UNIVERSITY OF CALIFORNIA

Santa Barbara

Cross-Layer Design for Efficient Wireless Medium Access Using Adaptive Multiuser Detection

A dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy
in Electrical and Computer Engineering

by

Kristoffer Norman Bruvold

Committee in charge:

Professor Upamanyu Madhow, Chair

Professor Elizabeth Belding-Royer

Professor Rene L. Cruz

Professor Jerry D. Gibson

Professor Kenneth Rose

Professor John J. Shynk

December 2005

The dissertation of Kristoffer Norman Bruvold is approved.

Elizabeth Belding-Royer

Rene L. Cruz

Jerry D. Gibson

Kenneth Rose

John J. Shynk

Upamanyu Madhow, Committee Chair

September 2005

Cross-Layer Design for Efficient Wireless Medium Access Using Adaptive Multiuser Detection

by

Kristoffer Norman Bruvold

Dedicated to my parents, Erik, Iva

and

Rally.

ACKNOWLEDGMENTS

This dissertation represents a major milestone in my life and it has been reached, and the next steps will be taken with the support and friendship of some people very special to me. I would like especially to thank my parents whose love, advice, and encouragement have always inspired me to take the next steps; my brother, Erik, whose art and creativity have been an inspiration; Iva, who has been an unmatched source of love, laughter and friendship; my Grandma, who, even though she missed the grade where you learn fractions, was the first person published in my family; Pat DuBois for her tireless support; Miki and Manjo for their warmth; and Rally, who has always kept me company and chased away the stress.

I would like to thank my advisor, Professor Madhow, for his vision, guidance, support and limitless patience. Thank you to my committee members, Professors Belding-Royer, Cruz, Gibson, Rose, and Shynk for their discussion and comments in bringing this project to fruition. And navigating all the channels at UCSB would have been impossible without Valerie De Veyra. I received continued motivation also from the students who were in my T.A. sections over the years.

There have been a few key individuals along the way who helped even to reach this point, and I am thankful to Ms. Koslove, Mrs. Michaels, and Prof. G.

Finally, this has been an education in life for me, and I am thankful for the people who have enriched all its aspects: Raghu for his unlimited capacity for debate on *any* topic; Noah for the calls to get to the gym; Suk-seung for the study hours; Kris C. and John K. for keeping a great environment; Nick Angelov for laughing (at me); Steve T. for his energy; John B. for that mid-western connection. I would like to thank a great friend and light in the darkness, Jelena. A host of others who all make my experience here special: Doca, Dragise, Jovana, Iva J., Ken S., DJ Roop (holla!), Roger S., Gwen, Sumit, Jaspreet, Zhenwen, Vebjorn, Una, Big Mike, Joong-ho & Mija, Taehyuk, JArko, Gmavs, Pdiggs, Neyba & Alleycat, Mad-Ox, Miller & Kreisel.

Thank you to you all, I am truly lucky,

Kristoffer Bruvold

VITA

KRISTOFFER NORMAN BRUVOLD

September 2005

EDUCATION

2005	*Doctor of Philosopy* (expected) Electrical and Computer Engineering University of California, Santa Barbara, California
2000	*Master of Science* Electrical and Computer Engineering University of California, Santa Barbara, California
1998	*Bachelor of Science* Electrical and Computer Engineering The Ohio State University, Columbus, Ohio

EXPERIENCE

2001-2005	*Research Assistant* Electrical and Computer Engineering University of California, Santa Barbara, California
1999-2005	*Engineering Intern* Communications Division Jet Propulsion Laboratory, Pasadena, California
1998-2001	*Teaching Assistant* Electrical and Computer Engineering University of California, Santa Barbara, California
Summer 1998	*Engineering Intern* Wireless Communications Xetron Corporation, Cincinnati, Ohio
Summer 1997	*Summer Intern* AT&T Wireless Services, Chicago, Illinois

PUBLICATIONS

Upamanyu Madhow, Kristoffer N. Bruvold, and Liping Julia Zhu, "Differential MMSE: A framework for robust adaptive interference suppression for DS-CDMA over fading channels," *IEEE Transactions on Communications*, vol. 53, no. 8, pp. 1377–1390, Aug. 2005.

Kristoffer N. Bruvold and Upamanyu Madhow, "A cross-layer design with multipacket reception for handoffs in pseudocellular networks," *in preparation.*

Kristoffer N. Bruvold, Raghuraman Mudumbai, and Upamanyu Madhow, "A QoS framework for stabilized collision channels with delay-constrained, mixed-priority traffic and multipacket reception," *in preparation*.

Kristoffer N. Bruvold, Raghuraman Mudumbai, and Upamanyu Madhow, "A QoS framework for stabilized collision channels with mulituser detection," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, Seoul, Korea, May 16–20, 2005, pp. 250–254.

Kristoffer N. Bruvold and Upamanyu Madhow, "Cross-layer optimization of the reservation channel in a pseudocellular network: Mobile-centric fast handoffs via multiuser detection," in *Proc. 37th Asilomar Conf. on Signals, Systems and Computers*, vol. 2, Pacific Grove, CA, Nov. 9–12, 2003, pp. 1733–1737.

Kristoffer N. Bruvold and Upamanyu Madhow, "Adaptive multiuser detection for mobile-centric fast handoffs in pseudocellular wireless networks," in *Proc. 58th IEEE Veh. Technol. Conf.*, vol. 2, Orlando, FL, Oct. 6–9, 2003, pp. 796–800.

ABSTRACT

Cross-Layer Design for Efficient Wireless Medium Access Using Adaptive Multiuser Detection

by

Kristoffer Norman Bruvold

We jointly optimize the physical (PHY) and medium access control (MAC) layers for enabling rapid, contention-based wireless medium access. The key new features of our cross-layer design are multipacket reception (MPR) capability and the ability to estimate the number of contending users, regardless of success or failure. Users contend by randomly choosing contention slots for transmitting, randomly choosing the "phase" of a common periodic training sequence known to the access point (AP), and randomly choosing short spreading sequences not known to the AP. Adaptive multiuser detection based on the Differential Minimum Mean Squared Error (DMMSE) criterion is employed, thus permitting reliable demodulation of an unknown number of contending users in a contention slot, despite multiuser interference (and possibly a near–far problem), fading, lack of carrier synchronization, and lack of knowledge of the users' spreading sequences. We show that even without coordination between the mobiles and the AP, multiple simultaneous contentions are successful in a contention frame with high probability, thus drastically reducing delay compared to contention based on classical, narrowband ALOHA. We characterize the throughput of the system and design dynamic stabilization policies based on estimates of the backlog, which are generated by extending Rivest's pseudo-Bayesian technique for classical Aloha to exploit the new information from the PHY on the number of transmissions.

The application focus of our cross-layer design is towards enabling rapid, mobile-centric, handoffs in *pseudocellular* networks with small AP coverage areas as in Wireless Local Area Network technology, but supporting real-time applications under vehicular

mobility as in cellular networks. To develop a design framework, we provide an analytical model of the cross-layer design that includes the new MPR and multiplicity feedback features and apply it to two classes of users: delay-constrained, *Hi*-priority users; and delay-tolerant, *Lo*-priority users whose throughput we wish to maximize, while guaranteeing QoS for the *Hi*-priority users. The channel throughput and the achievable QoS are characterized as functions of the arrival rates for *Hi* and *Lo*-priority users and we obtain contention policies that ensure QoS and stability. Finally, we apply these methods to simulations of the DMMSE-based cross-layer design, and show that the analytical model provides accurate guidelines for design and performance predictions.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Since the introduction of the ALOHA system by Abramson in 1970 [6], random multiple-access in wireless communication systems has mainly been limited to the single-packet reception (SPR) model, in which two or more simultaneous transmissions cannot be successful [6], [7], [11], [53], [85]. With the maturing of methods for multiuser detection, however, it is time for a fundamental reexamination [25], [93], of how the basic ALOHA model can be enhanced by a systematic design that exploits multipacket reception (MPR). In this dissertation, we present a cross-layer design of the physical (PHY) and medium access control (MAC) layers, exploiting adaptive multiuser detection to provide a robust mechanism for rapid, contention-based medium access that is significantly more efficient than narrowband ALOHA. Multiuser detection is provided at the PHY by the *Differential* MMSE (DMMSE) criterion introduced in [63], [113]. Several key enhancements to DMMSE introduced here (and published in [61]) enable its use in our cross-layer design. For the new MPR channel, we analytically determine the maximum attainable throughput and stabilization policies designed to operate near this throughput. These results, which apply to delay-insensitive traffic, show a significant improvement compared to narrowband ALOHA. However, the advantages of the design are even more apparent for the support of delay-constrained traffic, where they provide a mechanism for uncoordinated medium access with tight quality of service (QoS) guaran-

tees. A specific application that motivates this work is to *pseudocellular* networks, which employ a low-cost infrastructure of densely deployed wireless local area network (WLAN) access points (APs) to support functionalities similar to digital cellular telephony.

1.1 Pseudocellular Networks: An Application

802.11-based WLANs [3] have seen widespread adoption because of their low cost, ease of deployment, and significantly higher speeds compared to second and third generation cellular networks [1], [4], but are currently restricted to low-mobility data applications. The use of WLAN technology to support a broader range of applications, including a range of mobility profiles (e.g., quasistationary, pedestrian, vehicular) and traffic classes (e.g., voice, data, streaming audio/video) is an attractive option for deployment on academic and industrial campuses, or for emergency plug-and-play networks. Traditional approaches to such a range of user mobilities are based on an hierarchical structure, with large cells serving highly mobile users and small cells serving less mobile users [29], [48]–[50], [82], [102]. In contrast, a primary goal of this dissertation is to allow mobility on the scale of that supported by cellular networks using a flat, cost-effective architecture based on the small cells and unstructured geometries consistent with a WLAN-style deployment. Such a *pseudocellular* architecture (e.g., Fig. 1.1) is appropriate for plug-and-play networks, but, unlike in peer-to-peer ad hoc networks, mobile units, typically operating at lower power and complexity, need only to reach the nearest AP. The APs themselves may be more capable nodes connected via a wireless ad hoc network or through a wireline network.

The critical technical problem in realizing the preceding pseudocellular architecture is supporting the higher rate of handoffs for highly mobile users with real-time traffic: given the relatively small pseudocells, the frequency of handoffs for vehicular mobility is much higher than that of conventional cellular networks which employ significantly larger cells. With a reliable MPR design, the well-known short reaction times of mobile-controlled

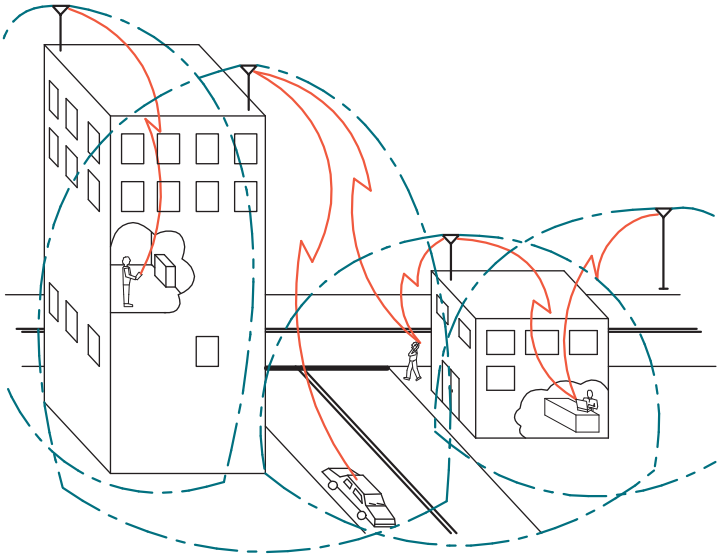

Figure 1.1. Pseudocellular network concept.

handoffs [94] may be leveraged by absorbing the task of handoff into the MAC: a mobile detecting a stronger beacon from an adjacent access point simply asks for a reservation. Assuming that the AP is capable of the sophisticated signal processing required for adaptive multiuser detection, joint PHY/MAC optimization of the medium yields a powerful mechanism for making multiple simultaneous, uncoordinated reservations. We can therefore achieve rapid, mobile-centric, handoffs using in-band, contention-based reservations, and can meet the tight handoff delay budgets required for sustaining a real-time application at vehicular speeds. Once a reservation is successful, we assume that the mobile has contention-free access (e.g., using polling or TDM/TDMA) until its resource requirements change or it leaves the pseudocell.

1.2 Cross-Layer Design

The cross-layer design we present in this dissertation uses a time-slotted contention channel, analogous to slotted-ALOHA [85]. As shown in Fig. 1.2, to contend in a *contention frame*, a user randomly chooses a *contention slot* in which to transmit and sends

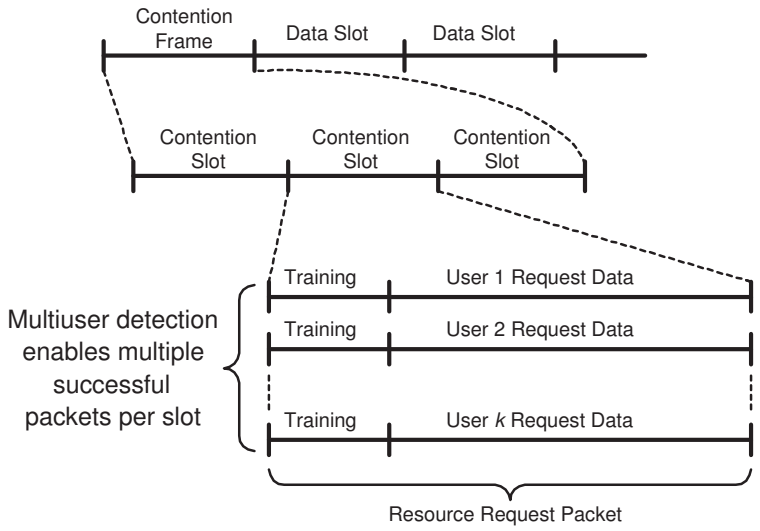

Figure 1.2. The structure of a contention frame.

a sequence of training symbols, followed by data symbols containing the reservation request. These symbols are spread using a "short" direct sequence spread spectrum (DSSS) spreading sequence whose period equals the symbol interval.

Since contention is the first communication from mobile to AP, prior coordination is unavailable. Although the set of training sequences from which the users randomly choose is established *a priori* (e.g., through broadcast in the AP's beacon), the AP does not know how many users are contending, nor which training sequences they choose. Through the use of DMMSE-based reception [61], [113], the AP receiver is capable of detecting and decoding an unknown number of transmitted packets simultaneously, *without prior coordination between the AP and the users.* The choice of DMMSE-based adaptive multiuser detection permits reliable demodulation of multiple contending users in a contention slot, despite multiple access interference (MAI), a possibly severe near–far problem, fading, lack of carrier synchronization, and lack of knowledge of the users' spreading sequences. As a result, even without coordination between the mobiles and the AP, multiple contentions are successful in a contention frame with high probability. Through analysis and simulation, we show that the capacity of this MPR channel

is significantly higher than traditional slotted-ALOHA, and that the maximum stable throughput supports much higher traffic arrival rates. Furthermore, the new extensions to the DMMSE receiver that enable the blind detection at the AP also enable accurate estimates of the number of simultaneous transmissions, regardless of whether they are successful or not. This is employed to obtain accurate backlog estimates by the AP, which in turn are used for dynamic stabilization policies that improve performance for delay-insensitive traffic.

1.3 QoS Provisioning for Mixed Traffic

The random access channel described in Section 1.2 yields a powerful mechanism for achieving rapid, mobile-centric, handoffs using in-band, contention-based reservations. However, the contention policies described in Section 1.2 for maximizing the throughput for delay-insensitive traffic do not ensure, for example, that a user in a vehicle moving into a neighboring pseudocell will complete a handoff before their voice connection is dropped. Therefore, we develop an analysis of the preceding cross-layer design in order to provide a design framework for the support of traffic with strict delay constraints in the random-access contention channel.

To obtain a general design framework for stabilization and QoS for a MPR-capable collision channel, we abstract an analytical model of the system in Section 1.2. The model is first applied to a single traffic class with strict delay constraints to obtain bounds on the system's performance, and then extended to a two-class mixture of delay-constrained and delay-tolerant traffic. The result is a multi-dimensional Markov chain description that specifies the throughput as a function of the traffic arrival rates. In the presence of mixed traffic, both the delay-constrained and delay-tolerant traffic may become backlogged. As for the case of delay-tolerant traffic only, lacking *any* controls, the throughput may go to zero, causing the real-time connections to be dropped. Therefore, we extend the backlog estimation method for homogeneous, delay-tolerant traffic to mixed-priority

traffic. From the backlog estimates, we are able to dynamically generate contention policies that stabilize the throughput while providing the required QoS to users with tight delay budgets, even in the presence of a heavy backlog.

