Adaptation, Coordination and Distributed Resource Allocation in Interference-Limited Wireless Networks

David Gesbert[†], Saad G. Kiani[†], Anders Gjendemsjø[*], Geir E. Øien[*]

[†]Mobile Communications Department, Eurecom Institute,

BP 193, F-06904 Sophia Antipolis, France, Email: {gesbert, kiani}@eurecom.fr

[*]Dept. of Electronics and Telecom. Norwegian Univ. of Science and Technology,

NO-7491 Trondheim, Norway, Email: {gjendems, oien}@iet.ntnu.no

Abstract

A sensible design of wireless networks involves striking a good balance between an aggressive reuse of the spectral resource throughout the network and managing the resulting co-channel interference. Traditionally this problem has been tackled using a "divide and conquer" approach. The latter consists in deploying the network with a static or semi-dynamic pattern of resource reutilization. The chosen reuse factor, while sacrificing a substantial amount of efficiency, brings the interference to a tolerable level. The resource can then be managed in each cell so as to optimize the per cell capacity using an advanced air interface design.

In this paper we focus our attention on the overall network capacity as a measure of system performance. We consider the problem of resource allocation and adaptive transmission in multicell scenarios. As a key instance, the problem of joint scheduling and power control simultaneously in multiple transmit-receive links, which employ capacity-achieving adaptive codes, is studied. In principle, the solution of such an optimization hinges on tough issues such as the computational complexity and the requirement for heavy receiver-to-transmitter feedback and, for cellular networks, cell-to-cell channel state information (CSI) signaling. We give asymptotic properties pertaining to rate-maximizing power control and scheduling in multicell networks. We then present some promising leads for substantial complexity and signaling reduction via the use of newly developed distributed and game theoretic techniques.

Index Terms

Interference Mitigation, Multicell Scheduling, Distributed Resource Allocation, Power Control, Link Adaptation, Fairness.

I. INTRODUCTION

The announced convergence between mobile and data access internet-based services, initiated in systems such as WiMax [1] and 3G-LTE [2], poses extraordinary challenges to the designers of future generation wireless networks who must cope with the scarcity of the spectral resource in areas with heavy user demand. It is widely admitted that, at the heart of this challenge, lies the ability to exploit the resource as efficiently as possible in all dimensions allowed by the multiple access scheme (e.g. time, frequencies, codes, power, beams, etc.). So far, the conventional approach for dealing with this problem, when deploying a network over a given geographical area necessitating many infrastructure sites or wireless nodes, has been a *divide and conquer* one, as outlined in the following:

Divide: First, network frequency (or, more generally, resource) planning is used to allow the fragmentation of the network area into smaller zones isolated from each other from a radio point of view. Within a cluster of neighboring links, the spectral resource is not reused at all (such as e.g. in GSM), or reused only partially (e.g. CDMA networks, where each cell limits the number of assigned codes to a fraction of the theoretical limit defined by the spreading factor). In ad-hoc networks, isolation of transmit-receive pairs from each other is also sought, via interference-avoidance multiple access control (MAC), typically by means of carrier sensing based protocols. The need for high efficiency figures however leads the system designer towards a planning featuring even more aggressive spectral reuse, for instance in the cellular case from a cluster size of 5 to 7 in early GSM deployments, down to close to 1 in today's available networks such as WiMax. Power control techniques and per-cell dynamic resource allocation (e.g. frequency hopping) methods help alleviate the problem of out-of-cell interference, but in practice aggressive resource reuse will still inevitably lead to an increased level of interference in the network, which undermines the link-level performance.

Conquer: In turn this loss (due to interference) of link efficiency for a given cell or for a local transmit-receive pair may be compensated, via a careful design of the radio air interface. The latter may exploit advanced processing such as efficient forward error correction (FEC) coding,

fast link adaptation protocols, multiple-antenna transceivers [3], and more recently channel aware scheduling techniques [4]. In the multiuser diversity approach, the scheduling protocol is designed towards a better utilization of the spectrum inside each cell by encouraging channel access for data-access users temporarily experiencing better propagation conditions, giving rise to the so-called multi-user diversity gain [5]. It is worth noting that this gain can be realized only if link adaptation techniques are available to take advantage of the improvement in channel conditions. Clearly multi-user diversity is gained at the expense of throughput fairness, which may be at least partially restored by modifying the scheduling criteria in one of several possible manners [6]. Interestingly, this idea of multi-user diversity, traditionally a single cell concept, is going to resurface in this article in a different form in the multicell context.

A. Voice-centric vs. data-centric models

To a large extent the divide-and-conquer approach outlined above is initially motivated by voice-centric considerations. Traditionally, multicell resource planning and power control are aimed at allowing the network users to operate under a common minimum carrier to interference level (C/I), that is compatible with the receiver's sensitivity or operating point[1] at the access points and the user terminals. Consequently, most power control algorithms are designed to reach an SINR target simultaneously for all interfering user terminals. This *SINR balancing* approach ensures a worst-case outage probability necessary for connection oriented voice calls, as was done in famous contributions such as [7]–[9].

The concept of a modem's operating point is becoming less relevant in modern networks designed for data-dominated traffic, as these typically feature adaptive coding and modulation protocols capable of adjusting the transmission rate to a wide range of channel conditions. Even if the number of coding rates remains limited in practice due to memory and complexity constraints, the strategy consisting of optimizing the spectral resource for a desired worst case interference level and then relying on advanced modem design alone for maximizing performance, is losing some relevance. This in turn shows the limitation of the divide and conquer approach when it comes to network wide optimization of performance. For best-effort data access (e.g. email, web browsing, multimedia messaging) the *sum network capacity*, defined as the sum of simultaneous

---

[1]The operating point is the level of SINR needed to operate on the link, below which the call may be dropped.

transmit-receive link capacities, appears as a more meaningful metric. However, additional constraints may be needed to include specific scenarios with QoS-driven traffic data (e.g. VoIP) into the resource optimization problem.

B. Coordinated multicell resource allocation: Challenges and leads

Instead of the traditional approach based on decoupling the multicell resource allocation from the optimization of the single cell capacity, one may naturally think that a *joint* optimization of resources in all cells *simultaneously* will give better system performance. When doing so, the per cell of optimization proposed previously involving e.g. code assignment, power control, multiple antenna beam design, and time/frequency channel-aware scheduling, is now expanded to take into account the dimension offered by the multiple cells of the network.

Evidently, such a joint multicell resource allocation offers an enormous number of degrees of freedom (governed by the number of cells, times the number of users, times the number of possible scheduling slots, codes, power levels etc.) that can be exploited to optimize the network performance at all times. As a key instance of such an optimization problem, we will be considering in later sections the problem of joint multi-user scheduling and power allocation in multiple cells for the purpose of maximizing the sum network capacity under an ideal link adaptation protocol.

The potential in coordinated resource allocation across cells also comes with several practical challenges. This includes among others the need for slot level synchronization for large network areas. However, this problem may be partly alleviated by *clustering* the optimization. Another severe problem is the need for the joint processing of traffic and channel quality parameters fed back by all network nodes to a central control unit (see Fig.1(a)), which necessitates significant computational power and huge signaling overhead. This makes global network coordination hard to realize in practice, especially in mobile settings where the control unit ought to cope with fast time-varying channels. Despite this important challenge, some recently published and promising methods have hinted at how some of the multicell coordination gains may be realized with limited complexity and/or limited centralized control, as discussed next.

C. Putting multicell coordination to practice

Following the recent literature, three leading and independent strategies may be identified in the effort toward making multicell coordination of resource more practical, though overall many interesting questions and challenges remain open. Some of these ideas are now briefly reviewed, while others are described in greater detail in the rest of this paper.

1) Structuring: One of the major difficulties associated with interference avoidance in packet access communications is the lack of predictability of interference coming from other transmit-receive links, due to burstiness of the traffic combined with the temporal channel variability. As an approach to counteract this effect, structure may be enforced on the resource planning grid to make interference more predictable. For instance, in the joint user scheduling and power allocation problem, a particular *power shaping* of the time frame can be exploited by allowing the AP to transmit with different powers in different portions of the frame, while users are allotted slots according to the amount of interference they can tolerate given their local channel conditions. This type of approach was pursued in e.g. [10], [11]. In an analogous strategy, power shaping over the cell sectors can be implemented by turning off sector beams according to a determined sequence, which permits users to measure the interference received and then tell their respective AP their preferred sub-frame for reception; this idea is referred to as *Time-Slot Reuse Partitioning* in [12]. In another approach, structure may be enforced by fixing the *order* in which time/frequency slots are being filled up with user packets. In the case of under-loaded systems, a predictable average portion of the slots remain unused (power-free) and the location of such slots on the multicell resource grid can be optimized to reduce interference for selected users [13]. The spatial position of users in the cell can also be used to coordinate intercell transmissions to avoid excessive interference [14]. Limited exchange of information between dominant interfering (neighboring) APs is yet another way of gaining knowledge about the worst-case interference, enabling the orthogonalization of these transmissions [15].

Such clever resource planning schemes are interesting as they offer additional flexibility in mitigating interference with very low complexity and little need for signaling. On the other hand, they are not fully exploiting the degrees of freedom provided by the joint multicell resource allocation problem, as the imposed structure tends to reduce the dimensions offered in the optimization.

2) Discretization: As certain quantities entering the resource allocation problem may be continuous, e.g. the transmit power levels, or the beamforming coefficients if multiple antennas are used, a potentially interesting tool in modifying the optimization problem consists of discretizing the optimization space so as to further reduce the number of potential solutions to search over, and also to reduce the feedback rate needed to communicate overhead data between network nodes. Discretization (via vector quantizing) of the optimal beamforming weights through the use of vector precoding has been proposed, but interestingly, mostly for the single cell scenario, and only for the purpose of feedback reduction (see e.g. [16]). In the case of beamforming weights, discretization can be applied posterior to beamforming weight computation. In the case of power control, discretization can be carried out prior to optimization, as a way to greatly simplify the power level search procedure. Remarkably, the discretization of power control, even to its extreme of binary on/off control, can be shown to yield quasi-optimal results in a number of cases [17], and as such constitutes a promising tool to making multicell coordination a reality. This is a central idea which is also developed in greater detail later in this article.

