Identifying Coverage Holes: Where To Densify?

Rebal Jurdi*, Jeffrey G. Andrews*, Dave Parsons†, and Robert W. Heath Jr.*
*Department of ECE, The University of Texas at Austin
†Crown Castle International
rebal@utexas.edu, {jandrews, rheath}@ece.utexas.edu, dave.parsons@crowncastle.com

***Abstract*—We develop a methodology to identify coverage holes in cellular markets, given demographic and socio-economic information of the population, detailed knowledge of the current macro-cellular infrastructure, and the spectrum capabilities of a cellular operator. Coverage holes are defined as areas where users contending for macro BS resources outnumber those who eventually get serviced with minimum acceptable average rates. We develop a method that begins with filtering out users who connect to the Internet through Wi-Fi or to their data networks through small cell BSs. Since the rate observed by a user is a function of their own SINR as well as that of other connected users, we generate an SINR *heat map* and use a fair scheduling algorithm that allows us to determine the effective load on a BS as a number of users. By comparing the cell load to the number of users with acceptable rates, we obtain a *coverage map* that identifies holes that should be targeted by further densification. We observe that more cells will be in major deficit ten years from now, with up to 25% of an operator's cells requiring more than four additional micro cell BSs to fill the gap. Finally, we give a list of extensions of this work.**

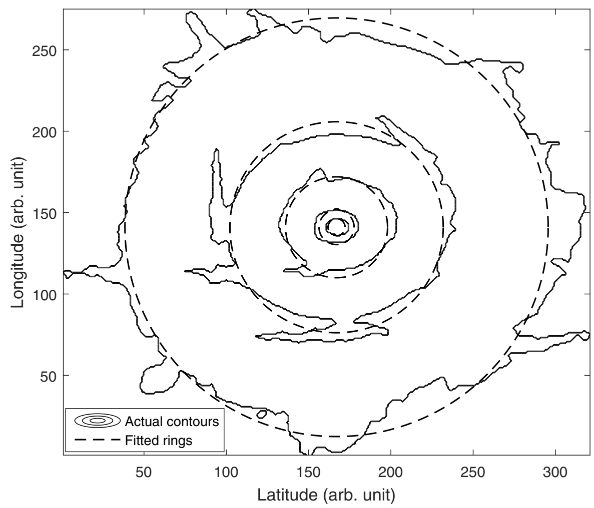

Fig. 1. Contours representing raw power attenuation plotted against power law path loss level contours at -50 down to -90 dB.

I. Introduction

It is unclear whether technological advancements will continue to accommodate the growth in demand on mobile data networks for the next ten years [1]. This deficit corresponds to the amount of network infrastructure densification required to bring back balance to the per-operator mobile network capacity "supply and demand". In this paper, we develop a methodology for predicting the size of the gap between network capabilities and user demand in 2025 *on a cell-by-cell basis*. The gap is precisely given as the difference between the number of users incident on the resources of the macro BS, and the number of users served with minimum acceptable average rates. Our definition stems from our assumption that small cells will be the first line of defense against growing user demand on data [2]. This demotes traditional macro BSs to a supporting role: "providing [additional] capacity when and where needed [3]".

Our forecast is driven by real infrastructure, spectrum, population, and enterprise data. However, our work is novel not only in its data-driven aspect, but also in that its outcome is a forecast made on a spatially granular scale in a number of cellular market areas (CMAs) rather than aggregated over an entire country. Armed with spatial demographic and economic statistics as well as current cellular infrastructure location and spectrum information, we can identify regions within every market where densification is required to fill the gap. We first evaluate the current capability of an operator's network in a given market by producing a *coverage map*, a tessellation that divides the cellular market into cells with different gap sizes. This evaluation is based on an operator's known BS deployments and spectrum holdings, level of household, neighborhood, and enterprise small cell penetration, and efficiency of radio access technologies. This will be the baseline. Next, we produce another coverage map that forecasts network capacity ten years from now with the introduction of 5G, addition of high-frequency spectrum, and increased small cell densification. Comparing the two maps highlights cells in which future user demand outweighs the cell's resource supply, i.e. cells which require further densification.

This work was supported by Crown Castle International.