1.4 Background

Our starting point is Abramson's classical work on ALOHA [6]. Previous studies [84], [96] of the standard, SPR slotted-ALOHA contention channel [6], [85] have shown that the backlog of users who have arrived in preceding time slots, but have not yet successfully contended, can be stabilized under certain conditions on the traffic arrival rate (namely, only when the rate of new arrivals is less than $e^{-1} \approx 0.367$ [96]). Further, even for a large backlog, the maximum throughput can be maintained through the use of a controlled access scheme [41], [84], [96]. There have been attempts to enhance the basic ALOHA model by including physical layer capabilities. By sensing the SPR channel as busy or idle, users may avoid collisions and the throughput may be improved [54], [92]. However, the potential delay from waiting for an idle channel [54] makes support of real-time traffic difficult. Another technique is capture [22], [37], [40], [58], [69], [79], [117], in which a stronger transmission may be successful even if there are some weaker simultaneous transmissions. The effect of capture in an ALOHA system can be enhanced by employing spread-spectrum techniques [8], [9], [22], [77], [83], in which users employ symbol spreading, or space-division multiple access (SDMA) [106], [116], wherein multiple antennas are used to isolate multiple received transmissions. While capture may allow a user to be decoded correctly even when there are multiple users present, the use of multiuser detection takes this one step further, allowing multiple users to be decoded despite potential near–far problems. Initial applications of multiuser detection to the multiple access channel have only partially addressed key issues. In [52], a simple, decision-feedback approach required coherent demodulation and would be inapplicable when the number of transmissions is not known *a priori*. Instead, adaptive techniques,

such as in [80], are required for demodulating an unknown number of received signals. However, unlike the DMMSE-based design presented in Chapter 3, an MMSE-based design as in [80] would not be able to support mobility-induced fading channels.

An abstract model for the enhancements in medium access possible using MPR capabilities, including stabilization techniques, has been investigated in [33], [34]. We build on these ideas for stabilization of our MPR random access scheme, using a backlog estimation procedure that extends the pseudo-Bayesian technique in [26], [84], which relied on only ternary feedback, to take advantage of the additional information available from the DMMSE receiver. However, our work differs from that in [33], [34] in two key aspects: (a) we provide a *detailed* PHY/MAC design, accompanied by extensive simulations for a specific adaptive multiuser detector that is well-matched to uncoordinated random access, and, (b) we consider support of delay-constrained traffic, in addition to our results on throughput and stabilization.

1.5 Organization

The remainder of this dissertation is organized as follows. Chapter 2 gives a description of DMMSE multiuser detection [61], [113], detailing the key new features that enable the joint PHY/MAC design described in Chapter 3. In Chapter 4, the capability of the PHY to detect the number of users who have collided, regardless of success or failure, is used for multiple-priority-class backlog estimation and the design of contention policies satisfying tight QoS constraints. Lastly, Chapter 5 discusses possible extensions of our joint PHY/MAC design to further aspects of pseudocellular wireless network design and concludes this work.

Chapter 2

Differential MMSE

Detecting and decoding wireless transmissions from a desired user in the presence of multiple interfering transmissions and noise is a well-studied problem [101]. For the pseudocellular system described in Chapter 1, we require an interference suppression technique that is robust to fading and frequency/phase asynchronism, and therefore can form the basis for efficient support of multiple, loosely-coordinated users. To achieve these capabilities, we employ reception based on the Differential Minimum Mean Squared Error (DMMSE) criterion by extending its functionalities for the unique features of the cross-layer design in Chapter 3. DMMSE was previously introduced in [63], [113]. In that work, the DMMSE criterion and its solution were derived, and its relationship to the standard MMSE receiver was described. For frequency-nonselective fading, adaptive implementations of the single-correlator DMMSE receiver were tested, and both synchronous and asynchronous simulations were done to show the robustness to fading and multiple access interference (MAI), and the capability for implicit timing acquisition. Finally for frequency selective fading, the *eigenrake* receiver was proposed, but not fully developed. In this thesis, we address several of the open questions on DMMSE by way of presenting several refinements of the previous theory, including some that are key to our cross-layer design in Chapter 3. In order to provide a self-contained presentation, we summarize key aspects of the previous work in-line with our new contributions. The

specifically new contributions we make are:

- For the flat-fading channel, we provide a new theorem which shows that the DMMSE criterion results in an interference suppressing receiver that is a scalar multiple of the standard MMSE receiver, inheriting its robustness to multiple access interference and the near–far problem.
- We give a detailed analysis of the convergence properties of gradient-descent methods for adaptive DMMSE reception under flat fading. This includes a characterization of the effects of the scaling for a Recursive Least Squares (RLS) adaptation with the previously prescribed scaling of [113], [115], as well as a new, less noisy scaling method we introduce.
- For the frequency-selective fading channel, we provide a new theorem that explicitly characterizes the structure of the DMMSE *eigenrake* receiver. This theorem is the natural extension to frequency-selective fading channels of the interference suppression theorem we introduce for flat-fading channels.
- We introduce a new capability for the eigenrake receiver to blindly detect multiple paths received from the desired user and automatically combine them noncoherently. The result is the new *selective eigenrake* receiver. The capability for blind detection of user transmissions is a crucial component that enables several key features of the cross-layer design in Chapter 3.

This Chapter is organized as follows. First, in Section 2.1, we discuss the motivation for DMMSE-based reception and describe the basic received signal and linear receiver models. We summarize the existing work on DMMSE for the flat-fading channel in Section 2.3, where we provide a new theorem to show that DMMSE inherits key properties of the MMSE interference suppression.In Section 2.4, we describe the DMMSE-based *eigenrake* receiver first introduced in [113] for frequency-selective fading channels. We provide a new theorem that clearly describes the structure of the eigenrake receiver and establishes the intuition for the new *selective eigenrake* receiver capable of blind detection of the multipath signal from the desired user as well as multiple transmissions from different users.

2.1 Motivation

It has been known for some time now that, for Direct Sequence (DS) Code Division Multiple Access (CDMA) systems with short spreading waveforms (i.e., in which the period of the spreading waveform equals the inverse of the symbol rate), the MAI has a cyclostationary structure (at the symbol rate) which can be learned and exploited by an adaptive receiver. The resulting adaptive multiuser detection schemes provide large potential gains over conventional matched-filter receivers without requiring explicit estimates of the MAI parameters. In particular, the linear Minimum Mean Squared Error (MMSE) receiver for a desired user can be implemented adaptively either using a training sequence for that user [5], [62], [66], [81], or (semi-)blindly by using knowledge of the desired user's spreading waveform and propagation channel [45]. However, standard training-based adaptation is known to break down in the presence of time-varying channels typical of wireless environments [114], while blind adaptation as in [45] exhibits a higher misadjustment than training-based adaptation under ideal conditions and is vulnerable to mismatch due to errors in the receiver's estimate of the desired user's propagation channel.

In [63], [113], a new approach to adaptive interference suppression over rapidly time-varying channels based on the *Differential* MMSE criterion was introduced. This was a reformulation of the classical linear MMSE criterion wherein the quantity being tracked is interpreted as the ratio of two successive elements of the desired data sequence, rather than the raw data sequence. The result is that the DMMSE criterion leads to adaptive interference suppression techniques that are robust to channel time variations. The key idea behind the DMMSE criterion is the avoidance of the problem of channel compensation by exploiting instead the observation that, even for rapidly varying channels, the channel fading gains in two consecutive observation intervals are approximately the same.

For flat fading channels, the DMMSE criterion yields a number of adaptive algo-

rithms robust to channel time variations, with complexities comparable to (but slightly larger than) that of analogous algorithms based on the MMSE criterion. Under standard assumptions, the DMMSE correlator is a scalar multiple of the MMSE correlator and inherits the well-known [59], [62], interference suppression properties of the MMSE correlator, including its immunity to the near–far problem. For frequency-selective fading, the DMMSE criterion provides the starting point for obtaining the *eigenrake* receiver, which provides diversity as well as interference suppression, implicitly acquiring the timing of the significant multipath components for the desired user. As with standard MMSE adaptation, DMMSE adaptive algorithms require an initial training period in which the symbols transmitted by the desired user are known to the receiver, and can subsequently operate in decision-directed mode. The receiver does not require explicit knowledge of the spreading waveforms and propagation channels for either the desired user or the interfering users. Since DMMSE-based algorithms do not explicitly track the channel, they must either be used with a noncoherent demodulation technique (e.g., differential demodulation), or the channel information required for coherent demodulation must be obtained by some other means (e.g., by using pilot symbols).

2.2 Received Signal Model and Linear MMSE Reception

As in much of the multiuser detection literature (see [101] and the references therein), we restrict attention to a synchronous CDMA system for developing the basic properties of the proposed methods. It is known (e.g., see [60] and Chapter 2 in [101]) that interference suppression algorithms based on the synchronous model apply to an asynchronous CDMA system by reducing the latter to an "equivalent synchronous discrete-time model" which depends on the receive filter, sampler, and the length of the observation interval used for each symbol decision (a tutorial description of how this is done is provided in [60]). Thus, in this thesis, we consider a discrete-time, complex baseband, synchronous CDMA system where there are K users, with the desired user labeled as user 1. We

define $\mathbf{r}[n]$ to be the N-dimensional vector of samples obtained from the nth observation interval. For a flat-fading channel, $\mathbf{r}[n]$ is given by

$$\mathbf{r}[n] = F_1[n]b_1[n]\mathbf{u}_1 + \sum_{k=2}^{K} F_k[n]b_k[n]\mathbf{u}_k + \mathbf{w}[n], \tag{2.1}$$

where, for $1 \leq k \leq K$, and $\{a_k[n]\}$ representing the kth user's binary, antipodal data sequence, the $\{b_k[n] = a_k[n]b_k[n-1]\}$ are the kth user's differentially encoded symbols (independent across time and users), $\mathbf{u}_k$ is the signal vector for user k, and $\{F_k[n]\}$ is the sequence of complex fading gains seen by user k (independent of the user symbols and independent across users). The complex vector $\mathbf{w}[n]$ is discrete-time, additive, white Gaussian noise (AWGN) with variance σ^2 per dimension. For Rayleigh fading, the gains $\{F_k[n]\}$ are modeled as wide-sense stationary, zero mean, circular Gaussian random processes and are assumed to be positively correlated in consecutive intervals for the ith user. It is also convenient to introduce the *faded symbol* sequence $B_k[n] = F_k[n]b_k[n]$, $1 \leq k \leq K$, for each user. We make the following assumptions:

1. Symbols are zero mean, and independent across time and users: $E\big[b_k[n]\big] = 0$ for all k, n, and $b_k[n]$ is independent of $b_j[m]$ for $k \neq j$ or $n \neq m$.
2. Symbols are independent of fading gains: $b_k[n]$ is independent of $F_j[m]$ for any k, j, n, m.
3. For each user k, the sequence of fading gains $\{F_k[n]\}$ is wide-sense stationary.
4. The fading gains for different users are independent: for $j \neq k$, $F_j[m]$ is independent of $F_k[n]$ for all m, n.

In general, a linear receiver computes a decision statistic of the form $\big\langle \mathbf{c}, \mathbf{r}[n] \big\rangle = \mathbf{c}^H \mathbf{r}[n]$, where $\mathbf{x}^H$ denotes the complex conjugate transposed for a vector $\mathbf{x}$. The standard linear MMSE receiver minimizes the Mean Squared Error (MSE) between the desired user's symbol sequence $b_1[n]$ and the receiver output, given by $E\big[\big|b_1[n] - \big\langle \mathbf{c}, \mathbf{r}[n] \big\rangle\big|^2\big]$. The MMSE correlator is given by the formula

$$\mathbf{c}_{mmse} = \mathbf{R}^{-1}\mathbf{p}, \tag{2.2}$$

where $\mathbf{R} = E\big[\mathbf{r}[n]\big(\mathbf{r}[n]\big)^H\big]$ and $\mathbf{p} = E\big[b_1^*[n]\mathbf{r}[n]\big]$.

For Rayleigh fading in the model (2.1), assuming that $b_1[n]$ is uncorrelated with $F_1[n]$ and $B_k[n]$ for $k \neq 1$, we obtain that

$$\mathbf{p} = E\big[\big|b_1[n]\big|^2\big]E\big[F_1[n]\big]\mathbf{u}_1 = \mathbf{0}. \tag{2.3}$$

This implies that $\mathbf{c}_{mmse} = \mathbf{0}$ when averaged over the desired user's Rayleigh fading coefficient. Adaptive implementations [44] of the MMSE correlator may be viewed as replacing statistical expectations with empirical averages; for example, replacement of the statistical expectation in (2.2) by a block-based empirical average leads to the Block Least Squares implementation, while replacement by an exponentially weighted average corresponds to the Recursive Least Squares (RLS) algorithm. Thus, if the averaging time constant used by an MMSE-based adaptive algorithm is comparable to or larger than the coherence time of the fading, as is the case in many outdoor mobile wireless environments, we should expect poor performance by virtue of (2.3). For example, a normalized Doppler spread[1] of 0.01 corresponds to a coherence time of 100 symbols and could result from operating at a symbol rate of 20 Ksymbols/sec, a carrier frequency of 2 GHz, and a relative velocity between transmitter and receiver of approximately 100 km/h. In this setting, RLS adaptation employing an exponential forget factor of 0.99, which effectively averages over hundreds of symbols (i.e., an interval of the order of the coherence time), fails (see the simulation results in Section 2.5). Decreasing the averaging time in the adaptive algorithm would alleviate this problem, but would then provide insufficient averaging to overcome the effect of noise and interference.

Most commercial systems employ known pilot codes or pilot symbols in order to track the channel and perform coherent demodulation. In this case, the receiver may be able to estimate the fading gain $F_1[n]$,[2] and the MMSE criterion can be modified so as to track the faded symbol $B_1[n]$ using the cost function $E\big[\big|F_1[n]b_1[n] - \big\langle \mathbf{c}, \mathbf{r}[n]\big\rangle\big|^2\big] =$

[1]The normalized Doppler spread is the product, $f_d T_s$, of the maximum Doppler frequency, f_d, and the symbol period, T_s.

[2]Of course, accurate channel estimation prior to interference suppression may not be easy, especially when there is a near–far problem.

$E\big[\big|B_1[n] - \big\langle \mathbf{c}, \mathbf{r}[n]\big\rangle\big|^2\big]$. This eliminates the need to compensate for the fading gain. Assuming now that $B_1[n]$ is uncorrelated with $B_k[n]$ for $k \neq 1$, the channel-compensated solution is given by

$$\mathbf{c}_{cc} = \mathbf{R}^{-1}\mathbf{p}_{cc}, \tag{2.4}$$

where

$$\mathbf{p}_{cc} = E\big[B_1^*[n]\mathbf{r}[n]\big] = E\big[\big|B_1[n]\big|^2\big]\mathbf{u}_1 \sim \mathbf{u}_1, \tag{2.5}$$

so that the overall solution is proportional to $\mathbf{R}^{-1}\mathbf{u}_1$.

The channel-compensated MMSE solution can be interpreted as a standard MMSE solution in a time-invariant setting, except that the data being tracked is the faded symbol $B_1[n]$ rather than the symbol $b_1[n]$. From well-known properties of the MMSE solution [62], we can infer that channel-compensated MMSE is effective in interference suppression. This approach to dealing with channel time variations has been considered in several recent publications [46], [67].

2.3 DMMSE for Flat-Fading

In contrast to the channel-compensated MMSE approach in (2.4)-(2.5), the DMMSE criterion does not require explicit estimation of the fading gains for the desired user. It relies instead on the assumption that $F_1[n] \approx F_1[n{-}1]$ even for "fast" fading environments to obtain a correlator equivalent to the channel-compensated MMSE correlator. The formal statement of the DMMSE criterion is as follows [61], [113]:

The DMMSE Criterion: Find a correlator $\mathbf{c}$ that minimizes the following cost function:

$$J(\mathbf{c}) = E\left[\left|b_1[n]\big\langle \mathbf{c}, \mathbf{r}[n{-}1]\big\rangle - b_1[n{-}1]\big\langle \mathbf{c}, \mathbf{r}[n]\big\rangle\right|^2\right], \tag{2.6}$$

subject to

$$E\left[\big|\langle \mathbf{c}, \mathbf{r}[n]\rangle\big|^2\right] = \mathbf{c}^H\mathbf{R}\mathbf{c} = 1. \tag{2.7}$$

The intuition behind the preceding optimization problem is as follows. Given the difficulty in tracking $F_1[n]$, the adaptive receiver aims to achieve a more modest goal,

that of suppressing the interference and recovering the faded sequence $B_1[n] = F_1[n]b_1[n]$ *up to an arbitrary complex multiple,* α. A correlator $\mathbf{c}$ that achieves this goal will satisfy

$$\begin{aligned}\big\langle \mathbf{c}, \mathbf{r}[n-1]\big\rangle &\approx \alpha B_1[n-1] \approx \alpha F_1[n-1]b_1[n-1],\\ \big\langle \mathbf{c}, \mathbf{r}[n]\big\rangle &\approx \alpha B_1[n] \approx \alpha F_1[n]b_1[n],\end{aligned}$$

so that, assuming $F_1[n] \approx F_1[n-1]$, we have

$$b_1[n]\big\langle \mathbf{c}, \mathbf{r}[n-1]\big\rangle - b_1[n-1]\big\langle \mathbf{c}, \mathbf{r}[n]\big\rangle \approx 0.$$

This implies that (2.6) is the natural cost function to minimize. The solution to this problem is known [61], [113], to be the eigenvector corresponding to the largest eigenvalue of the following generalized eigenvalue problem:

$$\mathbf{A}\mathbf{c} = \lambda \mathbf{R}\mathbf{c}, \tag{2.8}$$

where

$$\mathbf{R} = E\Big[\mathbf{r}[n]\big(\mathbf{r}[n]\big)^H\Big], \tag{2.9a}$$

$$\mathbf{A} = E\Big[b_1[n]b_1^*[n-1]\mathbf{r}[n-1]\big(\mathbf{r}[n]\big)^H + b_1^*[n]b_1[n-1]\mathbf{r}[n]\big(\mathbf{r}[n-1]\big)^H\Big]. \tag{2.9b}$$

We note here that, with Assumptions 1-4 and definitions (2.9), the cost function (2.6) reduces to

$$\begin{aligned}J(\mathbf{c}) &= 2\mathbf{c}^H\mathbf{R}\mathbf{c} - \mathbf{c}^H\mathbf{A}\mathbf{c}\\ &= 2 - \mathbf{c}^H\mathbf{A}\mathbf{c},\end{aligned} \tag{2.10}$$

where we have used the constraint (2.7) to obtain the second equality.

