Distributed Uplink Power Control for Optimal SIR Assignment in Cellular Data Networks

Prashanth Hande[1,2], Sundeep Rangan[2], Mung Chiang[1], Xinzhou Wu[2]
[1]Department of Electrical Engineering, Princeton University, NJ 08544, USA
[2]Qualcomm Flarion Technologies, NJ 07921, USA

***Abstract*—This paper solves the joint power control and SIR assignment problem through distributed algorithms in the uplink of multi-cellular wireless networks. The 1993 Foschini-Miljanic distributed power control can attain a given fixed and feasible SIR target. However, feasibility of SIR target cannot be readily determined a priori, and, in addition, SIR needs to be jointly optimized with transmit powers in wireless data networks. In the vast research literature since the mid-1990s, solutions to this joint optimization problem are either distributed but suboptimal, or optimal but centralized. For convex formulations of this problem, we report the first distributed and optimal algorithm.**

The main issue that has been the research bottleneck for many years is the complicated, coupled constraint set, and we resolve it through a re-parametrization via the *left* Perron Frobenius eigenvectors, followed by development of a *locally computable* ascent direction. A key step is a new characterization of the feasible SIR region in terms of the loads on the base stations, and an indication of the potential interference from mobile stations, which we term spillage. Based on this load-spillage characterization, we first develop a distributed algorithm that can achieve any Pareto-optimal SIR assignment, then a distributed algorithm that picks out a particular Pareto-optimal SIR assignment and the associated powers through utility maximization. Extensions to power-constrained and interference-constrained cases are carried out. The algorithms are theoretically sound and practically implementable: we present convergence and optimality proofs as well as simulations using 3GPP uplink evaluation tools.

I. Introduction

Power control in cellular networks has been extensively studied since the late 1980s as an important mechanism to control Signal-to-Interference Ratios (SIR), which in turn determine Quality-of-Service (QoS) metrics such as rate, outage, and delay. Uplink power control over multiple cells is particularly challenging. The basic version of CDMA power control solves the near-far problem by equalizing the received powers from mobile stations (MS) located at different parts of the cells. More sophisticated, iterative power control algorithms invented since the early 1990s find a transmit power vector so as to ensure that each MS attains the target SIR [2] while the overall power consumption is minimized. In particular, a distributed algorithm to achieve given fixed targets of SIR , if it is feasible, was proposed by Foschini and Miljanic [3] in 1993, with wide applicability in current cellular networks. This result has been followed by a large body of publications. In particular, the convergence properties of a more general class of similar distributed algorithms for power control was studied in [4].

This work was supported NSF Grants CCF-0448012, CNS-0417607, and CNS-0427677. Part of this work was presented at IEEE INFOCOM 2006.

However, there are two major limitations of the Foschini Miljanic power control algorithm. First, while in the power control algorithm in [3], the target SIRs are assumed to be within the feasibility region, a distributed mechanism to check the feasibility of the SIRs remains an open problem. A second limitation is that the fixed SIR approach to the power control problem is suitable only for voice networks where a given minimum SIR is necessary for stable communication at each MS. A cellular *data network*, however, can assign SIR according to traffic requirements, with higher SIRs implying better data rates and possibly greater reliability, while smaller SIRs can provide lower data rates for non-real-time data services. In addition, a cellular operator of a wireless data network might want to treat higher tariff paying users preferentially by allocating them to higher QoS classes, and the SIR target can be set differently for users in different QoS classes. It is therefore important to find a distributed mechanism to assign SIRs that belong to the feasible set and yet optimal according to criteria defined by the network operator. We provide a solution to this problem in this paper.

A main reason that such a solution has not been found over the last 16 years is that the feasible SIR region is coupled in a complicated way across MS in different cells. A standard way to describe the feasibility region is in terms of a condition on the spectral radius of a matrix with network-wide parameters as defined in 1992 [2], which seemingly requires a centralized computation. Initial attempts to assure some QoS performance but maintain the realistic distributed computation mechanism were made in the mid-1990s, e.g., in [5], [6]. Since the late 1990s, there have been two main threads of work tackling the problem of jointly optimizing SIR assignment and transmit powers over the feasibility region, e.g., in [7], [8], [9], [10], [11]. First, the work in [7] in 2001 proposes a utility framework by defining utility functions of the rates assigned to the MSs, with the rates being functions of SIRs, and the distributed solution converges to a Nash equilibrium that may be socially *suboptimal*. Second, in [12] in 2004, it is shown that the feasible SIR region is convex in the logarithm of SIR. This makes it possible to propose the utility framework for power control as a convex optimization problem. This has been the approach taken in [9], [10], [11] to obtain the global optimum, but the solutions require complex and *centralized* computation.

The main thread of this paper is the development of distributed algorithms that solves the convex formulations of the problem of jointly optimizing powers and SIR assignment over the feasibility region. Pictorially, the problem is to find the utility-optimal SIR assignment (where contours of the utility function are shown in dotted lines in Figure I) over the feasible SIR region (shown as the shaded region). The core idea behind all the developments and proofs is a novel re-parametrization of the feasibility region through the left, rather than the right, Perron-Frobenius eigenvectors of a network parameter matrix. Mathematically it amounts to a nonlinear change of coordinate that leads to decoupling through dual decomposition. The underlying engineering intuition is that the quality of a wireless uplink depends on both the strength of its direct channel gain and the weakness of interfering channel gains. The second key step in the solution, after the re-parametrization, is the development of an ascent direction that can be computed locally at each MS without global coordination.

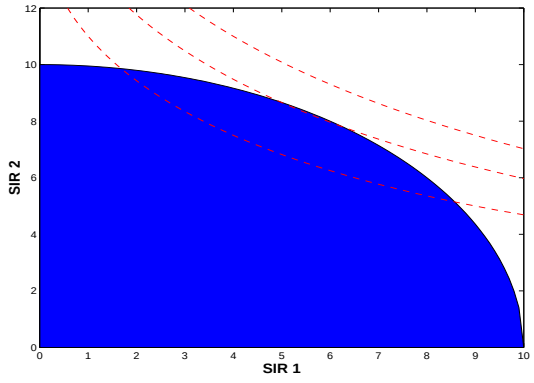

Fig. 1. Utility maximization over coupled feasibility region.

The key results of this paper are summarized as follows:

- *Characterize the Pareto-optimal boundaries of feasibility regions.* We define three different SIR feasibility regions. The first is a ρ-optimal subset where the power and interference are finite by a suitable choice of the parameter ρ. The other two feasibility sets are defined by imposing explicit constraints on the maximum allowed power and interference, respectively. We then derive the Pareto-optimal boundary of each of the three sets, with the intention of operating on those boundaries for SIR assignment.
- *Re-parameterize feasibility regions.* We propose an alternative view of the system, different from the standard one of power and interference. This is done by introducing two sets of parameters that can be interpreted as the *load* on the network, and an indication of the potential emitted interference from mobile stations, which we term *spillage*. It is of practical importance to note that the terms of load and spillage are used in a loose way in wireless industry. Here they have precise mathematical definitions, and the intuitions behind them are used in a provably optimal way.
- *Distributed algorithms to attain any point on the Pareto-optimal boundary.* Utilizing this load-spillage characterization, we propose a distributed algorithm that can attain any point on the Pareto-optimal boundary.
- *Distributed algorithms to "slide" along the Pareto-optimal boundary towards the utility-optimal point.* Further optimization over the Pareto-optimal boundary for unique operating point is posed as a utility maximization problem to maximize network utility. A distributed solution is proposed (the main algorithm labeled as Algorithm 4), thus overcoming the bottleneck of centralized computation of Perron-Frobenius eigenvalues and eigenvectors as required in current literature.
- In the process, we also introduce a new utility function, which we term the pseudo-linear utility function. It approximates the linear utility function at high SIRs but avoids MS starvation at low SIRs by precluding zero SIR allocation.
- Our algorithms can be readily implemented in today's cellular networks. Performance in realistic, large-sale networks is evaluated using the 3GPP uplink evaluation tool.

The rest of this paper is organized as follows. In Section II-A, we discuss the system setup and feasible SIR region, then introduce three different feasibility sets based on practical constraints, and then define the Pareto-optimal boundaries of the three sets. We develop the load-spillage characterization of the system in Section III-A, based on which we propose distributed algorithms of joint SIR assignment and power control for Pareto-optimal configuration in Section III. We discuss the utility framework in Section IV and propose distributed algorithms for utility optimization in Section V. Convergence and global optimality proofs are stated in the main text, with proofs of key lemmas summarized in various Appendices. The algorithm is extended to the power constrained and interference cases in Section V-B. Finally, we present simulations demonstrating the performance and robustness of our algorithms in Section VI. We conclude the paper in Section VII with a discussion of future research: the integration of multi-user detection, opportunistic scheduling, bandwidth assignment, soft handoff, and antenna beamforming into the load-spillage framework in this paper.

II. Feasibility and Pareto-Optimality

A. System Setup

Consider a general multi-cell setup where M MSs establish links to N BSs, as illustrated in Fig 2. We assume that each MS is served by one of the N BSs, thereby establishing M links. We let σ_i denote the receiving BS for link i.

Let C_i denote the set of links whose transmitted power appears as interference to link i. This definition allows us to consider both orthogonal and non-orthogonal uplinks. In a non-orthogonal uplink, such as CDMA, transmitted power from all links appear as interference, so we set $C_i = \{j \mid j \neq i\}$. For an orthogonal uplink, such as OFDM, links terminating on the same BS are orthogonal and do not contribute interference to one another. In this case, we set $C_i = \{j \mid \sigma_j \neq \sigma_i\}$.

Now let h^0_{kj} denote the absolute path gain from MS j to BS

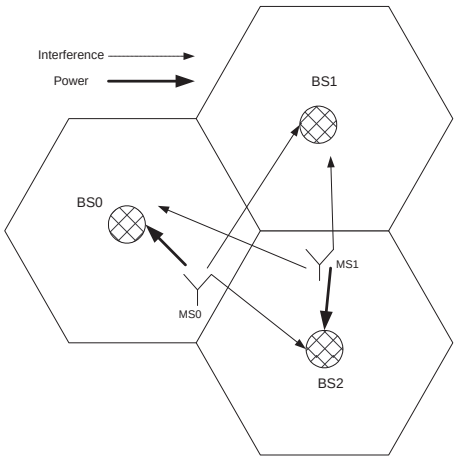

Fig. 2. An example of a multi-cellular network.

k, and define the normalized path gain by

$$h_{kj} = h^0_{kj}/h^0_{\sigma_j j}.$$

Define the $M \times M$ matrix $\mathbf{G}$ by,

$$G_{ij} = \begin{cases} h_{\sigma_i j} & \text{if } j \in C_i, \\ 0 & \text{if } j \notin C_i, \end{cases} \tag{1}$$

which represents the normalized path gain from MS on link j to the receiving BS on link i, when link j interferes with link i. Since an orthogonal uplink has $C_i = \{j \mid \sigma_j \neq \sigma_i\}$, and a non-orthogonal uplink has $C_i = \{j \mid j \neq i\}$, we can write both cases using a single expression:

$$G_{ij} = \begin{cases} h_{\sigma_i j} & \text{if } \sigma_i \neq \sigma_j \\ 0 & \text{if } i = j \\ \theta & \text{if } \sigma_i = \sigma_j, i \neq j \end{cases} \tag{2}$$

where $\theta = 1$ for the non-orthogonal case, and $\theta = 0$ for the orthogonal case.

Let p_j be the received signal power on link j at its serving BS σ_j. Since $h^0_{\sigma_j j}$ is the path gain from MS on link j to its serving BS, the MS on link j must transmit at a power of $p_j/h^0_{\sigma_j j}$. At any BS k, this signal will appear with a power of

$$h^0_{kj} p_j / h^0_{\sigma_j j} = h_{kj} p_j.$$

If $j \in C_i$, this transmission will appear as interference to link i with a power of $h_{\sigma_i j} p_j = G_{ij} p_j$. Using the fact that $G_{ij} = 0$ for $j \notin C_i$, the total interference and noise at the BS serving MS i is given by

$$q_i = \sum_{j \in C_i} G_{ij} p_j + \eta_i = \sum_{j=1}^{M} G_{ij} p_j + \eta_i, \tag{3}$$

where $\eta_i \geq 0$ is the power of noise other than interference from other links. In matrix notation, (3) can be written as,

$$\mathbf{q} = \mathbf{G}\mathbf{p} + \boldsymbol{\eta}. \tag{4}$$

Let γ_i be the SIR achieved by link i. With the above notation, $\gamma_i = p_i/q_i$, or equivalently,

$$\mathbf{p} = \mathbf{D}(\boldsymbol{\gamma})\mathbf{q}, \tag{5}$$

where $\mathbf{D}(\boldsymbol{\gamma}) = \text{diag}(\gamma_1, \ldots, \gamma_M)$. Combining (4) and (5), we get the following *basic equations*:

$$\mathbf{q} = \mathbf{G}\mathbf{D}(\boldsymbol{\gamma})\mathbf{q} + \boldsymbol{\eta}, \tag{6}$$

and

$$\mathbf{p} = \mathbf{D}(\boldsymbol{\gamma})\mathbf{G}\mathbf{p} + \mathbf{D}(\boldsymbol{\gamma})\boldsymbol{\eta}. \tag{7}$$

B. SIR Feasibility Region

Due to the interference between links, not all SIR vectors $\boldsymbol{\gamma}$ are achievable. An SIR vector $\boldsymbol{\gamma} \succ 0$ is called *feasible* if there exists an interference vector, $\mathbf{q} \succeq 0$, and power vector $\mathbf{p} \succeq 0$, satisfying (6) and (7), respectively. Let $\rho(\cdot)$ denote the spectral radius function [1] and assume that $\mathbf{G}$ is a primitive matrix [16] [2]. The following standard result gives a simple spectral radius characterization of SIR feasibility.