3) Greedy and iterative optimization: Due to the non-convexity of many of the multicell resource optimization problems, finding globally optimal solutions from standard techniques proves difficult, and an analytical formulation of the solution is often out of reach. In this case, heuristic approaches based on alternating optimization or greedy search may provide a good performance/complexity compromise. While greedy search techniques have been popularized over the last few years in the area of resource allocation in multiuser spatial division multiple access [18] and OFDMA scheduling [19], [20], their application to multicell resource allocation seems to have drawn attention only recently. Greedy multicell optimization operates by optimizing on a cell by cell basis, sequentially, just as individual users are optimized sequentially in the single cell scenario. At each cell visited, the resource is optimized based on local channel conditions and newly updated interference conditions originating from the other cells [21], [22]. Such techniques may also be applied in an iterative manner by revisiting a sequence of cells several times until capacity convergence is reached.

D. Distributed vs. centralized control

In most of the approaches above, including greedy optimization, the need may exist for centralized knowledge of all channel and interference state conditions for all nodes in the

network. In the case of the greedy approaches, the algorithm then only visits the cell *virtually*, and the actual computation takes place within the central control unit shown in Fig.1(a). Centralized channel state information for a many-cell many-user network involves immense signaling overhead and will not allow the extraction of diversity gains in fast-fading channel components. To circumvent this problem, the design of so-called *distributed* resource allocation techniques is crucial. Distributed optimization refers to the ability for each cell to manage its local resources (say e.g. rate and power control, user scheduling) based only on locally observable channel conditions such as the channel gain between the access point and a chosen user, and possibly locally measured noise and interference.

At first sight, joint multicell resource allocation does not lend itself easily to distributed optimization because of the strong coupling between the locally allocated resources and the interference created elsewhere in the network. Hence the maximization of the cell capacities taken individually will not in general result in the best overall network capacity, although we suggest later cases for which the outcomes for the per cell and multicell capacity optimization will differ little.

An interesting and recently explored path toward enforcing a distributed control of resource has been through the use of *game theoretic* concepts. Game theory, in its non-cooperative setting, pitches individual players in a battle, each seeking to maximize a utility function by selecting one of several available strategic actions. In the resource allocation framework, users can be terminals competing for access in a single cell, or interfering transmit-receive pairs of a multiple cell network or an adhoc network. The actions may be resource allocation strategies, and the utility may be capacity related. Non-cooperative game models allow transmit-receive pairs to maximize their capacity under reasonable guesses of what competing pairs might be doing [23]. In that respect, it naturally and beautifully lends itself to distributed optimization. The game theoretic framework is very well suited to network scenarios where infrastructure is sparse or completely absent, as in peer-to-peer and adhoc networks. In infrastructure-based networks like cellular, broadband access and to some extent WLAN networks, where a centralized operator retains control over the common resource, it remains to be seen if the purely non-cooperative model is overly pessimistic, as it may not be able to fully capture the gain that could be obtained from coordination. However, pricing-based game theoretic approaches have been proposed to alleviate this problem (see for instance [24]). These aspects are further detailed later in the

paper.

Finally, as an alternative to game theory techniques, this paper also investigates distributed forms of iterative multicell resource allocation. In such approaches, APs individually make a decision on their transmit power so as to optimize their contribution to the sum rate. At the core lies the idea that the interference behavior can be made more predictable by making the network larger or denser, and consequently the resource allocation problem in a given cell is made more dependent on the local channel conditions in that cell, thus facilitating distributed optimization.

E. Scope and organization

This paper reviews various interesting strategies for the joint multicell resource allocation problem. We consider transmitters with backlogged traffic, i.e. which always have data to send. We are taking a sum-rate point of view on the network performance, as opposed to considering packet delivery delays and other QoS issues. Issues related to queue-stabilizing routing and resource allocation protocols are therefore not addressed here but should rather come as possible extending direction for this study.

We formulate our ideas for a certain cellular network model, but we emphasize the strong links to the corresponding problem in adhoc networks. In Section II we review typical optimization models and first formulate the optimal, centralized solutions, focusing on the key problems of power control and user scheduling. We then give useful results on discretization of power control, including the case where power levels only admit two values $\{0, P_{\max}\}$, and show the central role that this type of solutions seems to be playing in our problem (Section VI). We refer to this strategy as *binary power control*.

Optimal (even binary) power control solutions necessitate by essence a centralized treatment. In Section VII we next turn to strategies allowing distributed optimization. We show that distributed strategies, which are optimal from a capacity scaling point of view, can be derived for multicell systems with an asymptotic number of users per cell. For a small to moderate number of users, we consider more practical alternatives, including game theoretic and iterative optimization techniques. We present some approaches using iterative optimization in the binary power control setting. We show how simple distributed algorithms can be derived from such techniques, particularly in the case of dense random networks. The performance of these and other existing schemes are compared. As in the single cell case, the multicell resource optimization

problem inevitably leads to a capacity vs. fairness debate. These issues are addressed briefly in Section VIII. We then conclude and discuss future areas of study in Section IX.

II. Network model and assumptions

We consider a wireless network featuring a number of transmitters and receivers. Among these, there are N active transmit-receive pairs, which are simultaneously selected for transmission by the scheduling protocol at any considered instant of time, while others remain silent. In this network the i-th transmitter, denoted T_i, (T_1, T_2,... are shown in Fig.2), sends a message which is intended to the i-th receiver R_i only. However R_i is being interfered by all $T_j, \forall j \neq i$ due to reuse of the spectral resource. This setup can be seen as an instance of the interference channel, the analysis of which is a famously difficult problem in information theory [25]. In practical terms, the situation depicted above can be that of a cellular network with reuse factor one (say e.g. the downlink with T_i being access points (AP) or base stations). It can also depict a snapshot of an adhoc network (see Fig.2(b)).

A. Signal Model

To facilitate exposition, we shall adopt a cellular terminology from here on. We thus consider N time-synchronized cells, and U_n users randomly distributed over each cell $n \in [1..N]$. In each cell, we consider an orthogonal multiple access scheme so that on any given *spectral resource slot* (where resource slots can be time or frequency slots in TDMA/FDMA, or code in orthogonal CDMA) a single user is supported[2]. Therefore, focus is on intercell interference rather than on intra-cell interference and the latter would come as a further extension of the study. On any given spectral resource slot, shared by all cells, let $u_n \in [1...U_n]$ be the index of the user that is granted access to the channel in cell n.

We denote the downlink channel pathloss gain and random complex fading coefficient between AP i and user u_n in cell n by $\gamma_{u_n,i}$ and $h_{u_n,i}$, respectively. We hereby focus on the downlink, but all ideas presented here carry over to the uplink as well, unless otherwise stated. We shall assume that the coherence time of the channel is long enough so that the receiver can estimate

---

[2]If the access point is equipped with multiple antennas we can generalize this model to multiple users per slot or code, giving rise to so-called spatial division multiple access (SDMA).

the gain (in each resource slot) and send this information to a local or global resource allocation unit via a feedback channel if necessary. The received signal Y_{u_n} at the user in a given resource slot is then given by

$$Y_{u_n} = \gamma_{u_n,n} h_{u_n,n} X_{u_n} + \sum_{i \neq n}^{N} \gamma_{u_n,i} h_{u_n,i} X_{u_i} + Z_{u_n},$$

where X_{u_n} is the message-carrying signal from the serving AP, $\sum_{i \neq n}^{N} \gamma_{u_n,i} h_{u_n,i} X_{u_i}$ is the sum of interfering signals from other cells, and Z_{u_n} is additive noise or additional interference. Z_{u_n} is modeled for convenience as complex AWGN, with power $\mathbb{E}|Z_{u_n}|^2 = \sigma^2$.

III. THE MULTICELL RESOURCE ALLOCATION PROBLEM

We now turn to the core problem of resource allocation. Given the orthogonal multiple access protocol described above, the resource allocation problem consists in *power allocation* and *user scheduling* subproblems. Importantly, we focus here on *capacity maximizing* resource allocation policies, rather than *fairness-oriented* ones. Fairness issues are very briefly touched upon later in the paper. In this setting, the optimization of resources in the various resource slots decouples, and we may consider the power allocation and user scheduling which maximize capacity in a particular slot, independently of others.

A peak transmit power constraint $P_{\max}$ is imposed at each AP and to simplify exposition, we shall assume that it is identical for all transmitters. In order to facilitate the problem formulation of the joint power allocation and scheduling problem, we state the following definitions:

Definition 1: A **scheduling vector U** for a given resource slot contains the set of users simultaneously scheduled across all cells:

$$\boldsymbol{U} = [u_1\, u_2\, \cdots\, u_n\, \cdots\, u_N]$$

where $[\boldsymbol{U}]_n = u_n$. Noting that $1 \leq u_n \leq U_n$, the feasible set of scheduling vectors is given by $\Upsilon = \{\boldsymbol{U} \mid 1 \leq u_n \leq U_n\ \forall\ n = 1, \ldots, N\}$.

Definition 2: A **transmit power vector P** for a given resource slot contains the transmit power values used by each AP to communicate with its respective user:

$$\boldsymbol{P} = [P_{u_1}\, P_{u_2}\, \cdots\, P_{u_n}\, \cdots\, P_{u_N}]$$

where $[\boldsymbol{P}]_n = P_{u_n} = \mathbb{E}|X_{u_n}|^2$. Due to the peak power constraint $0 \leq P_{u_n} \leq P_{\max}$, the feasible set of transmit power vectors is given by $\Omega = \{\boldsymbol{P} \mid 0 \leq P_{u_n} \leq P_{\max}\ \forall\ n = 1, \ldots, N\}$.

