"Wireless Networks Without Edges": Dynamic Radio Resource Clustering and User Scheduling

Yuhuan Du and Gustavo de Veciana
Department of Electrical and Computer Engineering, The University of Texas at Austin
Email: dyhuan123@gmail.com, gustavo@ece.utexas.edu

***Abstract*—Cellular systems using Coordinated Multi-Point (CoMP) transmissions leveraging clusters of spatially distributed radio antennas as Virtual Base Stations (VBSs) have the potential to realize overall throughput gains and, perhaps more importantly, can deliver substantial enhancement to poor performing "edge" users. In this paper we propose a novel framework aimed at fully exploiting the potential of such systems through dynamic radio resource clustering and user scheduling which maximize system utility. The dynamic clustering problem is modeled as a maximum weight clustering problem which is NP-hard, however, we show that by structuring the set of possible VBSs to be "2-decomposable" it can be efficiently computed. We also propose to optimize over a class of power allocation policies to radio resources, and thus VBSs, which allow dynamic user scheduling and flexible power allocations depending on instantaneous channel realizations. We use simulation to compare our approach with a state-of-the-art baseline which exploits dynamic frequency reuse and opportunistic user scheduling, but no clustering, and show edge users' throughput gains are as high as 80% without degrading the performance of others.**

I. Introduction

The fourth generation cellular systems based on OFDMA techniques currently being deployed achieve good coverage and high system throughput. Still there is interest in realizing even higher per user throughput, particularly for edge users.

Coordinated Multi-Point (CoMP) techniques provide a path to address this problem. They can be particularly advantageous when leveraging spatially distributed radio resources - we refer to these as Remote Radio Heads (RRHs). By clustering neighboring RRHs into Virtual Base Stations (VBSs) and encouraging cooperation among RRHs to reduce or eliminate mutual interference, edge users which were traditionally poorly served can become "central" to one or more VBSs, i.e., are well situated relative to their serving VBSs. Ideally, if there is enough freedom in choosing VBSs each user could be "central" - we refer to this as a "no-edge" wireless network.

Given the benefits of CoMP and assuming no computational constraints, one could in principle coordinate across all RRHs leveraging spatial diversity (in antennas) and removing all interference. Unfortunately, previous work (see e.g., [1]) has shown that coordination across the entire system does not bring as much benefit as expected because of measurement and signaling overheads. In practice cooperation is only possible, or desirable, within VBSs of limited size.

If we cluster RRHs into static sets of VBSs there may still be users at the edge of neighboring VBSs suffering from interference. To realize "no-edge" wireless networks it is thus necessary to use different VBSs in different sub-bands (or times) guaranteeing that each user can be "central" to a VBS. In the extreme case, on each time slot and sub-band, the system has the freedom to dynamically cluster RRHs into VBSs and to schedule users that are "central" to the VBSs. We call this "dynamic clustering". Advanced power control and user selection strategies are also essential to reap the benefits of CoMP to deliver good performance to every user. While this concept is attractive, there has not been much work on how to achieve good clustering and user scheduling in this setting.

A system employing CoMP and dynamic clustering could also leverage the current trend towards a Cloud-based Radio Access Network (C-RAN) compute infrastructure [2]. In addition to the flexibility to dynamically adjusting VBSs, a C-RAN based system could facilitate compute workload balancing among VBSs, is potentially more reliable and energy efficient by switching on/off RRHs and computational resources according to traffic loads.

In this paper, we focus on delay-tolerant best-effort traffic and propose a framework that enables dynamic clustering and advanced user scheduling (power control and user selection) algorithms to realize the promise of CoMP towards "wireless networks without edges."

Related Work. Let us briefly introduce CoMP techniques, see e.g., [3]. CoMP exploits cooperation among RRHs, or antennas co-located at a RRH, to provide better throughput to a single user or more aggressively to multiple users simultaneously. Dirty Paper Coding (DPC) [4] is known to be the optimal (capacity achieving) theoretical solution but is difficult to implement due to high complexity. In this paper, we consider a suboptimal CoMP technique called zero-forcing beamforming (ZFBF) that can be easily implemented. In ZFBF, a precoding vector is computed for each scheduled user by inverting the channel matrix of all scheduled users theoretically avoiding intra-VBS interference.

In OFDMA-based cellular systems with Single-Input and Single-Output (SISO) transmissions, cell edge users often see interference from adjacent cells, and/or users therein, sharing the same frequency band – inter cell interference. The dynamic Fractional Frequency Reuse (FFR) scheme described in [3], [5] offers perhaps the most efficient and flexible means to mitigate inter cell interference for such systems. A virtual scheduler and a power-control loop are proposed to adapt power across cells and sub-bands. However, Multi-User Multiple-Input Multiple-

978-1-4799-3360-0/14/$31.00 ©2014 IEEE

Output (MU-MIMO) techniques are generally employed to enable higher system capacity. In such systems users may also see interference associated with transmissions within the same cell – intra-cell interference. The work in [6], [7] extends the framework in [5] to MU-MIMO scenarios that adopt a fixed set of beams which limits the flexibility and is not considered to be efficient. Moreover, cooperation across RRHs is not considered in these works. This paper aims to address this challenge.

The work in [5] [6] [7] makes use of the result in [8] which generalizes the well-known proportional fair scheduler to an opportunistic gradient scheduling algorithm that aims to maximize a concave system utility function. This paper proves the asymptotic optimality of such an algorithm and also provides a theoretical foundation for our proposed framework.

In [9] the authors propose the use of different VBSs on different sub-bands so that each user can be central to its associated VBS, but the paper only lists a fixed set of VBSs for each sub-band to guarantee full coverage. They also do not consider systems which make clustering decisions dynamically and do not take into account opportunistic user scheduling.

Additional motivating work includes [10] which focuses on a single cell with multiple antennas that supports ZFBF. They propose an efficient semi-orthogonal user selection algorithm for simultaneous transmission and prove that such a framework achieves the same asymptotic sum rate as DPC when the number of users is large. However, it does not distinguish edge and central users and does not extend to larger systems.

The work in [11] aims to jointly design the base station clustering and the linear beamformers for all base stations to reduce coordination overhead of CoMP techniques. However, it does not consider user selection and power control.

Our Contributions. To our knowledge, this is the first work to propose a framework leveraging CoMP, via dynamic clustering and opportunistic user scheduling aimed at maximizing the overall system utility. To address the large number of degrees of freedom associated with solving highly-coupled problems of RRH clustering, user selection and power control, along with transmission precoding, and meet real-time computational constraints, we propose a decomposition of this complex problem, along with new structural properties that lead to efficient computation.

We provide an efficient solution to finding "optimal" clusters. In particular dynamic clustering is reduced to a Maximum Weight Clustering (MWC) problem which is proven to be NP-hard. However, in the wireless context where cooperation would happen amongst RRHs that are close by, we propose to structure the set of possible VBSs to satisfy "2-decomposability" property under which dynamic clustering in each sub-band can be solved with complexity $O(|R|^{1.5})$ where $|R|$ is the number of RRHs in the system. In our approach, dynamic clustering computation is centralized while the user scheduling for single VBSs can be decentralized and solved at possibly distributed compute resources associated with VBSs.

To exploit opportunistic gains associated with user selection and power control without increasing computational complexity, we propose a "flexible" power allocation policy which adapts power levels across radio resources and sub-bands. With these in hand, our scheduler can choose to assign power levels opportunistically to scheduled users. This enables the system to achieve good FFR patterns and at the same time allows the scheduler more flexible use of power to serve users and thus more freedom to achieve opportunistic benefits.

