Exploiting Dual Connectivity in Heterogeneous Cellular Networks

Narayan Prasad*, Sampath Rangarajan*
*NEC Labs America, Princeton, NJ e-mail: {prasad, sampath}@nec-labs.com

***Abstract*—We consider network utility maximization problems over heterogeneous cellular networks (HetNets) that permit dual connectivity. Dual connectivity (DC) is a feature that targets emerging practical HetNet deployments that will comprise of non-ideal (higher latency) connections between transmission nodes, and has been recently introduced to the LTE-Advanced standard. DC allows for a user to be simultaneously served by a macro node as well as one other (typically micro or pico) node and requires relatively coarser level coordination among serving nodes. For such a DC enabled HetNet we comprehensively analyze the problem of determining an optimal user association that maximizes the weighted sum rate system utility subject to per-user rate constraints, over all feasible associations. Here, in any feasible association each user can be associated with (i.e., configured to receive data from) any one macro node (in a given set of macro nodes) and any one pico node that lies in the chosen macro node's coverage area. We show that, remarkably, this problem can be cast as a non-monotone submodular set function maximization problem, which allows us to construct a constant-factor approximation algorithm. We then consider the proportional fairness (PF) system utility and characterize the PF optimal resource allocation. This enables us to construct an efficient algorithm to determine an association that is optimal up-to an additive constant. We then validate the performance of our algorithms via numerical results.**

I. Introduction

Traditional cellular wireless are rapidly transforming into dense HetNets that have discarded the classical structured layout of cells. Instead, these HetNets are characterized by the presence of a multitude of transmission nodes (or points) ranging from enhanced versions of the conventional high power macro base-station or NodeB (eNBs) to low power pico nodes, all deployed in a highly irregular fashion [1]. Indeed, the deployment of the low power nodes is done within the coverage area of an eNB to cater to emerging hot spots, thereby alleviating demand bottlenecks without being subject to many of the challenges in eNB site acquisition. However, a major hinderance in such deployments is that there is need for coordination among the transmission nodes (which becomes more acute as the density of such nodes rises) while at the same time the backhaul link between these nodes is often non-ideal. Consequently, for tractable resource allocation, a HetNet is partitioned into several coordination units or clusters with each cluster comprising of a set of high power eNBs along with a set of low power pico nodes assigned to each one of the high power nodes. Together, theses transmission points (TPs) cater to a given set of users. In addition, only semi-static coordination among TPs in a cluster is deemed feasible, wherein periodically (once in every frame of a few hundred milliseconds duration) there is coordination among serving nodes in the cluster. One popular method of coordination is load balancing or user association [1] where each user can be associated with only one TP at any given time. This load balancing requires limited coordination among TPs which is possible under a non-ideal backhaul, and it mitigates the undesirable scenario of TPs becoming overloaded due to too many users being associated with them. Combinations of load balancing with several resource management schemes have also received wide attention [2]–[8].

Our interest in this work is on dual connectivity (DC) that has been recently introduced to the 3GPP LTE-A standard [9], where the single-TP association constraint is relaxed and a user can be associated to a high power and a low power node. Such a user can simultaneously receive (different) data from both nodes. Schemes to fully exploit DC are being actively investigated and the potential challenges and good directions are summarized in [10]. The work in [11] considers a DC enabled uplink with one macro and one pico node and proposes optimal rate and power control solutions for a cost minimization problem with per-user minimum rate constraints. On the other hand, [12] considers resource partitioning at only the macro node in a DC enabled downlink to optimize the PF utility. [13] reuses existing algorithms for user association and investigates data forwarding and flow control problems. We consider a general DC enabled HetNet downlink with multiple users and TPs. Our key contributions are:

- We propose an efficient algorithm that yields a user association that is optimal for the PF system utility up-to an additive constant. To the best of our knowledge, this is the first such approximation algorithm for DC and the PF utility. Using this algorithm, we demonstrate the significant gains enabled by DC especially at low network loads.
- We also show that the user association problem to optimize the weighted sum rate utility subject to per-user rate constraints can be formulated as a constrained non-monotone submodular set function maximization. This allows us to obtain an efficient algorithm which guarantees a constant-factor approximation. We note that to prove submodularity we do not follow the direct approach of establishing the original definition, but instead we consider proving another cleverly obtained sufficient condition. The latter approach then requires us to characterize the (second order) change in the optimal solution of a linear program with respect to its parameters. We expect that our result and systematic derivation will have wider applications.

II. Problem Formulation

Let $\mathcal{U}$ denote the set of users with cardinality $|\mathcal{U}| = K$ and let $\mathcal{M}$ denote the set of Macro TPs. For each Macro TP $m \in \mathcal{M}$, let $\mathcal{B}_m$ denote the distinct set of pico TPs assigned to macro TP m. Here the set $\mathcal{B}_m$ of pico TPs facilitate the macro TP m to serve its associated users. We suppose that all indices in the set of all TPs, $\mathcal{S} = \mathcal{M} \cup \mathcal{P}$, where $\mathcal{P} = \cup_{m\in\mathcal{M}}\mathcal{B}_m$, are distinct. An illustrative schematic for DC is shown in Fig. 1. Notice that each user can be associated with any one macro and any one pico TP from the set of pico TPs assigned to the chosen macro. For each user $u \in \mathcal{U}$ and each TP $b \in \mathcal{S}$, we define $R_{u,b}$ to be the (average) peak rate user u can get (in bits per unit resource) when it alone is served by TP b. This average rate is a function of the slow fading parameters (e.g. path loss

978-3-901882-90-6/17 ©2017 IFIP

$$\begin{aligned}
&\max_{z_{u,m},x_{u,b}\in\{0,1\},\gamma_{u,b},\theta_{u,m}\in[0,1]\ \forall\ u\in\mathcal{U},b\in\mathcal{B}_m,m\in\mathcal{M}} \left\{\sum_{u\in\mathcal{U}}\sum_{m\in\mathcal{M}}\left(w_uR_{u,m}\theta_{u,m}+\sum_{b\in\mathcal{B}_m}w_ux_{u,b}R_{u,b}\gamma_{u,b}\right)\right\}\\
&\text{s.t.} \sum_{m\in\mathcal{M}}z_{u,m}\le 1;\ \sum_{b\in\mathcal{B}_m}x_{u,b}=z_{u,m},\ \forall\ u\in\mathcal{U},m\in\mathcal{M};\sum_{u\in\mathcal{U}}\theta_{u,m}\le 1\ \&\ \sum_{u\in\mathcal{U}}\gamma_{u,b}\le 1\ \forall\ b\in\mathcal{B}_m,m\in\mathcal{M};\\
&R_{u,m}\theta_{u,m}+\sum_{b\in\mathcal{B}_m}x_{u,b}R_{u,b}\gamma_{u,b}\in[z_{u,m}R_u^{\min},z_{u,m}R_u^{\max}],\ \forall\ u\in\mathcal{U}.
\end{aligned}\tag{1}$$

and shadowing) seen by user u but not the instantaneous fast-fading ones. Further, this average rate can be computed under the (widely used) assumption that all the other interfering TPs are always transmitting, or under the framework of [14] that each interfering TP transmits for a given fraction of the frame duration. We note that in the in-band DC case the picos and macros share the same spectrum band whereas in the out-of-band DC case they are assigned different bands. Our results are applicable to both scenarios and only the user peak rates have to be accordingly computed.We make the mild assumption that $R_{u,b}>0,\ \forall\ u\in\mathcal{U},b\in\mathcal{S}$. We also assume that $\frac{R_{u,m}}{R_{u,b}}\neq\frac{R_{u',m}}{R_{u',b}},\ \forall\ b\in\mathcal{B}_m,m\in\mathcal{M}\ \&\ u\neq u'$. Notice that these two assumptions hold true almost surely for all typical slow fading distributions. We first formulate a network utility maximization problem that adopts the weighted sum rate (WSR) system utility under per-user rate constraints. The WSR problem is posed in (1). $w_u>0,\ \forall\ u\in\mathcal{U}$ denotes any input weight or priority assigned to user u. In (1) the binary valued variable $z_{u,m}$ is one if user u is associated with macro TP m and zero otherwise, so that the first set of constraints in (1) ensures that each user is associated with at-most one macro. Further, exploiting dual connectivity, each user that is associated to the macro TP m is also associated with any one pico TP in $\mathcal{B}_m$. Indeed, the indicator variable $x_{u,b}$ is one if user u is associated to TP b and zero otherwise. Consequently, $\{x_{u,b}=1\}_{u\in\mathcal{U},b\in\mathcal{B}_m}$ yields the user set such that each user in that set is associated to any one TP in $\mathcal{B}_m$ as well with the macro TP m. The continuous variables $\{\gamma_{u,b},\theta_{u,m}\}$ are referred to here as allocation fractions and their respective sums are upper bounded by unity for each macro TP as well as each pico TP, as depicted in the second set of constraints. Although (1) does not enforce that each user must be associated with a macro TP, it does enforce (in the third set of constraints) that each user associated to any macro TP must be assigned at-least its minimum rate but should not exceed a maximum rate. Notice also that (1) is always feasible. [1] The minimum rate constraints are useful in enforcing that a minimum expected quality is provided to each served user and are based on the observation that in many scenarios serving a user at a rate below its minimum threshold is futile. On the other hand, the maximum rate limits can be used to cap the rates of any set of users such as those that have subscribed to a lower tier of service or another operator.