Lemma 1—Zander [2]: An SIR vector $\boldsymbol{\gamma} \succ 0$ is feasible if and only if

1) $\rho(\mathbf{G}\mathbf{D}(\boldsymbol{\gamma})) < 1$, when $\boldsymbol{\eta} \neq 0$
2) $\rho(\mathbf{G}\mathbf{D}(\boldsymbol{\gamma})) = 1$, when $\boldsymbol{\eta} = 0$.

If $\boldsymbol{\eta} \neq 0$, then the power allocation pertaining to $\boldsymbol{\gamma}$ is denoted by $\mathbf{p}(\boldsymbol{\gamma})$ and uniquely given by

$$\mathbf{p}(\boldsymbol{\gamma}) = (\mathbf{I} - \mathbf{D}(\boldsymbol{\gamma})\mathbf{G})^{-1}\mathbf{D}(\boldsymbol{\gamma})\boldsymbol{\eta}. \tag{8}$$

Correspondingly, the interference vector is denoted by $\mathbf{q}(\boldsymbol{\gamma})$ and uniquely given by

$$\mathbf{q}(\boldsymbol{\gamma}) = (\mathbf{I} - \mathbf{G}\mathbf{D}(\boldsymbol{\gamma}))^{-1}\boldsymbol{\eta}. \tag{9}$$

If $\boldsymbol{\eta} = 0$, then the power allocation $\mathbf{p}(\boldsymbol{\gamma})$ is the right eigenvector of $\mathbf{G}\mathbf{D}(\boldsymbol{\gamma})$ and the interference vector $\mathbf{q}(\boldsymbol{\gamma})$ is the right eigenvector of $\mathbf{D}(\boldsymbol{\gamma})\mathbf{G}$, both corresponding to eigenvalue 1. The eigenvectors are unique up to a scaling factor.

For a given feasible SIR allocation, the Foschini-Miljanic power control algorithm in [3] converges to the unique power vector as given in equation (8) when $\boldsymbol{\eta} \neq 0$ and to the right eigenvector of $\mathbf{G}\mathbf{D}(\boldsymbol{\gamma})$ when $\boldsymbol{\eta} = 0$. If γ_i is the target SIR and $\hat{\gamma}_i$ is the measured SIR at MS i, then the transmit power or effectively the received power is updated in the algorithm as

$$p_i[t+1] = p_i[t]\gamma_i/\hat{\gamma}_i. \tag{10}$$

However, the power control algorithm assumes a given SIR target without trying to jointly optimize over SIR assignments and transmit powers. Furthermore, the given SIR target is assumed to be feasible, but distributively checking feasibility is a difficult task on its own.

[1] Spectral radius is the maximum of the absolute value of the eigen values of a matrix.

[2] This is equivalent to saying that the network represented by $\mathbf{G}$ is connected, a reasonable assumption.

We assume in the rest of this paper that $\boldsymbol{\eta} \neq 0$ unless stated otherwise. Let $\mathbf{B} = \{\boldsymbol{\gamma} \succ 0 : \rho(\mathbf{GD}(\boldsymbol{\gamma})) < 1\}$ denote the set of all feasible SIR vectors $\boldsymbol{\gamma}$. As $\boldsymbol{\gamma}$ approaches the boundary of the feasibility region, the interference and power vectors tend to infinity in general. Practical networks, however, have finite limits on the interference and received power. To take care of this technical point, we devote several paragraphs to describe three alternative ways to ensure finite powers in practice.

Let $\mathbf{q}^m \succ 0$ and $\mathbf{p}^m \succ 0$ be maximum interference and power vectors, and define the corresponding feasibility sets of SIR vectors,

$$\begin{aligned} \mathbf{B}(\mathbf{p}^m) &= \{\boldsymbol{\gamma} \in \mathbf{B} \mid \mathbf{p}(\boldsymbol{\gamma}) \preceq \mathbf{p}^m\}, \\ \mathbf{B}(\mathbf{q}^m) &= \{\boldsymbol{\gamma} \in \mathbf{B} \mid \mathbf{q}(\boldsymbol{\gamma}) \preceq \mathbf{q}^m\}. \end{aligned}$$

The set $\mathbf{B}(\mathbf{p}^m)$ limits the received power, p_i of each mobile i. By scaling the received power by the path gain from the mobile to its serving BS, this bound can incorporate transmit power limits. The bound is most useful in coverage limited networks where the transmit power capabilities of the mobiles are the limiting factor in network capacity.

In capacity-limited networks, however, cells are typically sufficiently densely deployed, so that the absolute power capabilities of the mobiles are not the limiting factor. The set $\mathbf{B}(\mathbf{q}^m)$ limits the interference q_i at the BS serving MS i. In commercial network specifications, the interference limit $\mathbf{q}^m$ is often stated in the form $\mathbf{q}^m = \kappa\boldsymbol{\eta}$ for some constant $\kappa \geq 1$. With this definition, the interference, q_i, at each BS i, is not allowed to be larger than a factor κ greater than the thermal noise η_i. The factor, κ, is called the *rise over thermal (ROT)*, and typically quoted in dB,

$$\text{ROT} = 10\log_{10}(\kappa).$$

The ROT limit bounds the additional interference to the cell, and thereby limits the power required for new mobiles to access the network. Typical ROT values in commercial networks range from 3 to 10 dB and we will employ similar constraints in the simulations in Section VI.

Another constraint set that we will consider is

$$\mathbf{B}_\rho = \{\boldsymbol{\gamma} \succ 0 \mid \rho(\mathbf{GD}(\boldsymbol{\gamma})) \leq \rho\},$$

which is defined for any $\rho \in [0, 1)$ [3]. By Lemma 1, any $\boldsymbol{\gamma} \in \mathbf{B}_\rho$ is feasible. The constraint, $\boldsymbol{\gamma} \in \mathbf{B}_\rho$, results in finite power and interference for $\rho < 1$.

The sets $\mathbf{B}, \mathbf{B}(\mathbf{p}^m)$, $\mathbf{B}(\mathbf{q}^m)$ and $\mathbf{B}_\rho$ are not, in general, convex. However, a result in [12] shows that a certain transformation of the sets are. Given a set of feasible SIR vectors $\boldsymbol{\Gamma} \subseteq \mathbf{B}$, let $\log\boldsymbol{\Gamma} = \{\log\gamma \mid \gamma \in \boldsymbol{\Gamma}\}$.

Lemma 2—Boche and Stanczak [12]: The functions $\mathbf{q}(\boldsymbol{\gamma})$, $\mathbf{p}(\boldsymbol{\gamma})$ and $\rho(\mathbf{GD}(\boldsymbol{\gamma}))$ are convex in $\log\boldsymbol{\gamma}$ in the region $\log\mathbf{B}$. In particular, the sets, $\log\mathbf{B}(\mathbf{q}^m)$, $\log\mathbf{B}(\mathbf{q}^m)$ and $\log\mathbf{B}_\rho$ are convex.

[3] We tolerate the abuse of notation where ρ is both the spectral radius function as well as a parameter that limits the spectral radius.

C. SIR Pareto-Optimality Boundary

Let $\boldsymbol{\Gamma} \subseteq \mathbf{B}$ be a set of feasible SIR vectors. The selection of a feasible $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$ is, in general, a multi-objective optimization problem. Increasing the SIR, γ_i, for one mobile may require that the SIR, γ_j, for another mobile be reduced. A feasible SIR vector $\boldsymbol{\gamma} \succ 0$ is called *Pareto-optimal* if it is impossible to increase the SIR of any one link without at the same time reducing the SIR of some other link. The Pareto-optimal points of a set $\boldsymbol{\Gamma}$ form the Pareto-optimal boundary which we denote by $\partial\boldsymbol{\Gamma}$. The following theorem, proved in Appendices A and B, provides a simple characterization of Pareto-optimality for the feasible sets defined in Section II-B.

Theorem 1: Let $\boldsymbol{\eta} > 0$ be any positive noise vector, and let the gain matrix $\mathbf{G}$ be primitive. Then, the Pareto-optimal boundary $\partial\boldsymbol{\Gamma}$ of the following sets $\boldsymbol{\Gamma}$ is described as follows:

1) If $\boldsymbol{\Gamma} = \mathbf{B}(\mathbf{q}^m)$, $\boldsymbol{\gamma} \in \partial\mathbf{B}(\mathbf{q}^m)$ if and only if $\mathbf{q}(\boldsymbol{\gamma}) \preceq \mathbf{q}^m$ with at least one i such that $q_i(\boldsymbol{\gamma}) = q_i^m$
2) If $\boldsymbol{\Gamma} = \mathbf{B}(\mathbf{p}^m)$, $\boldsymbol{\gamma} \in \partial\mathbf{B}(\mathbf{p}^m)$ if and only $\mathbf{p}(\boldsymbol{\gamma}) \preceq \mathbf{p}^m$ with at least one i such that $p_i(\boldsymbol{\gamma}) = p_i^m$
3) If $\boldsymbol{\Gamma} = \mathbf{B}_\rho$, $\boldsymbol{\gamma} \in \partial\mathbf{B}_\rho$ if and only if $\rho(\mathbf{GD}(\boldsymbol{\gamma})) = \rho$.

A complete summary of the feasibility regions and Pareto-optimal boundaries is provided in Table I.

III. Distributed Power Control for Pareto-Optimal SIR Assignment

A. The Load-Spillage Characterization

Lemma 1 above shows that a candidate SIR vector $\boldsymbol{\gamma}$ is feasible if and only if $\rho(\mathbf{GD}(\boldsymbol{\gamma})) < 1$. Unfortunately, this has not been directly verifiable in a distributed manner. A direct computation of the spectral radius $\rho(\mathbf{GD}(\boldsymbol{\gamma}))$ will require a centralized controller that knows the entire candidate SIR vector $\boldsymbol{\gamma}$, as well as the complete connection matrix $\mathbf{G}$. In this section, we present an alternative, "dual" parameterization of the feasible SIR vectors $\boldsymbol{\gamma}$ that is more easily computed in a distributed manner. The parameterization is based on the following Lemma.

Lemma 3: An SIR vector $\boldsymbol{\gamma} \succ 0$ is feasible, *i.e.*, $\boldsymbol{\gamma} \in \mathbf{B}$, if and only if there exists a $\mathbf{s} \succ 0$ and $\rho \in [0, 1)$ such that

$$\mathbf{s}^T\mathbf{GD}(\boldsymbol{\gamma}) = \rho\mathbf{s}^T \tag{11}$$

Further, the SIR vector is ρ-optimal, *i.e.*, $\boldsymbol{\gamma} \in \partial\mathbf{B}_\rho$.

Proof: If $\boldsymbol{\gamma}$ is feasible, by Lemma 1, $\rho = \rho(\mathbf{GD}(\boldsymbol{\gamma})) \leq 1$. Let $\mathbf{s} \succ 0$ be the left eigenvector associated with $\rho(\mathbf{GD}(\boldsymbol{\gamma}))$, which is positive because $\mathbf{GD}(\boldsymbol{\gamma})$ is primitive [16].

