On the Efficiency tradeoffs in User-Centric Cloud RAN

Umair Sajid Hashmi*, Syed Ali Raza Zaidi†, Arsalan Darbandi* and Ali Imran*
*School of Electrical and Computer Engineering, University of Oklahoma, Tulsa, OK, USA
†School of Electronic and Electrical Engineering, University of Leeds, Leeds LS2 9JT, U.K.

***Abstract*—Ambitious targets for aggregate throughput, energy efficiency and ubiquitous user experience are propelling the advent of ultra-dense networks. Intercell interference and high energy consumption in an ultra-dense network are the prime hindering factors in pursuit of these goals. To address the aforementioned challenges, in this paper, we propose a novel user-centric network orchestration solution for Cloud RAN based ultra-dense deployments. In this solution, a cluster (virtual disc) is created around users depending on their service priority. Within the cluster radius, only the best remote radio head (RRH) is activated to serve the user, thereby decreasing interference and saving energy. We follow a stochastic geometry based approach to quantify the area spectral efficiency (ASE) and RRH power consumption models to quantity energy(EE) efficiency of the proposed user-centric Cloud RAN (UCRAN). Through extensive analysis, we observe that the cluster sizes that yield optimal ASE and EE are quite different. Subsequently, we propose a game theoretic self-organizing network (GT-SON) framework that can orchestrate the network between ASE and EE focused operational modes in real-time in response to changes in network conditions and the operator's revenue model, to achieve a Pareto optimal solution. A bargaining game is modeled to investigate the ASE-EE tradeoff through adjustment in the exponential efficiency weightage in the Nash bargaining solution (NBS). Results show that compared to current non user-centric network design, the proposed solution offers the flexibility to operate the network at multiple folds higher ASE or EE along with significant improvement in user experience.**

***Index Terms*—User-centric architectures, Cloud RAN, Poisson Point Process, Area Spectral Efficiency, Energy Efficiency, Nash Bargaining Solution**

I. Introduction

Cell-free user-centric networks are envisioned as enablers for interference management in ultra-dense 5th generation (5G) cellular networks. In particular, signal degradation for cell-edge users that is considered a limiting factor in LTE is addressed by structural evolution of the 5G networks designed from the users' (rather than base stations') perspective [1]. Operationally, each served user within a user-centric network is connected to one or more small cells in the vicinity defined by an elastic virtual user-centric cell boundary [2][3]. The virtual user-centric cell size is adaptable with respect to user traffic, channel environment and quality of service (QoS) requirements.

While the 5G systems target multiple fold increase in data rate, millisecond level latency and support for up to 500 km/hr user mobility; all this must be achieved with an improvement in spectral efficiency and reduced operational costs [4]. One enabling technology to meet these goals is Cloud-RAN (C-RAN) which is based on separating Baseband Units (BBUs) from the radio access units [5][6]. The BBUs are migrated to the cloud forming a BBU pool for centralized processing and resource allocation. C-RAN provides the network scalability for large scale remote radio head (RRH) deployment in dense networks at lower operational costs.

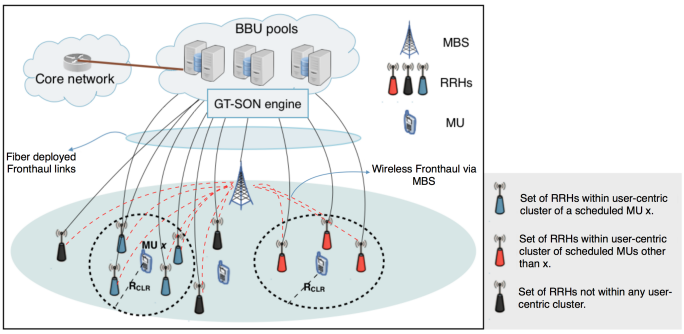

Fig. 1. User-centric C-RAN architecture

User-centric virtual cell approach coupled with centralized baseband processing via C-RAN deployment is an ideal merger to meet 5G's ubiquitous user experience targets within realistic energy and cost constraints. Fig.1 provides a graphical illustration of a User-centric Cloud RAN (UCRAN) network with virtual user-centric cell boundaries. The RRHs are connected to the pool of BBUs via flexible front haul. The front haul is usually an optical fiber where signaling is done using radio-over-fiber (RoF) or common public radio interface (CPRI) [6]. Most of the signal processing at baseband level is delegated to the BBUs. The key idea here is to dynamically select the best RRH within a circular area (virtual cell) with a pre-defined radius R_{CLR} around users selected for downlink transmission during each scheduling interval (used interchangeably with time slot and TTI). All other RRHs within the circle here after called cluster are kept OFF thereby minimizing the interference. The aforementioned UCRAN architecture provides two-fold benefits: i) on-demand centralized processing at the BBU pools caters to non-uniform user traffic that subsequently enables OPEX reduction by as much as 30% [7], ii) user-centric RRH clustering reduces the number of nearby interfering RRHs and eliminates cell-edge coverage issues, hence improving the overall user experience regardless of user location and movement [8]. The game theoretic self-organizing (GT-SON) engine in fig.1 enables dynamic adaptation of R_{CLR} in order to either enhance the overall system throughput or the energy efficiency (EE). The cluster size selection is dependent upon the

network operator's spatio-temporal revenue model which may include traffic intensity, time of the day and hotspot locations (e.g. cafes, stadiums) [9].

The key research question at hand is determining the optimal cluster size around a scheduled user. The cluster size $\mathcal{C} = \pi R_{CLR}^2$ determines the interference free region around each scheduled user. Increasing the RRH cluster size offers the following gains: 1) larger distances between MUs and interfering RRHs results in larger link SINR and thus, better link throughput, and 2) a larger RRH cluster size yields more macro diversity gain or cooperative gain through selection or cooperation among larger number of RRHs in the cluster, respectively. However, the downside of a larger RRH cluster size is reduced spectrum reuse and reduced number of MUs that can be served simultaneously which, in turn, reduces system level capacity. Hence with a larger cluster, there are fewer high quality links as opposed to many low bit-rate links (which occur with a smaller cluster).

In the back drop of these insights the goal of this paper is to investigate following research questions: 1) What are the optimal RRH cluster sizes that maximize KPIs of capacity (in terms of area spectral efficiency (ASE)), energy efficiency (EE) and the user quality of experience (QoE)? 2) Given that the optimal RRH cluster sizes for all the three KPIs is expected to be different, how to design the pareto-optimal solution that achieves the desired balance among aforementioned KPIs?

