Power Control and Capacity of Spread Spectrum Wireless Networks

S.V. Hanly [a,1], and D.N. Tse [b,2]

[a] *Dept. of EE, University of Melbourne, Vic. 3052, Australia, email: s.hanly@ee.mu.oz.au*

[b] *Dept. of EECS, University of California at Berkeley, CA 94720, USA, email: dtse@eecs.berkeley.edu*

Abstract

Transmit power control is a central technique for resource allocation and interference management in spread-spectrum wireless networks. With the increasing popularity of spread-spectrum as a multiple access technique, there has been significant research in the area in recent years. While power control has been considered traditionally as a means to counteract the harmful effect of channel fading, the more general emerging view is that it is a flexible mechanism to provide Quality-of-Service to individual users. In this paper, we will review the main threads of ideas and results in the recent development of this area, with a bias towards issues that have been the focus of our own research. For different receivers of varying complexity, we study both questions about optimal power control as well as the problem of characterizing the resulting network capacity. Although spread-spectrum communications has been traditionally viewed as a physical-layer subject, we argue that by suitable abstraction, many control and optimization problems with interesting structure can be formulated at the network layer.

Key words: Power control, CDMA

1 Introduction

With the introduction of the IS-95 Code-Division Multiple-Access (CDMA) standard ([63]), the use of spread-spectrum as a multiple-access technique in

[1] Supported by an ARC large grant

[2] Supported by a NSF CAREER award, and by the Air Force Office of Scientific Research under grant F49620-96-1-0199.

commercial wireless systems is growing rapidly in popularity. Unlike more traditional methods such as time-division multiple access (TDMA) or frequency-division multiple access (FDMA), spread-spectrum techniques are *broadband* in the sense that the entire transmission bandwidth is shared between all users at all times. This is done by the spreading of the users' signals onto a bandwidth much larger than an individual user's information rate. The advantages of spread-spectrum techniques include simpler statistical multiplexing without explicit scheduling of time or frequency slots, universal frequency reuse between cells, graceful degradation of quality near congestion, and exploitation of frequency-selective fading to avoid the harmful effects of deep fades that afflict narrowband systems.

Since spread-spectrum systems do not explicitly schedule time or frequency slots among the users, the central mechanism for resource allocation and interference management is *power control*. Each user varies its access to the resources by adapting its transmit power to the changing channel and interference conditions. While in the IS-95 standard, power control is used basically as a mechanism to keep the received powers of users equal so that the nearby users do not dominate over the far away users, the more general emerging view is that it is a flexible mechanism to provide different Quality-of-Service (QoS) to users with heterogeneous requirements.

In this survey paper, we would like to review the results on two sets of issues associated with resource allocation and power control problems in spread-spectrum systems:

- What are the appropriate power control algorithms which provide desired QoS requirements while minimizing the power usage?
- How can the capacity of a power controlled spread-spectrum network be characterized?

We will address these questions in two settings.

A typical way the wideband channel is shared among the users is through a spread-spectrum technique called *direct sequence CDMA* (DS-CDMA), where each user's information symbols are spread over the wideband channel by its unique *signature sequence*. To discriminate among the users, receivers of varying complexity can be implemented at the base-station. We will focus on the class of *linear receivers*, i.e. receivers that operate linearly on the total received signal to demodulate the symbols of a particular user. The simplest such receiver is the one which is matched to the signature sequence of the desired user. This matched filter receiver is the receiver used in the IS-95 standard. *Multiuser receivers* are more sophisticated receivers which take into account the signature sequences of the interfering users as well, thereby providing a better interference suppression capability.

For these linear receivers, we describe decentralized *power control algorithms* that enable resources to be rapidly reallocated among users without the need for centralized decision making. These algorithms are iterative in nature, and converge to the unique minimal possible power allocation that satisfies the demands of the users in the network. We will also show that these systems are all *interference-limited*, in the sense that the user capacity is limited even when there are no power constraints. We will present a unifying framework to characterize and compare network capacity for different receivers based on two notions:

- *effective interference*: a measure of the effect an interferer has on the desired user.
- *effective bandwidth*: a measure of the amount of network resources consumed by each user in the system;

The formulas for effective interference and effective bandwidths depend on the linear receivers. Thus, these notions serve as a unifying basis for performance comparison in this class of receivers. One can also think of the effective bandwidth of a user in a spread spectrum system as the analog of the number of time or frequency slots used by a user in a narrowband system. However, because the signals of users in a DS-CDMA system are superimposed on each other and discriminated by signal processing techniques, the existence of such notions is non-trivial, and indeed hold only under certain conditions which will be specified.

The above results are in the more practical setting of linear receivers for DS-CDMA systems. In the second setting, we take a more speculative point of view and ask a more fundamental question: what are the optimal power control and resource allocation schemes for spread-spectrum receivers that are *information theoretically* optimal? For the single-cell scenario, a complete answer to this question is described. Using multiuser information theory, the power control problem is formulated as the optimization of certain objective functions of the transmit powers subject to constraints imposed by the desired performance targets. By identifying *polymatroid* structure in the constraints, explicit greedy solutions to the optimal power allocation problems are derived. The solutions provide a nice contrast to the corresponding results for linear receivers, as they are very different in flavor.

The problem of power control in wireless networks has received much attention in recent years, and our survey here is by no means exhaustive. For a survey on power control that focusses on narrowband wireless networks we refer the reader to [7]. In the present paper, we focus on spread spectrum wireless systems, and the bias is towards the issues that we ourselves have studied in the past, and thus the paper is very much shaped by our outlook on the field.

While power control and spread-spectrum communication are traditionally thought of as physical layer subjects, we will show that many interesting network layer resource allocation problems require an understanding of the underlying spread spectrum physical layer that gives rise to optimization problems involving power control. Although we consider a variety of different resource allocation problems, we find two fundamental principles that are common to all: *monotonicity* and *conservation laws*. Monotonicity is crucial to the proofs of all convergence algorithms, and reflects the basic fact that if one user increases its share of the available network resources, then the remaining users obtain a smaller share of the resources. A stronger form of monotonicity is manifested in the other principle of resource conservation: there is always a total amount of network "resource", which can be shared out in various ways, but the total amount is fixed. The subtlety is in the appropriate definition of "resource", and this varies from problem to problem.

The rest of the paper is structured as follows. In Section 2, we review results on power control and capacity under the standard matched-filter CDMA receiver. We will then consider corresponding questions for linear multiuser receivers in Section 3. We then turn to power control problems for information theoretically optimal receivers in Section 4. Section 5 contains our conclusions and some open problems.

2 Conventional Matched Filter Receiver

2.1 Basic Model

In a spread-spectrum system, each of the user's information or coded symbols is spread onto a much larger bandwidth via modulation by its own *signature* or *spreading sequence*. The following is a sampled discrete-time model for a symbol-synchronous multi-access spread-spectrum system:

$$\mathbf{y} = \sum_{i=1}^{M} X_i \mathbf{s}_i + \mathbf{w}, \tag{1}$$

where $X_i \in \Re$ and $\mathbf{s}_i \in \Re^L$ are the transmitted symbol and signature spreading sequence of user i respectively, and $\mathbf{w}$ is $N(0, \sigma^2 I)$ background Gaussian noise. The length of the signature sequences is L, which gives the spreading ratio between the rate of narrowband information symbols (the X_i's) and the rate of the wideband spread-spectrum signals (the $X_i\mathbf{s}_i$); L is sometimes called the processing gain. The received vector is $\mathbf{y} \in \Re^L$. We assume the X_i's are independent and that $E[X_i] = 0$ and $E[X_i^2] = p_i$, where p_i is the received power of user i. Each sample is sometimes called a *chip*.

Rather than looking at *symbol-by-symbol* detection, we are interested in the more general problem of *demodulation*, extracting good estimates of the (coded) symbols of each user as soft decisions to be used by the channel decoder [57]. From this point of view, a relevant performance measure is the signal-to-interference ratio (SIR) of the estimates, which can be taken as a Quality-of-Service measure for the user. Strictly speaking, the SIR does not completely characterize performance such as bit error probability, since the interference from other users is not necessarily Gaussian. However, it is found in practice to be a reasonable measure, and its use is further justified rigorously in [76], [77], [96] for a large system with many interferers, using the Central Limit Theorem.

For convenience, let us focus on the demodulation of user 1's symbols, and the calculation of its SIR. The same approach can be taken to study the performance of any other user. The *conventional* CDMA receiver for demodulating user 1 is to perform the matched filtering $\mathbf{s}_1 \cdot \mathbf{y}$ on the received signal $\mathbf{y}$. This despreads the signal of user 1, inverting the original spreading operation at the transmitter, and results in the effective channel:

$$X_1 \to X_1(\mathbf{s}_1 \cdot \mathbf{s}_1) + \sum_{i=2}^{M} X_i \mathbf{s}_i \cdot \mathbf{s}_1 + \mathbf{s}_1 \cdot \mathbf{w}$$

The SIR for user 1 is the ratio of user 1's signal energy to that of the noise plus other users' interference at the output of the matched filter, and is given by: signal

$$\mathrm{SIR}_1 = \frac{(\mathbf{s}_1 \cdot \mathbf{s}_1)^2 p_1}{(\mathbf{s}_1 \cdot \mathbf{s}_1)\sigma^2 + \sum_{u=2}^{M} (\mathbf{s}_1 \cdot \mathbf{s}_u)^2 p_u} \tag{2}$$

In the context of this conventional receiver, the basic questions of power control and network capacity can be concretely stated as:

- Given a set of users with desired SIR requirements, does there exist transmit powers such that the requirements are met? If so, how can the powers be controlled?
- How do we characterize the number of users whose SIR requirements can be simultaneously met via appropriate power control?

2.2 Effective Interference

A natural question at this point is the choice of the signature sequences $\{\mathbf{s}_i\}$. To avoid interference, it is easily seen that the sequences can be chosen to be *orthogonal* to each other. In practice, this is usually not possible in an uplink CDMA system for several reasons. First, the underlying physical wireless channel may cause *multipath* distortion to the transmitted signal, such that several delayed replicas of the signal is superimposed together at the receiver. Hence, even if the transmitted signature sequences were chosen to be orthogonal, the received signatures would not be. Second, uplink CDMA systems are usually *asynchronous*, which means that there is a random relative delay between users so that a symbol of a user overlaps with two partial symbols of an interferer. Third, there may be more users than the processing gain L.

Rather than having a detailed model of these physical layer phenomena, we will stick to the simple synchronous channel (1) but assume that the signature sequences are *randomly* and independently chosen so as to capture the uncoordinated nature of spread-spectrum systems. In fact, practical CDMA systems often employ *pseudonoise* sequences which are a very close approximation to true random sequences (see for e.g. [80, Chapter 2]) for which our model is appropriate.

Under random spreading sequences, eqn. (2) can be approximated by:

$$SIR_1 \approx \frac{p_1}{\sigma^2 + \frac{1}{L}\sum_{u=2}^{M} p_u} \tag{3}$$

The factor $1/L$ can be thought of as the *processing gain advantage*. This approximation can be justified in two specific scenarios where random sequences are used. First, the spreading code of a user can be part of a *long* pseudo-noise sequence which spans many symbols (such as in the IS-95 system). Each entry of the sequence can be modeled as i.i.d. equally probable to be $+1/\sqrt{L}$ or $-1/\sqrt{L}$. Each term in the denominator of (3) is then the expected value of the interference $(\mathbf{s}_1 \cdot \mathbf{s}_u)^2 p_u$ due to interferer u, averaged over the randomness of the spreading sequences. For a system with coding over consecutive symbols, this SIR averaged over different symbols is more important than the instantaneous SIR during a single symbol period.

Another scenario for which the approximation (3) can be justified is for systems where the spreading sequence of each user is repeated from symbol to symbol, but it is randomly selected initially when the user enters the network. In this case, the SIR for each symbol is the same but random depending on the initial choice. It is proved (Prop. 3.3 in [67]) that in a large system with many users, i.e. $L, M \to \infty$, but number of users per unit processing gain M/L fixed at α, the random SIR_1 for user 1 converges in probability to the deterministic number:

$$\frac{p_1}{\sigma^2 + \alpha \int_0^\infty p\, dF(p)}$$

where F is the empirical distribution of the received powers of the interferers. This result thus supports (3) as a finite system approximation to the performance.

While the use of long pseudonoise sequences to average out the interference is reasonable for conventional CDMA systems which treat other users as white noise, repetition of the spreading sequence is more suitable for the implementation of more sophisticated receivers which try to adaptively exploit the structure of the interference provided by the signature sequences of the interfering users. (See [72,75] for further discussions and comparisons between these two approaches.) We will return to this point in Section 3 when we discuss multiuser receivers. For the purpose of the present section, however, we will take the SIR equation (3) as the starting point, abstracting away the underlying physical layer structure.

The abstraction (3) shows that we can ascribe an *effective interference* of p_u to user u, summarizing the effect of user u on other users. While this concept is almost trivial in the setting of the conventional receiver, we will see in Section 3.4 that this concept can be extended to multiuser receivers in a nontrivial manner.

2.3 Effective Bandwidth in a Power-Controlled Cell

We will now focus on a single-cell scenario where every user is received and power controlled to a single base-station, and we will derive the capacity of such a system.

Power control is almost a necessary feature of a CDMA system. Indeed, in the current implementation of CDMA in the IS-95 system, all users within the same cell control their transmit powers in such a way to be received at the same power at the cell site. The reasoning behind this type of power control is that if users do not control their powers in this way, then one user close to the receiver can completely dominate the others, and drown out the signals of the other users. This is known as the "near-far" problem (see [51]), so named because it is likely to occur, without power control, when an interferer is near to the desired signal's receiver, and the user himself is far.

The common power control policy in which users equalize their received powers is analyzed in many references ([18], [81], [82], [83], [84]). However, an implicit assumption is that the system under study is for one class of service (eg. voice) and it is not difficult to see that it can be extended to allow multiple classes of services to be accommodated simultaneously. This extension was undertaken in ([21], [22], [23]) and, independently, in [91], and leads to the notion of the *effective bandwidth* of a user within a class of service.

Consider now the situation when there are J different classes of service available. The different service classes might offer different bit rates, or different bit error rates, so that users in different classes have different SIR requirements. Let β_j be the SIR requirement of the users in class j and suppose there are M_j users in class j. If users in class j are received at power p_j, then the SIR achieved by those users is given by

$$\mathrm{SIR}_j = \frac{p_j}{\sigma^2 + \sum_{i=1}^{J} \alpha_i p_i - p_j/L} \quad j = 1, 2, \ldots, J \tag{4}$$

where $\alpha_j = \frac{M_j}{L}$ is the number of users in class j per unit processing gain. In a system with large processing gain L, the contribution of an individual interferer is negligible, and we can further simplify the above equation to:

$$\mathrm{SIR}_j = \frac{p_j}{\sigma^2 + \sum_{i=1}^{J} \alpha_i p_i} \quad j = 1, 2, \ldots, J \tag{5}$$

The *power control problem* then arises: how do the J classes choose their J respective received power levels in order to meet their desired SIRs? The basic requirement is that a solution in $\mathbf{p}$ can be found to the following linear inequalities:

$$\frac{p_j}{\sigma^2 + \sum_{i=1}^{J} \alpha_i p_i} \geq \beta_j \quad j = 1, 2, \ldots, J \tag{6}$$

but it can easily be shown that a solution exists if and only if a *minimal solution* exists, satisfying equality in every constraint. The minimal solution is given by

$$p_j^* = \frac{\sigma^2 \beta_j}{1 - \sum_{i=1}^{J} \alpha_i \beta_i} \quad j = 1, 2, \ldots, J \tag{7}$$

It is interesting to see how a notion of *effective bandwidth* arises from equation (7). Let β_j denote the effective bandwidth of a class j users. Equation (7) shows that it is not always possible to carry all the services simultaneously, and the condition for feasibility is precisely that the sum of the effective bandwidths does not exceed unity.

