# Joint network-wide opportunistic scheduling and power control in multi-cell networks
— **Source link** ⬈

Jeong-woo Cho, Jeonghoon Mo, Song Chong

**Institutions:** Norwegian University of Science and Technology, Yonsei University, KAIST

Related papers:

- Binary Power Control for Sum Rate Maximization over Multiple Interfering Links

- Joint Network-wide Opportunistic Scheduling and Power Control in Multi-cell Networks

- Fair end-to-end window-based congestion control

- Dynamic load balancing through coordinated scheduling in packet data systems

- Inter-cell scheduling in wireless data networks

# Joint Network-wide Opportunistic Scheduling and Power Control in Multi-cell Networks

Jeong-woo Cho, Jeonghoon Mo, and Song Chong, *Members, IEEE*

*Abstract*—We present a unified analytical framework that maximizes generalized utilities of a wireless network by network-wide opportunistic scheduling and power control. That is, base stations in the network jointly decide mobile stations to be served *at the same time as* the transmission powers of base stations are coordinated to mitigate the mutually interfering effect. Although the maximization at the first glance appears to be a mixed, twofold and nonlinear optimization requiring excessive computational complexity, we show that the maximization can be transformed into a pure binary optimization with much lower complexity. To be exact, it is proven that binary power control of base stations is *necessary and sufficient* for maximizing the network-wide utilities under a physical layer regime where the channel capacity is linear in the signal-to-interference-noise ratio. To further reduce the complexity of the problem, a distributed heuristic algorithm is proposed that performs much better than existing opportunistic algorithms. Through extensive simulations, it becomes clear that network-wide opportunistic scheduling and power control is most suitable for fairness-oriented networks and underloaded networks. We believe that our work will serve as a cornerstone for network-wide scheduling approaches from theoretical and practical standpoints.

*Index Terms*—Opportunistic scheduling, wireless network, power control, max-min fairness, proportional fairness.

## I. INTRODUCTION

RESOURCE allocation in wireless data network in a system perspective has drawn many attentions for the last ten years. As people from the networking research have paid more attentions to wireless networks, more progresses are made in the wireless network scheduling and power control. They have considered various system objectives such as total throughput maximization (or max-C/I) [1], max-min fairness [2], and proportional fairness [3] and proposed schedulers to achieve the given system objective. The max-min fair scheduler, though it is one of the most popular ones in wired

J. Cho is with the Centre for Quantifiable Quality of Service in Communication Systems, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. E-mail: jeongwoo@q2s.ntnu.no.

J. Mo (Corresponding Author) is with the Department of Information and Industrial Engineering, Yonsei University, Seoul 120-749, Korea. E-mail: j.mo@yonsei.ac.kr.

S. Chong is with the School of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea. E-mail: song@ee.kaist.ac.kr.

networks, turns out to be less efficient due to the *solidarity property* of wireless networks: the throughput of each user is equalized to the smallest user throughput [4]. They claimed proportional fair scheduler can be a better candidate in the wireless networks since it is easy to implement and provides a good balance between fairness and efficiency. Viswanath *et al.* [5] and Borst [6] showed that proportional fairness becomes equivalent to equal-time fairness under some assumptions.

Opportunistic scheduling concept was introduced in wireless data networks to increase the average throughput of wireless channels by exploiting the time-varying characteristics of wireless channel [1], [5]–[7]. Since channel conditions of users are good and bad randomly, higher throughput can be achieved by scheduling a flow whose instantaneous channel condition is relatively better than others. To exploit this diversity further, Viswanath *et al.* [5] proposed a proactive beamforming scheme in which fast channel fluctuations are artificially induced by multiple transmit antennas in a pseudorandom manner.

Even with aforementioned advances in wireless data networks, little work has been done in multi-cell network-wide scheduling. Both theoretical difficulty and impracticality prevent people from pursuing the multi-cell network scheduling problem. Instead, multi-cell problems with different objectives have been considered: load balancing among cells [8], [9], and a low data rate problem of cell boundary users [10].

Sang *et al.* considered multi-cell problem in a slightly different angle of load balancing in their seminal work [8]. They proposed an *intra-cell opportunistic downlink scheduling* algorithm where each BS (base station) exploits user diversity gain in the corresponding cell independently to achieve $(w, \alpha)$-proportional fairness, that was proposed by Mo *et al.* [11]. When $\alpha = 0$, an MS (mobile station) having the strongest channel is picked out by the scheduler (max-C/I). When $\alpha = 1$ and $\alpha \to \infty$, proportional fairness and max-min fairness are achieved, respectively. The central server participates in load balancing by adjusting $\alpha$ value of each cell, named as *cell breathing* in which $\alpha$ is a flexible knob to balance the time-slots allocated to MSs close to the BS and MSs at cell boundaries. For *load balancing*, each MS adapts to its channel variation and the load fluctuations by initiating the load-aware handoff and cell-site selection where their own condition is adopted rather than network-wide proportional fairness.

Bu *et al.* [9] also considered a similar problem and proposed *load balancing* schemes that achieve network-wide proportional fairness if the channel capacity tends to be linear in the SINR (signal-to-interference-noise), while the load balancing scheme proposed in [8] did not directly tackle the network-wide proportional fairness due to its complexity. They

showed that the general problem is NP-hard and provided efficient offline and online heuristic algorithms to solve an approximated problem. Kauffmann *et al.* [12] proposed two distributed algorithms, based on the annealed *Gibbs sampler* [13], for *channel selection* for interference mitigation, and *user association* (load balancing) for fair and optimal resource allocation in WLANs (Wireless Local Area Network). They also proved that the distributed algorithms lead to efficient spectrum usage and improved performance in the context of minimal potential delay fairness. As the proposed algorithms do not require explicit coordination among stations, the applicability latent within the Gibbs sampler technique is noticeable.

So called SSDT (Site Selection Diversity Transmission), which is an idea of BS coordination paying *no* regard to scheduling, was once considered in 3GPP WCDMA system [14]. In SSDT, an MS selects the BS with the largest SINR using rapid physical layer signaling. The main weak point of SSDT is that an MS to be served is decided by a scheduler (possibly controlling multiple BSs) in advance and SSDT decides only whether or not to turn off some BSs. In this sense, SSDT does not pertain to achieving network-wide fairness.

Recently, the insufficient data rate problem of cell boundary users in the OFDMA system attracted the most attentions ever due to popularity of new standards such as the IEEE 802.16e [15] and the evolved UTRA [16]. Many people considered different frequency reuse schemes. However, [17] observed high throughput loss with the frequency reuse schemes and recommended reuse factor of one. Our work can be considered as an extension of [17] from the fixed frequency reuse scheme to a dynamic one.

Bonald *et al.* [10] considered a novel notion of scheduling, that is to say, coordination of transmissions among BSs. Unlike [8], [9] where all BSs transmit at full power all the time, they formulated an optimization problem which opportunistically coordinates instantaneous transmission powers of BSs in order to provide favorable channels to MSs at the cell boundaries. However, they did not consider the objective of network-wide proportional fairness and opportunistic scheduling. In a similar context, Kiani *et al.* [18] addressed the problem of *distributed inter-cell coordination* where BS transmission can be switched off to maximize the total throughput. A unique feature that makes this work noteworthy is that each cell makes a decision regarding its transmission power based only on the SINR of the users within the cell. In addition, the proposed distributed algorithm lends itself to exploiting multiuser diversity.

From a quite different angle, Karakayali *et al.* [19] proposed a network-wide coordination of the BS transmissions so as to mitigate inter-cell interference. They showed that an improved form of signal-level coordination is obtained when a zero-forcing equalizer is combined with dirty paper coding. Since the performance metric for comparisons was the max-min rate achievable subject to per BS power constraints, the proposed signal-level coordination was turned out to improve the system capacity significantly. In Section IV-C, we explain why inter-cell interference mitigation is favorable for max-min fairness.

Recently, Li and Liu in their work [20], [21, Chapter 7] considered a semi-distributed radio resource allocation scheme in OFDMA networks where radio resource allocation is divided into two algorithms, *i.e.*, an RNC (radio network controller) algorithm and a BS algorithm. That is, RNC coordinates inter-cell interference between BSs at super-frame level and each BS makes its channel assignment decision on frame level based on users' traffic conditions. However, because a BS algorithm follows RNC's decision when all users in the BS have traffic to send in each frame, their scheme can be viewed as an RNC-centric radio resource allocation where each BS exploits only the *traffic diversity* [20]. Besides, they also did not consider the objective of network-wide proportional fairness but the total throughput maximization.

In this paper, we consider network-wide opportunistic scheduling problem under a multi-cell environment based on the optimization problem proposed in [8]. We would like to answer the following questions in particular:

*How much improvement can we make by considering inter-cell coordination? For what kind of network is inter-cell coordination most suitable?*

To answer the questions, we formulate an optimization problem which maximizes generalized utilities of a multi-cell network by network-wide opportunistic scheduling *and* network-wide opportunistic power control rather than by load balancing. Since load balancing without considering network-wide opportunistic power control is already too complicated [9], it is not treated in this paper to better focus on our problem.

Although the maximization in the first formulation appears to be a mixed, twofold and nonlinear optimization requiring excessive computational complexity, we prove that the maximization can be transformed into a pure binary optimization. Correctly speaking, we prove that binary power control of base stations is *necessary and sufficient* for maximizing the network-wide utilities under situations in which the channel capacity tends to be linear in the SINR. Consequently, network-wide opportunistic scheduling *absorbs* network-wide opportunistic power control. For distributed operation of network-wide optimization in real networks, a heuristic algorithm using local information of neighbor BSs is proposed. Through extensive simulations, we show that network-wide opportunistic scheduling and power control improve various performance indices considerably.