Conversely, if some $\mathbf{s} \succ 0$ and $\rho \in [0, 1)$ satisfies (11), then $\mathbf{s}$ is a positive left eigenvector of the non-negative, primitive matrix $\mathbf{GD}(\boldsymbol{\gamma})$ which implies that ρ is the maximal eigenvalue $\rho = \rho(\mathbf{GD}(\boldsymbol{\gamma}))$ [16]. Since $\rho < 1$, the corresponding $\boldsymbol{\gamma}$ is feasible. ■

Lemma 3 leads to a natural parameterization of the set of feasible SIRs. Given any $\mathbf{s} \succ 0$ and $\rho \in [0, 1)$, let

$$\boldsymbol{\gamma}(\mathbf{s}, \rho) = \rho\mathbf{s}/\mathbf{r}(\mathbf{s}), \tag{12}$$

where the division of the vectors is component-wise, and $\mathbf{r}(\mathbf{s})$ is the following M-dimensional vector:

$$\mathbf{r}(\mathbf{s}) = \mathbf{G}^T\mathbf{s}. \tag{13}$$

TABLE I

FEASIBILITY REGIONS AND THEIR ASSOCIATED PARETO-OPTIMAL BOUNDARIES

	Feasibility Region:	Pareto-optimal Boundary:
$\{\boldsymbol{\eta} = 0\}$	$\{\boldsymbol{\gamma} : \rho(\mathbf{GD}(\boldsymbol{\gamma})) = 1\}$	$\{\boldsymbol{\gamma} : \rho(\mathbf{GD}(\boldsymbol{\gamma})) = 1\}$
$\{\boldsymbol{\eta} \neq 0, \rho - \text{optimal}\}$	$\mathbf{B} = \{\boldsymbol{\gamma} : \rho(\mathbf{GD}(\boldsymbol{\gamma})) < \rho\}$	$\partial\mathbf{B} = \{\boldsymbol{\gamma} : \rho(\mathbf{GD}(\boldsymbol{\gamma})) = \rho\}$
$\{\boldsymbol{\eta} \neq 0, \mathbf{p}(\boldsymbol{\gamma}) \preceq \mathbf{p}^m\}$	$\mathbf{B}(\mathbf{p}^m) = \{\boldsymbol{\gamma} : 0 \preceq \mathbf{p}(\boldsymbol{\gamma}) \preceq \mathbf{p}^m\}$	$\partial\mathbf{B}(\mathbf{p}^m) = \{\boldsymbol{\gamma} : 0 \preceq \mathbf{p}(\boldsymbol{\gamma}) \preceq \mathbf{p}^m, \exists i : p_i(\boldsymbol{\gamma}) = p_i^m\}$
$\{\boldsymbol{\eta} \neq 0, \mathbf{q}(\boldsymbol{\gamma}) \preceq \mathbf{q}^m\}$	$\mathbf{B}(\mathbf{q}^m) = \{\boldsymbol{\gamma} : 0 \preceq \mathbf{q}(\boldsymbol{\gamma}) \preceq \mathbf{q}^m\}$	$\partial\mathbf{B}(\mathbf{q}^m) = \{\boldsymbol{\gamma} : 0 \preceq \mathbf{q}(\boldsymbol{\gamma}) \preceq \mathbf{q}^m, \exists i : q_i(\boldsymbol{\gamma}) = q_i^m\}$

Observe that (11) is equivalent to $\mathbf{r}(\mathbf{s})^T\mathbf{D}(\boldsymbol{\gamma}) = \rho\mathbf{s}^T$. Consequently, $\mathbf{s}$, $\boldsymbol{\gamma}$ and ρ satisfy (11) if and only if $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\mathbf{s}, \rho)$. Therefore, Lemma 3 shows that $\boldsymbol{\gamma}(\mathbf{s}, \rho)$ parameterizes the set of all feasible positive SIRs. Moreover, if $\rho \in [0, 1)$ is fixed, then the set of $\boldsymbol{\gamma}(\mathbf{s}, \rho)$ parameterizes the set of ρ-optimal SIRs.

Also, combining (13) with (11), we see that

$$\rho\mathbf{r}^T = \mathbf{r}^T\mathbf{D}(\boldsymbol{\gamma})\mathbf{G}. \tag{14}$$

Therefore, $\mathbf{r}$ and $\mathbf{s}$ are left positive eigenvectors of the matrices, $\mathbf{D}(\boldsymbol{\gamma})\mathbf{G}$ and $\mathbf{GD}(\boldsymbol{\gamma})$, respectively. This pair of left eigenvectors, $\{\mathbf{s}, \mathbf{r}\}$, have an interesting interpretation: $\mathbf{s}$ represents the load on the network to support an SIR $\boldsymbol{\gamma}$, and $\mathbf{r}$ represents the potential interference due to an SIR assignment. We call them the **load factors** and **spillage factors**.

This interpretation is made more clear by writing down the spillage r_i associated with link i in terms of the link matrix G_{ij} and the load factors. From (12) and (13), it follows that

$$\begin{array}{rrcl} \text{spillage:} & r_i & = & \sum_j G_{ji}s_j \\ \text{load:} & s_i & = & r_i\gamma_i/\rho \end{array} \tag{15}$$

The spillage factor r_i represents the effect of interference due to link i on other links in the network weighted by the loads of each link. Link i, with an SIR γ_i and responsible for spillage r_i, loads the network with $s_i = (1/\rho)r_i\gamma_i$ in the sense that it is less tolerant by a factor of s_i to interference from other links in the network.

B. Distributed SIR Assignment for ρ-Optimality

The characterization developed in Section III-A leads to a simple distributed algorithm for assigning feasible SIRs. To describe the assignment algorithm, suppose that we wish to achieve an SIR vector $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\mathbf{s}, \rho)$ for some $\mathbf{s} \succ 0$ and $\rho \in [0, 1)$. We let $\mathbf{r} = \mathbf{r}(\mathbf{s})$, then it follows from (15) that

$$r_i = \sum_j G_{ji}s_j = \sum_{j\neq i, \sigma_j = \sigma_i} G_{ji}s_j + \sum_{j\neq i, \sigma_j \neq \sigma_i} G_{ji}s_j \tag{16}$$

Using (2), we can write this as

$$r_i = \theta \sum_{j\neq i, j\in S_{\sigma_i}} s_j + \sum_{k\neq\sigma_i} h_{ki}\ell_k \tag{17}$$

where $S_k = \{j \mid \sigma_j = k\}$ is the set of mobiles connected to BS k, and

$$\ell_k = \sum_{j\in S_k} s_j, \tag{18}$$

which we term the **BS-load factor** associated with BS k. The spillage can therefore be factored as

$$r_i = r_i^{int} + r_i^{ext}, \tag{19}$$

where r_i^{ext} is the *external* component of the spillage due to MSs in other cells,

$$r_i^{ext} = \sum_{k\neq\sigma_i} h_{ki}\ell_k, \tag{20}$$

and r_i^{int} is the *internal* component due to spillage from mobiles in the same cell,

$$r_i^{int} = \begin{cases} \sum_{j\neq i, j\in S_{\sigma_i}} s_j & \text{if the uplink is non-orthogonal,} \\ 0 & \text{if the uplink is orthogonal.} \end{cases} \tag{21}$$

This computation leads to a natural distributed SIR assignment algorithm. Fix the load factors s_j allocated to the MSs, which in turn fixes the BS-load factors ℓ_k. Then the following algorithm assigns a ρ-optimal SIR.

Algorithm 1—**Distributed Assignment over $\partial\mathbf{B}_\rho$:**

Initialize: Fixed $\mathbf{s} \succ 0$ and $\rho \in [0, 1)$.

1) BS k broadcasts the BS-load factor ℓ_k.
2) Compute the spillage factor r_i according to (17).
3) Assign SIR values $\gamma_i = \rho s_i/r_i$.

Stop. The resulting SIR vector $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\mathbf{s}, \rho)$.

Remarks:

1) The algorithm consists of a distributed one-step computation (no iteration), and results in ρ-optimal SIR given a set of load-factors $\{s_i\}$. Increasing s_i for some i results in a higher γ_i at the expense of some other links in the system. But a proportional increase in every s_i results in no change in the SIR assignment.
2) BSs can broadcast ℓ_k on the downlink at a fixed power so that MSs in the vicinity can decode the BS load-factor. The BS load-factor transmission power can be fixed at a value such that all MSs that can potentially interfere with the BS can decode the load-factor transmitted. In addition, the transmission of the BS load-factor at fixed power will enable the MS to measure the relative path gains h_{ki} required for spillage calculation.
3) The initial load allocation can be made at the MS (*MS-Control*). Alternatively, it can also be made at the BS (*BS-Control*). BS-Control requires a mechanism to either convey the load s_i to the MS, or convey r_i^{ext} from the MS to the BS and then the assigned SIR from the BS to

the MS. BS-Control has a higher messaging overhead in comparison to MS-Control but has the advantage of better overall system control since the load allocation can be made based on a network-wide criterion.

Algorithm 1 is only the first step of the development. It results in a ρ-optimal point on the Pareto-optimal boundary of the feasible SIR region. By optimizing further over $\mathbf{s}$ along the right direction, the "best" point on the Pareto-optimal boundary can be picked out, as will be discussed in the next two sections.

For $\rho < 1$, the resulting power and interference vectors from Algorithm 1 are finite. We further quantify the resulting power and interference vectors as follows.

Lemma 4: Let $\gamma = \boldsymbol{\gamma}(\mathbf{s}, \rho)$ for some $\mathbf{s} > 0$ and $\rho \in [0, 1)$, and let $\mathbf{p} = \mathbf{p}(\gamma)$ and $\mathbf{q} = \mathbf{q}(\gamma)$ be the corresponding power and interference vectors. Then,

$$\mathbf{r}^T \left(\mathbf{p} - \frac{1}{1-\rho} \mathbf{D}(\boldsymbol{\gamma}) \boldsymbol{\eta} \right) = 0, \ \mathbf{s}^T \left(\mathbf{q} - \frac{1}{1-\rho} \boldsymbol{\eta} \right) = 0, \quad (22)$$

where $\mathbf{r} = \mathbf{r}(\mathbf{s})$ in (13).

Proof: Operating on the left by $\mathbf{r}^T$ on both sides of (7) and by $\mathbf{s}^T$ on both sides of (6) results in (22). ■

The following lemma is proved in Appendix C.

Lemma 5: Fix any $\mathbf{s} > 0$ and for $\rho \in [0, 1)$, let $\mathbf{q}_\rho = \mathbf{q}(\boldsymbol{\gamma})$ be the interference vector corresponding to $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\mathbf{s}, \rho)$. Then,

(a) There exists a unique vector $\tilde{\mathbf{q}} > 0$ such that, for all $\rho \in [0, 1)$, $\tilde{\mathbf{q}}$ is a positive right eigenvector of $\mathbf{G}\mathbf{D}(\boldsymbol{\gamma}(\mathbf{s}, \rho))$, normalized so that $\mathbf{s}^T \tilde{\mathbf{q}} = \mathbf{s}^T \boldsymbol{\eta}$.

(b) The interference vectors $\mathbf{q}_\rho$ satisfy the following limit:

$$\lim_{\rho \to 1} (1 - \rho) \mathbf{q}_\rho = \tilde{\mathbf{q}}. \quad (23)$$

To interpret Lemma 4, let $\mathbf{y} = \mathbf{s}\boldsymbol{\eta}$, where the multiplication of the vectors is component-wise. Then, the second equation in (22) can be rewritten as

$$\mathbf{y}^T \left(\frac{\mathbf{q}}{\boldsymbol{\eta}} - \frac{1}{1-\rho} \right) = 0.$$

Recalling from Section II-B that the ratio $\mathbf{q}/\boldsymbol{\eta}$ is the ROT vector for the system, Lemma 4 shows that a weighted average ROT is bounded by $1/(1-\rho)$. The parameter ρ is thus related to the ROT limit. Lemma 5 extends this further and shows that, as $\rho \to 1$, ROT $= O(1/(1-\rho))$. This fact will be used in a later proof of convergence and optimality.

C. Distributed Control for Pareto-Optimality with Power and Interference Constraints

While ρ-optimality with $\rho < 1$ guaranteed a finite power and interference vector, an alternative operating mode is to specify the power and interference constraints explicitly. In ρ-optimization, the set of feasible SIR vectors, $\boldsymbol{\gamma}$, were parameterized by the load vector $\mathbf{s}$. For the interference and power constrained case, we need to introduce a second parameter vector $\boldsymbol{\nu} \succeq 0$. The vector $\boldsymbol{\nu}$ will be called the **price vector**, and will play a role as Lagrange multipliers as will be seen in Section V-B.