Finally, we provide an initial performance evaluation based on simulations. By comparing with an aggressive benchmark that does not exploit dynamic clustering but does perform dynamic soft FFR (see e.g. [5]) and opportunistic user scheduling, we show that dynamic clustering improves the rate of edge users by 80.4% without degrading the performance of others. Moreover and interestingly, we show how dynamic clustering is prone to move "edges" to other locations.

Paper Organization. The paper is organized as follows: Section II introduces our system model, the problem we aim to solve and the conceptual overview of our approach. Section III discusses our problem decomposition and structural constraints to ensure efficient solution of dynamic clustering and user scheduling while Section IV addresses the adaptation of the power allocation policy. Simulation results are exhibited in Section V and Section VI points to future work.

II. System Model and Problem Formulation

In this paper, a vector $\mathbf{x}$ is by default a row vector. We use $\mathbf{x}^*$ to denote the conjugate, $\mathbf{x^T}$ the transpose and $\mathbf{x^H}$ the conjugate transpose of vector $\mathbf{x}$.

We consider the downlink of a multi-user multi-RRH OFDMA-based cellular network. The operational frequency band is divided into equal sub-bands of size WHz; J denotes the set of sub-bands. σ^2 is the noise power in a sub-band. The system operates in discrete time, over time-slots $t = 0, 1, \cdots$.

U and R represent finite sets of users and RRHs, respectively. We let $|S|$ denote the cardinality of the set S. The RRHs are indexed from 1 to $|R|$. At time t, each $r \in R$ is associated with a set of users $U_{\{r\}}(t) \subset U$. An example association policy[1] would be $u \in U_{\{r\}}(t)$ iff r is the closest RRH to u at time t. In each sub-band an RRH $r \in R$ could work by itself (in which case $\{r\}$ is a singleton VBS) or cooperate with neighboring RRHs to form a VBS. There may be constraints on forming VBSs, for example, the maximum number of RRHs in a VBS or the maximum distance between RRHs in a VBS. We denote by $\mathcal{V}$ the collection of all allowed VBSs. A VBS $V \in \mathcal{V}$ is a set of RRHs that can coordinate their transmissions. (Without loss of generality, we assume that all antennas of an RRH belong to the same VBS at a certain time in a given sub-band.) For example, let $V = \{r_1, r_2\} \in \mathcal{V}$ denote a "VBS V containing RRH r_1 and r_2" and $U_V(t) = \bigcup_{r \in V} U_{\{r\}}(t)$ denote the "set of users that could be served by VBS V at time t".

We let $\mathcal{P}_{t,j}$ denote a partition of R based on sets in $\mathcal{V}$ for use at time t in sub-band j, in other words, for any $V_1, V_2 \in \mathcal{P}_{t,j}$,

$$\text{if } V_1 \neq V_2 \text{ then } V_1 \cap V_2 = \emptyset,$$

[1]Other association policies could be considered, or in fact no such policy needs to be specified. Yet this simplifies description of the setting for now.

and

$$\bigcup_{V \in \mathcal{P}_{t,j}} V = R.$$

The set of all possible partitions of R induced by $\mathcal{V}$ is denoted $\mathcal{P}(R, \mathcal{V})$. Note that each user can be associated with more than one VBS but it can only be served by one at a given time per sub-band.

Suppose each RRH is equipped with n_t transmit antennas and each user has n_r receive antennas. In this paper we assume $n_r = 1$. ZFBF is employed to enable simultaneous transmissions to multiple users. In particular, k ($< n_t$) users per RRH can be served at the same time, thus, up to $k \cdot |V|$ users can be served simultaneously by VBS V in a given sub-band.

For simplicity we assume flat and fast fading, i.e., the same fading is experienced in all sub-bands but can vary across slots - the framework of this paper can be extended to frequency-selective fading scenarios. Let $\mathbf{h}_u^{\{r\}}(t) = (h_u^{r,1}(t), \ldots, h_u^{r,n_t}(t))$ represent the complex channel vector from RRH r to user u at time t and $\mathbf{h}_u^R(t) = (\mathbf{h}_u^{\{1\}}(t), \ldots, \mathbf{h}_u^{\{|R|\}}(t))$ to be the channels from all RRHs to u. We also define $\mathbf{h}_u^V(t) = (\mathbf{h}_u^{\{r\}}(t) | r \in V)$ to be the channel vector[2] from VBS V to user u at time t. Note that $\mathbf{h}_u^V(t)$ is the signal channel if $u \in U_V(t)$ and an interference "channel" otherwise.

We focus on best-effort traffic, so each user u has an associated concave, strictly increasing utility function $U_u(\bar{X}_u)$ of its long-term time average rate $\bar{X}_u$. The choice of utility functions U_u may take into account fairness, quality of service and priorities among different users. Additional requirements such as meeting a minimum throughput per user could also be included in our setting.

Our goal is to find a dynamic clustering and user scheduling strategy that maximizes the system utility given by

$$\mathcal{U} = \sum_{u \in U} U_u(\bar{X}_u)$$

and more generally tracks changes in the system while attempting to meet this goal.

A. Gradient Scheduler

Inspired by proportional-fair scheduler, [5], [8] introduced a gradient algorithm to tackle general utility maximization problems such as the above. For our setting, let d_j denote a clustering and user scheduling decision in sub-band j and D denote a finite set[3] of possible decisions. Also let $\{\bar{x}_u(t) | u \in U\}$ denote the long term average rate estimates up to time t and $R_{uj}(d_j, t)$ denote the rate user u would achieve at time t in sub-band j under decision d_j. The algorithm and the main theoretical result developed in [8] as applied to our problem is stated below.

[2]For consistency, the complex numbers in $\mathbf{h}_u^V(t)$ have the same orders as they appear in $\mathbf{h}_u^R(t)$.

[3]Although power is a continuous value, we can consider power as a discrete variable that has a large number of possible values.

***Theorem** 1:* Suppose the channels from the RRHs to users have stationary distributions and $U_u(\cdot)$ are concave increasing utility functions. In the convex set (because of time sharing) of all feasible long-term achieved rate vectors, let $\bar{\mathbf{x}}^{\text{opt}} = (\bar{x}_1^{\text{opt}}, \ldots, \bar{x}_{|U|}^{\text{opt}})$ be the vector maximizing system utility $\mathcal{U}$.

If at each time t and in sub-band j, the system chooses a decision d_j^{opt} such that

$$d_j^{\text{opt}} \in \arg\max_{d_j \in D} \sum_{u \in U} \left. \frac{\partial U_u}{\partial \bar{X}_u} \right|_{\bar{x}_u(t)} R_{uj}(d_j, t),$$

and for all u the average rate estimate $\bar{x}_u(t)$ is updated as:

$$\bar{x}_u(t+1) = (1-\beta)\bar{x}_u(t) + \beta J R_{uj}(d_j^{\text{opt}}, t),$$

with arbitrary initial value $\bar{x}_u(0)$. Then, as $\beta \to 0$, both the estimate rate $\lim_{t \to \infty} \bar{x}_u(t)$ and the long-term achieved average rate $\bar{X}_u$ converge to $\bar{x}_u^{\text{opt}}$ for every user u.