We can interpret

$$\sum_{j=1}^{J} \alpha_j \beta_j < 1 \tag{8}$$

as a capacity constraint on the network, when there are no limitations imposed on the transmit power levels of the users. As such, it reflects the fact that the CDMA system with conventional receiver is *interference-limited*, i.e. the number of users per unit processing gain cannot go unbounded even without any power constraints. It is also intuitively clear that power constraints should further reduce capacity. Let $\bar{P}_j$ be the power constraint on the received powers of the class j users. The effective bandwidth constraint then becomes:

$$\sum_{j=1}^{J} \alpha_j \beta_j < \min_{1 \leq j \leq J} \left[1 - \frac{\beta_j \sigma^2}{\bar{P}_j} \right] \tag{9}$$

2.4 Power Control in a Cellular Network

In this subsection we extend the power control results of Section 2.3 to the cellular network case. The focus is on deriving power control algorithms which yield minimal powers if it is at all feasible that the SIR requirements of the users can be simultaneously met. Moreover, in a cellular network, users in one cell create interference in all other cells, and thus the important issue of decentralization arises: it is essential that any power control algorithms should be based on localized information, and not on centralized decision-making. The problem of capacity characterization of cellular systems, which is considerably more complicated than the single cell scenario, will be discussed in Section 2.5.

A traditional cellular network is depicted in the top of Fig. 1. Each cell has a centrally located cell site and all users in the cell transmit to this cell site. This is the basic model for a narrowband, cellular network. In CDMA, however, the notion of "cell" is relaxed to obtain what is known as "soft handover". When a user is in soft handover between two (or more) cell sites, it really resides in both cells simultaneously, as depicted in the bottom of Fig. 1. Thus a user moving between two cells will spend a period in a soft-handover mode in which it sends signals to both cell sites, and in the IS-95 standard, the strongest signal at any one time is selected by the mobile switching center ("selection diversity").

We shall now formulate the general problem of optimal joint power control and cell-site selection. In what follows, we assume that during the communication period of interest, user i is in soft handover with a set of cell sites, which we denote by D_i. If $|D_i| = 1$, then the user is strictly within a particular cell. Each user has a unique SIR target corresponding to its own particular service requirement, and there are M users. Let K be the number of cell sites in the network.

We consider the network at an instant of time, and capture the random fadings from each user to each cell-site in the $M \times K$ matrix ?. Thus, $?[u, k]$ is the fading from user u to cell site k, which means that if user u transmits with

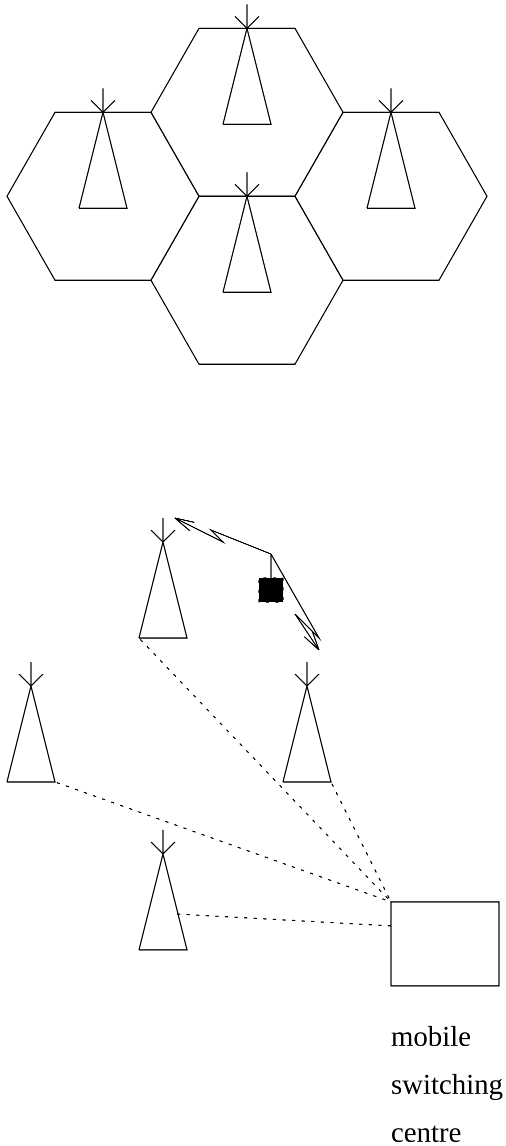

Fig. 1. Traditional cellular (top) vs CDMA cellular (bottom)

power p_u, it is received at cell site k at received power $p_u?[u,k]$. We assume that user i has SIR requirement $SIR_i \geq \beta_i$.

The *power control and cell-site selection problem* is to choose transmit power levels $(p_i)_{i=1}^M$, and cell-site selections $(c_i)_{i=1}^M$ such that

$$c_i \in D_i \tag{10}$$

$$\frac{p_i?[i,c_i]}{\frac{1}{L}\sum_{u\neq i} p_u?[u,c_i] + \sigma^2} \geq \beta_i \quad i = 1,2,\ldots,M \tag{11}$$

Note that this problem is more difficult than the single cell problem with simple linear inequalities in (6). In particular, it is only piecewise linear, and involves the potentially combinatorially explosive problem of finding an appropriate allocation of users to cell sites, a problem that requires knowledge of the entire system path gain matrix for its solution. There is no hope of writing down an explicit solution, except in simple cases. The resolution of this problem is to focus on an adaptive power control algorithm that rapidly converges to a *minimal solution* for transmitter powers. The optimal allocation of users to cell sites is then determined by the solution. Moreover, the algorithm can be implemented in a decentralized manner, in the sense that an individual user adapts its transmit power level based only on locally available information.

The algorithmic solution we now describe was obtained independently in [89] and [22], and both approaches are based on a *monotonicity* condition that applies to the problem. Following the elegant formulation of [90], let us define the interference function $I(\mathbf{p})$ by

$$\begin{aligned} I : \Re^M &\rightarrow \Re^M \\ \mathbf{p} &\rightarrow I(\mathbf{p}) \end{aligned}$$

where

$$I_i(\mathbf{p}) = \min_{k \in D_i} \left(\frac{1}{L} \sum_{u \neq i} p_u ?\,[u,k] + \sigma^2 \right) \frac{\beta_i}{?\,[i,k]} \quad i = 1, 2, \ldots, M \tag{12}$$

Then the inequalities (11) can be expressed as

$$\mathbf{p} \geq I(\mathbf{p}) \tag{13}$$

and we say that a vector $\mathbf{p}$ is *feasible* if and only if (13) is satisfied[3]. Note that if $\mathbf{p}$ is feasible, then for each i we obtain a feasible cell-site allocation c_i as the minimizing value of k in (12).

The power adaptation algorithm of interest is defined in discrete time, and for simplicity we assume users adapt their powers in a synchronous manner, although convergence can also be proved for asynchronous updates [90]. The algorithm is deterministic and iterative, so at time n, the new transmit powers can be defined in terms of the transmit powers at time $n-1$ in a recursive fashion.

Algorithm 1 *Start at time* 0 *with an arbitrary vector of positive transmit powers* $\mathbf{p}^{(0)}$. *Then the transmit powers at time* n *are defined by*

$$\mathbf{p}^{(n)} \equiv I^n(\mathbf{p}^{(0)})$$

[3] The inequality $\geq$ for vectors means greater than in every component.

so that

$$p_i^{(n+1)} = \min_{k\in D_i}(\frac{1}{L}\sum_{u\neq i} p_u^{(n-1)}?\,[u,k] + \sigma^2)\frac{\beta_i}{?\,[i,k]} \quad i = 1, 2, \ldots, M \qquad (14)$$

The associated cell-site selection algorithm is given by

$$c_i^{(n+1)} = \mathrm{argmin}_{k\in D_i}(\frac{1}{L}\sum_{u\neq i} p_u^{(n-1)}?\,[u,k] + \sigma^2)\frac{\beta_i}{?\,[i,k]} \quad i = 1, 2, \ldots, M$$

The following theorems characterize the convergence of this algorithm:

Theorem 2.1 *If the inequalities (13) have a solution, then there exists a minimal solution,* $(\mathbf{p}^*, \mathbf{c}^*)$*, with* $\mathbf{p} \geq \mathbf{p}^*$*, for any other solution* $(\mathbf{p}, \mathbf{c})$*. All inequalities are satisfied with equality at the minimal solution. ([89], [22])*

Theorem 2.2 *For any initial, non-negative vector of transmit powers* $\mathbf{p}^{(0)}$*, the following convergence result holds:*

$$I^n(\mathbf{p}^{(0)}) \rightarrow \mathbf{p}^* \quad as \quad n \rightarrow \infty$$

where p^* *is a minimal power solution. ([89], [22])*

There are three properties of the interference function that can be used to prove Theorems 2.1 and 2.2 ([90]). These are :

- $I(\mathbf{p}) > 0$ for all $\mathbf{p} \geq 0$. (positivity)
- if $\mathbf{p}^{(1)} \leq \mathbf{p}^{(2)}$ then $I(\mathbf{p}^{(1)}) \leq I(\mathbf{p}^{(2)})$. (monotonicity)
- $\alpha I(\mathbf{p}) > I(\alpha\mathbf{p})$ for $\alpha > 0$ (scalability)

In fact, it was shown in [90] that the two theorems hold for *any* function I satisfying the above three properties. Monotonicity is the most fundamental (see Section 4) but scalability is also a useful property that recurs in many other power control problems.

The key feature of scalability in the proof of convergence is that provided $\mathbf{p}^*$ exists, one can scale it up arbitrarily, to create an arbitrarily large feasible power vector, $\mathbf{u}$. Using monotonicity, it then follows that $I^n(\mathbf{u}) \downarrow \mathbf{p}^*$. Conversely, it is easy to see that monotonicity implies $I^n(\mathbf{0}) \uparrow \mathbf{p}^*$. Convergence from an arbitrary starting power vector $\mathbf{p}^{(0)}$ then follows by a sandwich argument.

A nice feature of the power adaptation algorithm, from an implementation point of view, is the way it provides a *decentralized solution* to the power control problem. Consider the right hand side of (14) which can be written as

$\min_{k\in D_i}(\frac{1}{L}I_k^{(n-1)} + \sigma^2)\frac{\beta_i}{?[i,k]}$ where $I_k^{(n-1)}$ is the total received power of all other users at cell site k at step $n-1$ of the algorithm. Thus, at each step of the algorithm, user i need only know its own desired β_i, its own path gains to the cell-sites in which it is is in soft-handover, and the total received interference at each of these cell sites. The user does *not* need to know any information about other cell sites in the network, and it does not need to know anything about the transmit power levels, or the path gains, of any other user.

Cell-site selection can be thought of in the following way. At each step of the algorithm, the user listens to broadcast interference levels from each of the cell-sites in its soft-handover set. The user then computes the transmit power levels that it would need if it were to send to each of these cell sites, and then transmits with the minimum computed power. Apart from this power control mechanism, the cell-site selection plays no other role. Each cell site in the soft handover set can still demodulate the user's signal and send it to the switching center. The switching center does the cell site selection, but the decision can be based on frame error rates, rather than on explicit knowledge of the selection that the mobile itself made when it transmitted the signal.

The above cell-site selection is also known as *selection combining*. If coherent demodulation is used, then a more sophisticated form of combining is *maximal ratio combining*. This type of combining, in the context of multiple cell sites, is called *macrodiversity*, and requires a link between each cell site to a centralized processor. Power control and capacity are studied for a macrodiversity network in [23].

To conclude this subsection, it should be remarked that the power control formulation described above assumes that all users have target SIR requirements and the transmit powers are controlled to meet those requirements. An alternative formulation, known as *power balancing* and adopted by much of the first works on power control ([2], [48], [92], [93]), is to maximize the minimum SIR achieved by all the users in the network. It follows from the monotonicity property that at the optimal solution, all users achieves a common SIR; thus, a better description of this formulation may be *SIR balancing*. In this formulation, there is no notion of network capacity; instead, a *best-effort service* is provided given the resources in the network and the current congestion level. It should also be noted that [2], [92], [93] in fact addressed the power control problem in the context of narrowband systems where users in the same cell have their own channel. In this case, the objective is to mitigate the co-channel interference between users in different cells to facilitate frequency re-use. However, conceptually the power control problem is very similar to the one in the CDMA context. Historically, the important paper [92] (see also [93]) introduced the notion of *adaptive, decentralized algorithms* for power control in the narrowband context.

Although it is natural to to include soft-handover in any model of a CDMA cellular network, early work on power adaptation for cellular CDMA systems [21] (see also [17] for the narrowband case) focussed on a network with a fixed cellular structure, *i.e.* no cell-site selection. In this case, an underlying linearity simplifies the power control problem and provides some further insight into the issue of congestion and effective bandwidth.

In this case, we can assume that for each user i there is a fixed cell site c_i corresponding to the cell in which user i resides. Then

$$I_i(\mathbf{p}) = (\frac{1}{L}\sum_{u \neq i} p_u ?\,[u, c_i] + \sigma^2)\beta_i / ?\,[i, c_i]$$

The fixed point equation $\mathbf{p} = I(\mathbf{p})$ then simplifies to the linear system of equations

$$(I - \mathcal{B}\ A)\mathbf{p} = \mathbf{b} \tag{15}$$

where I is the $M \times M$ identity matrix, $\mathcal{B}$ is a diagonal matrix with entries $\beta_1, \beta_2, \ldots, \beta_M$, and $\mathcal{A}$ is a $M \times M$ matrix with entries:

$$\mathcal{A}[i,j] = \begin{cases} \frac{1}{L}\frac{\Gamma[j,c_i]}{\Gamma[i,c_i]} & i \neq j \\ 0 & i = j \end{cases}$$

and $\mathbf{b} = \sigma^2 \mathcal{B}\mathcal{H}^{-1}\mathbf{1}$, where $\mathcal{H}$ is a diagonal matrix with ith entry $?\,[i, c_i]$, and $\mathbf{1}$ is the vector of all 1s. Note that $\mathcal{A}$ is not strictly positive, but it is *primitive* (its square is strictly positive) and so the Perron-Frobenius theory ([59]) applies. Indeed, it is well known from Perron-Frobenius theory that (15) has a positive solution, $\mathbf{p}^*$, if and only if $\lambda^* < 1$, where λ^* is the Perron-Frobenius eigenvalue of $\mathcal{B}A$. If $\lambda^* < 1$ then $(I - \mathcal{B}A)^{-1}$ exists and is positive, so we can express $\mathbf{p}^*$ as

$$\mathbf{p}^* = (I - \mathcal{B}A)^{-1}\mathbf{b} \tag{16}$$

Algorithm 1 reduces in the fixed cells scenario to the simple Jacobi iteration

$$\mathbf{p}^{(n+1)} = \mathbf{b} + \mathcal{B}A\mathbf{p}^{(n)} \tag{17}$$

and it is well known from Perron-Frobenius theory that (17) converges if and only if $\lambda^* < 1$, and if $\lambda^* < 1$ then it converges to $\mathbf{p}^*$. These results provide further insight into the performance of Algorithm 1, which behaves as (17) until a cell site re-selection occurs. If the final optimal cell site allocation were fixed and known *a priori*, then (17) would provide geometric convergence to the minimal power allocation, at a rate given by λ^*. Since the optimal allocation is not known, Algorithm 1 allows dynamic cell site selection, and this complicates the convergence analysis. However, recent work [33] has shown that Algorithm 1 does converge at a geometric rate, provided the allocation of users to cell sites at the optimal solution is unique (and this is true with probability 1, see [22]).

The capacity constraint $\lambda^* < 1$ suggests that λ^* itself might provide a measure of congestion. In the single cell case, we can easily compute that $\lambda^* = \frac{1}{L}\sum_{i=1}^{M}\beta_i$, under the approximation that the received power of each user's signal is negligible compared to the total received power of the other users. Note that λ^* is then precisely the sum of effective bandwidths, as in (8).