For ease of presentation, we define two cases, case *P* and case *Q*, with power consraint $\mathbf{p}^m$ and interference constraint $\mathbf{q}^m$, respectively.

$$\begin{array}{ll} \text{Case } P: & \{\boldsymbol{\gamma} \in \mathbf{B}(\mathbf{p}^m)\} \\ \text{Case } Q: & \{\boldsymbol{\gamma} \in \mathbf{B}(\mathbf{q}^m)\} \end{array} \quad (24)$$

Given a load vector $\mathbf{s} \succ 0$ and price vector $\boldsymbol{\nu} \succeq 0$, define vectors $\mathbf{r}^p(\mathbf{s}, \boldsymbol{\nu})$, $\boldsymbol{\gamma}^p(\mathbf{s}, \boldsymbol{\nu})$, $\mathbf{r}^q(\mathbf{s}, \boldsymbol{\nu})$ and $\boldsymbol{\gamma}^q(\mathbf{s}, \boldsymbol{\nu})$ as

$$\begin{array}{lrcl} \text{Case } P: & \mathbf{r}^p(\mathbf{s}, \boldsymbol{\nu}) & = & \mathbf{G}^T \mathbf{s} + \boldsymbol{\nu} \\ & \boldsymbol{\gamma}^p(\mathbf{s}, \boldsymbol{\nu}) & = & \mathbf{s} / \mathbf{r}^p(\mathbf{s}, \boldsymbol{\nu}) \end{array} \quad (25)$$

$$\begin{array}{lrcl} \text{Case } Q: & \mathbf{r}^q(\mathbf{s}, \boldsymbol{\nu}) & = & \mathbf{G}^T (\mathbf{s} + \boldsymbol{\nu}) \\ & \boldsymbol{\gamma}^q(\mathbf{s}, \boldsymbol{\nu}) & = & \mathbf{s} / \mathbf{r}^q(\mathbf{s}, \boldsymbol{\nu}) \end{array} \quad (26)$$

Notice the resemblance of the relationship (25) and (26) to that in equation (12) and (13). This resemblance enables us to interpret $\mathbf{s}$ as a load-factor and $\mathbf{r} = \mathbf{r}^p(\mathbf{s}, \boldsymbol{\nu})$ as the spillage associated with case *P* and $\mathbf{r} = \mathbf{r}^q(\mathbf{s}, \boldsymbol{\nu})$ as the spillage associated with case *Q*. In fact, for a given link i, the relationship can be expressed as

$$\begin{array}{lrcl} \text{Case } P: & r_i & = & \sum_j G_{ji} s_j + \nu_i \\ & s_i & = & r_i \gamma_i \end{array} \quad (27)$$

$$\begin{array}{lrcl} \text{Case } Q: & r_i & = & \sum_j G_{ji} (s_j + \nu_j) \\ & s_i & = & r_i \gamma_i \end{array} \quad (28)$$

The achieved SIR in each case is given by $\boldsymbol{\gamma} = \boldsymbol{\gamma}^p(\mathbf{s}, \boldsymbol{\nu})$ and $\boldsymbol{\gamma} = \boldsymbol{\gamma}^q(\mathbf{s}, \boldsymbol{\nu})$, respectively. For a fixed $\mathbf{s}$ and a suitable choice of the price vector $\boldsymbol{\nu}$, the achieved SIR $\boldsymbol{\gamma} = \boldsymbol{\gamma}^p(\mathbf{s}, \boldsymbol{\nu})$ lies on the Pareto-optimal boundary $\partial\mathbf{B}(\mathbf{p}^m)$ for case *P* and the achieved SIR $\boldsymbol{\gamma} = \boldsymbol{\gamma}^q(\mathbf{s}, \boldsymbol{\nu})$ lies on the Pareto-optimal boundary $\partial\mathbf{B}(\mathbf{q}^m)$ for case *Q*.

The prices $\{\nu_i\}$ correspond to local constraints and hence amenable to a simple subgradient-based update:

$$\text{Case } P: \quad \nu_i[t+1] = [\nu_i[t] + \delta[t](p_i[t] - p_i^m)]^+ \quad (29)$$

$$\text{Case } Q: \quad \nu_i[t+1] = [\nu_i[t] + \delta[t](q_i[t] - q_i^m)]^+ \quad (30)$$

where $\delta[t] = \delta_0/t, \delta_0 > 0$, is a suitable choice for the step size. Based on this update, we present the following algorithm to iteratively arrive at the appropriate price-vector.

Algorithm 2—**Distributed Assignment over $\partial\mathbf{B}(\mathbf{p}^m)$,$\partial\mathbf{B}(\mathbf{q}^m)$:**

Initialize: fixed $s_i[0] \succ 0, \nu_i[0] \succeq 0$.

1) Case *P*: Compute spillage $r_i[t]$ according to (27).
   Case *Q*: Compute spillage $r_i[t]$ according to (28).
2) Case *P*: Assign SIR $\gamma_i[t]$ for MS i according to (27).
   Case *Q*: Assign SIR $\gamma_i[t]$ for MS i according to (28).
3) Case *P*: Measure resulting power $p_i[t]$.
   Case *Q*: Measure resulting power $q_i[t]$.
4) Case *P*: Update power price $\nu_i[t]$ according to (29).
   Case *Q*: Update power price $\nu_i[t]$ according to (30).

Continue: $t := t + 1$.

While prices were additive to the spillage in case *P*, bringing down the SIR assignment to lie on $\partial\mathbf{B}(\mathbf{p}^m)$, the prices in case *Q* are additive to the load and have the same effect of bringing

down the SIR assignment to lie on $\partial\mathbf{B}(\mathbf{q}^m)$. When the prices converge, the resulting SIR is pareto-optimal as shown in the following theorem.

Theorem 2: If Algorithm 2 has a fixed point $\boldsymbol{\gamma}$, then $\boldsymbol{\gamma} \in \partial\mathbf{B}(\mathbf{p}^m)$ for case *P* and $\boldsymbol{\gamma} \in \partial\mathbf{B}(\mathbf{q}^m)$ for case *Q*.

Proof: Consider the case *P*. If $\boldsymbol{\nu}$ is the fixed point of Algorithm 2, then it cannot be that $\boldsymbol{\nu} = 0$ because that results in a $\boldsymbol{\gamma}$ requiring infinite power. So there is some i, such that $\nu_i > 0$ and for such a i, $p_i = p_i^m$, indicating that the resulting $\boldsymbol{\gamma}$ is indeed Pareto-optimal. The argument is similar for case *Q*. ■

IV. Utility Maximization Over Feasibility Regions

In Section III, we discussed distributed mechanisms to achieve any SIR vector $\boldsymbol{\gamma}$ on a Pareto-optimal boundary of the feasible SIR region. Picking a particular point on the Pareto-optimal boundary, out of an infinite number of choices, can be accomplished by specifying a network-wide objective function, generally denoted as the network utility $U(\boldsymbol{\gamma})$. Given a set of feasible SIR vectors, $\boldsymbol{\gamma} \in \boldsymbol{\Gamma} \subseteq \mathbf{B}$, the optimal SIR over this set is defined by $\boldsymbol{\gamma}^{opt} = \arg\max_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} U(\boldsymbol{\gamma})$.

Let $U_i(\beta_i)$ be a *utility*, representing the value to the overall network of allocating link i a QoS metric β_i, which has a bijective mapping to the SIR γ_i given by $\beta_i = \beta(\gamma_i)$. One such QoS metric of interest is the Shannon capacity, given by

$$\beta(\gamma_i) = w_i \log_2(1 + \gamma_i W/w_i) \tag{31}$$

where w_i is the bandwidth allocated to each user out of a total available bandwidth of W. We assume that w_i is fixed and equal for all users in this work. Joint optimization over SIR assignment and bandwidth allocation has also recently been studied in [13].

A useful set of utility functions to consider is the α-fair utility functions [14] defined on the QoS parameters $\beta_i = \beta(\gamma_i)$:

$$U_i(\beta_i) = \begin{cases} \log(\beta_i) & \text{if } \alpha = 1, \\ (1-\alpha)^{-1}\beta_i^{1-\alpha} & \text{if } \alpha \neq 1 \end{cases} \tag{32}$$

where $\alpha = 0$ represents *Linear Utility* and $\alpha = 1$ represents *Log Utility*. Increasing α leads to fairer allocation in the corresponding utility maximization problem with proportional fairness achieved for ($\alpha = 1$) and max-min fairness achieved as ($\alpha \to \infty$).

In addition to the α-fair utilities, we consider a novel utility function that we call the pseudo-linear utility function:

Pseudo-Linear Utility: $U_i(\beta_i) = \log(\exp(\beta_i/W) - 1)$. (33)

This utility function is fairer than the linear utility function in an important way: zero QoS assignment is infinitely penalized and not allowed. At high values of the QoS metric, the pseudo-linear utility approximates the linear utility, thus restoring the capability of maximizing a close approximation to sum QoS. At low values of the QoS metric, the pseudo-linear utility approximates the logarithmic utility and ensures non-zero QoS for all transmitting users. More precisely, for $U_i(\beta_i)$ in (33),

$$\begin{aligned} U_i(\beta_i) &\to \beta_i/W \quad \text{as} \quad \beta_i \to \infty \\ U_i(\beta_i) &\to -\infty \quad \text{as} \quad \beta_i \to 0. \end{aligned} \tag{34}$$

Pseudo-linear utility function is particularly suitable from the network-opeartor's point of view, since operators would typically like to maximize the sum throughput under normal loading of the cells, but also ensure no outage to any MS when the cells are heavily loaded. In this paper, instead of the linear utility function, we use this pseudo-linear utility function.

The bijective mapping β of the QoS metric β_i from the SIR γ_i allows us to view the utility as a function of the SIR. For ease of presentation, we tolerate some abuse of notation and represent this function as $U_i(\gamma_i)$ instead of the more appropriate $U_i(\beta(\gamma_i))$. We use the following notation: $U_i'(\gamma_i) = \partial U_i/\partial\gamma_i$ and $U_i''(\gamma_i) = \partial^2 U_i/\partial^2\gamma_i$. We will make the following assumptions on the utility functions throughout this paper.

1) The utility functions $U_i(\gamma_i)$ are strictly increasing, twice differentiable and strictly concave in $\log\gamma_i$.
2) $U_i'(\gamma_i)$ and $U_i''(\gamma_i)$ are both bounded for all $\gamma_i \geq 0$.
3) $U_i(\gamma_i)$ is fair in the sense that, as $\gamma_i \to 0$, $U_i(\gamma_i) \to -\infty$, so that zero SIR assignment is precluded in any solution to the utility maximization problem with non-empty constraint set.

The assumption on strict concavity in $\log\gamma_i$ can be expressed in an alternative form. It can be easily verified that $U_i(\gamma_i)$ is strictly concave in $\log\gamma_i$ if and only if the negative of the curvature is sufficiently large:

$$U_i''(\gamma_i) \leq -\frac{U_i'(\gamma_i)}{\gamma_i}.$$

In particular, the above condition is satisfied for α-fair utilities when $\alpha \geq 1$.

Together with Lemma 2, the above assumption shows that the maximization of the utility over any of the feasibility sets $\mathbf{B}_\rho$, $\mathbf{B}(\mathbf{q}^m)$, or $\mathbf{B}(\mathbf{p}^m)$ is a convex optimization problem. In particular, a local maximum of the constrained optimization is also a global maximum.

V. Distributed Power Control for Utility-Optimal SIR Assignment

In this section, we develop a distributed power control algorithm to solve the optimal SIR assignment problem of maximizing $U(\boldsymbol{\gamma}) = \sum_i U_i(\gamma_i)$ when the SIR is constrained to one of three feasibility sets: $\mathbf{B}_\rho, \mathbf{B}(\mathbf{p}^m), \mathbf{B}(\mathbf{q}^m)$.

Intuitively, any such optimization over the Pareto-optimal boundary should assign a higher SIR value to links in good channel conditions. A second criterion, which is less understood in existing literature, is to take into consideration the channel conditions of the *interfering paths* of each link. Good channel conditions on the interfering paths can carry over substantial interfering signal strength even if the transmit power of the link is low. In this case, we should assign a higher SIR to the link that has worse interfering channels. This intuition will be made precise in this section. The power-interference viewpoint of the system provides information on the condition of the channel that represents the primary connection of a link, whereas the load-spillage viewpoint of the system provides information on the channel conditions of the interfering paths associated with the link.

A. Distributed Optimization over ρ-Optimal Boundary

First consider the case where $\boldsymbol{\gamma}$ is a ρ-optimal assignment for some $\rho \in [0,1)$:

$$\begin{aligned} \text{maximize} \quad & U(\boldsymbol{\gamma}) \\ \text{subject to} \quad & \boldsymbol{\gamma} \in \mathbf{B}_\rho. \end{aligned} \tag{35}$$

While the above problem is convex optimization under the aforementioned assumptions, the constraint set is highly coupled. As a result, no distributed algorithms have been possible to date. Indeed, the problem has been considered before (e.g., in [9], [10]), but only centralized algorithms have been proposed. We now show that the load-spillage characterization makes distributed optimal algorithms possible.

Fixing $\rho \in [0,1)$ and using the parameterization $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\mathbf{s}, \rho)$ in Section III-A, we can view the utility maximization problem as an optimization over the load vector $\mathbf{s}$. The maximization problem can then be solved by application of any general ascent method [17] in the feasibility region of $\mathbf{s}$. We obtain an ascent direction that can be calculated locally by each MS, and employ the mechanism in Algorithm 1 to update the SIR associated with the links.