Related works address transmit power control [10], interference alignment [11], dynamic load balancing [12] and optimal cluster dimensioning strategies [8][13] in user-centric networks. [14] proposes a learning based approach to cluster RRHs within BBUs with the aim to enhance utilization rate and energy efficiency in C-RANs. Similarly, [15] adapts an efficient resource allocation scheme to improve energy efficiency in large scale C-RAN deployments. Through this work, we add another dimension to existing literature by simultaneously investigating the intertwined KPIs, i.e. ASE, EE and QoE in ultra-dense user-centric networks. The contributions and findings of this work are summarized as follows:

A. Contributions and Organization

- By employing well established stochastic geometry principles [16], we characterize the ASE and EE of a UCRAN system as a function of the mobile user (MU) and RRH deployment distributions. The analytical model takes into account both MU and RRH thinning arising from the user-centric RRH clustering performed in the centralized BBU pools during each TTI.
- The ASE-EE tradeoff in a UCRAN is modeled through a two-player bargaining problem. The performance metrics are modeled as virtual game players and a Nash bargaining solution is found that corresponds to a unique optimal cluster radius for a given set of network parameters.
- Based on our analysis, we advocate the gains of dynamic adaptation of the ASE-EE tradeoff by integrating a GT-SON engine within the BBU pools. Through an exponential weightage parameter, the GT-SON engine shifts the operator's preference between ASE and EE while ensuring higher SINR gains within a particular spatio-temporal zone.

The rest of the paper organization is as follows: in section II, we describe the spatial model, user-centric RRH clustering and the radio propagation model assumed in this work. Section III focuses on the analytical derivation of the ASE and EE of the UCRAN. In section IV, we present the proposed GT-SON model for adaptive cluster size adjustment based on network parameters and the operators' revenue model. Efficiency tradeoff analysis is performed through extensive simulations in section V. The paper closes with conclusions and future research directions in Section VI.

II. System Model

A. Spatial Model

We consider the downlink of a two-tier UCRAN consisting of one central macro base station (MBS) that has RRHs and MUs spatially distributed across its foot-prints. We model the spatial distributions of RRHs and MUs using two independent stationary Poisson point processes (SPPPs) $\Lambda_{RRH} \in \mathbb{R}^2$ and $\Lambda_{MU} \in \mathbb{R}^2$ with intensities λ_{RRH} and λ_{MU} respectively. Specifically, at an arbitrary time instant, the probability of finding $n_i \in \mathbb{N}, i \in \{RRH, MU\}$ RRHs / MUs inside a typical macro-cell with area foot-print $\mathcal{A} \subseteq \mathbb{R}^2$ follows the Poisson law with mean measure $\lambda_i v_2(\mathcal{A})$. The mean measure is characterized by the average number of RRHs / MUs per unit area (i.e., $\lambda_{RRH} \backslash \lambda_{MU}$) and the Lebesgue measure [18] $v_2(\mathcal{A}) = \int_{\mathcal{A}} d\boldsymbol{x}$ on $\mathbb{R}^2$, where if $\mathcal{A}$ is a disc of radius r then $v_2(\mathcal{A}) = \pi r^2$ is the area of the disc.

B. Channel Model

We model $h_{\boldsymbol{xy}} l(\|\boldsymbol{x}-\boldsymbol{y}\|)$ as the channel between an arbitrary MU $\boldsymbol{x} \in \Lambda_{MU}$ and an RRH $\boldsymbol{y} \in \Lambda_{RRH}$. Here $h_{\boldsymbol{xy}} \sim \varepsilon(1)$ is a unit mean exponential random variable that captures the effect of the small-scale fading between the MU and RRH as Rayleigh-distributed fading channel. In order to account for the large-scale fading we denote $l(|\boldsymbol{x} - \boldsymbol{y}|) = |\boldsymbol{x} - \boldsymbol{y}|^{-\alpha}$ where $|\boldsymbol{x} - \boldsymbol{y}|$ is the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ and $\alpha \geq 2$ is the path loss exponent. It is assumed that the same transmit power P_{RRH} is used for all RRHs.

C. User-centric RRH Clustering

The RRH clustering mechanism in the user-centric C-RAN is envisioned on a scenario where a high service priority MU is served by a RRH that provides the largest signal-to-noise-plus-interference ratio (SINR) within its cluster. Service priority to each MU is assigned using a random probability $p_{MU} \sim U(0, 1)$ which is incremented after every time slot during which the service is deferred to the MU because of presence of one or more higher preference MUs in the surroundings. For simplicity, we assume that each MU in our model is requesting service during all time slots (or TTIs). During each TTI, the GT-SON engine in the centralized BBU pools determines the optimal cluster radius "R_{CLR}" for existing network parameters and the operator's business specifications (e.g. high data rate or high energy efficiency). To avoid interference caused by simultaneous transmissions to nearby MUs, the user-centric

RRH clustering creates repulsion by avoiding spatial overlap between clusters. This implies that during a particular TTI, a scheduled UE will not have any other UE with higher service priority within a radial distance of $2R_{CLR}$. The joint RRH clustering and user scheduling scheme is summarized as algorithm 1. The symbol b(**x**, r) denotes a ball of radius r centered at a point **x**.