We now invoke the specific features of the CDMA model (2.1) to show that the DMMSE correlator is equivalent to the channel-compensated MMSE solution in (2.4), and hence suppresses interference.

Theorem 2.1 ***(DMMSE Interference Suppression)*****:** Suppose that the fading gains in consecutive intervals for the desired user are positively correlated; that is, defining $\rho = E\big[F_1[n]F_1^*[n-1]\big]$, we require that $\Re\{\rho\} > 0$.

Then the correlator $\mathbf{c}$ that minimizes $J(\mathbf{c})$ in (2.6) is a scalar multiple of $\mathbf{R}^{-1}\mathbf{u}_1$ and is therefore equivalent to the channel-compensated MMSE solution. That is, it is a scalar multiple of the MMSE solution for demodulating the desired user's faded symbol, $B_1[n] = F_1[n]b_1[n]$. The DMMSE solution therefore inherits the interference suppression properties of the MMSE solution.

Proof: We start by computing the matrix $\mathbf{A}$ in (2.9b) for the model (2.1). Letting $\mathbf{T}_1 = E\Big[b_1^*[n]b_1[n-1]\mathbf{r}[n]\big(\mathbf{r}[n-1]\big)^H\Big]$ and plugging in the model (2.1), we obtain

$$\mathbf{T}_1 = E\Bigg[b_1^*[n]b_1[n-1]\sum_{k=1}^{K}F_k[n]b_k[n]\mathbf{u}_k\sum_{j=1}^{K}F_j^*[n-1]b_j^*[n-1]\mathbf{u}_j^H\Bigg].$$

Under Assumptions 1-4 on the independence of our model, it is straightforward to verify that

$$E\Big[b_1^*[n]b_1[n-1]\big(F_k[n]b_k[n]\big)\big(F_j^*[n-1]b_j^*[n-1]\big)\Big] = 0$$

unless $j = k = 1$. Thus, we obtain that

$$\begin{aligned}\mathbf{T}_1 &= E\Big[\big|b_1[n-1]\big|^2\Big]E\Big[\big|b_1[n]\big|^2\Big]E\Big[F_1[n]F_1^*[n-1]\Big]\mathbf{u}_1\mathbf{u}_1^H \\ &= \rho\mathbf{u}_1\mathbf{u}_1^H.\end{aligned}$$

Similarly, $\mathbf{T}_1^H = E\Big[b_1[n]b_1^*[n-1]\mathbf{r}[n-1]\big(\mathbf{r}[n]\big)^H\Big] = \rho^*\mathbf{u}_1\mathbf{u}_1^H$. We therefore obtain from (2.9b) that

$$\begin{aligned}\mathbf{A} &= \mathbf{T}_1^H + \mathbf{T}_1 \\ &= 2\Re\{\rho\}\mathbf{u}_1\mathbf{u}_1^H.\end{aligned} \tag{2.11}$$

Since this is a rank one matrix, there is a unique nonzero eigenvalue for the generalized eigenvalue problem (2.8), which can be rewritten as

$$2\Re\{\rho\}\mathbf{u}_1\mathbf{u}_1^H\mathbf{c} = \lambda\mathbf{R}\mathbf{c}. \tag{2.12}$$

Multiplying each side by $\mathbf{R}^{-1}$, it is clear that the generalized eigenvector corresponding to the nonzero eigenvalue is a scalar multiple of $\mathbf{R}^{-1}\mathbf{u}_1$. The condition $\Re\{\rho\} > 0$ ensures that the unique nonzero eigenvalue is positive. If this were not the case, then

the largest eigenvalue for the solution to the DMMSE problem would be zero. The latter corresponds to the zero correlator, which is useless for demodulation. The necessity of $\Re\{\rho\} > 0$ for obtaining a useful DMMSE correlator is, of course, not surprising, since the formulation of the DMMSE criterion is based on the assumption that the fading gains in successive intervals are approximately equal. This concludes the proof. □

Remark 2.1: When there is no fading, the left-hand side side of (2.12) becomes $2\mathbf{u}_1\mathbf{u}_1^H\mathbf{c}$ and the DMMSE correlator is still a scalar multiple of the standard MMSE correlator. Therefore, except for the standard penalty incurred for differential demodulation, the performance of the DMMSE and MMSE correlators are identical in the absence of fading.

2.3.1 Properties of Unconstrained DMMSE

Having explored the basic properties of the DMMSE criterion, we can now comment in more detail on the role of the constraint $\mathbf{c}^H\mathbf{R}\mathbf{c} = 1$. It is most convenient to discuss this under the assumption that the received signal has been prewhitened. Using $\tilde{\mathbf{x}}$ to denote the whitened version of $\mathbf{x}$, we have

$$\tilde{\mathbf{r}}[n] = \mathbf{R}^{-\frac{1}{2}}\mathbf{r}[n], \tag{2.13}$$

$$\tilde{\mathbf{u}}_1 = \mathbf{R}^{-\frac{1}{2}}\mathbf{u}_1, \tag{2.14}$$

$$\tilde{\mathbf{R}} = \mathbf{I}, \tag{2.15}$$

$$\tilde{\mathbf{R}}_I = \mathbf{R}^{-\frac{1}{2}}\mathbf{R}_I\mathbf{R}^{-\frac{1}{2}}, \tag{2.16}$$

where

$$\begin{aligned}\mathbf{R}_I &= \sum_{k=2}^{K} E\left[\left|B_k[n]\right|^2\right]\mathbf{u}_k\mathbf{u}_k^H + 2\sigma^2\mathbf{I} \\ &= \mathbf{R} - E\left[\left|B_1[n]\right|^2\right]\mathbf{u}_1\mathbf{u}_1^H\end{aligned} \tag{2.17}$$

is the correlation matrix for the interference and noise. From (2.17), we have that $\tilde{\mathbf{R}} = \mathbf{I} = \tilde{\mathbf{u}}_1\tilde{\mathbf{u}}_1^H + \tilde{\mathbf{R}}_I$. Thus, as long as $\tilde{\mathbf{R}}_I$ (the noise and interference correlation matrix

in the whitened domain) is positive definite, we have that $\mathbf{I} - \tilde{\mathbf{u}}_1\tilde{\mathbf{u}}_1^H$ is positive definite. This can be used to show that $\|\tilde{\mathbf{u}}_1\|^2 < 1$.

Now, consider the DMMSE cost function (without the constraint) in the whitened domain, specializing equations (2.10) and (2.11):

$$J(\tilde{\mathbf{c}}) = 2\tilde{\mathbf{c}}^H\tilde{\mathbf{c}} - 2\Re\{\rho\}\tilde{\mathbf{c}}^H\tilde{\mathbf{u}}_1\tilde{\mathbf{u}}_1^H\tilde{\mathbf{c}}. \tag{2.18}$$

The gradient of the preceding cost function (with respect to $\tilde{\mathbf{c}}^*$) is given by $2\tilde{\mathbf{c}} - 2\Re\{\rho\}\tilde{\mathbf{u}}_1\tilde{\mathbf{u}}_1^H\tilde{\mathbf{c}}$, so that a gradient descent update is of the form:

$$\begin{aligned}\tilde{\mathbf{c}}[n+1] &= \tilde{\mathbf{c}}[n] - \beta_a\Big(\tilde{\mathbf{c}}[n] - \Re\{\rho\}\tilde{\mathbf{u}}_1\tilde{\mathbf{u}}_1^H\tilde{\mathbf{c}}[n]\Big) \\ &= (1-\beta_a)\tilde{\mathbf{c}}[n] + \beta_a\Big(\Re\{\rho\}\tilde{\mathbf{u}}_1^H\tilde{\mathbf{c}}[n]\Big)\tilde{\mathbf{u}}_1.\end{aligned} \tag{2.19}$$

It is now easy to see what happens in the absence of the constraint. Note that, from the results in [62], it was shown in [61], [113] that in the whitened domain, the desired MMSE solution is simply $\tilde{\mathbf{u}}_1$. From (2.19), the component of $\tilde{\mathbf{c}}[0]$ orthogonal to $\tilde{\mathbf{u}}_1$ gets attenuated exponentially, as $(1-\beta_a)^n$, which is exactly the desired behavior. However, the component of $\tilde{\mathbf{c}}$ along $\tilde{\mathbf{u}}_1$ is also shrinking (but more slowly). This happens because the desired signal vector $\tilde{\mathbf{u}}_1$ has norm strictly less than one in the whitened domain. To see this, suppose that $\tilde{\mathbf{c}}[0] = \beta\tilde{\mathbf{u}}_1$. Then

$$\tilde{\mathbf{c}}[n] = \beta\gamma^n\tilde{\mathbf{u}}_1,$$

where $\gamma = 1 - \beta_a + \beta_a\Re\{\rho\}\|\tilde{\mathbf{u}}_1\|^2 < 1$, since $\|\tilde{\mathbf{u}}_1\|^2 < 1$. Thus, while $\tilde{\mathbf{c}}[n]$ points in the right direction, its norm is shrinking to zero (although this shrinkage can be slow if γ is close to one). We will see the consequence of this on adaptive implementations when we discuss the scaled RLS algorithm in Section 2.3.3.

2.3.2 Scaled RLS Algorithm

The scaled RLS adaptation [115] is as follows:

First, compute

$$\begin{aligned}
\mathbf{k}[n] &= \frac{b_1[n-1]\mathbf{P}[n-1]\mathbf{r}[n]}{\beta + \big|b_1[n-1]\big|^2 \big(\mathbf{r}[n]\big)^H \mathbf{P}[n-1]\mathbf{r}[n]}, \\
\mathbf{P}[n] &= \beta^{-1}\mathbf{P}[n-1] - \beta^{-1} b_1^*[n-1]\mathbf{k}[n]\big(\mathbf{r}[n]\big)^H \mathbf{P}[n-1], \\
\xi[n] &= b_1[n]\big(\mathbf{c}_{sc}[n-1]\big)^H \mathbf{r}[n-1] - b_1[n-1]\big(\mathbf{c}_{sc}[n-1]\big)^H \mathbf{r}[n],
\end{aligned} \tag{2.20}$$

then update the scaled RLS receiver, $\mathbf{c}_{sc}$,

$$\mathbf{c}_{sc}[n] = \frac{\mathbf{c}_{sc}[n-1] + \beta_a \mathbf{k}[n]\xi^*[n]}{\big|\big(\mathbf{c}_{sc}[n-1]\big)^H \mathbf{r}[n-1]\big|}. \tag{2.21}$$

The factor $\beta_a \leq 1$ represents a slowing down of the RLS update, which appears to help alleviate the effect of unreliable estimates during deep fades of the desired signal. This can be seen by noting that, analogous to (2.19), the numerator of (2.21) is the gradient descent update for the unwhitened cost function (2.10):

$$\begin{aligned}
\mathbf{c}[n] &= \mathbf{c}[n-1] - \beta_a\Big(\mathbf{R}\mathbf{c}[n-1] - \Re\{\rho\}\mathbf{u}_1\mathbf{u}_1^H\mathbf{c}[n-1]\Big) \\
&= \mathbf{c}[n-1] - \beta_a\Big(E\big[F_1[n]F_1^*[n]\big] - \Re\{\rho\}\Big)\mathbf{u}_1\mathbf{u}_1^H\mathbf{c}[n-1] \\
&= \mathbf{c}[n-1] + \beta_a\Big(\Re\{\rho\} - 1\Big)\Big(\mathbf{u}_1^H\mathbf{c}[n-1]\Big)\mathbf{u}_1,
\end{aligned} \tag{2.22}$$

where the second equality assumes that correlator $\mathbf{c}$ is suppressing interference. While $\Re\{\rho\}$ is generally close to 1, deep fades of the desired signal may cause it to drop towards zero. In this case, or even if $\Re\{\rho\}$ temporarily becomes negative, $\beta_a < 1$ in the last line of (2.22) reduces the impact on the adaptation. Further, under the realistic assumption that $F_1[n] \approx F_1[n-1]$, these conditions are brief and are mitigated through the averaging with β_a (thus, in the simulations in this Chapter, $\beta_a = 0.99$ with a training period of 60 symbols was more than adequate).

2.3.3 Effect of Scaling

We can now illustrate the discussion in Section 2.3.1 of gradient descent on the unconstrained DMMSE cost function via a concrete adaptive implementation. To this end,

consider an unscaled RLS algorithm, which is identical to (2.20)-(2.21) except that the scaling in the denominator is eliminated. An algorithm equivalent to the latter was reported in [47], where it was called differential least squares. Since RLS can be viewed as stochastic gradient descent in the whitened domain, we expect, from the discussion of gradient descent in Section 2.3.1, that the unscaled RLS solution will "point in the right direction," but will ultimately converge to the zero correlator. To see this, consider the explicit unscaled RLS iteration, removing the scaling from (2.21):

$$\begin{aligned} \mathbf{c}_{us}[n] &= \mathbf{c}_{us}[n-1] + \beta_a \mathbf{k}[n]\xi^*[n] \\ &= \mathbf{Q}[n]\mathbf{c}_{us}[n-1], \end{aligned} \tag{2.23}$$

with

$$\mathbf{Q}[n] = \mathbf{I} + \beta_a \mathbf{k}[n]\Big(b_1[n]\mathbf{r}[n-1] \; - \; b_1[n-1]\mathbf{r}[n]\Big)^H.$$

As shown in Fig. 2.1, the unscaled RLS correlator does converge to zero, while the scaled RLS correlator does not. However, the correlator magnitude for the latter fluctuates wildly due to the stochastic scaling we employ (as discussed later, an averaged scale factor may be more appropriate for practical implementations).

In order to further explore the issue of scaling, we iterate (2.23) to obtain

$$\mathbf{c}_{us}[n] = \mathbf{V}[n]\mathbf{c}_{us}[0], \tag{2.24}$$

where $\mathbf{V}[n] = \prod_{i=1}^{n} \mathbf{Q}[i] = \mathbf{V}[n-1]\mathbf{Q}[n]$. In Fig. 2.2, we plot the largest eigenvalue of $\mathbf{V}[n]$; its decrease with n implies, from (2.24), that $\mathbf{c}_{us}[n]$ converges to the zero correlator.

Moreover, it is easy to see that the scaled RLS iteration (2.21) yields a scalar multiple of the unscaled correlator, as follows:

$$\mathbf{c}_{sc}[n] = \frac{1}{\left|\big(\mathbf{c}_{us}[n-1]\big)^H \mathbf{r}[n-1]\right|}\mathbf{c}_{us}[n]. \tag{2.25}$$

Thus, for implementations which represent the correlators with sufficient precision, the BER performance of the scaled and unscaled RLS correlators should be precisely the same. We have verified from our double precision floating point Matlab simulations that this is indeed the case.

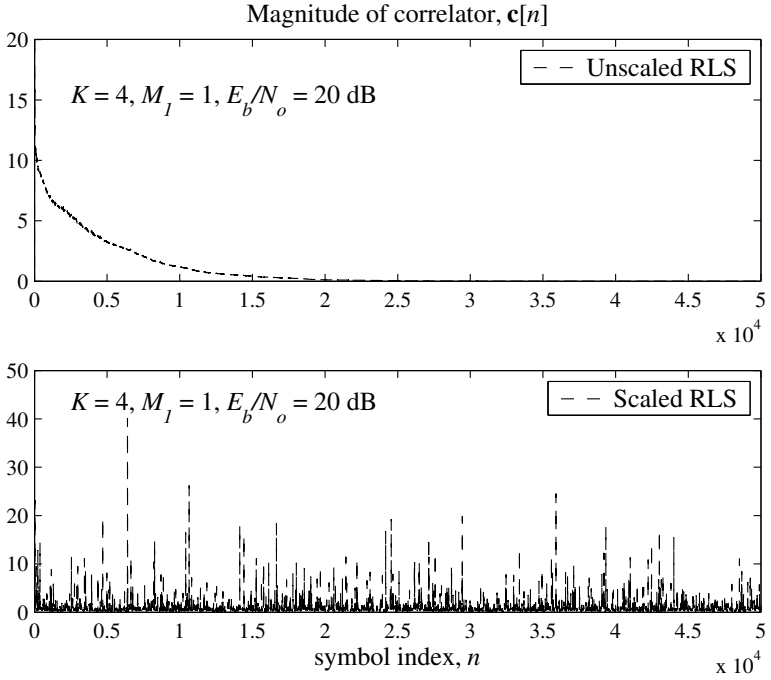

Figure 2.1. Magnitude of correlator, $\mathbf{c}[n]$, for unscaled (top) and scaled (bottom) RLS adaptations with $E_b/N_0 = 20\,\text{dB}$.

