frontier for enabling 5G access," *IEEE Veh. Technol. Mag.*, vol. 11, no. 2, pp. 71–78, Jun. 2016.

[21] H. Wang, S. Chen, M. Ai, and H. Xu, "Localized mobility management for 5G ultra dense network," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8535–8552, Sep. 2017.

[22] 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.

[23] H. Zhang, L. Song, Y. Li, and G. Y. Li, "Hypergraph theory: Applications in 5G heterogeneous ultra-dense networks," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 70–76, Dec. 2017.

[24] J. An, K. Yang, J. Wu, N. Ye, S. Guo, and Z. Liao, "Achieving sustainable ultra-dense heterogeneous networks for 5G," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 84–90, Dec. 2017.

[25] M.-S. Alouini and A. J. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1047–1066, Jul. 1999.

[26] A. Mahmud and K. A. Hamdi, "A unified framework for the analysis of fractional frequency reuse techniques," *IEEE Trans. Wireless Commun.*, vol. 62, no. 10, pp. 3692–3705, Oct. 2014.

[27] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Coverage and ergodic rate in K-tier downlink heterogeneous cellular networks," in *Proc. 49th Annu. Allerton Conf.*, Sep. 2011, pp. 1627–1632.

[28] V. Capdevielle, A. Feki, and A. Fakhreddine, "Self-optimization of handover parameters in LTE networks," in *Proc. Int. Symp. WiOpt*, May 2013, pp. 133–139.

[29] J. Zhu, "Capacity-power consumption and energy-efficiency evaluation of green wireless networks," *China Commun.*, vol. 9, no. 2, pp. 13–21, 2012.

[30] P. P. Dey, S. Chandra, and S. Gangopadhaya, "Speed distribution curves under mixed traffic conditions," *J. Transp. Eng.*, vol. 132, no. 6, pp. 84–90, Jun. 2006.

JINKANG ZHU (M'92–LM'18) received the B.S. degree in electrical engineering from Sichuan University, China, in 1966. He joined the University of Science and Technology of China in 1966, where he has been a Professor since 1992. He has been committed to research on wireless mobile communications and networks, signal processing for communications, and the future wireless technologies.

MING ZHAO (M'12) received the B.E., M.E., and Ph.D. degrees from the University of Science and Technology of China (USTC), Hefei, China, in 1999, 2002, and 2018, respectively, all in electronic engineering and information science. He is currently a Lecturer with the Department of Electronic Engineering and Information Science, USTC. His research interests include non-orthogonal multiple access, heterogeneous networks, and green communications.

SHENGLI ZHOU (S'99–M'03–SM'11–F'14) received the B.S. and M.Sc. degrees in electrical engineering and information science from the University of Science and Technology of China, Hefei, China, in 1995 and 1998, respectively, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, USA, in 2002. He is currently a Full Professor with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, USA.

His general research interests lie in the areas of wireless communications and signal processing. He received the 2007 ONR Young Investigator Award and the 2007 Presidential Early Career Award for Scientists and Engineers. He was an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2005 to 2007, the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2008 to 2010, and the IEEE JOURNAL OF OCEANIC ENGINEERING from 2010 to 2016.