We next consider the PF system utility and adopting the convention that $0\ln(0)=0$, we pose a mixed optimization problem given in (2). The first set of constraints in (2) ensures that each user is associated with exactly one macro. As before, exploiting DC each user that is associated with the macro TP m is also associated with any one pico TP in $\mathcal{B}_m$ for all $m\in\mathcal{M}$.

Note that our formulations assume an infinitely backlogged traffic model with no limits on buffer sizes at any TP. Coordination among TPs happens at frame boundaries where the user association can be altered. After a transient phase (whose length can be ignored), for each user distinct data streams are available for downlink transmission at its assigned macro as well as its assigned pico node. This setting bestows tractability while being relevant. Extending our results to a more realistic formulation with finite buffers entailing careful data forwarding (from each macro to each pico assigned to it) is an interesting topic for future work. Any proof that is missing from the following sections can be found in [17].

III. Characterizing Optimal Allocation Fractions

We begin our quest by characterizing the optimal allocation fractions of (1) and (2) for any given user association. This enables the design of approximation algorithms detailed in the subsequent section.

A. *Optimal allocation fractions of (1)*

We now proceed to characterize the solution of (1) for any given user association. We note that upon fixing the user association in (1) (i.e., upon fixing $\{x_{u,b},z_{u,m}\}$) the problem in (1) decouples into $|\mathcal{M}|$ sub-problems, one for each macro TP. Consequently, we focus our attention on the subproblem corresponding to any macro TP, say with index 1, and suppose that any subset of users $\mathcal{U}'\subseteq\mathcal{U}$ is associated to that macro by the given association. Then, for each $b\in\mathcal{B}_1$, let $\mathcal{U}^{(b)}=\{u\in\mathcal{U}:x_{u,b}=1\}$ denote the associated user set such that $\mathcal{U}^{(b)}\cap\mathcal{U}^{(b')}=\phi,b\neq b'$, where ϕ denotes the empty set and $\cup_{b\in\mathcal{B}_1}\mathcal{U}^{(b)}=\mathcal{U}'$. Let $\mathcal{B}_1'$ denote the set of all pico TPs in $\mathcal{B}_1$ with at-least one associated user. In addition, we consider a budget constraint for each pico TP, $\Gamma_b\in[0,1],\ \forall\ b\in\mathcal{B}_1'$ and another one for the macro, $\Gamma\in[0,1]$. With these in hand we pose the problem in (3), which we assume to be feasible. Note that without loss of generality we can assume that each pico TP is resource limited, i.e., $\sum_{u\in\mathcal{U}^{(b)}}\frac{R_u^{\max}}{R_{u,b}}>\Gamma_b$. This is because otherwise we can simply assign maximum possible resource from TP b to each user in $\mathcal{U}^{(b)}$ and remove those from further consideration as no macro resource will then be sought by those users. We will use the term *slack* to denote the resource assigned to a user in excess of its minimum rate requirement. For convenience, for each user $k\in\mathcal{U}^{(b)},b\in\mathcal{B}_1'$ we let $R_k=R_{k,1}\theta_{k,1}+R_{k,b}\gamma_{k,b}$ and supress the dependence of R_k on $\theta_{k,1},\gamma_{k,b}$. To analyze the linear program in (3) we offer the following result that can be derived by carefully manipulating the K.K.T. conditions.

[1]Note that single-TP only association is subsumed by (1). Indeed, with $x_{u,b}=1$ for some $u\in\mathcal{U},b\in\mathcal{B}_m$, by choosing $\theta_{u,m}=0$ ($\gamma_{u,b}=0$) we can ensure that user u will receive data only from the pico node b (macro node m). Also, each user u that is not associated with any macro TP gets zero rate and must have $\theta_{u,m}=0,\gamma_{u,b}=0,\forall\ b\in\mathcal{B}_m,m\in\mathcal{M}$ due to the third set of constraints.

$$\boxed{\begin{aligned}&\max_{z_{u,m},x_{u,b}\in\{0,1\},\gamma_{u,b},\theta_{u,m}\in[0,1]\ \forall\ u\in\mathcal{U},b\in\mathcal{B}_m,m\in\mathcal{M}}\left\{\sum_{u\in\mathcal{U}}\sum_{m\in\mathcal{M}}\sum_{b\in\mathcal{B}_m}x_{u,b}\ln\left(R_{u,m}\theta_{u,m}+R_{u,b}\gamma_{u,b}\right)\right\}\\ &\text{s.t.}\sum_{m\in\mathcal{M}}z_{u,m}=1;\sum_{b\in\mathcal{B}_m}x_{u,b}=z_{u,m},\ \forall\ u\in\mathcal{U},m\in\mathcal{M};\sum_{u\in\mathcal{U}}\theta_{u,m}\le 1\ \&\sum_{u\in\mathcal{U}}\gamma_{u,b}\le 1\ \forall\ b\in\mathcal{B}_m,m\in\mathcal{M}.\end{aligned}} \tag{2}$$

$$\boxed{\begin{aligned}&\max_{\gamma_{u,b},\theta_{u,1}\in[0,1]\ \forall\ u\in\mathcal{U}',b\in\mathcal{B}_1'}\left\{\sum_{b\in\mathcal{B}_1'}\sum_{u\in\mathcal{U}^{(b)}}\left(w_uR_{u,1}\theta_{u,1}+w_uR_{u,b}\gamma_{u,b}\right)\right\}\\ &\text{s.t.}\sum_{u\in\mathcal{U}'}\theta_{u,1}\le\Gamma\ \&\sum_{u\in\mathcal{U}^{(b)}}\gamma_{u,b}\le\Gamma_b\ \forall\ b\in\mathcal{B}_1';\ \ R_u\triangleq R_{u,1}\theta_{u,1}+R_{u,b}\gamma_{u,b}\in[R_u^{\min},R_u^{\max}],\ \forall\ u\in\mathcal{U}^{(b)},b\in\mathcal{B}_1'.\end{aligned}} \tag{3}$$

Proposition 1. *The following conditions must be satisfied by any optimal solution of (3):*

- *For any two distinct users associated with any TP, $k,j\in\mathcal{U}^{(b)},\ b\in\mathcal{B}_1'$, such that $\frac{R_{k,b}}{R_{k,1}}>\frac{R_{j,b}}{R_{j,1}}$, we must have*

$$\theta_{k,1}>0\Rightarrow\gamma_{j,b}=0 \tag{4}$$

- Slack ordering: *For any two distinct users associated with any TP, $k,j\in\mathcal{U}^{(b)},\ b\in\mathcal{B}_1'$, such that $w_kR_{k,b}>w_jR_{j,b}$, we must have*

$$R_j>R_j^{\min}\ \&\ \gamma_{j,b}>0\Rightarrow R_k=R_k^{\max}. \tag{5}$$

Similarly, for any two distinct users, $k,j\in\mathcal{U}'$, such that $w_kR_{k,1}>w_jR_{j,1}$, we must have

$$R_j>R_j^{\min}\ \&\ \theta_{j,1}>0\Rightarrow R_k=R_k^{\max}. \tag{6}$$

- *For any user $k\in\mathcal{U}^{(b)},\ b\in\mathcal{B}_1'$, such that $\gamma_{k,b}>0$, if there exists any other user $j\in\mathcal{U}'$ with $\theta_{j,1}>0$ and $R_j>R_j^{\min}$, we must have*

$$\frac{R_{k,b}}{R_{k,1}}\ge\frac{\max_{j\in\mathcal{U}^{(b)}\setminus k:R_j<R_j^{\max}}\{w_jR_{j,b}\}}{\min_{j\in\mathcal{U}'\setminus k:\theta_{j,1}>0\ \&\ R_j>R_j^{\min}}\{w_jR_{j,1}\}}. \tag{7}$$