The rest of the paper of organized as follows: The first formulation including network-wide opportunistic scheduling, power control and load balancing is presented and simplified in Section II. Since the simplified problem is still intractable, we propose a distributed heuristic algorithm with much lower complexity in Section III. The performance of our proposed algorithm is verified in Section IV. Then we conclude this work with answering the questions brought up in this section.

## II. NETWORK-WIDE UTILITY OPTIMIZATION

The first formulation of network-wide utility optimization incorporates not only network-wide opportunistic scheduling and power control but also load balancing, *i.e.*, an MS can be associated with any BS in the network.

## A. Problem Formulation

Suppose that there are $N$ BSs and $K$ MSs in a multi-cell network and the sets of BSs and MSs are denoted by $\mathcal{N} \doteq \{1, \ldots, N\}$ and $\mathcal{K} \doteq \{1, \ldots, K\}$, respectively. Each BS chooses only one MS for each time-slot $t$ for downlink transmission of data, *i.e.*, there is a one-to-one mapping of $\mathcal{N}$ into $\mathcal{K}$. Network-wide optimization of generalized network-wide utility $U(t)$ at time-slot $t$ can be formulated as follows:

$$\max_{\{I_{kn}(t)\}} \max_{\{p_n(t)\}} U(t), \tag{1}$$

$$\text{s.t.} \quad U(t) = \sum_{k \in \mathcal{K}} U_k(\bar{r}_k(t)), \tag{2}$$

$$\bar{r}_k(t) = \frac{1}{t} \sum_{\tau=1}^{t} \sum_{n \in \mathcal{N}} r_{kn}(\tau) I_{kn}(\tau), \tag{3}$$

$$\sum_{k \in \mathcal{K}} I_{kn}(t) = 1, \tag{4}$$

$$\sum_{n \in \mathcal{N}} I_{kn}(t) \leq 1, \tag{5}$$

$$r_{kn}(t) = B \cdot \mathrm{C}\left( \frac{p_n(t)g_{kn}(t)}{\sigma_k^2 + \sum_{i \in \mathcal{N}, i \neq n} p_i(t)g_{ki}(t)} \right), \tag{6}$$

$$0 \leq p_n(t) \leq \overline{p}, \tag{7}$$

where $U_k(\bar{r}_k(t))$ is the utility function of $\bar{r}_k(t)$, the average throughput of MS $k$ up to time-slot $t$. It is reasonable to let $U_k(\cdot)$ be a function of only $\bar{r}_k(t)$ as in [8], [9], [22]. Besides, $\bar{r}_k(t)$ is a function of the instantaneous channel rates from BSs to MSs, $\{r_{kn}(\tau) \mid 1 \leq \tau \leq t\}$, and the scheduling indicator matrices, $\mathcal{I}(t) \doteq \{I_{kn}(\tau) \mid 1 \leq \tau \leq t\}$ where

$$I_{kn}(t) = \begin{cases} 1, & \text{MS } k \text{ is assigned to BS } n \text{ at time-slot } t, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

The two constraints regarding $I_{kn}(t)$, (4) and (5), imply that each BS serves exactly one MS and each MS can be served by at most one BS, respectively, at each time-slot. The average throughput of MS $k$ up to $t$, $\bar{r}_k(t)$, is the ensemble average of the instantaneous channel rates, $r_{kn}(t)$, which are determined by the channel capacity model (6) where $B$ is the bandwidth in hertz, $\mathrm{C}(\cdot)$ is the normalized channel capacity in bits/s/hertz that is a function of the instantaneous SINR. In the SINR expression, $g_{kn}(t) > 0$ is the signal gain from BS $n$ to MS $k$ that characterizes the propagation loss of the transmission power $p_n(t)$, the fast Rayleigh fading, and the large time-scale log-normal fading. Assuming that Gaussian signalling is used by all nodes, the interference terms are also Gaussian and the channel capacity within a given time-slot follows AWGN Shannon capacity. The transmission powers of BSs, $p_n(t)$, are upper-bounded by $\overline{p}$ at all time slots as shown in (7), but it can be easily verified that all results in this paper hold even if the upper-bounds vary time. Before taking up the main subject, it is noteworthy that network-wide optimization is reduced to $N$ independent intra-cell scheduling problems if there is neither load balancing nor power control. (For readability, all proofs in this paper are in Appendix.)

**Observation 1.** *If we assume that all BSs are transmitting with its full power and each MS is associated with a predetermined*

BS, then network-wide optimization given by (1)∼(8) is reduced to $N$ independent intra-cell optimizations.

This observation signifies that the existing intra-cell opportunistic scheduling is the best way to optimize the network-wide utility if load balancing is not used and transmission powers of BSs are constant, *e.g.*, $p_n(t) = \overline{p}$. It should be remarked that the intra-cell opportunistic scheduling possesses a desirable convergence property. That is, it is shown in [22]–[24] that its gradient algorithm leads to long-term asymptotic optimality under various assumptions.

The given optimization problem is very complicated due to the nonlinearities of (1) and (6), combination of binary variables $\{I_{kn}\}$ and real variables $\{p_n(t)\}$, and twofoldness in (1). But we have found that the problem becomes more tractable if it is separated into two subproblems: (i) *the outer problem* – maximizing $U(t)$ by varying $\{I_{kn}(t)\}$ for a given $\{p_n(t)\}$ , and (ii) *the inner problem* – maximizing $U(t)$ by varying $\{p_n(t)\}$ for a given $\{I_{kn}(t)\}$. Note that the outer problem does not have (6) as constraints because $\{p_n(t)\}$ is given. Similarly, the inner problem does not have (4) and (5) as its constraints. Since the inner problem has a special characteristic which is to be revealed in Section II-B and has made our analysis possible, we will focus only on the inner problem for the moment until a simplified optimization is reformulated in Section II-C.

To simplify the optimization problem by assuming $\{I_{kn}\}$ is predetermined, let us define $k^{(n,t)}$ as the index of the MS to be served by BS $n$ at time-slot $t$. More formally, the set of pairs $(n, k^{(n,t)})$ is be defined as:

$$\Omega(t) \doteq \left\{ (n, k) \,\middle|\, I_{kn}(t) = 1, \ k \in \mathcal{K} \text{ and } n \in \mathcal{N} \right\}.$$

Since a BS chooses only one MS for each time-slot, the cardinality of $\Omega(t)$ is $N$. Whenever we use the notation $k^{(n,t)}$ in this section, we are assuming that $\Omega(t)$ is determined in advance and $(n, k^{(n,t)}) \in \Omega(t)$. By denoting $\epsilon_t \doteq \frac{1}{t}$, we can rewrite (3) compactly in the form of the following running average equation:

$$\bar{r}_{k^{(n,t)}}(t) = \bar{r}_{k^{(n,t)}}(t-1) + \left[ r_{k^{(n,t)}n}(t) - \bar{r}_{k^{(n,t)}}(t-1) \right] \epsilon_t. \tag{9}$$

If the initial condition $U(0) = 0$ is used, the network-wide utility becomes:

$$\lim_{t \to \infty} U(t) = \lim_{t \to \infty} \sum_{\tau=1}^{t} \left[ U(\tau) - U(\tau-1) \right]. \tag{10}$$

Here $\Delta U(t) \doteq U(t) - U(t-1)$ corresponds to the utility difference that should be maximized at each time-slot $t$ without the knowledge of $r_{kn}(\tau)$ for $\tau > t$. If we make use of the Taylor expansion of $\Delta U(t)$, it becomes

$$\begin{aligned} \Delta U(t) &= \sum_{k \in \mathcal{K}} U_k(\bar{r}_k(t)) - U_k(\bar{r}_k(t-1)) \\ &= \sum_{k \in \mathcal{K}} \frac{\partial U_k(\bar{r}_k(t-1))}{\partial \bar{r}_k} [\bar{r}_k(t) - \bar{r}_k(t-1)] + O(\epsilon_t^2) \\ &= \sum_{k \in \mathcal{K}} \frac{\partial U_k(\bar{r}_k(t-1))}{\partial \bar{r}_k} [r_k(t) - \underbrace{r_k(t-1)}_{\text{given at } t}] \epsilon_t + O(\epsilon_t^2) \quad (11) \end{aligned}$$

where the third equality holds by (9). Note that $\bar{r}_k(t-1)$ has no effect on the maximization of $\Delta U(t)$ since it is given at time-slot $t$. We adopt the generalized proportional fair utility function introduced in [11] where $U_k(\bar{r}_k(t))$ is given by:

$$U_k(\bar{r}_k(t)) = \begin{cases} w_k \frac{\bar{r}_k(t)^{1-\alpha}}{1-\alpha}, & \alpha \geq 0,\ \alpha \neq 1 \text{ and } w_k > 0, \\ w_k \log \bar{r}_k(t), & \alpha = 1 \text{ and } w_k > 0. \end{cases} \tag{12}$$

which is called $(w, \alpha)$-proportional fair utility. For $\alpha = 1$, $U_k(\bar{r}_k(t)) = w_k \log \bar{r}_k(t)$ is called proportional fair utility function. If we apply this substitution to (11) and neglect $O(\epsilon_t^2)$, maximizing $\Delta U(t)$ becomes equivalent to the following simple form:

$$\max_{\Omega(t)} \max_{\{p_n(t)\}} \sum_{n \in \mathcal{N}} w_{k(n,t)} \frac{r_{k(n,t)n}(t)}{[\bar{r}_{k(n,t)n}(t-1)]^\alpha}, \tag{13}$$

which bears a close resemblance to the Weighted Alpha-Rule scheduling algorithm in [8]. Note that the utility difference $\Delta U(t)$ is a linear combination of instantaneous channel rates $r_{k(n,t)n}(t)$ for given weights $\frac{w_{k(n,t)n}}{[\bar{r}_{k(n,t)n}(t-1)]^\alpha}$.