Underlying each decision d at each time t and sub-band j, we need to:

- Select a partition $\mathcal{P}_{t,j} \in \mathcal{P}(R, \mathcal{V})$ corresponding to a collection of VBSs.
- Select up to $k \cdot |V|$ users for each VBS V in $\mathcal{P}_{t,j}$.
- Assign power to each scheduled user.
- Determine precoding vectors for all scheduled users associated with a VBS.

Although we listed these four items sequentially, they correspond to coupled decisions and the challenge is to find computationally efficient approach that can be used in practice.

B. Parameterizing Power Allocation Policy

To decouple the dynamic clustering and user scheduling decisions across sub-bands and VBSs we will fix a power allocation policy but adapt its parameters.

***Definition** 1:* A **feasible power allocation policy** is a pair $(\mathbf{P}, \mathbf{\Phi})$. The RRH sub-band power allocation matrix $\mathbf{P}$ is

$$\mathbf{P} = (p_{j,r} | j \in J, r \in R),$$

where $p_{j,r}$ denotes the power allocated to RRH r on sub-band j and must satisfy $\sum_{j \in J} p_{j,r} \leq p$ where p is the power constraint per RRH. The power available to a VBS V on sub-band j is then given by $p_j^V = \sum_{r \in V} p_{j,r}$. Recall that V may serve up to $k \cdot |V|$ users simultaneously, for which it will need to assign transmit power levels. The VBS sub-band power level allocation matrix $\mathbf{\Phi}$ is given by

$$\mathbf{\Phi} = (\phi_{j,l}^V | j \in J, V \in \mathcal{V}, l = 1, \ldots, k \cdot |V|),$$

where $\phi_{j,l}^V$ is the ratio of the lth highest power level for VBS V in sub-band j to the total available power p_j^V and must satisfy $\sum_{l=1}^{k \cdot |V|} \phi_{j,l}^V \leq 1$ and $\phi_{j,l_1}^V \geq \phi_{j,l_2}^V$ for $l_1 < l_2$.

Under such a policy the scheduler can flexibly decide which users to serve and what power levels they will use allowing it to opportunistically exploit channel variations.

It may be desirable to turn off an RRH if there are few or no users associated with it. In this case the power allocation policy could set the associated allocation for that RRH to zero.

C. Conceptual Overview of Our Approach

Our approach contains two main "components" as shown in Fig.1.

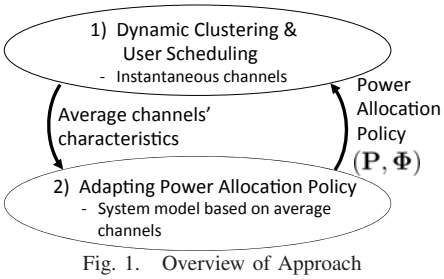


Fig. 1. Overview of Approach

The first component takes as inputs a power allocation policy for the system and performs dynamic clustering and user scheduling. To exploit channel variations and achieve opportunistic gains, it makes use of the instantaneous channels that are fed back from users to the RRHs.

The second component adapts the power allocation policy across radio resources and sub-bands. To do so, it approximates the gradients of system utility to the power allocation parameters and performs a gradient ascent algorithm.

III. Dynamic Clustering and User Scheduling

In this section, we assume we are given a power allocation policy $(\mathbf{P}, \mathbf{\Phi})$, the current long term average rate estimates $(\bar{x}_u(t)|u \in U)$ for users up to time t and instantaneous channel measurements $(\mathbf{h}_u^R(t)|u \in U)$. Our goal is to make a clustering and user scheduling decision for each sub-band j. To simplify notation, we will suppress the time index.

***Definition** 2:* A **clustering and user scheduling decision** d_j for sub-band j consists of a partition $\mathcal{P}(d_j) \in \mathcal{P}(R, \mathcal{V})$ for sub-band j and a user scheduling decision $\mathbf{s}_V(d_j)$ for each VBS $V \in \mathcal{P}(d_j)$. A user scheduling decision $\mathbf{s}_V$ for VBS V is a power allocation vector $\mathbf{s}_V = (l_1, l_2, \ldots, l_{|U|}) \in \{0, 1, \ldots, k \cdot |V|\}^{|U|}$ where

$$l_u = \begin{cases} l, \text{if } u \in U_V \text{ and } u \text{ is assigned power level } l \text{ in } \mathbf{s}_V, \\ 0, \text{otherwise.} \end{cases}$$

We require *no* two users can be assigned the same power *level*, but they may be allocated the same power, if two levels are assigned the same power. Thus, $\mathbf{s}_V$ has at most $k \cdot |V|$ non-zero elements.

Let D denote the finite set of possible clustering and user scheduling decisions and $D(V)$ denote the finite set of user scheduling decisions for VBS V. Following the result in Theorem 1, our goal is to find a decision satisfying

$$d_j^{\text{opt}} \in \arg\max_{d_j \in D} \sum_{u \in U} \left. \frac{\partial U_u}{\partial \bar{X}_u} \right|_{\bar{x}_u} R_{uj}(d_j). \tag{1}$$

A key issue here is to compute the achieved rate $R_{uj}(d_j)$ for user $u \in V$ under a given decision d_j. In particular, given $\mathbf{s}_{V'}(d_j)$ for VBS V' and using ZFBF, we can compute the precoding vector $\mathbf{w}_{u'}(\mathbf{s}_{V'}(d_j))$ (see e.g., [10]) for each scheduled user u' in V' using $\mathbf{0}$ for u' that are not scheduled. For user u along with $u' \in V'$ (where it is possible that $u = u'$), we define

$$g_u^{V',u'}(\mathbf{s}_{V'}(d_j)) = ||\mathbf{h}_u^{V'} \mathbf{w}_{u'}(\mathbf{s}_{V'}(d_j))^{\mathbf{T}}||^2$$

to be the effective gain from VBS V' to user u under the precoding vector of user u' associated with decision $\mathbf{s}_{V'}(d_j)$. Let $p_{uj}(\mathbf{s}_V(d_j))$ denote the power assigned to user u in sub-band j under decision $\mathbf{s}_V(d_j)$. Thus, $p_{uj}(\mathbf{s}_V(d_j))$ equals to $p_j^V \phi_{j,l}^V$ if l is the power level assigned to user u and equals to 0 if u is not scheduled under decision $\mathbf{s}_V(d_j)$.

Since multiple precoding vectors are used simultaneously, there may be some intra-VBS interference $I_{u,j}^{\text{intra}}$ from users in the same VBS and inter-VBS interference $I_{u,j}^{\text{inter}}$ from other VBSs. However, $I_{u,j}^{\text{intra}}$ should close to 0 if channel measurements are accurate and perfect ZFBF is used. So we have

$$R_{uj}(d_j) = W \log_2(1 + \frac{g_u^{V,u}(\mathbf{s}_V(d_j)) p_{uj}(\mathbf{s}_V(d_j))}{\sigma^2 + I_{u,j}^{\text{intra}} + I_{u,j}^{\text{inter}}}), \tag{2}$$

where

$$I_{u,j}^{\text{intra}} = \sum_{u' \in U_V, u' \neq u} g_u^{V,u'}(\mathbf{s}_V(d_j)) p_{u'j}(\mathbf{s}_V(d_j)),$$

$$I_{u,j}^{\text{inter}} = \sum_{V' \in \mathcal{P}(d_j), V' \neq V} \sum_{u'' \in U_{V'}} g_u^{V',u''}(\mathbf{s}_{V'}(d_j)) p_{u''j}(\mathbf{s}_{V'}(d_j)).$$

Problem (1) is difficult to solve because the number of decisions $|D|$ is large and the user scheduling decisions in different VBSs are coupled. Instead, we will explore a decomposition of this problem.