II. Available Data

We list the three main categories of data that were used in our forecast.

a) BS information: Includes actual locations of present-day macro towers and distributed antenna system (DAS) nodes. Every tower/DAS node is associated to a heat map that depicts power attenuation in its vicinity. Power attenuation values are function evaluations of appropriate path loss models (Okumara-Hata-like models) that take into consideration BS antenna height and local topography. Moreover, values in this

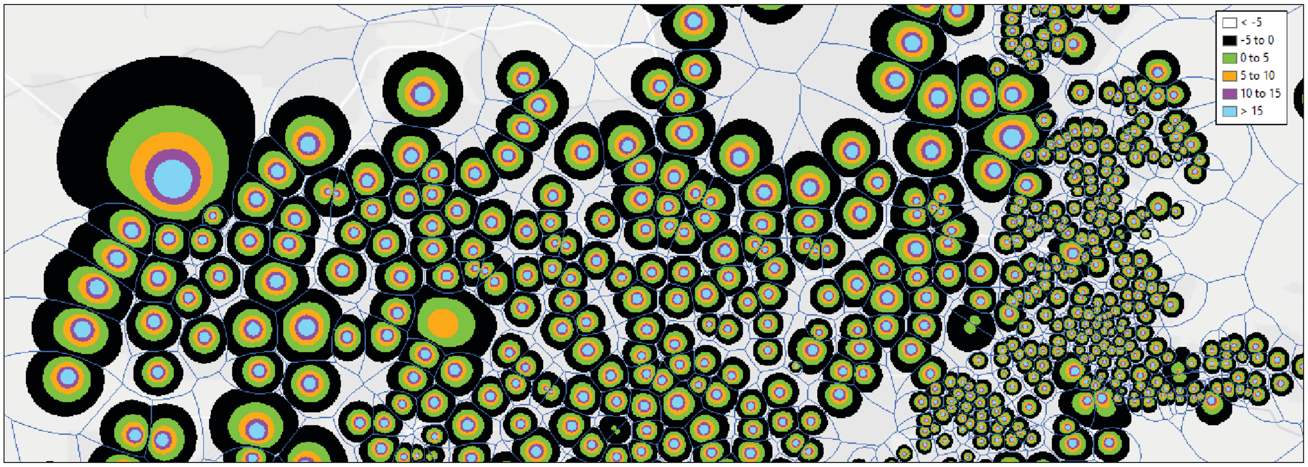

Fig. 2. SINR heat map for Phoenix metropolitan area over 30 km by 70 km area. SINR values are mapped to six quantization regions with cutoffs at -5, 0, 5, 10, and 15 dB.

dataset are quantized to five levels, -50 dB to -90 dB in jumps of 10 dB. To extend the reach of the power received from a BS to an arbitrary location, we resort to fitting the five-step staircase with a power law path loss function as seen in Figure 1. This is a key maneuver to obtain a computationally tractable SINR model whose importance will be addressed in the coming section.

b) Demographic tapestry: Contains population information for census block groups (BGs) such as population density and number of households. A BG is a geographical unit representing a residential neighborhood with a population ranging from a few hundreds to a few thousands, typically. Additionally, this dataset classifies BGs into different attributes based on demographic and socio-economic characteristics, as well as by the type of developed environment (for e.g., urban, suburban, and rural), also known as *urban-rural classification*. The last two attributes are key to our forecast, and we shall refer to those as the lifestyle indicator and population indicator.

c) Enterprise location: Describes the location of businesses and academic institutions scattered around the market area, with information such as number of occupants. These enterprises are in no way exhaustive, but they are assumed to be inclusive of businesses with a substantial number of employees.

III. Methodology

A. Computing Empirical SINR

The data rate is tightly related to the SINR seen by the user's device. Consequently, a critical step in predicting the size and location of the supply-demand gap is generating an *SINR coverage heat map* that represents the SINR seen by a user at an arbitrary location in the market, as shown in Figure 2. Practically speaking, a user's rate under a fair scheduling policy depends not only on their own SINR, but also on the SINR of other users who are being served. One way of obtaining the SINR of connected users is generating a pair of uniformly distributed coordinates and indexing the SINR raster with those coordinates, but a more computationally efficient way is sampling from an SINR distribution. Therefore, we obtain from the SINR map an empirical SINR distribution for every cell via histogram fitting. We observe that SINR follows a right-skewed distribution that fits relatively well to that of a negative-offsetted Gamma random variable.