It turns out that the form of the objective function given by (13) can be viewed as a gradient algorithm adopted in [22]–[24]. The only difference here is that (13) requires the two-dimensional network-wide selection of an index set, $\Omega(t)$, and transmission powers of BSs, $p_n(t)$, while [22]–[24] requires $N$ independent cell-wide selections of one MS in each cell. Note also that, to the best of our knowledge, the gradient algorithm represented by (13) is the only way to exploit multiuser diversity *simultaneously with* keeping full control of network-wide fairness through the flexible control knob $\alpha$ in multi-cell networks.

### B. On the Optimality of Binary Power Control

For a given $\Omega(t)$, the inner problem is equivalent to a problem of assigning transmission powers to BSs. To give a shape to the instantaneous channel rates $r_{kn}(t)$ which depend on normalized channel capacity function $C(\cdot)$, we consider the following channel capacity model.

- $C(x) = \frac{x}{\log 2}$: Since $\log_2(1+x) \approx \frac{x}{\log 2}$ for $x \approx 0$, this model closely approximates Shannon's channel capacity if SINR is not too large. Particularly, this model is well suited for wide-band systems [15], [16] (where $B$ is large) and low-power systems (where $\overline{p}$ is small) for which Shannon's channel capacity becomes more linear.

Even with linear channel capacity model, (6) still complicates the analysis due to its nonlinear relations to transmission powers of other BSs. However, we have found that the inner problem with the linear channel capacity model has an interesting property described by Theorem 1. To prove it, we need a proposition showing an advanced convexity of a certain function.

**Proposition 1.** *Let us assume that $a > 0$, $a_i \geq 0$, $b_i > 0$, $\forall i = 1, \ldots, M$. Consider the following problem:*

$$\max_x \ ax + \sum_{i=1}^{M} \frac{a_i}{b_i + x},$$
$$s.t. \ 0 \leq x \leq \overline{x}.$$

*The above objective is strongly quasiconvex [25], and maximized at either $x = 0$ or $x = \overline{x}$.*

**Theorem 1 (Binary Power Control).** *For given $\Omega(t)$, if $C(x) = \frac{x}{\log 2}$ and the error term $O(\epsilon_t^2)$ in (11) is neglected[§], the network-wide utility*

$$U(t) = \begin{cases} \sum_{k \in \mathcal{K}} w_k \frac{\bar{r}_k(t)^{1-\alpha}}{1-\alpha}, & \alpha \geq 0,\ \alpha \neq 1 \text{ and } w_k > 0, \\ \sum_{k \in \mathcal{K}} w_k \log \bar{r}_k(t), & \alpha = 1 \text{ and } w_k > 0. \end{cases} \tag{14}$$

*cannot be maximized if $p_n(t) \notin \{0, \overline{p}\}$ for any $n \in \mathcal{N}$.*

Considering a downlink CDMA data network and a multi-hop wireless network, Radunović *et al.* [26] and Bedekar *et al.* [27] respectively obtained a similar result for $\alpha = 0$ and the channel capacity model is linear, which coincides with a part of Theorem 1. Recently, Gjendemsjø *et al.* [28] generalized this result and proved that binary power control is optimal for $\alpha = 0$ and identical $w_i$ not only if SINR factors are very low but also if they are highly concentrated. However, little can be said for the case $\alpha \neq 0$ because it can be easily shown that the arithmetic mean-geometric mean approximation used in [28] depends on time-varying factors (weights in (13)) in this case. The main point may be summarized as follows.

- For all types of network-wide $(w, \alpha)$-proportional fairness, *only binary power control is optimal* if the underlying physical layer is in the linear regime.

It is worth noticing that multi-carrier systems employing binary power control can suffer from high PAPR (peak-to-average power ratio). A number of PAPR reduction techniques have been proposed and it is well-known that there is a tradeoff between data rate and PAPR reduction [29]. Therefore, the net effect of binary power control considering its adverse effect should be investigated so that we can decide whether to employ binary power control or not in a specific system. In this paper, we propose a nonnegative lower-bound for the transmission powers of BSs, $\underline{p} \geq 0$, so that transmission powers can be either $\underline{p}$ or $\overline{p}$. Then $\underline{p}$ can serve as a flexible control knob to balance the favorable and adverse effects of binary power control. Note that it is easy to show that Theorem 1 also holds for $\underline{p} \neq 0$.

### C. A Simplified Formulation

The original optimization problem given in Section II-A is extremely complicated and we are going to reduce its complexity by simplifying the power assignment: we assume that $p_n(t) \in \{0, \overline{p}\}$ by appealing to the fact that binary power control is optimal for linear channel capacity model though it is not true when the physical layer regime is nonlinear[¶]. In other words, if we revisit the original formulation laid out in (1)~(8), its complexity is alleviated due to the property of the inner problem revealed in Section II-B. The essential point in

---

[§]If $\alpha = 0$, the error term $O(\epsilon_t^2)$ in (11) becomes 0. Thus there is no approximation in this case.

[¶]If the channel capacity is logarithmic in the SINR, it can be proven that binary power control is optimal **if and only if** $\frac{w_k}{[\bar{r}_k(t-1)]^\alpha} = \frac{w_j}{[\bar{r}_j(t-1)]^\alpha}$ for all $k, j \in \mathcal{K}$ and there are at most two BSs in a wireless network. The "if" part of this proposition can be proven by applying techniques used in [28], [30] to $\Delta U(t)$. The other part can be easily proven by counterexamples.

the following formulation is that *network-wide opportunistic power control assimilates into network-wide opportunistic scheduling*.

$$\max_{\{\tilde{I}_{kn}(t)\}} \Delta U(t), \tag{15}$$

$$\text{s.t. } r_k(t) = B \sum_{n \in \mathcal{N}} C \left( \frac{g_{kn}(t)\tilde{I}_{kn}(t)}{\frac{\sigma_k^2}{\overline{p}} + \sum_{\substack{i \in \mathcal{N} \\ i \neq n}} g_{ki}(t)\tilde{I}_{\cdot i}(t)} \right), \tag{16}$$

$$\tilde{I}_{\cdot n}(t) \doteq \sum_{k \in \mathcal{K}} \tilde{I}_{kn}(t) \leq 1, \tag{17}$$

$$\sum_{n \in \mathcal{N}} \tilde{I}_{kn}(t) \leq 1. \tag{18}$$

Here the instantaneous channel rate $r_k(t)$ does not have the subscript regarding BSs anymore. We introduce *combined* scheduling indicator matrix, $\widetilde{\mathcal{I}}(t) = \{\tilde{I}_{kn}(t)\}$ where $\tilde{I}_{kn}(t) \doteq I_{kn}(t) \cdot \frac{p_n(t)}{\overline{p}}$. The binary control assumption, *i.e.*, $p_n(t) \in \{0, \overline{p}\}$, keeps $\tilde{I}_{kn}(t) \in \{0, 1\}$ satisfied. The constraint (17) allows the inequality that was not allowed in (4), because a BS with zero power does not serve any MS, *i.e.*, $\sum_{k \in \mathcal{K}} \tilde{I}_{kn}(t) = 0$, where $n$ is the index of the BS. It is remarkable that the mixed, twofold and nonlinear optimization (1) is now unfolded into a *flat binary optimization* (15) where the only variables to be optimized are $\{\tilde{I}_{kn}(t)\}$. The network-wide utility can be written as:

$$\Delta U(t) = \sum_{k \in \mathcal{K}} \frac{w_k r_k(t)}{[\bar{r}_k(t-1)]^\alpha}. \tag{19}$$

Note that this objective takes an opportunistic scheduler form that opportunistically picks out MSs having relatively high $r_k(t)$ at time-slot $t$ if $g_{kn}(t)$ is fluctuating.

### D. On the Complexity of the Simplified Formulation

To better understand the complexity of the simplified problem presented in Section II-C, we will derive the exact number of possible ways satisfying the two inequalities, (17) and (18). This will show that the simplified problem is still complex and it is necessary to develop a distributed heuristic algorithm.

Let us consider the set of all possible scheduler indicator matrices, $\mathbf{I} \doteq \{\widetilde{\mathcal{I}}(t)\}$ and denote its cardinality by $\varphi(N, K)$. Since (17) and (18) allow both equality and inequality, it is clear that $\varphi(N, K)$ is bigger than the cardinality of the set of the indicator matrices introduced in [8, Section 4.1], where power assignment problem was not considered at all.