Algorithm 1 RRH clustering and MU scheduling algorithm

Inputs: Λ_{RRH}, Λ_{MU}, R_{CLR}
Outputs: Λ'_{RRH}, Λ'_{MU}
1: Initialize the set of scheduled MUs and the RRHs serving within the user-centric clusters at any given time slot as Λ'_{MU}, $\Lambda'_{RRH} \leftarrow \emptyset$.
2: Update Λ'_{MU} and Λ'_{RRH} for the current time slot using the following conditions:
foreach $\boldsymbol{x} \in \Lambda_{MU}$ **do**
  if $\boldsymbol{y} \in b(\boldsymbol{x}, 2R_{CLR})$ *and* $p_{MU}^{\{\boldsymbol{x}\}} > p_{MU}^{\{\boldsymbol{y}\}}$, $\forall \boldsymbol{y} \in \Lambda_{MU}, \boldsymbol{y} \neq \boldsymbol{x}$ **then**
    $\Lambda'_{MU} \cup \{\boldsymbol{x}\}$
    foreach $\boldsymbol{r} \in \Lambda_{RRH}$ **do**
      if $\boldsymbol{r} \in b(\boldsymbol{x}, R_{CLR})$ **then**
        if $h_{rx} l(||\boldsymbol{r} - \boldsymbol{x}||) > h_{r'x} l(||\boldsymbol{r'} - \boldsymbol{x}||), \forall \boldsymbol{r'} \in \Lambda_{RRH}, \boldsymbol{r'} \in b(\boldsymbol{x}, R_{CLR}), \boldsymbol{r'} \neq \boldsymbol{r}$ **then**
          $\Lambda'_{RRH} \cup \{\boldsymbol{r}\}$
        end
      end
    end
  else
    continue.
  end
end
3: Serve all the scheduled users Λ'_{MU} from the associated RRHs and update scheduling priorities $p_{MU}^{\{\boldsymbol{x}\}}$ for all $\boldsymbol{x} \in \Lambda_{MU}$, i.e. increment $p_{MU}^{\{\boldsymbol{x}\}}$ if $\boldsymbol{x} \in \Lambda'_{MU}$ and decrement $p_{MU}^{\{\boldsymbol{x}\}}$ if $\boldsymbol{x} \notin \Lambda'_{MU}$.
4: Go to step 1.

III. Efficiency Metrics in User-centric C-RAN

A. Area Spectral Efficiency

Consider a scheduled user $\boldsymbol{x} \in \Lambda'_{MU}$. Let $\mathcal{S}_{cop}(\boldsymbol{x}, R_{CLR}) = \Lambda'_{RRH} \cap b(\boldsymbol{x}, R_{CLR})$ be the singleton set containing the RRH selected to serve $\boldsymbol{x}$ on the basis of the scheduling criteria (Algorithm 1). Furthermore, let $\Lambda_I = \Lambda'_{RRH} \backslash \mathcal{S}_{cop}(\boldsymbol{x}, R_{CLR})$ be the set of RRHs which are concurrently scheduled to serve $\boldsymbol{y} \neq \boldsymbol{x}, \forall \boldsymbol{y} \in \Lambda'_{MU}$. Let s_x and s_y be the desired and interference signals respectively at an arbitrary MU $\boldsymbol{x}$, then the received signal at $\boldsymbol{x}$ will be

$$\mathrm{r}_x = \sqrt{P_{RRH} \max_{\boldsymbol{i} \in \mathcal{S}_{cop}} h_{\boldsymbol{ix}} l(||\boldsymbol{x} - \boldsymbol{i}||)} s_{\boldsymbol{x}} + \sum_{\boldsymbol{y} \in \Lambda'_{MU}, \boldsymbol{y} \neq \boldsymbol{x}} \sqrt{P_{RRH} \max_{\boldsymbol{j} \in \Lambda'_{RRH} \cap (\boldsymbol{y}, R_{CLR})} h_{\boldsymbol{jx}} l(||\boldsymbol{x} - \boldsymbol{j}||)} s_{\boldsymbol{y}} + \varphi_{\boldsymbol{x}}, \quad (1)$$

where $\max_{\boldsymbol{i} \in \mathcal{S}_{cop}} h_{\boldsymbol{ix}} l(||\boldsymbol{x} - \boldsymbol{i}||)$ is the channel gain between the serving RRH $\boldsymbol{i}$ and the MU $\boldsymbol{x}$, $\max_{\boldsymbol{j} \in \Lambda'_{RRH} \cap (\boldsymbol{y}, R_{CLR})} h_{\boldsymbol{xj}} l(||\boldsymbol{x} - \boldsymbol{j}||)$ is the interference experienced at $\boldsymbol{x}$ due to a RRH $\boldsymbol{j}$ serving another MU $\boldsymbol{y}$, P_{RRH} is the transmit power employed by the RRHs and $\varphi_{\boldsymbol{x}}$ is the additive white Gaussian noise (AWGN) at $\boldsymbol{x}$'s receiver front end. Without loss of generality, we use the Silvnyak's theorem [16] and focus our analysis on the arbitrary MU $\boldsymbol{x}$ assumed to be located at the origin. Since ultra dense small cell networks are generally considered to be interference-limited, we may ignore the AWGN for our analytical analysis and express the signal-to-interference ratio (SIR) at MU $\boldsymbol{x}$ as

$$\Gamma_{\boldsymbol{x}} = \frac{\max_{\boldsymbol{i} \in \mathcal{S}_{cop}} h_{\boldsymbol{i}} l(r_{\boldsymbol{i}})}{\sum_{\boldsymbol{j} \in \Lambda_I} h_{\boldsymbol{j}} l(r_{\boldsymbol{j}})}, \quad (2)$$

where r_i and r_j are the relative distances of MU $\boldsymbol{x}$ with its DL scheduled and interfering RRHs respectively.

The primary hurdle in characterizing the SINR in a UCRAN arises from the fact that unlike Λ_{MU}, the point process of the scheduled MUs Λ'_{MU} is non-stationary. A closer inspection of Λ'_{MU} reveals that it is a modified version of Type II Matern Hard Core process [16]. Therefore, it can be approximated by an equidense SPPP with appropriate modified intensity [17][18] given by

$$\bar{\lambda}_{MU} = \frac{1 - e^{-4\pi\lambda_{MU}^2}}{4\pi R_{CLR}^2}. \quad (3)$$

Once the Λ'_{MU} distribution is characterized, the next step is to characterize the aggregate interference experienced by an arbitrary MU from the activated RRHs outside its user-centric cluster area.

Proposition 1. *The mean of the aggregate interference experienced by a typical MU under user-centric RRH clustering can be approximated as follows:*

$$\mathbb{E}(I) = \frac{2\pi\lambda_{RRH}[1 - \exp(-[1 - \exp(-4\pi\lambda_{MU} R_{CLR}^2)]/4)]}{(\alpha - 2)(R_{CLR})^{\alpha - 2}(\lambda_{RRH}\pi R_{CLR}^2)}, \quad (4)$$

where α is the terrain dependent pathloss exponent.

