Resource Allocation, Transmission Coordination and User Association in Heterogeneous Networks: a Flow-based Unified Approach

Jagadish Ghimire and Catherine Rosenberg, *Fellow, IEEE*

***Abstract*—In this paper, we formulate a flow-based framework for the joint optimization of resource allocation, transmission coordination, and user association in a heterogeneous network comprising of a macro base station and a set of pico base stations and/or relay nodes. By incorporating these three important network processes together and by unifying the analysis of pico base stations and relay nodes, our framework can act as an important engineering tool for understanding the performance of different configurations. We use the resulting formulations to characterize the performance of different combinations of resource allocation schemes and transmission coordination mechanisms. We obtained important engineering insights regarding the interplay of these network processes. In particular, under the deployment of pico base stations, we find that partially shared deployment outperforms the co-channel deployment, with or without transmission coordination. In contrast, the results also show that the deployment of relay nodes does not offer meaningful throughput gains for any choice of resource allocation scheme or transmission coordination mechanism.**

***Index Terms*—Heterogeneous cellular networks, resource allocation, transmission coordination, coordinated scheduling, user association, interference management.**

I. Introduction

Cellular network operators are facing an overwhelming growth in data traffic demands. A recent report reveals that the data traffic has already surpassed the voice traffic in mobile networks, as early as in 2009 [1]. As the capacity of a point-to-point link has almost reached the theoretical Shannon limit [2] thanks to very sophisticated physical layer techniques, cellular network operators are looking for new approaches and solutions. Adopting a "heterogeneous network (HetNet) architecture" is seen as one of such promising approaches. A heterogeneous network comprises of a set of low power base stations (BSs) overlaying the existing macro infrastructure. Examples of such low power BSs include pico base stations (PBSs), relay nodes (RNs) and femto base stations (FBSs). Pico BSs and relay nodes are operator-deployed and differ mainly in terms of the type of backhaul link. A pico BS is connected to the network core with a wired backhaul whereas a relay node is connected to the serving macro BS (MBS) via a wireless backhaul. Femto base stations on the other hand are usually deployed by users and are very low power access points intended for indoor coverage and usually connect to the network core using Internet connectivity at home. A heterogeneous architecture brings in a rich topology to the otherwise flat network architecture, but the deployment of different low power BSs over existing MBS coverage poses new challenges on important network processes including *resource allocation*, *transmission coordination* and *user association* which are all intricately linked to interference management and throughput performance. In our study, we focus on the *downlink* operation of a heterogeneous network comprising of one MBS overlaid with a number of pico BSs and/or relay nodes in an OFDM based system. Our study is relevant in the context of 3GPP LTE-Advanced air interface.

The literature is full of *resource allocation* (RA) schemes differing in complexity and efficiency including *co-channel deployment* (CCD) where all BSs (both MBS and PBSs/RNs) operate on the full set of subchannels (thereby, mutually interfering), and *orthogonal deployment* (OD) where PBSs are allocated a pool of subchannels orthogonal to the set of subchannels used for MBS operation. CCD is seen as an attractive approach, mainly due to its simplicity whereas OD is more flexible as the size of the pool of subchannels allocated to PBSs/RNs can be adjusted dynamically if required. A third type of resource allocation allows partial overlapping of the use of the subchannels. We consider a specific type of such RA and call it *partially shared deployment* (PSD) where the MBS is allocated a pool of subchannels for its dedicated use and can additionally transmit on the band allocated for PBS/RN operation at a lower power. This approach fits in well with the idea of carrier aggregation. Under a relay node deployment scenario, RA requires much more complex considerations as it affects how the backhauling is performed. Depending upon how we split the subchannels into orthogonal bands and how we allocate these bands to the backhaul links and other wireless links, a relay might perform differently. If backhaul links are allocated subchannels orthogonal to the relay downlinks, a full-duplex communication is possible at a relay which is not the case if these two types of links use the same band. These considerations do not apply to the pico base stations. In other words, even though we are trying to unify pico and relay BSs, the relay case requires careful considerations and is much more complex.

User scheduling (US) is another important network process that manages interference by scheduling users and do this by possibly coordinating the transmissions of the BSs. Such a *transmission coordination* (TC), in the most general form, can be carried out by scheduling the BSs in time together

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, 200 University Ave. West, Waterloo, Ontario, Canada ({jghimire,cath}@uwaterloo.ca)

This work was supported in part by the Natural Science and Engineering Research Council of Canada (NSERC) and by Research In Motion (RIM).

with power control at each BS, which is very complex. In its simplest form, a BS schedules its users independently of the other BSs at full power. This is the current practice. In our study, we focus on a simple type of TC called the *ON-OFF TC* where transmission coordination is carried out by scheduling BSs such that a BS can either be transmitting with the maximum available power or not transmitting. 3GPP considers such a coordination mechanism as a viable option in LTE-A networks [3]. Till recently, user scheduling was done locally by each BS on the channels allocated by RA, under some throughput criteria (e.g., proportional fairness (PF)). However, with transmission coordination among BSs, user scheduling becomes global to the macro cell. Thus, the tightly coupled nature of TC with user scheduling across multiple BSs mandates a global optimization approach. It is however not clear what magnitude of gains can be expected by the introduction of such sophisticated US.

A *user association* (UA) policy determines the typically unique BS via which a user communicates. In homogeneous networks, a user usually associates to the BS that offers the best signal strength. However, such a rule does not work well in HetNets, mainly because of the power disparity between the MBS and the low power BSs. User association policy is known to impact the performance of HetNets greatly [4]. Unfortunately, performing optimal or close-to-optimal UA is not trivial.

Each of these three network processes introduced so far impact the throughput performance of a HetNet greatly. More importantly, these network processes have a complex interplay, which is not clearly understood. Clearly many combinations of these processes are potentially viable options and we propose a *unified framework* to study them and to enable their fair comparisons. We call a configuration the exact choice of RA and US as well as the type of low power BSs. Generically, ***[X-Y-Z]*** denotes a configuration where X is the RA (either CCD, OD, or PSD), Y is the type of low power BS (either R for relay or P for pico[1]) and Z is the type of US (either O for ON-OFF TC or NC for no coordination). For example, ***[CCD-P-O]*** represents a configuration with pico BSs operated under CCD with ON-OFF TC. For each configuration, UA is either performed optimally or is based on some simple practical rules that are defined later.

User scheduling can be performed locally at each BS independently if there is no TC among the BSs, while it has to be done globally for all other configurations to guarantee global optimality. Hence, even though a problem formulation based on local scheduling as proposed in [4] is a simpler and efficient formulation for solving ***[PSD-P-NC]*** and ***[CCD-P-NC]***, such a formulation does not extend to other configurations. In our study, we have used the concept of "independent sets" to unify the performance analysis of different configurations.

Our framework is based on a flow-model for a system "snapshot". Under a given system "snapshot", the system parameters like the channel gains and the number of users are fixed and known. A flow[2] corresponds to a data stream from the network to a particular user (since we focus on the downlink). We only consider the active users in the network and hence assume that there is one flow per user. Moreover, we assume that the users are greedy and hence want to maximize their individual flow-rates. Our framework allows us to configure the network parameters to allocate optimal throughputs to these flows. This is an offline-static study and thus is intended to be used at the engineering and planning phase to compare many potential configurations and decide which ones to study further. To make our framework tractable, we have made a key assumption of multipath routing, which is equivalent to allowing users to associate with more than one BS. We later validate that the upper-bounds provided by this assumption are in fact tight. We are now ready to summarize the main contributions of this work.

- We formulate a network flow-based unified framework for the performance evaluation of a multi-rate heterogeneous network comprising of one macro and a number of pico base stations and/or relay nodes. By assuming multi-association, we are able to solve the problems involving co-channel deployment as well as the orthogonal deployment to optimality, whereas we can solve the problems involving partially-shared deployment approximately.
- Using numerical results, we provide important engineering insights on the throughput performance of different configurations under a global proportional fairness (PF) objective function.
  - The upper bounds obtained under our multi-BS association assumption are tight and hence allowing a user to associate with more than one BS will not offer significant performance gains.
  - Under a pico deployment, PSD performs very well even in the absence of sophisticated transmission coordination whereas transmission coordination is essential for the satisfactory performance of CCD.
  - Deployment of relay nodes does not offer significant throughput gains under the type of backhauling mechanism that we consider.
  - A simple *pico-cell first* user association performs well even with ON-OFF TC if properly tuned. Its effectiveness under *no coordination* was shown earlier in [4].