Consider the following load-update $\Delta\mathbf{s}$ to the load vector $\mathbf{s}$,

$$\Delta s_i = \frac{U_i'(\gamma_i)\gamma_i}{q_i} - s_i, \quad \forall i. \tag{36}$$

We first present an iterative, distributed algorithm based on this locally computable update, and then prove that (36) indeed represents an ascent direction and leads to an iteration whose fixed point is the optimal load distribution problem (35). The intuition behind the proposed load update will be further explained after the proof of convergence is presented.

Algorithm 3—**Utility Maximization over $\mathbf{B}_\rho$:**

- Parameters: step size $\delta > 0$ and utility functions $\{U_i(\gamma_i)\}$.
- Initialize: Arbitrary $\mathbf{s}[0] \succ 0$.

1) BS k broadcasts the BS-load factor $\ell_k[t] = \sum_{i \in S_k} s_i[t]$.
2) Compute the spillage-factor $r_i[t]$ according to (17).
3) Assign SIR values $\gamma_i[t] = \rho s_i[t]/r_i[t]$.
4) BS measures the resulting interference $q_i[t]$.
5) Update the load factor $s_i[t]$ in the ascent direction given by (36)
$$s_i[t+1] = s_i[t] + \delta \Delta s_i[t].$$

Continue: $t := t+1$.

Remarks: The load update can be done either at the BS(BS-Control) or at the MS(MS-Control).

1) The load update for MS-Control requires the calculation of the resulting interference based on the assigned SIR and the transmitted power, but otherwise results in a completely distributed implementation.
2) BS-Control requires a mechanism to have the MS calculate r_i^{ext} and convey that to the BS, but provides the flexibility of letting the BS decide on the utility function associated with the links it is serving. BS-Control is still distributed in the sense that each BS in the network decides on the SIR assignment of the links it is serving, with the overall SIR assignment turning out to be optimal for the entire multicellular network without the need for a centralized SIR assignment across the BSs in different cells.
3) The interference $q_i[t]$ can be measured after the power updates, e.g., by the Foschini-Miljanic method, converge. We empirically find in our simulations that convergence is achieved even when $q_i[t]$ is the interference resulting from a single update in the power level (10), i.e., even when power control is running at the same time-scale as SIR assignment.

From Lemma 5, as $\rho \to 1$, the following approximation becomes accurate:

$$\mathbf{q}[t] \approx \frac{1}{1-\rho}\tilde{\mathbf{q}}(\gamma[t]),$$

where $\tilde{\mathbf{q}}(\gamma[t])$ is a suitably normalized positive right eigenvector of $\mathbf{GD}(\gamma[t])$. Algorithm 4 converges to the globally utility-optimal SIR assignment when $\rho \to 1$. This includes the important, zero-noise case often considered in the research literature, for which $\rho = 1$ is the feasibility region.

Theorem 3: For $\rho \to 1$ and sufficiently small step size $\delta > 0$, Algorithm 3 converges to the globally optimal solution of problem (35).

Proof: Consider the sum utility that is being maximized in (35) as a function of $\mathbf{s}$, $F_\rho(\mathbf{s}) = U(\boldsymbol{\gamma}(\mathbf{s}))$. Let $\Delta\mathbf{s}$ be the update direction of Algorithm 3 as given in (36) with the load update given by

$$\mathbf{s}[t+1] = \mathbf{s}[t] + \delta\Delta\mathbf{s}[t]. \tag{37}$$

We show in Appendix D that (37) is a general ascent update. In addition, the Hessian's norm is bounded since the first and second derivatives of the utilities are bounded, the load $\mathbf{s} \succ 0$ at every iteration t (since zero SIR will not be assigned to any MS due to assumption 3 on utility function), and the connection matrix $\mathbf{G}$ is primitive. This implies the Lipschitz continuity property [18] on $F_\rho(\mathbf{s})$ and the existence of a non-zero step size δ such that $U(\boldsymbol{\gamma}[t+1]) > U(\boldsymbol{\gamma}[t])$. With Lipschitz property, the descent lemma in [18] implies convergence of this ascent algorithm.

If $\mathbf{s}^*$ is the point of convergence of the algorithm, then the corresponding spillage $\mathbf{r}^* = \mathbf{G}^T\mathbf{s}^*$ and SIR $\{\boldsymbol{\gamma}^* = \rho\mathbf{s}^*/\mathbf{r}^*\}$ satisfy $\Delta\mathbf{s}^* = 0$ so that

$$U_i'(\gamma_i^*) = r_i^*\tilde{q}_i^*(\boldsymbol{\gamma}^*), \qquad i = 1, \ldots, M. \tag{38}$$

This is precisely the KKT condition for $\{\boldsymbol{\gamma}^*, \nu^* = 1\}$ the optimization problem (35) as shown in Appendix E. Since the optimization problem (35) is convex after a change of variable on $\mathbf{s}$, any point that satisfies the KKT condition is indeed the optimal solution to the problem [17]. ■

Since $U_i'(\gamma_i)$ is decreasing in γ_i, we see that larger the spillage and larger the interference, the lower the optimal γ_i is. This makes intuitive sense, since a link with large interference requires a larger power allocated for a given SIR and results in

larger interference to neighboring links. Similarly, a link with larger spillage easily carries over interference into neighboring links and hence should be assigned a smaller SIR.

B. Utility Maximization with Power and Interference Constraints

The utility maximization problem over the power (Case *P*) and interference (Case *Q*) constrained SIR region can be written as follows

$$\text{Case } P: \quad \begin{array}{ll} \text{maximize} & \sum_i U_i(\gamma_i) \\ \text{subject to} & \mathbf{p}(\boldsymbol{\gamma}) \preceq \mathbf{p}^m \end{array} \tag{39}$$

$$\text{Case } Q: \quad \begin{array}{ll} \text{maximize} & \sum_i U_i(\gamma_i) \\ \text{subject to} & \mathbf{q}(\boldsymbol{\gamma}) \preceq \mathbf{q}^m \end{array} \tag{40}$$

The Lagrangian for this convex optimization problem can be written as

$$\text{Case } P: \quad \mathcal{L}_p(\boldsymbol{\gamma}, \boldsymbol{\nu}) = U(\boldsymbol{\gamma}) + \boldsymbol{\nu}^T(\mathbf{p}^m - \mathbf{p}) \tag{41}$$

$$\text{Case } Q: \quad \mathcal{L}_q(\boldsymbol{\gamma}, \boldsymbol{\nu}) = U(\boldsymbol{\gamma}) + \boldsymbol{\nu}^T(\mathbf{q}^m - \mathbf{q}) \tag{42}$$

where $\{\nu_i \geq 0, i = 1, \ldots, M\}$ are the Lagrange multipliers.

Using standard convex optimization theory [17], we attempt to move the SIR vector $\boldsymbol{\gamma}$ in a direction that increases the Lagrangian, thereby attempting to solve the weighted optimization

$$\max_{\boldsymbol{\gamma}} \mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\nu}),$$

where $\mathcal{L} = \mathcal{L}_p$ or $\mathcal{L}_q$ depending on the problem. Similar to Section V-A, we approach the problem in the domain of the load vectors $\mathbf{s}$ rather than the domain of SIR vectors $\boldsymbol{\gamma}$, and use the ascent direction $\Delta\mathbf{s}$ in (36) for this purpose. The prices are optimized by gradient updates. The proposed algorithm is as follows:

Algorithm 4—**Utility Maximization over $\mathbf{B}(\mathbf{p}^m)$ or $\mathbf{B}(\mathbf{q}^m)$:**

- Parameters: Step sizes $\delta_1 > 0$ for load update, $\delta > 0$ for price update, utility functions $U_i(\gamma_i)$, and interference or power constraints, $\mathbf{p}^m$ or $\mathbf{q}^m$.
- Initialize: Arbitrary $\mathbf{s}[0] \succ 0$ and $\boldsymbol{\nu}[0] \succeq 0$.

1) a) Case *P*: Compute spillage $r_i[t]$ according to (27)
   b) Case *Q*: Compute spillage $r_i[t]$ according to (28)
2) a) Case *P*: Assign $\gamma_i[t]$ according to (27)
   b) Case *Q*: Assign $\gamma_i[t]$ according to (28)
3) Measure resulting interference $q_i[t]$
4) Update load $s_i[t]$ in the ascent direction given by (36)

$$s_i[t+1] = s_i[t] + \delta_1 \Delta s_i[t].$$

5) a) Case *P*: Update price $\nu_i[t]$ according to (29)
   b) Case *Q*: Update price $\nu_i[t]$ according to (30)

Continue: $t := t + 1$

Theorem 4: For sufficiently small step size $\delta_1 > 0$ and $\delta > 0$, Algorithm 4 converges to the globally optimal solution of problem (39) for the power constrained case and (40) for the interference constrained case.

Proof: For a given $\boldsymbol{\nu}$, consider the Lagrangian as a function of the load:

$$F_p(\mathbf{s}) = \mathcal{L}_p(\boldsymbol{\gamma}(\mathbf{s}, \boldsymbol{\nu}), \boldsymbol{\nu}) \tag{43}$$

$$F_q(\mathbf{s}) = \mathcal{L}_q(\boldsymbol{\gamma}(\mathbf{s}, \boldsymbol{\nu}), \boldsymbol{\nu}) \tag{44}$$

The update direction $\Delta\mathbf{s}$ of the algorithm is given in (36) with the load update given by (37). We show in Appendix D that (37) is an ascent update with respect to $F_p(\mathbf{s})$ and $F_q(\mathbf{s})$. In addition, norms of the Hessians of $F_p(\mathbf{s})$ and $F_q(\mathbf{s})$ are bounded since the first and second derivatives of the utilities are bounded, the load $\mathbf{s} \succ 0$ at every iteration t, and the connection matrix $\mathbf{G}$ is primitive. This implies the Lipschitz continuity property [18] on $F_p(\mathbf{s})$ and $F_q(\mathbf{s})$ and the existence of a non-zero step size δ such that $\mathcal{L}(\boldsymbol{\gamma}[t+1], \boldsymbol{\nu}) > \mathcal{L}(\boldsymbol{\gamma}[t], \boldsymbol{\nu})$ where $\mathcal{L}$ represents $\mathcal{L}_p$ or $\mathcal{L}_q$. With Lipschitz property, the descent lemma in [18] implies convergence to the Lagrangian maximum.

The update for the price vectors $\boldsymbol{\nu}$ corresponds to the gradient for the Lagrange multipliers if we allow the convergence to the Lagrangian maxima over $\mathbf{s}$ at every price $\boldsymbol{\nu}[t]$. Of course, in practice, both the load factor $\mathbf{s}$ and price $\boldsymbol{\nu}$ are updated simultaneously. Such a joint update can be proven to converge to the optimum of the original constrained optimization problem if the Lagrangian is concave and the maximized Lagrangian for a given price is convex [19]. This is indeed true and therefore the algorithm converges to the optimal.

Consider now the interference constrained case. If $\{\mathbf{s}^*, \boldsymbol{\nu}^*\}$ is a fixed point of the algorithm, then the corresponding spillage $\mathbf{r}^* = \mathbf{G}^T(\mathbf{s}^* + \boldsymbol{\nu}^*)$ and SIR $\{\boldsymbol{\gamma}^* = \mathbf{s}^*/\mathbf{r}^*\}$ satisfy $\Delta\mathbf{s}^* = 0$ so that

$$U_i'(\gamma_i^*) = r_i^* q_i^*(\boldsymbol{\gamma}^*), \qquad i = 1, \ldots, M. \tag{45}$$

This is precisely the KKT condition at $\{\boldsymbol{\gamma}^*, \nu^* = 1\}$ for the optimization problem (40) as shown in Appendix E. Since the optimization problem (40) is convex after a change of variable, any point that satisfies the KKT condition is indeed the optimal solution to the problem [17]. ■

VI. Simulation Results

To illustrate our distributed power control algorithms for utility-optimal SIR assignment, we simulate the load control algorithm in a cellular network, using a simplified version of the realistic 3GPP uplink evaluation model [20] that is adopted by the wireless industry.