Similarly, for any user $k\in\mathcal{U}^{(b)},\ b\in\mathcal{B}_1'$, such that $\theta_{k,1}>0$, if there exists any other user $j\in\mathcal{U}^{(b)}$ with $\gamma_{j,b}>0$ and $R_j>R_j^{\min}$, we must have

$$\frac{R_{k,b}}{R_{k,1}}\le\frac{\min_{j\in\mathcal{U}^{(b)}\setminus k:\gamma_{j,b}>0\ \&\ R_j>R_j^{\min}}\{w_jR_{j,b}\}}{\max_{j\in\mathcal{U}'\setminus k:R_j<R_j^{\max}}\{w_jR_{j,1}\}}. \tag{8}$$

Let $\hat{O}(\Gamma,\{\Gamma_b\})$ denote the optimal objective value of (3) for the given budgets. We begin our quest to derive an efficient algorithm and establish a useful property for the problem in (3). Towards that end, without loss of generality, in this section we assume a labeling of user indices such that for any two users $k,j\in\mathcal{U}^{(b)}$ for any $b\in\mathcal{B}_1'$, $k<j\Rightarrow\frac{R_{k,b}}{R_{k,1}}\ge\frac{R_{j,b}}{R_{j,1}}$. Further, without loss of generality, the user indices in $\mathcal{U}^{(b)}$ for each $b\in\mathcal{B}_1'$, are assumed to be consecutive. Then, upon appyling primal decomposition on (3), we see that if we fix the share of the Macro resource that can be used by each TP $b\in\mathcal{B}_1'$ as Z_b, where $\sum_{b\in\mathcal{B}_1'}Z_b=\Gamma$, (3) decouples into $|\mathcal{B}_1'|$ sub-problems. In particular, the problem at hand for TP b is given by (9), and we let $\hat{O}(Z_b,\Gamma_b,b)$ denote its optimal objective value.

$$\begin{aligned}&\max_{\substack{\gamma_{u,b},\theta_{u,1}\in[0,1]\\ \forall\ u\in\mathcal{U}^{(b)}}}\left\{\sum_{u\in\mathcal{U}^{(b)}}\left(w_uR_{u,1}\theta_{u,1}+w_uR_{u,b}\gamma_{u,b}\right)\right\}\\ &\text{s.t.}\sum_{u\in\mathcal{U}^{(b)}}\theta_{u,1}\le Z_b\ \&\sum_{u\in\mathcal{U}^{(b)}}\gamma_{u,b}\le\Gamma_b;\\ &R_u\in[R_u^{\min},R_u^{\max}],\ \forall\ u\in\mathcal{U}^{(b)}.\end{aligned} \tag{9}$$

Hence, the original problem in (3) can be expressed as

$$\max_{\{Z_b\in\mathbb{R}_+\}:\sum_{b\in\mathcal{B}_1'}Z_b\le\Gamma}\left\{\sum_{b\in\mathcal{B}_1'}\hat{O}(Z_b,\Gamma_b,b)\right\} \tag{10}$$

A straightforward approach to determine the optimal Macro resource share among the TPs is to optimize $\{Z_b\}$ using the generic subgradient method. However, we will show that exploiting the structure of the problem at hand leads directly to a very simple algorithm. First, let us define the function, $\bar{h}:\mathbb{R}_+\times\mathcal{B}_1'\to\mathbb{R}_+$, such that $\bar{h}(\Gamma_b,b)$ for any given TP b and corresponding budget, Γ_b, yields the minimum Macro resource needed (in addition to available budget Γ_b for the pico TP b) to accommodate the minimum rates of all users in $\mathcal{U}^{(b)}$. Then, we can invoke Proposition 1 to explicitly detail $\bar{h}(\Gamma_b,b)$ after recalling the labelling we have adopted, as

$$\begin{cases}\frac{R_{\tilde{k}+1}^{\min}-\Xi_bR_{\tilde{k}+1,b}}{R_{\tilde{k}+1,1}}+\sum_{j\in\mathcal{U}^{(b)}:j>\tilde{k}+1}\frac{R_j^{\min}}{R_{j,1}}, & \text{If }\tilde{k}+1\in\mathcal{U}^{(b)}\\ 0, & \text{Otherwise}\end{cases} \tag{11}$$

where

$$\begin{aligned}\tilde{k}&=\max\left\{k:k\in\mathcal{U}^{(b)}\ \&\sum_{j\in\mathcal{U}^{(b)}:j\le k}\frac{R_j^{\min}}{R_{j,b}}\le\Gamma_b\right\},\\ \Xi_b&=\Gamma_b-\sum_{j\in\mathcal{U}^{(b)}:j\le\tilde{k}}\frac{R_j^{\min}}{R_{j,b}}\end{aligned}$$

In (11) we use the convention that $\tilde{k}+1$ reurns the user with the lowest index in $\mathcal{U}^{(b)}$ whenever $\tilde{k}$ is null on account of $\sum_{j\in\mathcal{U}^{(b)}:j\le k}\frac{R_j^{\min}}{R_{j,b}}>\Gamma_b$ for all $k\in\mathcal{U}^{(b)}$. In a similar manner we define $\underline{h}:\mathbb{R}_+\times\mathcal{B}_1'\to\mathbb{R}_+$, such that $\underline{h}(Z_b,b)$ for any given TP b and a given Macro budget, Z_b, yields the additional minimum resource needed by TP b to accommodate the minimum rates of all users in $\mathcal{U}^{(b)}$. Again invoking Proposition

1 we can explicitly detail $\underline{h}(Z_b, b)$, as

$$\begin{cases} \frac{R^{\min}_{\breve{k}-1} - \tilde{Z}_b R_{\breve{k}-1,1}}{R_{\breve{k}-1,b}} + \sum_{j \in \mathcal{U}^{(b)}: j < \breve{k}-1} \frac{R^{\min}_j}{R_{j,b}}, & \text{If } \breve{k}-1 \in \mathcal{U}^{(b)} \\ 0, & \text{Otherwise} \end{cases} \quad (12)$$

where

$$\breve{k} = \min\left\{ k : k \in \mathcal{U}^{(b)} \text{ \& } \sum_{j \in \mathcal{U}^{(b)}: j \geq k} \frac{R^{\min}_j}{R_{j,1}} \leq Z_b \right\},$$
$$\tilde{Z}_b = Z_b - \sum_{j \in \mathcal{U}^{(b)}: j \geq \breve{k}} \frac{R^{\min}_j}{R_{j,1}}$$

For any given Γ_b, we let $S(Z_b, \Gamma_b, b),\ b \in \mathcal{B}'_1, Z_b \geq \bar{h}(\Gamma_b, b)$ denote the slope of the function $\hat{O}(Z_b, \Gamma_b, b)$ at Z_b. In particular,

$$S(Z_b, \Gamma_b, b) = \lim_{\delta \to 0_+} \frac{\hat{O}(Z_b + \delta, \Gamma_b, b) - \hat{O}(Z_b, \Gamma_b, b)}{\delta} \quad (13)$$

Henceforth, without loss of generality, we assume $\underline{h}(1, b) \leq 1$ & $\bar{h}(1, b) \leq 1,\ \forall\, b \in \mathcal{B}'_1$.

Proposition 2. *For any fixed $\Gamma_b \geq \underline{h}(1, b)$, $\hat{O}(Z_b, \Gamma_b, b)$ is continuous, non-decreasing, piecewise linear and concave in $Z_b \in [\bar{h}(\Gamma_b, b), 1]$. For any fixed $Z_b \geq \bar{h}(1, b)$, $\hat{O}(Z_b, \Gamma_b, b)$ is continuous, non-decreasing, piecewise linear and concave in $\Gamma_b \in [\underline{h}(Z_b, b), 1]$.*