B. Accounting for the Effect of Heterogeneous Elements

To accurately account for the number of users requesting resources from the macro BS, we devise a method to filter out users whose connections reduce the load on the BSs. We shall refer to these types of connections as *alternative* connections. This category of users includes those who choose non-mobile broadband Internet connection (e.g. Cable or FTTH paired with Wi-Fi), those whose mobile data is offloaded through Wi-Fi, and those who connect to the data network through small cells. Given enterprise location and BG information in a given cell as in Figure 3, we estimate the number of users who connect to the Internet through indoor enterprise or household small cells or Wi-Fi APs. We denote the number of users within the kth cell pursuing data connections from home as H_k, workplace as W_k, and at a public venue or while in transit as X_k. We also denote by the constants $H_k^{\max}$ and $W_k^{\max}$ the maximum number of users that can be present at home or work. We use BG population information to determine a value for $H_k^{\max}$, and business information to give a value for $W_k^{\max}$. We define $H_k(t)$ to be the number of users within the kth macro cell trying to connect from home at some time t, and we similarly define $W_k(t)$ and $X_k(t)$. Finally, we make a simplifying assumption that balances the number of users of different categories, in the different macro cells covering the cellular market area:

$$H_k(t) + W_k(t) + X_k(t) = \max(H_k^{\max}, W_k^{\max}). \qquad (1)$$

This is similar to saying, users who live inside a macro cell work and travel only inside that macro cell, and no users enter the cell from outside. Now let $\alpha_H(t)$ be the fraction of people at home, and $\alpha_W(t)$ the fraction of employees at work.

Suppressing the subscript k, we get the following breakdown of users:

$$\begin{aligned} H(t) &= \alpha_H(t)H^{\max},\ W(t) = \alpha_W(t)W^{\max}, \\ X(t) &= (1-\alpha_H(t))H^{\max} - \alpha_W(t)W^{\max}, \end{aligned} \tag{2}$$

We further reduce $H(t)$, $W(t)$, and $X(t)$ by the factors β_H, β_W, and β_X that represent the fraction of users with alternative connections at home, work, or public places. Additionally, we denote the fraction of users with alternative connections at work in BG b as $\beta_{H,b}$, and the number of users in BG b trying to connect from home as $H_b(t)$. We can now give β_H as

$$\beta_H H(t) = \sum_b \beta_{H,b} H_b(t), \tag{3}$$

where b ranges over the BGs inside the cell. Of course, $\beta_{H,b}$ is directly correlated to the level of small cell and Wi-Fi penetration in households of BG b, which in turn is strongly dependent on factors such as average household size, average family income, total number of households, and urban-rural classification. β_W can be similarly defined, but with respect to other criteria such as institution size, type of business, and whether the institution is a likely adopter of new technology. As for β_X, it has a more explicit definition

$$\beta_X X(t) = U \sum_b A_b D_{b,P} D_{b,L} D^{\max}, \tag{4}$$

where U is the number of users that can be simultaneously served by one small cell, A_b is area of the part of BG b overlapping with the cell, $D^{\max}$ is the maximum possible small cell density, $D_{b,P}$ is a density multiplier based on population indicator, and $D_{b,L}$ is a density multiplier based on lifestyle indicator. Finally, the number of users *trying* to access the data network through the macro BS is

$$M(t) = (1-\beta_H)H(t) + (1-\beta_W)W(t) + (1-\beta_X)X(t).$$

We shall refer to this group of users as *contending users*.

For simulation purposes, we set $D^{\max} = 9$ /km^2 (i.e. about one outdoor SC node every three 100 m $\times$ 100 m city blocks) and $U = 3$ for current-year capacity estimation, and $D^{\max} = 100$ /km^2 (i.e., one node every block) and $U = 5$ for future-year forecast. We set $\alpha_H = 0.2$ and $\alpha_W = 0.8$ for morning hours, $\alpha_H = 0.8$ and $\alpha_W = 0.2$ for evening hours, and $\beta_W = 0.2$ at any time. The density multipliers $D_{b,P}$ and $D_{b,L}$ takes values between 0 and 1 and have a direct relationship with population and lifestyle.

C. Computing Gap Size

Having determined the number of users contending for the resources of the BS serving the cell, we now proceed to determine the fraction that are served with satisfactory average rates to ultimately compute the gap. First of all, we budget for a user's data consumption during peak hour; we assume that a user is conscious of their data consumption and that they are able to ration their daily and hourly quota under a monthly data plan. In other words, we assume that the BS has a queue for every user, with a size equal to their hourly budget.