A. Decoupling Dynamic Clustering and User Scheduling

Note that the achieved rate $R_{uj}(d_j)$ for user $u \in V$ not only depends on $\mathbf{s}_V(d_j)$ but also depends on the clustering decision $\mathcal{P}(d_j)$ and the user scheduling decisions in other VBSs through $I_{u,j}^{\text{inter}}$. Our approach is to find an approximation for $I_{u,j}^{\text{inter}}$ and thus for $R_{uj}(d_j)$ such that problem (1) can be decomposed into sub-problems which in turn can be efficiently solved. The optimal decision $\tilde{d}_j^{\text{opt}}$ determined under our approximation should still generate a large value for (1).

Suppose each RRH $r \notin V$ transmits independently, i.e., without mutual cooperation, then an achievable upper bound for $I_{u,j}^{\text{inter}}$ is given by

$$\tilde{I}_{u,j}^{\text{inter}} \triangleq \sum_{r \notin V} ||\mathbf{h}_u^{\{r\}}||^2 p_{j,r}. \tag{3}$$

We argue that $\tilde{I}_{u,j}^{\text{inter}}$ is a good estimate, because our algorithm is geared at choosing appropriate clustering and user scheduling decisions to make $I_{u,j}^{\text{inter}}$ small in the first place which means a small error in $I_{u,j}^{\text{inter}}$ should not lead to a large bias in $R_{uj}(d_j)$.

By replacing $I_{u,j}^{\text{inter}}$ with $\tilde{I}_{u,j}^{\text{inter}}$, we get the approximation for instantaneous achieved rate for user u served by VBS V that depends only on $\mathbf{s}_V(d_j)$, i.e., decisions local to VBS V rather than all decisions across the network, i.e.,

$$\tilde{R}_{uj}(\mathbf{s}_V(d_j)) = W \log_2(1 + \frac{g_u^{V,u}(\mathbf{s}_V(d_j)) p_{uj}(\mathbf{s}_V(d_j))}{\sigma^2 + I_{u,j}^{\text{intra}} + \tilde{I}_{u,j}^{\text{inter}}}). \tag{4}$$

Now the problem to be solved can be approximated and written as

$$\tilde{d}_j^{\text{opt}} \in \arg\max_{d_j \in D} \sum_{V \in \mathcal{P}(d_j)} \sum_{u \in U_V} \frac{\partial U_u}{\partial \bar{X}_u}\bigg|_{\bar{x}_u} \tilde{R}_{uj}(\mathbf{s}_V(d_j)), \quad (5)$$

i.e., a sum over VBSs in partition $\mathcal{P}(d_j)$ of the marginal utilities under user scheduling decision $\mathbf{s}_V(d_j)$.

Let us define

$$w_j(V) = \max_{\mathbf{s}_V \in D(V)} \sum_{u \in U_V} \frac{\partial U_u}{\partial \bar{X}_u}\bigg|_{\bar{x}_u} \tilde{R}_{uj}(\mathbf{s}_V)$$

to be the weight of VBS V in sub-band j which captures the maximum marginal utility for VBS V, then solving (5) is equivalent to finding a weight for each VBS and finding a partition that gives maximum sum weight.

In Subsection III-B we tackle the problem of picking a partition assuming "weight" for each VBS is known and in Subsection III-C we propose a suboptimal scheme to compute approximations of the VBSs' weights.

B. Maximum Weight Clustering Problem

Picking an optimal partition is called a Maximum Weight Clustering Problem which we define as follows:

Maximum Weight Clustering (MWC) Problem: The Maximum Weight Clustering (MWC) Problem $(\mathcal{V}, \mathbf{w}_j)$, where $\mathcal{V}$ is the collection of allowed VBSs and $\mathbf{w}_j = (w_j(V)|V \in \mathcal{V})$ represents the associated non-negative weights for VBSs, is to determine a partition $\mathcal{P}_j^{\text{opt}}$ such that

$$\mathcal{P}_j^{\text{opt}} \in \arg\max_{\mathcal{P} \in \mathcal{P}(R,\mathcal{V})} \sum_{V \in \mathcal{P}} w_j(V).$$

Fig.2 exhibits an MWC Problem. The sets R, $\mathcal{V}$ and two possible partitions $\mathcal{P}1, \mathcal{P}2$ are shown while the weights are not given in the figure. For this simple example, $\mathcal{P}(R,\mathcal{V}) = \{\mathcal{P}1, \mathcal{P}2\}$. The goal is to find which partition has maximum weight.

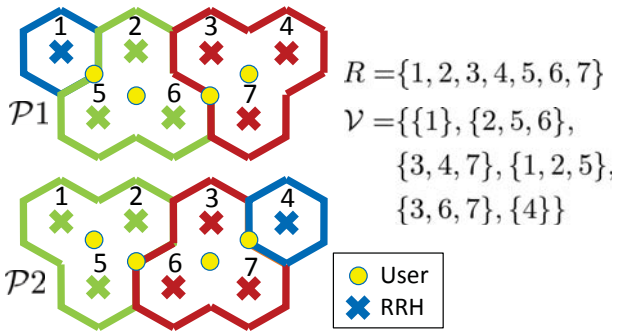


Fig. 2. An example MWC Problem.

***Theorem** 2:* MWC is an NP-hard problem.

Proof: MWC can be transformed from the problem of three-dimensional matching [12] which is NP-hard. ■

In general, a brute-force algorithm which goes over all possible partitions in $\mathcal{P}(R,\mathcal{V})$ would have a time complexity of $\mathrm{O}(2^{|\mathcal{V}|})$. Knuth's Algorithm X [13] provides a backtracking algorithm that finds all possible partitions. However, the time complexity is still exponential in $|\mathcal{V}|$ which is not sufficiently efficient to be applied each time slot on each sub-band.

To simplify the solution to dynamic clustering, we consider constraining the set of possible VBSs, $\mathcal{V}$, so that MWC can be solved efficiently. Some definitions are needed to understand our approach.

***Definition** 3:* Two RRHs r_1 and r_2 are **equivalent** under $\mathcal{V}$, written as $r_1 \overset{\mathcal{V}}{\sim} r_2$, iff $\forall\ V \in \mathcal{V}$, if $r_1 \in V$ and $|V| > 1$, then $r_2 \in V$.

In other words, for any possible partition in $\mathcal{P}(R,\mathcal{V})$, r_1 and r_2 are either singletons or in the same VBS. For the example in Fig.2, RRH 2 and RRH 5 are equivalent. In $\mathcal{P}1$, they belong to VBS $\{2,5,6\}$ and in $\mathcal{P}2$ they are in $\{1,2,5\}$. RRH 3 and RRH 7 are also equivalent.

***Collorary** 1:* The above equivalence relation is reflexive, symmetric and transitive, so it partitions R into a set of equivalence classes $\mathcal{E}_\mathcal{V} = \{E_1, E_2, \dots\}$ where

- $\mathcal{E}_\mathcal{V}$ is a partition of R,
- and $\forall i$ and $\forall r_1, r_2 \in E_i, r_1 \overset{\mathcal{V}}{\sim} r_2$.