Proof. We only prove the first claim since proof for the second one follows along similar lines. The continuity and non-decreasing properties are straightforward to verify. It can be shown that the conditions stated in Proposition 1 provide necessary and sufficient conditions to determine an optimal set of allocation fractions for the problem in (9). To verify the other two properties, we start at $Z_b = \bar{h}(\Gamma_b, b)$. Then, if $\bar{h}(\Gamma_b, b) = 0$, the slack at the pico TP b, $\Gamma_b - \sum_{j \in \mathcal{U}^{(b)}} \frac{R^{\min}_j}{R_{j,b}}$ must be distributed among users in the decreasing order $\{w_k R_{k,b}\}_{k \in \mathcal{U}^{(b)}}$ subject to their respective maximum rate limits (cf. slack ordering in Proposition 1). On the other hand, when $\bar{h}(\Gamma_b, b) > 0$ there is no slack at the pico TP for this Z_b. The next key observation we use is the one in Proposition 1 pertaining to the order in which macro resources are assigned to users in $\mathcal{U}^{(b)}$. Following our labelling, we see that when $Z_b = \bar{h}(\Gamma_b, b) > 0$ either user $\tilde{k}+1$ (when $\Xi_b > 0$) or user $\tilde{k}$ (when $\Xi_b = 0$) is the user with the largest index in $\mathcal{U}^{(b)}$ to be assigned a positive resource by TP b. Let user k' be this user so that users $k \in \mathcal{U}^{(b)} : k < k'$ are assigned resource only by TP b at this $Z_b = \bar{h}(\Gamma_b, b)$. The slope of $\hat{O}(Z_b, \Gamma_b, b)$, $S(\bar{h}(\Gamma_b, b), \Gamma_b, b)$, can be determined as the maximum of two terms $\max\left\{ \frac{w_k R_{k,b} R_{k',1}}{R_{k',b}} : k \in \mathcal{U}^{(b)} \text{ \& } k < k' \right\}$ and $\max\{w_k R_{k,1} : k \in \mathcal{U}^{(b)} \text{ \& } k \geq k'\}$. Then, as Z_b is increased to $Z_b + \delta$, for any arbitrarily small $\delta > 0$, the slack is put to the user yielding the slope $S(\bar{h}(\Gamma_b, b), \Gamma_b, b)$ (i.e., offering the maximum bang-per-buck). If such a user is some $\hat{k} \geq k'$ then the additional available Macro resource, δ, is directly assigned as slack to it. Otherwise, the additional available Macro resource is first assigned to user k', which frees up resource $\frac{\delta R_{k',1}}{R_{k',b}}$ at the pico TP b (while maintaining the minimum rate of user k'). This freed up pico resource is assigned as slack to the user $\hat{k}$ yielding the slope $S(\bar{h}(\Gamma_b, b), \Gamma_b, b)$. As δ is increased the slack is continuously assigned to the user $\hat{k}$ till the slope changes and that user is not the one yielding the maximum bang-per-buck. This happens either if user $\hat{k}$ attains its maximum rate upon which it is removed from the candidate list of users which can be assigned slack, or if user k' is no longer assigned any resource by pico TP b. In the former case, the slope changes to the maximum of $\max\left\{ \frac{w_k R_{k,b} R_{k',1}}{R_{k',b}} : k \in \mathcal{U}^{(b)} \setminus \hat{k} \text{ \& } k < k' \right\}$ and $\max\{w_k R_{k,1} : k \in \mathcal{U}^{(b)} \setminus \hat{k} \text{ \& } k \geq k'\}$, whereas in the latter case the slope changes to the maximum of $\max\left\{ \frac{w_k R_{k,b} R_{k'-1,1}}{R_{k'-1,b}} : k \in \mathcal{U}^{(b)} \text{ \& } k < k'-1 \right\}$ and $\max\{w_k R_{k,1} : k \in \mathcal{U}^{(b)} \text{ \& } k \geq k'-1\}$. Thenceforth additional Macro resources are assigned as slack to the user yielding the new slope and the process continues. Note that at every change the slope decreases because either users are removed from candidate list or the gain term multiplying the weight of each user served exclusively by the pico reduces. This demonstrates the piecewise linearity and concavity. The same arguments can be applied when $\bar{h}(\Gamma_b, b) = 0$. In particular, we begin at $Z_b = \bar{h}(\Gamma_b, b) = 0$ after determining user k' that has the highest index among those that have been assigned a positive resource by the pico TP and after removing users that have achieved their maximum rates from the candidate pool. The subsequent process proceeds as before and we can deduce the piecewise linearity and concavity. □

TABLE I
Algorithm I: WSR optimal allocation fractions

1: Initialize with $\mathcal{B}'_1, \mathcal{U}^{(b)},\ b \in \mathcal{B}'_1$. Set $Z_b = \bar{h}(\Gamma_b, b),\ \forall\ b \in \mathcal{B}'_1$ and $C = \Gamma - \sum_{b \in \mathcal{B}'_1} Z_b$.
2: For each pico TP $b \in \mathcal{B}'_1$, if $Z_b = 0$ then distribute the slack at that pico, $\Gamma_b - \sum_{j \in \mathcal{U}^{(b)}} \frac{R^{\min}_j}{R_{j,b}}$, among users in $\mathcal{U}^{(b)}$.
3: **Repeat**
4: Determine $\hat{b} = \arg\max_{b \in \mathcal{B}'_1} \{S(Z_b, \Gamma_b, b)\}$
5: Determine $\hat{\Delta} = \sup\{\Delta \in \mathbb{R}_+ : S(Z_{\hat{b}} + \Delta, \Gamma_{\hat{b}}, \hat{b}) = S(Z_{\hat{b}}, \Gamma_{\hat{b}}, \hat{b})\}$.
6: Increment $Z_{\hat{b}} = Z_{\hat{b}} + \min\{C, \hat{\Delta}\}$ and update $C = \max\{0, C - \hat{\Delta}\}$.
7: **Until** $C = 0$.
8: Output $\{Z_b\}, \forall\ b \in \mathcal{B}'_1$, the corresponding allocation fractions $\{\theta_{u,1}, \gamma_{u,b}\},\ \forall\ u \in \mathcal{U}^{(b)}, b \in \mathcal{B}'_1$.

Corollary 1. *For any fixed $\Gamma_b \geq \underline{h}(1, b)$, $\hat{O}(Z_b, \Gamma_b, b)$ can be computed as*

$$\hat{O}(Z_b, \Gamma_b, b) = \sum_{k \in \mathcal{U}^{(b)}} R^{\min}_k + H(\Gamma_b, b) + \int_{\bar{h}(\Gamma_b, b)}^{Z_b} S(z_b, \Gamma_b, b) dz_b,\ \forall\ Z_b \in [\bar{h}(\Gamma_b, b), 1], \quad (14)$$

where $H(\Gamma_b, b) = 0$ whenever $\bar{h}(\Gamma_b, b) > 0$ and when $\bar{h}(\Gamma_b, b) = 0$, it yields the weighted sum rate obtained by distributing the excess pico resource as slack among users in $\mathcal{U}^{(b)}$ in the decreasing order $\{w_k R_{k,b}\}$ subject to their respective maximum rate limits.

We now propose Algorithm I to determine the Macro allocations that optimize (10) (and hence (3)).

Proposition 3. *The optimal solution to (3) can be determined using Algorithm I whenever the necessary and sufficient condition for feasibility, $\sum_{b \in \mathcal{B}'} \bar{h}(\Gamma_b, b) \leq 1$, holds. For any fixed budgets $\{\Gamma_b\}\ \forall\ b \in \mathcal{B}'_1$ satisfying the feasibility condition,*

$\hat{O}(\Gamma, \{\Gamma_b\})$ *is continuous, non-decreasing, piecewise linear and concave in* $\Gamma \in [\sum_{b\in\mathcal{B}_1'} \bar{h}(\Gamma_b, b), 1]$.

Proof. First note that using Proposition 2 with (10), we see that we are maximizing $|\mathcal{B}_1'|$ piecewise linear concave functions subject to a linear budget constraint. Notice that in Algorithm I we always choose the highest slope and assign it as much resource as possible till the point that maximal slope changes. This greedy strategy is optimal for the problem at hand because: (i) each slope curve $S(Z_b, \Gamma_b, b), Z_b \geq \bar{h}(\Gamma_b, b)$ is a piecewise constant function in Z_b and (ii) any Macro resource assigned to any TP b' has no influence on the slope curve of any other TP $b \neq b', b \in \mathcal{B}_1'$. More formally, (10) can be shown to equivalent to the maximization of a modular function subject to a cardinality constraint for which the greedy strategy is optimal. Next, the claimed properties of $\hat{O}(\Gamma, \{\Gamma_b\})$ directly follow from the facts in Proposition 2 that in (10) each $\hat{O}(Z_b, \Gamma_b, b)$ is continuous, non-decreasing, piecewise linear and concave in Z_b. $\square$

Let $\mathbf{\Gamma} = [\Gamma_b]_{b\in\mathcal{B}_1'}$ denote any vector of all pico budgets and let $S(Z, \mathbf{\Gamma})$ denote the slope curve of $\hat{O}(Z, \mathbf{\Gamma}) = \hat{O}(Z, \{\Gamma_b\})$ for $Z \geq \sum_{b\in\mathcal{B}_1'} \bar{h}(\Gamma_b, b)$, which from Proposition 3 we know to be piecewise constant and non-increasing in Z. We then have the following corollary.