A. Utility-optimal resource allocation

The merit associated with a particular choice of a scheduling vector and power allocation vector is measured via the help of a *network utility function*, denoted by $F(\boldsymbol{U}, \boldsymbol{P}) : \Upsilon \times \Omega \longrightarrow \mathbf{R}_+$. Because N pairs are served in parallel, the total utility typically decomposes into the sum $F(\boldsymbol{U}, \boldsymbol{P}) = \sum_n f_n(\boldsymbol{U}, \boldsymbol{P})$, where $f_n(\cdot)$ is the utility enjoyed by the network in cell n. A logical choice for the utility in the above interference limited system is to pick a function of the signal to noise and interference ratio (SINR), $f_n(\boldsymbol{U}, \boldsymbol{P}) = f(\Gamma([\boldsymbol{U}]_n, \boldsymbol{P}))$, where $\Gamma([\boldsymbol{U}]_n, \boldsymbol{P})$ refers to the SINR experienced by the receiver u_n in cell n as a result of power allocation in all cells. This SINR is given by

$$\Gamma([\boldsymbol{U}]_n, \boldsymbol{P}) = \frac{G_{u_n,n} P_{u_n}}{\sigma^2 + \displaystyle\sum_{i \neq n}^{N} G_{u_n,i} P_{u_i}}, \tag{1}$$

where $G_{u_n,i} = \gamma_{u_n,i}^2 |h_{u_n,i}|^2$ is the channel power gain from cell i to receiver u_n.

1) Capacity optimal resource allocation: In connection-oriented communication, a typical utility function is a step function of the SINR with the SINR threshold dictated by the receiver's sensitivity. In data-centric applications however, where rate adaptation is implemented, a more reasonable choice of utility is a monotonically piece-wise increasing function of the SINR, reflecting the various coding rates implemented in the system. Assuming an idealized link adaptation protocol, i.e assuming Shannon capacity can be achieved at any SINR in any resource slot, the utility eventually converges to a smooth function reflecting the user's instantaneous rate in bits/sec/Hz. For the overall network utility we thus define the *sum capacity* [25] as

$$\mathcal{C}(\boldsymbol{U}, \boldsymbol{P}) \triangleq \frac{1}{N} \sum_{n=1}^{N} \log\left(1 + \Gamma([\boldsymbol{U}]_n, \boldsymbol{P})\right). \tag{2}$$

The capacity optimal resource allocation problem can now be formalized simply as:

$$(\boldsymbol{U}^*, \boldsymbol{P}^*) = \arg \max_{\substack{\boldsymbol{U} \in \Upsilon \\ \boldsymbol{P} \in \Omega}} \mathcal{C}(\boldsymbol{U}, \boldsymbol{P}), \tag{3}$$

The optimization problem above can be seen as generalizing known approaches in two ways: First, the capacity-maximizing scheduling problem is well-studied for a single cell scenario, but traditionally not jointly over multiple cells. Second, the problem above extends the classical multicell power control problem (which usually rather aims at achieving SINR balancing) to

include joint optimization with the scheduler. Despite its promise, solving (3) presents the system designer with several serious challenges.

The problem above is non-convex, and standard optimization techniques do not apply directly. On the other hand, an exhaustive search for the $(\boldsymbol{U}^*, \boldsymbol{P}^*)$ pair over the feasible set is prohibitive. Finally, even if computational issues were to be resolved, the optimal solution still requires a central controller updated with instantaneous inter-cell channel gains which would create acute signaling overhead issues in practice. The central question of this paper thus arises: Can we extract all or some of the gain related to multicell resource allocation using the solution of (3), within reasonable complexity and signaling constraints? Inspection of the recent literature reveals that this is a hot research issue with many possible tracks of investigation. We do not pretend to exhaust all of them here by any means. Instead we simply draw the reader's attention to a few promising ideas. As a first approach we examine so-called scaling laws of the network capacity under the optimal resource allocation solution. This study reveals in fact a much simpler structure for the problem (3) in the asymptotic regime (in the number of users per cell). For a fixed (moderate) number of users, however, we must consider techniques allowing for a sub-optimal solution to (3), using different approaches in Section V. These techniques however, remain mostly non-distributed. Next we turn to some distributed techniques, and offer some comparisons, in Section VII.

IV. NETWORK CAPACITY FOR ASYMPTOTIC NUMBER OF USERS

Let us consider a system with a large number of users in each cell. For the sake of exposition we shall assume in this particular section $U_n = U$ for all n, where U is asymptotically large, while N remains fixed. We expect a growth of the sum capacity $\mathcal{C}(\boldsymbol{U}^*, \boldsymbol{P}^*)$ thanks to the *multicell multiuser diversity gain*[3]. Thus we are interested in how the *expected* sum capacity *scales* with U. To this end a series of bounding arguments can be used [26].

A. Bounds on network capacity

First we present some simple bounds which hold in both the asymptotic and non-asymptotic regimes.

---

[3]The multicell multiuser diversity gain is a straightforward generalization of the conventional multiuser diversity [5] to multicell scenarios with joint scheduling.

Upper bound: An upper bound (ub) on the capacity for a given (not necessarily optimal) resource allocation vector is obtained by simply ignoring intercell interference effects:

$$\mathcal{C}(\boldsymbol{U}, \boldsymbol{P}) \leq \frac{1}{N} \sum_{n=1}^{N} \log\left(1 + \frac{G_{u_n,n} P_{u_n}}{\sigma^2}\right). \tag{4}$$

In the absence of interference, the maximum capacity is clearly reached by transmitting at a level equal to the power constraint, i.e. $\mathbf{P}_{\max} = [P_{\max}, .., P_{\max}]$ and selecting the user with largest channel gain in each cell (maximum rate scheduler), thus giving the following upper bound on capacity:

$$\mathcal{C}(\boldsymbol{U}^*, \boldsymbol{P}^*) \leq \mathcal{C}^{ub} = \frac{1}{N} \sum_{n=1}^{N} \log\left(1 + \Gamma_n^{ub}\right). \tag{5}$$

where the upper bound on SNR is given by the so-called MAX-SNR scheduler:

$$\Gamma_n^{ub} = \max_{u_n=1..U} \{G_{u_n,n}\} P_{\max}/\sigma^2 \tag{6}$$

Lower bound: A lower bound on the optimal capacity (in the presence of interference) $\mathcal{C}(\boldsymbol{U}^*, \boldsymbol{P}^*)$ can be derived by restricting the domain of optimization. Namely, by restricting the power allocation vector to full power $P_{\max}$ in all transmitters, we have

$$\mathcal{C}(\boldsymbol{U}^*, \boldsymbol{P}^*) \geq \mathcal{C}^{lb} = \mathcal{C}(\boldsymbol{U}_{FP}^*, \mathbf{P}_{\max}) \tag{7}$$

where $\boldsymbol{U}_{FP}^*$ denotes the optimal scheduling vector assuming full power everywhere, defined by

$$\boldsymbol{U}_{FP}^* = \arg\max_{\boldsymbol{U} \in \Upsilon} \mathcal{C}(\boldsymbol{U}, \boldsymbol{P}_{\max}), \tag{8}$$

Note that the n-th cell's user in $\boldsymbol{U}_{FP}^*$ is found easily via:

$$[\boldsymbol{U}_{FP}^*]_n = \arg\max_{\boldsymbol{U} \in \Upsilon} \Gamma_n^{lb} \tag{9}$$

where Γ_n^{lb} is a lower bound on the best SINR given by:

$$\Gamma_n^{lb} = \max_{u_n=1..U} \frac{G_{u_n,n} P_{\max}}{\sigma^2 + \sum_{i \neq n}^{N} G_{u_n,i} P_{\max}} \tag{10}$$

B. Scaling laws with number of users

We now investigate the rate of scaling with U for both upper and lower bounds on capacity. The scaling must depend on the statistics of the channel power gains and interference level, and in turn, on the assumed geometry of the network. In [26], an asymptotic analysis is carried out for two network geometries. In the first one, users are assumed to be located at a similar distance from their AP and at least at a distance R from interfering APs, where R plays the role of the cell radius. In this case, the following result is obtained for the scaling laws of $\mathcal{C}^{ub}$ and $\mathcal{C}^{lb}$ respectively:

Theorem 1: Let $G_{u_n,n} = \gamma_{u_n,n}^2 |h_{u_n,i}|^2$, $u_n = 1...U, n = 1...N$, where $\gamma_{u_n,n} = \gamma$ under constant path loss. Assume that $|h_{u_n,n}|^2$ is Chi-square distributed with 2 degrees of freedom ($\chi^2(2)$) (i.e. $h_{u_n,n}$ is a unit-variance complex normal random variable). Assume the $|h_{u_n,n}|^2$ are i.i.d across users. Then for fixed N and U asymptotically large, the upper bound on the SINR in cell n, and on the capacity scale respectively like

$$\Gamma_n^{ub} \approx \frac{P_{\max}\gamma^2}{\sigma^2} \log U \tag{11}$$

$$E(\mathcal{C}^{ub}) \approx \log \log U \tag{12}$$

where the expectation is taken over the complex fading gains.

The capacity scaling for the full-interference case (lower bound) is obtained as follows:

Theorem 2: Let $G_{u_n,i} = \gamma_{u_n,i}^2 |h_{u_n,i}|^2$, $u_n = 1..U, n = 1..N$, where $\gamma_{u_n,n} = \gamma$, $\gamma_{u_n,i}^2 = \beta d_{u_n,i}^{-\epsilon}$, for $i \neq n$, where $d_{u_n,i}$ is the distance between user u_n and AP i, β is a positive constant and ϵ is the path loss exponent. Assume $|h_{u_n,i}|^2$ is Chi-square distributed with 2 degrees of freedom ($\chi^2(2)$). Assume that the $|h_{u_n,i}|^2$ are i.i.d across users and cells. Then for fixed N and U asymptotically large, the lower bound on the SINR in cell n and on the capacity scale respectively like

$$\Gamma_n^{lb} \approx \frac{P_{\max}\gamma^2}{\sigma^2} \log U \tag{13}$$

$$E(\mathcal{C}^{lb}) \approx \log \log U \tag{14}$$

Note that the results above exploit tools from extreme value theory [27], which have been used before in the context of single cell opportunistic scheduling [28], [29].