The model consists of 19 cells arranged in a three ring hexagonal structure as shown in Fig. 3. Each cell is divided into three identical 120 degree sectors for a total of 57 base station sectors. The BS angular antenna sectorization pattern is based on the commercial Decibel DB 932DG65T2E antenna, which has a 65 degree 3 dB bandwidth, 15 dB antenna gain, and 20 dB front-to-back rejection. This antenna, used in the PCS 1.9 GHz band, is typical of commercial antennae used in cellular wireless networks. The mobiles use single omni-directional antennae. We adopt the path loss model with a path loss exponent of

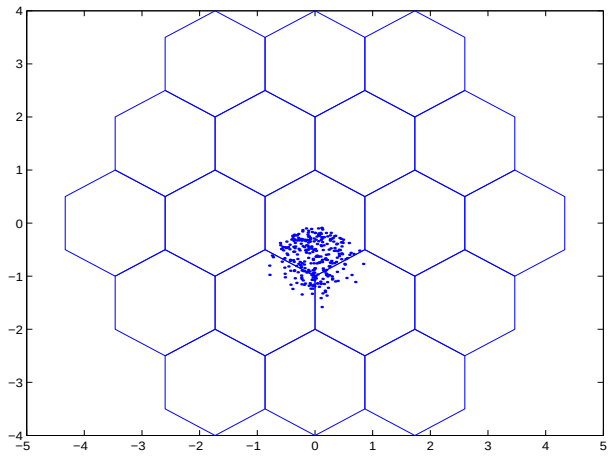

Fig. 3. Hexagonal cellular network for the uplink simulation consisting of 19 cells with wrap around. Cells are three-way sectorized for a total of 57 base station sectors. Path loss accounts for log-normal shadowing, and angular antenna pattern. Plotted are the location of 300 random mobiles connected to the lower sector of the center cell. Cell radii are normalized to unity.

3.7 and log-normal shadowing of 8.9 dB. To avoid edge effects, the sector boundaries at the edge of the hexagonal structure are "wrapped around" to the opposite cell. Mobiles are distributed uniformly in the 19 cell region and then connect to the base station sector to which the path loss is minimum. Fig. 3 shows the location of 300 random mobiles connected to the lower sector of the middle cell under this path model.

In all the simulations below, each realization of the network consists of 10 randomly selected mobiles in each base station sector, for a total of 570 mobiles. As a measure of the QoS, we will use the capacity formula (31), where we assume that each mobile is allocated equal uplink bandwidth fraction, $w_i/W = 1/10$ where W is the total bandwidth. To represent the sum utility in a meaningful unit, we plot the inverse of the average of utilities of QoS factors, which is a quantity whose units are the same as the QoS factor and is denoted as "QoS factor" on the y-axis in Figures 4 and 5. For example, the QoS factor thus computed for logarithmic utility is the geometric mean of the user QoS factors:

$$\exp\left[\frac{1}{M}U(\boldsymbol{\gamma})\right] = \exp\left[\frac{1}{M}\sum_{i=1}^{M}\log\beta_i(\gamma_i)\right] = \prod_{i=1}^{M}\beta_i(\gamma_i)^{1/M}.$$

We first illustrate the convergence of the distributed algorithm, Algorithm 3, using the logarithmic utility defined in Section IV. Fig. 4 shows the convergence of the distributed algorithm as a function of the iteration number for a particular distribution of the MSs. The algorithm is initialized with a random positive load vector $\mathbf{s}$, and the step size is taken as $\delta_1 = 0.1$. The ROT limit, $-10\log(1-\rho)$ is set to 10 dB. Also plotted is the utility which was numerically attained by centralized computation. It can be seen that the utility from the distributed algorithm is almost identical to the maximal utility within about 30 iterations.

In Fig. 5, we plot the convergence of the algorithm with various utility functions. As expected, the pseudo-linear utility converges to the maximum QoS value while the α-fair utility with $\alpha = 3$ converges to the minimum. The difference between the pseudo-linear and logarithmic utility is minimal since this experiment is run under heavy loaded scenario.

TABLE II

SECTOR CAPACITY AND FAIRNESS UNDER VARIOUS UTILITY FUNCTIONS.

Utility function	Sector capacity (bps / Hz / sector)	10% User capacity (bps / Hz)
Pseudo-Linear	1.77	0.054
$\alpha = 1$ (Log)	1.76	0.057
$\alpha = 2$	1.56	0.076
$\alpha = 3$	1.45	0.086

Fig. 6 shows the cumulative distribution of the user capacities after the final iteration of Algorithm 4. As discussed earlier, one of the important features of the utility approach is that this QoS distribution can be varied by appropriately adjusting the utility. Fig. 6 shows the resulting QoS distribution for the pseudo-linear utility, log utility and the α-fair utilities with $\alpha = 2$ and $\alpha = 3$, all with the same realization of MS distribution in the network. It can be seen that the variance in the mobile's QoS with the $\{2, 3\}$-fair distributions is significantly smaller than with the logarithmic and pseudo-linear utility functions. Consequently, fewer mobiles are seen with either very high or very low achievable capacities.

As is widely known in the research literature and discussed in Section IV, the coefficient α in the α-fair exponential distribution can be seen as a fairness parameter, with larger α resulting in greater equalization of allocated QoS. However, in general, while increasing fairness improves the performance of worst case mobiles, it tends to reduce the average sector capacity. To illustrate this fairness-efficiency tradeoff in wireless cellular networks, Table II shows the average sector capacity and performance of the 10% worst mobiles under the various utility functions. It can be seen that average sector capacity drops from 1.77 to 1.46 bps/Hz/sector when we go from pseudo-linear utility to α-fair utility with $\alpha = 3$. However, the QoS for the 10% worst mobile increases from 0.054 to 0.086 bps/Hz. By appropriately adjusting the utility function, the network operator can tradeoff fairness and total sector capacity.

Finally, we simulated Algorithm 4 with interference constraints. We assumed that the noise vector η_i was constant for all mobiles, and the interference constraint was taken as $\mathbf{q}^m = \kappa\boldsymbol{\eta}$, where $\kappa > 1$ is the ROT limit. We used the logarithmic utility, and ran Algorithm 4 with step sizes $\delta_1 = 0.1$ and $\delta = 0.01$.

Fig. 7 shows the convergence of the algorithm for various desired ROT levels for logarithmic utility function. Plotted is the target ROT and the actual maximum ROT over the 57 BS sectors after 25 iterations of the algorithm. The algorithm converges to within a fraction of a dB of the target within the 25 iterations. Fig. 8 plots the convergence of the interference over time for the logarithmic utility and the 2-fair utility at a ROT level of 10dB. The convergence is slower for the fairer utility function but still, we notice convergence within 40 iterations. Fig. 9 shows the

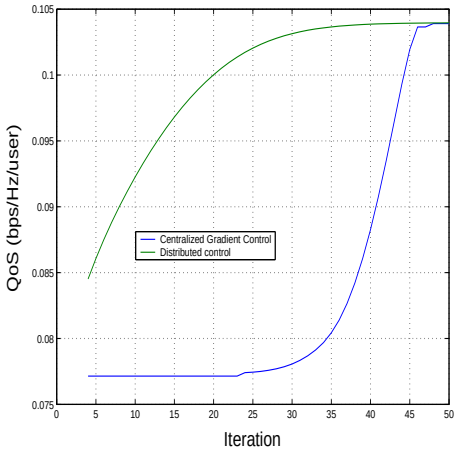

Fig. 4. Convergence of Algorithm 4 with the logarithmic utility. The simulation is based on one realization of the cellular network in Fig. 3 with 10 mobiles per sector for a total of 570 mobiles. The logarithmic utility is plotted as the geometric mean user capacity.

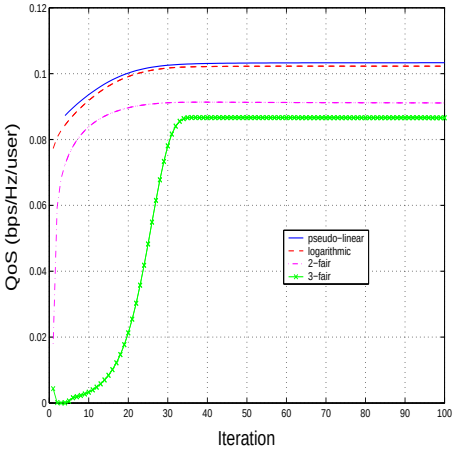

Fig. 5. Convergence of utility functions for different utilities. Pseudo-linear utility functions attains the maximum utility and the 3-fair utility attains the minimum.

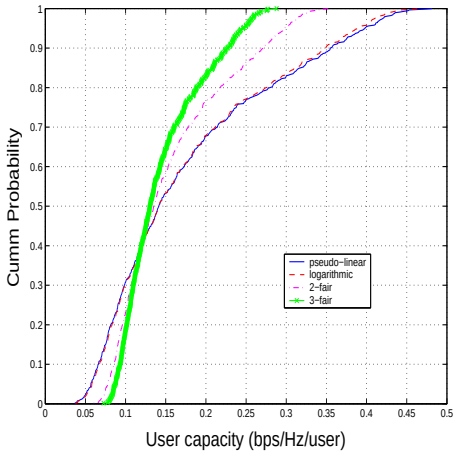

Fig. 6. User capacity distribution after 50 iterations of the distributed algorithm under different utility functions: α-fair distribution with $\alpha = 1, 2, 3$, and the pseudo-linear utility. The final distribution is based on one realization of randomly located mobiles, with 10 mobiles per sector.

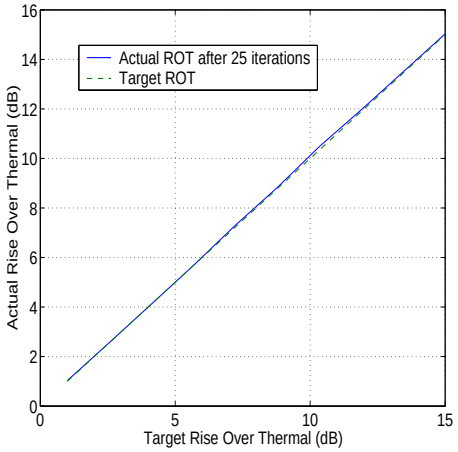

Fig. 7. Convergence with interference constraints expressed as ROT limits. Plotted is the desired and actual maximum ROT level.

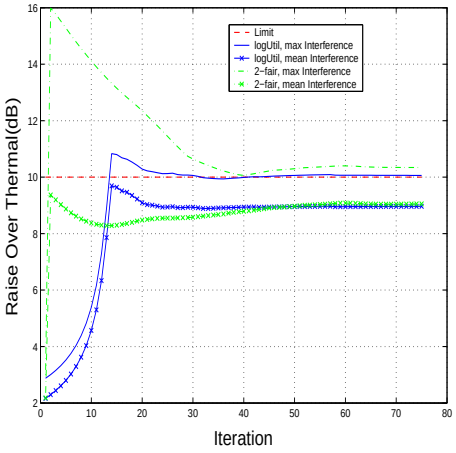

Fig. 8. Convergence of the interference expressed as ROT for different utilities.

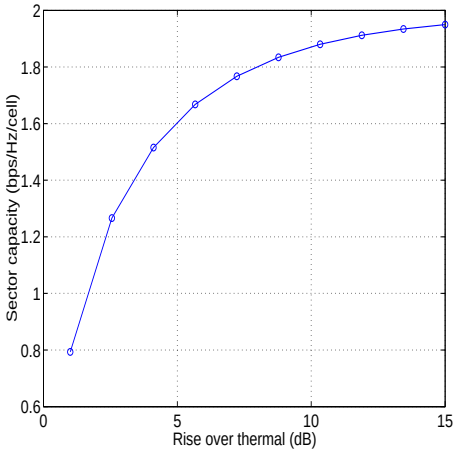

Fig. 9. Mean sector capacity as a function of ROT limit.

mean sector capacity achieved with the distributed algorithm as a function of the ROT limit, κ. The graph quantifies the intuition that raising the ROT limit increases the maximum spectral efficiency, with a diminishing marginal impact. In this way, the network can be configured to tradeoff ROT level (which affects the absolute power usage of the mobiles) and the overall spectral efficiency.

We see from these simulations that, in a practical cellular model with a large number of base station sectors, the distributed power control algorithms are stable and converge to the globally optimal SIR solution within a small number of iterations. Moreover, the algorithms can be applied with a general utility function. This flexibility provides network operator several adjustable knobs to trade-off various considerations such as fairness, total throughput, loading effects, differentiated user QoS and ROT levels.

VII. Conclusion

We solved the following problem in this paper: distributed uplink power control over multiple cells for globally optimal SIR assignment. This covers a variety of earlier results since the early 1990s as special cases, including power control methods for fixed target SIR, for suboptimal Nash equilibrium in SIR optimization, and for optimal SIR assignment obtained through centralized computation. The first key step is the load-spillage characterization (which provides a complementary angle to the traditional power-interference view on the feasible SIR region and its Pareto-optimal boundary), and the second key step is the locally computable ascent direction to "slide" along the Pareto-optimal SIR boundary. We also quantify the intuition that a jointly optimal SIR assignment and power control should take into consideration both spillage and interference for each link, with the spillage being a measure of the emitted interference.

This characterization enables us to decouple the optimization problem with complex coupling across all MS in many cells.

This work clears the bottleneck of "global coupling" towards solving the general resource allocation problem for multicellular uplink in wireless cellular networks. The algorithms can be readily implemented in commercial systems such as the Flash-OFDM system [21], and the convergence, optimality, and efficiency-fairness tradeoff properties of the algorithms verified through 3GPP evaluation tools.