The multiple cell scenario is much more complicated, as it has a spatial aspect, and "congestion" no longer just depends on each user's own effective bandwidths, but also on the path gains of each user in the network. While it is unrealistic to expect that a single number can capture all aspects of congestion in a spatial model, it is of interest to see what λ^* might be measuring in the multi-cell case. This is the subject of investigation in [28] in which λ^* is related to other measures of congestion, such as total received powers at cellsites (*i.e.* interference levels), "power warfare" (the sensitivity of power levels in the network to new traffic), and actual traffic levels in the network (as measured by effective bandwidths). Indeed, lower bounds on λ^* are provided which are sums of effective bandwidths in regions, where the regions range from single cells, to the whole network itself. A recent paper [4] shows a very strong correlation between small fluctuations in λ^* and outage events in a simulated cellular model with maximum power constraints on the users. Nevertheless, the interpretations here for λ^* are not as clear-cut as in the single cell case, and for that matter, it is difficult to be precise about what "congestion" means in the multiple cell context. When we look at soft-handover, for example, the eigenvalue that corresponds to the cell site allocation that minimizes the powers in the network is not necessarily the same as the smallest eigenvalue out of all cell site allocations. The minimum power allocation reduces the congestion as measured by power consumption, yet the minimum eigenvalue allocation might provide more capacity when the network load increases; one can make a distinction between "power warfare" congestion, and "capacity congestion". Fortunately, this is not a practical problem, since cell site allocation need not be fixed, but can instead be dynamic, as it is in Algorithm 1, and the minimum power allocation is then always optimal.

In the case of macrodiversity [23] (i.e. network-wide soft-handover, with maximal ratio combining) a network-wide notion of congestion arises that makes this distinction between "capacity congestion", and "power warfare" congestion particularly clear. In contrast to the cellular models, capacity can be completely characterized by an effective bandwidth constraint:

$$\sum_{j=1}^{J} \alpha_j \beta_j < K \tag{18}$$

where K is the number of cell sites in the network. Note that the total effective bandwidth $\sum_{j=1}^{J} \alpha_j \beta_j$ does not depend on the path gains in the network. However, this measure of congestion does not characterize a kind of "power warfare" congestion that can arise when too many users are located in close proximity, and it is not an adequate measure of congestion when power constraints are imposed.

It may well be that power consumption is the bottleneck resource in a radio network, and congestion then needs to be measured with respect to this resource. In addition to λ^*, measures such as the total received power at each cell site in the network may be important ([21]). How all these measures relate to each other, and which will actually prove useful in admissions control and flow control applications is not yet well understood.

2.6 Extensions to Basic Model

2.6.1 Power Constraints

The basic decentralized power control algorithms we have considered so far can easily be extended to include maximum power constraints on the transmit powers of the users [90]. In this case, the only necessary modification is to limit the transmit power of the users at each step of the algorithm; they use the minimum of the power the unconstrained algorithm would specify and their maximum power level. The constrained algorithm is guaranteed to converge, and will satisfy all the users' SIR requirements if this is possible under the power constraints. If not, those users with final transmit powers below their constraints will at least achieve their target SIRs.

Although it is not the focus of the present paper, it is also possible to consider the effect of power constraints on the "best effort" power control mentioned at the end of Section 2.4. It is clear that the problem of maximizing the minimum SIR subject to power constraints still amounts to SIR balancing, as it was without power constraints (see [20]). It is not possible to do this in a totally decentralized way, and a centralized algorithm is obtained in [20]. In future work, it may be of interest to see if better algorithms can be obtained which involve only limited communication between cell sites. Another approach might be to include constraints on total received power levels at cell sites, as a way of controlling congestion in the network.

2.6.2 Asynchronicity, measurement errors, stochastic effects, and quantization

So far we have assumed that power control operates in a synchronous fashion. The basic model in [17] is relaxed in [44] to allow asynchronism: users adapt their powers in discrete time, but it is no longer necessary for all users to update their powers at each time point. Propagation delays are also modeled: it is assumed that for each user there is a delay between the interference measurement at the cell site, and when this information is available to the user. Nevertheless, [44] shows that a power adaptation algorithm, similar in form to (17), is robust in the face of these relaxations, although it is shown that the Perron-Frobenius eigenvalue, λ^*, increases if there is asynchronism, and hence some configurations diverge that in the synchronous case would have converged. The important issue of rate of geometric convergence is addressed in [44], and it is shown that asynchrony slows down the rate of convergence (precisely because λ^* increases).

Another practical issue that is not properly addressed in the earlier algorithms is that of *measurement*. For example, in the power control scheme actually implemented in the IS-95 CDMA standard (see Section 2.7 for more details of their implementation), power levels are not measured directly at the cell site, but are extracted from matched filter outputs, and the issue of measurement error arises. A theoretical study that extends the algorithm in (17) along these lines is [70]. Consider the deterministic iteration:

$$\mathbf{p}^{(n+1)} = \mathcal{B}A\mathbf{p}^{(n)} + \mathbf{b}$$

but which they write, perhaps more conveniently as:

$$\mathbf{p}^{(n+1)} = \mathcal{B} \quad H^{-1}(A\mathbf{p}^{(n)} + \sigma^2\mathbf{1})$$

where $\mathcal{H}$ is the diagonal matrix of $?[i, c_i]$s, $\mathbf{1}$ is the vector of all 1s, and A is defined by

$$A[i,j] = \begin{cases} \frac{1}{L}?\,[l, c_i] & i \neq j \\ 0 & i = j \end{cases}$$

The next step is to consider a relaxed version:

$$\mathbf{p}^{(n+1)} = (1 - a_n)\mathbf{p}^{(n)} + a_n\mathcal{B} \quad H^{-1}(A\mathbf{p}^{(n)} + \sigma^2\mathbf{1}) \tag{19}$$

where (a_n) is a sequence satisfying $a_n < 1$ for all n. However, this assumes that the users can measure the interferences in $A\mathbf{p}^{(n)} + \sigma^2\mathbf{1}$, so [70] extends the model to the case where the users only measure the squared value of their matched filter outputs, taken over some finite window. Randomness enters because the information bits of the users are random, as is the external noise, and so it follows that the matched filter outputs are random. Indeed, the random vector of averaged matched filter outputs, $\mathbf{v}$, can be written as

$$\mathbf{v} = (A + \mathcal{H})\mathbf{p} + \sigma^2 \mathbf{1} + \mathbf{w}$$

where $\mathbf{w}$ is a zero-mean random variable. The stochastic adaptation based on $\mathbf{v}$ can be written as:

$$\mathbf{p}^{(n+1)} = (1 - a_n)\mathbf{p}^{(n)} + a_n(-\mathcal{B}\mathbf{p}^{(n)} + \mathcal{B} \quad H^{-1}\mathbf{v}^{(n)})$$

where we note that if we replace $\mathbf{v}^{(n)}$ by its expected value, we would recover the deterministic iteration (19). Stochastic convergence results are obtained under various assumptions; one result being, for instance, that if the sequence $a_n = \frac{1}{n}$ is used, then $\mathbf{p}^{(n)}$ still converges to the solution $\mathbf{p}^*$ in (16).

Another feature of the IS-95 power control (again, see Section 2.7), is that frame error rate measurements are used. A theoretical study that extends the type of algorithm considered above to the case in which power updates are based on bit error rate (BER) is provided in [39].

An important assumption behind all the power adaptation algorithms we have mentioned so far, from the original narrowband work in [92] to Algorithm 1 in the present paper, is that updates occur so quickly that the channel gains (the ?$[i,j]$s) can be assumed to be fixed. This does not mean that the algorithms only apply when users are immobile, but rather that there is a separation of timescale between power updates and changes in propagation conditions. The issue that the path gains themselves may be stochastic (*i.e.* not be known precisely) is considered briefly in [70]. The dynamics of the iteration

$$\mathbf{p}^{(n+1)} = (1 - a_n)\mathbf{p}^{(n)} + a_n(-\mathcal{B}\mathbf{p}^{(n)} + \mathcal{B}\hat{H}^{-1}\mathbf{v}^{(n)})$$

are studied, where $\mathcal{H}$ is replaced with an estimate $\hat{H}$. Nevertheless, this work still assumes that the gains are time-invariant. An interesting topic for future research might be to study stochastic power control algorithms when the channel is described by a stochastic process.

So far, the theoretical power control models we have considered do not capture the discrete nature of the transmit power levels that are available to mobiles in practical systems such as IS-95. A recent paper [3] addresses this issue and shows that notions of optimality and convergence are less straightforward. Two discrete algorithms are presented, both being modifications of the continuous power control algorithm; one rounds up to the nearest grid point (the "ceiling" algorithm) and one rounds down (the "floor" algorithm). Both are shown to converge in a weaker sense to an envelope of powers, but of course oscillations are possible. It is shown that oscillations can be avoided by first running the floor algorithm , followed later by the ceiling algorithm, and that this converges to an optimal power vector.

An earlier power control paper to consider quantization is [82]. This paper provides an analysis of the closed-loop control of IS-95 power control, which only allows mobiles to increase or decrease their transmit powers by 1 dB based on 1 bit feedback from the cell sites. We will examine this control, and the analysis in [82], in more detail in Section 2.7.

2.6.3 Bursty Traffic and Rate Control

With the rapidly growing level of data in modern communications, it is natural to look for new power control algorithms that are more appropriate for data. Two important properties of data traffic come into play here: it tends to be highly bursty, and it usually has less stringent delay requirements so that there is often the flexibility of being able to do some form of dynamic rate allocation.

In [45], an approach is taken which assumes that burstiness of the traffic occurs on a very fast time-scale, and that powers do not have time to adapt to the dynamically changing interference in the way the previous algorithms did. Instead, transmit power levels need to be found that satisfy the SIR requirement *in a statistical sense*, taking into account that the random, bursty nature of the interferers. The basic power adaptation algorithm (17) is adjusted by including a measurement of the *variance* of the interference, and it is shown to converge (in an appropriate large bandwidth limiting regime) to a solution in which users obtain their desired SIRs a sufficient proportion of the time.

The approach in [45] is sender-driven in the sense that the receiver does not attempt to control the traffic rate of the users. An alternative approach considers the scenario in which the bursty traffic can be queued at the transmitter and explicit rate allocation and power control performed in adaptation to congestion level, channel conditions and traffic demands from the users. For example, [60] and [52], [53] study the problem in which users do not require fixed target rates, but can adapt the processing gain to keep the required SIR fixed. Thus, the SIR target for a particular user i remains fixed, but now L can be adapted. Indeed, different users can have different spreading factors, so we should write L_i for user i. The assumption is still that the overall rate of chips/sec is held fixed, and hence the spectrum occupied by the signals is fixed, but the number of chips/information symbol L_i is variable. By varying L_i, user i therefore varies the rate of information in symbols/sec. More sophisticated approaches can even consider the target SIR as variable, as would be the case with adaptive coding. This type of "best effort" bandwidth allocation may be very appropriate for many types of data, and indeed mirrors the sort of flow control provided by TCP on the Internet. The papers [52], [53] also consider the issue of dynamic rate allocation, in which it is shown that throughput can be increased if delay tolerant packets are scheduled so as to avoid collisions *i.e.* schemes with a hybrid TDMA/CDMA flavour are considered. This type of approach, but with focus on fading issues rather than burstiness, is also taken in [36], [37]. We will also consider this approach in more detail in Section 4.4.

A related study, although not specifically on power control, is [49], [50]. These authors study a random access model of data in a CDMA environment and show that processing gain control eliminates bistability. In fact, they show that there is an optimal spreading gain for each level of backlogged traffic (which increases linearly with the amount of backlogged traffic) and that asymptotically, in the limit of large backlog and spreading gain, the optimal retransmission probability of a backlogged packet is 1. The connection with power control is provided by the final section of the paper, in which energy constraints are imposed. This work seems open to be extended to the case of joint power control and spreading gain control, and indeed preliminary work in this direction can be found in [32].

Recent research on flow control for the internet has advocated an economic approach to resource allocation based on utility functions and the pricing of resources (see [34]). This approach can also be applied to power control, and has been taken up recently in [15], and in an information-theoretic context in [66], [26]. We will review the latter two papers in Section 4, when we return to the problem of joint rate and power allocation in the context of information-theoretically optimal schemes.

2.6.4 Admission Control

The interesting idea that power control should be intimately connected with *call admissions control* is taken up in [6] (see also [7]). Bambos and Pottie [6] formulate a notion of *active link protection* and incorporate this idea into the power control procedure. In this approach, a small, positive constant, δ, defines the protection margin for the active calls that are already in progress. Thus, if all users require the minimal SIR of β, then it is assumed that they can actually achieve a higher SIR, namely $(1+\delta)\beta$. When a new call arrives, the network must decide whether or not to admit the call into the active set, and power control is intimately related to this call admissions procedure. While the active links use the basic power adaptation algorithm, with target SIRs marginally above the minimal requirements, as described above, the new user is only allowed to increase its power at a slow, linear rate prescribed by the parameter δ. It is shown that no active user will ever drop below its minimal SIR requirement, and that if it is possible for the new user to become a member of the active set, then it will do so eventually. However, if after a certain time-out period has elapsed, the new user has not yet become an active link, the network will drop the call.

2.7 Power Control in IS-95

To conclude this section, we will discuss the implementation of power control and handover in the IS-95 CDMA standard ([18], [81], [82], [83], [84], [80]), and compare it with the theoretical results surveyed above.

An interesting feature of IS-95 is the ability of a user to go into "soft handover" between several cell sites simultaneously. The approach to power control and soft handover in [18] is to assume that at any instant of time, a user selects the cell site to which its current path gain is *strongest*. All users in the network then control their transmit powers in such a way to be received at a fixed received power level at their chosen cell site. The level of received power is fixed, and common to all cell sites in the network. Under these assumptions, applicable to a system with a single class of service, users will achieve a variable SIR, with a level dependent on the random, fluctuating level of interference at their chosen cell site. In [18], capacity is then calculated on the basis that traffic is uniformly distributed in the network, and that users have a minimum tolerable SIR (common to all users) that they must achieve a sufficient proportion of the time. Further papers in this direction consider the benefits of soft handover [84], the effects of other-cell interference ([83]), and the random arrival and departure of users in the system ([81]).

It is interesting to compare this power control algorithm to Algorithm 1. Rather than fixing the received power level of all users at their chosen cell site, Algorithm 1 takes into account the interference levels at each cell site, and received power levels *adapt* to fluctuating interference levels. This is clearly an advantage in the case in which traffic is *nonuniformly* distributed in the network, and we note that the algorithm will adapt to changing traffic patterns. On the other hand, Algorithm 1 may be *too adaptive* in some scenarios, when interference levels are very volatile, and power control algorithms do not have time to converge. A lowpass filter can be used to overcome this ([1])) where the bandwidth is chosen to provide the desired level of smoothing of channel fluctuations. A power control algorithm that adapts received powers on a slow time-scale (such that the received powers are constant on the fast time-scale of fading and burstiness) is described in [24].

The actual power control in IS-95 has an open-loop component that attempts to keep the received power near the target level. The way this is achieved is by allowing the cell site to transmit a pilot signal on the forward link, in addition to the users' data, and then the mobile can measure the strength of the pilot signal and so estimate the path gain from cell site to the mobile. This can be used to adjust the transmit power of the mobile to compensate for large changes in signal strength brought about by shadow fading and cell geometry. However, because IS-95 is implemented in a frequency division duplex mode, the reverse link channel is not identical to the forward link channel due to fast varying multipath effects. As a result, a faster, closed-loop control is also needed to try to compensate for these effects that change rapidly, and yet can't be measured by the mobile.

The closed-loop power control operates at 800 Hz and captures fairly fast changes in propagation conditions, as well as rapid changes in interference levels from other mobiles, but it is not fast enough to allow direct measurement and feedback of multipath fading effects. Instead, the impact of multipath is indirectly measured in an outer loop that tracks frame error rates. The inner loop involves 1 bit feedback from the cell site to the mobile, based on measured SIR values; if the measured SIR is above a setpoint, the command is to decrease power by 1 dB, and if the SIR is below a setpoint, then the command is to increase power by 1 dB. The received power measurement is based on matched filter outputs, rather like the model considered in [70], but in addition to measurement errors, the accuracy of the power control is also limited by the 1-bit quantization. An outer loop varies the SIR setpoint as a function of frame error rates, and in this respect is similar to the model studied in [39]. An important point, however, is that even though 1 bit feedback occurs at a high rate (800 Hz), this is only fast enough to track shadow fading at vehicular speeds, and there is no separation of timescale between updates and the fast end of this fading process. The way the closed-loop control deals with soft handover is that the mobile will always decrease its transmit power by 1 dB if at least one of the soft handover cell sites instructs it to do so; in other words, the minimum transmit power is always used. Thus, Algorithm 1 captures this aspect of the IS-95 power control. It should be noted that in current third-generation wideband CDMA proposals, mobiles transmit a pilot signal on the reverse link as well as on the forward link, and this enables the cell site to directly measure received power from the pilot.