There are $_N\mathbf{C}_i$ possible different ways to choose $i$ BSs among $N$ BSs. Let us assume that the transmission powers of $i$ BSs are $\overline{p}$ and those of $N - i$ BSs are 0, without loss of generality. For each of $_N\mathbf{C}_i$ ways, there exist $_K\mathbf{P}_i$ ways to associate $i$ BSs to $K$ MSs. We provide a more tractable expression of $\varphi(N, K)$ in the following theorem with the help of the *Rodrigues representation* of the *associated Laguerre polynomials* [31]:
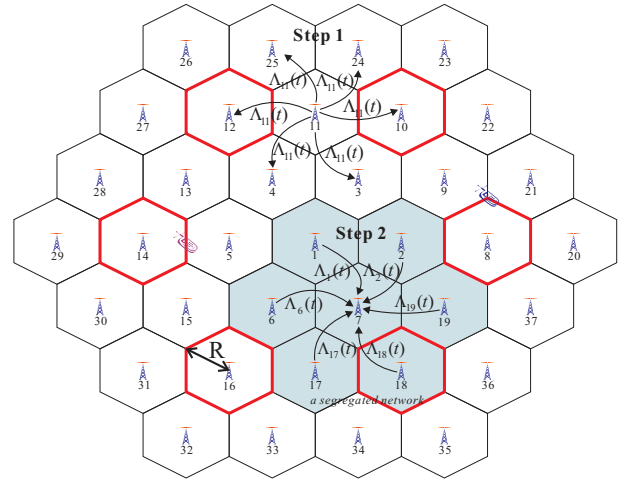


Fig. 1.   A multi-cell network composed of 37 hexagonal cells in four tiers.

**Theorem 2.** *Assuming that the inequality constraints (17) and (18) hold, the total number of possible ways to associate BSs with MSs is*

$$\varphi(N, K) = \left| \frac{d^N}{dx^N} \left( e^{-x^{-1}} x^K \right) \right|_{x=-1}. \tag{20}$$

We can see from (26) that $\varphi(N, K)$ is larger than $_K\mathbf{P}_N$, which is the number of possible ways satisfying the constraints in [8, Section 4.1]. Moreover, (20) shows that a centralized optimization cannot be deployed in practical systems. For instance, $\varphi(37, 740) \approx 6.1 \times 10^{105}$.

### III. A DISTRIBUTED HEURISTIC ALGORITHM

The intractable complexity of the network-wide optimization formulated by (15)∼(19) makes us feel keenly the necessity of a distributed heuristic algorithm. First of all, we assume that each MS can be associated only with the BS that is closest to the MS. By assuming this, we are transferring the task of load balancing to a centralized entity that should be capable of transacting a huge amount of computation. Therefore, our heuristic algorithm tries to approximate *only* network-wide opportunistic scheduling that has already absorbed network-wide opportunistic power control in Section II-C. If we divide the set of MSs $\mathcal{K}$ into $N$ disjoint sets as we did in the proof of Observation 1, we get a new constraint, $\tilde{I}_{kn}(t) = 0$ if $k \notin \mathcal{K}_n$, which greatly simplifies the optimization. However, the number of possible ways satisfying the constraints on $\{\tilde{I}_{kn}(t)\}$ is $\varphi'(N, |\mathcal{K}_i|) \doteq \prod_{i \in \mathcal{N}}(|\mathcal{K}_i| + 1)$, which is still too large. For example, if there are 37 BSs and each BS has 20 MSs, $\varphi'(37, 20) \approx 8.4 \times 10^{48}$.

To reduce the complexity of the simplified problem, we propose a heuristic algorithm named Neighbor-Assisted Network-wide Opportunistic Scheduling algorithm (NANOS), which is simpler in two regards. Each BS considers only its neighbor BSs. Furthermore, only a subset of MSs in its neighbors are considered instead of all MSs in the proposed heuristic. For example, BS 7 makes use of information from BSs 1, 2, 6, 19, 17 and 18 in Fig. 1. The heuristic consists of three steps.

**Step 1. Local selection of MS candidates**: Each BS $n$ makes a list $\Lambda_n$ of candidates that might possibly maximize the sum of utilities of BS $n$. We first select a candidate that maximizes the sum of utilities when all of its neighbors are transmitting at full power $\bar{p}$. We repeat the selection process for the cases in which only one (or none) of its neighboring BSs in ($=\widetilde{\mathcal{N}}_n - \{n\}$) is OFF. In the example of Fig. 1, there can be up to 7 possible candidates if each candidate from the above cases is distinct from the rest. In other words,

$$\Lambda_n(t) \doteq \{\kappa_n^j(t) \mid \kappa_n^j(t) = \arg\max_{k\in\mathcal{K}_n} \Delta U_k^j(t), \; j \in \widetilde{\mathcal{N}}_n\}, \quad (21)$$

where the per-MS utility $\Delta U_k^j(t)$ is defined as:

$$\Delta U_k^j(t) \doteq \frac{w_k}{[\bar{r}_k(t-1)]^\alpha} \cdot \mathrm{C}\left(\frac{g_{kn}(t)}{\frac{\sigma_k^2}{\bar{p}} + \Psi_{kn}^j(t)}\right). \quad (22)$$

Here, the other-cell interference is defined as:

$$\Psi_{kn}^j(t) \doteq \begin{cases} \sum_{i\in\mathcal{N}, i\neq n} g_{ki}(t), & \text{if } j = n \text{ (All BSs are ON)};\\ \sum_{i\in\mathcal{N}, i\neq n, i\neq j} g_{ki}(t), & \text{otherwise (Only } j \text{ is OFF).} \end{cases} \quad (23)$$

The first case in which $j = n$ is when all neighbor BSs are transmitting while the other case is when one of neighbor interference terms, $g_{kj}(t)$, is excluded from the normalized interference sum of MS $k$. Therefore, $\kappa_n^j(t)$ corresponds to the MS index which maximizes (19) if neighbor BS $j$ is turned off. Similarly, $\kappa_n^n(t)$ is obtained if no neighbor BS is turned off. In brief, Step 1 serves to decrease the number of MSs to be considered in Step 3.

**Step 2. Broadcasting of list**: Each BS $n$ broadcasts the candidate list $\Lambda_n(t)$ to its neighbor BSs ($=\widetilde{\mathcal{N}}_n - \{n\}$) through the backhaul network.

**Step 3. Finding a locally optimal solution**: In this step, each BS $n$ finds a local optimal solution by solving a problem composed of the neighbor BS set $\widetilde{\mathcal{N}}_n$ and MSs $\widetilde{\mathcal{K}}_n \doteq \bigcup_{n\in\widetilde{\mathcal{N}}_n} \Lambda_n(t)^\|$. The local optimization at each BS $n$ is solved as if the BS set $\mathcal{N}$ and the MS set $\mathcal{K}$ were replaced by $\widetilde{\mathcal{N}}_n$ and $\widetilde{\mathcal{K}}_n$, respectively. Note that the cardinality of index set $\Lambda_n(t)$ is less than or equal to $\min\left[|\mathcal{K}_n|, |\widetilde{\mathcal{N}}_n|\right]$. In addition to this, since it is likely that $\kappa_n^j(t)$ are repeated for different $j'$, it can be even smaller. Once a local optimal solution $\{\tilde{I}_{km}^*(t)|k \in \bigcup_{n\in\widetilde{\mathcal{N}}_n}\Lambda_n(t), m\in\widetilde{\mathcal{N}}_n\}$ is determined, BS $n$ serves an MS whose index $k$ satisfies $\tilde{I}_{kn}^*(t) = 1$. Note that BS $n$ may not transmit at all if $\tilde{I}_{kn}^*(t) = 0$ for all $k \in \Lambda_n(t)$. Even though BS $n$ solves the problem including neighboring BSs, it only uses a subset of the solution and disregards all solutions of neighbor BSs.

The MS selection algorithm is based on the following observations. Firstly, if the set $\Lambda_n(t)$ is decided under the assumption that all BSs transmit at full power, MSs located around cell boundaries are rarely selected since $r_k(t)$ of them are underestimated due to interference signal from neighbor BSs. That is, MSs that might be served more frequently after coordinating transmission powers of BSs are discriminated against. This certainly diminishes the network-wide utility.

$^\|$We assume that BSs not in $\widetilde{\mathcal{N}}_n$ are turned on.

Secondly, the situations in which two consecutive BSs do not transmit anything rarely coincide with the optimal power assignments**. For instance, consider an MS in the middle of the two BSs as shown in Fig. 1, where the strongest interferers of the MS correspond to BS 5 and 14 due to their closeness to the MS. If BS 5 serves the MS and BS 14 does not transmit, the instantaneous channel rate of the MS becomes quite large because the main interferer is removed. Therefore, it is more plausible that the network-wide utility is increased if one of them transmits data at full power. Thus it is imperative that we should introduce $\Psi_{kn}^j(t)$.

Even though the proposed heuristic is simpler than the optimal algorithm, we should admit there are many challenges in implementing it. For example, if $K$ is sufficiently large, $|\Lambda_n(t)|$ approaches $|\widetilde{\mathcal{N}}_n|$ and the complexity of an exhaustive search becomes $(|\widetilde{\mathcal{N}}_n|+1)^{|\widetilde{\mathcal{N}}_n|} \approx 2\times10^6$ with $|\widetilde{\mathcal{N}}_n| = 7$. However, it is *virtually impossible to finish the global optimization* laid out in (15)$\sim$(19) even if there is no load balancing, due to its extreme complexity, *e.g.*, $\varphi'(37,2) \approx 4.5\times10^{18}$, $\varphi'(37,4) \approx 7.3\times10^{25}$ and $\varphi'(37,20) \approx 8.4\times10^{48}$. As of now, we claim that our heuristic is the only possible way to evaluate the performance of network-wide opportunistic scheduling.

We see that there are many possible ways to decrease the complexity of the proposed heuristic algorithm. For example, let us assume that a cell can be divided into 3 or 6 sectors such that each of the sectors is allocated $1/3$ or $1/6$ of the total bandwidth. Then, a sector needs to consider 2 or 1 adjoining sectors, respectively. In this case, the complexity of the proposed algorithm for each cell becomes $(|\widetilde{\mathcal{N}}_n|+1)^{|\widetilde{\mathcal{N}}_n|}$ = 9 or 2, respectively. However, as this sectorization scheme leads to less optimal results, we decided not to adopt this scheme.