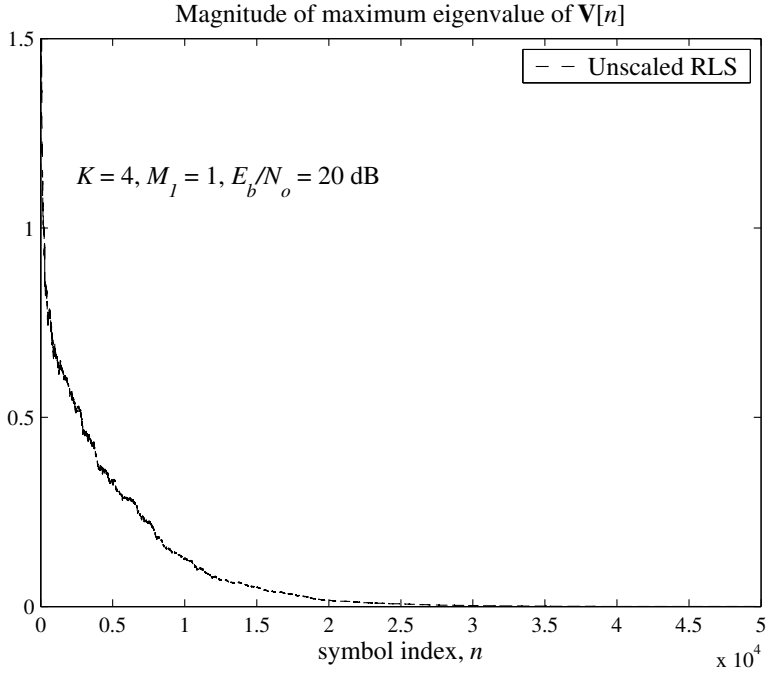

Figure 2.2. Magnitude of maximum eigenvalue of $\mathbf{V}[n]$ for the unscaled RLS adaptation with $E_b/N_0 = 20\,\text{dB}$.

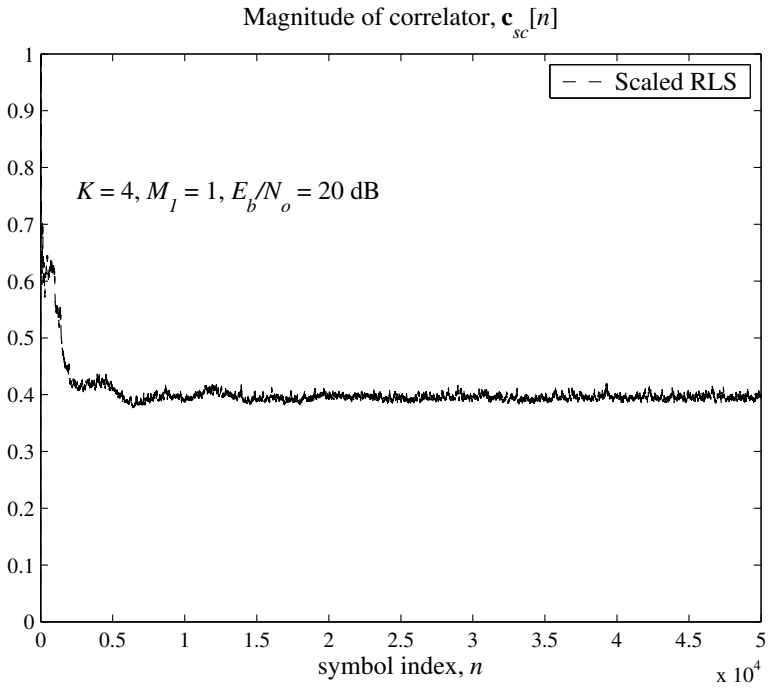

Figure 2.3. Magnitude of correlator, $\mathbf{c}_{sc}[n]$, using scaling of Eqn. 2.26 with $E_b/N_0 = 20\,\mathrm{dB}$.

There are several practical implications of the preceding observations. First, any scaling that keeps $\mathbf{c}_{us}[n]$ away from zero will work. The particular scaling in the scaled RLS algorithm introduced in [115] and reproduced in (2.20) is actually quite noisy, and a more appropriate scaling might be based on averaging the power at the output of the correlator using an empirical estimate of the correlation matrix, $\hat{\mathbf{R}}[n] = \frac{1}{n}\sum_{i=1}^{n} \mathbf{r}[i](\mathbf{r}[i])^H$, and the constraint in (2.7), as follows:

$$\mathbf{c}_{sc}[n] = \frac{\mathbf{Q}[n]\mathbf{c}_{sc}[n-1]}{\sqrt{\left|\left(\mathbf{c}_{sc}[n-1]\right)^H \hat{\mathbf{R}}[n-1]\mathbf{c}_{sc}[n-1]\right|}}. \tag{2.26}$$

The smoothed response of the correlator's magnitude to this scaling is shown in Fig. 2.3.

Second, the convergence to zero of the unscaled correlator is often slow enough that, for packetized communication with a small enough number of bits, scaling of the correlator may not be required. For example, in Fig. 2.1, although the correlator norm decreases by a factor of nine over 10000 user data symbols, it is still greater than one,

which may be acceptable for moderate sized packets.

For the remainder of this dissertation, we focus on block rather than recursive implementations of the DMMSE criterion since these extend more readily to multipath channels.

2.3.4 Block Power Updates

To compute the correlator $\mathbf{c}$, the receiver replaces the statistical averages $\mathbf{R}$ and $\mathbf{A}$ in (2.9) with empirical averages computed over the training interval of length T:

$$\hat{\mathbf{R}} = \frac{1}{T}\sum_{n=1}^{T} \mathbf{r}[n](\mathbf{r}[n])^H, \tag{2.27a}$$

$$\begin{aligned}\hat{\mathbf{A}} &= \frac{1}{T}\sum_{n=1}^{T}\Big\{ b_1[n]b_1^*[n-1]\mathbf{r}[n-1]\big(\mathbf{r}[n]\big)^H + b_1^*[n]b_1[n-1]\mathbf{r}[n]\big(\mathbf{r}[n-1]\big)^H\Big\} \\ &= \frac{1}{T}\sum_{n=1}^{T} a_1[n]\mathbf{r}[n-1]\big(\mathbf{r}[n]\big)^H + a_1^*[n]\mathbf{r}[n]\big(\mathbf{r}[n-1]\big)^H.\end{aligned} \tag{2.27b}$$

The correlator $\mathbf{c}$ is thus computed without knowledge of the user's spreading sequence. The desired user's symbols are then decoded using the decision statistic

$$\hat{a}_1[n] = \mathrm{sign}\Big(\Re\Big\{\big\langle \mathbf{c}, \mathbf{r}[n]\big\rangle\big\langle \mathbf{c}, \mathbf{r}[n-1]\big\rangle^*\Big\}\Big). \tag{2.28}$$

As described in [113], the generalized eigenvalue problem for the pair $\big(\hat{\mathbf{A}}, \hat{\mathbf{R}}\big)$ may be solved by standard techniques such as the QZ method [36]. However, since only the dominant generalized eigenvector is required, an efficient algorithm would be to apply the power method [36] to the matrix $\mathbf{M} = \hat{\mathbf{R}}^{-1}\hat{\mathbf{A}}$ until convergence.

2.4 Eigenrake Reception for Frequency-Selective Fading

The preceding section described the fundamentals of the DMMSE criterion in the context of the flat fading model (2.1). We now discuss the extension of DMMSE concepts to frequency-selective fading channels and the blind detection of user transmissions and/or

multipath signals. The received vector for the nth symbol decision now takes the following form:

$$\mathbf{r}[n] = b_1[n]\mathbf{u}_1[n] + \sum_{k=2}^{K} b_k[n]\mathbf{u}_k[n] + \mathbf{w}[n], \tag{2.29}$$

where the time-varying, multipath signal vector for the kth user, $1 \leq k \leq K$, is given by

$$\mathbf{u}_k[n] = \sum_{l=1}^{L_k} F_{k,l}[n]\mathbf{v}_{k,l}.$$

Here, L_k is the number of resolvable fading paths for the kth user, $F_{k,l}[n]$ is the time-varying channel gain for the lth resolvable path of the kth user, and $\mathbf{v}_{k,l}$ is the effective spreading waveform for the lth path of the kth user.

For fast fading channels, uncorrelated multipath components from an interfering user appear as different "users" to a linear MMSE or DMMSE receiver, which therefore tries to separately suppress the interference corresponding to each multipath component. This *interferer multiplication* phenomenon leads to increased noise enhancement, as is well known [67], [114]. This penalty is unavoidable for any linear interference suppression scheme, unless the multipath components of the interfering users are tracked and combined prior to interference suppression. We do not focus on this issue here since our interest is in the effect of multipath fading for the desired user (and in the absence of knowledge or estimates of the desired user's channel at the receiver as in [57]). Thus, it is convenient to rewrite the model (2.29) as follows, hiding the structure of the interference due to other users and noise in a single vector, $\mathbf{i}[n]$:

$$\begin{aligned} \mathbf{r}[n] &= b_1[n]\mathbf{u}_1[n] + \mathbf{i}[n] \\ &= b_1[n]\sum_{l=1}^{L_1} F_{1,l}[n]\mathbf{v}_{1,l} + \mathbf{i}[n]. \end{aligned} \tag{2.30}$$

We have seen in Section 2.3 that the DMMSE-based algorithm avoids tracking the fading gain for a single path by the use of differential demodulation. However, if the desired user undergoes multipath fading, in order to automatically combine two paths, the DMMSE-based algorithms must track the time-varying linear combination $F_{1,1}[n]\mathbf{v}_{1,1} + F_{1,2}[n]\mathbf{v}_{1,2}$, which amounts to tracking the relative complex gain

$F_{1,2}[n]/F_{1,1}[n]$ with the single-path techniques of Section 2.3. This imposes a limit on the automatic multipath combining capability of the DMMSE-based algorithm as the fading rate increases [114]. It is necessary, therefore, to extend the basic DMMSE criterion to a multipath setting.

The idea is to convert the frequency-selective fading channel into several parallel frequency nonselective fading subchannels, apply the basic DMMSE algorithm for obtaining an interference-suppressing correlator for each subchannel, and to then noncoherently combine the correlator outputs for each subchannel to obtain the decision statistic. Further, the preceding should be accomplished without knowledge of the subchannels. Therefore, we describe next a DMMSE-based approach termed the *eigenrake* receiver, which achieves interference suppression and diversity combining without requiring explicit information regarding the multipath fading gains or timing.

2.4.1 The Eigenrake Receiver

Consider the generalized eigenvalue problem (2.8), where the matrices $\mathbf{R}$ and $\mathbf{A}$ are defined as in (2.9). Let $\{\mathbf{c}_l,\, 1 \le l \le L_e\}$, denote the eigenvectors corresponding to the positive eigenvalues. The eigenrake receiver employs a subset of these as correlators to obtain both interference suppression and diversity. Specifically, application of the lth correlator yields the decision statistic

$$Z_l[n] = \big\langle \mathbf{c}_l, \mathbf{r}[n] \big\rangle = \hat{\mathbf{F}}_{1,l}[n] b_1[n] + x_l[n], \tag{2.31}$$

where

$$\hat{\mathbf{F}}_{1,l}[n] = \big\langle \mathbf{c}_l, \mathbf{u}_1[n] \big\rangle \tag{2.32}$$

is the *effective* fading gain on the lth subchannel, and $x_l[n]$ is the residual interference plus noise at the output of the lth correlator. The outputs of these subchannels are then combined to generate the following decision statistic for differential demodulation:

$$Z[n] = \sum_{l=1}^{L_e} \alpha_l \big\langle \mathbf{c}_l, \mathbf{r}[n] \big\rangle \big\langle \mathbf{c}_l, \mathbf{r}[n-1] \big\rangle^*, \tag{2.33}$$

where the $\alpha_l \geq 0$ are combining gains. For differentially encoded data, this statistic is fed to a slicer. For example, for binary DPSK, the bit estimates are given by

$$\hat{a}[n] = \text{sign}\Big(\Re\big\{Z[n]\big\}\Big). \tag{2.34}$$

Equations (2.33)-(2.34) result in the *eigenrake* receiver, and we discuss its structure and properties in the next section.

2.4.2 Eigenrake Structure

For the signal model (2.30), assuming that the fading gains for different multipath components of the desired user are uncorrelated, we have

$$\mathbf{R} = \sum_{l=1}^{L_1} \gamma_l \mathbf{v}_{1,l}\mathbf{v}_{1,l}^H + \mathbf{R}_I, \tag{2.35}$$

$$\mathbf{A} = 2\sum_{l=1}^{L_1} \Re\{\rho_l\}\mathbf{v}_{1,l}\mathbf{v}_{1,l}^H, \tag{2.36}$$

where, for the lth multipath component of the desired user (with L_1 total multipath components), $\gamma_l = E\big[\big|F_{1,l}[n]\big|^2\big]$ denotes the average strength and $\rho_l = E\big[F_{1,l}[n]F_{1,l}^*[n-1]\big]$ denotes the correlation between the fading gains in successive symbol intervals. As was done for the flat fading environment Section 2.3.1, it is useful to develop the frequency-selective fading notation in the whitened domain for the derivation of the eigenrake receiver and its properties.

Whitened Domain: Using (2.13)-(2.16), as well as the whitened representation of $\mathbf{A}$ given by

$$\tilde{\mathbf{A}} = \mathbf{R}^{-\frac{1}{2}}\mathbf{A}\mathbf{R}^{-\frac{1}{2}}, \tag{2.37}$$

the corresponding eigenvalue problem (2.8) becomes

$$\tilde{\mathbf{A}}\tilde{\mathbf{c}} = \lambda\tilde{\mathbf{R}}\tilde{\mathbf{c}} = \lambda\tilde{\mathbf{c}}. \tag{2.38}$$

Note that there is a one-to-one correspondence between the eigenvectors of (2.8) and (2.38): $\mathbf{c}$ is an eigenvector of (2.8) if and only if $\tilde{\mathbf{c}} = \mathbf{R}^{\frac{1}{2}}\mathbf{c}$ is an eigenvector of (2.38)

with the same eigenvalue. For $1 \leq l \leq L_e$, let $\tilde{\mathbf{c}}_l = \mathbf{R}^{\frac{1}{2}}\mathbf{c}_l$ denote the eigenvectors corresponding to the positive eigenvalues for the whitened problem (2.38). Thus, the decision statistics in both domains are identical; that is,

$$\big\langle \mathbf{c}, \mathbf{r}[n] \big\rangle = \big\langle \tilde{\mathbf{c}}, \tilde{\mathbf{r}}[n] \big\rangle.$$

In the whitened domain, the received signal model (2.30) can be rewritten as

$$\begin{aligned} \tilde{\mathbf{r}}[n] &= b_1[n]\tilde{\mathbf{u}}_1[n] + \tilde{\mathbf{i}}[n] \\ &= b_1[n] \sum_{l=1}^{L_1} F_{1,l}[n]\tilde{\mathbf{v}}_{1,l} + \tilde{\mathbf{i}}[n], \end{aligned} \tag{2.39}$$

where $\tilde{\mathbf{v}}_{1,l} = \mathbf{R}^{-\frac{1}{2}}\mathbf{v}_{1,l}$, $1 \leq l \leq L_1$.

We are now ready to formally state the properties of the eigenrake receiver in the form of the following theorem.

Theorem 2.2 ***(Eigenrake Receiver)*****:** The eigenvectors $\{\mathbf{c}_l, 1 \leq l \leq L_e\}$, satisfy the following properties:

(a) The number of branches in the eigenrake receiver is at most equal to the number of multipath components for the desired user; that is, $L_e \leq L_1$.

(b) For each $1 \leq l \leq L_e$, the correlator $\mathbf{c}_l$ is an interference-suppressing, near–far resistant, linear receiver providing an estimate of the desired symbol sequence (up to complex scaling).

(c) As long as $\Re\{\rho_l\}$ are approximately equal for all l, the effective fading gains $\hat{\mathbf{F}}_{1,l}[n]$ are approximately uncorrelated for different l, $1 \leq l \leq L_e$. That is, the eigenrake receiver provides L_e-fold diversity.

Proof: Using (2.35) and (2.36), we obtain that

$$\mathbf{R}^{-1}\mathbf{A} = \sum_{l=1}^{L_1} \mathbf{d}_l \mathbf{v}_{1,l}^H,$$

where

$$\begin{aligned} \mathbf{d}_l &= 2\Re\{\rho_l\}\mathbf{R}^{-1}\mathbf{v}_{1,l} \\ &= 2\Re\{\rho_l\}\left(\mathbf{R}_I + \sum_{j=1}^{L_1} \gamma_j \mathbf{v}_{1,j}\mathbf{v}_{1,j}^H\right)^{-1} \mathbf{v}_{1,l}. \end{aligned} \tag{2.40}$$

The form of (2.40) is that of an MMSE correlator for which the desired signal is the lth multipath component of the desired user and the interference corresponds to the interference due to other users, as well as the other multipath components of the desired user. Thus, $\mathbf{d}_l$ inherits the classical interference suppression properties of the MMSE correlator (including its near–far resistance).

Now, consider an eigenvector $\mathbf{c}$ of $\mathbf{R}^{-1}\mathbf{A}$ corresponding to a nonzero eigenvalue λ. Such an eigenvector must satisfy

$$\begin{aligned}\mathbf{c} &= \frac{1}{\lambda}\,\mathbf{R}^{-1}\mathbf{A}\mathbf{c} \\ &= \frac{1}{\lambda}\sum_{l=1}^{L_1}\left(\mathbf{v}_{1,l}^H\mathbf{c}\right)\mathbf{d}_l,\end{aligned}$$

so that $\mathbf{c}$ is a linear combination of the interference-suppressing correlators $\{\mathbf{d}_l, 1 \le l \le L_1\}$. Hence, $\mathbf{c}$ also suppresses interference, is near–far resistant, and produces a scaled version of the desired symbol sequence. Further, since each eigenvector corresponding to a positive eigenvalue must be a linear combination of $\{\mathbf{d}_l, 1 \le l \le L_1\}$, the number of such vectors that can be linearly independent is at most L_1. Thus, the number of eigenvectors L_e is at most L_1. This completes the proof of properties (a) and (b).