Corollary 2.

$$\hat{O}(\Gamma, \mathbf{\Gamma}) = \sum_{b\in\mathcal{B}_1'} \sum_{k\in\mathcal{U}^{(b)}} R_k^{\min} + \sum_{b\in\mathcal{B}_1'} H(\Gamma_b, b) + \int_{\sum_{b\in\mathcal{B}_1'} \bar{h}(\Gamma_b,b)}^{Z} S(z, \mathbf{\Gamma}) dz, \ \forall\, Z \in \left[\sum_{b\in\mathcal{B}_1'} \bar{h}(\Gamma_b, b), 1\right]. \quad (15)$$

To illustrate our results, in Fig. 2 we consider a macro TP serving $|\mathcal{U}'| = 30$ users, with $|\mathcal{B}_1| = 10$ pico nodes assigned to it and where each user is associated with the pico in $\mathcal{B}_1$ from which it sees the strongest signal strength. We obtained the user peak rates by emulating a realistic deployment (the details are defered to the simulation results section) and imposed no maximum rate limits. A unit budget at each pico was assumed and the minimum rate of each user was chosen to be a scalar times its peak rate from the macro TP (with the scalar being identical for each user). We considered several values for this scalar and in each case plot $\hat{O}(\Gamma, \{1\}), \Gamma \in [\sum_{b\in\mathcal{B}_1'} \bar{h}(1, b), 1]$ (computed using Algorithm I). As predicted by Proposition 3, each curve is non-decreasing, piece-wise linear and concave and as expected the obtained value matches the one obtained by solving (3) via a generic LP solver. Notice as the minimum rate requirements become more stringent, more macro resource is needed to satisfy them and the optimized utility value decreases.

To offer our next key result, we introduce some notation. For any two pico TPs $b_1, b_2 \in \mathcal{B}_1'$, we let $\mathbf{e}_{b_1}, \mathbf{e}_{b_2}$ define $|\mathcal{B}_1'| \times 1$ unit vectors that have a zero on all their entries except the ones corresponding to b_1, b_2, respectively, which are both one. We also assume that the per-user maximum rate constraints are vacuous, i.e., $R_u^{\max} \geq R_{u,1} + R_{u,b}, \ \forall\, u \in \mathcal{U}^{(b)}, b \in \mathcal{B}_1'$.

Proposition 4. *For any non-negative scalars* $\delta, \tilde{\delta}, \delta_{b_1}, \tilde{\delta}_{b_2}$ *and budgets* $\Gamma, \mathbf{\Gamma}$ *such that* $\Gamma \geq \sum_{b\in\mathcal{B}_1'} \bar{h}(\Gamma_b, b)$, *we have that*

$$\hat{O}(\Gamma, \mathbf{\Gamma}) - \hat{O}(\Gamma + \delta, \mathbf{\Gamma} + \delta_{b_1}\mathbf{e}_{b_1}) \leq \hat{O}(\Gamma + \tilde{\delta}, \mathbf{\Gamma} + \tilde{\delta}_{b_2}\mathbf{e}_{b_2}) - \hat{O}(\Gamma + \tilde{\delta} + \delta, \mathbf{\Gamma} + \delta_{b_1}\mathbf{e}_{b_1} + \tilde{\delta}_{b_2}\mathbf{e}_{b_2}) \quad (16)$$

B. Optimal allocation fractions of (2)

As before we obtain the sets $\mathcal{U}', \mathcal{B}_1'$ from the given assoication and let $N_b = |\mathcal{U}^{(b)}|$ denote the cardinality or the number of users associated with TP $b \in \mathcal{B}_1'$. Consider the PF system utility optimization problem (restricted to the user pool $\mathcal{U}'$) for the the given user association, depicted in (17).

$$\max_{\substack{\gamma_{u,b},\theta_{u,1}\in[0,1] \\ \forall\, u\in\mathcal{U}', b\in\mathcal{B}_1'}} \left\{ \sum_{u\in\mathcal{U}'} \sum_{b\in\mathcal{B}_1'} x_{u,b} \ln\left(R_{u,1}\theta_{u,1} + R_{u,b}\gamma_{u,b}\right) \right\}$$
$$\text{s.t.} \sum_{u\in\mathcal{U}'} \theta_{u,1} \leq 1 \ \& \sum_{u\in\mathcal{U}'} \gamma_{u,b} \leq 1 \ \forall\, b \in \mathcal{B}_1'. \quad (17)$$

Note that (17) is a purely continuous optimization problem.

Next, for each $b \in \mathcal{B}_1'$ define $\mu_{1,b} = \min_{u\in\mathcal{U}^{(b)}} \{\frac{R_{u,1}}{R_{u,b}}\}$. Similary, let $\mu_{k,b}, k \in \{2, \cdots, N_b\}$ denote the k^{th} smallest ratio in the set $\{\frac{R_{u,1}}{R_{u,b}}\}_{u\in\mathcal{U}^{(b)}}$ and recall that these ratios are all strictly positive and distinct. Then, defining $\mu_{N_b+1,b} = \infty$, we have $0 < \mu_{1,b} < \mu_{2,b} < \cdots < \mu_{N_b,b} < \mu_{N_b+1,b} = \infty$. Next, we define two functions $h : \mathbb{R}_{++} \times \mathcal{B}_1' \to \mathbb{R}_+$ and $g : \mathbb{R}_{++} \times \mathcal{B}_1' \to \mathbb{R}_+$ as in (18). The following result follows upon carefully analyzing the K.K.T conditions for the convex optimization problem in (17).

Theorem 1. *The optimal objective value of (17) is given by*

$$\sum_{b\in\mathcal{B}_1'} \left(g(\hat{\lambda}, b) + \sum_{k\in\mathcal{U}^{(b)}} \ln(R_{k,b}) \right), \quad (19)$$

where $\hat{\lambda} \in (0, \infty)$ *is the unique solution to the relation*

$$1 + \sum_{b\in\mathcal{B}_1'} h(\lambda, b) = \sum_{b\in\mathcal{B}_1'} N_b/\lambda, \quad (20)$$

and can be determined via bisection search.

Corollary 3. *Suppose the optimal* $\hat{\lambda}$ *satisfying (20) is given. Then, if* $\hat{\lambda} \in ((m-1)\mu_{m,b}, m\mu_{m,b})$ *for some* $m = 1, \cdots, N_b$, *the optimal solution comprises of assigning an identical resource share* $\gamma_{k,b} = \mu_{m,b}/\hat{\lambda}$ *with* $\theta_{k,1} = 0$ *for all users* $k \in \mathcal{U}^{(b)} : \frac{R_{k,1}}{R_{k,b}} < \mu_{m,b}$, *whereas all users* $k \in \mathcal{U}^{(b)} : \frac{R_{k,1}}{R_{k,b}} > \mu_{m,b}$ *are assigned* $\theta_{k,1} = 1/\hat{\lambda}$ *with* $\gamma_{k,b} = 0$. *The user* $k \in \mathcal{U}^{(b)} : \frac{R_{k,1}}{R_{k,b}} = \mu_{m,b}$ *is assigned* $\theta_{k,1} = m/\hat{\lambda} - 1/\mu_{m,b}$ *with* $\gamma_{k,b} = 1 - (m-1)\mu_{m,b}/\hat{\lambda}$. *On the other hand, if* $\hat{\lambda} \in [(m-1)\mu_{m-1,b}, (m-1)\mu_{m,b}]$ *for some* $m = 2, \cdots, N_b+1$, *the optimal solution comprises of assigning an identical resource share* $\gamma_{k,b} = 1/(m-1)$ *with* $\theta_{k,1} = 0$ *for all users* $k \in \mathcal{U}^{(b)} : \frac{R_{k,1}}{R_{k,b}} < \mu_{m,b}$, *whereas all users* $k \in \mathcal{U}^{(b)} : \frac{R_{k,1}}{R_{k,b}} \geq \mu_{m,b}$ *are assigned* $\theta_{k,1} = 1/\hat{\lambda}$ *with* $\gamma_{k,b} = 0$.

We next introduce another useful result that will be invoked to establish the performance guarantee of an algorithm proposed later for (2) in the sequel. Towards that end, we introduce the problem in (21) where we recall that $\{x_{u,b}\}_{u\in\mathcal{U}', b\in\mathcal{B}_1'}$ are given.