Finally, we note that the proposed algorithm necessitates additional measurements in the network. Provided that the approximate values of $\sigma_k^2$ and $\overline{p}$ are known to MSs, each MS should be able to measure or infer the values of $g_{ki}(t)$ where $k \in \mathcal{K}_n$ and $i \in \widetilde{\mathcal{N}}_n$. These values should be regularly delivered to BSs so that they can solve the optimization problems laid out in Step 1 and 3.

## IV. SIMULATION RESULTS

To evaluate the performance of opportunistic scheduling algorithms in multi-cell networks, various scenarios having different fairness objectives are considered by varying $\alpha$. The weights of MSs are assumed to be equal, *i.e.*, $w_i = 1$. A four-tier multi-cell network composed of $37(= N)$ hexagonal cells shown in Fig. 1 is adopted, where each cell is circumscribed by a circle of radius $R = 1$ km. The number of MSs, $K$, is varied to change MS density in the network. Since we assume the uniformity of MS location distribution in Section IV-A, the probability distribution of the number of MSs associated with BS $i$ is given by Prob$\{|\mathcal{K}_i| = k\} = {}_K\mathrm{C}_k \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{K-k}$. Nonuniform MS distribution is considered in Section IV-B. For

**For simplicity, our algorithm excludes situations in which two consecutive BSs are turned off. For instance, if an MS is located at the vertex where hexagonal cells of BS 8, 9 and 21 meets as shown in Fig. 1, there is some possibility of improving the network-wide utility by turning off two of them. As $\alpha$ increases, this possibility becomes higher.

(a) Network-wide utility difference.



(b) Cell throughput.



(c) Sum of MS throughputs below the 5<sup>th</sup> percentile.
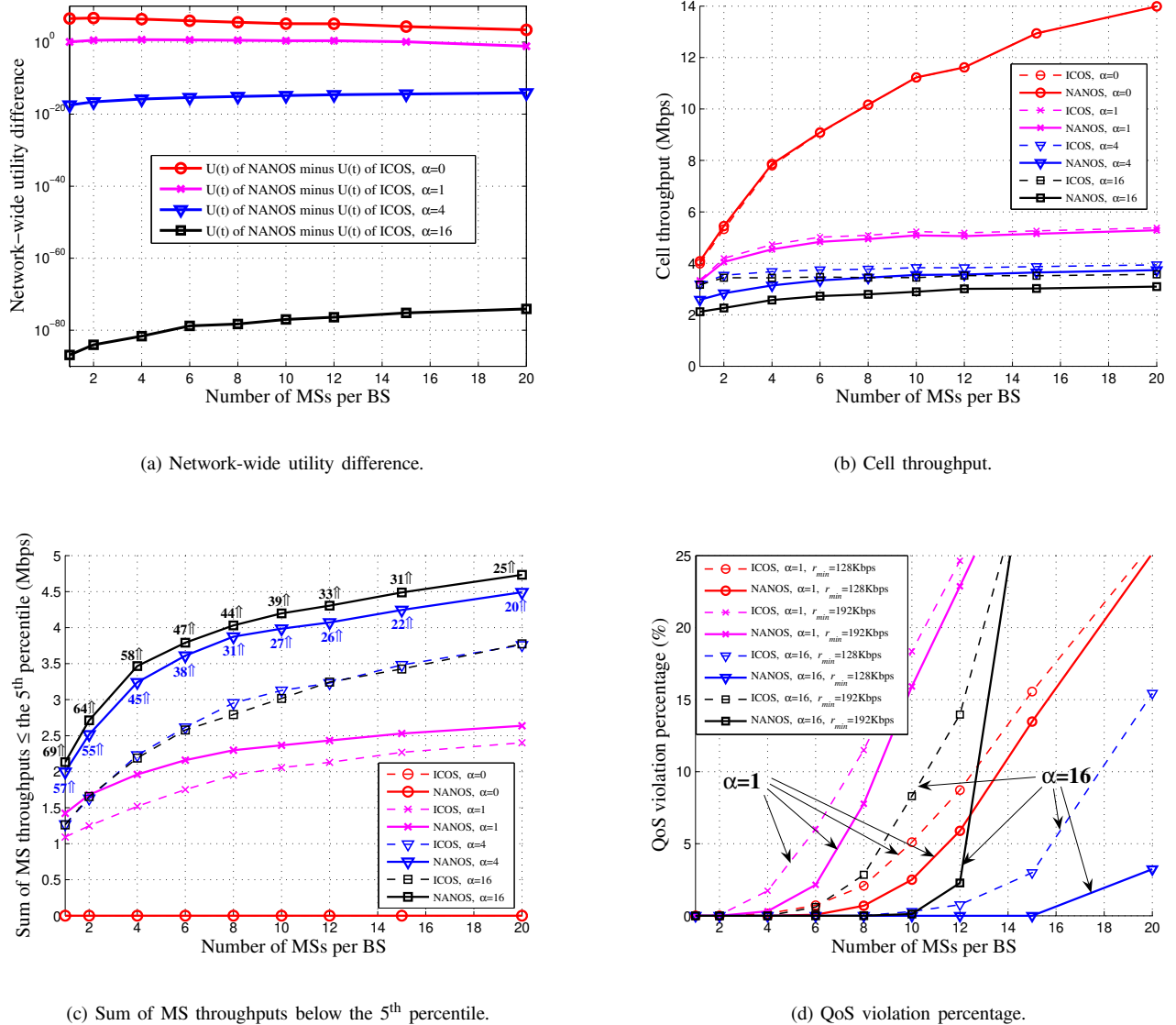


(d) QoS violation percentage.

Fig. 2. Utility, throughput and QoS performance of network-wide and intra-cell opportunistic scheduling algorithms.

each given parameter set, we execute 50∼100 runs where each run has 4,000∼16,000 time-slots and MSs are located over the network at the beginning of each run. Intra-cell opportunistic scheduling (which we denote by ICOS) and neighbor-assisted network-wide opportunistic scheduling (which we denote by NANOS), are evaluated under the same configuration.

Regarding the physical model, $B = 1.5\text{MHz}$ and $\text{C}(x) = \log_2(1 + x)$ are used. All BSs have the same maximum transmission power $\overline{p} = 40\text{dBm}$ when they transmit data and the same thermal noise $\sigma_k^2 = -100\text{dBm}$ which represents 3G networks. The signal gain from BS $n$ to MS $k$ is defined as $g_{kn}(t) \doteq \min\{1, s_{kn}(t) \cdot \Gamma(d_{kn})\}$ where $s_{kn}(t)$ is a unit-mean log-normal fading variable such that $10\log_{10} s_{kn}(t)$ is a memoryless Gaussian random variable with a standard deviation $\sigma_s = 4\text{dB}$ and $\Gamma(d_{kn}) = -130 - 35\log_{10}(d_{kn})$ dB ($d_{kn}$ is the distance from BS $n$ to MS $k$ in kilometers). Note that this corresponds to a path loss exponent of 3.5. The NANOS

parameter $\widetilde{\mathcal{N}}_n$ is defined as $\widetilde{\mathcal{N}}_n \doteq \{i \in \mathcal{N} \mid d'_{in} \leq 2R \cdot \cos\frac{\pi}{6}\}$ ($d'_{in}$ is the distance between BS $i$ and $n$) so that each BS located in the interior of the network has 6 neighbor BSs excluding itself and $|\widetilde{\mathcal{N}}_n| = 7$.

### A. Utility, Throughput, QoS and Fairness Performance

It is our understanding that the network-wide utility $U(t)$ itself is not meaningful since there is no direct relevance between $U(t)$ and performance indices. For instance, performance indices of NANOS with $\alpha > 1$ are greatly improved even with slight increment of $U(t)$. However, to make it clear that NANOS is superior to ICOS, we present Fig. 2(a) where $\overline{r}_k(t)$ in bits per seconds (bps) is used for calculating $U(t)$. NANOS *does* result in higher $U(t)$ than ICOS for all combinations of $K/N$ and $\alpha$.
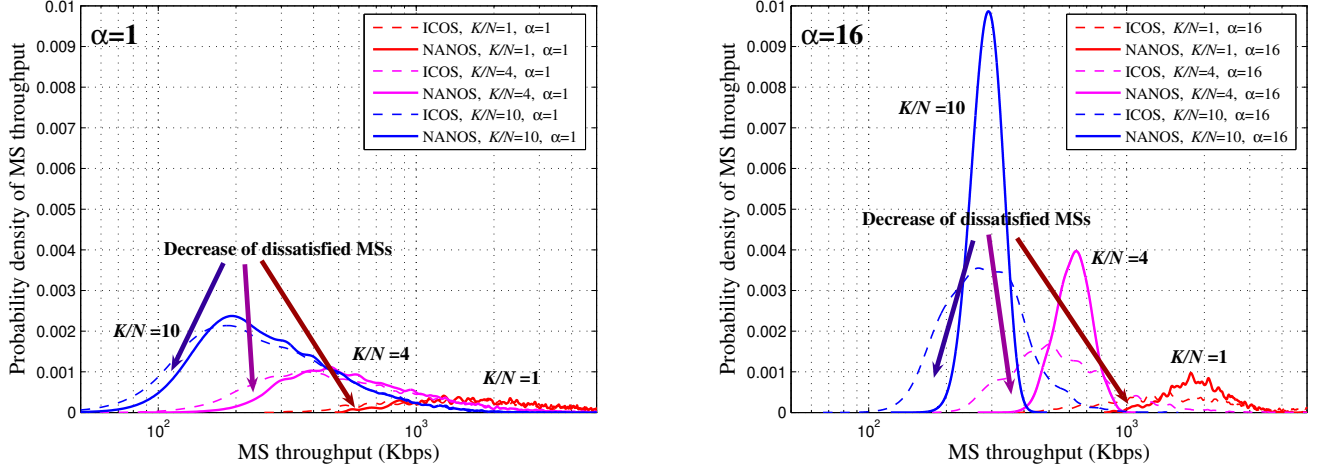
Fig. 3. Probability densities of MS throughput of network-wide and intra-cell opportunistic algorithms.