Proof: Consider the SPPP Λ_{RRH}, then under the user-centric RRH clustering algorithm, for each scheduled MU, only a single RRH which resides in the vicinity as well as provides maximum channel gain to that MU is activated by the MBS. A natural implication of this policy is that the resulting PPP Λ'_{RRH} is non-stationary. However, like Λ'_{MU}, it can be approximated with an equivalent SPPP with modified density $\lambda_{RRH} p_{ACT}$. Here p_{ACT} is the activation probability for the RRH and can be computed as follows:

$$\begin{aligned} p_{ACT} &\overset{(a)}{=} Pr\{\Lambda'_{MU} \cap b(\boldsymbol{r}, R_{CLR}) \neq \emptyset | \boldsymbol{r} \in \Lambda'_{RRH}\}. \\ &\quad Pr\{h_{\boldsymbol{r}} l(r_{\boldsymbol{r}}) > h_{\boldsymbol{j}} l(r_{\boldsymbol{j}}) | \boldsymbol{j} \in \Lambda'_{RRH}, \boldsymbol{j} \neq \boldsymbol{r}\}, \\ &= \left[1 - Pr\{\Lambda'_{MU} \cap b(\boldsymbol{r}, R_{CLR}) = \emptyset | \boldsymbol{r} \in \Lambda'_{RRH}\}\right]. \\ &\quad Pr\{h_{\boldsymbol{r}} l(r_{\boldsymbol{r}}) > h_{\boldsymbol{j}} l(r_{\boldsymbol{j}}) | \boldsymbol{j} \in \Lambda'_{RRH}, \boldsymbol{j} \neq \boldsymbol{r}\}, \\ &= \left[1 - \exp(-\bar{\lambda}_{MU}\pi R_{CLR}^2)\right].(1/[\lambda_{RRH}\pi R_{CLR}^2]), \\ &= \frac{1 - \exp\left(-[1 - \exp\left(-4\pi\lambda_{MU} R_{CLR}^2\right)]/4\right)}{\lambda_{RRH}\pi R_{CLR}^2}, \end{aligned} \quad (5)$$

where (a) follows from the fact that a RRH is only activated if: i) there is a scheduled user within a distance of R_{CLR}, and ii) there is no other RRH within a distance of R_{CLR} from that user providing better channel gain. Now noticing that $\Lambda_I = \Lambda'_{RRH} \backslash \mathcal{S}_{cop}(\boldsymbol{o}, R_{CLR})$, we can precisely describe $\Lambda_I = \Lambda'_{RRH} \backslash b(\boldsymbol{o}, R_{CLR})$. Hence the mean interference can be computed using Campbell's theorem [16] as follows

$$\begin{aligned} \mathbb{E}(I) = \mathbb{E}(I) &= \mathbb{E}\left(\sum_{\boldsymbol{j} \in \Lambda'_{RRH} \backslash b(\boldsymbol{o}, R_{CLR})} h_{\boldsymbol{j}} l(r_{\boldsymbol{j}})\right), \\ &= 2\pi\lambda_{RRH} p_{ACT} \int_{RCLR}^{\infty} \mathbb{E}(H) r^{1-\alpha} dr. \end{aligned} \quad (6)$$

Substituting $\mathbb{E}(H) = 1$ in (6) concludes the proof. ■

Once the interference is characterized, we can approximate the link success probability which represents the percentage of users with adequate link channel quality with the connected RRHs for DL.

Proposition 2. *The link success probability of the probe MU served under the proposed user-centric clustering and RRH selection scheme algorithm can be lower-bound as*

$$\mathbb{P}_{suc_{\boldsymbol{x}}} \geq 1 - \exp\left(-\frac{2\pi\lambda_{RRH}}{\alpha\gamma_{th_{\boldsymbol{x}}}^{2/\alpha}\mathbb{E}(I)^{2/\alpha}}\gamma(2/\alpha, \gamma_{th_{\boldsymbol{x}}}\mathbb{E}(I)R_{CLR}^{\alpha}) \right), \tag{7}$$

where $\gamma_{th_{\boldsymbol{x}}}$ is the MU $\boldsymbol{x}$'s SIR threshold for reliable DL transmission and $\gamma(a,b) = \int_0^b t^{\alpha-1}\exp(-t)dt$ is the lower incomplete Gamma function. $\mathbb{P}_{suc_{\boldsymbol{x}}} = Pr\{\Gamma_{\boldsymbol{x}} > \gamma_{th}\}$ is $\boldsymbol{x}$'s successful transmission probability, i.e. probability that the received SIR at $\boldsymbol{x}$ is higher than γ_{th}. The derivation of coverage probability is in same spirit as [13] and in the interest of space left for the journal version of the paper.

Considering a constant bitrate system, the system wide ASE can simply be lower bounded using transmission success probability as

$$\mathcal{ASE}(R_{CLR}) \geq \bar{\lambda}_{MU}\log_2(1+\gamma_{th})\mathbb{P}_{suc}(\gamma_{th}, R_{CLR}^2). \tag{8}$$

A thorough investigation of (8) reveals that the effective number of scheduled users will increase as cluster size shrinks. On the other hand, increasing the cluster size decreases co-tier interference and thus enhances $\Gamma_{\boldsymbol{x}}$. This discussion implies that there exists an optimal cluster radius that maximizes system wide ASE.

B. Energy Efficiency

The power consumption of a stand-alone small cell was investigated in the award winning project EARTH [19]. The model was extended by parameterization for C-RAN [20]. Taking inspiration from [21] and [20], the power consumption per unit area can be written in simplified form as

$$P_{CRAN} = \lambda_{RRH}p_{ACT}\left(M\theta P_0 + \Delta_u P_u\right), \tag{9}$$

where M is the mean RRH activation per cluster, P_0 is the fixed power consumption of an active RRH, Δ_u is the coefficient that lumps together frequency dependent response of a power amplifier and several other factors, and P_u denotes the load (active MU density) dependent RRH transmit power. $0 \leq \theta \leq 1$ parameterizes the UCRAN implementation efficiency with $\theta = 1$ indicating least energy efficienct deployment. The mathematical expression for determining average number of RRHs in each cluster (M) is given in Lemma 1.

Lemma I: The average number of activated RRHs within an arbitrary user-centric cluster, i.e. M, is the complement of the void probability of the RRHs, i.e. $M = 1 - e^{-\lambda_{RRH}\pi R_{CLR}^2}$.