Note that although $\mathcal{E}_\mathcal{V}$ is a partition of R it need not be in $\mathcal{P}(R,\mathcal{V})$. For the example in Fig.2, there are five equivalence classes, i.e. $E_1 = \{1\}, E_2 = \{2,5\}, E_3 = \{6\}, E_4 = \{3,7\}, E_5 = \{4\}$.

***Definition** 4:* A collection of VBSs $\mathcal{V}$ is said to be **2-decomposable** if for any $V \in \mathcal{V}$, there exists $E_i, E_j \in \mathcal{E}_\mathcal{V}$ such that $V \subseteq E_i \cup E_j$.

Theorem 3 below shows the benefit of satisfying this property. It is proven in the Appendix.

***Theorem** 3:* Given a MWC Problem $(\mathcal{V}, \mathbf{w}_j)$, if $\mathcal{V}$ is 2-decomposable, then it is equivalent to a maximum weight matching problem which is solvable in polynomial-time.

For the example in Fig.2, $\mathcal{V}$ is 2-decomposable.

Constructing the set of VBSs $\mathcal{V}$ that satisfy 2-decomposability still allows flexibility towards achieving a "no-edge" network and thus realizing the benefits of dynamic clustering. First, including all singleton VBSs to a 2-decomposable $\mathcal{V}$ would maintain this property but provide freedom to the scheduler. Second, this property allows VBSs including more than two RRHs, such as VBS $\{1,2,5\}$ in Fig.2. As mentioned in Section I, VBSs of interest generally have limited size making it more likely they satisfy 2-decomposability.

Taking Fig.2 as an example, by switching between $\mathcal{P}1$ and $\mathcal{P}2$, all four users can be scheduled and served by VBSs with respect to which they are central. By including more VBSs in $\mathcal{V}$ without losing 2-decomposability, for example, all singleton VBSs and VBS $\{2,5,3,7\}$, more areas can be guaranteed good coverage. Thus, "no-edge wireless network" can increasingly be achieved under this choice of $\mathcal{V}$ in this example.

The maximum weight matching problem is well studied in graph theory and can be solved efficiently using Edmond's matching algorithm (see e.g., [14]). The analysis in the Appendix shows the time complexity would be $\mathrm{O}(|R|^{1.5})$.

C. User Scheduling for Each VBS

We refer to the computation of the weight $w_j(V)$ for VBS V in sub-band j as a User Scheduling Problem. Recall that the weights for all $V \in \mathcal{V}$ serve as inputs to the MWC Problem considered in the previous subsection.

User Scheduling (US) Problem: The User Scheduling (US) Problem $(\mathbf{p}_j, \boldsymbol{\phi}_j^V, \bar{\mathbf{x}}, \mathbf{h}, U_V)$, where $\mathbf{p}_j = (p_{j,r} | r \in R)$, $\boldsymbol{\phi}_j^V = (\phi_{j,l}^V | l = 1, \ldots, k \cdot |V|)$, $\bar{\mathbf{x}} = (\bar{x}_u | u \in U_V)$ and $\mathbf{h} = (\mathbf{h}_u^R | u \in U_V)$, is to determine an optimal decision $\mathbf{s}_V^{\text{opt}}$ consisting of an optimal user group S^{opt} and power level assignment that gives a maximum weight

$$w_j(V) = \max_{\mathbf{s}_V \in D(V)} \sum_{u \in U_V} \left. \frac{\partial U_u}{\partial \bar{X}_u} \right|_{\bar{x}_u} \tilde{R}_{uj}(\mathbf{s}_V).$$

Solving this requires an exhaustive search over all possible user groups. The time complexity for this task would be $\sum_{i=1}^{k \cdot |V|} \binom{|U_V|}{i} C(i, |V|)$ where $C(i, |V|)$[4] captures the complexity of evaluating a user group of size i in a VBS of size $|V|$, i.e., determining precoding vectors and power level assignments. Thus, a suboptimal algorithm that is easy to implement is of interest.

We propose a two phase approach. In Phase 1 we use a greedy algorithm to select user group S_0 and in Phase 2 we find the optimal power level assignment for S_0. As a result, we compute the marginal utility for this decision which is only an approximation for $w_j(V)$.

Phase 1 - Suboptimal User Selection: Inspired by [10], we claim that in dense networks where there are many users in each VBS[5], the optimal user group S^{opt} consists of users that have nearly orthogonal channels, i.e., the precoding vectors for such users have similar direction as their channels, resulting in large effective gains and thus large rates. If users were perfectly orthogonal, the precoding vectors would be $\frac{\mathbf{h}_u^{V*}}{||\mathbf{h}_u^V||}$.

By using $\frac{\mathbf{h}_u^{V*}}{||\mathbf{h}_u^V||}$ as precoding vectors in (4) and assuming total power p_j^V equally split among $M = \min[k \cdot |V|, |U_V|]$ users, we get an estimate $R'_{uj}(S)$ for the rate of user $u \in S$ when user group S is selected,

$$R'_{uj}(S) = W \log_2(1 + \frac{||\mathbf{h}_u^V||^2 \frac{p_j^V}{M}}{\sigma^2 + I'_{\text{intra},j} + \tilde{I}_{u,j}^{\text{inter}}}),$$

where

$$I'_{\text{intra},j} = \sum_{u' \in S, u' \neq u} \text{EI}(\mathbf{h}_u^V, \mathbf{h}_{u'}^V) \frac{p_j^V}{M},$$

with

$$\text{EI}(\mathbf{h}_u^V, \mathbf{h}_{u'}^V) = ||\mathbf{h}_u^V \frac{\mathbf{h}_{u'}^{V\,\mathbf{H}}}{||\mathbf{h}_{u'}^V||}||^2, \tag{6}$$

representing the Effective Intra-VBS interference (EI) to u generated by u' and $\tilde{I}_{u,j}^{\text{inter}}$ given in (3). In this computation we may generate non-zero $I'_{\text{intra},j}$ by using $\frac{\mathbf{h}_u^{V*}}{||\mathbf{h}_u^V||}$ as precoders. However, once user group S_0 is determined, we can afford to compute actual precoding vectors which gives zero intra-VBS interference in the power assignment process.

Further, we define

$$w'_j(V, S) = \sum_{u \in S} \left. \frac{\partial U_u}{\partial \bar{X}_u} \right|_{\bar{x}_u} R'_{uj}(S),$$

to be the estimate of marginal utility if user group S is selected and power is equally split among M users.

The suboptimal user group S_0 is constructed as below: We start with an empty S_0. At each step we pick a user u^{opt} such that

$$u^{\text{opt}} \in \arg\max_{u \in U_V \setminus S_0} w'_j(V, S_0 \cup \{u\})$$

and let $S_0 \leftarrow S_0 \cup \{u\}$. The algorithm terminates when $|S_0| = M$.

Phase 2 - Power Level Assignment: In Phase 2, the user selection S_0 is determined and we can compute the precoding vector for each $u \in S_0$. Since $I_{u,j}^{\text{intra}}$ is 0, $\tilde{R}_{uj}(\mathbf{s}_V)$ in (4) depends on $\mathbf{s}_V$ only through the power level assigned to user u under scheduling decision $\mathbf{s}_V$. In this phase we use $\tilde{R}_{ujl}$ to denote $\tilde{R}_{uj}(\mathbf{s}_V)$ for $\mathbf{s}_V$ that assigns power level l to user u.