**Observation 2.** *If the network objective is throughput maximization ($\alpha = 0$) and the number of MSs is not too small, network-wide optimization does not help.*

Observe that in Fig. 2(b) that the two topmost lines of ICOS and NANOS are overlapping when $\alpha = 0$. To maximize throughput, an MS with the best channel condition would be chosen in each BS in ICOS. Since the MS with the best channel is likely to be close to the BS, the impact of other-cell interference may be negligible, which is why the cell throughputs of NANOS and ICOS are almost the same.

**Observation 3.** *As a network pursues fairness-oriented network objectives ($\alpha > 0$), NANOS results in a bit lower total throughput than ICOS. However, cell boundary MSs get more throughput.*

In Fig. 2(b), when $\alpha=1$, 4 and 16, NANOS results in slightly lower throughput than those of ICOS. When $\alpha > 0$, there is no plausible reason for NANOS to achieve higher throughput than ICOS does. The network just does its own duty to maximize its utility in each cell (ICOS) or in a network-wide way (NANOS). As the network pursues fairness-oriented utilities with larger $\alpha$, the throughput gap between two scheduling algorithms becomes more obvious because a network operated by NANOS algorithm gradually approaches a max-min fair network where all MSs have the same throughput.

To observe benefits of NANOS that are received by cell boundaries MSs, let us consider the sum of MS throughputs not exceeding the $5^{\text{th}}$ percentile[††]of MS throughput distribution which is shown in Fig. 2(c). It is noteworthy that throughput performance of dissatisfied users in NANOS is significantly improved (Fig. 2(c)) whereas the average throughput performance of NANOS is similar to that of ICOS (Fig. 2(b)). For example, when $\alpha = 4$ and the number of MSs is six, NANOS is 38% better than ICOS. The throughput increases in percentage for $\alpha = 4$ and $\alpha = 16$ are written down in

Fig. 2(c). Anther noteworthy point is that the throughput gap between two scheduling algorithms gets smaller as the number of MSs becomes larger. This phenomenon is due to the fact that the oversatisfied users are more reluctant to return their spare utilities which might be used to increase utilities of dissatisfied users as the network gets more loaded.

**Observation 4.** *Throughput performance of ICOS is saturated for large $\alpha$ while NANOS mitigates this saturation.*

If we look into Fig. 2(b) again, the cell throughputs of ICOS are almost the same for $\alpha = 4$ and $\alpha = 16$, but on the other hand those of ICOS become smaller as a network pursues fairness-oriented objectives. In other words, for $\alpha \geq 4$, ICOS *cannot* increases its network-wide utility even if it reduces its cell throughput while NANOS *does* increase its network-wide utility at the sacrifice of its cell throughput. Note also that the throughput performance of dissatisfied users shown in Fig. 2(c) *does not* improve for ICOS and *does* improve for NANOS. From these results, we can conclude that generalized proportional fair scheduling without inter-cell interference coordination becomes meaningless as $\alpha$ gets larger ($\alpha > 1$).

To strengthen our claim that network-wide opportunistic scheduling with a large value of $\alpha$ helps cell boundaries MSs, let us compute the QoS violation probability with a minimum-throughput condition, $\bar{r}_k(t) > r_{min}$. In Fig. 2(d), we show the percentages of MSs whose rate is lower than the threshold for $r_{min} = 128$ and 192 Kbps. It is an expected result that the violation ratio decreases with a larger $\alpha$. But it is noteworthy that the violation percentage of NANOS is much lower than that of ICOS for large $\alpha$ ($\alpha = 16$). This implies that network-wide opportunistic scheduling pertains to fairness-oriented network where users demands their average throughput to be larger than a certain threshold.

NANOS with large $\alpha$ enables MSs in the network to receive benefits of BS coordination equally while ICOS fails to do that even with large $\alpha$. If we look into Fig. 3, where probability densities of MS throughput for $\alpha = 1$ and $\alpha = 16$ are shown
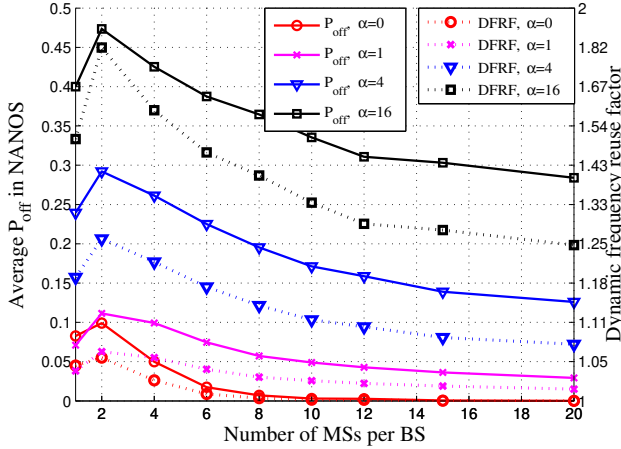
---

[††]The $p^{\text{th}}$ percentile cuts off lowest $p\%$ data.

Fig. 4. Average BS off probabilities and frequency reuse factor in NANOS.



(a) Without hot-spot cells (Uniform MS density).



(b) With hot-spot cells (Nonuniform MS density).

Fig. 5. Per-BS off probabilities in NANOS with and without hot-spot cells.

for various numbers of MSs per BS, such phenomenon is revealed more clearly. NANOS significantly increases throughputs of dissatisfied MSs while slightly sacrificing those of oversatisfied MSs in underloaded networks. Therefore, we conclude that NANOS gives preferential treatments to discontented MSs with large $\alpha$ by proving itself an able coordinator to better channel conditions around cell boundaries.

We can summarize the aforementioned results as follows: Network-wide opportunistic scheduling remarkably improves the throughput performance of users around cell boundaries by offering fairer chances to them while the aggregate throughput is maintained. Network-wide opportunistic scheduling is most suitable for fairness-oriented or underloaded networks.

Let us turn our attention on the probability that BS does not transmit data, *i.e.*, $P_{off} \doteq \text{Prob}\{p_n(t) = 0\}$. It is clear from Fig. 4 that NANOS controls BS powers more aggressively as $\alpha$ becomes larger and $K/N$ gets smaller except the case $K/N = 1$ where too many BSs ($\approx 0.363 \times 37$) having no MSs in their cells are excluded from the calculation of $\text{Prob}\{p_n(t) = 0\}$. Moreover, it should be remarked that $\text{DFRF} \doteq (1 - P_{off})^{-1}$ can be interpreted as *dynamic frequency reuse factor* where $\text{DFRF}^{-1} \cdot B$ represents the portion of bandwidth utilized for achieving the network objective. As shown in Fig. 4, DFRF increases with $\alpha$ and decreases with $K/N$.

### B. NANOS as an Implicit Load Balancer

Compared with existing frequency reuse schemes where all BSs should follow a global rule that specifies the usage of the whole bandwidth in the network and cannot be adapted dynamically, NANOS can cope with hot-spot scenarios in which several cells (called hot-spot cells) are more crowded than the other cells in the network.

To avoid repetitive figures and results, only the probability that each BS is turned off in NANOS is shown in Fig. 5 for two cases: (i) without hot-spot cells, and (ii) with hot-spot cells. The number of MSs per BS is 10 and $\alpha = 16$ in all cases. In Fig. 5(a), the MS density function in each BS is given by the same function used in Section IV-A. In Fig. 5(b), it is assumed that each of six hot-spot cells, *i.e.*, BS 8, 10, 12,
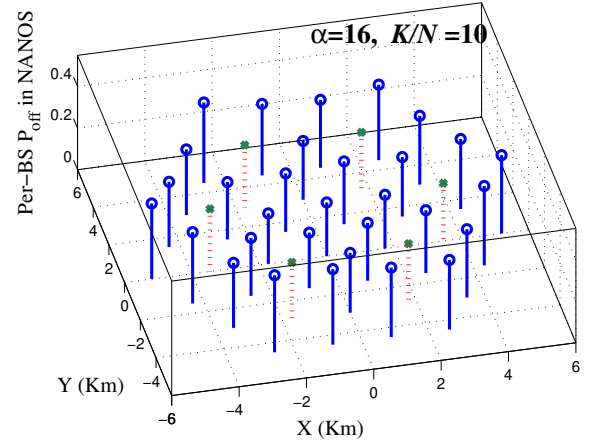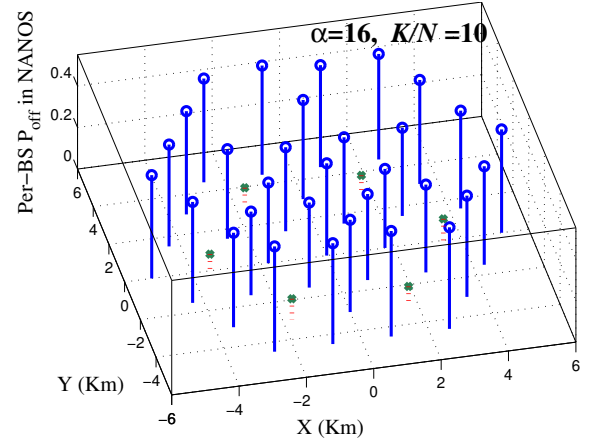
14, 16 and 18 in Fig. 1 has twice as many MSs as each of the other cells has in average while the total number of MSs in the two networks is 370 ($= 37 \times 10$) for all cases.