The second bottleneck of "non-convexity" remains a difficult challenge, especially when antenna beamforming, base station assignment (due to mobility), and opportunistic scheduling (under fast fading) are performed jointly with power control for optimal SIR assignment. Some of these additional degrees of freedom have recently been incorporated [13] into the framework of distributed solution developed in this paper.

Appendix A: Gradient Evaluation

We state the following results on gradients, skipping the proof due to space limitation. If $\rho(\mathbf{D}(\boldsymbol{\gamma})\mathbf{G})$ is the spectral radius $\mathbf{D}(\boldsymbol{\gamma})\mathbf{G}$, then

$$\frac{\partial \rho}{\partial \gamma_i} = \mathbf{r}^T \mathbf{D}(\mathbf{e}_i)\tilde{\mathbf{q}} \tag{46}$$

where $\mathbf{D}(\mathbf{e}_i)$ has zero everywhere, except the ith diagonal element which is 1, $\mathbf{r}$ is the left eigenvector of $\mathbf{D}(\boldsymbol{\gamma})\mathbf{G}$, and $\tilde{\mathbf{q}}$ is the right eigenvector of $\mathbf{GD}(\boldsymbol{\gamma})$. If $\mathbf{p}(\boldsymbol{\gamma})$ and $\mathbf{q}(\boldsymbol{\gamma})$ are given according to (8) and (9), then

$$\frac{\partial \mathbf{p}}{\partial \boldsymbol{\gamma}} = (\mathbf{I} - \mathbf{D}(\boldsymbol{\gamma})\mathbf{G})^{-1}\mathbf{D}(\mathbf{q}) \tag{47}$$

$$\frac{\partial \mathbf{q}}{\partial \boldsymbol{\gamma}} = (\mathbf{I} - \mathbf{GD}(\boldsymbol{\gamma}))^{-1}\mathbf{GD}(\mathbf{q}). \tag{48}$$

If $\boldsymbol{\gamma} = \boldsymbol{\gamma}(\mathbf{s}, \rho)$ for some fixed ρ, then

$$\frac{\partial \boldsymbol{\gamma}}{\partial \mathbf{s}} = \mathbf{D}(1/\mathbf{r})(\rho\mathbf{I} - \mathbf{GD}(\boldsymbol{\gamma}))^T. \tag{49}$$

If $\boldsymbol{\gamma} = \boldsymbol{\gamma}^q(\mathbf{s}, \boldsymbol{\nu})$ for some fixed vector $\boldsymbol{\nu}$, then

$$\frac{\partial \boldsymbol{\gamma}}{\partial \mathbf{s}} = \mathbf{D}(1/\mathbf{r})(\mathbf{I} - \mathbf{GD}(\boldsymbol{\gamma}))^T. \tag{50}$$

Appendix B: Pareto-Optimality Proof

We first consider $\partial\mathbf{B}_\rho$. From (46), we have $\partial\rho/\partial\gamma_i = r_i\tilde{q}_i$. Since $\mathbf{GD}(\boldsymbol{\gamma})$ and $\mathbf{D}(\boldsymbol{\gamma})\mathbf{G}$ are primitive, and primitive matrices have positive left and right eigenvectors associated with the simple eigenvalue ρ, $r_i > 0, \tilde{q}_i > 0$, we have $\partial\rho/\partial\gamma_i > 0$. This implies that the Pareto-optimal boundary of $\mathbf{B}_\rho$ is indeed given by $\rho(\mathbf{GD}(\boldsymbol{\gamma})) = \rho$.

Next, we prove the result for $\mathbf{B}(\mathbf{p}^m)$. Suppose $\boldsymbol{\gamma}$ is a SIR assignment that satisfies the Pareto-optimal constraint given in the theorem but is not Pareto-optimal, then γ_i cannot be increased without decreasing any other SIR assignment since p_i is already at its maximum. So let $\gamma_j, j \neq i$ be a SIR assignment that can be increased, the primitivity assumption on $\mathbf{G}$ implies that the network is fully connected, so p_i has to increase to maintain the same γ_i. This is impossible, thus proving the claim by contradiction. The proof for $\mathbf{B}(\mathbf{q}^m)$ follows similarly.

Appendix C: Proof of Lemma 5

For part (a), let $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}(\mathbf{s}, 1)$ and let $\tilde{\mathbf{q}}$ be a positive right eigenvector with $\mathbf{GD}(\boldsymbol{\gamma}_1)\tilde{\mathbf{q}} = \tilde{\mathbf{q}}$. We can normalize $\tilde{\mathbf{q}}$ so that $\mathbf{s}^T\tilde{\mathbf{q}} = \mathbf{s}^T\boldsymbol{\eta}$. For $\rho < 1$, (12) shows that $\boldsymbol{\gamma}(\mathbf{s}, \rho) = \rho\boldsymbol{\gamma}_1$, hence $\mathbf{D}(\boldsymbol{\gamma}(\mathbf{s}, \rho)) = \rho\mathbf{D}(\boldsymbol{\gamma}_1)$. Consequently,

$$\mathbf{GD}(\boldsymbol{\gamma}(\mathbf{s}, \rho))\tilde{\mathbf{q}} = \rho\mathbf{GD}(\boldsymbol{\gamma}_1)\tilde{\mathbf{q}} = \rho\tilde{\mathbf{q}},$$

so $\tilde{\mathbf{q}}$ is a positive right eigenvector of $\mathbf{GD}(\boldsymbol{\gamma}(\mathbf{s}, \rho))$.

Part (b) follows from the technique used in the expansion of $(\mathbf{I} - \mathbf{GD}(\boldsymbol{\gamma}))^{-1}$ in [15]. From [15], we can write $(\mathbf{I} - \mathbf{GD}(\boldsymbol{\gamma}))^{-1}$ as

$$(\mathbf{I} - \mathbf{GD}(\boldsymbol{\gamma}))^{-1} = \frac{1}{1-\rho}\mathbf{Z}_{1,1} + \sum_{k=2}^{n}\sum_{j=1}^{m_k}\frac{(j-1)!}{(1-\lambda_k)^j}\mathbf{Z}_{k,j} \tag{51}$$

where n is the number of distinct eigen-values, m_k is the algebraic multiplicity of eigen value λ_k and $\mathbf{Z}_{k,j}$ are known as the principal component matrices. Further, we have from [15] that

$$\mathbf{Z}_{1,1} = \tilde{\mathbf{q}}\mathbf{s}^T/(\mathbf{s}^T\boldsymbol{\eta}) \tag{52}$$

Since $\mathbf{q}_\rho = (\mathbf{I} - \mathbf{GD}(\boldsymbol{\gamma}))^{-1}\boldsymbol{\eta}$, Lemma 5(b) follows.

Appendix D: Proof of Ascent Direction

Consider the functions

$$\begin{aligned} F_\rho(\mathbf{s}) &= U(\boldsymbol{\gamma}), & \mathbf{s}^T\mathbf{GD}(\boldsymbol{\gamma}) &= \rho\mathbf{s}^T \\ F_q(\mathbf{s}) &= U(\boldsymbol{\gamma}) + \boldsymbol{\nu}^T(\mathbf{q}^m - \mathbf{q}(\boldsymbol{\gamma})), & (\mathbf{s}^T + \boldsymbol{\nu}^T)\mathbf{GD}(\boldsymbol{\gamma}) &= \mathbf{s}^T \\ F_p(\mathbf{s}) &= U(\boldsymbol{\gamma}) + \boldsymbol{\nu}^T(\mathbf{p}^m - \mathbf{p}(\boldsymbol{\gamma})), & (\mathbf{s}^T\mathbf{G} + \boldsymbol{\nu}^T)\mathbf{D}(\boldsymbol{\gamma}) &= \mathbf{s}^T, \end{aligned} \tag{53}$$

based on which we define

$$J_\rho(\mathbf{s}) = \nabla F_\rho^T\Delta\mathbf{s}, J_q(\mathbf{s}) = \nabla F_q^T\Delta\mathbf{s}, J_p(\mathbf{s}) = \nabla F_p^T\Delta\mathbf{s} \tag{54}$$

where $\Delta\mathbf{s}$ is given by (36). Let $F(\mathbf{s})$ represent any one of the functions defined in (53) and $J(\mathbf{s})$ represent the corresponding function in (54). We need to prove that $\Delta\mathbf{s}$ is an ascent direction for $F(\mathbf{s})$. To prove this, we need to show that for all $\mathbf{s} > 0$, $J(\mathbf{s}) \geq 0$ with $J(\mathbf{s}) = 0$ if and only if $\nabla F(\mathbf{s}) = 0$. To this end, we first state the following Lemma, the proof of which is omitted due to space limitation.

Lemma 6—Positivity Lemma*:* Suppose $\mathbf{T}$ is a matrix with positive entries, i.e. $T_{ij} > 0$ for all i and j.

(a) Suppose that $\mathbf{q} > 0$ and $\mathbf{s} > 0$ are vectors satisfying,

$$\mathbf{Tq} = \rho\mathbf{q}, \quad \mathbf{T}^T\mathbf{s} = \rho\mathbf{s}, \tag{55}$$

for some eigenvalue $\rho > 0$. Define the function,

$$H(\mathbf{x}) = \mathbf{x}^T\mathbf{D}(\mathbf{s})(\rho\mathbf{I} - \mathbf{T})\mathbf{D}(\mathbf{q})\mathbf{x}. \tag{56}$$

Then, $H(\mathbf{x}) \geq 0$ for all vectors $\mathbf{x}$, with $H(\mathbf{x}) = 0$ if and only if $\mathbf{x} = \beta\mathbf{1}$ for some scalar β.

(b) Suppose that $\mathbf{q} > 0$ and $\mathbf{s} > 0$ are vectors satisfying,

$$\mathbf{T}\mathbf{q} < \rho\mathbf{q}, \quad \mathbf{T}^T\mathbf{s} < \rho\mathbf{s}, \tag{57}$$

for some $\rho > 0$. Then, $H(\mathbf{x}) \geq 0$ for all vectors $\mathbf{x}$, with $H(\mathbf{x}) = 0$ if and only if $\mathbf{x} = 0$.

Continuing with the proof, we first consider $F = F_q$ and set $\mathbf{T} = \mathbf{G}\mathbf{D}(\boldsymbol{\gamma})$. We now evaluate $\partial F(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}$. We let $U'(\boldsymbol{\gamma}) = \partial U/\partial\boldsymbol{\gamma}$ and use result from (48) to derive that

$$\begin{aligned}(\partial F/\partial\boldsymbol{\gamma})^T &= U'(\boldsymbol{\gamma})^T - \boldsymbol{\nu}^T\frac{\partial\mathbf{q}}{\partial\boldsymbol{\gamma}} \\ &= U'(\boldsymbol{\gamma})^T - \boldsymbol{\nu}^T(\mathbf{I}-\mathbf{T})^{-1}\mathbf{G}\mathbf{D}(\mathbf{q}). \end{aligned} \tag{58}$$

Now

$$\mathbf{r} = \mathbf{G}^T(\mathbf{s}+\boldsymbol{\nu}) = \mathbf{G}^T(\mathbf{D}(\boldsymbol{\gamma})\mathbf{r}+\boldsymbol{\nu}),$$

and consequently,

$$\begin{aligned}\mathbf{r} &= (\mathbf{I}-\mathbf{G}^T\mathbf{D}(\boldsymbol{\gamma}))^{-1}\mathbf{G}^T\boldsymbol{\nu} \\ &= \mathbf{G}^T(\mathbf{I}-\mathbf{D}(\boldsymbol{\gamma})\mathbf{G}^T)^{-1}\boldsymbol{\nu} = \mathbf{G}^T(\mathbf{I}-\mathbf{T}^T)^{-1}\boldsymbol{\nu}.\end{aligned}$$

Substituting this identity into (58), we see that

$$(\partial F/\partial\boldsymbol{\gamma})^T = U'(\boldsymbol{\gamma})^T - \mathbf{r}^T\mathbf{D}(\mathbf{q}). \tag{59}$$

Now combining (59) with (50), we have

$$\begin{aligned}\nabla F^T(\mathbf{s}) &= (\partial F/\partial\boldsymbol{\gamma})^T\frac{\partial\boldsymbol{\gamma}}{\partial\mathbf{s}} \\ &= (U'(\boldsymbol{\gamma}) - \mathbf{r}\mathbf{D}(\mathbf{q}))^T\mathbf{D}(1/\mathbf{r})(\mathbf{I}-\mathbf{T})^T \\ &= (\frac{U'(\boldsymbol{\gamma})}{\mathbf{r}\mathbf{q}} - \mathbf{1})^T\mathbf{D}(\mathbf{q})(\mathbf{I}-\mathbf{T})^T \\ &= \mathbf{x}^T\mathbf{D}(\mathbf{q})(\mathbf{I}-\mathbf{T})^T, \end{aligned} \tag{60}$$

where

$$\mathbf{x} = \frac{U'(\boldsymbol{\gamma})}{\mathbf{r}\mathbf{q}} - \mathbf{1}. \tag{61}$$