We construct a bipartite graph [15] $G = (S_0, Y, L)$ where S_0 is the suboptimal user group, Y is the set of power levels and there is an edge in L connecting each user u and each power level l with associated weight $\left. \frac{\partial U_u}{\partial \bar{X}_u} \right|_{\bar{x}_u} \tilde{R}_{ujl}$. The power assignment problem in Phase 2 now becomes that of finding the maximum weight matching (definition can be found in Appendix) in the bipartite graph G which can be efficiently solved using Hungarian algorithm (see e.g., [16]). Finally we compute the marginal utility of this suboptimal decision and pass it to MWC.

Now we analyze the complexity of our approach. In Phase 1, there are at most $k \cdot |V|$ iterations and in each iteration it needs $O(|U_V|)$ operations. So time complexity is $O(k \cdot |V| \cdot |U_V|)$. In Phase 2, the computation for edge weights takes $O(k^2 \cdot |V|^2)$ and solving maximum weight matching takes $O(k^4 \cdot |V|^4)$. Thus, time complexity for user scheduling for VBS V in each sub-band is $O(k \cdot |V| \cdot |U_V| + k^4 \cdot |V|^4)$.

IV. Adaptation of Power Allocation Policy

Our goal in this section is to adapt the power allocation policy $(\mathbf{P}, \boldsymbol{\Phi})$ so as to increase system utility. Doing so requires computing the gradients of system utility to the power allocation policy's parameters. This is hard because utility and gradients depend on stochastic channel variations. To address this we introduce a virtual system based on average channel measurements and estimate the gradients for the virtual system. This is still difficult because power affects system utility implicitly through average rates which in turn depend on dynamic clustering, opportunistic user selection and power assignment. Let $D_{j,r}$ and $D_{j,l}^V$ be the gradient estimates of virtual system utility to $p_{j,r}$ and $\phi_{j,l}^V$, respectively. Given a clustering and user scheduling decision, it is easy

[4] A reasonable estimate would be that computing precoding vectors takes $O((i \cdot n_t \cdot |V|)^3)$ and power assignment takes $O(i^4)$.

[5] Even if there are only a few users in the network, our algorithm attempts to pick as many (possibly suboptimal) orthogonal users as possible.

to express achievable rates as a function of power allocation parameters and thus easy to compute gradients under this decision. However, the virtual system utility is maximized by taking each decision with some fraction of time and thus estimating $D_{j,r}$ and $D^V_{j,l}$ requires averaging the gradients for each decision by the time fraction it will be taken.

The work in [5] provides a solution to this problem. The virtual system runs virtual scheduler which implicitly captures the time fractions associated with decisions and computes fraction-weighted gradients.

Specifically, the virtual scheduler runs a fixed number n_v virtual slots. On each virtual slot in sub-band j, it uses average channels to make a clustering and user scheduling decision $\hat{d}_j^{\text{opt}}$ and computes virtual rates $\hat{R}_{uj}(\hat{d}_j^{\text{opt}})$ according to (2). It then uses small averaging parameters β_1 and β_2 to update virtual average rate $\hat{x}_u$ for each u,

$$\hat{x}_u = (1-\beta_1)\hat{x}_u + \beta_1 J \hat{R}_{uj}(\hat{d}_j^{\text{opt}}),$$

and $D_{j,r}$ for each $r \in R$,

$$D_{j,r} = (1-\beta_2)D_{j,r} + \beta_2 \sum_{u\in U} \frac{\partial U_u}{\partial \bar{X}_u}\bigg|_{\hat{x}_u} \frac{\partial \hat{R}_{uj}(\hat{d}_j^{\text{opt}})}{\partial p_{j,r}},$$

and $D^V_{j,l}$ for each $V \in \mathcal{V}$ and $l = 1, \ldots, k\cdot|V|$,

$$D^V_{j,l} = (1-\beta_2)D^V_{j,l} + \beta_2 \sum_{u\in U} \frac{\partial U_u}{\partial \bar{X}_u}\bigg|_{\hat{x}_u} \frac{\partial \hat{R}_{uj}(\hat{d}_j^{\text{opt}})}{\partial \phi^V_{j,l}}.$$

However, unlike the SISO scenario in [5], in the ZFBF context we select users (in III-C) to exploit instantaneous orthogonality by using normalized channels as precoding vectors while in virtual scheduler the orthogonality between average channels may not be representative for real world. The challenge is to capture the degree of instantaneous orthogonality based on average channels. For example, users with nearly orthogonal average channels (e.g., in VBS $\{r_1, r_2\}$ a user close to r_1 and a user close to r_2) are very likely orthogonal while users with non-orthogonal average channels are less likely to remain so.

Let us divide $[0, \frac{\pi}{2}]$ into m angle ranges. Let $(U^1, U^2, V^{1,2})$ be a random triplet representing a "typical" pair of users concurrently scheduled on the same VBS $V^{1,2}$. We use $\mathbf{h}^{V^{1,2}}_{U^1}, \mathbf{h}^{V^{1,2}}_{U^2}$ and $\bar{\mathbf{h}}^{V^{1,2}}_{U^1}, \bar{\mathbf{h}}^{V^{1,2}}_{U^2}$ for their instantaneous and average channels, respectively. Let θ_{U^1,U^2} denote the angle of their average channels computed by $\theta_{U^1,U^2} = \arccos\left(\frac{||\bar{\mathbf{h}}^{V^{1,2}}_{U^1} \bar{\mathbf{h}}^{V^{1,2}\mathbf{H}}_{U^2}||}{||\bar{\mathbf{h}}^{V^{1,2}}_{U^1}||\cdot||\bar{\mathbf{h}}^{V^{1,2}}_{U^2}||}\right)$. We define

$$\gamma_i = \mathrm{E}\left[\frac{\mathrm{EI}(\mathbf{h}^{V^{1,2}}_{U^1}, \mathbf{h}^{V^{1,2}}_{U^2})}{\mathrm{EI}(\bar{\mathbf{h}}^{V^{1,2}}_{U^1}, \bar{\mathbf{h}}^{V^{1,2}}_{U^2})}\bigg| \theta_{U^1,U^2} \in \left[\frac{(i-1)\pi}{2m}, \frac{i\pi}{2m}\right)\right],$$

where the function EI is given in (6) and $\left[\frac{(i-1)\pi}{2m}, \frac{i\pi}{2m}\right)$ is the ith angle range. The vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$ represents a set of averaged quantities that the system keeps track of for use in the virtual system.

Given a pair of users u and u', we let $\gamma(\theta_{u,u'}) = \gamma_i$ where $\theta_{u,u'}$ falls into the ith angle range. Now in virtual scheduler to find $\hat{d}_j^{\text{opt}}$, when we select suboptimal user group S_0 in VBS V, we replace $\mathrm{EI}(\mathbf{h}^V_u, \mathbf{h}^V_{u'})$ with

$$\hat{\mathrm{EI}}(\bar{\mathbf{h}}^V_u, \bar{\mathbf{h}}^V_{u'}) = ||\bar{\mathbf{h}}^V_u \frac{\bar{\mathbf{h}}^{V\ \mathbf{H}}_{u'}}{||\bar{\mathbf{h}}^V_{u'}||}||^2 \gamma(\theta_{u,u'}).$$

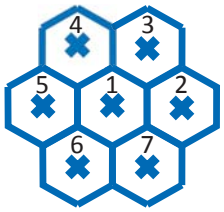
Adaptation of Power Allocation Policy

Given the gradient estimates, we update $(\mathbf{P}, \boldsymbol{\Phi})$ as below.