Due to the upper and lower bounds showing similar scaling laws, the average of the network capacity with optimum power control and scheduling scales like

$$E(\mathcal{C}(\boldsymbol{U}^*, \boldsymbol{P}^*)) \approx \log\log U \tag{15}$$

The results above suggest that, in a symmetric multicell network, the capacity obtained with optimal multicell scheduling in both an interference-free environment and an environment with full interference power have identical scaling laws in $\log\log U$. This result bears analogy to the results in [29] which indicate that in a single cell broadcast channel with random beamforming and opportunistic scheduling, the degradation caused by inter-beam interference becomes negligible when the number of users becomes large. Here the multicell interference becomes negligible because the optimum scheduler tends to select users who have both large instantaneous SNR and *small* interference power.

Of course, it is better to take interference into account when doing the resource allocation. However, a system where the full power is allocated at all transmitters will be asymptotically optimal. As a result, a simple procedure based on (9) must be also asymptotically optimal. Interestingly, this algorithm is completely distributed as only local CSI is exploited by each user and fed back to its serving AP only. The SINRs can be computed during a preamble phase where all APs are asked to transmit pilot or data symbols at full power.

We may wonder what the scaling law becomes in a non-symmetric network where users are located at a random distance to their serving AP. This problem can also be investigated using extreme value theory and a surprising result is reached. In this case, the capacity is dominated by the selection of users located close to their serving AP. For a uniform distribution of users in a disc around the the AP, the growth rate obtained with optimum user scheduling and power control is shown to be [26]:

$$\mathcal{C}(\boldsymbol{U}^*, \boldsymbol{P}^*) \approx \frac{\epsilon}{2}\log U \tag{16}$$

Hence a much faster growth rate than in the case of symmetrically located users. In this case too, the power allocation based on transmitting at full power everywhere is asymptotically optimal and gives rise to a distributed user selection algorithm based on maximum SINR.

For a small to moderate number of users per cell, the analysis above may not be valid, and other approaches must be investigated. In particular the intuition is that *some* (maybe a small number of) selected transmitters then ought to back off from full power, for the good of the

network, and even possibly shut down completely for the considered resource slot. To pursue this idea in more depth we now consider the specific problem of capacity optimal power allocation. First we consider below centralized schemes in Section V and Section VI, then finally turn to distributed solutions in Section VII.

V. APPROXIMATING THE SOLUTION FOR A FINITE NUMBER OF USERS

In this and the next section we imagine that a candidate scheduling vector $\boldsymbol{U}$ has been selected, and focus on optimal power allocation. Even given a set of scheduled users, the simultaneous optimization of transmission rates and power with the goal of maximizing the sum network capacity is still a difficult problem, which perhaps explains why the problem has received only modest attention in the past. However, by restricting our attention to the high or low SINR regime, the power control problem can be efficiently solved.

A. Low SINR regime

In this regime we can apply an approximation of the achievable rate of each user, thus simplifying the problem. Specifically, when the SINR is low, the following approximation holds: $\log(1 + \Gamma([\boldsymbol{U}]_n, \boldsymbol{P})) \approx \frac{\Gamma([\boldsymbol{U}]_n, \boldsymbol{P})}{\ln 2}$. Thus, we have

$$\mathcal{C}(\boldsymbol{U}, \boldsymbol{P}^*) = \sum_{n=1}^{N} \log(1 + \Gamma([\boldsymbol{U}]_n, \boldsymbol{P})) \approx \frac{1}{\ln 2} \sum_{n=1}^{N} \frac{G_{u_n,n} P_{u_n}}{\sigma^2 + \sum_{i \neq n}^{N} G_{u_n,i} P_{u_i}}. \tag{17}$$

Inspecting (17), we see that there is a linear relationship between each user's achievable rate and power, equivalent to the setup in [30]. Then, it is clear that the sum rate in (17) is convex in each variable P_{u_n}[4] [30], and hence each base station, when on, should transmit with full power for optimality, i.e., on/off power control is optimal.

B. High SINR regime

On the other hand, if a high SINR assumption can be made we are again able to simplify the problem. In particular, when the SINR is much larger than 1, we can apply the following

[4]Note that this does not necessarily imply that the function itself is convex.

approximation [31][5], [32]:

$$\log(1 + \Gamma([\boldsymbol{U}]_n, \boldsymbol{P})) \approx \log(\Gamma([\boldsymbol{U}]_n, \boldsymbol{P})). \tag{18}$$

By virtue of this approximation, it can be shown that the sum rate maximization problem now is convex, and can be efficiently solved using geometric programming techniques [32].

The approximation is valid when the signal level is much higher than the interference level, or in CDMA systems where the spreading gain is large. However, this approximation by construction prohibits completely turning off any base stations at any time slot. This extra constraint may in fact cause the resulting power vector to steer away from the optimum solution. Indeed, as we will see in the next section, the ability to turn off base stations in certain time slots can be instrumental in approaching the maximum network capacity.

VI. BINARY VS. OPTIMAL POWER CONTROL

Designing wireless networks where the resource is to be optimized over many dimensions is a very complex task, and the complexity grows as fast as the number of variables in the design space increases. As discussed in the introduction, to reduce the complexity one idea is to limit the network parameter design space, for example by letting the transmit power take only quantized values. When reducing the design space, obviously a loss in system capacity can occur, and there will be a tradeoff between capacity and complexity. Although this trade-off is still by far an open problem, there has been some recent understanding as to how by a very coarsely quantized transmit power allocation, namely a binary or an "on-off" one, we can actually come very close to optimal power allocation. Besides complexity reduction, an important additional benefit of quantizing the power space is to allow distributed optimization. However this aspect is addressed later in the paper.

A. Binary vs. optimal power control for $N = 2$

We now consider the case of a cellular network with a small to moderate number of users, so the asymptotic results of Section IV do not apply. We start here by considering $N = 2$ cells. Note that the network may have more than two cells, in which case the cells are simply clustered

---

[5]In [31] instead of using this approximation, the authors present an alternative objective function that has certain properties, not commenting on the fact that the new objective function is identical to this high SINR approximation.

into adjacent groups of two cells, over which the power allocation is carried out. The two-cell case is treated separately here as it allows for an analytical derivation of a surprising result for the optimal power allocation. In particular the following result holds [17]:

Theorem 3: For the two-cell case, the capacity maximizing power allocation is binary. Mathematically, for any $\boldsymbol{U} \in \Upsilon$,

$$\arg \max_{(P_1,P_2) \in \Delta\Omega^2} \mathcal{C}(\boldsymbol{U},(P_1,P_2)) = \arg \max_{(P_1,P_2) \in \Omega} \mathcal{C}(\boldsymbol{U},(P_1,P_2)) \tag{19}$$

where $\Delta\Omega^2 = \{(P_{\max},0),(0,P_{\max}),(P_{\max},P_{\max})\}$ is the set of corner points in the feasible power domain. Interestingly, a similar result was independently reported also in [33].

In fact, this result remains true even when the noise variance is different in each cell, thus possibly accounting for out-of-cluster interference. Hence for the two-cell cluster case, the sum rate optimal power allocation has a remarkably simple nature: Depending on the noise variances and channel gains, *either* both APs should transmit at full power, *or* one of them should be shut down completely. From a practical point of view it means that the transmit power range per cell can be quantized to two values, either on or off, *without* loss in capacity! The optimal decision however requires simultaneous CSI from both cells.

B. Binary vs. optimal power control for $N > 2$

Without resorting to the high or low SINR approximations discussed in Section V, in the general SINR regime the power control problem proves to be very difficult due to the lack of convexity and the fact that the analysis for the two-cell case does not generalize to N cells. One of the first paper to treat the problem in detail was [31], giving an iterative procedure to solve the sum throughput maximization. However, due to the inherent non-convexity of the problem, no guarantees about the quality of the solution can be given. Another approach is to take advantage of recent developments in the mathematical framework of geometric programming. It can be shown that the sum of rates maximization subject to peak power constraints can be very well approximated by a geometric program. *Geometric programs* are a class of non-linear optimization problems characterized by objective and constraint functions that have a special form [34]. Then by solving a series of such geometric programs, at each step improving the approximation, the power control problem becomes manageable [32], [35].

Motivated both by the optimality of binary power allocation for the two-cell case and its simplicity, we also investigate the properties of binary power control also in the N-cell case. Binary power control for N cells is done by evaluating $\mathcal{C}(\boldsymbol{U}, \mathbf{P})$ at the corners of Ω. Mathematically,

$$\mathbf{P}_{\text{bin}} = \arg \max_{\mathbf{P} \in \Delta\Omega^N} \mathcal{C}(\boldsymbol{U}, \mathbf{P}), \tag{20}$$

where $\Delta\Omega^N$ is the set of $2^N - 1$ corner points of Ω (minus the all-zero point).

Unfortunately, a pessimistic theoretical result is obtained there: It can be shown that binary power allocation is no longer optimal for $N > 2$. However, it appears to be very well approximating the capacity obtained by the optimal solution resulting from continuous power control, as illustrated below.

Example 1: We simulated a $N = 3$ cell network with the following parameters. Common peak and minimum power constraints of $P_{\max} = 10^{-3}$, and $P_{\min} = 0$, respectively, assuming identical noise figures for the different receivers, the AWGN power is found as kT_0B, where k is Boltzmann's constant, $T_0 = 290$ Kelvin is the ambient temperature, and $B = 1$ MHz is the equivalent noise bandwidth, i.e., $\sigma_{Z_1}^2 = \sigma_{Z_2}^2 = \sigma_{Z_3}^2 = 4.0039 \times 10^{-15}$. As an example of the randomly generated channel gain matrix, based on path loss, shadowing and multipath effects we have

$$G = 10^{-9} \times \begin{pmatrix} 0.0432 & 0.0106 & 0.0012 \\ 0.0004 & 0.2770 & 0.0043 \\ 0.0045 & 0.0137 & 0.1050 \end{pmatrix}.$$

Then, by the best binary power allocation $(P_1, P_2, P_3) = (1, 1, 1)P_{\max}$, a sum throughput of 9.4555 bits/s/Hz is obtained, while by assigning the optimal powers $(P_1, P_2, P_3) = (1, 0.8595, 1)P_{\max}$ we get a throughput of 9.4590 bits/s/Hz. As we will see later, this example is quite typical in the sense that binary power control, though suboptimal, very often yields a throughput close to that obtained by optimally allocating the powers. While achieving only marginally higher sum throughput under the given power constraints, optimal continuous control can however offer some savings in terms of sum transmit power.