Proof: Consider that Λ_{RRH} is a SPPP with intensity λ_{RRH}, then under user-centric scheme, the average number of RRHs within a circular area of radius R_{CLR} is given by $\lambda_{RRH}\pi R_{CLR}^2$. Since each user-centric cluster can have at most one activated RRH, the average number of activated RRHs is the complement of the probability that an arbitrary cluster would at least one RRH within its foot-prints, i.e.

$$\begin{aligned} M &= Pr\{\Lambda_{RRH} \cap b(\boldsymbol{x}, R_{CLR}) \neq \emptyset | \boldsymbol{x} \in \Lambda'_{MU}\}, \\ &= 1 - Pr\{\Lambda_{RRH} \cap b(\boldsymbol{x}, R_{CLR}) = \emptyset | \boldsymbol{x} \in \Lambda'_{MU}\}, \\ &= 1 - \exp\{-\pi\lambda_{RRH}R_{CLR}^2\}. \end{aligned}$$

■

Considering unity bandwidth, the system energy efficiency '$\mathcal{EE}(R_{CLR})$' (bits/s/Joule) for a UCRAN system can be expressed as the ratio of sustainable system throughput (8) and the total power consumed by the activated RRHs (9).

IV. GT-SON Framework for RRH Cluster Size Optimization

The GT based SON engine is embedded within the centralized BBU pool for real-time adjustment of R_{CLR} to optimize a system level efficiency parameter of interest with respect to terrain environment, user demographics, RRH deployment scenario and network operator's spatio-temporal revenue model (see fig.2). The variation in the cluster size models the dynamic tradeoff between ASE and EE in our bargaining game model. The proposed GT-SON framework with the sequence of steps in dynamic cluster size adjustment for modeling the ASE-EE tradeoff is given in fig. 2.

To analytically express the ASE-EE tradeoff, we formulate a two-player cooperative bargaining game where both ASE and EE are modelled as virtual game players that independently estimate the best cluster size for maximizing their respective utility functions. We will see later that due to a large dissimilarity in cluster size preferences of the players, each player's payoff is affected by the cluster size selection made by the other player. However, both players can mutually benefit through the cooperative game where they negotiate for the R_{CLR} that achieves optimal ASE-EE tradeoff. Using Nash's axiomatic model, it is well known that the Nash bargaining solution (NBS) achieves a pareto-optimal solution, i.e. the optimal tradeoff in the utilities of the players in such cooperative games [22]. If

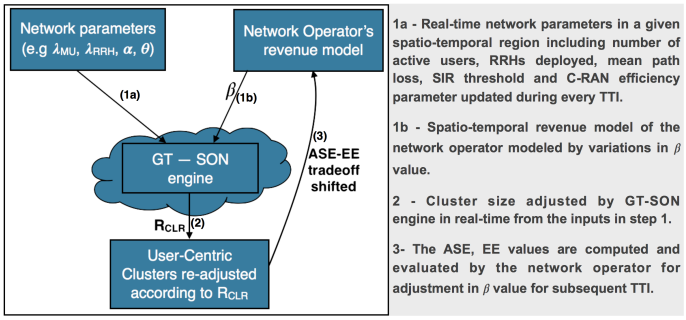

Fig. 2. GT-SON Framework in UCRAN

the players can be denoted by the set $\mathbb{N} = \{1, 2\}$, where player $i = 1$ denotes ASE, player $i = 2$ denotes EE and S_i denotes the set of all feasible payoffs to an arbitrary MU i as

$$S_i = \{U_i | U_i = U_i(R_{CLR}), R_{CLR} \in \mathbb{R} : R_{CLR} > 0\}. \tag{10}$$

Let us define the space $\mathcal{S}$ as the set of all feasible payoffs that players $i \in \mathbb{N}$ can achieve when they collaborate, i.e.

$$\mathcal{S} = \{U = (u_1, u_2) | u_1 \in S_1, u_2 \in S_2\} \tag{11}$$

where $u_1(x_1)$ is the utility of the first player and $u_2(x_2)$ is the utility of the second player such that

$$s_1 = u_1(x_1) = [\mathcal{ASE}(R_{CLR})]^{\beta}, \tag{12}$$

$$s_2 = u_2(x_2) = [\mathcal{EE}(R_{CLR})]^{1-\beta} \tag{13}$$

and $x_1 = x_2 = R_{CLR} \in \mathbb{R} : R_{CLR} > 0$. $\beta \in [0,1]$ is the exponential bias factor in NBS that defines the bargaining power (or the tradeoff) division between the two players. We also define the disagreement space $\mathcal{D} \in \mathcal{S}$ as the set of the two disagreement points $d = (d_1, d_2)$ where $d_1 = u_1(\mathcal{D})$ and $d_2 = u_2(\mathcal{D})$ represent the payoffs for the two players if the bargaining process fails and no outcome is reached. For our game, we set $d = (0,0)$ thus giving both players uniform leeway to improve their utilities. [23] shows that the NBS in such parametric cooperative games exists only if the utility functions for the players form convex and compact sets.

Proposition 3. *The utility and disagreement spaces in the proposed GT-SON framework constitute a two-player bargaining problem defined by $(\mathcal{S}, d)$ where $\mathcal{S} \in \mathbb{R}^2$, $d \in \mathcal{S}$ and the resulting unique bargaining outcome is pareto-optimal.*

Proof: A bargaining problem can be defined as the pair $(\mathcal{S}, d)$ if: i) $\mathcal{S}$ is a convex and compact set, ii) There exists some $s \in \mathcal{S}$ such that $s > d$, i.e. $s_1 > d_1$ and $s_2 > d_2$. It is quite obvious that $\mathcal{S}$ is compact and since $d = (0,0)$, positive utilities for our players satisfies the 2nd condition. This leaves behind the question whether $\mathcal{S}$ is convex which holds true if: $\forall \epsilon : 0 \leq \epsilon \leq 1$, if $U^a = (u_1^a, u_2^a) \in S_1$ and $U^b = (u_1^b, u_2^b) \in S_2$, then $\epsilon U^a + (1-\epsilon)U^b \in \mathcal{S}$. From (8), we see that $\epsilon u_1^a + (1-\epsilon)u_1^b = [\bar{\lambda}_{MU} \log_2(1+\gamma_{th})\bar{P}]^\beta$ where $\bar{P} = [\epsilon(\mathbb{P}_{suc}^a)^\beta + (1-\epsilon)(\mathbb{P}_{suc}^b)^\beta]$ and since we know that $0 \leq \mathbb{P}_{suc}^a, \mathbb{P}_{suc}^b, \beta \leq 1$, the sum in (14) forms a convex set, i.e.