1) Update $p_{j,r}$: for each RRH, we increase the power allocated to the sub-band with largest positive gradient and decrease the power allocated to the sub-band with smallest negative gradient. (see [5] for details).
2) Update $\phi^V_{j,l}$: for each VBS V and sub-band j, we find the level l_{small} with smallest gradient and level l^{large} with largest gradient and exchange a small amount between them while maintaining the ranking of the power levels.

V. Simulations

We consider a grid of 7 RRHs as shown in Fig.3. It is easy to check that the set of VBSs $\mathcal{V}$ exhibited in Fig.3 is 2-decomposable and there are four equivalence classes, i.e. $E_1 = \{1\}, E_2 = \{2,3\}, E_3 = \{4,5\}, E_4 = \{6,7\}$. The propagation and transmission parameters are listed in Table I. For these parameters, the SNR for a user in between two neighboring RRHs turns out to be 10dB.



$R = \{1,2,3,4,5,6,7\}$
$\mathcal{V} = \{\{1\},\{2\},\{3\},\{4\},$
$\{5\},\{6\},\{7\},\{2,3\},$
$\{4,5\},\{6,7\},\{1,2,3\},$
$\{1,4,5\},\{1,6,7\}\}$

Fig. 3. RRH setup and set of VBSs used in simulations.

Parameters	Values
Inter-RRH Distance	1 Km
Path Loss Model	$L = 133.6 + 35\log_{10}(d)$
Bandwidth	1 MHz
Noise Power Density	-174 dBm/Hz
Tx Power per RRH	40 dBm
Tx Antenna Gain	15 dB
Rx Antenna Gain	76 dB
Rx Noise Figure	7 dB
Num of Antennas per RRH (n_t)	2
k	2

TABLE I
Propagation and Transmission Parameters in Simulations.

Suppose the operational bandwidth is divided into 3 sub-bands. We randomly and uniformly generate the locations of 100 users which are stationary. The full buffer model where all users always have traffic to send is used in all simulations. We assume flat and fast Rayleigh fading conditions and use the log utility function. An averaging parameter $\beta = 0.01$ is used in the actual scheduler. All simulations are run for 3000 time slots. The power allocation policy is adapted every

10 time slots and $n_v = 30$ virtual slots are performed in virtual scheduler. The two averaging parameters are chosen to be $\beta_1 = 0.005, \beta_2 = 0.01$.

We use PC to signify a system that adapts the power allocation policy to achieve good FFR patterns and DC to represent a system using dynamic clustering. Our approach is thus denoted PC/DC. To evaluate the benefits of these we consider three baseline systems.

PC/NO-DC: This system uses PC but no DC. This represents an aggressive state-of-the-art approach which supports ZFBF, explores dynamic soft FFR and opportunistic user scheduling, similar to e.g., [5].

NO-PC/DC: This system uses DC but no PC.

NO-PC/NO-DC: This system uses neither PC nor DC.

For NO-DC systems we assume no cooperation among RRHs and for NO-PC systems we set the power allocation policy such that power p for a RRH is split equally among sub-bands and then further equally split among power levels.

Note the channel realizations are the same in all systems for comparison.

Moving Edges: Suppose the 10 users experiencing the worst throughput under each system roughly represent the "location" of the edges. Fig.4 shows the edges for three simulated systems. The edge users in NO-PC/NO-DC can be seen to be at traditional edges lying at the borders between RRH cells. PC/NO-DC aims to mitigate interference by exploiting soft FFR patterns but does not change the locations of edge users by much. Under our PC/DC, most of the traditional edge users are well covered, thus dynamic clustering "moves" the "edges" to other locations. These new "edges" depend on the selected set of VBSs $\mathcal{V}$. By adding more VBSs to $\mathcal{V}$ without breaking 2-decomposability, we can provide good coverage to any location such that even "edge" users could get high throughput.

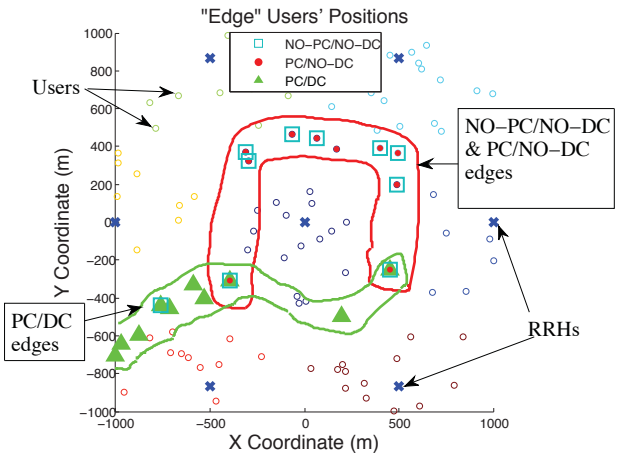


Fig. 4. "Edge" users' positions in three systems.

Throughput Improvement for Traditional Edge Users: To quantitatively evaluate the throughput gains achieved by dynamic clustering for traditional edge users, we focus on the worst 10 users under PC/NO-DC and compare those *same* users' rates under PC/DC. By ranking the users according to their average rates under PC/NO-DC, Fig.5 exhibits the average rates for all users in PC/NO-DC and PC/DC and zooms in the comparison for worst 10 users. We can see that dynamic clustering substantially improves the throughput of edge users in PC/NO-DC without degrading the performance of others. More precisely, the mean throughput of 10 edge users is increased by 80.4% and 81 users get better rates in PC/DC while there are only 3 users whose rates decrease by more than 5%. The Jain's fairness of average user rates also increases from 0.65 to 0.73.

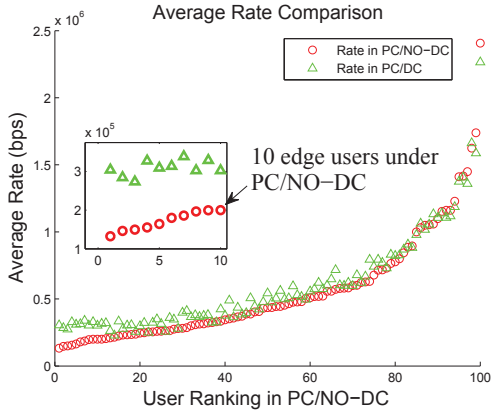


Fig. 5. Average rates for users in PC/NO-DC and PC/DC.

Benefits of DC vs. Benefits of PC: Both DC and PC aim at improving the edge throughput which in turn results in higher fairness. We use 10-percentile mean throughput (i.e. the mean throughput of worst 10% users) as a measure of the edge throughput for a system. To evaluate and compare the benefits of DC and PC separately, we compute the 10-percentile mean throughput and Jain's fairness of all users' average rates in four systems and compare the gains of adding DC or PC in Fig.6. It turns out that both techniques bring significant benefits to edge users but DC provides the higher gain for our simulation setup.

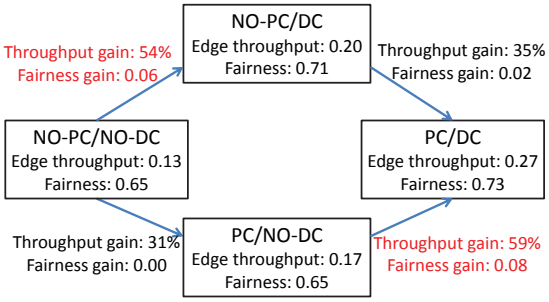


Fig. 6. 10-percentile mean throughput and Jain's fairness in four systems. The unit for edge throughput is 10^6bps.