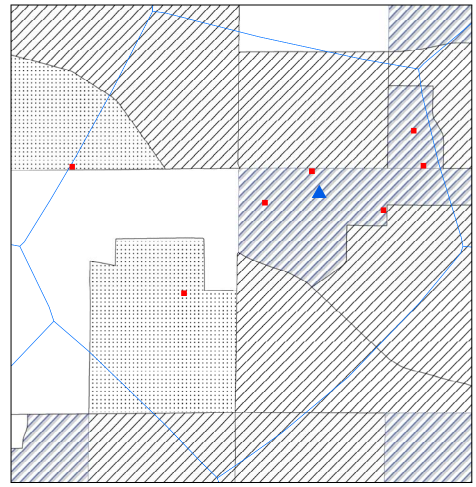

Fig. 3. Block groups shown by urban zone classification inside the boundary of a macro cell. Different hatchings correspond to different population densities. Blue triangle indicates the location of BS. Red squares indicate the locations of select businesses.

Next, we consider a traffic model that emulates the arrival of download requests to the BS scheduler. We start with N users who *jointly* access the network over the course of an hour and have a quota of B MBs, each. We consider a slotted access mechanism that divides time into R slots of duration T. Every user *expects* to receive their data through packets of size P at most once per slot. To emulate the event of users accessing the network from different areas of the cell, every user is associated to an SINR level that is drawn according to the empirical distribution.

Lastly, we run a proportional-fair-like algorithm to schedule these users. A proportional-fair algorithm is a scheduling algorithm that compromises between maximizing total throughput and maximizing the number of users given an acceptable level of service [4]. If a user receives *most* of their data within an hour, they are added to the pool of served users. We refer to the eventual number of users in the pool as *supported users*. Note that we assume all users have the same category of mobile devices, i.e. we assume everyone can receive the same peak rate. This assumption could easily be lifted by assigning different mobile devices to users at random. The size of the gap is thus computed as the difference between the number of contending users and the number of supported users. Cells with positive gaps correspond to coverage holes the size of the cell area. These cells become targets for small cell or micro cell densification, i.e. low-power radio access nodes which create cells that do not reshape the existing macro cell tessellation.

IV. Coverage Map and Gap

Now, we show an application example of the proposed methodology to generate coverage maps that show the *present* coverage holes in the networks of three major operators in one of the biggest US markets. One of these maps is shown in Figure 4. Along with these maps, we also obtained the number of cells for different gap sizes for the current year

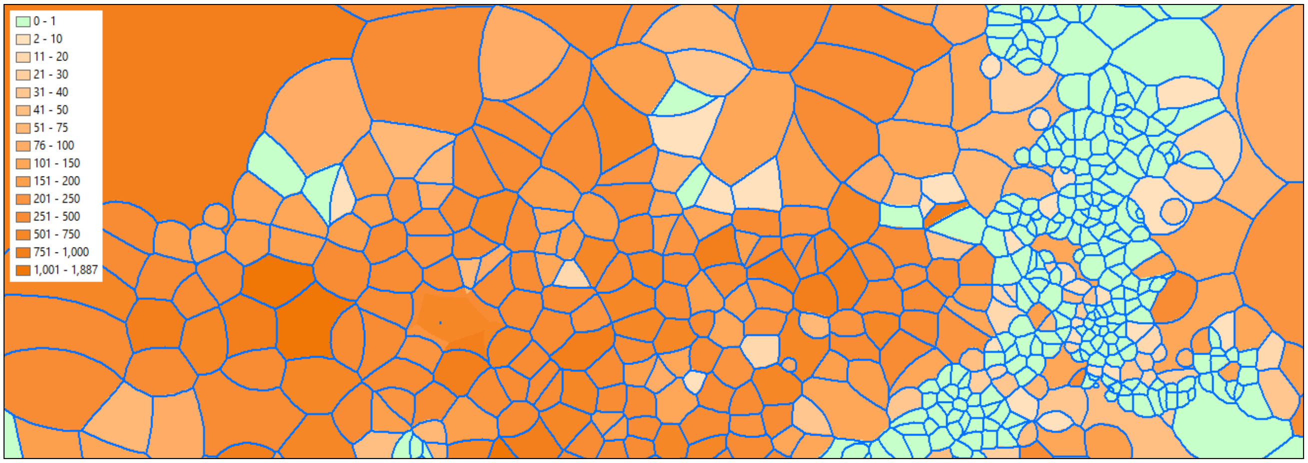

Fig. 4. Coverage map showing a tessellation of cells created by BSs (mostly macros) over the same market shown in Figure 2. Shades of orange correspond to a cell deficit, i.e. more contending users than supported users. Darker shades mean bigger deficit. Light green corresponds to a neutral/possible surplus.

and for 2025; see Figure 5. The first observation is that more cells will be in deficit ten years from now. Operator A has more than 60% cells in deficit, and Operator B and Operator C have more than 90% cells in deficit. We note that, for this particular market, Operator A has almost double the number of BSs of Operator B and Operator C combined. One possible way of filling these coverage holes is through the addition of strategically-placed, low-power micro cells which maintain the original macro cell boundaries. We observe that less than 4% of Operator A's cells require more than 4 micro sites to be added, and up to 25% of Operator B's cells and 35% of Operator C's cells require more than 4 micro sites to be added.