To show diversity, it is easier to work in the whitened domain, where (2.36) reduces to

$$\tilde{\mathbf{A}} = 2\sum_{l=1}^{L_1}\Re\{\rho_l\}\tilde{\mathbf{v}}_{1,l}\tilde{\mathbf{v}}_{1,l}^H.$$

Comparing this with the covariance matrix for the whitened desired signal vector, given by

$$\begin{aligned}\tilde{\mathbf{R}}_d &= E\left[\tilde{\mathbf{u}}_1[n]\tilde{\mathbf{u}}_1^H[n]\right] \\ &= \sum_{l=1}^{L_1}E\left[\left|F_{1,l}[n]\right|^2\right]\tilde{\mathbf{v}}_{1,l}\tilde{\mathbf{v}}_{1,l}^H,\end{aligned}$$

we see that, as long as $\Re\{\rho_l\}$ is approximately the same for each l, $\tilde{\mathbf{A}}$ is approximately a scalar multiple of $\tilde{\mathbf{R}}_d$. Thus, the eigenvectors of $\tilde{\mathbf{A}}$ provide an approximate Karhunen-Loeve (KL) decomposition of the whitened desired signal vector $\tilde{\mathbf{u}}_1[n]$, which implies that

the effective fading gains along the directions of these eigenvectors are approximately uncorrelated (and hence approximately independent, if the fading coefficients are jointly complex Gaussian). This proves part (c), and completes the proof of the theorem. □

Remark 2.2: The condition that $\Re\{\rho_l\}$ be approximately the same for all l is satisfied in practice, since $\Re\{\rho_l\} \approx 1$ for typical fading rates (the fading gains for a given multipath component are roughly equal across successive symbols).

Remark 2.3: The eigenrake receiver achieves implicit timing acquisition, interference suppression and diversity, providing a KL decomposition of the faded signal vector in the whitened domain, without requiring explicit estimation of the location or strengths of the multipath components for the desired user.

Remark 2.4: In order to get the full performance benefit from the eigenrake receiver, it is important to employ correlators that coincide with the eigenvectors with positive eigenvalues. Simply choosing correlators lying in the subspace spanned by these eigenvectors does not work as well. Thus, an application of subspace tracking methods such as [87], [103], [104], [109] (or the orthogonal iteration of [36], [113]), will, at best, yield results comparable to the eigenrake receiver.

2.4.3 Blind Detection for the Selective Eigenrake

In this section, we describe a technique by which a DMMSE-based receiver may blindly detect the number of multipaths for the desired user. Further, since prior to transmission by the user, the receiver only has knowledge of the frame synchronization and the training symbol sequence, the technique applies more generally to the blind detection of any user transmission: in detecting the presence of received multipaths, the receiver inherently is detecting the presence of *any* received path.

To introduce the new *selective eigenrake* receiver, we must specify the combining gains, $\{\alpha_l\}$, in (2.33). If the number of multipaths of the desired user, L_1, were known, then by Theorem 2.2, the eigenrake receiver should use at most the L_1 eigenvectors corresponding to the L_1 largest eigenvalues of (2.8) for demodulation in (2.33). Among

these correlators may be some that should be deselected because their output is of poor quality (due to low signal strength or bad interference patterns for the corresponding multipath component). However, the eigenrake receiver has no prior knowledge of the number of multipath components or the quality of the corresponding correlators and employs instead a selection strategy based on the eigenvalues of (2.8) to choose which correlators to use. Once this set is selected, we have found by experimentation that equal gain combining is the most effective approach. Specifically, the combining rule we use is

$$\alpha_l = \begin{cases} 1 & \text{if } \lambda_l \geq t_0, \\ 0 & \text{if } \lambda_l < t_0. \end{cases} \tag{2.41}$$

To choose the threshold t_0, consider the DMMSE constrained cost function $J(\mathbf{c})$ in (2.6). If a correlator $\mathbf{c}_l$ is working well, then we must have $J(\mathbf{c}_l) \approx 0$. But, from (2.10), $J(\mathbf{c}_l) = 2\mathbf{c}_l^H \mathbf{R}\mathbf{c}_l - \mathbf{c}_l^H \mathbf{A}\mathbf{c}_l = 2 - \lambda_l$, which means that $\lambda_l \approx 2$ for a correlator that is producing a good reproduction of the desired symbol sequence. From our numerical results, we find that $t_0 = 1$, which balances the tradeoff between false indication (i.e., incorrectly indicating a specific path is present) and failed detection (i.e., not detecting the presence of a particular path), works well. We leverage this capability next in Chapter 3 to blindly detect the presence or absence of any users using specific sets of training sequences in a random access contention channel.

Alternatively, it is possible to optimize the combining gains, α_l, based on an estimate of the SIR on each branch. Denote the desired signal power at the output of the lth correlator by P_l, and interference plus noise power by σ_l^2. At the output of correlator $\mathbf{c}_l$, the average power of the net received signal is $\mathbf{c}_l^H \mathbf{R}\mathbf{c}_l = 1$ (by virtue of the normalization we have imposed). The average power of the desired signal is $P_l = E\big[|\mathbf{c}_l^H \mathbf{u}_1[n]|^2\big] \approx \frac{1}{2}\mathbf{c}_l^H \mathbf{A}\mathbf{c}_l = \frac{\lambda_l}{2}$, assuming that $\Re\{\rho_l\} \approx 1$. Hence, the averaged interference plus noise power is $\sigma_l^2 = 1 - \frac{\lambda_l}{2}$. It can be shown that the maximal ratio combining coefficients α_l in (2.33) should be set as $\alpha_l^2 = \frac{P_l}{\sigma_l^4}$ (i.e., the same as classically defined in [12] and recently reissued in [13]), approximating the outputs of the fingers of the eigenrake

as independent, differentially coherent Rayleigh fading channels with Gaussian noise. However, based on our simulations, such optimization does not improve upon the simpler equal gain, selective combining strategy of (2.41).

2.5 Simulations

To demonstrate the receivers described in this thesis and compare them with other standard approaches, we simulate several MMSE and DMMSE receiver implementations as well as a standard Rake receiver. For DMMSE reception, results are obtained using the following methods: block power update of Section 2.3.4; eigenrake of Section 2.4.1 with knowledge of the number of desired user multipaths present; selective eigenrake of Section 2.4.3 without multipath knowledge; scaled RLS as described in Section 2.3.2; and the unscaled RLS from Section 2.3.3. Both unscaled and scaled RLS DMMSE receivers are simulated using differential decoding where the decoded symbol estimates are given by

$$\hat{a}_1[n] = \text{sign}\Big(\Re\Big\{\big\langle \mathbf{c}[n], \mathbf{r}[n]\big\rangle \big\langle \mathbf{c}[n-1], \mathbf{r}[n-1]\big\rangle^*\Big\}\Big). \tag{2.42}$$

Additionally, to demonstrate the standard MMSE receiver in (2.2), a standard RLS MMSE algorithm [44] is simulated using direct decoding, where the recursive correlator updates are computed via

$$\begin{aligned}
\mathbf{k}[n] &= \frac{\mathbf{P}[n-1]\mathbf{r}[n]}{\beta + \mathbf{r}^H[n]\mathbf{P}[n-1]\mathbf{r}[n]}, \\
\mathbf{P}[n] &= \beta^{-1}\mathbf{P}[n-1] - \beta^{-1}\mathbf{k}[n]\mathbf{r}^H[n]\mathbf{P}[n-1], \\
\xi[n] &= b_1[n] - \mathbf{c}^H[n-1]\mathbf{r}[n], \\
\mathbf{c}[n] &= \mathbf{c}[n-1] + \beta_a \mathbf{k}[n]\xi^*[n],
\end{aligned}$$

and the decoded symbol estimates are computed as

$$\hat{b}_1[n] = \text{sign}\Big(\Re\Big\{\big\langle \mathbf{c}[n-1], \mathbf{r}[n]\big\rangle\Big\}\Big). \tag{2.43}$$

For benchmarks, included are results obtained using instantaneous, ideal versions of the MMSE solution and standard Rake. Both of these techniques assume perfect knowledge

of all the users spreading waveforms and fading gains and are computed at every symbol time. The instantaneous, ideal MMSE solution is computed as:

$$\mathbf{c}[n] = \mathbf{R}_I^{-1}[n]\left(\sum_{l=1}^{L_1} F_{1,l}[n]\mathbf{v}_{1,l}\right), \tag{2.44}$$

where $\mathbf{R}_I[n]$ is computed as in (2.17), appropriately modified for multipath and perfect channel knowledge:

$$\mathbf{R}_I[n] = \sum_{k=2}^{K}\sum_{l=1}^{L_k}\left|F_{k,l}[n]\right|^2\mathbf{v}_{k,l}\mathbf{v}_{k,l}^H + 2\sigma^2\mathbf{I}.$$

The instantaneous, ideal Rake solution is computed as:

$$\mathbf{c}[n] = \sum_{l=1}^{L_1} F_{1,l}[n]\mathbf{v}_{1,l}. \tag{2.45}$$

In both cases, the symbols are differentially decoded via

$$\hat{a}_1[n] = \operatorname{sign}\left(\Re\left\{\left\langle\mathbf{c}[n],\mathbf{r}[n]\right\rangle\left\langle\mathbf{c}[n],\mathbf{r}[n-1]\right\rangle^*\right\}\right). \tag{2.46}$$

For all results reported here, we consider a synchronous CDMA system, given by (2.1) for flat fading, and (2.35) for multipath fading. The processing gain $N = 10$, with $K = 1$ and 4 users, each with a fixed but randomly chosen spreading sequence. The Rayleigh fading coefficients are generated independently for different paths and users using a modified Jakes' simulator [51], [75], [108], with a normalized Doppler spread of 0.01 (see the discussion in Section 2.2 of parameters that might correspond to such a system). When there are multiple users present, each interfering user has average power 20 dB higher than that of the desired user. We thereby demonstrate that DMMSE interference suppression is robust under severe near–far conditions, which is to be expected given the theoretical results on near–far resistance proven in Section 2.4.2.

First, we consider frequency-nonselective fading for each user. Since the signal-to-noise ratio (SNR) for the desired transmission varies with time due to fading, the ability of an adaptive algorithm to track the channel and to suppress interference is gauged by the difference between the signal-to-interference ratio (SIR) [113], [114], and SNR, rather than the raw value of SIR. Thus, with the AWGN power set so that the desired

user's average $E_b/N_0 = 20\,\text{dB}$, we plot SIR/SNR (dB) averaged over 250 adaptations in Fig. 2.4 for four schemes: ideal MMSE, standard RLS MMSE with direct decoding, block power DMMSE, and scaled RLS DMMSE. The algorithms have a training length of $T = 60$ symbols, after which the block power receiver is fixed and both the standard and scaled RLS receivers switch to decision-directed mode. Both RLS receivers were simulated with $\beta = \beta_a = 0.99$. Clearly, the standard RLS MMSE algorithm with direct decoding cannot track the fading at all, and actually begins to diverge immediately after the training period. The block power DMMSE algorithm is able to track the desired users signal, but at a level $1.5 - 2\,\text{dB}$ worse than the scaled RLS DMMSE algorithm. In Figs. 2.5 and 2.6, we plot the average BER for $K = 1$ and 4, respectively, over 2000 simulations at each E_b/N_0 for six schemes: ideal MMSE and Rake, block power DMMSE, standard RLS MMSE, unscaled and scaled RLS DMMSE, and eigenrake DMMSE (using only one correlator). For the simulations, the training length was $T = 100$ symbols per packet and the packet length was 2000 symbols. In Fig. 2.5 with $K = 1$, all of the receivers perform similarly well except for the standard RLS MMSE with direct decoding. The ideal Rake and MMSE receivers match exactly since there are no interferers. Further, the block power and eigenrake receivers match exactly since they are equivalent when there is no multipath and are only slightly worse than the ideal MMSE receiver. Both the unscaled and scaled RLS DMMSE receivers perform just in the gap (and the same, as discussed in Section 2.3.3) between the eigenrake and ideal MMSE, while the standard RLS MMSE fails completely to track the channel fading gains. In the presence of other users ($K = 4$) in Fig. 2.6, the results are similar to those for the single-path, no-interferer case with one exception: the ideal Rake fails completely due to MAI. The MAI causes an approximately 3 dB decrease in performance for the group.

Next, we consider a situation in which the desired user has two independently faded multipath components while each interferer still sees a single-path fading channel. Average BER is plotted in Figs. 2.7 and 2.8 for the case of $L_1 = 2$ multipath components for the desired user, and 2000 symbol packets with training length $T = 100$ symbols. In

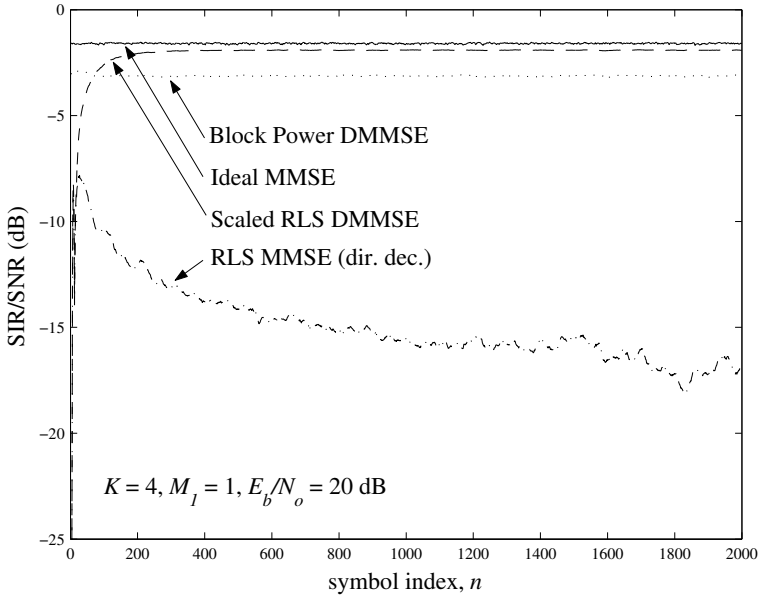

Figure 2.4. $\frac{SIR}{SNR}$ averaged over 250 adaptations for $K = 4$ and a flat fading channel using ideal MMSE, standard RLS MMSE (with direct decoding), block power DMMSE, and scaled RLS DMMSE. The training period is $T = 60$ symbols, and the RLS adaptations switch to decision-directed mode after training.

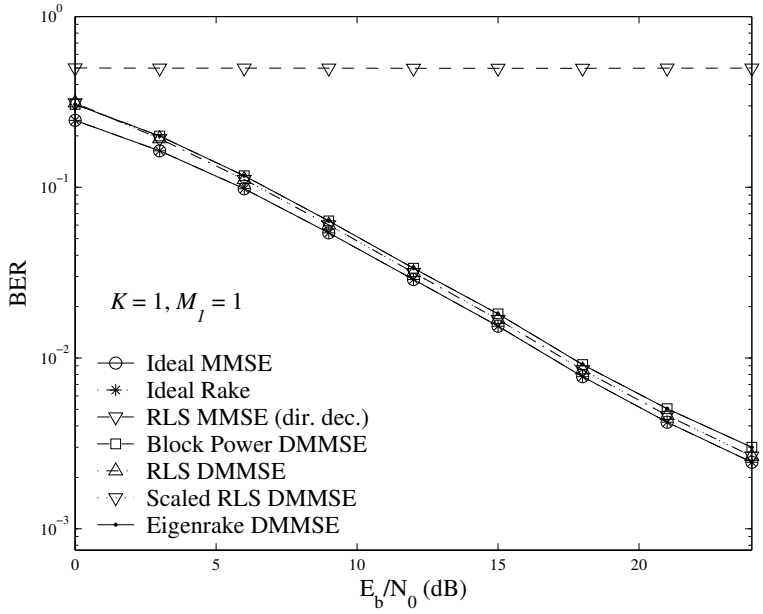

Figure 2.5. BER for $K = 1$ user in a flat fading channel using ideal MMSE and Rake, standard RLS MMSE (with direct decoding), block power DMMSE, unscaled and scaled RLS DMMSE, and eigenrake DMMSE receivers. The training period is $T = 100$ symbols, and the RLS adaptations switch to decision-directed mode after training.

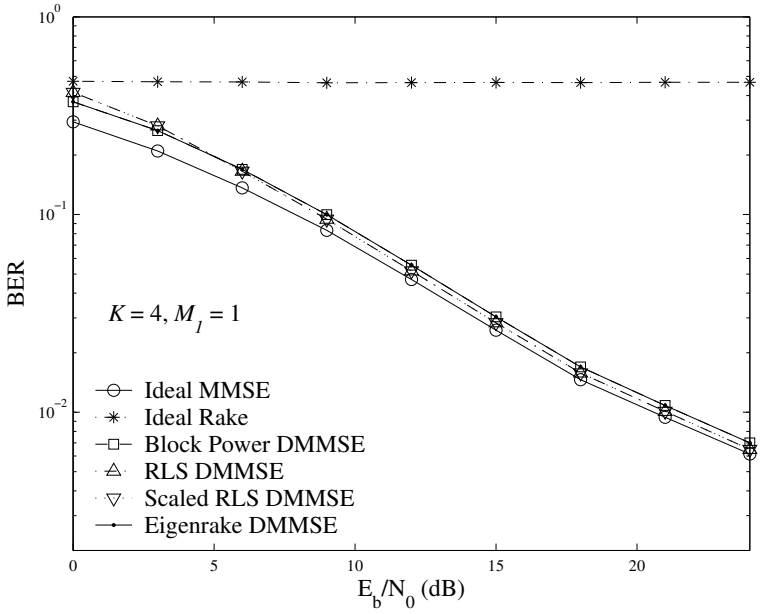

Figure 2.6. BER for $K = 4$ users in a flat fading channel using ideal MMSE and Rake, block power DMMSE, unscaled and scaled RLS DMMSE, and eigenrake DMMSE receivers. The training period is $T = 100$ symbols, and the RLS adaptations switch to decision-directed mode after training.

addition to the receivers in the previous plots, the selective eigenrake receiver is tested. For $K = 1$, Fig. 2.7 shows the diversity gains obtained by the eigenrake and selective eigenrake algorithms, while the single-correlator implementations of the other techniques grow progressively worse in performance as the SNR improves. This is expected since the single-correlator techniques actually experience the extra paths of the desired user as interference. Here again, the ideal Rake and MMSE receivers are equivalent, but only approximately 3 dB better than the eigenrake receivers operating with knowledge only of the desired user's training sequence. In Fig. 2.8 for $K = 4$, we see that both the standard and selective eigenrake receivers are robust to strong MAI (the ideal Rake receiver is omitted, due to its poor performance). At higher E_b/N_0, when the effect of MAI is dominant, adaptive selection of eigenrake correlators is $0.5 - 1$ dB better than fixing the number of correlators to equal the number of paths. We attribute this to the ability of the selective eigenrake to adaptively "deselect" paths suffering too great a level of interference.