Proposition 5. *The optimal solution determined from (21) yields an objective value for (17) that is no less than the optimal objective value of (17) minus* $\min\{|\mathcal{B}_1'|, |\mathcal{U}'|\}\ln(2)$.

$$h(\lambda,b) = \begin{cases} \frac{1}{\mu_{m,b}}, & m=1,\cdots,N_b : \lambda \in ((m-1)\mu_{m,b}, m\mu_{m,b}) \\ \frac{m-1}{\lambda}, & m=2,\cdots,N_b+1 : \lambda \in [(m-1)\mu_{m-1,b}, (m-1)\mu_{m,b}] \end{cases} \tag{18}$$

$$g(\lambda,b) = \begin{cases} \sum_{j=m}^{N_b} \ln(\mu_{j,b}/\lambda) + (m-1)\ln(\mu_{m,b}/\lambda), & m=1,\cdots,N_b : \lambda \in ((m-1)\mu_{m,b}, m\mu_{m,b}) \\ -(m-1)\ln(m-1) + \sum_{q=m}^{N_b} \ln(\mu_{q,b}/\lambda), & m=2,\cdots,N_b+1 : \lambda \in [(m-1)\mu_{m-1,b}, (m-1)\mu_{m,b}] \end{cases}$$

$$\boxed{\begin{aligned} \max_{\tilde z_{u,1}\in\{0,1\}\ \forall\ u\in\mathcal{U}'} \Bigg\{ \sum_{u\in\mathcal{U}'} \Bigg(\tilde z_{u,1}\ln(R_{u,1}) + \sum_{b\in\mathcal{B}_1'} (1-\tilde z_{u,1}) x_{u,b} \ln(R_{u,b}) \Bigg) - \\ \Bigg(\Bigg(\sum_{k\in\mathcal{U}'} \tilde z_{k,1}\Bigg) \ln\Bigg(\sum_{k\in\mathcal{U}'} \tilde z_{k,1}\Bigg) + \sum_{b\in\mathcal{B}_1'} \Bigg(\sum_{k\in\mathcal{U}'} (1-\tilde z_{k,1}) x_{k,b}\Bigg) \ln\Bigg(\sum_{k\in\mathcal{U}'}(1-\tilde z_{k,1})x_{k,b}\Bigg)\Bigg)\Bigg\} \end{aligned}} \tag{21}$$

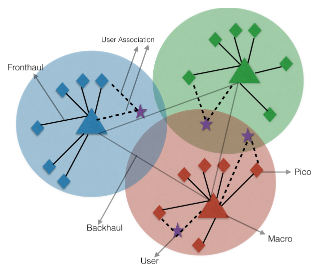

Fig. 1. Dual Connectivity Schematic

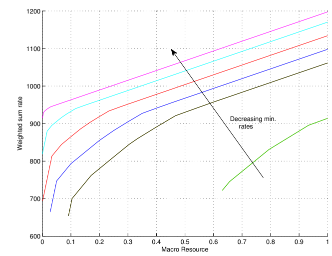

Fig. 2. Optimized WSR vs macro budget for different min. rates

IV. Approximation Algorithms

We are now ready to propose approximation algorithms for the problems in (1) and (2). We begin with the WSR maximization problem in (1). Let us define a ground set $\underline{\Omega} = \{(u,b), u \in \mathcal{U}, b \in \mathcal{P}\}$ where (u,b) conveys the association of user u with pico TP b. The tuple also implicitly indicates the association of u to the Macro TP m where $b \in \mathcal{B}_m$. Without loss of generality we suppose that only a tuple (u,b) for any $u \in \mathcal{U}$ & $b \in \mathcal{B}_m, m \in \mathcal{M}$ for which $R_{u,m} + R_{u,b} \geq R_u^{\min}$ is included in $\underline{\Omega}$. This is because any tuple not satisfying this assumtion will never be selected as its minimum rate cannot be met even when the assigned macro and the pico TPs fully allocate their resources to that user. Let $\underline{\Omega}^{(m)} = \{(u,b) \in \underline{\Omega} : b \in \mathcal{B}_m\}$ denote all possible associations to any pico TP in $\mathcal{B}_m$, the set of pico TPs assigned to macro TP $m \in \mathcal{M}$, and let $\underline{\Omega}_{(u')} = \{(u,b) \in \underline{\Omega} : u = u'\}$ denote all possible associations of a user $u' \in \mathcal{U}$. Define a family of sets $\underline{\mathcal{I}}$ as as the one which includes each subset of $\underline{\Omega}$ such that the tuples in that subset have mutually distinct users. Formally,

$$\underline{\mathcal{A}} \subseteq \underline{\Omega} : |\underline{\mathcal{A}} \cap \underline{\Omega}_{(u)}| \leq 1\ \forall\ u \in \mathcal{U} \Leftrightarrow \underline{\mathcal{A}} \in \underline{\mathcal{I}}. \tag{22}$$

Further, define a family, $\underline{\mathcal{J}}$, contained in $\underline{\mathcal{I}}$ that comprises of each member of $\underline{\mathcal{I}}$ for which (1) is feasible. It can be shown that while $\underline{\mathcal{I}}$ defines a matroid, $\underline{\mathcal{J}}$ is a downward closed family but need not satisfy the exchange property and hence need not define a matroid. Next, we define a non-negative set function on $\underline{\mathcal{J}}$, $f^{\text{wsr}} : \underline{\mathcal{J}} \to \mathbb{R}_+$ such that it is normalized, i.e., $f(\phi) = 0$, and for any non-empty set $\underline{\mathcal{G}} \in \underline{\mathcal{J}}$, we have

$$f^{\text{wsr}}(\underline{\mathcal{G}}) = \sum_{m\in\mathcal{M}} f_m^{\text{wsr}}(\underline{\mathcal{G}} \cap \underline{\Omega}^{(m)}). \tag{23}$$

Each $f_m^{\text{wsr}} : \underline{\mathcal{J}}^{(m)} \to \mathbb{R}_+$ in (23) is a normalized non-negative set function that is defined on the family $\underline{\mathcal{J}}^{(m)}$ which comprises of each member of $\underline{\mathcal{J}}$ that is contained in $\underline{\Omega}^{(m)}$, as follows. For any set $\underline{\mathcal{A}} \in \underline{\mathcal{J}}^{(m)}$, we define $f_m^{\text{wsr}}(\underline{\mathcal{A}}) = \hat{O}(1,\mathbf{1})$, where $\hat{O}(1,\mathbf{1})$ is computed as described in Algorithm I in Section III-A for the macro TP m and the set of pico TPs $\mathcal{B}_m$ assigned to it, using unit budgets and the given association in $\underline{\mathcal{A}}$. We recall that a simple necessary and sufficient condition to determine feasibility of the minimum rates for the given association and budgets is provided in Proposition 3. With these definitions in hand, can re-formulate the problem in (1) as the following constrained set function maximization problem.

$$\max_{\underline{\mathcal{G}}\in\underline{\mathcal{J}}} \{f^{\text{wsr}}(\underline{\mathcal{G}})\} \tag{24}$$

We offer our first main result that characterizes $f^{\text{wsr}}(.)$.

Theorem 2. *The set function $f^{\text{wsr}}(.)$ is normalized, non-negative and possibly non-monotone. It is also submodular when $R_u^{\max} \geq \max_{m\in\mathcal{M}} \{R_{u,m} + \max_{b\in\mathcal{B}_m}\{R_{u,b}\}\},\ \forall\ u.$*

Proof. The set function $f^{\text{wsr}}(.)$ in (23) defined on the family $\underline{\mathcal{J}}$ is normalized and non-negative by construction. Due to the presence of minimum rate limits in (1) this function need not be monotone, i.e., there can exist members $\mathcal{A},\mathcal{B} \in \underline{\mathcal{J}}$: $\mathcal{A} \subseteq \mathcal{B}$ for which $f^{\text{wsr}}(\mathcal{A}) > f^{\text{wsr}}(\mathcal{B})$. Simultaneously, there can exist members $\mathcal{A}',\mathcal{B}' \in \underline{\mathcal{J}}$: $\mathcal{A}' \subseteq \mathcal{B}'$ for which $f^{\text{wsr}}(\mathcal{A}') \leq f^{\text{wsr}}(\mathcal{B}')$. Then, to establish submodularity of $f^{\text{wsr}}(.)$ on the family $\underline{\mathcal{J}}$, it suffices to show that each $f_m^{\text{wsr}}(.)$ is submodular on the family $\underline{\mathcal{J}}^{(m)}$. Without loss of generality, we consider macro TP 1 and will prove that forall $\underline{\mathcal{E}} \subseteq \underline{\mathcal{F}} \in \underline{\mathcal{J}}^{(1)}$, $(u_1,b_1) \in \underline{\Omega} \setminus \underline{\mathcal{F}} : \underline{\mathcal{F}} \cup (u_1,b_1) \in \underline{\mathcal{J}}^{(1)}$,

$$\begin{aligned} f_1^{\text{wsr}}(\underline{\mathcal{E}} \cup (u_1,b_1)) - f_1^{\text{wsr}}(\underline{\mathcal{E}}) \geq \\ f_1^{\text{wsr}}(\underline{\mathcal{F}} \cup (u_1,b_1)) - f_1^{\text{wsr}}(\underline{\mathcal{F}}). \end{aligned} \tag{25}$$