V. Conclusion

A. Summary

We have described a methodology to forecast the gap between user demand and network capability in ten years on a spatially granular scale, given information on BS locations and on the distribution and characteristics of census block groups. The gap is the difference between users contending for the resources of the BS serving the macro cell and users that receive acceptable service; the bigger the gap, the bigger the deficit in resources. Gaps are presented visually through the coverage map, a tessellation of cells that are shaded according to the size of their gap. We compared the percentage of cells experiencing deficit for three operators in one cellular market. A major observation is that more cells will be in deficit by 2025. We determined not only how much deficit there will be, but where this deficit will be.

B. Future Work

a) Improved SINR map: The cornerstone to our prediction was the SINR coverage map whose values were computed assuming no BS co-operation on the downlink. Future work can consider contemporary methods for interference reduction, which can diminish low SINR regions [5], [6] and improve data rates.

b) Addition of resources: On the one hand, our forecast accounted for up to seven-fold growth in data consumption. On the other hand, it accounted for up to ten-fold increase in the number of small cells, and a modest two-fold increase in peak data rates supported by a mobile device. It is interesting to determine how much capacity is further opened up by the addition of UHF spectrum (for e.g., 600 MHz) and macro BSs.

c) Massive MIMO: Massive MIMO promises enhanced reliability, reduced interference, and increased data rates [7]. Currently, trials and demos are being conducted as a prelude to the mass deployment of this technology. It is important to carefully model how a Massive MIMO BS serves connected users to get a more accurate figure for the number of supported users in the future.

d) Small cell effect: Although coverage by 2025 will be extended through already existing macro cells and an additional layer of small cells, our modeling treated these two elements differently. Apart from the fact that small cell location information is unavailable or even non-existing, the main reason for the different treatment is ensuring computational tractability: Small cells tessellating a market would immensely outnumber macro cells tessellating the same market. Hence, it would be computationally inefficient to compute the SINR distribution, collect demographic statistics, generate user traffic and run a scheduling algorithm *for every* small cell. This drove us to examine the net effect of indoor, enterprise, and outdoor/public small cells on reducing the number of contenders within the greater macro cell. It is interesting to observe how robust this net effect is to changing the density of outdoor small cells (see (4)) and the number of users supported per cell, to within reasonable numbers.

e) Accurate traffic model: User-generated traffic is a crucial quantity to understand since it determines how much network capacity is needed. Our system level simulation of user traffic assumed that a pre-determined number of users, referred to as contending users, have identical download request processes; download requests from each user were independent and