$$\epsilon u_1^a + (1-\epsilon)u_1^b \in S_1. \tag{14}$$

Similarly, from (9), we see that $\epsilon u_2^a + (1-\epsilon)u_2^b = [\frac{\bar{\lambda}_{MU} \log_2(1+\gamma_{th})\bar{P}}{\lambda_{RRH} p_{\bar{ACT}} (M\theta P_0 + \Delta_u P_u)}]^{(1-\beta)}$, where the numerator is convex from (14) and denominator is convex since $p_{\bar{ACT}} = \epsilon(p_{ACT}^a)^{1-\beta} + (1-\epsilon)(p_{ACT}^b)^{1-\beta}$ and $0 \leq p_{ACT}^a, p_{ACT}^b, \beta \leq 1$. Therefore,

$$\epsilon u_2^a + (1-\epsilon)u_2^b \in S_2. \tag{15}$$

From (14) and (15), we conclude that $\epsilon U^a + (1-\epsilon)U^b \in \mathcal{S}$ which satisfies the conditions for convexity for set $\mathcal{S}$. According to Nash's axiomatic approach [22], there exists a unique solution for the two-player bargaining problem which is the pair of utilities (s_1^*, s_2^*) that solves the following optimization problem:

$$\max_{(s_1,s_2)} (s_1 - d_1)(s_2 - d_2), (s_1, s_2) \in \mathcal{S} \geq (d_1, d_2). \tag{16}$$

■

Proposition 3 implies that for an arbitrary MU $\boldsymbol{x}$, the optimal cluster size "$R_{CLR,x}^{opt}$" is obtained through the solution of a convex optimization problem (also known as Nash Product (NP)) which for our model can be given by

$$R_{CLR,\boldsymbol{x}}^{opt} = \max_{R_{CLR,\boldsymbol{x}}} [\mathcal{AEE}(R_{CLR,\boldsymbol{x}})]^\beta [\mathcal{EE}(R_{CLR,\boldsymbol{x}})]^{1-\beta}. \tag{17}$$

Notice that the computational complexity of the GT-SON engine is a function of the cluster size granularity, i.e. $\mathcal{O}(N_{CLR})$ where N_{CLR} denotes the number of distinct cluster sizes over which the optimization in (17) is performed. As the processing times are independent of MU or RRH densities, real-time implementation of the GT-SON optimization framework is practically realizable and scalable throughout the network.

V. Simulation Results and Discussion

In this section, we discuss the analytical trends and Monte Carlo simulation results by employing a 3GPP standard compliant LTE network simulator. For simulation, we consider a two tier HetNet with a tri-sector hexagonal MBS of radius 500 m. We consider a single sector of the MBS covering an area of 73850 m^2 where MUs and small cell RRHs are uniformly distributed according to their independent SPPPs. Without loss of generality, the channel power gains between all MUs and RRHs are assumed unity. We assume uniform transmit power of 30 dBm for all RRHs. Other power consumption parameters are taken from [23]. Simulation results are averaged over 1000 Monte Carlo trials.

A. Impact of β on ASE, EE in a UCRAN

From the analytical results in (8), (9) and (17), we investigate the variation in the optimal cluster size and the efficiency metrics as β is shifted between ASE-optimal ($\beta = 1$) and EE-optimal ($\beta = 0$) points. The GT-SON engine optimizes R_{CLR} on the following fixed network parameters: $\lambda_{MU} = 10^{-2}/\text{m}^2$, $\theta = 0.5$, $\gamma_{th} = 4$ dB, and $0 < R_{CLR} \leq 100$ m. The ASE results in fig. 3a indicate around the same ASE-optimal cluster size of 5m for variations in pathloss exponent and RRH deployment densities. It is seen that higher RRH densities yield superior system throughput which is understandable considering p_{ACT} is expected to increase with λ_{RRH}. It is also noted that $\alpha = 4$ yields more than two-fold increase in ASE as compared to $\alpha = 3$. Since mmWave network propagation studies [24] have indicated higher pathloss due to blocking effects, the UCRAN is expected to yield more system capacity at mmWave spectrum by virtue of relatively larger MU-interfering RRH distances. EE results in fig. 3b indicate optimal R_{CLR} to be the highest possible cluster size because of the combined effect of inreased throughput and reduced power consumption with increase in R_{CLR}. Like ASE, the maximum EE is achieved at higher RRH density and pathloss exponents. This implies that the GT-SON engine will most effectively utilize the ASE-EE tradeoff with gain variations of over 100% through appropriate β adjustment in ultra-dense mmWave networks.

B. User QoE Analysis in a UCRAN

Users' QoE analysis is conducted through SINR distribution between MUs with network parameters: $\lambda_{MU} = 10^{-2}/\text{m}^2$, $\lambda_{RRH} = 10^{-3}/\text{m}^2$, α=4, $\theta = 0.5$, $\gamma_{th} = 4$ dB and bandwidth B=1 Hz. Both the MU and RRH deployments are performed using uniform PPPs and average performance results are obtained via Monte Carlo simulations. We use two variants of the proposed user-centric approach: i) RRH cluster size deployment that maximizes ASE henceforth referred as UC(ASE), and ii) cluster size deployment that maximizes EE henceforth referred as UC(EE). To compare the performance with a standard non user-centric PPP deployment, we follow the approach in [25] and represent it as NUC. Results in fig.4 show that even with the most data throughput efficient user-centric design, we obtain a SINR gain of over 20 dB for almost 50% of the users. The ruggedness in the cdf graph of UC(EE) in comparison to the other two CDFs is because of lower number of users in the thinned PPP $\bar{\lambda}_{MU}$ which is a direct consequence of