An analysis of the IS-95 closed loop is provided in [82]. This paper focusses on more detailed physical layer issues than the other papers we have reviewed so far. On the other hand, the previous analyses take into account the interaction that occurs between users *i.e.* the fact that when one user increases its power, the interference to other users also increases, and causes them to increase their power. In [82], the multi-access aspects are not addressed. The analysis in [82] is in discrete time, and T_n refers to the transmit power of the user at time n in dB. Let L_n be the propagation loss, and E_n the received power, so that

$$E_n = T_n - L_n \quad \text{in dB}$$

In the IS-95 closed loop, the transmit power at time $n+1$ is an explicit function of feedback from the cell-site, but implicitly this is dependent on the received power at time $n-1$ (this lag takes account of measurement and propagation delays). Thus,

$$T_{n+1} = T_n + C(E_{n-1})\Delta \tag{20}$$

where $C(E_{n-1})$ is the power command from the cell site (± 1) and Δ is the discrete change in power level (*e.g.* 1 dB). To be more precise, C is not a function of E_{n-1} , but takes the value ± 1 with a probability that depends on E_{n-1}. The analysis includes the calculation of the probability for a particular low-rate encoded orthogonal modulation scheme, as used in the Qualcomm system. Included in the probability is the chance that the mobile receives the command incorrectly. If we rewrite (20) as

$$E_{n+1} = E_n + C(E_{n-1})\Delta - (L_{n+1} - L_n) \tag{21}$$

and if we assume the increments $L_{n+1} - L_n$ are independent, as in fairly fast fading, then we have a nonlinear difference equation with independent-increment driving function. It is reasonable to assume that the increments are Gaussian, under the assumption that the shadow fading is log-Normally distributed. A simulation study in [82] then obtains the probability distribution for E_b/N_0 under various multipath fading assumptions. Finally, coded error performance is obtained from the probability distribution for E_b/N_0.

3 Linear Multiuser Receivers

3.1 Multiuser Receivers

It was shown in Section 2.2 that under the matched filter receiver and random signature sequences, the *effective interference* of an interferer u is equal to its received power p_u. A direct consequence of this fact is the *near-far problem*: users with strong received powers will drown out the weak users. In a conventional CDMA system, the only counter-measure available is power control.

It has been well appreciated for some time ([74]) that the near-far problem is actually not intrinsic to a direct sequence CDMA system, but is due to the sub-optimality of the matched-filter receiver. The matched-filter only depends on the signature sequence of the user to be demodulated, and is optimal (in the sense that its outputs are sufficient statistics) only when the signature sequences of the users are orthogonal to each other. When this is not the case (such as for pseudonoise sequences), performance gain can be achieved by taking into account the signature sequences of the other users as well.

The first multiuser receiver, proposed by Verdu [73], has the property of minimizing the probability of symbol detection error. However, its complexity grows exponentially with the number of users in the system. Linear multiuser receivers were later proposed [40,41] which have lower complexity and retain much of the performance advantage over the conventional matched-filter receiver. In this part of the paper, we will focus on linear multiuser receivers.

Although a significant amount of work has been done on the performance of multiuser receivers, most have focused on evaluating their ability to reject the *worst-case* interference from other users (a notion called *near-far resistance* [40]). Here, we are concerned with the questions of how power control should be done in conjunction with multiuser receivers and the resulting network capacity. The latter gives a basis for comparing the performance gain over the conventional receiver.

3.2 MMSE Receiver

A linear receiver for user 1 is specified by a L dimensional vector $\mathbf{r}_1$ such that the the demodulated symbol is $\mathbf{r}_1 \cdot \mathbf{y}$, where $\mathbf{y}$ is the received signal. The SIR associated with a given receiver is given by:

$$SIR_1 = \frac{(\mathbf{r}_1 \cdot \mathbf{s}_1)^2 p_1}{(\mathbf{r}_1 \cdot \mathbf{r}_1)\sigma^2 + \sum_{u=2}^{M} (\mathbf{r}_1 \cdot \mathbf{s}_u)^2 p_u},$$

using the same channel model (1) as in the previous section.

The conventional receiver is obtained by picking $\mathbf{r} = \mathbf{s}_1$. Given the signature sequences and received powers of all the users, the optimal linear receiver that maximizes the SIR is called the Minimum Mean-Square Error (MMSE) receiver [42,54,57]. This receiver is thus given by:

$$\mathbf{r}_1^* = \text{argmax}_{\mathbf{r} \in \Re^L} \frac{(\mathbf{r}_1 \cdot \mathbf{s}_1)^2 p_1}{(\mathbf{r}_1 \cdot \mathbf{r}_1)\sigma^2 + \sum_{u=2}^{M} (\mathbf{r}_1 \cdot \mathbf{s}_u)^2 p_u}.$$

The formulae for the MMSE and its performance are well known:

$$\mathbf{r}_1^* = \frac{p_1}{1 + p_1 \mathbf{s}_1^t (S_1 D_1 S_1^t + \sigma^2 I)^{-1} \mathbf{s}_1} (S_1 D_1 S_1^t + \sigma^2 I)^{-1} \mathbf{s}_1 \tag{22}$$

and the signal to interference ratio for user 1 is

$$SIR_1^* = p_1 \mathbf{s}_1^t (S_1 D_1 S_1^t + \sigma^2 I)^{-1} \mathbf{s}_1 \tag{23}$$

where $S := [\mathbf{s}_2, \ldots, \mathbf{s}_M]$ and $D := \mathrm{diag}(p_2, \ldots, p_M)$. Observe that the MMSE receiver, unlike the conventional receiver, depends not only on the signature sequence of the user to be demodulated but also on the sequences and powers of all the other users. In practice, this information is obtained by adaptive algorithms that enable the receiver to learn about the structure of the interference (see for e.g. [31]). Because of its optimality, we will focus on the MMSE receiver in this section. However, we will also be comparing it with other multiuser receivers.

3.3 Power Control

Consider now a cellular network with demodulation by MMSE receivers at the base-stations. As in the conventional receiver scenario, we are interested in the power control problem: how does one find the appropriate powers and cell sites for the users to satisfy given desired SIR requirements? This problem was studied by [38,69,55] and they showed that it can be solved naturally in the framework considered in Section 2.4. Basically, the ability to choose the optimal linear receiver provides additional degrees of freedom in the optimization beyond cell site selection.

Using the same notation as introduced in Section 2.4, the problem can be formulated as follows: given SIR requirement β_i for user i, $i = 1, \ldots M$, choose transmit power levels $(p_i)_{i=1}^M$, cell site selection $(c_i)_{i=1}^M$ *and* receivers $(\mathbf{r}_i)_{i=1}^M$, such that:

$$c_i \in D_i \tag{24}$$

$$\frac{(\mathbf{r}_i \cdot \mathbf{s}_i)^2 p_i ? [i, c_i]}{\sum_{u \neq i} (\mathbf{r}_i \cdot \mathbf{s}_u)^2 p_u ? [u, c_i] + (\mathbf{r}_i \cdot \mathbf{r}_i)\sigma^2} \geq \beta_i \quad i = 1, 2, \ldots, M \tag{25}$$

As before, the set D_i contains the base-stations to which user i is currently in soft-handover. Compared to (11), we have explicitly included the signature sequences in the SIR equations, as we are now concerned also with choosing the best linear receiver for given sequences. Observe also that for a given user i, the interferers can be from inside and outside the cell of user i, so that the receiver for user i will depend on out-of-cell interference as well. As a matter of fact, there is no distinction between in-cell and out-of-cell interference in this formulation.

Again, the key to the solution of this problem is to define the appropriate interference function I_i. In this problem, we can define for given power vector

p:

$$I_i(\mathbf{p}) = \min_{k \in D_i} \min_{\mathbf{r} \in \Re^L} \frac{\left(\sum_{u \neq i} (\mathbf{r} \cdot \mathbf{s}_u)^2 p_u ?\,[u,k] + (\mathbf{r} \cdot \mathbf{r})\sigma^2\right)}{(\mathbf{r} \cdot \mathbf{s}_i)^2 ?\,[i,k]} \beta_i \quad i = 1, 2, \ldots, M \qquad (26)$$

It is easily seen that inequalities (25) are equivalent to $p_i \geq I_i(\mathbf{p})$ for all i. Straightforward calculations show that this interference function satisfies the three properties of positivity, monotonicity and scalability. Hence Theorems 2.1 and 2.2 hold for this problem. If there exists powers, cell site selections and receivers such that the SIR requirements of the users are satisfied, then there exists a minimal solution $(\mathbf{p}^*, \mathbf{c}^*, \mathbf{r}^*)$ such that $\mathbf{p} \geq \mathbf{p}^*$ for any other feasible solution and for which the SIR requirements are met with equality. It can also be seen that the optimal receiver for user i is the MMSE receiver for the chosen cell site and transmit power levels. Moreover, the optimal solution can be obtained by iterating I starting with any arbitrary non-negative powers.

These results show how optimal powers can be computed for the MMSE receiver, *if* the SIR requirements are feasible. This still leaves the question of characterizing the feasible SIR requirements, i.e. the capacity region. To answer this question, a better qualitative understanding of the performance of the MMSE receiver in a power controlled system is needed, particularly the effect of an interferer to be demodulated. This is a more difficult problem than in the conventional receiver case, since the MMSE receiver depends on signature sequences and received powers of all users. While there is no known solution to the cellular network capacity characterization problem, we will present in the next few sections some results which shed some light on this question.

3.4 Effective Interference

For the conventional receiver, it was observed that for random spreading sequences, a simple approximate SIR equation (2) can be written down. The effect of the interferers can be decoupled into a sum of effective interference terms, each term being equal to the received power of the interferer. This simple SIR equation forms a basis for the derivation of capacity constraints for the conventional receiver, as well as give a simple abstraction of the interfering effect without worrying about specific signature sequences.

The decoupling for the conventional receiver is a consequence of the fact that the receiver depends only on the signature sequence of user 1 and nothing else. The situation is not as simple for the MMSE receiver, as it depends on *all* the sequences and received powers of users (see eqn. (22)). Somewhat surprisingly, in a large system with many users and large processing gain, some sort of decoupling does occur even for the MMSE receiver.

Recall that the random sequence model is that each entry is randomly and independently chosen to be $+1/\sqrt{L}$ or $-1/\sqrt{L}$. Since the sequences are random, the performance SIR_1, being a function of the sequences, is also random. The following result describes the asymptotic distribution of SIR_1 [67]. Its proof makes use of results from random matrix theory [43,62].

Theorem 3.1 *Suppose the number of users M and the processing gain L both go to infinity, with $M/L \to \alpha$, and the empirical distribution of the received powers of the interferers converge to a limiting distribution F. Then* SIR_1 *converges to β_1^* in probability, where β_1^* is the unique solution to the equation:*

$$\beta_1^* = \frac{p_1}{\sigma^2 + \alpha \int_0^\infty I(p, p_1, \beta_1^*) dF(p)} \tag{27}$$

and

$$I(p, p_1, \beta_1^*) \equiv \frac{p p_1}{p_1 + p\beta_1^*}$$

Heuristically, this means that in a large system, the SIR β_1 is deterministic and approximately satisfies:

$$\beta_1 = \frac{p_1}{\sigma^2 + \frac{1}{L}\sum_{u=2}^{M} I(p_u, p_1, \beta_1)} \tag{28}$$

where as before p_u is the received power of user u. Thus, under the MMSE receiver and in a large system, the total interference can be decoupled into a sum of the background noise and an interference term from each of the other users. (The factor $\frac{1}{L}$ results from the processing gain advantage of user 1.) The interference term depends only on the received power of the interfering user, the received power of user 1 and the attained SIR. It does not depend on the other interfering users except through the attained SIR β_1.

One must be cautioned not to think that this result implies that the interfering effect of the other users on a particular user is additive across users. It is not, since the interference term $I(p_u, p_1, \beta_1)$ from interferer u depends on the attained SIR which in turn is a function of the entire system. However, it can be shown that the equation:

$$x = \frac{p_1}{\sigma^2 + \frac{1}{L}\sum_{u=2}^{M} I(p_u, p_1, x)} \tag{29}$$

has a unique fixed point x^*, and moreover the equation has the following monotonicity property: for any x, $x^* \geq x$ if and only if

$$\frac{p_1}{\sigma^2 + \frac{1}{L}\sum_{u=2}^{M} I(p_u, p_1, x)} \geq x \tag{30}$$

It follows then that to check if the target for user 1's SIR, β_T, can be met for a given system of users, it suffices to check the following condition:

$$\frac{p_1}{\sigma^2 + \frac{1}{L}\sum_{u=2}^{M} I(p_u, p_1, \beta_T)} \geq \beta_T$$

Based on this interpretation, it seems justified to consider $I(p_u, p_1, \beta_T)$ as the *effective interference* of user u on user 1, at a target SIR of β_T.

The correspondence between eqns. (3) and (28) is somewhat striking. For the matched filter, the interference due to user u is simply p_u in place of $I(p_u, p_1, \beta_1)$. Since the matched filter is independent of the signature sequences of the other users, it is not surprising that the interference is linear in the received powers of the interferers. In the case of the MMSE receiver, the filter does depend on the signature sequences of the interferers, thus resulting in the interference being a non-linear function of the received power of the interferer. Also, observe that $I(p_u, p_1, \beta_1) < p_u$, which is expected since the MMSE receiver maximizes the SIR among all linear receivers. But more importantly, in the matched filter case the interference grows unbounded as the received power of the interferer increases, yet for the MMSE receiver, the effective interference from user i is bounded and approaches $\frac{p_1}{\beta_1}$ as p_u goes to infinity. Thus, while the SIR of the matched filter receiver goes to zero for large interferers' powers, the SIR of the MMSE receiver does not. This is the well-known *near-far resistance* property of the MMSE receiver [42]. The intuition is that as the power of an interferer grows to infinity, the MMSE receiver will null out its signal.

A graphical comparison of the effective interferences of the matched filter and the MMSE receiver is shown in Fig. 2. This figure also shows the performance of the *decorrelator*. This is another multiuser receiver [40,41] which operates by nulling out the directions spanned by the signature sequences of the interferers. More precisely, rather than projecting the received signal onto the signature sequence of the desired user, as would the matched filter, the decorrelator receiver projects the received signal onto the orthogonal complement to the space spanned by the signature sequences of all the other users. This receiver is only well defined if the dimension of the space spanned by the interferers is less than the total processing gain L. The decorrelator is the *zero forcing* linear filter and the effective interference is the effect of the background noise through the filter. The effective interference under the decorrelator is p_1/β_1 ([67]), and does not depend on the received power of any interferer.

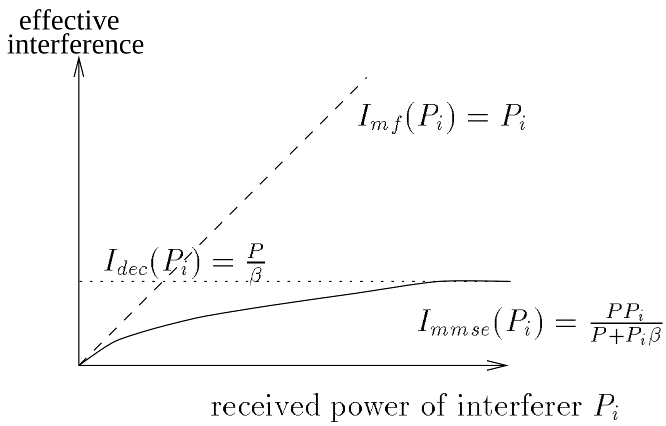

Fig. 2. Effective interference for the 3 receivers as a function of interferer's received power P_i. Here,P is the received power of the user to be demodulated, and β is the SIR achieved.

The effective interference results discussed here for the ideal CDMA model are extended to symbol-asynchronous systems (i.e. symbols of different users are not necessarily time-aligned) [35] and to channels with multipath fading [14].

3.5 Effective Bandwidth in A Power Controlled Cell

The notion of effective interference captures the effect of an individual interferer on the user to be demodulated, and is valid for both in-cell and out-of-cell interferers. Specializing now to a single power-controlled cell allows us to develop a notion of *effective bandwidth* to characterize the capacity under the MMSE receiver [67]. This is in parallel to the development in Section 2.3 for the matched filter.