The rest of the paper is organized as follows. In Section II, we summarize the related work. In Section III, we present the system model. In Section IV, the problem formulations are presented. In Section V, we discuss some simple user association rules and mention how these rules will be used to show the tightness of the upper-bounds provided by the multi-BS association assumption. In Section VI, we present some numerical results and discuss various engineering insights.

II. Related work

Even though RA, UA and TC are well-known network processes, these network processes are not usually studied

[1]We do not mix the types of BS in this study even though our framework allows such mixed deployments.

[2]This notion of flow is similar to the notion used in multi-commodity network-flow problems [5]. It is the same notion used in the existing literature on wireless networks in similar contexts (e.g., [6], [7]).

jointly. The literature however is rich with a large body of work on either one or few of these network processes. These studies either focus on pico deployment or relay deployment and these two scenarios are seldom considered together. We first summarize the related work on pico deployment. Under a conventional UA rule where a user associates to the BS offering the highest SINR, [8] studies the performance of different RA schemes. This UA rule is shown to result in poor performance [4]. Combining both RA and UA together and solving to optimality, however is not easy. Optimal UA problem alone has been studied in [9]. Requiring a user to associate to one BS results in Integer Problems and [10] studies a number of approximation algorithms to such problems by identifying them as generalized assignment problems (GAP). A joint UA and RA problem is studied in [11] with transmit power, UA, and US as variables, resulting in a complex combinatorial problem. [4] and [12] perform the joint optimization of RA, UA, and reuse pattern. Under optimal UA and local scheduling, they show that OD outperforms CCD. Their model however cannot be extended to systems with TC. Moreover, all of the above-mentioned works are applied to pico deployment and cannot handle relay deployment. A disjoint set of work is available for relay deployment scenarios. [13] studies the throughput-optimal scheduling policy derived from [14] where they find the queue-aware optimal scheduling policy that maximizes the user throughput, while maintaining the system stable. Some recent works in the literature extend the idea of UA from pico deployment to relay deployment. Under a local PF (proportional fair) scheduling, [15] formulates a user routing problem and presents a method to obtain a sub-optimal UA. [16] extends this work by considering the effect of backhaul link resource allocation and solves the problem approximately. [17] is a rare effort to compare the performance of pico base station deployment and relay node deployment. Their framework however does not incorporate some important network processes like the UA. To the best of our knowledge, there is no framework in the literature that can unify both the pico and the relay deployments, let alone the mixed deployments. Our framework tries to fill this gap. The notion of *independent sets* as used in our optimization framework is motivated from [18] where a throughput optimization model has been presented for wireless mesh networks. Adopting this approach allowed us to obtain exact solutions for small to medium size networks under a realistic SINR-based interference model. Efficient solution techniques might be warranted for large networks. [7] presents an efficient computation technique based on *column generation*.

III. General system description

We consider a system with one macro base station (MBS), X short range base stations (pico BSs or RNs), and N user equipments (UEs). MBS is represented as node 0. $\mathcal{P} = \{1, 2, \cdots, X\}$ represents the set of the short range base stations, which can either be pico base stations (PBSs) with wired backhaul to the network core, or relay nodes (RNs), or a mix of PBSs and RNs. *Pico deployment* refers to the network comprising of only the PBSs, and *relay deployment* refers to the network comprising of only the relay nodes. $\mathcal{U}$ and $\mathcal{N}$ represent respectively the set of all UEs, and all nodes (i.e., the MBS, PBSs/RNs, and UEs). We assume that there is a downlink flow corresponding to each UE originated at the MBS. $\mathcal{F}$ is the set of flows. Any flow $f \in \mathcal{F}$ is characterized by f_d (its destination node). The system has a set of M OFDM subchannels on a given frequency band with a per subchannel bandwidth of b. A set of commonly used notations as well as abbreviations are summarized in Table I.

TABLE I
Table of common notations and abbreviations

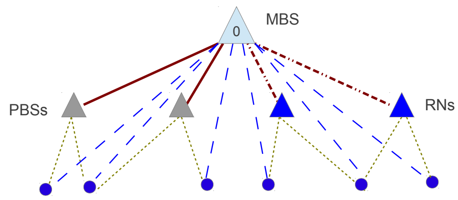
Notation/Abbreviation	Description
0	MBS
$\mathcal{P}$	Set of PBSs/RNs
X	Number of PBSs/RNs
$\mathcal{U}$	Set of UEs
N	Number of UEs
$\mathcal{N}$	Set of all nodes
$\mathcal{F}$	Set of Flows
$\mathcal{R}, \boldsymbol{\beta}$	Set of supported data rates and the corresponding SNR thresholds
$l = (o(l), d(l), r_m)$	A link with source $o(l)$, destination $d(l)$ and link rate r_m
L	Set of wired backhaul links
$\overline{L}$	Set of wireless backhaul links
$\overline{\mathcal{L}}_j$	Set of wireless links from BS j
P_{MBS}	Total power available to MBS
P_{PBS}	Total power available to PBS/RN
P_j	Power per subchannel for BS j
N_0	Noise power per subchannel
M	The number of OFDM subchannels
$G_{i,j}$	Channel gain between node i and j
RA	Resource Allocation
CCD	Co-channel Deployment (RA)
OD	Orthogonal Deployment (RA)
PSD	Partially Shared Deployment (RA)
US	User Scheduling
UA	User Association
TC	Transmission coordination
NC	No coordination (TC)
O	ON-OFF (TC)
[X-Y-Z]	A configuration of RA, TC, and deployment
$\mathcal{I}^X_{Z(Y)}$	Set of ISets for **[X-Y-Z]**

A. Assumptions

In order to simplify our formulations, we take the following assumptions.

- **[A1]** A BS, when transmitting, transmits on all subchannels that are allocated to it.
- **[A2]** Power allocated to a given BS is equally divided among the allocated subchannels. P_m represents the per-subchannel power of node m. The details are discussed in Section III-D.
- **[A3]** Channels are flat, i.e., the channel gains across different subchannels between a pair of nodes are equal. $G_{m,n}$ represents the channel gain between node m and n, assumed to be known either by measurements or a channel model. Besides this, we make no restricting assumptions on the channel gains.

It is easy to observe that the following is a consequence of **[A1]**, **[A2]** and **[A3]**.



Symbol	Name	Type	Subchannels for CCD	Subchannels for OD	Notation
	Backhaul links	Wired	Not applicable	Not applicable	L
	Backhaul links	Wireless	***M***	***M-K***	$\overline{L}$
	Direct links	Wireless	***M***	***M-K***	$\overline{\mathcal{L}}_0$
	Access links	Wireless	***M***	***K***	$\overline{\mathcal{L}}_j$

Fig. 1. Different types of links

Observation 1: **[A1]**, **[A2]**, and **[A3]** together make the individual subchannels across a pair of nodes identical and thus any scheduling based on both time and frequency domain can be equivalently reduced to a time domain scheduling where a BS allocates all of its subchannels to one UE at a given time.
This observation significantly reduces the complexity of the resulting optimization problems as will be evident while defining wireless links. However, this also means that we cannot exploit the channel-dependent scheduling aspect of an OFDM system in our framework.

B. Links

There are three types of links in the system (see Fig. 1). *MBS-UE links* are the wireless links from MBS to UEs. They are also called the *direct links*. *PBS-UE links* (respectively *RN-UE links*) are the wireless links from PBSs to UEs (respectively from RNs to UEs), which are also called the *access links*. *MBS-PBS links* (respectively *MBS-RN links*) are the *backhaul links* from MBS to PBSs (respectively MBSs to RNs). *MBS-PBS* backhaul links are wired. They are the logical representations of the wired connection of the PBSs to the backhaul network. On the contrary, MBS-RN backhaul links are wireless. Depending upon the RA mechanism, these wireless backhaul links might compete for the wireless resources (subchannels and/or time) with other wireless links in the network. Next, we define these links more formally.