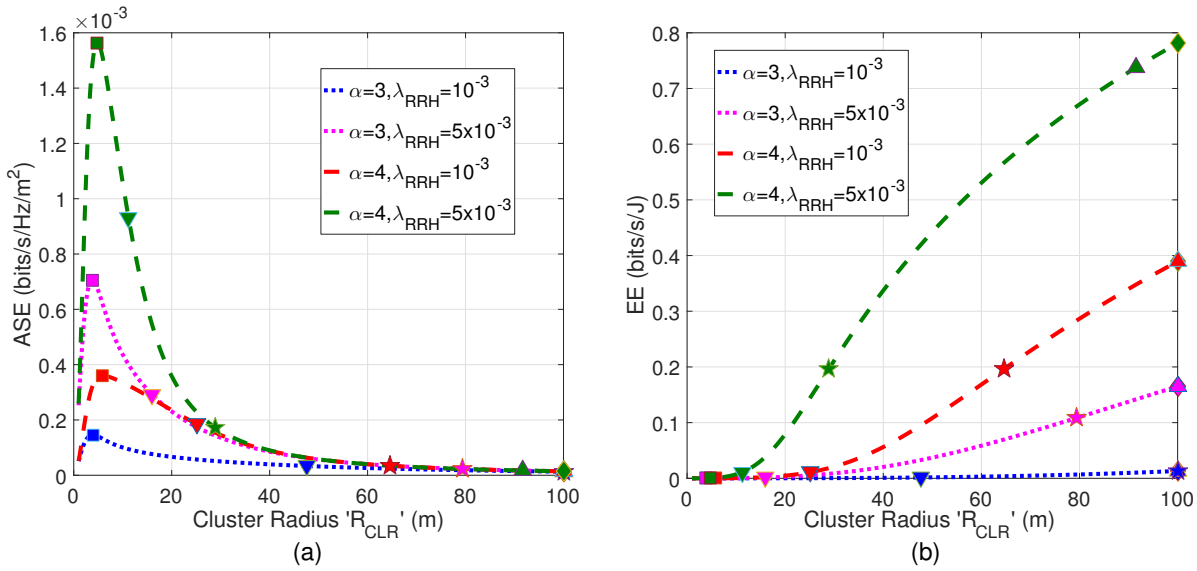

Fig. 3. (a) ASE and (b) EE at different β values when varying the λ_{RRH} and α. The NBS for each use case is shown separately for ASE and EE at β values of 0, 0.25, 0.5, 0.0.75 and 1 and denoted by ♦, ▲, ★ , ▼ and ■ respectively.

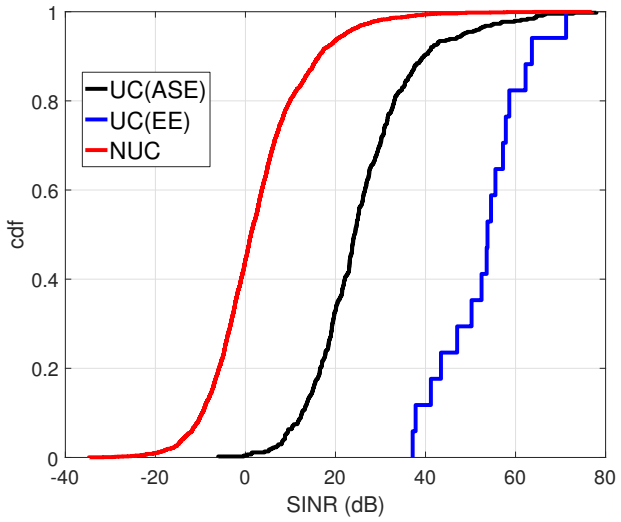

Fig. 4. Downlink SINR cdf comparison between user-centric and non user-centric approaches

the larger cluster sizes in EE optimization. The 5 percentile SINR performance (for the cell-edge users with worst SINR in conventional networks) is also significantly improved for user-centric approaches with about 20 dB and 40 dB gain with UC(ASE) and UC(EE) respectively. Clearly the user-centric approach eliminates cell-edge degradation and guaranteed high QoE for every user regardless of its physical location.

C. ASE, EE v/s λ_{RRH} in a UCRAN

Fig. 5 compares the system wide ASE and EE of the user-centric approaches with the baseline scheme at different RRH densities and $\lambda_{MU} = 10^{-2}/\text{m}^2$, α=4, $\theta = 0.5$ and $\gamma_{th} = 4$ dB. Fig. 5a reveals that as the RRH deployment density increases, UC(ASE) emerges as the most data efficient scheme. While NUC exhibits uniform ASE, UC(ASE) by virtue of increased $\mathbb{P}_{suc}$ exhibits highest system capacity. On the other hand, UC(EE), though not throughput efficient by any regards, yields more than 5 times power efficient network as compared to NUC approach (fig. 5b). This observation highlights the inherent ASE-EE tradeoff available to the network operator by adjusting β via the GT-SON and choosing the appropriate RRH cluster size.

VI. Conclusion

In this paper, we proposed a user-centric Cloud RAN orchestration framework capable of offering higher system capacity, better energy efficiency and improved received signal quality in dense deployment scenarios, compared to non user-centric conventional Cloud RAN architectures. We derived expressions for the area spectral and energy efficiency parameters as a function of system wide RRH cluster size in the user-centric network. Analytical results revealed that while ASE is optimized at low cluster sizes, EE becomes optimal at a large cluster size as large cluster sizes ensure lower interference and reduced power consumption through smaller number of activated RRH. Consequentially, the ASE-EE tradeoff manifests itself in terms of dimensioning of the cluster radius in UCRAN. We then propose a game theoretic framework to achieve Pareto optimal solution and show that a SON engine within the centralized BBU pools can be used to dynamically configure the optimal cluster size. Simulation results indicate that: i) the SON mechanism allows more than 100% efficiency variation particularly at dense RRH deployments and high pathloss exponents, and ii) significant SINR gains can be realized in both ASE and EE operating modes by virtue of interference-free RRH cluster zones around each scheduled user. Future directions include investigations of methods to group multiple users into clusters based on their spatial proximity and service class.