Further, it suffices to prove (25) for $\underline{\mathcal{F}} = \underline{\mathcal{E}} \cup (u_2,b_2)$ so that $|\mathcal{F}| = |\mathcal{E}| + 1$ and $(u_2,b_2) \in \underline{\mathcal{J}}^{(1)}$. Then, we evaluate

$f_1^{\rm wsr}(\underline{\mathcal{F}} \cup (u_1, b_1))$ as described in Section III-A and in the obtained optimal allocation fractions let the share of pico TP b_1 resource assigned to user u_1 in tuple (u_1, b_1) be δ_{b_1} and the share of macro TP resource assigned to that user be δ. Similarly, let the share of pico TP b_2 resource assigned to user u_2 in tuple (u_2, b_2) be $\tilde{\delta}_{b_2}$ and the share of macro TP resource assigned to that user be $\tilde{\delta}$. Define $\Gamma = 1 - \delta - \tilde{\delta}$ and $\mathbf{\Gamma} = \mathbf{1} - \delta_{b_1}\mathbf{e}_{b_1} - \tilde{\delta}_{b_2}\mathbf{e}_{b_2}$. Thus, we have that

$$f_1^{\rm wsr}(\underline{\mathcal{F}} \cup (u_1, b_1)) = \hat{O}(\Gamma, \mathbf{\Gamma}) + w_{u_1}R_{u_1,b_1}\delta_{b_1} + w_{u_1}R_{u_1,1}\delta + w_{u_2}R_{u_2,b_2}\tilde{\delta}_{b_2} + w_{u_2}R_{u_2,1}\tilde{\delta}, \quad (26)$$

where $\hat{O}(\Gamma, \mathbf{\Gamma})$ is evaluated for the tuples in $\underline{\mathcal{E}}$ under the budgets Γ and $\mathbf{\Gamma}$. Further, we can readily verify the relations in (27). Using (26) and (27) in (25), it is now seen that a sufficient condition for (25) to hold is for (16) to be true. The latter is assured by Proposition 4. $\square$

Motivated by the result in Theorem 2, we employ an algorithm proposed in [16] for general non-monotone submodular set function maximization under a matroid constraint, over the problem at hand (24). That algorithm, referred to here as the local search (LS) algorithm, comprises of an LS stage consisting of addition, deletion and swap operations. At the termination of the LS stage we obtain the primary choice $\breve{\underline{\mathcal{G}}}$. Then, the LS stage is repeated over the complement set, $\underline{\Omega} \setminus \breve{\underline{\mathcal{G}}}$ to generate an alternate choice, $\tilde{\underline{\mathcal{G}}}$. Finally, the choice yielding the larger weighted sum rate utility among the primary and alternate choices is chosen. The worst-case complexity of the LS algorithm scales polynomially in $|\Omega|/\epsilon$, where $\epsilon > 0$ is an input parameter used to set a threshold used by the algorithm. We proceed to derive performance guarantee for the LS Algorithm. Towards that end, we introduce an assumption pertaining to the feasibility of the minimum rates. We emphasize that this assumption is only needed for deriving a performance guarantee but not for implementing the algorithm.

• *Admission control assumption: Each macro TP $m \in \mathcal{M}$ can itself simultaneously meet twice the minimum rates of all users u that are present in at-least one tuple $(u, b) \in \underline{\Omega}$ for any $b \in \mathcal{B}_m$.* We now offer the following result.

Theorem 3. *The LS algorithm yields a constant factor ($\frac{1}{4+\epsilon}$) approximation to (1) over all input instances for which the admission control assumption holds and the per-user maximum rate constraints are vacuous.*

Proof. Let $\tilde{\underline{\mathcal{I}}}$ denote the family of sets obtained by taking the pairwise union of members of $\underline{\mathcal{I}}$. We define an extended set function as $\tilde{f}^{\rm wsr}(\underline{\mathcal{G}}) = \sum_{m\in\mathcal{M}} \tilde{f}_m^{\rm wsr}(\underline{\mathcal{G}} \cap \underline{\Omega}^{(m)}), \ \forall \ \underline{\mathcal{G}} \in \tilde{\underline{\mathcal{I}}}$. Here, we define $\tilde{f}_m^{\rm wsr}(\underline{\mathcal{G}} \cap \underline{\Omega}^{(m)}) = \hat{O}(1, \mathbf{1})$, where $\hat{O}(1, \mathbf{1})$ is computed as described in Section III-A for the macro TP m and its set of pico TPs $\mathcal{B}_m$ using unit budgets and the given association in $\underline{\mathcal{G}} \cap \underline{\Omega}^{(m)}$, with the following caveat. In particular, now in obtaining the user sets $\{\mathcal{U}^{(b)}\}$ we treat the user in each tuple $(u, b) \in \underline{\mathcal{G}} \cap \underline{\Omega}^{(m)}$ as a distinct virtual user. Hence, if (u, b_1) and (u, b_2) belong to $\underline{\mathcal{G}} \cap \underline{\Omega}^{(m)}$, we suppose that two distinct virtual users with their own separate peak rates and associated maximum and minimum rate limits are specified. These peak rates and limits are of course identical, respectively, to those of user u and we have $\tilde{f}^{\rm wsr}(\underline{\mathcal{G}}) = f^{\rm wsr}(\underline{\mathcal{G}}), \ \forall \ \underline{\mathcal{G}} \in \underline{\mathcal{J}}$. Notice that under the admission control assumtion each member of $\underline{\mathcal{I}}$ is feasible so that $\underline{\mathcal{J}} = \underline{\mathcal{I}}$. Further, each member in $\tilde{\underline{\mathcal{I}}}$ is also feasible . Then, we can verify from the arguments used to prove Theorem 2 that $\tilde{f}^{\rm wsr}(.)$ is a normalized non-negative submodular set function over $\tilde{\underline{\mathcal{I}}}$. With this understanding, we can re-formulate (1) as the following constrained set function maximization problem.

$$\max_{\underline{\mathcal{G}} \in \underline{\mathcal{I}}} \{\tilde{f}^{\rm wsr}(\underline{\mathcal{G}})\} \quad (28)$$

Let $\hat{\underline{\mathcal{G}}}$ be the set returned by the LS algorithm and let $\hat{\underline{\mathcal{O}}}$ denote any optimal solution of (28). Notice that $\hat{\underline{\mathcal{G}}} \cup \hat{\underline{\mathcal{O}}} \in \tilde{\underline{\mathcal{I}}}$ so that the extended set function is defined and is submodular over all subsets of $\hat{\underline{\mathcal{G}}} \cup \hat{\underline{\mathcal{O}}}$. This enables us to invoke the arguments presented in [16] to prove the aproximation guarantee and thereby establish our desired result. $\square$

Let us now focus on the problem in (2). In order to design an approximation algorithm, we consider the problem in (29), where we recall our convention that $0\ln(0) = 0$. Note that (29) imposes an orthogonal split on (2) and allows for each user to be associated to (and served by) exactly one node. The problem in (29) has been widely considered before and seeks to optimize the PF utility over user associations but does not permit dual connectivity. There are several approaches to solve (29), including an efficient optimal one [4] and approximately optimal ones with lower complexity [3], [4], [15]. In Algorithm II we propose a method to solve (2) where we can leverage any of the available approaches to solve (29). Once a user association is so obtained, we enhance it by exploiting dual connectivity. Hence, all users associated to any pico node $b \in \mathcal{B}_m$ for any $m \in \mathcal{M}$ are also connected to the macro m. Further, each user associated to a macro TP is also associated to a pico TP in the set of pico TPs assigned to that macro. Then, the allocation fractions are optimized as described in Section III-B. The performance guarantee of Algorithm II is established below, where we let $\Pi \geq 0$ denote the (additive) guarantee pertaining to the approach used to solve (29), i.e., the value yielded by the obtained output is no less than the corresponding optimal objective value of (29) minus Π (so that $\Pi = 0$ for the optimal algorithm [4]).