The wired backhaul link from MBS to PBS j is represented as L_j with the link capacity C_j. L represents the set of all such wired backhaul links. A *wireless link* l is a tuple (j, i, r_m) (any direct, access or wireless backhaul link) where j and i are the origin node (denoted as $o(l)$) and destination node (denoted as $d(l)$) of the link, respectively. r_m is the associated link-rate (denoted as R_l). We note here that our assumptions have allowed us to define wireless links without the need to associate them with a specific subchannel. A wireless link l is *feasible* if the signal-to-noise ratio (SNR) received by $d(l)$ from $o(l)$ is greater than the threshold SNR required for supporting the rate R_l as defined in the modulation and coding Scheme (MCS). We take an MCS with a set of supported data rates $\mathcal{R} = \{r_1, r_2, \cdots, r_D\}$ and the corresponding SNR threshold set given as $\boldsymbol{\beta} = \{\beta(r_1), \beta(r_2), \cdots, \beta(r_D)\}$. The exact choices of $\mathcal{R}$ and $\boldsymbol{\beta}$ depend upon the rate function (representing the rate supported for a given SNR value) of the underlying physical layer technology. Our framework can thus be applied to different systems with different underlying physical technologies including the ones employing MIMO techniques for enhanced link rates. Under this MCS, the set of feasible wireless MBS-RN (*backhaul*) links ($\overline{L}$) is given as follows.

$$\overline{L} = \{(0, j, r_m) : j \in \mathcal{P}, r_m \in \mathcal{R}, \frac{P_0 G_{0,j}}{N_0} \geq \beta(r_m)\} \quad (1)$$

where N_0 is the additive white Gaussian noise power. $\mathcal{P}$, as defined earlier represents the set of RNs. We can also define the set of all feasible wireless links from BS j to the UEs (denoted as $\overline{\mathcal{L}}_j$) as follows.

$$\overline{\mathcal{L}}_j = \{(j, i, r_m) : i \in \mathcal{U}, r_m \in \mathcal{R}, \frac{P_j G_{j,i}}{N_0} \geq \beta(r_m)\}, \quad \forall j \in \mathcal{P} \cup \{0\} \quad (2)$$

The set of feasible *access* links is given as $\overline{\mathcal{L}} = \cup_{j \in \mathcal{P}} \overline{\mathcal{L}}_j$ and the set of feasible *direct* links is given as $\overline{\mathcal{L}}_0$. We note here that the links are logical and thus we can have multiple links for a given transmitter-receiver pair (these logical links differ in terms of the rate being supported). By definition, all wired links L_j are feasible. $\mathcal{L}$ represents the set of all feasible links. We do not allow links between any two PBSs/RNs or between any two UEs.

C. Resource allocation

RA dictates the exact way in which subchannels are allocated to the wireless links. Under pico deployment, RA affects two types of links, the *direct* and the *access* links. Under relay deployment however, an RA affects all three types of links (i.e., the *backhaul* links in addition to the *direct* and the *access* links). This in effect requires more complex considerations to be taken while dealing with the relay deployment, as will be evident below. For a given system, we study three types of RA: *co-channel deployment*, *orthogonal deployment*, and *partially shared deployment*.

1) Co-channel deployment (CCD): Under CCD, all wireless links are operating over all the M subchannels. Hence, there is no resource allocation parameter to configure. For each wireless link l, the number of subchannels allocated is given as $K_l = M$. Under pico deployment, the PBSs (which are the intermediate nodes in the downlink path) receive an incoming flow from the MBS via a wired link and hence a PBS can transmit to a UE at the same time while it is receiving a flow from the MBS. This is however not possible under relay deployment. Under relay deployment with CCD, the backhaul links and the access links operate on the same band and hence when a relay is receiving a flow from the MBS, it cannot transmit to a UE, and vice-versa. This is the result of the half-duplex communication constraints of the relay nodes. 3GPP identifies such an operation as *in-band* relay operation.

2) Orthogonal deployment (OD): Under pico deployment, OD corresponds to *channel splitting* where a set of K subchannels is allocated for PBS operation (i.e., the access links) and the remaining set of $M-K$ subchannels is dedicated for MBS operation (i.e., the direct links). Such an orthogonal set of frequencies at the two tiers allows for low interference operation. Additionally, a frequency reuse pattern could be used among the PBSs so as to guarantee low interference at the PBS-tier also. However, [4] has shown that if other network processes are chosen optimally, an aggressive full frequency reuse performs better than more conservative frequency reuse patterns. Accordingly, in our work, we consider that all K subchannels are used by each PBS (i.e., a frequency reuse factor of 1). Under this RA, K is a parameter to be configured and we call it the *channel split parameter*. More formally, with $\mathcal{P} = \{1, 2, \cdots, X\}$, K_l, the number of subchannels on which wireless link l can operate, is given as follows.

$$K_l = (M-K)\mathbf{1}_{\{o(l)=0\}} + K\mathbf{1}_{\{o(l)\in\mathcal{P}\}} \tag{3}$$

where $\mathbf{1}_{\{A\}}$ is an indicator function evaluating to 1 if statement A is true, and 0 otherwise. For relay deployment, we assume that a relay is half-duplex on the same band but is full duplex across bands (i.e., it can transmit or receive independently on each band). In that case, we could define an OD based on a splitting in three bands, i.e., K_1 subchannels for the access links, K_2 for the backhaul links and $M-K_1-K_2$ for the direct links. In that case, the relays could transmit and receive independently on each type of links. If we define only 2 bands K and $M-K$, we have to allocate the same band to two types of links. There are three possible combinations which will have different impact on the performance. In this study, we choose what we believe is a natural form of channel splitting in two bands where the direct and the backhaul links operate on $M-K$ subchannels whereas the access links operate on the remaining K subchannels. In this case, K_l is defined by (3). This choice results in the *out-of-band* operation where a relay is allowed to receive from the MBS while it is transmitting to an UE. Other splitting choices will be studied in the future.

3) Partially shared deployment (PSD): Under wired deployment, with PSD, K subchannels are allocated to each PBS and the remaining $M-K$ subchannels are dedicated to the MBS, as in OD. However, the MBS can also transmit in the K subchannels allocated to the PBSs at a lower power. Clearly, OD can be viewed as a special case of PSD when the MBS does not transmit at the K subchannels. Under relay deployment, PSD could be carried out in many ways, corresponding to all different ways in which OD could be carried out, as discussed before. In this work, similar to OD, we choose a particular splitting choice in which K subchannels are allocated to the RNs whereas the MBS can use these K subchannels (but at lower power) in addition to its dedicated $M-K$ subchannels.

For our modeling convenience, we introduce a dummy BS corresponding to the MBS when it is transmitting on the K shared subchannels. This dummy BS is represented as $0'$ and can be viewed as an additional PBS that is connected to the MBS (node 0) with a wired link $L_{0'}$ of infinite capacity, i.e., $C_{0'} = \infty$. Clearly, the set of PBSs/RNs under PSD includes $X+1$ elements, i.e., $\mathcal{P}' = \{1, 2, \cdots, X\} \cup \{0'\}$. The channel gains of the dummy BS correspond to the channel gains of the MBS, i.e., $G_{0',n} = G_{0,n}$.

D. Power allocation

MBS can transmit at the maximum total power of P_{MBS} and each PBS/RN can transmit at the maximum total power of P_{PBS}. Under CCD, the power per subchannel is chosen by assigning equal power to all of the allocated subchannels. Hence, for both pico and relay deployment, it is simply given by,

$$P_0 = \frac{P_{MBS}}{M}; \qquad P_j = P = \frac{P_{PBS}}{M}, \qquad \forall j \in \mathcal{P} \tag{4}$$

where P_j represents the power per subchannel for a BS j.

Under PSD, MBS allocates P' for transmission on the shared K subchannels and the remaining power $(P_{MBS} - P')$ for transmission on the dedicated $M-K$ subchannels. The power per subchannel for different BSs (MBS and PBSs/RNs) is simply given by,

$$P_{0'} = \frac{P'}{K}; \quad P_0 = \left(\frac{P_{MBS} - P'}{M-K}\right); \quad P_j = P = \frac{P_{PBS}}{K}, \quad \forall j \in \mathcal{P} \tag{5}$$

Recall that we decomposed MBS into node 0' (resp. node 0) transmitting on K (resp. $M-K$) subchannels. Clearly, OD corresponds to PSD with $P' = 0$.

E. Independent sets and US

US can be modeled using the notion of independent sets (ISets). We first describe our notion of independent set[3] before describing how we were able to translate the problem of different types of US and TC into the problem of scheduling a set of ISets. An ISet can be defined as a subset of feasible links which can be activated simultaneously without harmful interference at any of the destination nodes (i.e., all destination nodes can decode their received signals). The definition of an ISet should take into account the following facts.

- There is the "half-duplex" communication constraint for a wireless node on the same set of subchannels.
- Each link of an ISet has to yield a successful decoding which in turn requires each link to see an SINR greater than or equal to the minimum SINR required for the link-rate.
- Wired links can always be included in an ISet without causing infeasibility of other links.
- The characteristics of a particular US dictate the structure of an ISet (i.e., what kinds of and how many links can be included in an ISet).