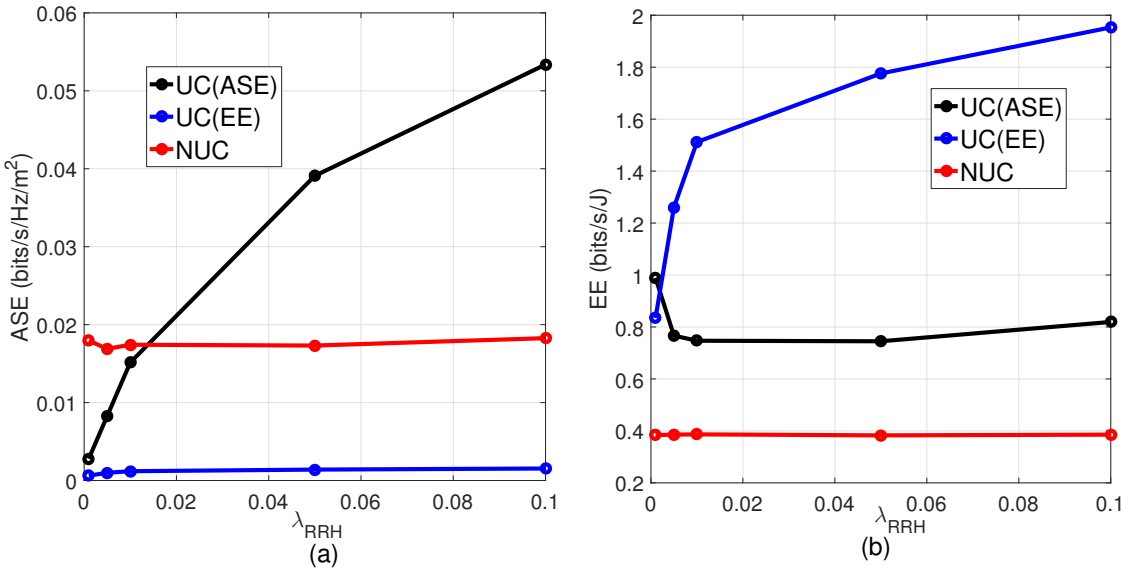

Fig. 5. (a) ASE and (b) EE comparison of UC(ASE), UC(EE) and NUC approaches with different RRH densities

ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under Grant Number 1619346 and Grant Number 1559483.

REFERENCES

[1] S. Zaidi, S. Affes, U. Vilaipornsawai, L. Zhang, and P. Zhu, "Wireless Access Virtualization Strategies for Future User-Centric 5G Networks," in *2016 IEEE Globecom Workshops (GC Wkshps)*, Dec 2016, pp. 1–7.

[2] L. Wang and Q. Liang, "Optimization for user centric massive mimo cell free networks via large system analysis," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–1.

[3] B. Romanous, N. Bitar, S. A. R. Zaidi, A. Imran, M. Ghogho, and H. H. Refai, "A Game Theoretic Approach for Optimizing Density of Remote Radio Heads in User Centric Cloud-Based Radio Access Network," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[4] V. Mohanan, R. Budiarto, and I. Aldmour, *Powering the Internet of Things With 5G Networks*, ser. Advances in Wireless Technologies and Telecommunication. IGI Global, 2017. [Online]. Available: https://books.google.com.pk/books?id=XYwtDwAAQBAJ

[5] R. Wang, H. Hu, and X. Yang, "Potentials and Challenges of C-RAN Supporting Multi-RATs Toward 5G Mobile Networks," *IEEE Access*, vol. 2, pp. 1187–1195, 2014.

[6] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for Mobile Networks x2014;A Technology Overview," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.

[7] C. Mobile, "C-RAN: the road towards green RANs," *White Paper, ver, vol. 2*, 2011.

[8] Y. Zhang and Y. J. Zhang, "User-centric virtual cell design for Cloud Radio Access Networks," in *Signal Processing Advances in Wireless Communications (SPAWC), 2014 IEEE 15th International Workshop on*, June 2014, pp. 249–253.

[9] A. Imran, M. Imran, and R. Tafazolli, "A novel Self Organizing framework for adaptive Frequency Reuse and Deployment in future cellular networks," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2010 IEEE 21st International Symposium on*, Sept 2010, pp. 2354–2359.

[10] H. Zhang, Z. Yang, Y. Liu, and X. Zhang, "Power Control for 5G User-Centric Network: Performance Analysis and Design Insight," *IEEE Access*, vol. 4, pp. 7347–7355, 2016.

[11] C. Li, J. Zhang, M. Haenggi, and K. B. Letaief, "User-Centric Intercell Interference Nulling for Downlink Small Cell Networks," *IEEE Transactions on Communications*, vol. 63, no. 4, pp. 1419–1431, April 2015.

[12] S. Bassoy, M. Jaber, M. A. Imran, and P. Xiao, "Load Aware Self-Organising User-Centric Dynamic CoMP Clustering for 5G Networks," *IEEE Access*, vol. 4, pp. 2895–2906, 2016.

[13] S. Zaidi, A. Imran, D. McLernon, and M. Ghogho, "Characterizing Coverage and Downlink Throughput of Cloud Empowered HetNets," *Communications Letters, IEEE*, vol. PP, no. 99, pp. 1–1, 2015.

[14] N. Saxena, A. Roy, and H. Kim, "Traffic-aware cloud ran: A key for green 5g networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 1010–1021, April 2016.

[15] P. R. Li and K. T. Feng, "Channel-aware resource allocation for energy-efficient cloud radio access networks under outage specifications," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7389–7403, Nov 2017.

[16] J. M. D. Stoyan, W. S. Kendall and L. Ruschendorf, *Stochastic geometry and its applications*. Wiley Chichester, 1995, vol. 2.

[17] M. Haenggi, "Mean Interference in Hard-Core Wireless Networks," *Communications Letters, IEEE*, vol. 15, no. 8, pp. 792–794, August 2011.

[18] Z. Chen, C.-X. Wang, X. Hong, J. Thompson, S. Vorobyov, X. Ge, H. Xiao, and F. Zhao, "Aggregate Interference Modeling in Cognitive Radio Networks with Power and Contention Control," *Communications, IEEE Transactions on*, vol. 60, no. 2, pp. 456–468, February 2012.

[19] [Online]. Available: https://www.ict-earth.eu/

[20] R. Gupta, E. Calvanese Strinati, and D. Ktenas, "Energy efficient joint DTX and MIMO in cloud Radio Access Networks," in *Cloud Networking (CLOUDNET), 2012 IEEE 1st International Conference on*, Nov 2012, pp. 191–196.

[21] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *Wireless Communications, IEEE*, vol. 18, no. 5, pp. 40–49, October 2011.

[22] Z. Han, D. Niyato, W. Saad, T. Basar, and A. Hjorungnes, *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*. Cambridge University Press, 2012.

[23] B. Romanous, N. Bitar, S. A. R. Zaidi, A. Imran, M. Ghogho, and H. H. Refai, "A Game Theoretic Approach for Optimizing Density of Remote Radio Heads in User Centric Cloud-Based Radio Access Network," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[24] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056, Sept 2015.

[25] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3122–3134, November 2011.