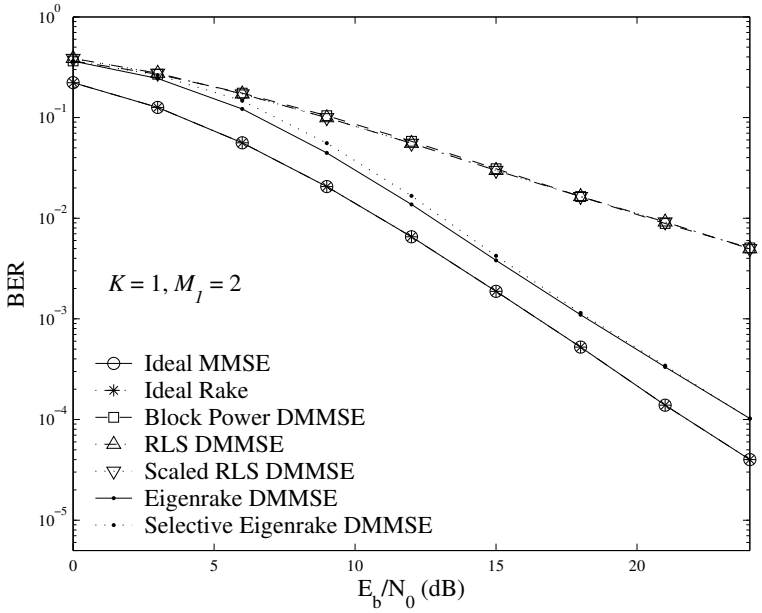

Figure 2.7. BER for $K = 1$ user with $L_1 = 2$ multipath components using ideal MMSE and RAKE, block power DMMSE, unscaled and scaled RLS DMMSE, and eigenrake and selective eigenrake DMMSE receivers. The training period is $T = 100$ symbols, and the RLS adaptations switch to decision-directed mode after training.

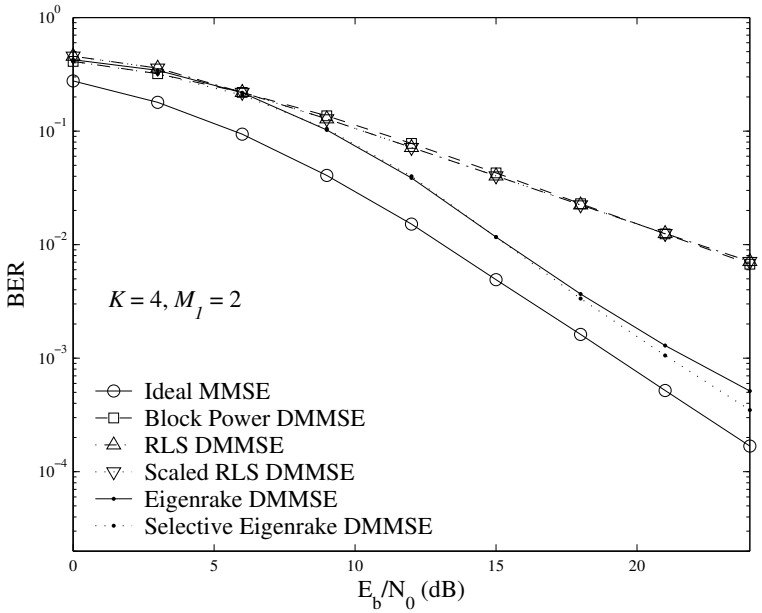

Figure 2.8. BER for $K = 4$ users with $L_1 = 2$ multipath components for the desired user using ideal MMSE, block power DMMSE, unscaled and scaled RLS DMMSE, and eigenrake and selective eigenrake DMMSE receivers. The training period is $T = 100$ symbols, and the RLS adaptations switch to decision-directed mode after training.

Finally, Figs. 2.9 and 2.10 show the BER for the eigenrake receiver (with equal gain combining) as the number of correlators and the number of multipath components for the desired user are varied, without recourse to an adaptive selection mechanism. There are three strong interferers ($K = 4$), each with a single path. In Fig. 2.9, L_1 is varied from 1 to 4, with the number of correlators always set as $N_{Ci} = L_1$. The performance of the ideal MMSE receiver improves monotonically with L_1, while the performance of the eigenrake improves until $L_1 = 3$, but degrades for $L_1 = 4$. This is probably due to the fact that as the number of paths increases with $N = 10$ fixed, the likelihood of one of the eigenrake correlators seeing a bad crosscorrelation pattern also increases. On the other hand, the ideal MMSE receiver sees a single effective spreading waveform, regardless of the number of paths. In Fig. 2.10, we fix $L_1 = 3$, and vary N_{Ci} from 1 to 4. The performance improves as N_{Ci} increases, until $N_{Ci} = L_1 = 3$. However, when $N_{Ci} = 4 > L_1$, the additional correlator cannot be linearly independent by Theorem 2.2, and the performance degrades significantly compared to that for $N_{Ci} = 2, 3$, while still showing diversity gains relative to a single correlator. We conclude, therefore, that adaptive selection of the number of correlators, as done by the selective eigenrake, is crucial to balancing the effects of diversity gains and residual interference, and that mismatch between the number of correlators and paths severely degrades performance.

2.6 Discussion

We have shown that the DMMSE criterion leads to adaptive linear receivers which are robust to rapid channel time variations, unlike adaptive algorithms based on the conventional MMSE criterion. The near–far resistance and interference suppression properties of the resulting DMMSE solution are shown by establishing equivalence with a channel-compensated MMSE solution. For frequency-selective fading, the DMMSE criterion is extended to obtain the *eigenrake* and *selective eigenrake* receivers, which provide implicit timing acquisition, diversity, and interference suppression.

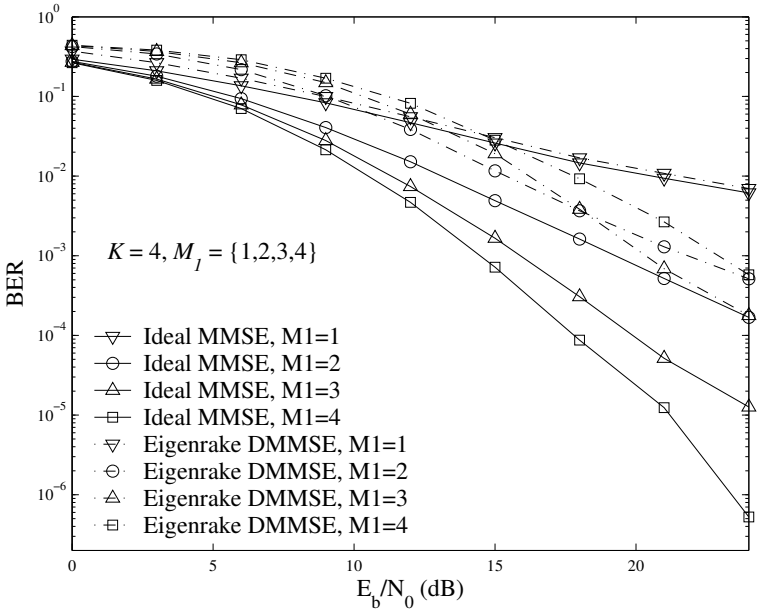

Figure 2.9. BER for $K = 4$ users with $L_1 = \{1, 2, 3, 4\}$ multipath components for the desired user using the ideal MMSE and the eigenrake DMMSE receivers with the number of eigenrake correlators fixed as the number of multipaths for the desired user.

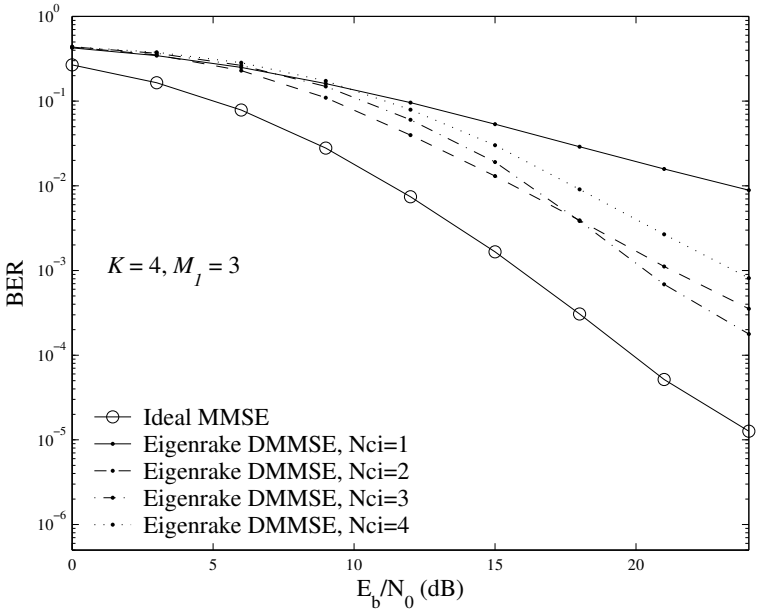

Figure 2.10. BER for $K = 4$ users with $L_1 = 3$ multipath components for the desired user using the eigenrake DMMSE receiver with $N_{Ci} = \{1, 2, 3, 4\}$ correlators for detection, with the ideal MMSE solution included for comparison.

In addition to its robustness to fading, DMMSE reception relaxes traditional needs for strict power controls and is also robust to lack of carrier synchronization. Thus, the DMMSE Eigenrake receiver is an attractive technique for packetized transmission in rapidly varying network topologies or networks with multiple uncoordinated users simultaneously contending in a wireless contention channel. In Chapter 3, the information on the number of eigenvalues of (2.8) resulting from DMMSE training is used to blindly detect these multiple users and decode their resource requests. Further, in Chapter 4, the information on the number of received transmissions is employed to generate backlog estimates for computing stabilizing and QoS-providing contention policies.

Beyond the initial exposition of the DMMSE criterion in this thesis, much further work remains on detailed receiver design, comparison with other approaches, and development of efficient numerical techniques. For example, it is necessary to resolve practical issues such as whether to use noncoherent techniques in conjunction with DMMSE, possibly with multiple symbol detection [24], [107], or whether to use DMMSE for robust interference suppression, followed by separate channel gain recovery for coherent detection. Another possible direction for future work is integration of DMMSE with sophisticated coding techniques, such as those used for turbo multiuser detection [72], [105], and single-user noncoherent communication based on joint channel and data estimation [16], [91].

Chapter 3

Cross-Layer Design

Since the introduction of the ALOHA system by Abramson in 1970 [6], random multiple-access in wireless communication systems has been limited to the single-packet reception (SPR) model. In this dissertation, we present a new design, capable of multipacket reception (MPR), for random, contention-based wireless medium access. The design uses a time-slotted contention channel, analogous to slotted-ALOHA [85]. The receiver at the access point (AP) utilizes Differential MMSE-based (DMMSE) reception, described in Chapter 2 and [61], which is capable of detecting *and* decoding an unknown number of transmitted packets simultaneously, *without prior coordination between the AP and the users*. To contend, users randomly choose a training symbol sequence from a known set of sequences and prefix their packet with the training symbols. The resulting contention packet is transmitted after all of the symbols are spread using a fixed, but randomly chosen, spreading sequence. By leveraging the special feedback capabilities of the DMMSE receiver described in Chapter 2, the AP can determine the number of transmissions received and decode those that have not collided by having chosen the same training sequences. The multiple access interference (MAI) suppression properties of DMMSE-based reception result in a random access contention channel with MPR, and without constraints such as capture requirements, centralized power control, or strict carrier synchronization. Before describing the details of the design, we first discuss existing work

in order to motivate the current approach.

3.1 Motivation

In the wireless communications sector today, the low cost and ease of deployment of 802.11 [3] based wireless local area networks (WLANs) has lead to their widespread adoption. The attainable data rates of such networks are much higher than those in second and third generation cellular networks [1], [4]. While WLANs are currently restricted to low-mobility data applications, the support of cellular functionalities such as high mobility and real-time traffic using a WLAN-type infrastructure is an attractive alternative for applications such as coverage of academic or industrial campuses, or for emergency and plug-and-play deployments. The object would be to support a range of user mobilities and traffic using, instead of an hierarchical structure as in [29], [48], a flat, cost-effective architecture based on small cells and unstructured geometries of WLANs. Such a *pseudocellular* architecture is appropriate for plug-and-play networks and permits lower power and complexity mobile units, while the more capable AP nodes may be connected via a wireless ad hoc network or through a wireline network.

In this chapter, we present a cross-layer design for the contention channel that solves the critical technical problem of enabling the higher rate of handoffs between the smaller pseudocells for highly mobile users with delay-constrained traffic. In our design, a mobile detecting a stronger beacon from an adjacent access point simply asks for a reservation. Joint optimization of the physical (PHY) layer and the MAC, with a novel use of adaptive multiuser detection, achieves rapid, mobile-centric, handoffs using in-band, contention-based reservations, and can meet the tight handoff delay budgets required for sustaining a real-time application at vehicular speeds. Once a reservation is successful, we assume that the mobile has contention-free access (e.g., using polling or TDM/TDMA) until its resource requirements change or it leaves the pseudocell.

Previous studies [84], [96] of the standard, SPR slotted-ALOHA contention channel

[6], [85] have shown that the backlog of users who have arrived in preceding time slots, but have not yet successfully contended, can be stabilized under certain conditions on the traffic arrival rate. Further, even for a large backlog, the maximum throughput can be maintained through the use of a controlled access scheme [41], [84]. In [33], [34], these concepts were extended to the multipacket reception (MPR) case, where it was shown that a dynamic control policy and constraints on the arrival process are required to stabilize the system. We design such a contention policy for our MPR system, which maintains the maximum throughput under heavy traffic loads and stabilizes the backlog when conditions on the traffic arrival rates are satisfied. A technique for estimating the backlog size, needed in the stabilization procedures, is developed through an extension of pseudo-Bayesian backlog estimation [84].

The cross-layer design is described in Section 3.2. Section 3.3 provides analytical insight into the performance of the design, which is extended to derive throughput and stability properties in Section 3.4. In Section 3.4, a new technique for backlog estimation is given. Simulation results are provided in Section 3.6. Section 3.7 contains concluding remarks.

3.2 System Design

Mobile-controlled handoff is well-known to provide short reaction times [94], but the problem of multiuser interference due to uncoordinated transmissions by different mobiles has not been satisfactorily addressed. This issue is solved here by a joint PHY/MAC design in which the AP can successfully decode multiple simultaneous users on the contention channel.

The contention channel utilizes *contention frames* interspersed among contention-free data slots. As shown in Fig. 3.1, each contention frame may contain multiple *contention slots*, the number of which may be dynamically allocated by the AP based on traffic estimates (e.g., as in [111] and [41]). To contend in a contention frame, a user randomly

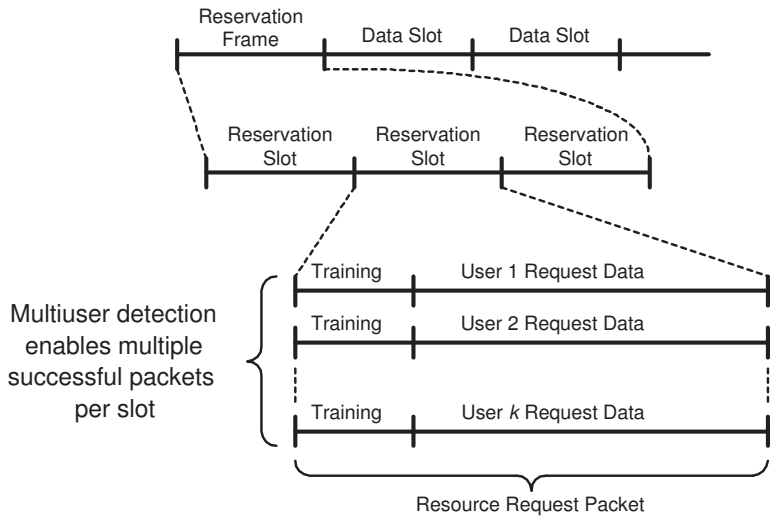

Figure 3.1. The structure of a contention frame.

chooses a contention slot in which to transmit. Each user sends a sequence of training symbols followed by data symbols containing the reservation request. These symbols are spread using direct sequence spread spectrum (DSSS), using a "short" spreading sequence whose period equals the symbol interval. As described in Section 3.2.1, the training sequence choice by each user defines a *virtual subslot*.