Next, we construct the set of ISets for different configurations, starting with the case without coordination and then the case of ON-OFF coordination. The set of ISets for the configuration ***[X-Y-Z]*** is denoted as $\mathcal{I}^{X}_{Z(Y)}$.

[3]The notion of an independent set, as used in the paper, is different from the notion of independent set in graph theory.

1) No coordination: Under no-coordination, we activate all BSs all the time with their maximum power, as long as it is possible (as explained before, under CCD, a relay node cannot transmit when it is receiving from the MBS). We assume that each user has at least one feasible link to a BS when all BSs are transmitting simultaneously. *For pico deployment*, with a no-coordination scheme, an ISet under both CCD and PSD has a simple form. It contains all the wired links in L and one wireless link l from each BS (CCD has X PBSs and 1 MBS whereas PSD has $X+1$ PBSs and 1 MBS). Moreover, we require that no two wireless links of an ISet share a common destination node. More formally, we can define the set of ISets for ***[PSD-P-NC]*** as follows[4].

$$\begin{aligned}&\mathcal{I}^{PSD}_{NC(P)} = \{L \cup s \cup \{\tilde{l}\} : s \subset \overline{\mathcal{L}}, |s| = X+1, \tilde{l} \in \overline{\mathcal{L}}_0,\\ &\forall (l_i, l_j \in s, l_i \neq l_j),\\ &\qquad d(l_i) \neq d(l_j), o(l_i) \neq o(l_j), d(l_i) \neq d(\tilde{l}) \ \&\ (7)\}\end{aligned} \tag{6}$$

where (7) is the decoding constraint for the PSD operation, given as

$$\forall l_i \in s, \quad \frac{P_{o(l_i)} G_{o(l_i),d(l_i)}}{N_0 + \sum_{\substack{l_j \in s,\\ l_j \neq l_i}} P_{o(l_j)} G_{o(l_j),d(l_i)}} \geq \beta(R_{l_i}) \tag{7}$$

This constraint checks whether the SINR for a given link exceeds the minimum threshold required for supporting the link-rate (this is different from the feasibility constraint defined earlier, which is based on SNR). As interference does not affect the MBS-UE links, we check the decoding constraint of the PBS-UE links only. We have imposed a cardinality constraint on the set of PBS-UE links to be included in each ISet. We take exactly one PBS-UE link per PBS. We also impose the half-duplex communication constraint on a given band. The set of ISets for ***[CCD-P-NC]*** can also be defined similarly, except that there are X PBSs (i.e., $|s| = X$), and the decoding constraint of a link now needs to take into account the interference caused by MBS transmissions on the PBS transmissions, and vice-versa, i.e,

$$\forall l_i \in s \cup \{\tilde{l}\}, \quad \frac{P_{o(l_i)} G_{o(l_i),d(l_i)}}{N_0 + \sum_{\substack{l_j \in s \cup \{\tilde{l}\},\\ l_j \neq l_i}} P_{o(l_j)} G_{o(l_j),d(l_i)}} \geq \beta(R_{l_i}) \tag{8}$$

Under a relay deployment with no coordination, the structure of an ISet is more involved than that of pico deployment. Under relay deployment with CCD (***[CCD-R-NC]***), "some coordination" is required even under this so-called "no coordination" case. Under such a configuration, a RN has to "time-share" the reception from MBS and the transmission to UEs because the reception from MBS and the transmission to UEs take place in the same set of subchannels since the *in-band* mode is assumed. In this case, an ISet either contains *one MBS-UE link and X RN-UE links*, or *one MBS-RN link and $X-1$ RN-UE links*. More formally,

$$\begin{aligned}&\mathcal{I}^{CCD}_{NC(R)} = \{s \cup \{\tilde{l}\} : s \subset \overline{\mathcal{L}}, |s| = X - \mathbf{1}_{\{\tilde{l} \in \overline{L}\}}, \tilde{l} \in \overline{\mathcal{L}}_0 \cup \overline{L}\\ &\forall (l_i, l_j \in s, l_i \neq l_j),\\ &\qquad d(l_i) \neq d(l_j), o(l_i) \neq o(l_j), d(l_i) \neq d(\tilde{l}), o(l_i) \neq d(\tilde{l}) \ \&\ (8)\}\end{aligned} \tag{9}$$

Unlike under CCD, a relay under PSD can receive from MBS (operating on $M-K$ subchannels) *at the same time* as when it is transmitting to a UE (on the orthogonal K subchannels). This is the *out-of-band* operation. The resulting model contains one PBS (node $0'$) with a wired backhaul of infinite capacity in addition to the X RNs with wireless backhaul. Under this mixed mode of deployment with both wired and wireless backhaul links, an ISet either contains *one MBS-UE link, $X+1$ access links and one wired backhaul link*, or *one wireless MBS-RN link, $X+1$ access links and one wired backhaul link*. More formally,

$$\begin{aligned}&\mathcal{I}^{PSD}_{NC(R)} = \{s \cup \{\tilde{l}\} \cup \{L_{0'}\} : s \subset \overline{\mathcal{L}}, |s| = X+1, \tilde{l} \in \overline{\mathcal{L}}_0 \cup \overline{L}\\ &\forall (l_i, l_j \in s, l_i \neq l_j),\\ &\qquad d(l_i) \neq d(l_j), o(l_i) \neq o(l_j), d(l_i) \neq d(\tilde{l}) \ \&\ (7)\}\end{aligned} \tag{10}$$

2) ON-OFF coordination: Under ON-OFF US, different groups of BSs can be activated for different fractions of time. An optimal ON-OFF coordination obtains the exact proportion of times for which these groups of BSs have to be activated for maximizing a chosen objective. Under PSD, the PBS/RN transmissions do not interfere with an MBS (node 0) transmission and vice-versa. This means that the MBS can be activated all the time in the $M-K$ subchannels without interfering on the other links (i.e., the access links). Hence, even under ON-OFF coordination, the MBS transmits all the time and a group of PBSs/RNs $g \subseteq \mathcal{P}$ is activated for a proportion of time α_g. Using this property, we can construct the set of ISets for the PSD. For example,

$$\begin{aligned}&\mathcal{I}^{PSD}_{O(R)} = \{s \cup \{\tilde{l}\} \cup \{L_{0'}\} : s \subset \overline{\mathcal{L}}, \tilde{l} \in \overline{\mathcal{L}}_0 \cup \overline{L},\\ &\forall (l_i, l_j \in s, l_i \neq l_j),\\ &\qquad d(l_i) \neq d(l_j), o(l_i) \neq o(l_j), d(l_i) \neq d(\tilde{l}) \quad \&\ (7)\}\end{aligned} \tag{11}$$

Compared to (10), we now do not restrict the set of access links in each ISet to have the cardinality of $X+1$. We can define the set of ISets for ***[PSD-P-O]*** similarly. Under CCD however, the MBS interferes with the PBS/RN transmissions (and vice-versa). Thus, with ON-OFF coordination, at any time a group of BSs $g \subseteq \{0\} \cup \mathcal{P}$ is activated for a proportion of time α_g. We thus define the set of ISets for CCD by allowing the MBS to potentially turn-off if required and by removing the restriction on cardinality. For example,

$$\begin{aligned}&\mathcal{I}^{CCD}_{O(R)} = \{s : s \subset \overline{\mathcal{L}} \cup \{\tilde{l}\}, \tilde{l} \in \overline{\mathcal{L}}_0 \cup \overline{L},\\ &\forall (l_i, l_j \in s, l_i \neq l_j),\\ &\qquad d(l_i) \neq d(l_j), o(l_i) \neq o(l_j), d(l_i) \neq d(\tilde{l}), o(l_i) \neq d(\tilde{l}) \ \&\ (8)\}\end{aligned} \tag{12}$$

We can define the set of ISets for ***[CCD-P-O]*** similarly.

Clearly, ON-OFF transmission coordination provides a greater degree of freedom in terms of interference mitigation

[4] $|\mathcal{A}|$ represents the cardinality of set $\mathcal{A}$.

as compared to the systems without coordination (noting that $\mathcal{I}^A_{NC(B)} \subseteq \mathcal{I}^A_{O(B)}$ for a given RA A and deployment B) and hence ***[PSD-P-O]***, ***[PSD-R-O]***, ***[CCD-P-O]***, and ***[CCD-R-O]*** can only perform better than their respective "no coordination" counterparts. Now, after defining precisely the set of ISets corresponding to each configuration, we are ready to model the US using the idea of ISet activation schedule. If α_s is the fraction of time ISet s is scheduled, then for a given set of ISets $\mathcal{I}$, the problem of user scheduling (in ***[PSD-P-NC]***, ***[PSD-R-NC]***, ***[CCD-P-NC]***, and ***[CCD-R-NC]***) and the problem of user scheduling and ON-OFF transmission coordination (in ***[PSD-P-O]***, ***[PSD-R-O]***, ***[CCD-P-O]***, and ***[CCD-R-O]***) are equivalent to finding the optimal values of $(\alpha_s)_{s\in\mathcal{I}}$ to optimize the objective function.