Consider as before the situation when there are J different classes of service available. Let β_j be the SIR requirement of the users in class j, and suppose there are M_j users in class j. Focus on the asymptotic regime where $L, M_j \to \infty$ and $M_j/L \to \alpha_j$, the number of users per degree of freedom in each class. The result in Section 3.3 tells us that if the SIR requirements are feasible, there is a minimum power solution. From (27), it is also clear that the minimal received powers of users in the same class should be the same; let p_j^* be the power for class j. Then the power control equations become

$$\frac{p_j^*}{\sigma^2 + \sum_{i=1}^{J} \alpha_i I(p_i^*, p_j^*, \beta_j)} = \beta_j \quad j = 1, 2, \ldots, J \tag{31}$$

where, as in Theorem 3.1, $I(p_i, p_j, \beta_j) \equiv \frac{p_i p_j}{p_j + p_i \beta_j}$. Solving these equations give:

$$p_j^* = \frac{\beta_j \sigma^2}{1 - \sum_{i=1}^J \alpha_i \frac{\beta_i}{1+\beta_i}} \quad \text{j= 1, 2, \ldots, J.} \tag{32}$$

The capacity constraint for the MMSE receiver with J classes is therefore given by

$$\sum_{j=1}^J \alpha_j \frac{\beta_j}{1+\beta_j} < 1 \tag{33}$$

which is linear in $\alpha_1, \ldots, \alpha_J$. This shows that even under the MMSE receiver, the system is still interference limited, and the interference-limited capacity region under random sequences is given by (33).

As in the matched filter case, maximum power constraints provide tighter capacity constraints, and in this context we note that (32) implies that

$$\sum_{i=1}^J \alpha_j \frac{\beta_i}{1+\beta_i} = 1 - \frac{\beta_j \sigma^2}{p_j^*} \qquad j = 1, 2, \ldots, J.$$

Thus if $p_j^* \leq \bar{P}_j$ is a maximum power constraint on class j, then the linear constraint

$$\sum_{j=1}^J \alpha_j \frac{\beta_j}{1+\beta_j} \leq \min_{1 \leq j \leq J} \left[1 - \frac{\beta_j \sigma^2}{\bar{P}_j} \right] \tag{34}$$

defines the restricted capacity region of the system. It seems very reasonable to define the effective bandwidth of class j users to be $e_{mmse}(\beta_j)$ degrees of freedom per user, where

$$e_{mmse}(\beta_j) \equiv \frac{\beta_j}{1+\beta_j}.$$

Comparing eqn. (34) to the corresponding capacity constraint (9), we note that the effective bandwidth of a user in class j under the matched filter is β_j degrees of freedom. The effective bandwidth concept thus forms a basis for comparing the performance of different receivers; see Fig. 3. Also shown is the effective bandwidth under the decorrelator; it is precisely 1 degree of freedom, irrespective of the SIR requirement of the user (see [67]).

We note that the conventional receiver is more efficient than the decorrelator when β is small, and far less efficient when β is large. Intuitively, at high SIR requirements, a user has to transmit at high power, thus causing a lot of

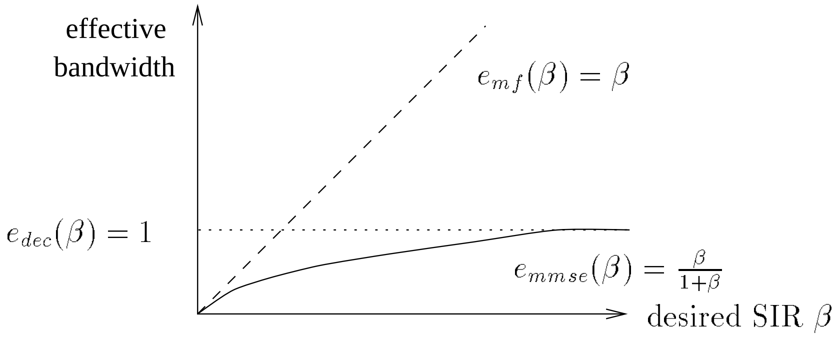

Fig. 3. Effective bandwidths for 3 receivers as a function of SIR.

interference to other users under the conventional receiver. Not surprisingly, since it is by definition optimal, the MMSE filter is the most efficient in all cases. When β is small, it operates more like the conventional receiver, allowing many users per degree of freedom, but when β is large, each user is decorrelated from the rest, much as in the decorrelator receiver, and therefore the interferers can still occupy no more than 1 degree of freedom per interferer. The performance gain afforded by the MMSE receiver over the conventional receiver depends on the SIR at which the system is to be operated, and this in turn depends on the data rate, the amount of coding and the symbol constellation size. However, due to the superior performance of the MMSE receiver over a wide range of SIRs, it can be seen that it is particularly suitable in a heterogeneous network with multiple traffic types.

Linearity in the capacity constraints in the matched filter case is a straightforward consequence of the fact that powers add. However, the MMSE effective bandwidth results are rather surprising, as the receiver itself depends on the signature sequences and the received powers of the users. We will provide a partial explanation of this phenomenon in the next subsection.

3.6 Conservation Laws

In this section, we will show that the linearity of the capacity constraints under the MMSE receiver is partially a consequence of underlying deterministic *conservation laws* governing the tradeoff between the performance of different users under the MMSE receiver. This understanding in turn allows us to extend some of the above asymptotic results to finite systems with arbitrary signature sequences, not necessarily random.

Suppose $\mathbf{s}_1, \ldots, \mathbf{s}_M$ are given signature sequences and $p_1, \ldots, p_M$ given received powers of the users in a system with processing gain L. The following lemma relating the performance of different users under the MMSE receiver is proven in [78].

Lemma 3.2 *Let* SIR_i *be the performance of user* i *under the MMSE receiver. Then:*

$$\sum_{i=1}^{M} \frac{SIR_i}{1+SIR_i} = \sum_{j=1}^{L} \frac{\lambda_j}{\lambda_j+\sigma^2} \tag{35}$$

where λ_j*'s are the eigenvalues of* SDS^t *and* $S = [\mathbf{s}_1, \ldots, \mathbf{s}_M]$, $D = \mathrm{diag}(p_1, \ldots, p_M)$.

The following result yields the capacity region in a finite system, for given signature sequences but with power control [27].

Theorem 3.3 *For any subset* U *of the users, let* $S(U)$ *be the matrix whose columns are the signature sequences of the users in* U. *Let* $\beta_1, \ldots, \beta_M$ *be the SIR requirements of the users. They can be satisfied by some choice of transmit powers if and only if:*

$$\sum_{i\in U} \frac{\beta_i}{1+\beta_i} < \; rank\,(S(U)) \qquad \forall U \subseteq \{1,2,\ldots,M\} \tag{36}$$

The necessity of these constraints follows directly from Lemma 3.2. The sufficiency is verified by showing that an iterative power control algorithm, similarly to the one presented in Section 3.3, converges to a finite minimum power solution whenever the SIR requirements satisfy (36).

The constraints (36) reflect the basic conservation laws trading off the performance of one user and the other. When the signature sequences are chosen such that any subset of L or fewer users has linearly independent sequences, the constraints (36) collapse into a single constraint:

$$\sum_{i=1}^{M} \frac{\beta_i}{1+\beta_i} < L \tag{37}$$

This deterministic result for fixed spreading sequences provides a clear explanation of the interference-limited capacity region (33) for random sequences; we see here that the role of the random sequences is to ensure that the linear independence conditions are met with high probability in a large system. However, in the case when there are power constraints on the users, there is no known capacity characterization analogous to (36) for finite system with arbitrary signature sequences. Indeed the power-constrained capacity characterization (34), and the notion of effective interference from which it is derived, depends heavily on the randomness of the signature sequences.

It should also be noted that no such deterministic conservation laws exist for the matched filter. Thus, the effective bandwidth characterization of the capacity region under the matched filter is valid only for random sequences, unlike the MMSE case.

3.7 Extensions: imperfect power control, symbol-asynchronism, multipath fading, and antenna arrays

The effective bandwidth and capacity results of Section 3.5 are for systems with perfect power control. In practice, due to channel fading, feedback delay and errors, the received powers of users in the same class cannot be kept identical. The effect of imperfect power control on capacity is analysed in [94] and [96] for the matched filter and the MMSE receivers respectively.

The effective interference results discussed in Section 3.4 for the ideal CDMA model are extended to symbol-asynchronous systems (i.e. symbols of different users are not necessarily time-aligned) in [35], and to channels with multipath fading [14]. Effective bandwidth results are also obtained for the symbol-asynchronous system in [35].

Although we have described the application of the MMSE receiver and the decorrelator receiver to a CDMA system, these receivers can be applied to any system with spatial diversity, and in particular to antenna array processing ("beam-forming"). Work on antenna capacity using the decorrelator receiver can be found in [86]. It is shown in [67] that the notion of effective interference and effective bandwidth carry over to antenna arrays. A study of joint power control, and beam-forming for antenna arrays is undertaken in [55], [56].

Recent work has also considered the combination of CDMA and antenna array signal processing (*e.g.* [47], [23], [64]). Typically, standard rake signal processing (matched filtering, using maximal ratio combining of the antenna outputs) is assumed. In this scenario, it is shown in [23] that effective bandwidth results carry over from the single antenna, matched filter scenario, to the antenna array case. In [27], effective bandwidth and power control results are obtained for CDMA antenna arrays, using the multi-user MMSE receiver.

4 Information Theoretic Optimal Receivers

In the previous section, we have discussed the problems of power control and capacity characterization for linear multiuser receivers. Although they are likely candidates for implementation in the next generation CDMA systems, it should be noted that neither the direct-sequence modulation format nor the linear receiver structure is optimal from an information theoretic point of view. In this section, we will take a more speculative look at the problem of power control and resource allocation for information theoretically optimal systems. It will be seen that the solutions to these problems are quite different in flavor from the counterpart in DS-CDMA systems with linear multiuser receivers.

The focus in this section will be exclusively on the problem for a single cell, where all the users are power-controlled to the same receiver. The problem of power control in the cellular case is at this time completely open.

We first introduce the multiuser Shannon capacity region for the Gaussian multi-access channel modeling the uplink. We will then use the characterization as constraints for the optimal power control problem.

4.1 Multiuser Shannon Capacity Region

The *Shannon capacity* [4] of a point-to-point channel is the maximum rate at which information can be transmitted reliably with arbitrarily small probability of error. Analogously, the *Shannon capacity region* of a M-user multi-access channel is the set of rate vectors $\mathbf{R} = (R_1, \ldots, R_M)$ that can be simultaneously transmitted reliably from the M users to the single receiver. We focus on channels with additive Gaussian noise. The capacity of a discrete-time point-to-point Gaussian channel with power constraint p and noise power σ^2 is well known:

$$C = \frac{1}{2}\log(1 + \frac{p}{\sigma^2}).$$

For the additive Gaussian multi-access channel model in (1), it is known that Shannon capacity can be achieved when bandwidth expansion is done by coding alone, i.e. the processing gain is set to be 1. This maximizes the rate for a given total bandwidth, i.e. the rate per chip. The resulting Shannon capacity region is given by (see e.g. [11]):

$$\mathcal{C}(\mathbf{p}) = \left\{ \mathbf{R} : \sum_{i \in U} R_i \le \frac{1}{2}\log\left(1 + \frac{1}{\sigma^2}\sum_{i\in U} p_i\right) \qquad \text{for all } U \subset \{1, \ldots, M\} \right\}, \tag{38}$$

where σ^2 is the background noise power and p_i is the received power of user i. Here, R_is are measured in terms of bits per chip (or per sample) and the log is to the base 2.

[4] To distinguish this from our general use of capacity as the number of users that can be supported in the network, we will use "Shannon capacity" to refer to the information rate.

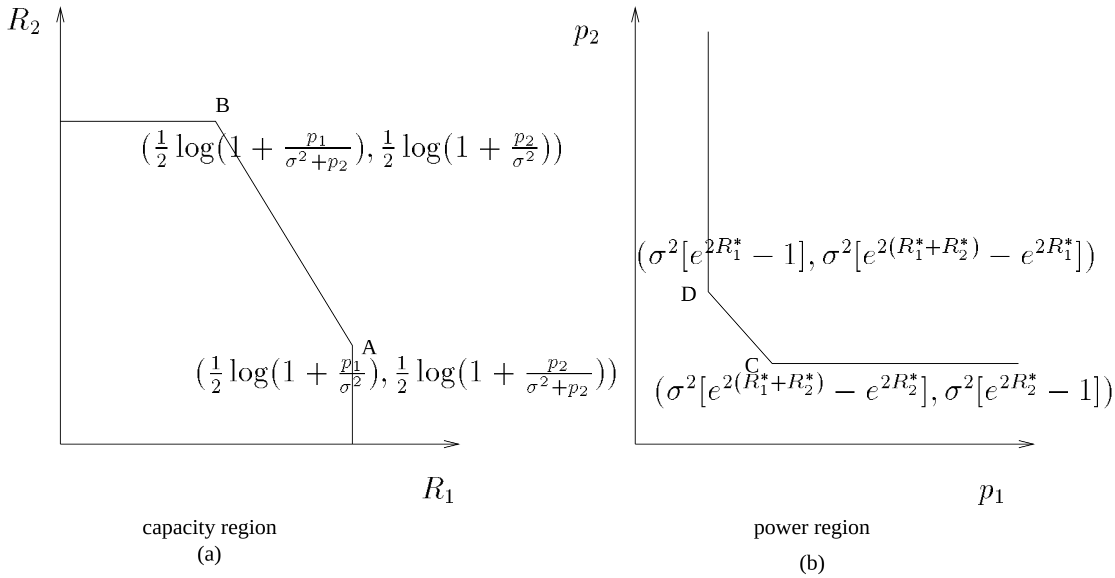

Fig. 4. (a) Two-user capacity region inside the pentagon, for given power constraints p_1 and p_2. Point A is achieved by decoding user 2 first and then user 1; point B vice versa. Priority is given to user 1 at point A so that user 1 gets better rate than at point B. (b) Two-user power region to the outside of the three constraints, for given target rates R_1^*, R_2^*. Point C is achieved by decoding user 2 first and then user 1; point D vice versa. Priority is given to user 2 at point C so that user 2 needs less power than at point D.

This Shannon capacity region is a polyhedron with an interesting structure. It is characterized by $2^M - 1$ constraints, one for each subset of users. There are exactly $M!$ vertices in the positive orthant. The rate vector at each of the vertices can be achieved by a technique called *successive decoding*. First fix an ordering of users. Decode the first user, treating other-user interference as Gaussian noise. Then subtract the known transmitted signal of this user from the total received waveform and repeat the process with the second user. The last user is decoded with the interference coming only from the background noise, the other users' interference having been cancelled out in previous stages. Assume ideal coding for each user, this procedure achieves for user i an information rate equal to:

$$\frac{1}{2}\log\left(1+\frac{p_i}{\sigma^2+\sum_j p_j}\right)$$

where the sum is over the powers of the users that are not yet cancelled when user i is decoded. Note that this is also the capacity of a point-to-point Gaussian channel with received power p_i and noise power $\sigma^2+\sum_j p_j$. Applying this technique to every possible ordering of the users yields all the vertices of the capacity region. Fig. 4(a) shows a two-user capacity region.

It is observed in [66] that the Shannon capacity region of the Gaussian multi-access channel is a *polymatroid*, a class of combinatorial objects first studied by Edmonds [13]. Since this structure is central to our power control problems, we will review the general definition here.