It is shown in Fig. 5(b) that NANOS with hot-spot cells provides more scheduling chances to MSs in hot-spot cells by preventing hot-spot cells from being turned off. That is, NANOS can be viewed as an *implicit load balancer* which improves the network-wide utilities of fairness-oriented networks through inter-cell power control. For reference, $P_{off}$ of hot-spot cells and the others are $0.090$ and $0.462$, respectively, as shown in Fig. 6. In addition, it is shown in Fig. 5(a) that NANOS differentiates $P_{off}$ of BSs even if there are no hot-spot cells. As outer cells are interfered with by less number of cells than inner cells are, there is an upward tendency of $P_{off}$ with distance from the center. As shown in Fig. 6, $P_{off}$ of BS 1 (tier 1), BS 2~7 (tier 2), BS 8~19 (tier 3), and BS 20~37 (tier 4) are $0.259$, $0.272$, $0.288$ and $0.346$, respectively.
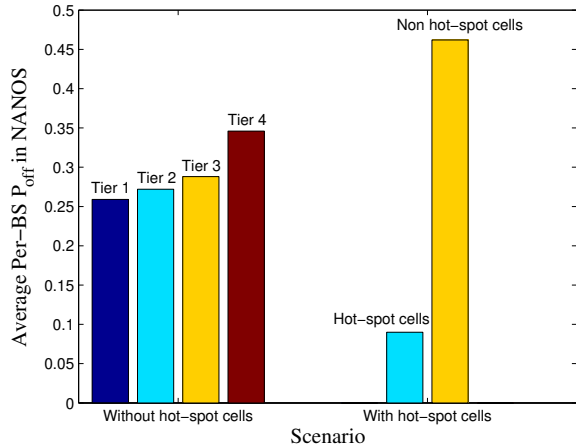
Fig. 6. Comparison of Per-BS off probabilities in NANOS.

### C. Discussion on Simulation Results

Unlike wired networks, wireless networks inclining to max-min fair bandwidth allocation suffer from solidarity property [4] that forces all users to have the same throughput. For instance, let us consider a wireless network where all BS transmit data all the time, *i.e.*, $p_n(t) = \overline{p}$, and MS $i \in \mathcal{K}_n$ achieves average throughput $\hat{r}_i$ when it monopolizes BS $n$. If we neglect the opportunistic gain caused by user diversity and denote the portion of time-slots allocated to MS $i$ by $\tau_i$, $\hat{r}_i$s should satisfy $\hat{r}_i \tau_i = \hat{r}_j \tau_j$ for any $i, j \in \mathcal{K}_n$, and $\sum_{i \in \mathcal{K}_n} \tau_i = 1$. In this case, it is easy to show that the actual average throughputs of all MSs are equalized as $\bar{r} = 1/\sum_{i \in \mathcal{K}_n} \hat{r}_i^{-1}$, which illustrates how max-min wireless networks suffer from their fairness objective. Thus it is important to improve $\hat{r}_i$s of the tail-enders when fairness is preferable to efficiency.

Besides, it sounds plausible that oversatisfied users are willing to return their utilities in lightly-loaded networks, whereas it does not make sense to divert utilities of more satisfied users to less satisfied users in heavily-loaded networks. That is, fairness is preferable to efficiency in lightly-loaded networks.

## V. CONCLUSION

Prospective high speed wireless networks [15], [16] are faced with wireless environments where downlink signals from different base stations are randomly interfering with each other while those from the same BS are orthogonal to each other. Aggregate throughput increases as wireless networks adopt more advanced wireless techniques and broader frequency bands, but on the other hand users at cell boundaries are still suffering from the inter-cell interference problem. Therefore, it is imperative that a sophisticated inter-cell scheduling scheme should provide satisfactory throughput to users who have been treated unfairly for their disadvantageous positions. It is also remarkable that there is a consensus of opinion in the evolved UTRA [32] that cell throughput can be sacrificed for users at cell boundaries not to mention that several inter-cell interference mitigation schemes are in the design stage.

Our contribution is four-fold: Firstly, to the best of our knowledge, this is the first work dealing with an analytical framework that optimizes generalized utilities of multi-cell networks through network-wide opportunistic scheduling and power control. Secondly, we have shown that the optimization can be simplified because network-wide opportunistic power control can be absorbed into the network-wide opportunistic scheduling when the underlying physical layer is in the linear regime. Thirdly, we have proposed a heuristic algorithm which makes distributed and simplified operations of base stations possible. Fourthly, we have obtained remarkable findings which answer the questions brought up in Section I by simulating various scenarios. In particular, network-wide opportunistic scheduling brings on significant performance improvement to discontented mobile stations at cell boundaries and is most suitable for

1) networks having a preference for fairness to efficiency because users demand their throughputs to be larger than a certain threshold and
2) underloaded networks where excessively satisfied users do not mind reducing their utilities.

It should be remarked that the generalized proportional fair scheduling, which covers most well-known fairness concepts, becomes *meaningless* for networks pursuing fairness criteria fairer than proportional fairness, *i.e.*, $\alpha > 1$, if there is no interference coordination between neighboring base stations. Our last finding is that network-wide opportunistic scheduling is capable of elevating fairness in multi-cell networks with hot-spots while existing frequency reuse schemes are useless for localized overloading.

Though we have barely managed to finish evaluating performance of network-wide opportunistic scheduling by our heuristic algorithm, it is necessary to develop an advanced algorithm which has lower computational complexity and results in better performance.

## APPENDIX: PROOFS

### A. Proof of Observation 1

If there is no load balancing, the set of MSs, $\mathcal{K}$, is divided into $N$ disjoint sets, $\mathcal{K}_1, \ldots, \mathcal{K}_N$, such that

$$\bigcup_{i \in \mathcal{N}} \mathcal{K}_i = \mathcal{K} \quad \text{and} \quad \mathcal{K}_i \bigcap \mathcal{K}_j = \emptyset \text{ for } i \neq j,$$

which also implies that $I_{kn}(t) = 0$ if $k \notin \mathcal{K}_n$. Since this condition forces MSs to belong to predetermined BSs, (4) becomes equivalent to $\sum_{k \in \mathcal{K}_n} I_{kn}(t) = 1$, (5) becomes needless, and (3) is reduced to $\bar{r}_k(t) = \frac{1}{t} \sum_{\tau=1}^{t} r_{kn}(\tau) I_{kn}(\tau)$ where $k \in \mathcal{K}_n$. If transmission powers of BSs are not adjustable, it is easy to see that $\sum_{k \in \mathcal{K}_n} U_k(\bar{r}_k(t))$ depends only on $I_{kn}(t)$ where $k \in \mathcal{K}_n$. Therefore, each BS can maximize its own objective $\sum_{k \in \mathcal{K}_n} U_k(\bar{r}_k(t))$ independently.

### B. Proof of Proposition 1

Let us denote the given objective function by $f(x)$. If $\frac{df}{dx} = a - \sum_{i=1}^{M} \frac{a_i}{(b_i+x)^2} = 0$ holds, at least one $a_i$ should be positive by the assumption $a > 0$. Thus it is straightforward to see that

$$\frac{df}{dx} = 0 \quad \Longrightarrow \quad \frac{d^2 f}{dx^2} > 0. \tag{24}$$

This means that whenever the function $\frac{df}{dx}$ crosses the value 0, it is strictly increasing. Therefore it can cross the value 0 at most once. If $\frac{df}{dx}$ does not cross the value 0 at all, then $f(x)$ is either strictly increasing or strictly decreasing, *i.e.*, $\frac{df}{dx} > 0$, $\forall x \in [0, \overline{x}]$ or $\frac{df}{dx} < 0$, $\forall x \in [0, \overline{x}]$. Otherwise it must cross the value 0 exactly once, say at $\tilde{x} \in [0, \overline{x}]$. Since $\frac{d^2 f}{dx^2}(\tilde{x}) > 0$, it follows that $\frac{df}{dx} < 0$, $\forall x \in [0, \tilde{x})$, and $\frac{df}{dx} > 0$, $\forall x \in (\tilde{x}, \overline{x}]$. Thus, it can be easily shown that, for each *distinct* $x_1$, $x_2$ in the interval $[0, \overline{x}]$ with $\frac{df}{dx}(x_1) \cdot (x_2 - x_1) \geq 0$, we have $f(x_2) > f(x_1)$. Therefore, $f(x)$ is *strictly pseudoconvex* [25, pp. 113]. Since every differentiable strictly pseudoconvex function is *strongly quasiconvex* [25, pp. 112], $f(x)$ satisfies

$$f(x) < \max\{f(0), f(\overline{x})\}, \ \forall x \in (0, \overline{x}).$$

### C. Proof of Theorem 1

If the factor $\frac{1}{\log 2}$ in $C(x) = \frac{x}{\log 2}$ is ignored, the objective function in (13) becomes:

$$\sum_{n \in \mathcal{N}} \frac{Bw_{k(n,t)}}{[\overline{r}_{k(n,t)n}(t-1)]^\alpha} \cdot \frac{p_n(t)g_{k(n,t)n}(t)}{\sigma^2_{k(n,t)} + \sum_{\substack{i=1 \\ i \neq n}}^{N} p_i(t)g_{k(n,t)i}(t)}. \quad (25)$$

If we examine the above expression carefully, we can see that Proposition 1 is applicable to (25) with the following substitutions:

$$N \to M+1, \quad p_m(t) \to x, \quad \overline{p} \to \overline{x},$$

$$\frac{Bw_{k(m,t)}g_{k(m,t)m}(t)}{[\overline{r}_{k(m,t)m}(t-1)]^\alpha \left(\sigma^2_{k(m,t)} + \sum_{\substack{i=1 \\ i \neq m}}^{N} p_i(t)g_{k(m,t)i}(t)\right)} \to a$$

$$\frac{Bw_{k(n,t)}p_n(t)g_{k(n,t)n}(t)}{[\overline{r}_{k(n,t)n}(t-1)]^\alpha g_{k(m,t)m}(t)} \to a_n$$

$$\left(\sigma^2_{k(n,t)} + \sum_{\substack{i=1 \\ i \neq n, \ m}}^{N} p_i(t)g_{k(n,t)i}(t)\right) \bigg/ g_{k(m,t)m}(t) \to b_n.$$

Note that $p_m(t)$ appears at no constants other than variable $x$. In this way, we can show that, for each variable $p_m(t)$, the maximum of the objective function in (13) is achieved at $p_m(t) = 0$ or $p_m(t) = \overline{p}$. We can also see that the conclusion is independent of $\Omega(t)$. This completes the proof.