Also, since $\gamma_i = s_i/r_i$, we can rewrite Δs_i as

$$\Delta s_i = \frac{U_i'(\gamma_i)\gamma_i}{q_i} - s_i = s_i\left(\frac{U_i'(\gamma_i)}{q_i r_i} - 1\right) = s_i x_i,$$

or in vector notation,

$$\Delta\mathbf{s} = \mathbf{D}(\mathbf{s})\mathbf{x}. \tag{62}$$

Substituting (60) and (62) into (54),

$$J(\mathbf{s}) = \mathbf{x}^T\mathbf{D}(\mathbf{q})(\mathbf{I}-\mathbf{T})^T\mathbf{D}(\mathbf{s})\mathbf{x}.$$

Now observe that,

$$\mathbf{s} = \mathbf{D}(\boldsymbol{\gamma})\mathbf{r} = \mathbf{D}(\boldsymbol{\gamma})\mathbf{G}^T(\mathbf{s}+\boldsymbol{\nu}) = \mathbf{T}^T(\mathbf{s}+\boldsymbol{\nu}) > \mathbf{T}^T\mathbf{s},$$

and note that

$$\mathbf{T}\mathbf{q} = \mathbf{q} - \boldsymbol{\eta} < \mathbf{q}.$$

Therefore, $\mathbf{s}$ and $\mathbf{q}$ satisfy (57) in Lemma 6(b). Hence, $J(\mathbf{s}) \geq 0$ for all $\mathbf{s}$. Moreover, $J(\mathbf{s}) = 0$ if and only if $\mathbf{x} = 0$. But, using (61), $\mathbf{x} = 0$ if and only if,

$$U_i'(\gamma_i) = r_i q_i.$$

From (60), this condition is equivalent to $\partial F/\partial\boldsymbol{\gamma} = 0$. We conclude that

$$J(\mathbf{s}) = \nabla F^T\Delta\mathbf{s} \geq 0,$$

with equality if and only if $\nabla F(\mathbf{s}) = 0$. A similar computation proves the result for $F = F_p$.

For $F = F_\rho$,

$$(\partial F/\partial\boldsymbol{\gamma})^T = U'(\boldsymbol{\gamma})^T \tag{63}$$

Now combining (63) with (49),

$$\begin{aligned}\nabla F^T(\mathbf{s}) &= (\partial F/\partial\boldsymbol{\gamma})^T\frac{\partial\boldsymbol{\gamma}}{\partial\mathbf{s}} \\ &= (U'(\boldsymbol{\gamma})\mathbf{D}(1/\mathbf{r})(\rho\mathbf{I}-\mathbf{T})^T \\ &= \mathbf{x}^T\mathbf{D}(\mathbf{q})(\rho\mathbf{I}-\mathbf{T})^T, \end{aligned} \tag{64}$$

where

$$\mathbf{x} = \frac{U'(\boldsymbol{\gamma})}{\mathbf{r}\mathbf{q}}. \tag{65}$$

Therefore, we have

$$\Delta\mathbf{s} = \mathbf{D}(\mathbf{s})\mathbf{x} - \mathbf{s}. \tag{66}$$

Substituting (64) and (66) into (54),

$$J(\mathbf{s}) = \mathbf{x}^T\mathbf{D}(\mathbf{q})(\mathbf{I}-\mathbf{T})^T\mathbf{D}(\mathbf{s})\mathbf{x}.$$

Since $\mathbf{q}$ and $\mathbf{s}$ are, respectively, right and left eigenvectors of $\mathbf{T}$, Lemma 6(a) shows that $J(s) \geq 0$ for all $\mathbf{s}$. Moreover, $J(\mathbf{s}) = 0$ if and only if $\mathbf{x} = \beta\mathbf{1}$ for some scalar constant β. which implies that $\mathbf{x}^T\mathbf{D}(\mathbf{q})(\rho\mathbf{I}-\mathbf{T})^T = \beta\mathbf{q}^T(\rho\mathbf{I}-\mathbf{T})^T = 0$. From (64), this condition is equivalent to $\nabla F(\mathbf{s}) = 0$. We conclude that

$$J(\mathbf{s}) = \nabla F^T\Delta\mathbf{s} \geq 0,$$

with equality if and only if $\nabla F(\mathbf{s}) = 0$.

APPENDIX E: KKT CONDITION

We derive the KKT condition [17] for the utility optimization problems. The Lagrangian for the optimization problem (35) is given by

$$\mathcal{L}(\boldsymbol{\gamma},\nu) = U(\boldsymbol{\gamma}) - \nu(\rho(\mathbf{D}(\boldsymbol{\gamma})\mathbf{G}) - \rho) \tag{67}$$

with $\nu \geq 0$ being the Lagrange multiplier. The KKT condition is given by $\nabla\mathcal{L} = 0$ which implies that $\nabla U = \nu\nabla\rho(\mathbf{D}(\boldsymbol{\gamma})\mathbf{G})$. From (46), this can be written as

$$U_i'(\gamma_i) = \nu r_i(\boldsymbol{\gamma})\tilde{q}_i(\boldsymbol{\gamma}) \tag{68}$$

where $\mathbf{r}$ is the left eigenvector of $\mathbf{D}(\boldsymbol{\gamma})\mathbf{G}$ and $\tilde{\mathbf{q}}$ is the right eigenvector of $\mathbf{G}\mathbf{D}(\boldsymbol{\gamma})$.

The Lagrangians for the power and interference constrained cases are given in (41) and (42) respectively. The KKT condition can be written as $\nabla U = \boldsymbol{\nu}^T \nabla(\mathbf{p})$, which using (47) and (48) respectively, we simplify to

$$U_i'(\gamma_i) = r_i(\boldsymbol{\gamma}) q_i(\boldsymbol{\gamma}) \tag{69}$$

where $\mathbf{q}$ is given by (9) and $\mathbf{r} = \mathbf{r}^p$ as given by (25) for the power constrained case and $\mathbf{r} = \mathbf{r}^q$ as given by (26) for the interference constrained case.

ACKNOWLEDGMENTS

The authors would like to acknowledge very helpful discussions with Arnab Das at Qualcomm Flarion Technologies.

REFERENCES

[1] P. Hande, S. Rangan, M. Chiang, "Distributed Uplink Power Control for Optimal SIR Assignment in Cellular Data Networks", *IEEE INFOCOM 2006*, April 2006.

[2] J. Zander, "Performance of Optimum Transmitter Power Control in Cellular Radio Systems", *IEEE Transactions on Vehicular Technology*, VT-41, no. 1, pp. 57-62, February 1992.

[3] G. J. Foschini and Z. Miljanic, "A Simple Distributed Autonomous Power Control Algorithm and it's Convergence", *IEEE Transactions on Vehicular Technology*, VT-42, no. 4, pp. 641-646, April 1993.

[4] R. D. Yates, "A Framework for Uplink Power Control in Cellular Radio Systems", *IEEE Journal on Selected Areas in Communications*, Vol. 13, no. 7, pp. 1341-1347, September 1995.

[5] N. Bambos, S. C. Chen, G. J. Pottie, "Radio Link Admission Algorithms for Wireless Networks with Power Control and Active Link Quality Protection", *Proc. IEEE INFOCOM* 1995: 97-104

[6] D. Mitra and A. Morrison, "A distributed power control algorithm for bursty transmissions in cellular, spread spectrum wireless networks", *Proc. 5th WINLAB Workshop*, Rutgers University, New Brunswick, NJ, 1995.

[7] C. Saraydar, N. Mandayam, and D. Goodman, "Pricing and power control in a multicell wireless data network", *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 1883-1892, October 2001.

[8] T. Javidi, "Decentralized Rate Assignments in a Multi-Sector CDMA Network", *IEEE Globecom Conference*, December 2003.

[9] M. Chiang and J. Bell, "Balancing supply and demand of wireless bandwidth: Joint rate allocation and power control", *Proc. IEEE INFOCOM*, Hong Kong, China, March 2004.

[10] D. O'Neill, D. Julian, and S. Boyd, "Adaptive management of network resources", *Proc. IEEE VTC*, October 2003.

[11] M. Xiao,N. B. Shroff, E. Chong, "A utility-based power-control scheme in wireless cellular systems", *IEEE/ACM Transactions on Networking*, Volume 11, Issue 2, pp. 210 - 221, April 2003.

[12] H. Boche and S. Stanczak,"Convexity of some feasible QoS regions and asymptotic behavior of the minimum total power in CDMA systems", *IEEE Transactions on Communications*, Volume 52, Issue 12, pp. 2190 - 2197, December 2004.

[13] T. Lan, P. Hande, and M. Chiang, "Joint uplink power control, bandwidth allocation, and beamforming for optimal QoS in wireless networks", Preprint, 2006.

[14] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control", *IEEE/ACM Transactions on Networking*, Volume 8, Issue 5, pp. 556 - 567, October 2000.

[15] E. Deutsch and M. Neumann, "Derivatives of the Perron root at an essentially nonnegative matrix and the group inverse of an M-matrix", Journal of Mathematical Analysis and Applications, vol. I-29, no. 102, 1984.

[16] R. A. Horn, C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[17] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[18] D. P. Bertsekas, *Nonlinear Programming*, 2nd Ed., Athena Scientific, 1999.

[19] R. T. Rockafellar, *Saddle-points and convex analysis,* Differential Games and Related Topics, H. W. Kuhn and G. P. Szego, Eds. North-Holland, 1971.

[20] 3GPP TR 25.896 (2004-02), *Feasibility Study for Enhanced Uplink for UTRA FDD*, Document R1-040346, V1.3.0.

[21] Flash-OFDM, `http://www.qualcomm.com/qft/`

Prashanth Hande Prashanth Hande is a Graduate student at Princeton University and a Staff Engineer at Qualcomm Flarion Technologies. He received the B.Tech in Electrical Engineering at the Indian Institute of Technology, Mumbai in 1998 and his M.S in Electrical Engineering at Cornell University in 2000. Prashanth Hande's interests are in wireless communications, network architectures and boradband access networks.

Sundeep Rangan Dr. Sundeep Rangan is a Director of Engineering at Qualcomm Flarion Technologies. He received the B.A.Sc. in Electrical Engineering at the University of Waterloo, Canada in 1992 and his M.S and Ph.D. in Electrical Engineering at the University of California, Berkeley in 1995 and 1997, respectively. After a postdoctoral research fellowship at the University of Michigan, Ann Arbor, he joined Bell Labs, Lucent Technologies in 1998. At Bell Labs, Dr. Rangan worked on the development of Flash-OFDM, a novel air interface for broadband, cellular wireless. The Bell Labs department was spun off to form a separate company, Flarion Technologies, in October 2000. Dr. Rangan was one of four founding members. Flarion was acquired by Qualcomm in Jan 2006, and Dr. Rangan remains involved in physical layer design and implementation. Dr. Rangan's interests are in wireless communications, information theory, control theory and signal processing.

Mung Chiang Mung Chiang (S'00 – M'03) received the B.S. (Hon.) degree in electrical engineering and in mathematics, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1999, 2000, and 2003, respectively. He is an Assistant Professor of Electrical Engineering, and an affiliated faculty of Applied and Computational Mathematics at Princeton University. He conducts research in the areas of optimization of communication systems, theoretical foundations of network architectures, algorithms in broadband access networks, and information theory.

Professor Chiang has been awarded a Hertz Foundation Fellowship and received the Stanford University School of Engineering Terman Award for Academic Excellence, the SBC Communications New Technology Introduction contribution award, the National Science Foundation CAREER Award, and the Princeton University Howard B. Wentz Junior Faculty Award. He is the Lead Guest Editor of the IEEE J. Sel. Area Comm. special issue on Nonlinear Optimization of Communication Systems, a Guest Editor of the IEEE Trans. Inform. Theory and IEEE/ACM Trans. Networking joint special issue on Networking and Information Theory, and the Program Co-Chair of the 38th Conference on Information Sciences and Systems.

Xinzhou Wu Xinzhou Wu was born in Jiangyin, China, on September 23, 1975. He attended Tsinghua University, Beijing, China, where he received the bachelor of engineering degree in electronics engineering in 1998. He received his M.S. in 2000 and Ph.D. in 2004, both in electrical engineering from the University of Illinois at Urbana-Champaign. For the academic year 2003-2004, he was a recipient of the Vodaphone Fellowship. He was a member of technical staff at Flarion technologies from 2005 to 2006 and he is currently a senior engineer at Qualcomm Flarion technologies. His current research interests are in multiple antenna systems, resource allocation in cellular communication systems, and information theory