Theorem 4. *Algorithm II provides an output that yields an objective value for (2) that is no less than the optimal objective value of (2) minus $\Pi + \min\{K, \sum_{m\in\mathcal{M}} |\mathcal{B}_m|\}\ln(2)$.*

Proof. We first note that (29) is equivalent to (2) with the additional constraint that $\gamma_{u,b}\theta_{u,m} = 0, \ \forall \ u \in \mathcal{U}, b \in \mathcal{B}_m, m \in \mathcal{M}$. Suppose that (29) is solved using an approach that offers a guarantee of Π.[2] Then, note that the obtained solution is feasible for (2) and invoking Proposition 5 (once for each macro TP together with its assigned set of pico TPs and the users associated to them) we can conclude that the attained objective value is no less than the optimal one minus the claimed additive factor. The remaining steps of Algorithm II further improve the solution at hand and hence further reduce the gap to optimal, which proves the theorem. $\square$

We remark that another way to view the performance guarantee of Algorithm II (when $\Pi = 0$) is as follows. Let us scale all the peak rates of any set of $\min\{K, \sum_{m\in\mathcal{M}} |\mathcal{B}_m|\}$ users by 2 and obtain an output by Algorithm II. Then, the

[2]Note that since the objective function in (29) and (2) can be negative, we can only offer additive guarantees instead of multiplicative ones.

$$f_1^{\rm wsr}(\underline{\mathcal{F}}) \geq \hat{O}(\Gamma+\delta, \mathbf{\Gamma}+\delta_{b_1}\mathbf{e}_{b_1}) + w_{u_2}R_{u_2,b_2}\tilde{\delta}_{b_2} + w_{u_2}R_{u_2,1}\tilde{\delta}; \quad f_1^{\rm wsr}(\underline{\mathcal{E}}) = \hat{O}(\Gamma+\delta+\tilde{\delta}, \mathbf{\Gamma}+\delta_{b_1}\mathbf{e}_{b_1}+\tilde{\delta}_{b_2}\mathbf{e}_{b_2})$$
$$f_1^{\rm wsr}(\underline{\mathcal{E}}\cup(u_1,b_1)) \geq \hat{O}(\Gamma+\tilde{\delta}, \mathbf{\Gamma}+\tilde{\delta}_{b_2}\mathbf{e}_{b_2}) + w_{u_1}R_{u_1,b_1}\delta_{b_1} + w_{u_1}R_{u_1,1}\delta. \quad (27)$$

$$\boxed{\max_{x_{u,b}\in\{0,1\}\ \forall\ u\in\mathcal{U},b\in\mathcal{S}} \left\{ \sum_{u\in\mathcal{U}}\sum_{b\in\mathcal{S}} x_{u,b}\ln(R_{u,b}) - \sum_{b\in\mathcal{S}}\left(\sum_{k\in\mathcal{U}} x_{k,b}\right)\ln\left(\sum_{k\in\mathcal{U}} x_{k,b}\right)\right\} \text{ s.t.} \sum_{b\in\mathcal{S}} x_{u,b} = 1,\ \forall\ u\in\mathcal{U}.} \quad (29)$$

TABLE II
Orthogonal Split Processing based Algorithm (OSPA)

1: Initialize with $\mathcal{U}, \mathcal{M}, \mathcal{B}_m, \forall\ m\in\mathcal{M}$.
2: Set $\mathcal{S} = \mathcal{M}\cup(\cup_{m\in\mathcal{M}}\mathcal{B}_m)$ and determine user associations $\{x_{u,b}\}, u\in\mathcal{U}, b\in\mathcal{S}$ by solving (29)
3: **For** each macro TP $m\in\mathcal{M}$ **Do**
4: Consider each user with $x_{u,m}=1$ and associate that user with the pico TP in $\mathcal{B}_m$ yielding the strongest received power for that user.
5: Using the obtained association for TPs in $\mathcal{B}_m$ obtain the optimal allocation fractions using Theorem 1.
6: **End For**
7: Output the user associations and allocation fractions.

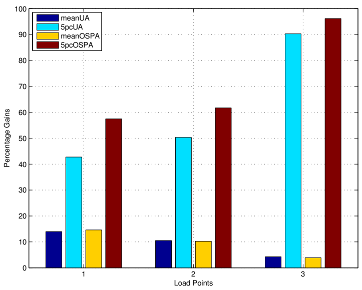

Fig. 3. In-band scenario

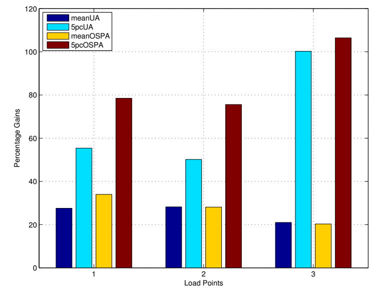

Fig. 4. Out-of-band scenario

objective value in (2) yielded by the solution at hand, will be no less than the one yielded by the optimal solution using the original peak rates.

V. Simulation Results

We now present our simulation results obtained for an LTE HetNet deployment. We emulate a HetNet comprising of 57 macro cells with 10 pico cells being assigned to each macro and with a full buffer traffic model. Each macro base-station transmits with a power of $46\ dBm$ whereas the transmit power at each pico node is $40\ dBm$ and the system bandwidth is 10 MHz. A noise PSD of $-174\ dB/Hz$ with a noise figure of $9\ dB$ were assumed. The other major parameters such as the distributions used to drop users, macro and pico nodes are all as per 3GPP evaluation guidelines. We consider an in-band scenario as well as an out-of-band scenario. Due to space constraints we only report the performance of OSPA which optimizes the PF utility (2). To benchmark the performance of this algorithm, we determine the average and the $5-$percentile spectral efficiency (SE) yielded by a baseline scheme (without DC) in which each user independently associates to the TP from which it can obtain the highest average rate. This association scheme is also referred to as the maximum SINR association [1]. Next, we determine the average and $5-$percentile SE values yielded by the user association (UA) algorithm from [4] that optimizes the PF utility without exploiting DC (29). Finally, we use that algorithm as a module in OSPA to optimize (29), with the obtained output being further refined by exploiting DC. The obtained results are plotted in Figs. 3 and 4 as relative percentage gains over the respective baseline counterparts, for the in-band and out-of-band scenarios, respectively. In each figure we consider three different load points, such the first load point emulates a HetNet with 342 users, the second one has 684 users and the last load point has 1368 users, respectively. From the results in these figures, we see that DC can be quite beneficial at low to moderate loads, which agrees with our intuition.

VI. Conclusions

We considered the problem of maximizing the weighted sum rate and the proportional fairness utility over dual connectivity enabled HetNets by exploiting load balancing. We constructed efficient algorithms to solve the resulting mixed optimization problems and proved that they yield approximately optimal solutions.

References

[1] J. Andrews, et.al., "An overview of load balancing in HetNets: Old myths and open problems," in *IEEE commun. mag.(submitted)*, July 2013
[2] K. Son, et.al., "Dynamic association for load balancing and interference avoidance in multi-cell networks," in *IEEE Trans. Wireless Comm.*, 2009.
[3] K. Shen and W. Yu, "Distributed pricing-based user association for downlink heterogeneous cellular networks," in *IEEE Journal Sel. Areas. Commun.*, Jun. 2014.
[4] N. Prasad, et.al., "Exploiting cell dormancy and load balancing in LTE hetnets: Optimizing the proportional fairness utility," in *IEEE Trans. on Commun.*, Oct. 2014.
[5] E. Altman, et.al., "Spatial sinr games of base station placement and mobile association," *IEEE Infocom.*, 2009.
[6] R. Madan, et.al., "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE J. Sel. Areas Comm.*, Dec. 2010.
[7] A. Bedekar and R. Agrawal, "Optimal muting and load balancing for eICIC," *Proc. IEEE WiOPT*, May 2013.
[8] Q. Ye, et. al., "On/Off Macrocells and Load Balancing in Heterogeneous Cellular Networks", in *IEEE Globecom.*, Dec. 2013
[9] "3GPP Technical Report 36.842. Small cell enhancements for E-UTRA and E-UTRAN Higher Layer Aspects", V1.0.0 (2013-11) www.3gpp.org
[10] S. Jha, et.al., "Dual Connectivity in LTE Small Cell Networks," in *IEEE Globecom.*, Dec. 2014.
[11] Y. Wu, et.al., "Joint Scheduling and Power Allocations for Traffic Offloading via Dual-Connectivity," in *ArXiv.*, Sep. 2015.
[12] S. Singh, et.al.,"Proportional Fair Traffic Splitting and Aggregation in Heterogeneous Wireless Networks," in *IEEE Commun. Let.*, May. 2016.
[13] H. Wang, et.al., "Dual connectivity for LTE-advanced heterogeneous networks," in *Springer, Wireless Networks*, May. 2016.
[14] V. Singh, et.al.,"Optimizing user association and activation fractions in heterogeneous wireless networks," in *IEEE WiOpt*, June. 2015.
[15] Q. Ye, et.al., "User association for load balancing in heterogeneous cellular networks," in *IEEE Trans. on Wireless Comm.*, June 2013.
[16] J. Lee, et.al., "Non-monotone submodular maximization under matroid and knapsack constraints," in *STOC*, 2009.
[17] N. Prasad and S. Rangarajan, "Exploiting Dual Connectivity in Heterogeneous Cellular Networks," Extended version in *https://www.dropbox.com/home/DualConnectjsb?preview=DualConnectjgjTR.pdf*, Jan. 2017.