3.2.1 Transmission

A user's resource reservation request is transmitted as a reservation packet during the randomly chosen contention slot. The reservation request would typically be quite short, containing limited information such as the amount, duration or priority of the resources required. While our PHY/MAC design differs significantly from existing 802.11 systems, some chosen design parameters are similar to those in the 802.11b standard [3]. Thus, our AP receiver requires 100-120 training symbols for good performance, comparable to the 128 SYNC bits in 802.11b. Also as in 802.11b, differential encoding is applied to the training and payload symbols. They are then spread with a short spreading sequence, as in the 1 or 2 Mbps modes in 802.11b. However, our design differs from 802.11b in several crucial aspects, as described in the following.

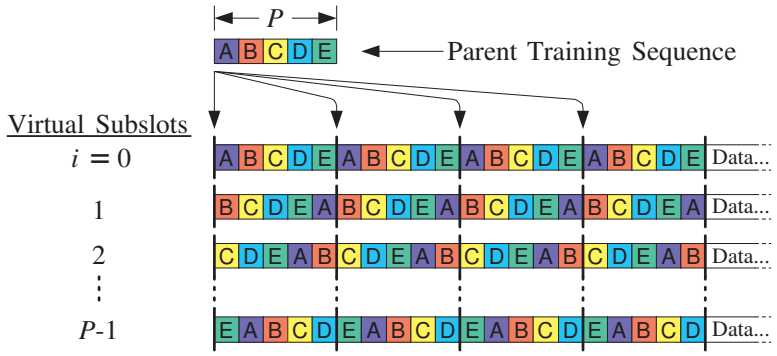

Figure 3.2. Training sequence construction using a single "parent" sequence with good autocorrelation properties.

First, unlike the fixed training sequence used in 802.11b, each user chooses a random, cyclic, shift of a "parent" sequence of length P with good autocorrelation properties (e.g., Gold sequences [35], [78]). The training sequences are constructed by concatenating several periods of the parent sequence. Thus, if $t_P[n], n = 0, 1, ..., P-1$, is the parent sequence, a user selecting a random "phase" $i \in \{0, 1, \ldots, P-1\}$ employs the training sequence

$$t_i[n] = t_P\big[(n-i) \bmod P\big], \quad n = 0, 1, \ldots, T-1, \tag{3.1}$$

where T is the total length of the training period (typically an integer multiple of P). Thus, i indexes the virtual subslot chosen by the user out of the P possible choices and the $\{t_i[n]\}$ have low correlation across i. The resulting virtual subslots are shown in Fig. 3.2.

Second, each user employs a randomly chosen "short" spreading sequence of length N, rather than all users employing the same Barker code as in 802.11b. That the spreading sequence is unknown to the AP is inconsequential because the AP employs an adaptive multiuser detection technique that requires only a sequence of training symbols.

Since there are P virtual subslots (corresponding to the P possible phases of the training sequence), two users who choose the same contention slot are unlikely to choose the same virtual subslot. The random spreading sequences further reduce correlation between user transmissions in a contention slot. Under these conditions, the multiuser

detection algorithm in Section 3.2.2 enables the processing of a large number of simultaneous reservation requests without requiring prior coordination with the AP. In contrast, support for multiple simultaneous users in 802.11b is poor because all users employ the same training and spreading sequence.

3.2.2 Multipacket Reception

The AP receiver knows only the common parent training sequence, $t_P[n]$, used by all the users. *Unknown* to the receiver are:

a) the number of contending users,

b) the contention slots in which users are contending,

c) the random phase shifts, i, of the parent training sequence, $t_P[n]$, and

d) the randomly chosen spreading sequences of the contending users.

Further, the receiver employs a fixed local oscillator which is not synchronized to the carrier of any specific user. For typical Doppler shifts and oscillator tolerance, however, the carrier phase is well approximated as constant over several symbol intervals. Since the design is intended to serve mobile users, it must also be capable of overcoming the effects of Rayleigh fading without the availability of channel information.

Based on the preceding constraints, a natural strategy is to employ adaptive multiuser detection based on the DMMSE criterion of Chapter 2, which is robust to both fading and carrier phase uncertainty, and differential decoding of the data based on the decision statistics at the output of the DMMSE correlator. The results in Chapter 2 on the selective eigenrake receiver imply that the presence of a user can be detected by checking whether the largest eigenvalue of the eigenproblem $\mathbf{Ac} = \lambda\mathbf{Rc}$ in (2.8) exceeds a threshold. We exploit this to obtain the following receiver algorithm.

To detect the presence of users in all the virtual subslots, the receiver computes the eigenvalues of (2.8) for each possible phase i in (3.1) of the parent training sequence.

Thus, the receiver trains P DMMSE correlators in parallel and, at the end of training, checks the eigenvalues of (2.8). For each training phase i, there are three possibilities:

Case 1: Only one eigenvalue greater than the threshold;

Case 2: Two or more eigenvalues greater than the threshold;

Case 3: No eigenvalues greater than the threshold.

In Case 1, the receiver decides that a user is present and continues to demodulate the remainder of the reservation packet symbols. In Case 2, the receiver assumes that a collision has occurred and discontinues demodulation for that virtual subslot[1] since, if more than one user chooses the same virtual subslot, they cannot be reliably decoded. In Case 3, the receiver decides that the virtual subslot is unoccupied and discontinues demodulation. If a user's transmission falls into Case 2 (collision) or Case 3 (not detected due to poor channel gain), then the user can re-contend in subsequent contention frames until its delay budget, if it has one, expires. The feedback for such a user could be implicit (e.g., not being polled by the AP) or explicit (e.g., the AP broadcasting the outcome of each virtual subslot).

It is important to note that the complexity of training P correlators can be significantly reduced by exploiting commonalities in the required computations. For example, the matrix $\hat{\mathbf{R}}$ given in (2.27a) is the same for all correlators. Note also that the same terms $\mathbf{r}[n]\big(\mathbf{r}[n-1]\big)^H$ and $\mathbf{r}[n-1]\big(\mathbf{r}[n]\big)^H$ (which are themselves transposed conjugates) are required for the computation of $\hat{\mathbf{A}}$ in (2.27b) for all the correlators, while the terms $b[n-1]b^*[n]$ and $b^*[n-1]b[n]$ may both be reduced to $a[n] = t_i[n]$ for a given i.

3.3 Collision Analysis

In order to characterize key system design and performance parameters in the following sections, we need a description of the collision mechanisms of the contention channel

[1]While we consider frequency-nonselective fading for simplicity, for multipath channels, several large eigenvalues may result from a single user. In such settings, the receiver would continue demodulation in Case 2 and decide at the end of the demodulation whether or not a collision has occurred by use of an error detection and/or correction code.

capable of having both multiple successes and failures simultaneously. We do this by first deriving the probability of a user's transmission failing to be successfully decoded at the AP for a generalized PHY. We then introduce a simple, idealized model of the PHY in Section 3.2.2, based on the DMMSE receiver of Chapter 2, to obtain an analytical expression for the system's probability of failure. Results using this model are then applied in Section 3.4 to discuss the throughput, stability, and backlog estimation capabilities of the design. The model is further applied in Chapter 4 for provisioning QoS for delay-constrained traffic.

3.3.1 Collision Model

Our goal is to derive the probability of failed contention for a tagged user, user i, for a general PHY model and then apply the analysis to a simplified model of the PHY design in Section 3.2.2. The generalization of the following analysis results from the interpretation of the number of virtual subslots as simply a limit to the number of unique ways users may transmit a contention packet to the AP, and these may be provided by any careful PHY design, not just that of Section 3.2. To begin, we define the following

terms:

$$
\begin{aligned}
M &\triangleq \text{number of contention slots per contention frame,} \\
U &\triangleq \text{number of additional users in user } i\text{'s contention slot,} \\
\varepsilon_{cs} &\triangleq \text{failure due to contention slot interference,} \\
&= \begin{cases} 1 & \text{if failure due to contention slot interference,} \\ 0 & \text{otherwise,} \end{cases} \\
\varepsilon_{vs} &\triangleq \text{failure due to virtual subslot collision,} \\
&= \begin{cases} 1 & \text{if failure due to virtual subslot collision,} \\ 0 & \text{otherwise,} \end{cases} \\
\varepsilon &\triangleq \text{failure due to interference,} \\
&= \varepsilon_{cs} \cup \varepsilon_{vs}.
\end{aligned}
$$

Given that there are n users contending in the contention frame, let ρ_n denote the probability of success conditioned on n for the tagged user. Then, conditioning further on U, the probability of failure for the tagged user is given by

$$
\begin{aligned}
\phi_n &= 1 - \rho_n \\
&= P(\varepsilon = 1 | n \text{ users contending}) \\
&= \sum_{k=1}^{n-1} P(\varepsilon = 1 | U = k) P(U = k | n).
\end{aligned}
\tag{3.2}
$$

When $M > 1$, the probability that the number of additional users in user i's contention slot is $U = k$ is the standard binomial distribution. When $M = 1$, the probability that $U = k$ is 1 for $k = n - 1$ and 0 for all other k:

$$
P(U = k | n) = \binom{n-1}{k} \left(\frac{1}{M}\right)^k \left(1 - \frac{1}{M}\right)^{n-1-k} \mathrm{I}_{\{M>1\}} + \mathrm{I}_{\{M=1\}} \mathrm{I}_{\{k=n-1\}}, \tag{3.3}
$$

where $\mathrm{I}_{\{.\}}$ is the indicator function.

We may write the first term in the summation of (3.2) as the union of "contention slot interference" and "virtual subslot interference" failure:

$$\begin{aligned} P(\varepsilon = 1|U = k) &= P(\varepsilon_{cs} = 1 \cup \varepsilon_{vs} = 1|U = k) \\ &= 1 - P(\varepsilon_{vs} = 0|U = k)P(\varepsilon_{cs} = 0|U = k). \end{aligned} \tag{3.4}$$

The last equality in (3.4) results from the observation that ε_{vs} and ε_{cs} are conditionally independent given $U = k$.

Failure due to virtual subslot interference occurs if at least one other user chooses the same virtual subslot as the tagged user. Thus, the conditional probability of no collision in the tagged user's virtual subslot is given by

$$P(\varepsilon_{vs} = 0|U = k) = \left(1 - \frac{1}{P}\right)^k, \tag{3.5}$$

and we may write (3.4) as

$$P(\varepsilon = 1|U = k) = 1 - \left(1 - \frac{1}{P}\right)^k P(\varepsilon_{cs} = 0|U = k). \tag{3.6}$$

The conditional probability of failure due to contention slot interference, $P(\varepsilon_{cs} = 1|U = k)$, is solely determined by the receiver's ability to overcome "non-virtual-subslot interference." In other words, $P(\varepsilon_{cs} = 1|U = k)$ is a function of the ability of the PHY to suppress noise and multiple access interference (MAI) from other users in different virtual subslots. Therefore, for a specific PHY, $P(\varepsilon_{cs} = 0|U = k) = 1 - P(\varepsilon_{cs} = 1|U = k)$ must be obtained from detailed bit error rate (BER) analysis, simulations, or some suitable performance measure that incorporates the applicable definition of packet failure (e.g., a BER of 3%).

For the analysis of the PHY of Section 3.2.2, we assume a model in which failure due to contention slot interference occurs if and only if the number of interfering users in the tagged user's contention slot exceeds $N - 1$. If the number of interfering users is $N - 1$ or smaller, failure may still occur, but only due to virtual subslot interference. Under this simplifying assumption, we have

$$P(\varepsilon_{cs} = 0|U = k) = \mathrm{I}_{\{k<N\}}. \tag{3.7}$$

Inserting (3.7) into (3.6), we obtain the result

$$P(\varepsilon = 1|U = k) \;=\; 1 - \Big(1 - \frac{1}{P}\Big)^{k} \mathrm{I}_{\{k<N\}}. \tag{3.8}$$

Using (3.3) and (3.8), the expression for the probability of the ith user failing in contention in (3.2) can now be written as

$$\phi_n \;=\; \sum_{k=1}^{n-1} \Bigg\{ \Big(1 - \big(1 - \tfrac{1}{P}\big)^{k} \mathrm{I}_{\{k<N\}}\Big) \times \Big(\tbinom{n-1}{k} \big(\tfrac{1}{M}\big)^{k} \big(1 - \tfrac{1}{M}\big)^{n-1-k} \mathrm{I}_{\{M>1\}} + \mathrm{I}_{\{M=1\}} \mathrm{I}_{\{k=n-1\}}\Big) \Bigg\}. \tag{3.9}$$

Thus, we have computed the probability of failure, ϕ_n, when n users transmit for a generalized PHY via the combination of (3.2) and (3.6). For the PHY described in Section 3.2.2, we have applied a simplified model to obtain (3.9). These results are used next to analyze the stability of the contention channel and are compared with simulations in Section 3.6.

3.4 Performance Framework

In this section, we characterize system performance. Using the results of Section 3.3, we are able to analyze the throughput and stability properties of the contention channel. However, as is well known, a key requirement to maintaining system stability or throughput is some kind of knowledge or estimate of the number of active users who have not yet successfully contended for resources. To obtain these estimates in the context of multiple simultaneous successes and collisions, we leverage the unique capability of the receiver in Section 3.2.2 for providing feedback on the collision multiplicity.

3.4.1 Throughput

Using the probability of failure when n users contend in a frame in (3.9), we will now find an expression for the throughput, $\eta(\nu)$, given Poisson transmissions with rate ν. Then, for a system with an infinite population of users with Poisson traffic with aggregate

arrival rate λ, we will apply the results of [33], [34], to find the stability region, in terms of the backlog, and the contention policy that maintains the maximum throughput when λ is outside the stability region.

Following the notation of [33], let $\{\epsilon_{nk} \colon 0 \leq k \leq n\}$ be the probabilities of receiving k packets correctly when n are transmitted. From (3.2), we have

$$\epsilon_{nk} = \binom{n}{k} (1 - \phi_n)^n \phi_n^{n-k}. \tag{3.10}$$

The average number of packets correctly received when n are transmitted is then given by $C_n = \sum_{k=1}^{n} k\epsilon_{nk}$ and

$$\begin{aligned} \eta(\nu) &= \sum_{n=1}^{\infty} C_n P(n|\nu) \\ &= \sum_{n=1}^{\infty} C_n \frac{\nu^n}{n!} e^{-\nu} \end{aligned} \tag{3.11}$$

is the throughput for rate ν Poisson transmissions. Fig. 3.3 shows $\eta(\nu)$ for the model in (3.9) when there are $P = 31$ virtual subslots and $M = 1$ contention slot per contention frame, with $N = 11$. It is interesting to note that the curve $\eta(\nu)$ is shaped very similar to the well-known throughput curve for single-packet reception (SPR) slotted ALOHA [7], [11], [53], [56]. Thus, much of the intuition into the throughput and stability characteristics of the SPR channel is the same for the MPR channel of Section 3.2.

3.4.2 Stability

For the stability analysis of the design, we assume a delayed first-transmission (DFT) model wherein newly arrived users must wait until the next full contention frame to contend. The aggregation of these new arrivals forms a Poisson process with rate λ. Let X_t be the backlog of users who have not yet successfully contended prior to frame t.[2] With S_t and B_t denoting the number of successful transmissions and the number

[2]Note that we now assume the contention model where newly arriving users to the service set of the AP contend in the contention channel with a single contention packet.

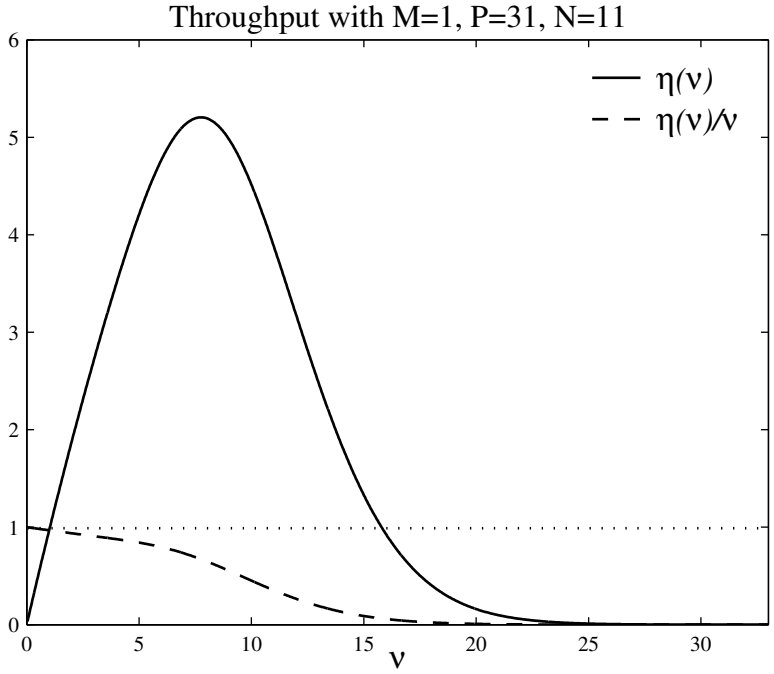

Figure 3.3. Throughput, $\eta(\nu)$, for rate ν Poisson transmissions with $M = 1$, $P = 31$, and $N = 11$.

of new arrivals, respectively, during frame t, the evolution of the backlog is given by $X_{t+1} = X_t + B_t - S_t$. As is standard for random access systems, we define the system to be stable if the homogeneous Markov chain $(X_t)_{t\geq 0}$ is ergodic.