We note that the above results are representative of different realizations of fading channels for uniformly distributed users. However, it will of interest to investigate the results for non-uniform user distributions.

VI. Future Work

There are several interesting topics for future work. Since we focused on downlink and delay-tolerant best-effort traffic, it would be of interest to integrate CoMP and network layer scheduling for uplink and other types of traffic. Evaluating the framework by taking into account user mobility and non-uniform user distribution are also goals for future work.

Appendix
Proof of Theorem 3

We begin by introducing maximum weight matching.

***Definition** 5:* In an edge-weighted graph $G = (N, L)$ where N is the set of vertices and L represents the set of edges, a matching is a collection of edges without common vertices. And a **Maximum Weight Matching (MWM)** is defined as the matching that has the maximum sum weights of the edges in the matching.

As shown in the left figure of Fig.7, the edges in dashed lines form a matching.

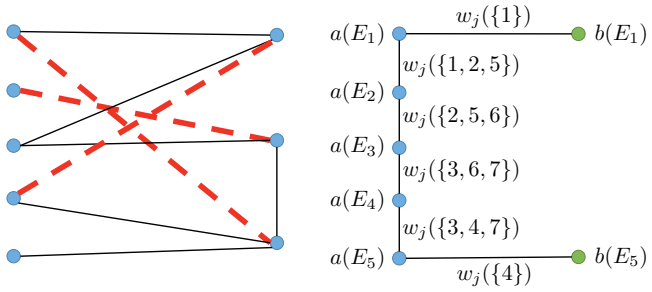


Fig. 7. A matching in a general graph and the converted graph for the MWC Problem in Fig.2.

Suppose we are given Maximum Weight Clustering (MWC) Problem $(\mathcal{V}, \mathbf{w}_j)$ where $\mathcal{V}$ is 2-decomposable. A set $E \in \mathcal{E}_\mathcal{V}$ is said to be singleton-decomposable if $\forall r \in E, \{r\} \in \mathcal{V}$, i.e., each RRH in E is a candidate VBS in $\mathcal{V}$. We define a new set of weights $\omega(E)$ for each $E \in \mathcal{E}_\mathcal{V}$ as below,

$$\omega(E) = \begin{cases} \text{If } E \in \mathcal{V} \text{ and } E \text{ is singleton-decomposable,} \\ \qquad \max[\sum_{r\in E} w_j(\{r\}), w_j(E)]. \\ \text{If } E \in \mathcal{V} \text{ and } E \text{ is not singleton-decomposable,} \\ \qquad w_j(E). \\ \text{If } E \notin \mathcal{V} \text{ and } E \text{ is singleton-decomposable,} \\ \qquad \sum_{r\in E} w_j(\{r\}). \\ \text{If } E \notin \mathcal{V} \text{ and } E \text{ is not singleton-decomposable,} \\ \qquad 0. \end{cases}$$

Next, we construct a graph $G = (N, L)$ for this MWC Problem as follows. The right figure in Fig.7 is the graph for the example problem in Fig.2.

- For each element $E \in \mathcal{E}_\mathcal{V}$ that has an positive weight $\omega(E)$, we add two vertices $a(E)$ and $b(E)$ to N.
- For each element $E \in \mathcal{E}_\mathcal{V}$ whose $\omega(E)$ is 0, we add only one vertex $a(E)$ to N.
- For each element $E \in \mathcal{E}_\mathcal{V}$ that has an positive weight $\omega(E)$, we put an edge between $a(E)$ and $b(E)$ and associate weight $\omega(E)$ to that edge. For example, the edge between $a(E_1)$ and $b(E_1)$ in Fig.7.
- For each VBS $V \in \mathcal{V}$, if V is the union of two equivalence classes, say $V = E_i \cup E_j$, we add an edge to connect $a(E_i)$ and $a(E_j)$ with weight $w_j(V)$. For example, the edge between $a(E_1)$ and $a(E_2)$ in Fig.7.

After constructing the graph, we claim that solving MWM in graph G gives us the answer for the MWC problem. It is not difficult to prove this claim and we omit it to save space.

Let m denote the number of vertices and n denote the number of edges in a graph, the time complexity for solving MWM is $O(n\sqrt{m})$. In our scenario, $m \leq 2\times|\mathcal{E}_\mathcal{V}| \leq 2\times|R| = O(|R|)$, $n \leq |\mathcal{E}_\mathcal{V}| + |\mathcal{V}| \leq |R| + |\mathcal{V}|$. In practical systems of interest, an RRH cooperates with neighboring RRHs rather than RRHs that are far away which means each RRH belongs to a small number of VBSs. Thus, $|\mathcal{V}| = O(|R|)$ which results in time complexity in our scenario becoming $O(|R|^{1.5})$.

Acknowledgement

This research was supported by Huawei Technologies Co. Ltd. The authors would like to thank Andrés García Saavedra, Alan Gatherer, Robert W. Heath Jr. and Namyoon Lee for their comments and feedback on this work.

References

[1] A. Lozano, R. W. H. Jr., and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Trans. Inf. Theory*, vol. PP, p. 1, March 2013.
[2] China Mobile, "C-RAN The Road Towards Green RAN," Oct 2011.
[3] E. Pateromichelakis, M. Shariat, A. ul Quddus, and R. Tafazolli, "On the evolution of multi-cell scheduling in 3gpp lte / lte-a," *IEEE Commun. Surv. Tutor.*, vol. 15, pp. 701–717, May 2013.
[4] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, pp. 439–441, May 1983.
[5] A. L. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse for best-effort traffic through distributed inter-cell coordination," in *Proc. INFOCOM'09*, April 2009, pp. 1287–1295.
[6] G. Wunder, M. Kasparick, A. Stolyar, and H. Viswanathan, "Self-organizing distributed inter-cell beam coordination in cellular networks with best effort traffic," *WiOpt 2010*, pp. 295–302.
[7] M. Kasparick and G. Wunder, "Autonomous distributed power control algorithms for interference mitigation in multi-antenna cellular networks," *European Wireless 2011, Vienna, Austria*, April 2011.
[8] A. L. Stolyar, "On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation," *Operations Research*, vol. 53, no. 1, pp. 12–25, January-February 2005.
[9] W. Mennerich and W. Zirwas, "User centric coordinated multi point transmission," in *VTC 2010-Fall*, September 2010, pp. 1–5.
[10] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, March 2006.
[11] M. Hong, R. Sun, H. Baligh, and Z. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 31, February 2013.
[12] R. M. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations*. New York: Springer, 1972, pp. 85–103.
[13] D. E. Knuth, "Dancing links," *Millenial Perspectives in Computer Science*, pp. 187–214, 2000.
[14] A. Schrijver, in *Combinatorial Optimization: Polyhedra and Efficiency*, ser. Algorithms and Combinatorics. Springer, 2003, vol. 24.
[15] A. S. Asratian, T. M. J. Denley, and R. Häggkvist, in *Bipartite Graphs and Their Appliations*. Cambridge University Press, 1998, vol. 131.
[16] J. Munkres, "Algorithms for the assignment and transportation problems," *J. Soc. Indust. Appl. Math.*, vol. 5, no. 1, pp. 32–38, 1957.