### D. Proof of Theorem 2

According to the arguments we provided in Section II-D, $\varphi(N, K)$ can be written as follows.

$$\varphi(N, K) = \sum_{i=0}^{N} {}_N\mathbf{C}_i \cdot {}_K\mathbf{P}_i = \sum_{i=0}^{N} {}_N\mathbf{C}_i \cdot {}_K\mathbf{P}_{N-i} \quad (26)$$

$$= \sum_{i=0}^{N} \frac{N! \cdot K!}{(N-i)! \cdot i! \cdot (K-N+i)!}$$

$$= N! \cdot L_N^{K-N}(-1) \quad (27)$$

where the equality in (27) holds by the definition of the associated Laguerre polynomial [31, pp. 832]:

$$L_n^k(x) = \sum_{i=0}^{n} \frac{(n+k)!}{(n-i)! \cdot (k+i)! \cdot i!}(-x)^i.$$

It is also shown in [31, pp. 833] that the Rodrigues representation of $L_n^k(x)$ can be obtained as

$$L_n^k(x) = \frac{e^x x^{-k}}{n!} \frac{d^n}{dx^n}\left(e^{-x} x^{k+n}\right).$$

Therefore, we can simplify (27) further as follows.

$$\begin{aligned} \varphi(N, K) &= e^{-1}(-1)^{-(K-N)} \frac{d^N}{dx^N}\left(e^{-x} x^K\right)\bigg|_{x=-1} \\ &= \left|\frac{d^N}{dx^N}\left(e^{-x-1} x^K\right)\right|_{x=-1}. \end{aligned}$$

### REFERENCES

[1] R. Knopp and P. A. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. IEEE ICC*, Seattle, WA, Jun. 1995, pp. 331–335.

[2] D. Bersekas and R. Gallager, *Data Networks*, 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall, 1992.

[3] F. Kelly, A. Maullo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, Jul. 1998.

[4] B. Radunović and J.-Y. LeBoudec, "Rate performance objectives of multihop wireless networks," *IEEE Trans. Mobile Comput.*, vol. 3, no. 4, pp. 334–349, Oct. 2004.

[5] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1277–1294, Jun. 2002.

[6] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *IEEE/ACM Trans. Netw.*, vol. 13, no. 3, pp. 636–647, Jun. 2005.

[7] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 2053–2064, Oct. 2001.

[8] A. Sang, X. Wang, M. Madihian, and R. D. Gitlin, "Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems," in *Proc. ACM MobiCom*, Philadelphia, PA, Sep. 2004, pp. 302–314.

[9] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *Proc. IEEE Infocom*, Barcelona, Spain, Apr. 2006.

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

[11] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[12] B. Kauffmann, F. Baccelli, A. Chaintreau, V. Mhatre, K. Papagiannaki, and C. Diot, "Measurement-based self organization of inter 802.11 wireless access networks," in *Proc. IEEE Infocom*, Anchorage, AK, May 2007.

[13] P. Brémaud, *Markov Chains, Gibbs Field, Monte Carlo Simulation and Queues*. Springer-Verlag, 1999.

[14] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Physical layer procedures (FDD) (Release 4), 3GPP Std. Tech. Rep. 25.214, Mar. 2003.

[15] Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems–Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, IEEE Std. 802.16e, 2005.

[16] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Feasibility study for evolved UTRA and UTRAN (Release 7), 3GPP Std. Tech. Rep. 25.912, Sep. 2006.

[17] Downlink inter-cell interference co-ordination/avoidance–evaluation of frequency reuse, 3GPP TSG-RAN WG1 Contribution R1-061374, 2006. [Online]. Available: http://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_45/Docs/R1-061374.zip

[18] S. G. Kiani, G. E. Øien, and D. Gesbert, "Maximizing multicell capacity using distributed power allocation and scheduling," in *Proc. IEEE WCNC*, Hong Kong, Mar. 2007.

[19] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun. Mag.*, vol. 13, pp. 56–61, Aug. 2006.

[20] G. Li and H. Liu, "Downlink radio resource allocation for multi-cell OFDMA system," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3451–3459, Dec. 2006.

[21] H. Liu and G. Li, *OFDM-Based Broadband Wireless Networks: Design and Optimization*. Wiley-Interscience, 2005.

[22] A. L. Stolyar, "On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation," *Oper. Res.*, vol. 53, no. 1, pp. 12–25, Jan. 2005.

[23] R. Agrawal and B. Subramanian, "Optimality of certain channel aware scheduling policies," in *Proc. 40th Annual Allerton Conf. Comm., Control, and Comput.*, Monticello, IL, Oct. 2002.

[24] H. J. Kushner and P. A. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1250–1259, Jul. 2004.

[25] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 2nd ed. John Wiley & Sons Inc., 1993.

[26] B. Radunović and J.-Y. LeBoudec, "Power control is not required for wireless networks in the linear regime," in *Proc. IEEE WoWMoM*, Taormina, Italy, Jun. 2005.

[27] A. Bedekar, S. Borst, K. Ramana, P. Whiting, and E. Yeh, "Downlink scheduling in CDMA data networks," in *Proc. IEEE Globecom*, Rio de Janeiro, Brazil, Dec. 1999.

[28] A. Gjendemsjø, D. Gesbert, G. E. Øien, and S. G. Kiani, "Binary power control for sum rate maximization over multiple interfering links," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3164–3173, Aug. 2008.

[29] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wireless Commun. Mag.*, vol. 12, no. 2, pp. 56–65, Apr. 2005.

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

[31] G. B. Arfken and H. J. Weber, *Mathematical Methods For Physicists*, 4th ed. Academic Press, 1995.

[32] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Physical layer aspects for evolved UTRA (Release 7), 3GPP Std. Tech. Rep. 25.814, Sep. 2006.

PLACE PHOTO HERE

**Song Chong** (S'93-M'95) received the B.S. and M.S. degrees in Control and Instrumentation Engineering from Seoul National University, Seoul, Korea, in 1988 and 1990, respectively, and the Ph.D. degree in Electrical and Computer Engineering from the University of Texas at Austin in 1995. Since March 2000, he has been with the School of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, where he is a Professor and the Director of the Communications and Computing Group of the school. Prior to joining KAIST, he was with the Performance Analysis Department, AT&T Bell Laboratories, New Jersey, as a Member of Technical Staff. His current research interests include wireless networks, future Internet, human mobility and performance evaluation. He has published more than 70 papers in international journals and conferences and holds three U.S. patents in these areas.

He is an Editor of the Journal of Communications and Networks and has served on the Technical Program Committee of a number of key international conferences including IEEE INFOCOM, ACM CoNEXT, ITC etc. He is the General Chair of WiOpt '09. He is currently the Chair of Wireless Working Group of the Future Internet Forum and the Vice President of Information and Communication Society of Korea.

PLACE PHOTO HERE

**Jeong-woo Cho** (S'01-M'06) received his B.S., M.S., and Ph.D. degrees in Electrical Engineering and Computer Science from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2000, 2002, and 2005, respectively. From September 2005 to July 2007, he was with the Telecommunication R&D Center, Samsung Electronics, Suwon, Korea, as a Senior Engineer. From August 2007 to July 2008, he was a Senior Researcher in the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. From August 2008, he has been a Postdoc at the Centre for Quantifiable Quality of Service in Communication Systems, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

His research interests include flow/congestion control, mean field theories, network calculus, and cognitive radio networks.

PLACE PHOTO HERE

**Jeonghoon Mo** joined the Department of Information and Industrial Engineering at Yonsei University in March 2008. Before joining the Yonsei, he was an Associate Professor at Information and Communications University, Korea, where he led the WINE LAB. He also worked for the AT&T Bell Laboratories, New Jersey, and a couple of start-up companies in San Jose, CA. He got his Ph.D. degree at University of California at Berkeley in the spring of 1999.

He has been serving as Technical Program Committee members of the IEEE INFOCOM 2007, 2008, COIN-ACOFT 2007, IEEE ICC 2005 and so on. He is an editor of the Journal of Communications and Networks and was a secretary for COMSOC AP Board for the years of 2005-2007. He is affiliated with the WiMAX Forum, Future Internet Forum and more. He is a member of IEEE COMSOC and KICS.