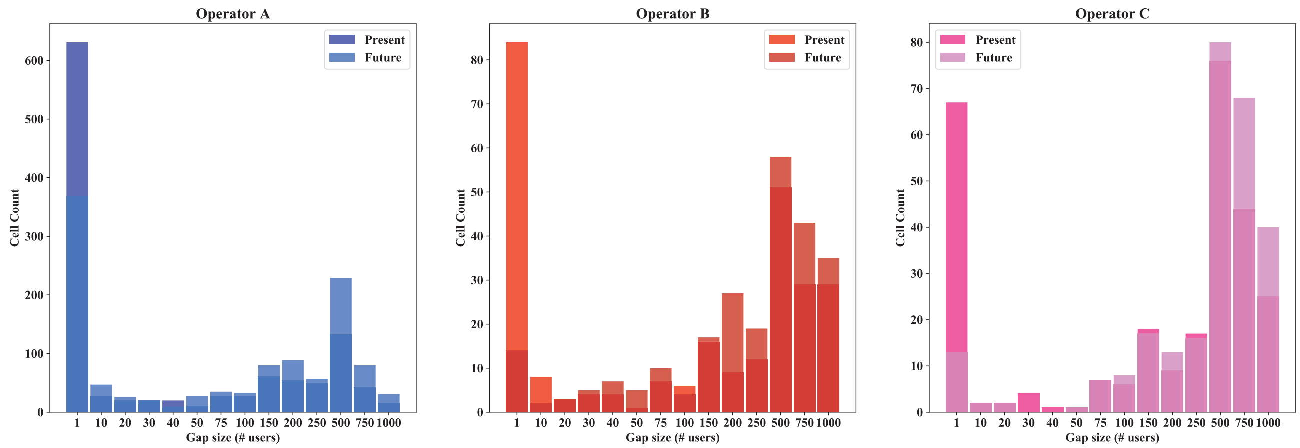

Fig. 5. Histogram showing the present and future number of cells per gap size for three major operators serving a US metropolitan area.

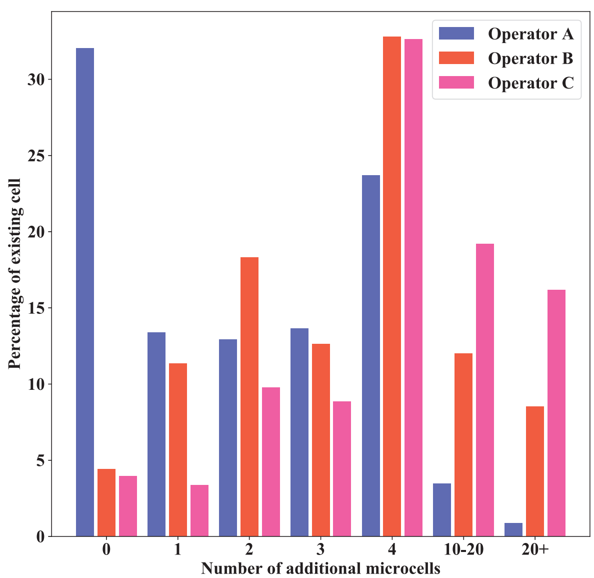

Fig. 6. Histogram showing the present and future number of cells per gap size for three major operators serving a US metropolitan area.

uniformly distributed over the access period. Alternatively, one can consider different arrival processes: a process modeling a long session of light web browsing, and another modeling a session of sparse download activity mimicking short video streaming. A more general traffic model could allow users to access the network during non-overlapping time windows which reduces contention over the scheduler's resources and gives users higher data rates.

f) Unified analytical framework: Traffic could be studied using a different framework altogether that permits analyzing quantities such as average user rate. Previous work took the analytical route to evaluate throughput and capacity [8] or balance cell loads [9] under a spatial *elastic traffic* model. In that model, user demand is assumed uniform over space, and file transfer requests are assumed to follow Poisson processes. In [10], the sum rate of a network comprised of macro and small cells was computed under the empirical assumption that instantaneous traffic within a macro cell was Gamma distributed. Even though these papers marry a traffic model with a methodology to analyze system performance, it is not clear how versatile these frameworks are when different heterogeneous elements are added and different spatial traffic characteristics.

References

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[2] A. Ghosh, J. G. Andrews, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. A. Thomas, P. Xia, H. S. Jo, H. S. Dhillon, and T. D. Novlan, "Heterogeneous cellular networks: From theory to practice," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 54–64, June 2012.

[3] "White paper: Ten key rules of 5G deployment," Nokia, Tech. Rep. C401-0120041-WP-201609-1-EN, 2016. [Online]. Available: http://resources.alcatel-lucent.com/asset/200006

[4] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[5] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," vol. 28, no. 9, pp. 1380 –1408, Dec. 2010.

[6] X. Zhang and M. Haenggi, "A stochastic geometry analysis of inter-cell interference coordination and intra-cell diversity," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6655–6669, Dec. 2014.

[7] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, February 2014.

[8] T. Bonald and A. Proutière, "Wireless downlink data channels: User performance and cell dimensioning," in *Proc. MobiCom*. ACM, 2003, pp. 339–352.

[9] H. Kim, G. de Veciana, X. Yang, and M. Venkatachalam, "Distributed $alpha$ -optimal user association and cell load balancing in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 1, pp. 177–190, Feb 2012.

[10] K. Sakaguchi, G. Tran, H. Shimodaira, S. Nanba, T. Sakurai, K. Takinami, I. Siaud, E. Strinati, A. Capone, I. Karls, R. Arefi, and T. Haustein, "Millimeter-wave evolution for 5G cellular networks," *IEICE Trans. Commun.*, vol. E98–B, no. 3, pp. 388–402, March 2015.