For a numerical comparison, we consider a hexagonal cellular system with an operating frequency of 1.8 GHz. Gains for *all inter-cell and intra-cell* AP-UT links are based on the COST-231 [36] path loss model including lognormal shadowing with standard deviation of 10 dB, as well as fast fading which is assumed i.i.d. with distribution $\mathcal{CN}(0, 1)$. The peak power

constraint is given by $P_{\max} = 1$ watt. Fig. 3 depicts the network capacity, in bits/s/Hz/cell, versus the number of cells. It is clear that introducing power control improves the throughput performance, and that binary power control is optimal for one and two cells. However, note that only marginal improvement is obtained in going from binary power control to optimal GP power control based on geometric programming for $N \geq 3$. As an example of how instrumental it is to be able to operate some cells at minimum power, we see that the total network capacity in bits/s/Hz/cell is less for two cells than for one cell when using full power. However, using binary and GP power control, we observe an increase in the total network capacity when going from one to two cells, due to better management of interference.

Although close to optimal, the binary power allocation based on (20) remains nonetheless exponentially complex in N, as well as centralized. In addition, it must be coupled with a proper scheduling algorithm. However the discrete nature of the power optimization opens a door onto simpler, iterative and/or even distributed approaches. One interesting way of cutting the search complexity for large values of N is *clustering*. For small cluster sizes, (20) can easily be solved by checking all corner points, based on the knowledge of or assumptions about interference created by the remaining clusters. For a given cluster of size $K << N$, the interference from the remaining $N - K$ cells then contributes as noise, i.e. the sum throughput of a cluster Q is given as,

$$\mathcal{C}_{\text{cluster}, Q} = \sum_{q \in Q} \log_2 \Big(1 + \frac{P_q G_{q,q}}{\sigma_{Z_q}^2 + \sigma_{I_q}^2 + \sum_{\substack{j \neq q \\ j \in Q}} P_j G_{q,j}} \Big), \tag{21}$$

where $\sigma_{I_q}^2 = \sum_{j \notin Q} P_j G_{q,j}$ is the interference from out-of-cluster. Assuming this interference term can be estimated or averaged, from the knowledge of the power activity in other clusters, then the following problem is solved for each cluster Q:

$$\mathbf{P}_Q = \arg \max_{\mathbf{P} \in \Delta \Omega^K} \mathcal{C}_{\text{cluster}, Q}. \tag{22}$$

VII. Distributed resource allocation techniques

Except in the asymptotic number of users regime, and to some rare exceptions (such as [32]), the methods presented so far require CSI of all cells to be centrally collected and processed (Fig.1(a)). As the network size grows, this turns into a serious impairment, especially in high mobility scenarios where the coherence time of the channel gain may not exceed a few tens of milliseconds. In a *distributed* scenario (Fig.1(b)), a resource allocation unit is instead located in

each cell, handling scheduling and resource management for the cell users based on local CSI and possibly *limited* remote CSI. Note that local CSI includes any information that is observable by a local user, such as the gain from its preferred AP as well as the power from the various interference sources. Though such information is enough to evaluate the capacity within the observed cell, it is not sufficient to evaluate the impact of the interference on the rest of the network, and in turn, on global capacity. Therefore, a framework allowing each cell to maximize its capacity without concern for interference created elsewhere must be suited to a distributed implementation. This is precisely the idea followed in *game theoretic* approaches, especially in so-called *non-cooperative* settings. This interesting line of work is examined below.

A. *Game theoretic resource allocation*

In recent years, game theory which has its roots in economics (see [37] for an overview), has been applied to several problems in communications [23]. Game theory involves a set of competing players looking to maximize their own utility, based upon certain actions they can take. When the users have no knowledge *a priori* about what others are simultaneously doing, e.g. in a distributed system, the game is *non-cooperative*, as users have no information on the basis of which they could cooperate. Non-cooperative game theory was first proposed as an efficient framework to formulate the power control problem for wireless data networks without centralized control, e.g. in *ad-hoc* networks [24].

Consider a set $\mathcal{N}$ of transmit-receive pairs communicating over a shared medium. Each transmit-receive pair represents a player, where each player n can adjust its transmit power $0 \leq p_n \leq P_n^{\max}$ to maximize its individual utility function f_n. Different utility functions can be considered, but for mutually interfering links, these are invariably dependent upon the user SINR. The user rate is an obvious candidate for utility. The SINR in turn depends on the transmit powers of all users in the network. Using the standard notation in the game theory literature, we denote the transmit power vector of users *other than user n* by $\mathbf{p_{-n}}$. The non-cooperative power control game can then be written as

$$\max_{0 \leq p_n \leq P_n^{\max}} f_n(p_n, \mathbf{p_{-n}}) \ \forall \ n.$$

In this setting, the resulting optimal outcomes of the game are categorized in two different ways, as seen below.

Nash Equilibrium and *Pareto Optimality*: A set of strategies (in this case transmit powers) are said to be at Nash equilibrium if no user can gain individually by *unilaterally* altering its own strategy. Mathematically, a transmit power allocation vector $(\widehat{p}_n, \widehat{\mathbf{p}}_{-\mathbf{n}})$ is at Nash equilibrium if for every n,

$$f_n(\widehat{p}_n, \widehat{\mathbf{p}}_{-\mathbf{n}}) \geq f_n(p_n, \widehat{\mathbf{p}}_{-\mathbf{n}}) \quad \forall \ p_n \leq P_n^{\max}.$$

Unfortunately, it is well known that the Nash equilibrium is not efficient in terms of sum utility i.e. $\sum_n f_n$. In particular, if the utility is the rate, it has been shown that players tend to act selfishly by increasing their transmit power $P_n^{\max}$ beyond what is reasonable [24]. By altering the strategy of some players, one may achieve a more *socially favorable* outcome. We thus define a Pareto optimal power allocation vector $\mathbf{p}^*$, i.e. a vector $\mathbf{p}^*$ for which there is no other power vector $\mathbf{p}$ such that for all $n \in \mathscr{N}$, $f_n(\mathbf{p}) \geq f_n(\mathbf{p}^*)$ and for some $n \in \mathscr{N}$, $f_n(\mathbf{p}) > f_n(\mathbf{p}^*)$. In other words, some players benefit while some players remain unaffected by this altered strategy.

In order to find a Nash equilibrium that is more Pareto efficient, *pricing* mechanisms have been investigated. Pricing aims at forcing users to act in a way that is more socially beneficial, in other words, to improve the sum of the utilities of the players. Pricing is typically incorporated as a penalty paid by a player in the utility function that is to be optimized. In our application, it is natural to suppose that the penalty a player pays should depend on its transmit power, since this determines the interference it is creating to the rest of the network. There are various ways of integrating a power-related cost in the utility. In [38] for instance, the utility is the ratio of the goodput by the power level itself. Additive costs are also possible, in which case, denoting the cost paid by each user by $c_n(p_n)$, we can rewrite the non-cooperative game with pricing as a maximization of the net utility:

$$\max_{0 \leq p_n \leq P_n^{\max}} \ \{f_n(p_n, \mathbf{p}_{-\mathbf{n}}) - c_n(p_n)\} \ \forall \ n.$$

There is a large body of literature considering various choices of utility and pricing mechanisms. In voice-oriented systems, utility functions are step function or sigmoid-like, geared toward trying to achieve a target SINR at each user as in [8]. In that case pricing may be used to stabilize power consumption when the SINR targets are close to the non-feasible region [39].

In data-oriented settings, the utility is usually a smoothly increasing function of the SINR. For instance the authors in [24], [40], [41] consider a function giving the amount of information successfully transferred per unit energy by each player, while the incurred cost is a linear function

of the transmit power. An iterative algorithm is proposed which maximizes the net utility by updating individual transmit powers assuming other players' power vectors to be constant. The downlink of a two-cell CDMA data network is studied in [42], with the goal of finding the optimal transmit powers for utility and revenue maximization. The AP announces a price to the users, which then demand certain powers based on maximization of the net utility.

Power control for transmit-receive pairs in an ad-hoc network is considered in [43]. Here, the cost is not a constant function, but is based on prices announced by the players to each other. Interestingly, the players charge each other for the interference created. The iterative algorithm updates the power and prices at every step, but this is not completely distributed as it requires channel gain information, as well as price updates, from all other users in the network. A truly distributed setting is obtained by making the pricing a simple linear function of the consumed power, as considered in some of the approaches discussed above. Clearly, an issue with pricing is that it should eventually be a function of the macroscopic parameters, like the number of cells, users, cell size etc. Finally, it is worth noting that, although significant work on resource allocation using game theoretic frameworks can be found, it appears that the problem of user scheduling in cellular networks has been little or not addressed in this framework, a fact probably due to the historic ties between game theory approaches and adhoc networks. In Section VII-D we investigate the performance of game theoretic power allocation using a pricing mechanism paired with a reasonable distributed user scheduling algorithm.