Definition 1 *Let* $E = \{1, \ldots, M\}$ *and* $f : 2^E \to \Re_+$ *be a set function. The polyhedron*

$$\mathcal{B}(f) \equiv \{(x_1, \ldots, x_M) : \sum_{i \in U} x_i \leq f(U) \quad \forall U \subset E, \quad x_i \geq 0 \quad \forall i\} \tag{39}$$

is a polymatroid *if the set function* f *satisfies: 1)* $f(\emptyset) = 0$ *(normalized), 2)* $f(U) \leq f(T)$ *if* $U \subset T$ *(nondecreasing), and 3)* $f(U) + f(T) \geq f(U \cup T) + f(U \cap T)$ *(submodular). The polyhedron*

$$\mathcal{G}(f) \equiv \{(x_1, \ldots, x_M) : \sum_{i \in U} x_i \geq f(U) \quad \forall U \subset E\}$$

is a contra-polymatroid *if* f *is normalized, nondecreasing and satisfies* $f(U) + f(T) \leq f(U \cup T) + f(U \cap T)$ *(supermodular).*

If f *satisfies the three properties,* f *is called a* **rank** *function in both cases.*

It can be shown that a polymatroid has exactly $M!$ vertices in the positive orthant, each of which is the intersection of M of the constraints, corresponding to a sequence of nested subsets. Polymatroid structure arises in many resource allocation and scheduling problems as a consequence of underlying strong *conservation laws* [58]. By giving different priority orders to the users in the scheduling of resources, one can achieve all the extreme points of the polymatroid performance region.

For the present problem, the strong conservation laws are the following. First restrict the operating points to those for which the sum constraint is tight, ie:

$$\sum_{i=1}^{M} R_i = \frac{1}{2} \log \left(1 + \frac{1}{\sigma^2} \sum_{i=1}^{M} p_i \right)$$

Note that at these operating points, the *sum* rate over all users is conserved. Then, for any subset of users, U, the information theoretic constraint says that the achievable rates among this subset of users must satisfy:

$$\sum_{i \in U} R_i \leq \frac{1}{2} \log \left(1 + \frac{1}{\sigma^2} \sum_{i \in U} p_i \right)$$

with equality if and only if *strict priority* is given to the users in U over the remaining users (this means that users in U are always decoded *after* the other users have been decoded). It is also clear that from a power control perspective there is no point in considering any operating points that do not achieve equality in the overall constraint ([30]).

It is interesting to note in passing that the achievable region (36) in the linear receiver setting is also a polymatroid, in terms of the performance measure SIR/(1 + SIR). There, prioritization among users is achieved by power control together with the MMSE receiver, with strict priority given to a subset of users by allocating arbitrarily large powers to these users; the remaining users then null them out. In the information theoretic setting, priority is done through the successive decoding order together with power control.

4.3 Optimal Power Control

In the information theoretic setting, a natural QoS measure for a user is its achievable information rate. This is the analog of the SIR requirement used in the formulation for DS-CDMA systems with linear receivers. The power control question is then : how can one "optimally" control the received powers $p_1, \ldots, p_M$ to achieve a target rate vector $\mathbf{R}^* = (R_1^*, \ldots, R_M^*)$?

This power control problem is fundamentally different from the one in the linear receiver setting. In particular:

- The system viewed as a single cell is *not* interference-limited because powerful users can be cancelled out after they are decoded. Nevertheless, in practice, using the minimum power to guarantee the desired level of QoS is still necessary to reduce interference in adjacent cells and to conserve battery power.
- There is no solution which minimizes the required power of *all* users. Unlike the linear receiver case, increasing the power of one user *benefits* the others because it can be decoded and cancelled more easily. In other words, the *monotonicity* property, which was central to the power control problems for linear receivers, does not hold in this setting.

A sensible formulation then is to minimize a weighted sum of the users' *transmit* powers while at the same time ensuring that the target rates $\mathbf{R}^*$ can be met [26,46]. Denote $?_i$ as the path gain from user i to the receiver. The power control problem can be precisely stated as:

$$\min_{\mathbf{p}} \sum_{i=1}^{M} \frac{\lambda_i}{?_i} p_i \qquad \text{subject to} \qquad \mathbf{R}^* \in \mathcal{C}(\mathbf{p}). \tag{40}$$

The coefficients $\lambda = (\lambda_1, \ldots \lambda_M)$ are the weights for the transmit powers of the users. We will address the choice of λ in a moment, but let us first focus on solving the optimization problem (40). The constraints define a feasible *power region*:

$$\mathcal{P}(\mathbf{R}^*) \equiv \left\{ \mathbf{p} \in \Re_+^M : \sum_{i \in U} p_i \geq \sigma^2 [\exp(2 \sum_{i \in U} R_i^*) - 1] \right\},$$

It can be directly verified that the power region is a *contra-polymatroid*. See Fig. 4(b) for an example of a two-user power region.

To find the optimal solution, we observe that (40) is a *linear programming problem*, so it follows that the optimal solution corresponds to a *vertex* of the power region $\mathcal{P}(\mathbf{R}^*)$. Each vertex of the power region corresponds to one of $M!$ possible successive decoding order, with the powers such that the target rates $\mathbf{R}^*$ can be achieved when the users are decoded in that order. More explicitly, the vertex $\mathbf{p}$ corresponding to successive decoding order π is given by :

$$p_{\pi(i)} = \begin{cases} \sigma^2 [\exp(2 R^*_{\pi(1)}) - 1] & \text{if } i = 1 \\ \sigma^2 [\exp(2 \sum_{m=1}^{i} R^*_{\pi(m)}) - \exp(2 \sum_{m=1}^{i-1} R^*_{\pi(m)})] & i = 2, \ldots, M \end{cases}$$

(The interpretation of π is such that user $\pi(M)$ is decoded first, user $\pi(1)$ is decoded last.)

The optimal solution to the problem (40) must be at one of these $M!$ vertices, corresponding to the $M!$ possible successive decoding orders. A well-known result in polymatroid theory [13] says that the decoding ordering should be in increasing value of the coefficients $\lambda_i/?_i$, i.e. the user with smallest $\lambda_i/?_i$ decoded first, the user with largest $\lambda_i/?_i$ decoded last. Note that the optimal ordering does not depend on the target rates $\mathbf{R}^*$, although the optimal powers do.

Thus, even though the power region has exponentially large number of constraints (in M), a simple explicit solution can be obtained. Here again, it is useful to think of the successive decoding order π as a way to give priority to different users in the scheduling of resources; a user decoded *later* in the ordering is given higher priority than a user decoded earlier. This is because users need more transmit power to support their target rates when they are decoded earlier. What polymatroid theory tells us is that the optimal solution can be obtained in a *greedy* manner: always decode the user with the "cheapest" power first, where the cost is measured by the coefficient $\lambda_i/?_i$. This rule is analogous to the classic $c-\mu$ rule in scheduling theory (see eg. [58]), as both arise from the polymatroid structure of the underlying optimization problems.

The weights λ_i's can be thought of as "power prices". In the special case when they are all set to be equal, the optimal strategy minimizes the total transmit power and takes on the simple form of decoding the user with the best channel first and the one with the worst channel last [25]. The decoding order thus adapts to the fading state $(?_1, \ldots, ?_M)$. Contrast this with the strategy of keeping received powers equal in conventional CDMA schemes, we see that this optimal scheme leads to a much lower transmit power for the user with the weakest channel as it need not compete with any of the other users who have better channels. This improvement in performance is a direct consequence of the flexibility of the successive decoding technique. In a cellular system, this optimal power control strategy has the further advantage of reducing the inter-cell interference, leading to an increase in its interference-limited capacity ([9,85]).

Given fixed λs, we can think of the greedy solution to (40) as a *fast time-scale* power allocation algorithm. On a fast time-scale, during which the channel can be thought of as fixed, the greedy solution determines the optimal successive decoding order and the allocation of transmit power levels to the users, for the current channel state.

On a *slow time-scale*, over which channel variations occur, this gives rise to an *average*, or long-term power consumption by the users. A more general formulation is to impose *average* transmit power constraints $\bar{p}_1, \ldots, \bar{p}_M$ on the users, averaged over the random time variation of the fading state $(?_1, \ldots, ?_M)$, and to require that a target rate vector $\mathbf{R}^*$ is achieved at all fading states. In [26], it is shown that in this problem formulation, there is no loss in generality in restricting attention to power control strategies that solve (40), for some choice of λ. Suppose the fading state has a certain stationary distribution. Given a target rate vector $\mathbf{R}^*$ and power prices λ, let $\bar{p}_i(\mathbf{R}^*, \lambda)$ be the average transmit power for user i when applying the power control which solves (40) at each fading state $?$. However, for a different power price vector, we would get a different average power consumption, and so the issue of *fairness* arises. Setting λ_is to be all the same in the above example minimizes the total average power consumption, but this may not be a *fair* allocation if users have different rate and power requirements.

In [26] the issue of *min max* fairness is considered. The problem is to find power prices λ which minimizes the maximum of the average transmit powers of the users, weighted by the respective average power constraints:

$$\inf_{\lambda > 0} \max_{1 \leq i \leq M} \frac{\bar{p}_i(\mathbf{R}^*, \lambda)}{\bar{p}_i}. \tag{41}$$

Note that if the optimal value is less than 1, then the target rates are achievable within the given average power constraints, but otherwise they are not.

To solve (41), the fading state distribution is needed, and in practice this may not be known explicitly. However, in [26], an iterative algorithm is provided to solve (41) that be implemented *adaptively*, where the updates on the power prices can be driven by measuring the actual average power consumed. An interesting feature of the algorithm is that the proof that it converges to the minmax solution is based on a monotonicity property that holds for the underlying mapping that defines the iteration.

The power control we have sketched in the present section is a *two-time-scale* resource allocation scheme. At the slow time-scale over which the channel variations occur, the algorithm iteratively updates the power prices to meet average power constraints. At the fast time-scale during which the channel can be thought of as fixed, the solution to (40) for the current power prices is used to control the powers and successive decoding order.

4.4 Optimal Rate and Power Control

It was observed that the optimal power control strategy considered in the previous section provides flexibility by prioritizing users according to their fading states. However, users in deep fade will still require a large amount of transmit power to ensure that the target rates are met. In fact, for some fading distributions, such as the Rayleigh distribution, meeting a fixed target rate at every fading state would require an infinite amount of transmit power. If instead one is interested only in maximizing the long-term rate, *averaged over time* as the fading state varies, then an alternative strategy is to dynamically vary both the *rate* and the power over time: more power is used to send at a higher rate when the channel is good and less or even no power when the channel is bad. While this does not guarantee a constant rate at all fading states, it can yield a better long-term average rate for a given average power constraint by exploiting the *time-diversity* in the system. For applications, such as data, with delay requirements longer than the time-scale of the channel fluctuations, this may suffice.

Goldsmith and Varaiya [19] formulate this idea for *point-to-point* fading channels, and pose the question: given a time-varying fading channel, what is the optimal power control policy which maximizes the long-term average rate subject to an average power constraint? The optimal transmit power to use at fading state ? is given by:

$$\max\left(\lambda - \frac{1}{?}, 0\right),$$

where λ is the Lagrange multiplier (power price) chosen such that the average power constraint is met. If one considers a sample path of the fading process $\{?(t)\}$ and plots the curve $1/?(t)$, then this optimal power allocation has the interpretation of filling water on this curve up a level of λ such that the average amount of water (power) per unit time is equal to the power constraint. Note that this solution has the qualitative properties of a good policy described above.

The problem of optimal rate and power control is formulated and solved for the general multi-access scenario in [65,66]. The optimal solution has the following structure. At each fading state $? = (?_1, \ldots, ?_M)$, the optimal rate and power allocation solves the following optimization problem:

$$\max_{(\mathbf{R},\mathbf{p})} \sum_{i=1}^{M} \left(\mu_i R_i - \frac{\lambda_i}{?_i} p_i \right) \qquad \text{subject to} \qquad \mathbf{R} \in \mathcal{C}(\mathbf{p}) \tag{42}$$

The weights μ_is can be interpreted as *rate rewards* prioritizing the users, while the λ_is are Lagrange multipliers associated with the average power constraints. They can also be interpreted as *power prices*. Contrast this with the optimization problem (40), where the rates are fixed and the optimization is only over the powers. In the formulation of the previous section, powers are optimally controlled to meet a target rate vector $\mathbf{R}^*$ at every fading state, subject to average power constraints. In the formulation here, both rate and power can be varied to adapt to channel conditions, in such a way as to maximize the long term rates (averaged over the fading state) subject to average power constraints. Varying the rate rewards μ_is allows the tradeoff between the long-term rates achieved by different users. Exploiting the underlying convexity of the problem, it is shown in [66] that for given average power constraints, all achievable long term average rates can be obtained by an appropriate choice of the rate rewards μ.

Let us consider the heart of the problem, which is to solve (42) for the optimal rate and power allocation for given rate rewards μ, power prices λ and fading state $?$. This would yield the "fast time-scale" rate and power control at a fixed fading state. Unlike the optimization problem (40), this is not a standard problem in polymatroid theory, and requires a new solution [66]. Nevertheless, the solution still retains the simple greedy flavor that one would expect from the underlying polymatroid structure of the constraints. Define:

$$\begin{aligned}
g(z) &\equiv \frac{1}{2}\log(1 + \frac{z}{\sigma^2}) \\
u_i(z) &\equiv \mu_i g'(z) - \frac{\lambda_i}{?_i} = \frac{\mu_i}{2(\sigma^2 + z)} - \frac{\lambda_i}{?_i} \\
u^*(z) &\equiv \left[\max_i u_i(z)\right]^+,
\end{aligned}$$

where $x^+ \equiv \max(x, 0)$. The optimal value for problem (42) is given by

$$\int_0^\infty u^*(z)dz.$$

The optimal solution is again achieved by successive decoding and can be interpreted as follows. Think of $\sigma^2 + z$ as the current "interference level" due to background noise and received powers of users not yet cancelled, and think of $u_i(z)$ as the *marginal utility* obtained from allocating unit received power to user i at interference level $\sigma^2 + z$. Starting with $z = 0$, at each z we allocate a marginal *received* power δp to the user i^* with the largest positive $u_i(z)$. Stop when $u_j(z) < 0$ for all j. The marginal increase in rate of user i^* is $g'_{i^*}(z) \cdot \delta p$, decoding at interference level $\sigma^2 + z$. The value $u^*(z) \cdot \delta p$ is therefore the marginal increase in the value of the overall objective function $\sum_i \mu_i R_i - (\lambda_i / ?_i) p_i$ by allocating power δp to the user that will benefit most at the interference level $\sigma^2 + z$. The procedure is thus *greedy*. Integrating over all z gives the optimal rate and power allocation to all the users. Moreover, it is guaranteed that the resulting solution can be achieved by successive decoding, with the ordering of the users implicitly given by the above procedure. See Fig. 5 for an example. We observe that some users may get no power and therefore no rate in the optimal solution. This means that the current fading state is too unfavorable for those users to transmit information.

The special case when the rate rewards and power prices are the same for all users was earlier solved by Knopp and Humblet [36]: the optimal strategy has the interesting structure that only the user with the best channel transmits at any one time. This follows directly from the general solution above: when the μ_i's are the same, the functions $u_i(\cdot)$ are all parallel and that of the user with the best channel dominates for all z. Moreover, the transmit power used by the strongest user also has the waterfilling interpretation described above. When some users have weaker channels a lot of the time, this strategy may have some fairness problems. Assigning unequal rate rewards to users can yield a fairer policy in that case.