By properly defining the set of ISets, we have effectively unified *pico* and *relay* deployment scenarios, different types of RA, and different types of US. The rest of the model elements and ideas apply equally to different configurations of both pico and relay deployment.

IV. Problem formulation

In this section, we first model user association as flow-routing and present the optimization problems for PSD ($\tilde{\mathbf{P}}_{\mathbf{PSD}}$) and for CCD ($\tilde{\mathbf{P}}_{\mathbf{CCD}}$).

A. Routing variables under multipath routing

We incorporate user association into our framework by introducing the "routing variables" $(x^f_l)_{f\in\mathcal{F},l\in\mathcal{L}}$ where x^f_l represents the amount of flow f routed through link l. Typically a user associates to exactly one BS. Such a *single-association* would then impose single-path routing constraints on the routing variables which would thus result in an Integer Problem (IP), which is very hard to solve. While formulating our framework, for tractability, we make the a priori unrealistic assumption that a user can associate to multiple BSs. Clearly, such a *multi-association* can be modeled under a multipath routing framework. Such an assumption yields a much more tractable model and the solution based on optimal multipath routing is an upper bound to the optimal single-association solution. It is however unclear a priori if such an upper-bound is tight. We will later show that it is indeed the case.

B. Optimization problem for PSD

Let $\mathcal{I}$ be the set of ISets for a given PSD configuration characterized by the type of US. Given $\mathcal{I}$, P_{MBS}, P_{PBS}, the set of flows $\mathcal{F}$, the MCS (with rates $\mathcal{R}$ and the SINR thresholds $\boldsymbol{\beta}$) in a network comprising of one MBS, X (physical) PBSs/RNs and N UEs with known channel gains $(G_{j,i})$ and M subchannels each with a bandwidth of b, $\tilde{\mathbf{P}}_{\mathbf{PSD}}$ represents the problem to find the optimal parameters α_s (the fractions of time each ISet is scheduled), x^f_l (the routing parameters), K (the resource allocation parameter), and P' (the power used by the MBS on the shared band) such that $\sum_{f\in\mathcal{F}}\log(\lambda_f)$ is maximized where λ_f represents the throughput allocated to flow f. Such a throughput allocation is called a global proportional fair allocation.

$$\tilde{\mathbf{P}}_{\mathbf{PSD}}: \max_{(\alpha_s),(x^f_l),K,P',(\lambda_f)} \sum_{f\in\mathcal{F}}\log(\lambda_f) \tag{13}$$

subject to:

$$\sum_{l\in\mathcal{L}:o(l)=n} x^f_l - \sum_{l\in\mathcal{L}:d(l)=n} x^f_l = \lambda_f \mathbf{1}_{\{n=0\}} - \lambda_f \mathbf{1}_{\{n=f_d\}}, \quad \forall n\in\mathcal{N}, \forall f\in\mathcal{F} \tag{14}$$

$$\sum_{f\in\mathcal{F}} x^f_l \le bK_l \sum_{s\in\mathcal{I}} R_l\alpha_s \mathbf{1}_{\{l\in s, l\notin L\}} + C_{d(l)}\mathbf{1}_{\{l\in L\}}, \forall l\in\mathcal{L} \tag{15}$$

$$K_l = (M-K)\mathbf{1}_{\{o(l)=0\}} + K\mathbf{1}_{\{o(l)\in\mathcal{P}\}}, \qquad \forall l\in\mathcal{L} \tag{16}$$

$$\sum_{s\in\mathcal{I}}\alpha_s \le 1 \tag{17}$$

Eqs. $(1);(2);(5);$

$$K\in\{0,1,\cdots,M\}; x^f_l, \alpha_s, \lambda_f \ge 0, \forall f\in\mathcal{F}, \forall l\in\mathcal{L}, \forall s\in\mathcal{I} \tag{18}$$

(14) are the flow conservation constraints. They are the fundamental constraints that the routing variables have to satisfy. Additionally, the amount of flow that can be routed through a given link is further constrained by the link capacity constraint (15) which guarantees that the sum of the amount of all flows at a link cannot exceed the scheduled link capacity. The scheduled capacity of a wireless link is given by the product of the number of subchannels allocated to it, the per subchannel bandwidth and the link-rate multiplied by the fraction of times it is scheduled. The scheduled capacity of a wired link is simply the capacity of the link since it is scheduled all the time. These capacity constraints couple the flow routing (user association) with transmission coordination and user scheduling. (16) represents the channel splitting mechanism. (17) is the scheduling constraint. By taking the appropriate definition for $\mathcal{I}$, $\tilde{\mathbf{P}}_{\mathbf{PSD}}$ results in the optimization problem for a specific configuration.

$\tilde{\mathbf{P}}_{\mathbf{PSD}}$ is a non-convex problem and thus solving it to global optimality is difficult. The non-convexity of the problem is related to the following.

- Power allocation (P'): P' is the power to be used by the MBS to transmit in the overlapped set of K subchannels. As long as it is a variable, the problem becomes largely intractable as it is not possible to determine a priori the set of feasible wireless links.
- Power splitting: The SINR of a link is a non-linear function of the channel-split parameter K. The feasible region is thus not convex as long as K is a variable, due to the general lack of convexity of the rate function on non-linear SINR functions.
- Non-convex capacity constraint (15).

The first level of simplification can be obtained by setting P' to a fixed value. Let $\tilde{\mathbf{P}}_{\mathbf{PSD}}(P')$ be the problem with a fixed P'. Even by fixing P', the problem still remains non-convex. However, if we fix K in addition to P', we can define a set of parameterized problems $\mathbf{P}_{\mathbf{PSD}}(K,P')$. $\mathbf{P}_{\mathbf{PSD}}(K,P')$

contains only the linear constraints (obtained by fixing K) with a strictly concave maximization function, hence, it is a convex optimization problem and thus it can be solved efficiently for small to medium sized problems. The solution to $\tilde{\mathbf{P}}_{\mathbf{PSD}}(P')$ can then be derived as $\mathbf{P}_{\mathbf{PSD}}(K^*, P')$ where $K^* = \arg\max_{K\in\{0,1,\cdots,M\}} \mathbf{P}_{\mathbf{PSD}}(K, P')$. It is important to note here that as P' is a continuous variable, it is not further possible to represent the solution of $\tilde{\mathbf{P}}_{\mathbf{PSD}}$ as the maximal solution of the set of parameterized problems $\tilde{\mathbf{P}}_{\mathbf{PSD}}(P')$. We will restrict the values of P' to a discrete set.

C. Optimization problem for CCD

CCD does not have a resource allocation parameter nor the power allocation variable P'. All M subchannels are allocated to all the BSs. This simplifies the resulting problem greatly. The optimization problem for CCD can be obtained by replacing (16) with $K_l = M$ for all $l \in \mathcal{L}$ and setting $P' = 0$ on $\tilde{\mathbf{P}}_{\mathbf{PSD}}$. The lack of channel split variable and P' makes the resulting problem (represented as $\tilde{\mathbf{P}}_{\mathbf{CCD}}$) a convex optimization problem.

The number of ISets grows in the order of $O(2^{|\mathcal{L}|})$, making the problem NP-hard. However, the size of an ISet in our cellular system is upper-bounded by the number of transmitters (i.e., $X+1$ for CCD and $X+2$ for PSD). Thus, our model can be used to obtain exact solutions for networks of practical size, i.e., with a number of PBSs/RNs less than 8.