B. Distributed techniques using "on-off" power allocation

In this section, we exploit lessons learned from Section VI in terms of the near optimality of an "on-off" power allocation strategy in order to come up with a distributed optimization strategy. Hence we take on the discretized problem defined by:

$$(\boldsymbol{U}^*, \boldsymbol{P}^*) = \arg \max_{\substack{\boldsymbol{U} \in \Upsilon \\ \boldsymbol{P} \in \Delta\Omega^N}} \mathcal{C}(\boldsymbol{U}, \boldsymbol{P}) \tag{23}$$

In order for a solution to be practically feasible, computationally simple and distributed (although possibly sub-optimal) algorithms are desirable. In the following sections, we discuss various distributed approaches. A central idea behind all of these approaches is that of *interference-averaging*, which helps make the total interference more predictable in the case where the number

of interference sources is large. It also makes the SINR less dependent on CSI information not locally available at each cell.

C. Modeling interference in dense networks

Dense networks lend themselves to modeling of the total interference experienced by any user, thanks to the large number of interference sources being averaged at the receiver. Based on the observation that interference to any user in a large dense network is only *weakly* dependent on the user's position in the cell, we can approximate the interference term in (1) by an average interference gain, denoted by G, which is independent of the user location multiplied by the total transmit power of active interferers:

$$\sum_{i \neq n}^{\tilde{N}} G_{u_n,i} P_i \approx G \sum_{i \neq n}^{\tilde{N}} P_i = G(\tilde{N} - 1) P_{\max}, \tag{24}$$

where G is independent of u_n, and is a certain function of pathloss, link budget parameters, etc., and $\tilde{N}$ is the number of active "on" cells. Though only an approximation, this model is supported by simulations [44]. More importantly, it proves a remarkably useful tool for algorithm design in large networks. The estimation of G can be made from simulations for a given network. It can be also used as a tuning parameter, expressing the overall weight we want to give to co-channel interference in the resource allocation. To this end, worst case and best case values can be obtained from statistical analysis [44]. However, a nicer and more practical result is reported below.

1) Iterative power allocation in the interference limited regime: A general approach for distributed power allocation under the framework of (23) consists in letting each cell take a activate/deactivate decision corresponding to a transmission with full or zero power. From the point of view of sum capacity, each cell will exploit locally available CSI in order to determine whether its activation creates a capacity gain that outweighs (or not) the capacity loss caused by its interfering power to the rest of the network. From manipulations of (2), (1), and assuming both high SINR in active cells and a high interference-to-noise ratio, the best decision regarding

activation of cell m is taken upon inspection of the following inequality :

$$\frac{G_{m,m}P_m}{\displaystyle\sum_{\substack{i\neq m \\ i\in\mathcal{N}}} G_{m,i}P_i} \geq \frac{\displaystyle\prod_{\substack{n\in\mathcal{N} \\ n\neq m}} G \sum_{\substack{i\neq n \\ i\in\mathcal{N}}} P_i}{\displaystyle\prod_{\substack{n\in\mathcal{N} \\ n\neq m}} G \sum_{\substack{i\neq n\neq m \\ i\in\mathcal{N}}} P_i}.$$

where the average interference model (24) was used, yet the knowledge of the average interference gain G is not needed, suggesting robustness with respect to the model (24). If all "on" cells transmit with same power $P_{\max}$, and denoting $|\mathcal{N}| = \tilde{N}$ where $\mathcal{N}$ is the set of active cells, cell m will activate itself if

$$\frac{G_{m,m}}{\displaystyle\sum_{\substack{i\neq m \\ i\in\mathcal{N}}} G_{m,i}} > \left(\frac{\tilde{N}-1}{\tilde{N}-2}\right)^{(\tilde{N}-1)} \approx e, \tag{25}$$

where the right hand side approximation is valid for a large enough network (more than, say, 10 cells). Note that the left hand side is the signal-to-interference (SIR) level. Depending on whether (25) is used with or without the approximation, semi- or fully-distributed power control can be obtained. In the semi-distributed case, the number of active cells from the previous iteration must be communicated to the new cells. However this information could also be measured locally from pilot signals.

The overall algorithm works as follows: Starting with an all "on" power allocation vector, each cell simultaneously measures the SINR of the best user and based on (25) remains active or inactive during the next iteration. For a cell to have the best chance to be active and thus contribute to the network capacity, at every iteration it should schedule the user which has the best chance of satisfying the inequality (25) i.e. the user with the maximum left-hand-side of (25), based on the power allocation resulting from the previous iteration. Thus, along with power allocation, the technique above suggests a corresponding scheduling policy. This is in contrast to the game theoretic approach described previously in which only power allocation is addressed.

As a result of this algorithm, a subset of the total number of cells transmit simultaneously during a given scheduling period. Though some cells may stay silent, they may be active during the next scheduling period. This approach can also be considered as a distributed mechanism for *dynamic spectral reuse*. In contrast with traditional cellular networks, the reuse pattern obtained

with this method is random, possibly highly irregular as illustrated in Fig. 4, and varies from one scheduling period to the next as a function of the channel state information of the cell users.

D. Performance issues

In this paper, we have essentially discussed two different approaches for multicell resource allocation in wireless networks equipped with link adaptation capabilities:

1) The *game theoretic formulation*, allowing users (cells) to behave selfishly by individually maximizing their capacity and then inducing a sort of social benefit (for network capacity) through pricing. This approach is by nature distributed, as the measure of performance is based on individual cell capacity which can be calculated locally. However, maximizing individual capacity does not necessarily maximize the network capacity. Pricing does try to move the behavior to benefit the network, but this entails finding the right cost function, which sets up another optimization problem to solve. Moreover, the game theoretic approaches in the literature deal mostly with power allocation. To the best of our knowledge, coupling user scheduling with power allocation for multiuser diversity gain has been overlooked so far in the game theoretic framework.

2) *Iterative Binary Power Allocation* on the other hand, tries to directly maximize the network capacity by exploiting the interference-averaging effect in large (in the number of nodes) networks. Thus, it can be truly said to address the joint user scheduling and power allocation problem. However, it relies on a restricted feasible power set, as well as on an approximate model for interference in large wireless networks.

Due to the differing philosophies of tackling the multicell resource allocation problem, the outcome of a quantitative comparison of the two methodologies is difficult to predict. To this end, we compare the two through Monte Carlo simulations. However, in order to have a fair comparison, we need to adapt the game theoretic approach to handle user scheduling as well. Thus, given a multicell system, the user with the best SINR is selected in each cell and then the power allocation game with pricing is played. As our goal is capacity maximization, for this game, f_n is the cell capacity given by,

$$f_n(p_n, \mathbf{p_{-n}}) = \log_2 \left(1 + \frac{p_n G_{n,n}}{\sigma^2 + \sum_{i \neq n} p_i G_{n,i}} \right).$$

As in much of the current literature, pricing is considered to be a linear function of the transmit power,

$$c_n(p_n) = \alpha p_n,$$

where α is the *cost factor*. An iterative algorithm is adopted, where at each iteration the transmit power which maximizes the net utility is found, considering the interference to be constant. As already stated, an issue that arises in the game theoretic approach is tuning the value of α. Here, as the number of users vary, the net utility is calculated for a range of cost factors. The α that gives the best net utility is considered the optimal cost factor, and the corresponding net utility is used to calculate the network capacity.

We consider a 19-cell (hexagonal layout) system with the same parameters as described in Section VI-B. We measure the average network capacity as the number of users per cell is varied. We also plot the network capacity for a network employing maximum SINR scheduling with full reuse (maximum power at all AP). The results show (Fig. 5) that the game theoretic approach and the iterative binary power allocation in this case both perform better than full reuse, and that they give quite similar average network capacity. There is roughly a 50 percent capacity gain, which decreases when the number of active users increases. It is important to note that the number of active users corresponds to users simultaneously active within a scheduling window (of a few 10s of ms perhaps), therefore this number is likely to be low in all cases.

As the number of users increases, both implicitly converge to full reuse (Fig. 6). This result is expected and can be explained as follows. From (3), ideally we want to schedule N users at full power that have good communication links and are not too affected by interference. Then, with an increasing number of users, the probability of finding such users increases. This is also the sense of the asymptotic results reported in Section IV.

VIII. Fairness issues

In single cell scenarios with channel aware scheduling, it is known that maximizing the cell capacity is done at the expense of the fairness between the various users' throughputs, as users endowed with favorable channel conditions are allocated a relatively greater share of the spectral resource. Analogous conclusions will be obtained in a multicell context. For instance, the iterative binary power control algorithms tends to activate cells for which the best user enjoys good

channel gain and little interference, while cells for which users suffer from worse SIR will have a greater chance of shutdown. Although sum-rate optimal, this type of policy is not giving a fully fair treatment to the various cells, and consequently to the users in it. For instance, some cells might experience long periods of silence due to prolonged detrimental fading conditions or a poor user distribution. However, over a longer timescale the system is likely to be fair across cells for a symmetric user distribution over the cells. To maintain some minimum QoS at every cell, different strategies may be imagined, although in the multicell context these remain open research topics. One is to modify the utility measured by each cell to take into account the accumulated throughput by each of the users, in a similar way to the proportional-fair type utility used in certain single cell scheduling techniques [28], [45]. The buffer state at each AP could also be considered to prevent overflow. Finally, an interesting way to restore some fairness is to open the resource space to generate additional scheduling slots, over which the channel quality is different and cells get an additional chance of activation. This can be done using scheduling in both time and frequency slots (multicell OFDMA).

IX. FURTHER DISCUSSIONS

We have given some leads towards optimizing the resource in an interference limited multicell network, via joint power allocation and channel-aware user scheduling in view of maximizing the sum of user rates. Although difficult to solve in general settings, this problem promises significant gains in network capacity. It also admits simple properties in certain cases that can substantially reduce the complexity and make it practically relevant. If the number of users grows asymptotically, the transmitters should all operate at the power constraint level and the best user in each cell can be scheduled, in a fully distributed fashion. The loss due to interference then does not affect how the network capacity scales with the number of users. For small to moderate number of users, power control simply based on an on-off policy yields near-optimal performance for more than 2 cells and exactly optimal for two cells. The on-off power control policy can be exploited to derive simple and to a large extent distributed resource allocation algorithms. When the number of cells is large, a cell can decide of its own activation based on a condition involving purely locally measurable data, resulting in a fully distributed algorithm for power control at the base stations. Unlike the centralized solutions, the distributed solutions involve users feeding back channel gain information to their serving AP only, without the need

for sharing this information across many APs. Channel gain feedback is the price to pay to achieve coordination gains. Note however that this type of feedback is required in any case by the link adaptation protocol to select the appropriate coding and modulation schemes, and thus may be reused by the resource allocation protocol as well. Game theoretic approaches can also yield distributed algorithms for resource allocation. Typical utilities used there depart from the network's sum capacity. However, an optimization procedure can be used to modify the per-user utility so as to improve the social optimality of the Nash equilibrium, in terms of sum rate. For other choices of utilities (such as energy efficiency [46]) the difference between the Nash equilibrium and the Pareto optimal solution has been shown to be small.