The greedy solution presented above to the optimization problem (42) is intimately tied to a certain *dual* polymatroid structure of the constraints: given received powers $\mathbf{p}$, the Shannon capacity region $\mathcal{C}(\mathbf{p})$ is a polymatroid; on the other hand, given target rate vector $\mathbf{R}$, the feasible power region $\mathcal{P}(\mathbf{R})$ is a contra-polymatroid. In fact, a more general class of polymatroids shares this structure. This is the class of polymatroids with *generalized symmetric* rank functions, i.e. rank functions f of the form:

$$f(U) = g(\sum_{i \in U} y_i)$$

where g is an increasing concave function and $y_1, \ldots y_M$ are scalars. It is clear

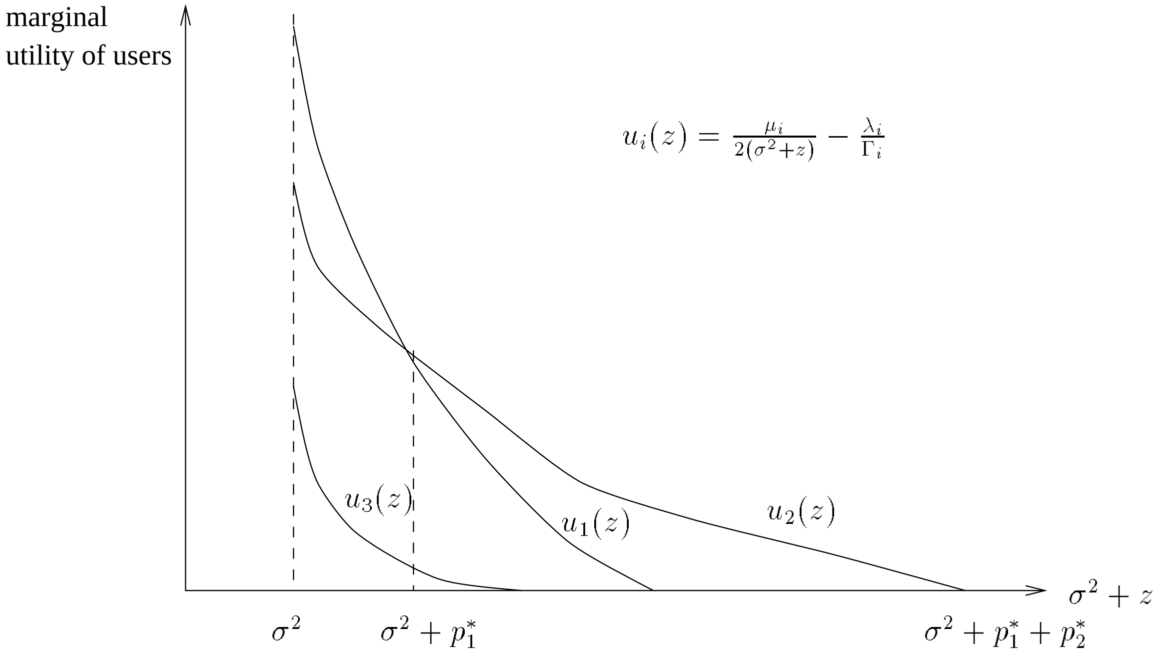

Fig. 5. A 3-user example illustrating the greedy power allocation. The x-axis represents the received interference level and y-axis the marginal utility of each user at the interference levels. At each interference level, the user who is selected to transmit is the one with the highest marginal utility. Here, user 1 gets decoded after user 2, and user 3 gets no power at all. The optimal received powers for user 1 and user 2 are p_1^* and p_2^* respectively.

that the multi-access Gaussian capacity region belongs to this class, with g specializing to the log function. If we now consider the following optimization problem for polymatroids with generalized symmetric rank function:

$$\max_{(\mathbf{x},\mathbf{y})} \mu \cdot \mathbf{x} - \lambda \cdot \mathbf{y} \qquad \text{subject to} \qquad \sum_{i \in U} x_i \leq g(\sum_{i \in U} y_i) \qquad \forall U \subset \{1, \ldots, M\}, \tag{43}$$

then it can be shown that the greedy algorithm presented above solves this more general problem as well. This may be of independent interest for other resource allocation problems involving this class of polymatroids (see [16,58] for some examples.)

We remark here that at the optimal solution to (43), the overall constraint

$$\sum_{i=1}^{M} x_i \leq g(\sum_{i=1}^{M} y_i)$$

will be satisfied with equality. In the present problem, this says that there is a total amount of resources provided by the users, $\sum_{i=1}^{M} p_i$, and this dictates the total amount of rate, $\sum_{i=1}^{M} R_i$ that can be allocated among the users. Given we decode some user first, then the total amount of rate remaining to be allocated is dictated by the remaining total amount of received power, and so on. These are the conservation laws referred to in Section 4.2; they involve *conservation* since it is the sum of rates that is conserved among many possible rate allocation policies.

We have considered the optimal power and rate allocation at a fading state, for a given μ and λ. As in Section 4.3, we consider the case in which there is a stationary fading process, and we note that for fixed μ and λ, the optimal rate and allocation solution gives rise to *average* rate and power vectors $\bar{\mathbf{R}}(\mu, \lambda)$, and $\bar{\mathbf{p}}(\mu, \lambda)$ respectively. In [66], iterative algorithms are developed that adjust μ and/or λ to solve various resource allocation problems.

Consider, for example, the situation in which there is a desired average power vector, $\bar{\mathbf{p}}$, and a vector of rate rewards, μ, is given. Consider the following iterative algorithm for updating λ such that the average power constraints are met and $\sum_i \mu_i \bar{\mathbf{R}}(\mu, \lambda)$ is maximized.

Algorithm 2 *Start with an arbitrary power price vector $\lambda^{(0)}$. Generate a sequence of power price vectors $(\lambda^{(n)})_{n=1}^{\infty}$ as follows. Given $\lambda^{(n-1)}$, vary λ_i, holding the other $\lambda_j^{(n-1)}$s fixed, until $\bar{p}_i(\mu, \lambda) = \bar{p}_i$. This defines the new value of $\lambda_i^{(n)}$. Each user does this simultaneously, and we obtain a new vector $\lambda^{(n)}$.*

It is proven in [66] that $\bar{\mathbf{p}}(\mu, \lambda^{(n)}) \to \bar{\mathbf{p}}$, as $n \to \infty$. A key element of the proof is the following monotonicity property: if λ_i increases, then $\bar{p}_i(\mu, \lambda)$ decreases, but all other $\bar{p}_j(\mu, \lambda)$s increase, for $j \neq i$.

A "dual" algorithm in [66] deals with the case in which a desired rate vector $\bar{\mathbf{R}}$ is given, together with a vector of power prices λ. An algorithm for adapting μs is given, such that $\bar{\mathbf{R}}(\mu^{(n)}, \lambda) \to \bar{\mathbf{R}}$, as $n \to \infty$, while minimizing $\sum_i \lambda_i \mathbf{p}(\mu, \lambda)$.

Summarizing, we have presented a two-time-scale resource allocation scheme for optimal rate and power control in multi-access fading channels. At the slow time-scale, power prices or rate rewards are updated to meet average power or rate constraints. At the fast time-scale, the greedy solution to the optimization problem (42) yields the optimal rate and power allocation at the current fading state and current power prices and rate rewards.

The results we have presented here are for the time-varying, flat fading Gaussian multiaccess channel. Dual results hold for the time-invariant, frequency selective Gaussian multiaccess channel [66]. In this scenario, the channel response is not flat over frequency and the optimal solution involves power allocation accross frequencies, rather than over time. Our solution generalizes earlier work in [10], which considered this problem in the special case of two users. In [66] we also treat the more general case of a time-varying, frequency selective channel. As remarked above, in the flat fading scenario, the optimal power and rate allocation to achieve the maximum sum rate was found in [36]. This work is also extended to the frequency selective case in [37].

5 Conclusions and Open Problems

It is important to first emphasize the *differences* between the linear receiver CDMA power control solutions, and the information-theoretic resource allocation solutions. The latter are *two-time-scale* resource allocation schemes. At the fast time-scale, during which the channel can be thought of as fixed, fast greedy algorithms allocate rates and/or powers among the users. At the slow time-scale over which the channel variations occur, the algorithm iteratively updates the rate rewards and/or power prices to meet average constraints. Contrast this with the power control strategies for linear receivers. They can be viewed as single time-scale algorithms since there is a strict component-wise optimal solution for each fading state in that problem; a property which does not hold in the information theoretic formulation. In the linear receivers case, users directly control their access to the "available bandwidth" through their transmit power levels. In the information-theoretic formulation, increasing power can *benefit* other users, and it is this lack of monotonicity in power space that requires us to incorporate performance constraints that are averaged over the fading distribution.

At a more fundamental level, however, we note that there *is* monotonicity in λ-space, in the information-theoretic formulation of Section 4.3. If a user increases its power price, then it will benefit, since it will use less average power, but all other users will use more. In Section 4.4, if a user increases its rate reward, then it will benefit since it will get more long-term rate, but all the others will get less rate. Users control access to the "available resources" through their power prices (and in Section 4.4, rate rewards). This enables very similar iterative procedures to be applied to compute the appropriate rate rewards and/or power prices, as were used in Section 2.4 to compute the optimal transmit power levels.

We also note that conservation laws arise in both problems. Section 3.6 shows that the totality of *effective bandwidths* of the users is always bounded by the processing gain, no matter what power control is used. Furthermore, for each subset U of users in the system, there is a conservation constraint imposed by the dimension of the space spanned by the signature sequences of the users in U. In the information-theoretic single-cell models, the notion of "available resources" is more subtle, since the system is not interference limited, and resources can be increased or decreased through power control. Nevertheless, there are indeed conservation laws that still apply, as explained in Section 4.2, and these have a strikingly similar form to those in Section 3.6. It should be emphasized though that the existence of effective interference and effective bandwidths under power constraints is based not only on the conservation laws but also on the randomness of the signature sequences.

There are many open problems in the area of power control and its relation to network capacity. First of all, little attention has been payed to multiple time-scales that arise in both models of fading, and models of data source behavior. For example, all the works reviewed in this paper, with the exception of the Shannon-theoretic work on fading channels, and the essentially single-user analysis of the IS-95 closed loop power control in [82], assume that the fadings (path gains) are held fixed for the duration of the algorithm. This is unrealistic even for indoor wireless systems in which the terminals might be more or less immobile, because mutipath fading effects still occur due to a time-varying environment in which the terminals are located. On this point, it is often necessary to model fading as occurring on two time-scales; the slow timescale of shadow fading is often of the order of seconds, and the fast timescale of multipath fading, due to the constructive and destructive effects of multipath self-interference, can occur on the order of milliseconds. Both time-scales, of course, depend on the carrier frequency, bandwidth, and speed of the mobiles. The stochastic work of [70] briefly considers the issue of imperfect knowledge of the channel gains, but it seems the problem may become much more challenging if the *dynamics* of the fading are taken into account.

The issue of channel measurement and its relation to power control, and channel feedback, is an area that warrants more study. Note that if one measures interference at a high rate then one can update power levels at a high rate, but the measurements tend to be more noisy. This raises the question as to whether it is better to make accurate measurements, and use up the feedback bandwidth this way, or make coarser measurements at a higher rate, and use up the feedback bandwidth with a high rate 1 bit feedback, as in the closed-loop of IS95 [82]. We note that the proposals for third generation power control involve more precise interference measurements being sent back to the mobile ([12]). The correct approach will depend on the timescale of the fading effects.

The issue of channel measurement also raises the question as to whether it is better to measure and then adapt to the actual realizations of fading, or to adapt to the statistics of the fading only. Again, this question depends on the time-scale of the fading effects. For example, with multipath, one might measure over a window short enough that the statistics remain constant, but long enough to obtain averaging. Clearly, there is a limit to the rate of fading, if power adaptations are to be functions of channel realizations, and in IS95 it is assumed that multipath effects are too fast for this at vehicular speeds [80].

In the theoretical power control models we have considered in the present paper, power updates and interference measurements are assumed to occur at about the same rate. Similarly, the IS-95 closed loop causes the mobile's transmit power to adapt to realizations of the shadow fading of the user, and also attempts to measure interference at the same rate. However, it updates the E_b/N_0 setpoint on a slower time-scale based on frame error rates, which can be thought of as an adaptation to interference statistics, and the statistics of multipath effects. Note that the power updates occur more rapidly than the measurements of these statistics. In the future third-generation systems, it may be possible to more directly measure the multipath fading realizations, since there will be coherent reception on the uplink. Coherent detection requires a tracking of the multipath fading effects at the receiver, and this is possible in third generation systems because a pilot signal will be used on the uplink ([12]). It is in principle possible for the transmit power updates to be very fast and actually track the multipath fading, especially if a high rate 1 bit feedback is used. However, it may not be possible to accurately measure and feed back the interference effects at this same rate, and therefore it may still be preferable to measure the statistics of the interference over a longer timescale. Thus, there may need to be a separation of time-scale between power updates and interference measurements. This aspect is not currently a feature of power control analyses.

The works reviewed on power control for data traffic seem to be of a rather preliminary nature. Here again, time-scales are important; some traffic types *e.g.* voice and video, are very sensitive to the slightest delay, and would rather lose packets than suffer any delay; others, such as web browsing, can tolerate small delays, but are inelastic on longer time-scales; and emails and off-line file transfers may be quite elastic, with most emphasis on accuracy, and very little on delay. The problem of power control for data is clearly intimately connected with that of flow control, and a holistic approach is required in which both teletraffic and radio propagation issues are considered together. See [15] for some preliminary work in this direction.

Mobility and the variation over time of traffic patterns occur on a relatively slow timescale, yet are clearly important issues for call admissions. A paper that tries to relate power control to the traffic pattern is [28], but this work does not consider time-evolution at all. Nevertheless, it suggests that power control and call admissions may well be intimately connected. This is also the view taken in [6], but again, call admission based on channel probing along the lines described in [6] makes the assumption that the users are essentially immobile; this may be a reasonable assumption for some important wireless systems, but it remains true that combined study of power control and call admissions taking into account mobility, remains an open area. In summary, very little attention has been paid to date to the interaction between power control (which is usually thought of as a physical layer control) and networking layer issues.

In the information-theoretic paradigm, the power and rate control problem formulations we surveyed all assume instantaneous and perfect channel state information at both the transmitters and the receiver. In practice, such information is obtained via measurement and feedback from the receiver to the transmitter. In fast time-varying channels, the channel state information may be inaccurate due to measurement errors and delay in the feedback link. An interesting problem is the impact of such imperfection on capacity and on the optimal power and rate control strategies. Interesting results are obtained recently [79] in the context of a point-to-point time-varying channel with delayed feedback.

The information theoretic capacity of a network of cells is still very open. Indeed, if the interference from other cells is to be treated as noise (rather than decoded jointly via a connected "antenna array" of base stations, as in [29], [87]) then the problem is a hybrid between a *multiple access channel* (because of the users within the cell) and an *interference channel* (because of the other-cell users) [11]. As such, it is an open problem to characterize the Shannon capacity region for fixed powers, even if we ignore the issue of resource allocation altogether. Indeed, the characterization of capacity for the general interference channel has been an open problem in information theory for many years [71].

References

[1] Adachi F., Sawahashi M. (1998) "Wideband DS-CDMA for Next-generation Mobile Communications Systems" *IEEE Communications Magazine* Vol. 36., No. 9., September, pp56-69.

[2] Aein J.M. (1973) "Power balancing in system employing frequency reuse" *COMSAT Tech. review* Vol 3. 1973

[3] M. Andersin, Z. Rosberg, and J. Zander (1998) "Distributed Discrete Power Control in Cellular PCS" *Wireless Personal Communications* Vol. 6, No. 3.

[4] Andrew L., Hanly S.V. (1999) "Performance of a global congestion measure for CDMA networks" *to appear in IEEE VTC 99*

[5] Bambos N., Pottie G.J. (1992) "Power control based admission policies in cellular radio networks" *IEEE Global Telecomm. Conf. GLOBECOM92*

[6] Bambos N., Chen S., Pottie G. (1994) "Radio link admission algorithms for wireless networks with power control and active link quality protection" *Tech. Report UCLA-ENG-94-25, UCLA Scool of Engineering and Applied Science.*

[7] Bambos N. (1998) "Toward Power-Sensitive Network Architectures in Wireless Communications" *IEEE Personal Communications* Vol. 5, No. 3, June: 50–59

[8] Carnahan B, Luther H.A., Wiles J.O. "Applied numerical anaylsis" Wiley

[9] Chan, C.C. and S.V. Hanly (1997) "The capacity improvement of an integrated successive decoding and power control scheme" *Int. Conf. Univ. Personal Commun. ICUPC97*, October, San Diego, pp800-804.

[10] R. Cheng, S. Verdu (1993) "Gaussian Multiaccess Channels With ISI: Capacity Region and Multi-user Water-Filling" *IEEE Trans. on Information Theory*, Vol. 39, May: 773–785.

[11] Cover, T. and J. Thomas (1991) "Elements of Information Theory" Wiley.

[12] Dahlman E., Beming P., Knutsson J., Ovesjo F., Persson M., Roobol C. (1998) "WCDMA - The Radio Interface for Future Mobile Multimedia Communications" *IEEE Trans. on Vehic. Techn.* Vol. 47., No. 4., Nov. pp1105-1117

[13] Edmonds, J. (1969) "Submodular functions, matroids and certain polyhedra" *Proc. Calgary Intl. Conf. on Combinatorial structures and appl.* Calgary, Alberta, pp. 69-87, June.

[14] Evans, J.S. and D.N. Tse. (1999) " Linear Multiuser Receivers in Multipath Fading Channels", submitted to *IEEE Transactions in Information Theory*, Feb.