V. Simple UA rules and validation of the models

So far, we have assumed multi-association. However, conventionally, a UE associates to *one* BS only. For the results based on multi-association to be useful, we need to validate that the upper-bounds obtained under this assumption are tight. In an attempt to do so, we study three simple UA rules, namely *SINR-based*, *Range Extension* and *Pico-cell first(δ)*, each of which defines simple rules to determine the unique BS to which an UE associates to. Under *SINR-based* UA, an UE associates to the BS that offers the highest SINR. This approach is shown to work poorly due to the power disparity between MBS and the PBSs, thereby resulting in overloaded MBS [4]. Range extension (RE) tries to solve this problem of power disparity by associating a user to the BS with the smallest path-loss [19]. A third rule called the *pico-cell first* (PCF) rule with a parameter δ, first introduced in [4], says that if a UE finds a PBS offering an SINR greater than or equal to δ, it should associate to the PBS even if MBS might offer a greater SINR. All of these three rules are simple in the sense that they do not involve any real-time load-balancing and are easy to calculate (each UE can do it itself). They also provide feasible single-association solutions and thus provide the lower bounds on the optimal single-association solution. These UA rules can be applied to our earlier problem by translating the association structure into the routing variables (x_l^f) of our model. As an example, let UE i associates to BS j under the given association rule. Then the corresponding flow routing variable x_l^f (where flow f is the downlink flow to user i and thus $f_d = i$) will be 0 for all wireless links l that do not belong to BS j (i.e., $o(l) \neq j$). Once x_l^f captures the user association structure imposed by this rule, we can easily compute the other parameters by using our problem formulations.

Studying these simple UA rules serves us with two purposes. The first is to obtain lower-bounds so that we can validate our upper-bounds. The second is to understand how these simple UA rules perform. In the absence of transmission coordination, [4] already shows that PCF(δ) works well. Our study allows us to see whether this observation extends to the case of ON-OFF TC as well.

VI. Numerical Results

We consider a $500m \times 500m$ square as the user deployment area with an MBS placed at the center. We consider scenarios with $X = 2$, 3, 4 and 6 PBSs/RNs deployed as shown in Fig. 2. The path loss $\gamma_{j,i}$ for the transmitter-receiver pair (j, i) separated by a distance d_{ji} (m), which is a path-loss model recommended by 3GPP, is given in Table II, together with the appropriate values of antenna-gains and miscellaneous losses [3]. We further apply a log-normal shadowing with zero mean and standard deviation of 8 dB to obtain the random path-loss $\overline{\gamma}$. i.e., $\overline{\gamma}_{j,i} = \gamma_{j,i} + \mathbf{N}(0, 8)$ where $\mathbf{N}(\mu, \sigma)$ is a normal random variable with mean μ and standard deviation σ. The channel gains can then be obtained as $G_{j,i} = 10^{-\frac{\overline{\gamma}_{j,i}}{10}}$. We take $P_{MBS} = 46dBm$, $N_0 = -112.4245dBm$, and $M = 100$ subchannels each with $b = 180KHz$ bandwidth. We consider an adaptive MCS with 15 discrete rates. The rates and the corresponding required threshold SNRs are listed in Table III. The rates have the unit of bits per symbol. An OFDM symbol of a particular system is constant and hence our rates can be seen as being normalized by the OFDM symbol duration. We assume now that the wired MBS-PBS backhaul links are not the bottleneck and thus we consider these wired links to be of infinite capacity. For each scenario of X PBSs/RNs and N UEs, a network realization is obtained by generating N uniformly distributed random user positions in the deployment area. For each X and N, we have studied 100 such random realizations of the network. We obtain the numerical results by solving the convex optimization problems formulated earlier to global maximum for each realization by using the commercial solver, Minos™ [20].

PF is known to maximize the geometric mean (GM) of the throughput of the users, given as $\left(\prod_{f\in\mathcal{F}} \lambda_f^*\right)^{\frac{1}{N}}$. Thus, we take the GM throughput as the performance metric to compare different configurations. Obtaining the GM throughput results for PSD is not as straightforward as in CCD or OD as the performance of PSD depends upon the power P'. However, as we discussed earlier, solving for optimal P' is a difficult problem and our models developed so far can obtain the GM throughput only when P' is given. In order to obtain good performance gains for PSD, we coarsely tune P' by selecting the best power from the set of power choices from -10 dBm to 30 dBm at 1 dBm interval. All the results shown for PSD are obtained for the best P' from this set. For comparison, we also obtain results for OD (i.e., PSD with $P' = 0$ W).

Throughput gain for each configuration on a particular realization is computed over the case when PBSs/RNs are not deployed. For a particular realization i, the throughput

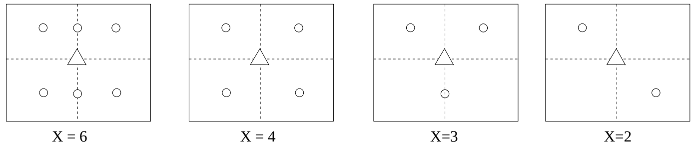


Fig. 2. X PBSs or relay nodes placed in a grid layout on a macro coverage of a $500m \times 500m$ square

TABLE II
PATH-LOSS MODEL

Transmitter	Link (j, i)	Path-loss at the medium $(\phi_{j,i})$ (dB)	Antenna gain (AG_j) (dB)	Cable and other losses (ζ_j) (dB)
MBS	$(0, i)$	$128.1 + 37.6 \log_{10}\left(\frac{d_{0i}}{1000}\right), d_{0i} \geq 35m$	15	20
PBS	$(j, i) : j \in \{1, 2, \cdots X\}$	$140.7 + 36.7 \log_{10}\left(\frac{d}{1000}\right) (dB), d_{ji} \geq 10m$	5	20
Total path-loss $(\gamma_{j,i})$				
$\gamma_{j,i} = \phi_{j,i} + \zeta_j - AG_j$				

TABLE III
AVAILABLE RATES AND THE CORRESPONDING SNR THRESHOLDS

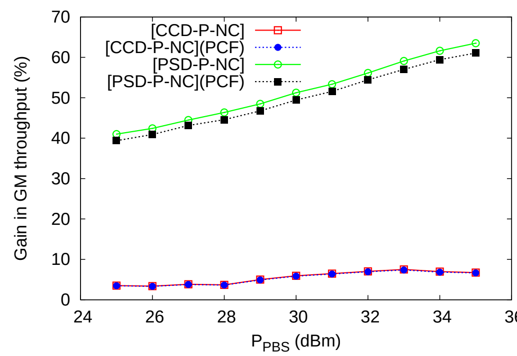
Threshold SNR (dB)	-6.5	-4	-2.6	-1	1	3	6.6	10	11.4	11.8	13	13.8	15.6	16.8	17.6
Efficiency (bits/symbol)	0.15	0.23	0.38	0.60	0.88	1.18	1.48	1.91	2.41	2.73	3.32	3.9	4.52	5.12	5.55

gain obtained by configuration Y is given by $\mathcal{G}_Y(i) = 100 \times \frac{\lambda_Y^{GM}(i) - \lambda_0^{GM}(i)}{\lambda_0^{GM}(i)}$ where $\lambda_Y^{GM}(i)$ is the GM throughput of realization i under configuration Y. $Y = 0$ corresponds to the base-case with the MBS only. In order to characterize the average gain in throughput performance of each configuration, we obtained the average gain in GM throughput over all of our 100 random realizations.

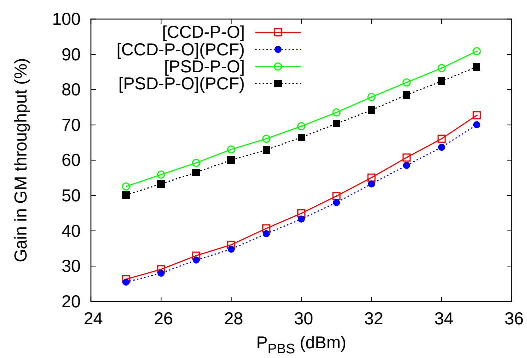
A. Validation of the upper bounds

Before continuing with the performance comparison of different configurations, we validate our assumption of multi-association with the help of a feasible single-association solution as discussed below.