In the techniques above, joint power control and scheduling may be interpreted as a form of cooperation between cells, which amplify or attenuate their power in the best interest of the network capacity. Alternatively, forms of multicell cooperation which involve some degree of coding or coherent signal combining between the base stations have been proposed [47], [48]. One way to optimize the link between several APs and one or more user is to imagine a distributed multiuser MIMO system, over which standard multi-user MIMO techniques such as Dirty Paper precoding, linear precoding or decoding, multiuser detection and space-time codes. Although these are promising techniques, some of the immediate challenges associated with this form of cooperation are the need to route data traffic to several sites in the network, the need for accurate CSI at the APs, the fact that some of these algorithms do not lend themselves easily to a distributed implementation (as opposed to MAC-layer oriented cooperation schemes), and finally the lack of scalability when the number of AP gets large. As a result, an interesting direction of research may lie in the combined use of cooperation concepts at the PHY layer (multibase MIMO etc.) on a small scale basis and coordination at the MAC layer on a larger scale basis.

REFERENCES

[1] *Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands*, IEEE Std. Std 802.16e-2005, 2005.

[2] 3rd generation - Long Term Evolution. 3rd Generation Partnership Project. [Online]. Available: http://www.3ggp.org

[3] A. Goldsmith, *Wireless Communications*, Cambridge Univ. Press, 2005.

[4] G. Song and Y. Li, "Utility-based resource allocation and scheduling in OFDM-based wireless broadband networks," *IEEE Communications Magazine*, vol. 43, Dec. 2005.

[5] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. on Comm.*, Seattle, June 1995, pp. 331–335.
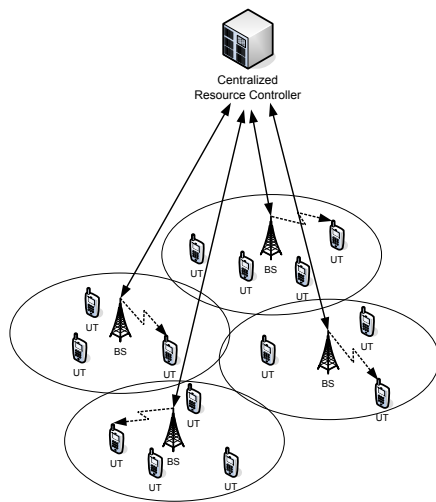
[6] M. Andrews, K. Kumaran, K. Ramanan, A. Stoylar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.*, pp. 150–154, Feb. 2001.

[7] J. Zander, "Distributed cochannel interference control in cellular radio systems," *IEEE Trans. Veh. Technol.*, vol. 41, pp. 305–311, Aug. 1992.

[8] G. J. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, 1993.

[9] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1341–1347, 1995.

[10] V. Tralli, R. Veronesi, and M. Zorzi, "Power-shaped advanced resource assignment for fixed broadband wireless access systems," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 2207–2220, Nov. 2004.

[11] R. Veronesi, V. Tralli, J. Zander, and M. Zorzi, "Distributed dynamic resource allocation for multicell SDMA packet access networks," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 1–12, 2006.

[12] K. Chawla and X. Qiu, "Quasi-static resource allocation with interference avoidance for fixed wireless systems," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 493–504, March 1999.

[13] K. K. Leung and A. Srivastava, "Dynamic allocation of downlink and uplink resource for broadband services in fixed wireless networks," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 990–1006, May 1999.

[14] T. Bonald, S. Borst, and A. Proutière, "Inter-cell scheduling in wireless data networks," in *Proc. European Wireless*, Cyprus, April 2005.

[15] M. H. Ahmed, H. Yanikomeroglu, and S. Mahmoud, "Interference management using basestation coordination in broadband wireless access networks," *Wireless Communications and Mobile Computing*, vol. 6, pp. 95–103, 2006.

[16] D. J. Love, R. W. Heath, W. Santipach, and M. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Communications Magazine*, Oct. 2004.

[17] A. Gjendemsjø, D. Gesbert, G. E. Øien, and S. G. Kiani, "Optimal power allocation and scheduling for two-cell capacity maximization," in *Proc. Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, Boston, MA, April 2006, pp. 1–6.

[18] K. N. Lau, "Analytical framework for MIMO uplink space-time scheduling design with convex utility functions," *IEEE Trans. on Wireless Communications*, pp. 1832–1843, Sept. 2004.

[19] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1747–1758, Oct. 1999.

[20] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 171–178, Feb. 2003.

[21] G. Li and H. Liu, "Downlink dynamic resource allocation for multi-cell OFDMA system," in *Proc. IEEE VTC*, Oct. 2003.

[22] H. Kim, Y. Han, and J. Koo, "Optimal subchannel allocation scheme in multicell OFDMA systems," in *Proc. IEEE VTC*, May 2004.

[23] E. Altman, T. Boulogne, R. El-Azouzi, T. Jimenez, and L. Wynter, "A survey on networking games in telecommunication," *Comput. Oper. Res.*, vol. 33, pp. 286–311, Feb. 2006.

[24] D. Goodman and N. Mandayam, "Power control for wireless data," *IEEE Personal Commun. Mag.*, vol. 7, pp. 48–54, April 2000.

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[26] D. Gesbert and M. Kountouris, "Resource allocation in multicell wireless networks: Capacity scaling laws and distributed techniques," in *IEEE Workshop on Resource Allocation in Wireless Networks (RAWNET WiOpt)*, 2007.

[27] H. A. David and H. N. Nagaraja, *Order Statistics, third ed.* New York: John Wiley and Sons, 2003.

[28] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.

[29] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channel with partial side information," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[30] A. Bedekar, S. Borst, K. Ramanan, P. Whiting, and E. Yeh, "Downlink scheduling in CDMA data networks," in *Proc. IEEE Glob. Telecom. Conf.*, Rio de Janeiro, Brazil, Dec. 1999, pp. 2653–2657.

[31] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 884–895, June 1999.

[32] M. Chiang, "Geometric programming for communications systems," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 1/2, pp. 1–156, Aug. 2005.

[33] M. Ebrahimi, M. A. Maddah-Ali, and A. K. Khandani, "Power allocation and asymptotic achievable sum-rates in single-hop wireless networks," in *Proc. CISS*, Princeton, NJ, March 2006.

[34] S. Boyd, S. J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," to appear in Optimization and Engineering 2007. [Online]. Available: http://www.stanford.edu/~boyd/gp_tutorial.html

[35] C. W. Tan, D. P. Palomar, and M. Chiang, "Solving nonconvex power control problems in wireless networks: Low SIR regime and distributed algorithms," in *Proc. IEEE Glob. Telecom. Conf.*, St. Louis, MO, Nov.-Dec. 2005, pp. 3445–3450.

[36] *Urban Transmission Loss Models for Mobile Radio in the 900 and 1800 MHz Bands*, EURO-COST Std. 231, 1991.

[37] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.

[38] F. Meshkati, V. Poor, S. Schwartz, and N. Mandayam, "An energy-efficient approach to power control and receiver design in wireless data networks," *IEEE Trans. on Communications*, Nov. 2005.

[39] M. Xiao, N. B. Shroff, and E. K. P. Chong, "A utility-based power-control scheme in wireless cellular systems," *IEEE/ACM Trans. Networking*, vol. 11, pp. 210–221, April 2003.

[40] C. Saraydar, N. B. Mandayam, and D. J. Goodman, "Pricing and power control in a multicell wireless data network," *IEEE Trans. Commun.*, vol. 19, pp. 1883–1892, Oct. 2001.

[41] ——, "Efficient power control via pricing in wireless data networks," *IEEE Trans. Commun.*, vol. 50, pp. 291–303, Feb. 2002.

[42] C. Zhou, M. L. Honig, and S. Jordan, "Utility-based power-control for a two-cell CDMA data network," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2764–2776, Nov. 2005.

[43] J. Huang, R. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Select. Areas Commun.*, vol. 24, pp. 1074–1084, May 2006.

[44] S. G. Kiani and D. Gesbert, "Interference modelling in full reuse wireless networks," *IEEE Trans. Wireless Comm., submitted. Available upon request.*

[45] F. Kelly, "Charging and rate control for elastic traffic," *Eur. Trans. on Telecommunications*, vol. 8, pp. 33–37, 1997.

[46] F. Meshkati, V. Poor, and S. Schwartz, "Energy-efficient resource allocation in wireless networks: An overview of game-theoretic approaches," *To appear in the IEEE Signal Processing Magazine: Special Issue on Resource-Constrained Signal Processing, Communications and Networking*, May 2007.

[47] C. Rose, S. Ulukus, and R. Yates, "Wireless systems and interference avoidance," *IEEE Transactions on Wireless Communications*, vol. 1, Jul. 2002.

[48] M. K. Karakayali, G. J. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Communications*, Aug. 2006.