[15] D. Famolari, N. Mandayam, D. Goodman, and V. Shah (1999) "A New Framework for Power Control in Wireless Data Networks: Utility and Pricing" *Wireless Multimedia Network Technologies*, Kluwer Academic Publishers.

[16] Federgruen, A. and H. Groenevelt (1986). "The Greedy Procedure for Resource Allocation Problems:Necessary and Sufficient Conditions for Optimality". *Operations Research* 34, 909-918.

[17] Foschini G. J., Milzanic Z. (1993) "A simple distributed autonomous power control algorithm and its convergence". *IEEE Trans. Vehic. Techn*, Vol. 42., No. 4.:641–646.

[18] Gilhousen, K. S., Jacobs, I. M., Padovani, R., Viterbi, A. J ., Weaver, L. A., Wheatley, C. E. (1991) "On the capacity of a cellular CDMA system" *IEEE. Transactions on Vehicular Technology*, Vol. 40 No. 2 May.:303- -312

[19] Goldsmith A. and P. Varaiya (1995) "Capacity of Fading Channel with Channel Side Information" *IEEE Trans. on Information Theory, Vol. 43, No. 6*, pp. 1986-1992, November.

[20] S. Grandhi, J. Zander, and R. Yates (1995) "Constrained Power Control" *Wireless Personal Communications*, Vol. 2, No. 3, Aug. , Kluwer.

[21] Hanly S.V. (1993) "Information Capacity of Radio Networks", PhD Thesis, Cambridge University, Aug.

[22] Hanly S. V. (1995) "An algorithm for combined cell site selection and power control to maximize cellular spread spectrum capacity". *IEEE Journal on Selected Areas, special issue on the fundamentals of networking, Vol. 13, No. 7* September.

[23] Hanly, S.V. (1996) "Capacity and power control in spread spectrum macrodiversity radio networks" *IEEE Trans. on Communications, Vol. 44, No. 2* Feb.

[24] Hanly S.V., Chan C.C. (1999) "Slow Power Control in a CDMA Network" *To appear in Proceedings of IEEE VTC 99.*

[25] Hanly,S.V. and D.N. Tse (1995) "Multi-Access Fading Channels: Shannon and Delay-Limited Capacities". *Proceedings of the 33rd Allerton Conference*, Monticello, IL, Oct.

[26] Hanly, S. V. and D.N. Tse (1998), "Multi-access fading channels: Part II: Delay-limited capacities", *IEEE Trans. on Info. Theory*, v. 44, No. 7, November, pp. 2816-2831.

[27] Hanly, S. V. and D.N. Tse (1999) "Resource Pooling and Effective Bandwidths in CDMA Networks with Multiuser Receivers and Spatial Diversity" *Submitted for publication*

[28] Hanly S.V. (1999) "Congestion measures in DS-CDMA networks". *IEEE Trans. on Communications*, Vol. 47, No. 3, March:426–437.

[29] Hanly S.V. and Whiting P.A. (1993) "Information theoretic capacity of multi-receiver networks" *Telecommunication Systems* 1:1–42.

[30] Hanly S.V. and Whiting P.A. (1994) "Constraints on capacity in a multi-user channel" *International symposium on information theory* Trondheim, Norway.

[31] Honig, M., U. Madhow and S. Verdu (1995) "Blind adaptive multiuser detection". *IEEE Trans. on Information Theory*, July, pp.944-960.

[32] Honig, M. and Kim. (1996) "Allocation of DS-CDMA parameters to achieve multiple rates and qualities of service". Globecom 96, London, Nov., 50a2.

[33] Huang C.Y., Yates R.D. (1998) "Rate of convergence for minimum power assignment algorithms in cellular radio systems" *Baltzer/ACM Wireless Networks*, Vol 4, No. 3, April: 223-231

[34] Kelly F.P. (1997) "Charging and Rate Control for Elastic Traffic" *European Transactions on Communications* Vol. 8, : 33-37.

[35] Kiran and D.N. Tse. (1998) "Effective Bandwidths and Effective Interference for Linear Multiuser Receivers in Asynchronous Systems". Submitted to *IEEE Transactions on Information Theory*, Nov.

[36] Knopp, R. and P.A. Humblet (1995) "Information capacity and power control in single-cell multiuser communications." *IEEE Int. Conf. on Communications*, Seattle, Wash., June.

[37] Knopp, R., and P.A. Humblet (1995) "Multiple Accessing over Frequency Selective Fading Channels" *Proc. IEEE Int. Symp. Personal, Indoor, Mobile radio Commun. PIMRC 95*, Toronto, Canada.

[38] Kumar, P. and J. Holtzman (1995) "Power control for a spread-spectrum system with multiuser receivers". *Proc. 6th IEEE Int. Symp. Personal, Indoor, Mobile radio Commun. PIMRC 95*, Sept, pp.955-959.

[39] Kumar P.S., Yates R.D., Holtzman J. (1995) "Power control based on bit error rate (BER) measurements". Milcom 95.

[40] Lupas, R. and S. Verdu (1989) "Linear multiuser detectors for synchronous code-division multiple access". *IEEE Trans. on Information Theory*,IT-35, Jan., pp.123-136.

[41] Lupas, R. and S. Verdu (1990) "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. on Communications*, COM-38, Apr, pp. 496-508.

[42] Madhow, U. and M. Honig (1994) "MMSE interference suppression for direct-sequence spread-spectrum CDMA". *IEEE Trans. on Communications*, Dec.,pp.3178-3188.

[43] Marcenko, V.A., Pastur L.A. (1967) "Distribution of eigenvalues for some sets of random matrices" *Math. USSR-Sb* 1. pp457–483.

[44] Mitra D. (1993) "An asynchronous distributed algorithm for power control in cellular radio systems" *Fourth winlab workshop on third generation wireless information networks*:249–257

[45] Mitra D., J. Morrison. (1996) "A distributed power control algorithm for bursty transmissions on cellular, spread spectrum wireless networks" *Wireless Information Networks* ed:J. Holtzman, Kluwer.

[46] Muller,R.R., A. Lampe and J.B. Huber. (1998) "Gaussian multiple-access channels with weighted energy constraints," *Proc. of IEEE Inform. Theory Workshop*, Killarney, Ireland, June, pp. 106-107.

[47] A.F. Naguib, A. Paulraj and T. Kailath (1994) "Capacity Improvement With Base-Station Antenna Arrays in Cellular CDMA" *IEEE Trans. on Vehic. Techn.*, Vol. 43, No. 3, Aug.:691–698.

[48] Nettleton, R. W., Alavi H. (1983) "Power control for spread spectrum cellular mobile radio system" *Proc. IEEE Veh. Tech. Conf.*:242–246

[49] Oh and Wasserman (1998) "Adaptive resource management for DS-CDMA networks subject to energy constraints". IEEE INFOCOM 98, San Francisco, March.

[50] Oh and Wasserman (1999) "Dynamic Spreading Gain Control in Multi-service CDMA Networks" *IEEE Journal on Sel. Areas. in Comm.* Vol. 17, No. 5: 918–927.

[51] Pickholtz, R. L., Schilling, D. L., Milstein L.B. (1982) "Theory of spread spectrum communications – a tutorial" *IEEE Transactions on Communications* Vol. Com30, No. 5 May

[52] Ramakrishna S., Holtzman J.M. (1997) "A Scheme for Throughput Maximization in a Dual-Class CDMA System" *Int. Conf. Univ. Personal Commun.* ICUPC97, San Diego, October.

[53] Ramakrishna S., Holtzman J.M. (1998) "A Scheme for Throughput Maximization in a Dual-Class CDMA System" *IEEE Journal on Sel. Areas in Commun.* Vol 16, No. 6: 830–844.

[54] Rapajic, P. and B. Vucetic (1994) "Adaptive receiver structures for asynchronous CDMA systems." *IEEE Journal on Sel. Areas in Commun.*, May, pp. 685-697.

[55] Rashid-Farrokhi, F., L. Tassiulas, and K.J.R. Liu (1996) "Joint Power Control and Beamforming for Capacity Improvement in Wireless Networks with Antenna Arrays." *Proc. IEEE GLOBECOM*, pp. I-555-559, London, Nov.

[56] Rashid-Farrokhi, F., L. Tassiulas, and K.J.R. Liu (1996) "Joint Optimal Power Control and Beamforming in Wireless Networks Using Antenna Arrays." *IEEE Transactions on Communications* Vol. 46, No. 10, Oct.: 1313–1324.

[57] Rupf, M., F. Tarkoy and J. Massey, (1994) "User-separating demodulation for code-division multiple access systems." *IEEE Journal on Sel. Areas in Commun.*, June pp.786-795.

[58] Shantikumar, J.G. and D.D. Yao (1992) "Multiclass queueing systems: Polymatroid structure and optimal scheduling control" *Operations Research*, 40(2):293-299.

[59] E. Seneta, "Non-negative matrices and markov chains," *Springer-Verlag, second edition,* 1981.

[60] Sampath A, Kumar P., Holtzman J. (1995) "Power control and resource management for a multimedia CDMA wireless system" *Proc. IEEE Int. Symp. Personal, Indoor, Mobile radio Commun.* PIMRC95, Toronto, Canada

[61] Shenker S. (1995) "Fundamental design issues for the future internet" *IEEE Journal Sel. Areas in Comms.* 13:1141–1149.

[62] Silverstein J.W., Bai Z.D (1995) "On the empirical distribution of eigenvalues of a class of large dimensional random matrices". *Journal of Multivariate Analysis* 54(2), pp. 175-192.

[63] TIA/EIA/IS-95-A (1995) "Mobile station - Base station Compatability Standard for Dual-mode Wideband Spread Spectrum Cellular System" *Tech. Report, Telecommunications Industry Association.*

[64] J.S.Thompson and P.M. Grant and B. Mulgrew (1996) "Smart Antenna Arrays for CDMA Systems" *IEEE Personal Communications*, Vol. 3, No. 5, Oct.: 16–25.

[65] Tse,D.N. and S.V. Hanly (1996) "Capacity Region of the Multi-Access Fading Channel under Dynamic Resource Allocation and Polymatroid Optimization." *Proc. of the IEEE Information Theory Workshop*, Haifa, Israel, June.

[66] Tse, D.N. and S.V. Hanly (1998) "Multi-acess fading channels: Part I: Polymatroid structure, optimal resource allocation and throughput capacities." *IEEE Trans. on Info. Theory*, v. 44, No. 7, Nov., pp. 2796-2815.

[67] Tse D.N., Hanly S.V. (1999) "Linear Multiuser Receivers: Effective Interference, Effective Bandwidth and Capacity." *IEEE Trans. on Information Theory*, Vol. 45, No. 2, Mar. : 641–657.

[68] Tse D.N., Hanly S.V. (1998) "Effective bandwidths in wireless networks with multi-user receivers." *IEEE INFOCOM 98, San Francisco*, March.

[69] Ulukus,S. and R.D. Yates (1998) "Adaptive power control and MMSE interference suppression," *Baltzer/ACM Wireless Networks* Vol. 4(6) : 489–496.

[70] Ulukus S., Yates R.D. (1998) "Stochastic Power Control for Cellular Radio Systems" *IEEE Trans. on Communications* Vol 46, No. 6:784–798

[71] E.C. Van der Meulen (1994) "Some reflections on the Interference Channel" *Communications and Cryptography: Two sides of one tapestry*, Kluwer Academic Publishers.

[72] Vembu,S. and A.J. Viterbi (1996) "Two Philosophies in CDMA - a comparison." *Proc. IEEE Vehicular Technology Conference*, pp. 869-873.

[73] Verdu,S. (1996) "Minimum probability of error for asynchronous Gaussian channels." *IEEE Trans. on Information Theory*, IT-32, Jan., pp.85-96.

[74] Verdu, S., (1986) "Optimum multiuser asymptotic efficiency." *IEEE Trans. on Comm.*, COM-34, Sept, pp. 890-897.

[75] Verdu, S. (1997) "Demodulation in the presence of multiuser interference: progress and misconceptions." *Intelligent Methods in Signal Processing and Communications*, Eds. D. Docampo et. al., Birkhauser, Boston.

[76] Verdu S., Shamai S. (1997) "Multi-user detection with random spreading and error correction codes: fundamental limits" *Allerton conference, 1997*

[77] Verdu S., Shamai S. (1999) "Spectral Efficiency of CDMA with Random Spreading" *IEEE Trans. on Information Theory*, Vol. 45, No. 2: 622–640.

[78] Viswanath, P., V. Anantharam and D.N. Tse. "Optimal sequences, power control and capacity of spread-spectrum systems with multiuser receivers", to appear in *IEEE Trans. on Information Theory*, Sept. 1999.

[79] Viswanathan,H. (1998) "Capacity of Markov Channels with Receiver CSI and Delayed Feedback." *Proc. of Int. Symp. on Info. Theory*, Cambridge, MA, Aug., p. 238.

[80] Viterbi A.J. (1995) "CDMA:Principles of Spread Spectrum Communications" *Addison-Wesley*

[81] Viterbi A.M., Viterbi A.J (1993) "Erlang capacity of a power controlled CDMA system" *IEEE J. Select. Areas Comms., Vol. 11:892–900* Aug.

[82] Viterbi A.J., Viterbi A.M., Zehavi E. (1993) "Performance of power-controlled wideband terrestrial digital communication" *IEEE Trans. on Comms* Vol. 41, No. 4, April, pp 559-569.

[83] Viterbi A.J., Viterbi A.M., Zehavi E. (1994) "Other-cell interference in cellular power-controlled CDMA" *IEEE Trans. on Communications* Vol. 42, No. 2/3/4 Feb/March/April.

[84] Viterbi A.J., Viterbi A.M., Gilhousen K.S., Zehavi E. (1994) "Soft handoff extends CDMA cell coverage and increases reverse link capacity" *IEEE J. Sel. Areas in Comm.* Vol. 12:1281–1288

[85] Warrier,D. and U. Madhow (1998) "On the capacity of cellular CDMA with controlled power disparities." *Proc. of VTC*, Ottawa, May, pp. 1873-1877.

[86] J.H. Winters, J. Salz and R.D. Gitlin (1994) "The Impact of Antenna Diversity on the Capacity of Wireless Communication Systems" *IEEE Trans. on Commun.* Vol. 42, No. 2/3/4, Feb/Mar/Apr: 1740–1751.

[87] A. Wyner (1994) "A Shannon-theoretic Approach to a Gaussian Cellular Multiple-access Channel" *IEEE Trans. on Inforamtion Theory*, Vol. 40., Nov. : 1713–1727.

[88] Xie, Z., R. Short and C. Rushforth (1990) "A family of suboptimum detectors for coherent multi-user communications." IEEE Journal on Sel. Areas in Commun., May, pp.683-690.

[89] Yates R.D, Huang C.Y. (1995) "Integrated power control and base station assignment." *IEEE Trans. Vehic. Techn.* Vol. 44 No.3:638–644.

[90] Yates R. (1995) "A framework for uplink power control in cellular radio systems." *IEEE Journal on Selected Areas, special issue on the fundamentals of networking, Vol 13, No. 7.*

[91] L.C. Yun and D.G. Messerschmitt, (1994) "Power Control for Variable QOS on a CDMA Channel." *Proc. IEEE MILCOM*, Fort Monmouth, NJ, Oct pp. 178-182.

[92] Zander, Z. (1991) "Near-optimum transmitter power control in cellular radio systems" *Radio Communications Systems, Royal Institute of Sweden, research report*

[93] Zander, Z. (1992) "Distributed cochannel interference control in cellular radio systems" IEEE Transactions on Vehicular Technology, Vol 41., No. 3., August, 1992, pp305-311.

[94] Zhang,J. and E.K.P. Chong (1998). "CDMA Systems in Fading Channels: Admissibility, Network Capacity and Power Control", in *Proc. of 36th Allerton Conference*, Monticello, Illinois, Sept.

[95] Zhang,J. and E.K.P. Chong. "CDMA Systems in Fading Channels: Admissibility, Network Capacity and Power Control". To appear in *IEEE Trans. on Information Theory.*

[96] Zhang,J., E.K.P. Chong and D.N. Tse (1998). "Output MAI distribution of linear MMSE multiuser receivers". Submitted for publication.