As will be discussed later, PCF(δ) yields the best performance of the three UA schemes that we studied. Hence, we present the results obtained under PCF(δ) and show these results along-side the results obtained with optimal multi-association, averaged over the 100 realizations . In order to get the "best" lower bound, for each configuration and each realization, we select the value of δ that provides the best performance in terms of the GM throughput. In Figs. 3 and 4, we plot the average gain in GM throughput for different configurations with a fine-tuned PCF(δ) as well as with the optimal multi-association for the scenario with $X = 4$ PBSs/RNs and $N = 75$ UEs. In Fig. 3, the four low power BSs are pico base stations while they are relays in Fig. 4. In our framework, we assume multi-association, even when in practice, a user associates to a single BS. Our optimal multi- association yields an upper bound to single-association whereas the (sub-optimal) PCF based association provides a feasible single-association and hence yields a lower bound to the optimal single-association. The results in Figs. 3 and 4 show that the performance of PCF, in terms of the gain in GM throughput with respect to the base-case, averaged over 100 realizations, is within 4% of the performance with optimal multi-association, across all configurations. Moreover, the gap between the lower-bound and the upper-bound was less



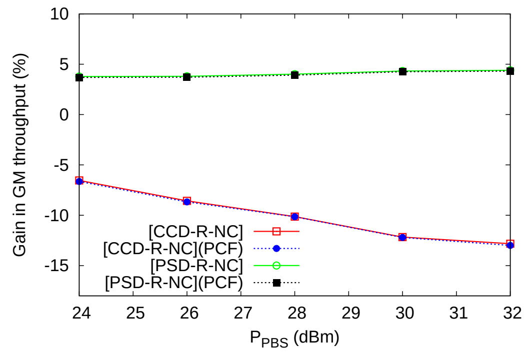
(a) **No coordination**



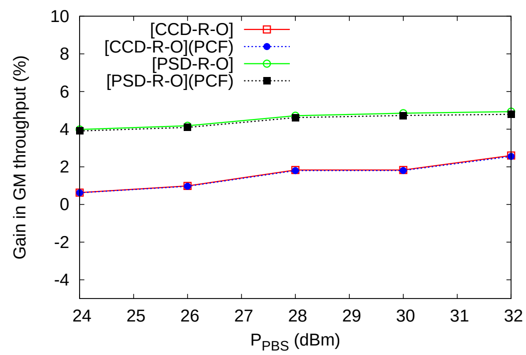
(b) **ON-OFF TC**

Fig. 3. Pico deployment: Average gain in GM throughput over 100 realizations for optimal and PCF association - $X = 4$ and $N = 75$

than 5% for at least 95% of the realizations that we studied. The numerical closeness of the two bounds thus validates that the results obtained by considering multi-association are tight bounds for the optimal single-association. Moreover, it also means that the optimal multi-association does not provide much performance gains over the optimal single-association. Hence, introducing multi-association capabilities will not offer significant performance gains. We conducted similar compu-



(a) **No coordination**



(b) **ON-OFF TC**

Fig. 4. Relay deployment: Average gain in GM throughput over 100 realizations for optimal and PCF association - $X = 4$ and $N = 75$

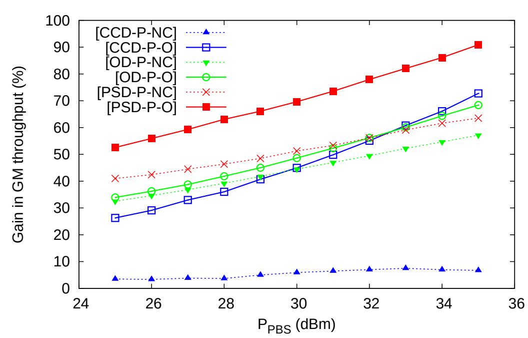


Fig. 5. Average gain in GM Throughput over 100 realizations under pico deployment - $X = 4$ and $N = 75$

tations for 100 cases of non-uniformly distributed users and randomly deployed PBSs/RNs, and obtained similar results.

B. Engineering insights: pico deployment

We present the average gain in GM throughput obtained by different configurations in Fig. 5 as a function of P_{PBS} for $X = 4$ and $N = 75$. Next, we discuss these results.

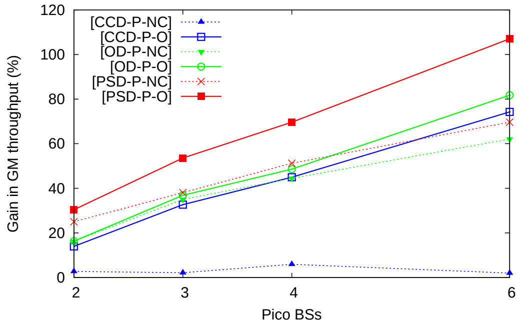
1) No coordination case: When P' is chosen properly, PSD clearly offers the best throughput performance among all the three RA mechanisms that we have considered. As evident from Fig. 5, PSD outperforms OD. The gains obtained by PSD over OD can simply be attributed to the added flexibility of allowing MBS to use more channels at a carefully chosen power P'. It is however important to stress that any PSD is not guaranteed to perform better than OD if the power P' is not chosen carefully. CCD, on the other hand, performs very poorly in the case of no transmission coordination. Both PSD and OD outperform CCD significantly. In fact, the deployment of PBSs under CCD provides very little gains (less than 8%) to the MBS-only deployment. Clearly, co-channel deployment, though attractive due to its simplicity, is an unacceptable resource allocation mechanism in the absence of transmission coordination.

2) ON-OFF Transmission coordination case: For a given RA, allowing ON-OFF TC can only improve over the case with no coordination. Our results show that the magnitude of improvements brought by ON-OFF TC are significant, especially for CCD. Under ON-OFF transmission coordination, PSD continues to perform significantly better (15 to 20 %) than CCD. More important perhaps is the observation that the relative performance of CCD under ON-OFF transmission coordination is very different from its performance under no coordination. CCD is a simple resource allocation mechanism as it does not require the configuration of any resource allocation parameter. The good performance of CCD under ON-OFF transmission coordination might motivate us to consider CCD as a favorable choice. However, we have seen that CCD requires transmission coordination, or otherwise performs too poorly to justify its simplicity. PSD, on the other hand, performs very well even without transmission coordination as evident from the comparison of the performance of ***[PSD-P-NC]*** with the performance of ***[CCD-P-O]***. ON-OFF transmission coordination involves a problem of exponential complexity and requires a much fine-grained control as compared to computing the optimal channel split parameter K (with no coordination). Thus, our results favor PSD over CCD.

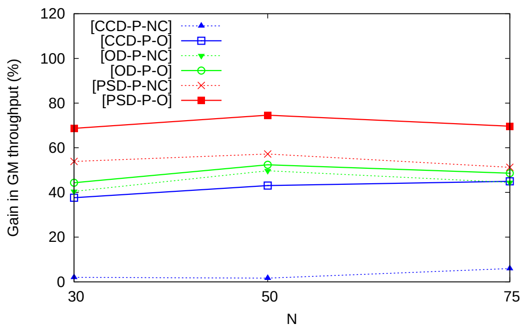
3) Different number of PBSs and UEs: Fig. 6(a) shows the performance of different configurations for $N = 75$ and $P_{PBS} = 30dBm$ for different numbers of PBSs deployed. The performance of all configurations except ***[CCD-P-NC]*** improve with more PBSs deployed. Notable is the result that with increasing number of PBSs, the gains due to ON-OFF TC increases for each RA scheme.

Fig. 6(b) shows the performance of different configurations for $X = 4$ and $P_{PBS} = 30dBm$ for different values of N. The results show that the performance in terms of throughput gains do not change significantly with the number of UEs in the system.

4) Performance of different UA rules: In Fig. 7, we show the performance of the three simple UA rules along with the optimal multi-association for ***[PSD-P-O]*** and ***[CCD-P-O]*** for $X = 4$ and $N = 75$. The results for PCF(δ) are obtained for a fine-tuned δ. This result shows that if properly configured, the performance of PCF is adequate and that it outperforms both SINR-based and range extension based UA rules. Similar conclusion was reported in [4] for the case of no coordination.