LIST OF FIGURES

(a) A multicell system managed by centralized resource controller. This controller processes all network information jointly.
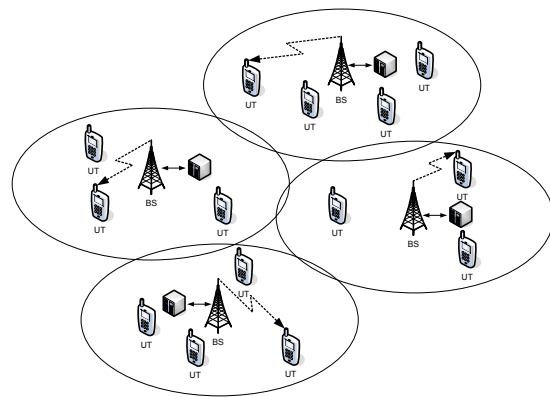
(b) A distributed multicell system requires no centralized control. Each cell performs resource allocation based on local channel knowledge (and possibly limited intercell information).

Fig. 1.   Centralized vs. Distributed Network Control.

(a)

(b)
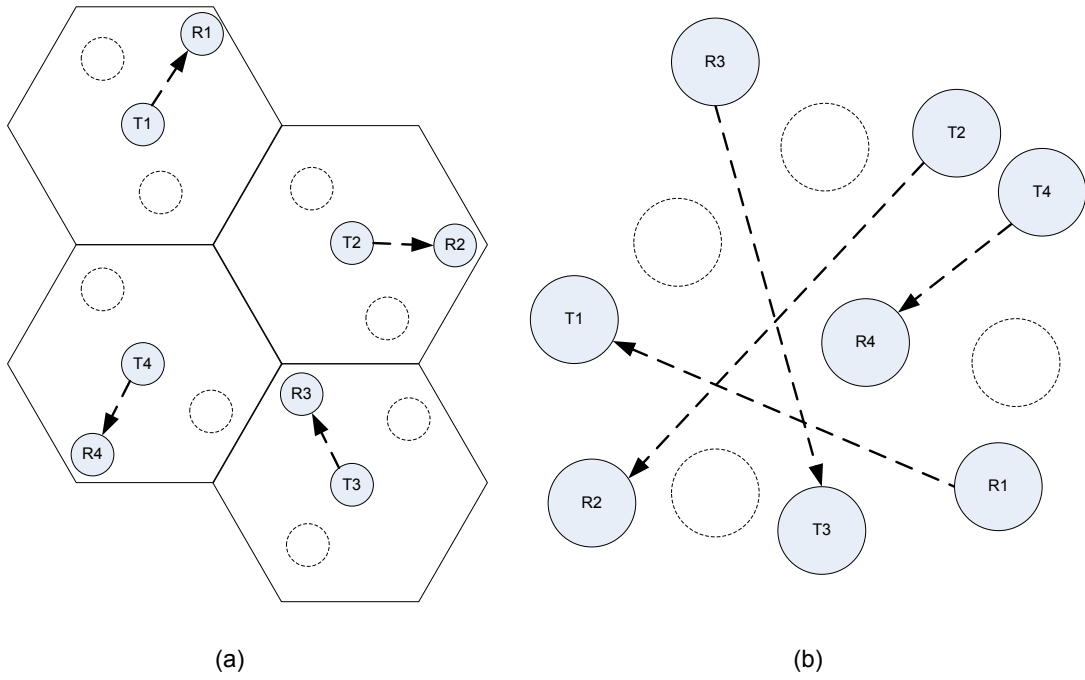
Fig. 2. Snapshot of network model, with $N = 4$ interfering pairs of transmitters T_i and receivers R_i. The cellular model (a) and the single-hop peer-to-peer or adhoc model (b) give rise to equivalent mathematical models. Dashed circles refer to silent users while solid circles refer to access points or users selected by the scheduler in a given resource slot.
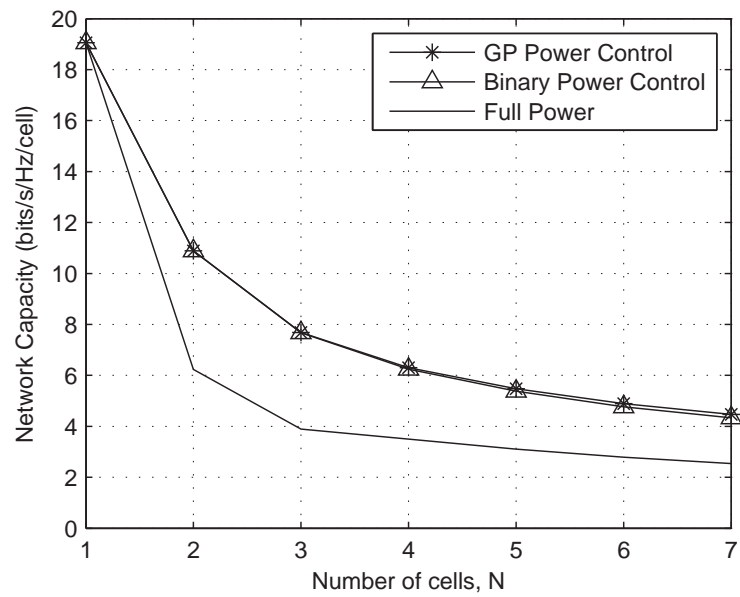
Fig. 3.   Network capacity vs. number of cells. GP power control based on geometric programming provides only marginal improvement as compared with binary power control.
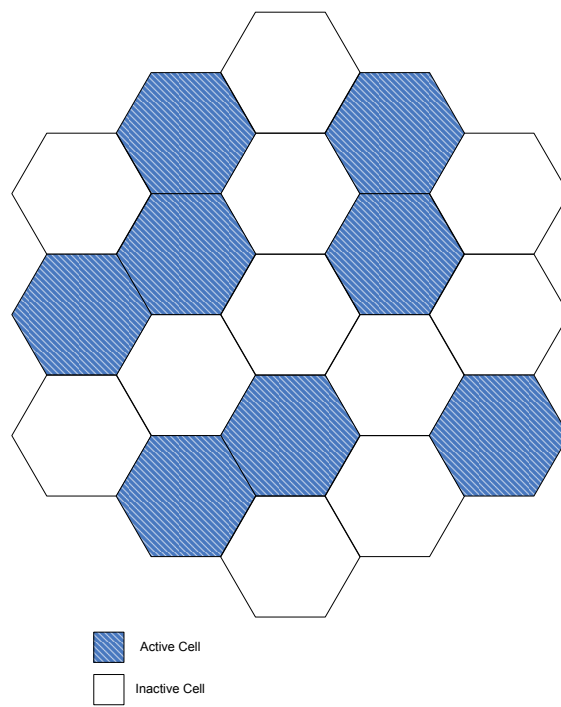
Fig. 4.   Possible irregular reuse pattern at a given scheduling period due to dynamic spectral reuse.
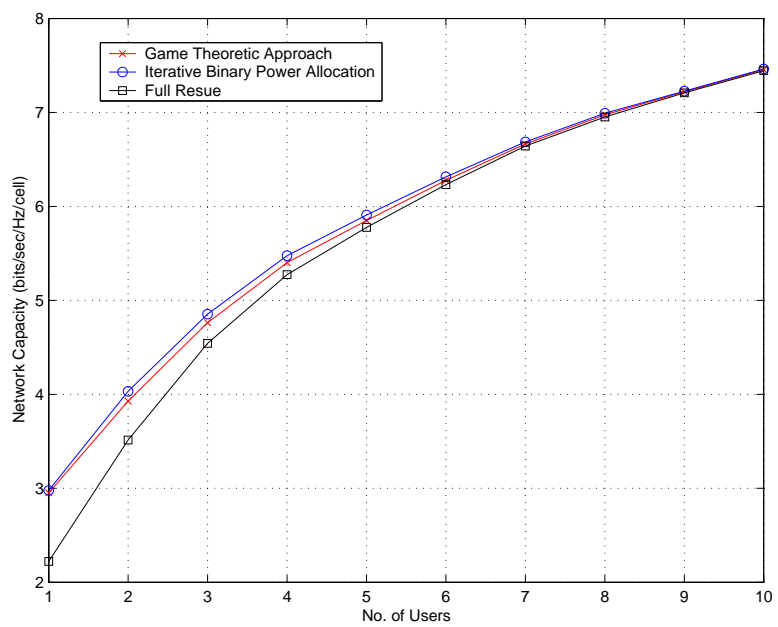
Fig. 5.   Network capacity vs. number of users for hexagonal cellular system with 19 cells. Game theoretic and Iterative Binary Power Allocation perform quite similar and converge to Full Reuse as the number of users increase.
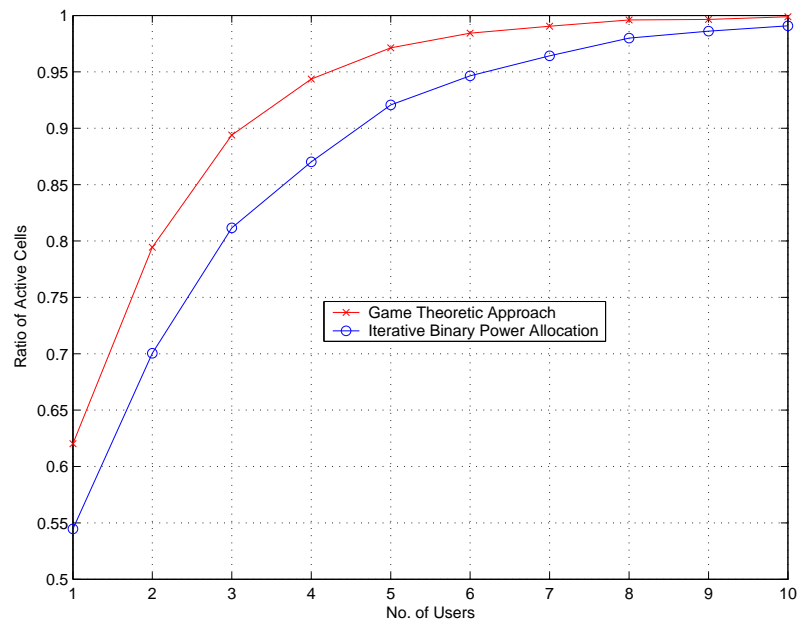
Fig. 6. Ratio of active cells vs. number of users. The game theoretic approach keeps more cells active than iterative binary power allocation. As the number of users increases, the majority of cells are active.