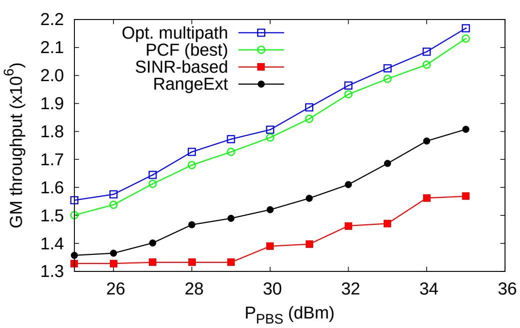


(a) Different number of PBSs (X) with $N = 75$

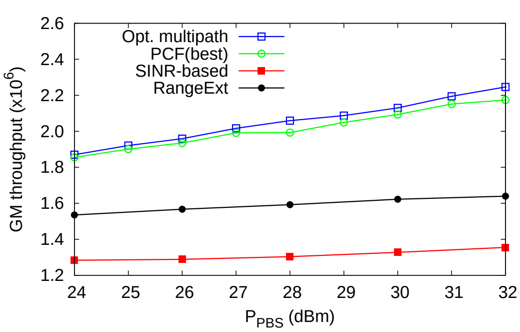


(b) Different number of UEs (N) with $X = 4$

Fig. 6. Average gain in GM throughput over 100 realizations for pico deployment, with $P_{PBS} = 30dBm$

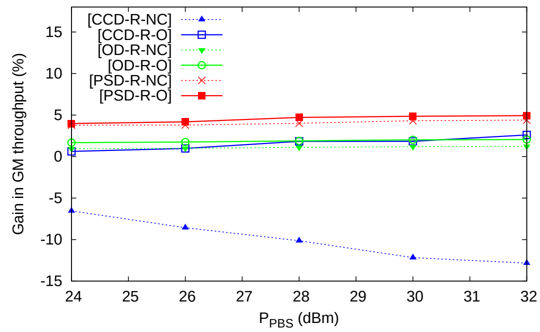


(a) ***[CCD-P-O]***

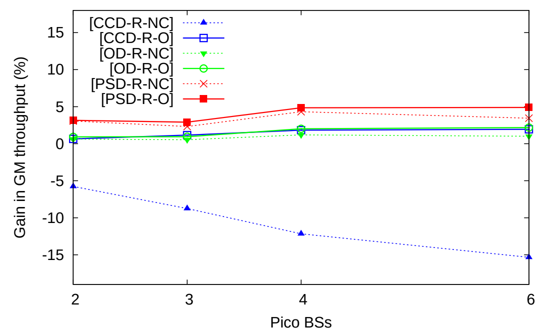


(b) ***[PSD-P-O]***

Fig. 7. Comparison of different UA rules (pico deployment) with $X = 4$ and $N = 75$ - one realization (PCF is carried out for a fine-tuned δ)



(a) Different values of P_{PBS}, $X = 4$, $N = 75$



(b) Different number of RNs, $P_{PBS} = 30dBm$, $N = 75$

Fig. 8. Average gain in GM throughput over 100 realizations for relay deployment

C. Engineering insights: relay deployment

We also carried out the performance analysis for the relay deployment. As evident from Fig. 8(a), which shows the gains for $X = 4$ and $N = 75$, the deployment of relays does not offer significant gain on the geometric mean throughput over the base-case in any of the configurations that we have considered. Moreover, Fig. 8(b) shows the gains for scenarios with different number of RNs. The reported gains are marginal ($\leq 5\%$) even for the best configuration ***[PSD-R-O]***. It is important to note however that we have considered a particular form of channel splitting while defining OD and PSD, and the performance of other channel splitting choices are unknown. This negative result means that even if deploying relay nodes might be easier and less costly than deploying regular pico base stations, network operators should not expect a gain in network throughput. Since this insight is based on the upper-bounds, this negative conclusion is fairly robust. We had also obtained results for different values of N. Similar to Fig. 6(b), the impact of N on the throughput-gain was found to be small.

VII. Conclusion

In this paper, we formulated a flow-based framework for the joint optimization of resource allocation, transmission coordination, and user association in a heterogeneous network comprising of a macro base station and a set of pico base stations and/or relay nodes. In addition to providing a useful framework for the offline study of heterogeneous networks, we also obtained important engineering insights. Our results showed that the gain offered by multi-association as compared to the optimal single-association is small for both pico and relay deployments. In pico deployment scenario, our numerical results showed that co-channel deployment requires transmission coordination for a satisfactory performance whereas partially shared deployment performs well even in the absence of sophisticated transmission coordination mechanism. PSD, thus can still be a better practical approach as compared to CCD which incurs complexity on transmission coordination. Under relay deployment, no meaningful gain on throughput performance was obtained for any configuration that we had studied. We have chosen a particular way of channel splitting for relay deployment. Other splitting choices are also possible and might perform differently.

REFERENCES

[1] UMTS Forum, "Mobile traffic forecasts, 2010-2020 report, report number 44," January 2011.

[2] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *Wireless Communications, IEEE*, vol. 18, no. 3, pp. 10 –21, June 2011.

[3] 3GPP-TSG-RAN-WG1, "Evolved universal terrestrial radio access (EU-TRA)," *3GPP, Tech. Rep. TR 36.814*, 2010.

[4] D. Fooladivanda, A. A. Daoud, and C. Rosenberg, "Joint channel allocation and user association for heterogeneous wireless cellular networks," in *Proc. IEEE PIMRC*, Sept. 2011.

[5] J. L. Kennington, "A survey of linear cost multicommodity network flows," *Operations Research*, vol. 26, no. 2, pp. pp. 209–236, 1978. [Online]. Available: http://www.jstor.org/stable/169631

[6] S. Shabdanov, P. Mitran, and C. Rosenberg, "Cross-layer optimization using advanced physical layer techniques in wireless mesh networks," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 4, pp. 1622 –1631, april 2012.

[7] J. Luo, C. Rosenberg, and A. Girard, "Engineering wireless mesh networks: Joint scheduling, routing, power control, and rate adaptation," *IEEE/ACM Trans. Netw.*, vol. 18, no. 5, pp. 1387–1400, 2010.

[8] W. Lei, W. Hai, Y. Yinghui, and Z. Fei, "Heterogeneous network in LTE-advanced system," in *Proc. IEEE ICCS*, Nov. 2010.

[9] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *Proc. IEEE INFOCOM*, April 2006, pp. 1 –12.

[10] D. Amzallag and D. Raz, "Resource allocation algorithms for the next generation cellular networks," in *Algorithms for Next Generation Networks*. Springer London, 2010, pp. 99–129.

[11] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1479 –1489, december 2010.

[12] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous wireless cellular networks," *accepted in IEEE trans. on Wireless Comm.*, Oct. 2012.

[13] H. Viswanathan and S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling," in *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 3, Oct. 2003, pp. 1923 – 1928 Vol.3.

[14] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," in *Decision and Control, 1990., Proceedings of the 29th IEEE Conference on*, Dec 1990, pp. 2130 –2132 vol.4.

[15] W.-G. Ahn and H.-M. Kim, "Proportional fair scheduling in relay enhanced cellular OFDMA systems," in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, Sept. 2008, pp. 1 –4.

[16] Z. Ma, W. Xiang, H. Long, and W. Wang, "Proportional fair-based in-cell routing for relay-enhanced cellular networks," in *Wireless Communications and Networking Conference (WCNC), 2011 IEEE*, March 2011, pp. 381 –385.

[17] A. B. Saleh, S. Redana, J. Hämäläinen, and B. Raaf, "Comparison of relay and pico eNB deployments in LTE-Advanced," in *VTC Fall*, 2009.

[18] A. Karnik, A. Iyer, and C. Rosenberg, "Throughput-optimal configuration of fixed wireless networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 5, pp. 1161–1174, 2008.

[19] A. Khandekar, N. Bhushan, J. Tingfang, and V. Vanghi, "LTE-Advanced: Heterogeneous networks," in *Wireless Conference (EW), 2010 European*. IEEE, 2010, pp. 978–982.

[20] "Minos 5.5," [Online], Website, http://www.sbsi-sol-optimize.com/asp/solproductminos.htm.

Jagadish Ghimire received his B.E. degree in Electronics and Communications Engineering from Tribhuvan University (TU), Nepal, in 2007. He received the M.Sc. degree in Communication Networks and Services from Telecom SudParis, France, in 2009, and his M.E. degree in Information and Communication Technologies from Asian Institute of Technology (AIT), Thailand, in 2009. He is currently a PhD candidate in the Department of Electrical and Computer Engineering at the University of Waterloo, Waterloo, Canada. His current research activities focus on modeling and performance evaluation of heterogeneous wireless networks.

Catherine Rosenberg was educated in France (Ecole Nationale Supérieure des Télécommunications de Bretagne, Diplôme d'Ingénieur in EE in 1983 and University of Paris, Orsay, Doctorat en Sciences in CS in 1986) and in the USA (UCLA, MS in CS in 1984), Dr. Rosenberg has worked in several countries including USA, UK, Canada, France and India. In particular, she worked for Nortel Networks in the UK, AT&T Bell Laboratories in the USA, Alcatel in France and taught at Purdue University (USA) and Ecole Polytechnique of Montreal (Canada). Since 2004, Dr. Rosenberg is a faculty member at the University of Waterloo where she now holds a Tier 1 Canada Research Chair in the Future Internet. Her research interests are broadly in networking with currently an emphasis in wireless networking and in smart energy systems. She has authored over 150 papers and has been awarded eight patents in the USA. She is a Fellow of the IEEE. More information can be found at http://ece.uwaterloo.ca/∼cath/.