We first consider the scheme where all users transmit their packets with the same fixed probability p, $0 \leq p \leq 1$, i.e., the uncontrolled MPR channel. In this setting, Ghez et al. [33] showed that $(X_t)_{t\geq 0}$ is ergodic only for arrival rates satisfying $\lambda < \lim_{n\to\infty} C_n$. However, for the general model in (3.6) with finite P, $\lim_{n\to\infty} \phi_n = 0$, and indeed for the simplified model in (3.8), $\phi_n = 0$ for $n > N$.[3] Therefore, $\lim_{n\to\infty} C_n = 0$ and, for any fixed contention policy p, the system is unstable for $\lambda > 0$. This result is the same as for the uncontrolled SPR channel with an infinite user population [11], [53], [86], [97].

Since a fixed contention policy is inadequate for stabilizing this system, we consider a dynamically controlled policy, $p_t = F(X_t)$, that is a function of the backlog (or estimated backlog, $\hat{X}_t$) at the beginning of frame t. The goal of this approach is similar to those

[3]For the PHY of Section 3.2.2, the DSSS processing gain is the limiting factor on ϕ_n since the number, P, of virtual subslots in a contention slot can, in general, be made large enough.

for the SPR channel [41], [56], [84], however we must apply the analysis of Ghez et al. in [34] for dynamic control in the MPR channel.

We briefly summarize the results from [34] relevant to our analysis and then apply them to our design. First, define $t_n(q)$ as the average number of successes in a slot given a backlog of size $X_t = n$ and transmission probability q:

$$t_n(q) = \sum_{k=1}^{n} \binom{n}{k} q^k (1-q)^{n-k} C_k. \tag{3.12}$$

By Theorem 1 of [34], there exists a transmission probability q_n^* that minimizes the expected backlog increase when the backlog is equal to n. Further, with this transmission probability, the system is stable for $\lambda < \eta_{max}$ and unstable for $\lambda > \eta_{max}$, where

$$\begin{aligned} \eta_{max} &= \lim_{n\to\infty} t_n(q_n^*) \\ &= \sup_{\nu \geq 0} \sum_{n=1}^{\infty} C_n \frac{\nu^n}{n!} e^{-\nu} \\ &= \sup_{\nu \geq 0} \eta(\nu) \end{aligned} \tag{3.13}$$

is the maximum throughput determined in (3.11). In other words, as the backlog grows large, the Bernoulli transmission process of (3.12) goes to the rate $\nu = nq_n^*$ Poisson limit and the transmission probability q_n^* maintains the throughput at its maximum, η_{max}. In addition, the system backlog is stabilized when $\lambda < \eta_{max}$. To specify the control policy, $p_t = F(X_t)$, we apply Theorem 2 of [34], which states that if $\eta_{max} > C$, there exists a constant $A > 0$ such that the control $p_t = A/X_t$ yields the maximum throughput, η_{max}, when $X_t > A$. In the proofs of Theorems 1 and 2, Ghez et al. [34] show that A is given by the value for which $\eta(A) = \eta_{max}$. This results in a control policy that yields the expected number of transmissions in a frame that achieves the maximum throughput possible. Intuitively, this policy maintains the throughput at its maximum value for large backlogs so that, as long as the average arrival rate is less than the maximum throughput, the average backlog will tend towards zero.

The results of the preceding discussion for the system in Section 3.2 are summarized as follows:

1. For a fixed contention policy, p, the system is unstable for any arrival rate $\lambda > 0$.
2. For a dynamic control policy, the system can be stabilized only for $\lambda < \eta_{max}$.
3. For backlog X_t at the beginning of frame t, the policy that minimizes the backlog increase, and stabilizes the system when $\lambda < \eta_{max}$, is $p_t = A/X_t$, where A is determined from $\eta(A) = \eta_{max}$.

To find the value of A for the new MPR channel, we must find η_{max} for (3.11). Although (3.11) becomes analytically cumbersome with increasing N for the systems (3.6) and (3.8), $\eta(\nu)$ does quickly go to its limit $C = 0$. Thus, a simple numerical search can be performed to find η_{max} and the value of A for which $\eta(A) = \eta_{max}$. For the model of (3.8) shown in Fig. 3.3, the value $A = \nu \approx 7.76$ results in the maximum throughput of $\eta_{max} \approx 5.20$! This result is a significant improvement over 802.11 systems [3] based on the standard SPR slotted-ALOHA system [6], [85], where the well known maximum *stable* throughput is $\eta_{max} = e^{-1} \approx 0.367$.[4]

3.5 New Feedback for Backlog Estimation

For the stabilization method described in the previous section, a technique is required for estimating the backlog in an uncoordinated fashion. To do this, we extend Rivest's pseudo-Bayesian estimation framework in [84] to our MPR system. Further, we leverage the receiver's capability to determine how many users transmitted packets in a given frame, regardless of the number who succeeded or failed.[5]

In general, the backlog estimate at the start of frame $t+1$, $\hat{X}_{t+1}$, is the previous estimate, $\hat{X}_t$, plus the expected number of new arrivals, λ, plus an evolution factor, $c\big(f[t]\big)$, that is a function of the outcome in frame t:

$$\hat{X}_{t+1} = \hat{X}_t + \lambda + c\big(f[t]\big). \tag{3.14}$$

[4]Note that standard SPR slotted-ALOHA is a simple case of the general model in [33], [34], and the analysis therein confirms the $\eta_{max} = e^{-1}$ result for the policy $p_t = 1/X_t$.

[5]In practice, the DMMSE receiver is capable detecting the number of received transmissions so long as the total number is below a threshold of roughly $2N$, where N is the DS-CDMA processing gain.

In [84], $f[t] = \{0, 1, e\}$ is the feedback for 0, 1, or more than one transmission, respectively, in frame t, and the pseudo-Bayesian update is given by

$$c\big(f[t]\big) = \begin{cases} -1 & \text{for } f[t] = \{0, 1\}, \\ \frac{1}{e-2} & \text{for } f[t] = e. \end{cases} \tag{3.15}$$

For the system described in Section 3.2, we must modify (3.15) for richer feedback and a more generalized notion of under- or over-utilized channel resources: not only is the DMMSE receiver capable of determining the number of user transmissions, but now holes (unused, but available, resources), successes, and collisions are possible (and not necessarily equal to 1) in the same slot.

In the MPR contention channel, the desired number of received transmissions in frame t is A, and is achieved by the transmission policy $p_t = A/\hat{X}_t$. Using a new backlog estimation correction factor, ξ_t, and the actual number of received transmissions, Y_t, we define the new form of $c(f[t])$ as:

$$c\big(f[t]\big) = \xi_t - S_t \tag{3.16}$$

where the feedback $f[t] = (Y_t, S_t)$. The new correction factor is given by

$$\xi_t = Y_t - A. \tag{3.17}$$

Thus, the previous notion of holes is now interpreted as the number of "missing" transmissions from the expected number under transmission policy $p_t = A/\hat{X}_t$. In similar fashion, the previous notion of collisions is interpreted as the number of "extra" transmissions. Simulation results of this technique are presented in Section 3.6.

3.6 Performance Evaluation: Simulations

In this section, we present some simulation results to demonstrate the performance of the system and verify the accuracy of the analyses in Sections 3.3-3.5. We use $P = 31$ in all of our simulations. The reservation packets are 2000 symbols in length with the

first $T = 4 \times P = 124$ fixed as the training symbols (increasing the training period beyond this does not yield appreciable performance gains). The processing gain is set as $N = 11$, as in 802.11b. All users see independent, flat Rayleigh-fading channels, generated using the modified Jakes' Sum of Sinusoids simulator [51], [75], with the Doppler offset–symbol period product set to $f_m T_s = 0.01$. This corresponds to a vehicular velocity of approximately 120 kph with a 900 MHz carrier. Also, since the AP's receiver does not have carrier synchronization with the transmitting users prior to contention, the receiver's local oscillator is simulated such that it has an independent, random synchronization offset with each user of up to 50 parts per million. Finally, the simulated symbol energy to noise ratio is $E_b/N_0 = 24$ dB and contention is successful if a packet is demodulated with a maximum uncoded bit error rate (BER) of 3%. In addition to the DMMSE receiver under the scenario given by the above parameters, we also simulate, for exactly the same traffic arrival processes, the simplified ideal reception model as given in (3.9) (i.e., failures are caused only by virtual subslot collisions or the number of users exceeding N in a contention slot, but there is no degradation due to RF MAI).

For the first experiment, the performance in a single contention frame is simulated for $M = \{1, 3, 5\}$ contention slots per frame for the two different receivers. Fig. 3.4 shows the expected number of successes, C_n, given n users contending in a contention frame for the two receivers, while a third curve (dots only) shows the result analytically derived for the ideal PHY model in Section 3.4.1. The plot shows the dramatic performance improvement relative to standard ALOHA [6], [11]: for $M = 1$, the number of successes for ALOHA would be equal to 0 for $n > 1$, whereas our method can easily sustain good performance for n up to 10. As expected, each configuration does reach a maximum point; if n is too large, then the performance deteriorates due to higher rates of collisions in virtual subslots as well as contention slot interference too large for the MAI suppression capabilities of the DMMSE receiver. In Fig. 3.5, the three curves show the corresponding probabilities of successful contention (where the analytical curve is obtained from (3.9)). The exact match between the analytically derived probability and the simulation for the

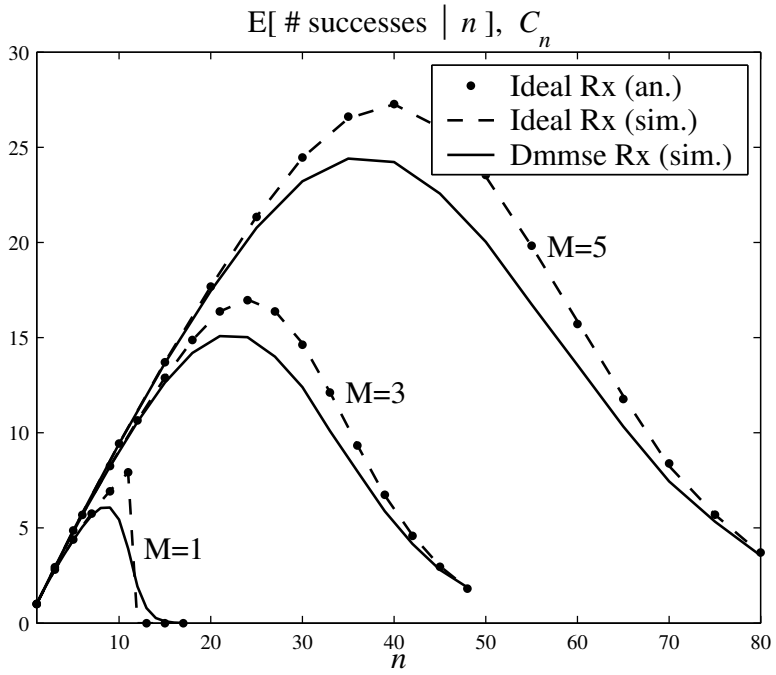

Figure 3.4. Expected number of successes in a contention frame, C_n, given n user transmissions: analytical and simulated data for ideal reception, and simulated data for DMMSE reception.

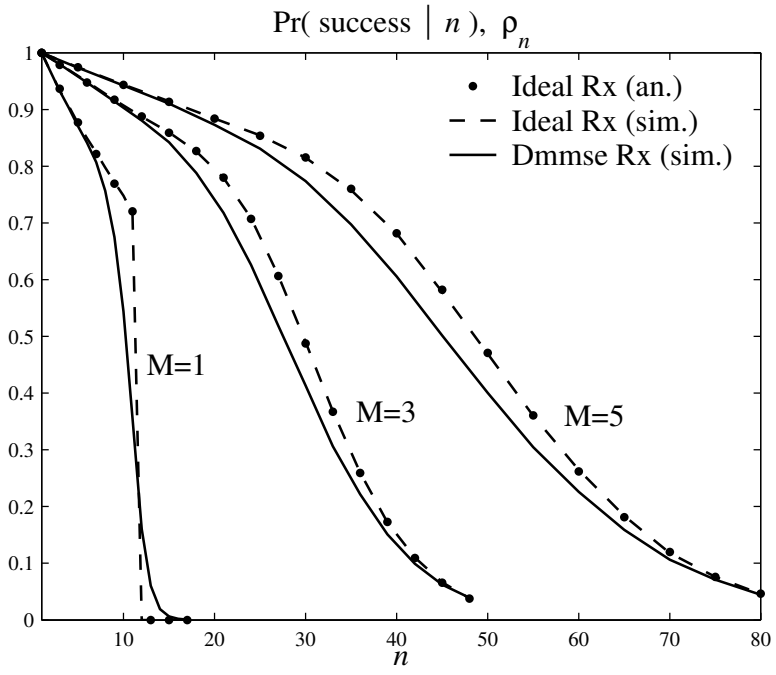

Figure 3.5. Probability of success in a contention frame, ϕ_n, given n users contending: analytical and simulated data for ideal reception, and simulated data for DMMSE reception.

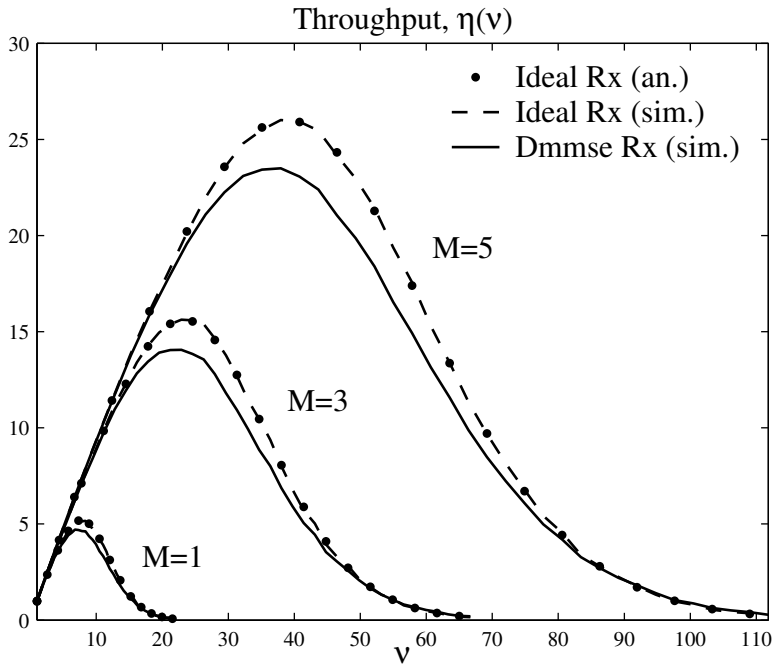

Figure 3.6. Throughput, $\eta(\nu)$, given Poisson transmissions with rate ν: analytical and simulated data for ideal reception, and simulated data for DMMSE reception.

ideal PHY model validates the analysis. More importantly, the close match between the simulations of the DMMSE receiver and its idealized model indicates that the ideal PHY model employed in the analysis adequately describes the mechanisms causing contention failure. The small difference in the results is attributable to the hard threshold model used in the analysis of contention slot interference: in practice, the transition to failure is smoother since the performance of the DMMSE receiver does deteriorate with increasing U, even when $U < N$, and the DMMSE receiver can occasionally perform successful decoding even when $U \geq N$.

The second experiment tests the system when the transmission process is Poisson with rate ν when there are $M = \{1, 3, 5\}$ contention slots per frame for the two different receivers. Fig. 3.6 shows the resulting throughput, $\eta(\nu)$, and the third curve (dots only) shows the throughput analytically derived in (3.11) for the ideal PHY model. Again, the analysis matches the simulation for the idealized model, while the DMMSE receiver performs close to the ideal reception model. From these curves, we obtain the (A, η_{max})

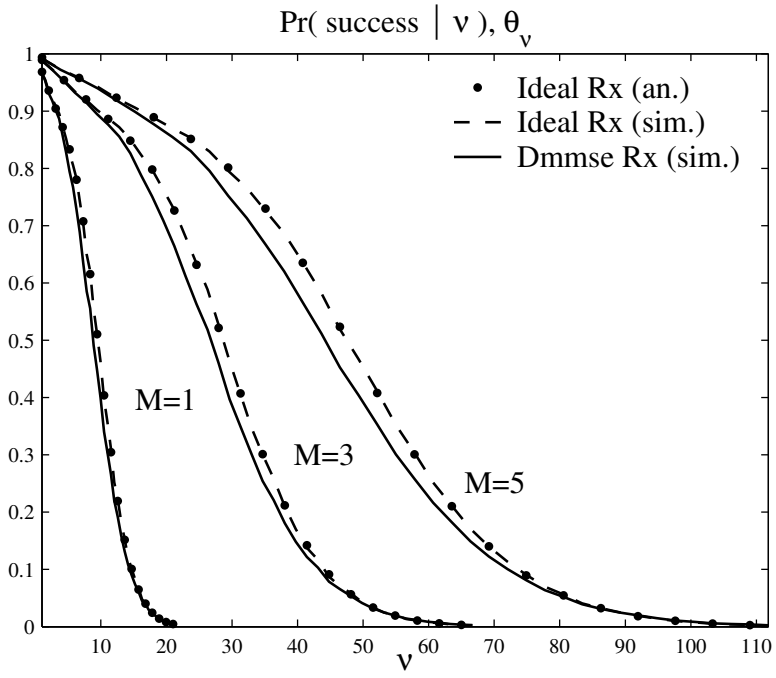

Figure 3.7. Probability of success, θ_ν, given Poisson transmissions with rate ν: analytical and simulated data for ideal reception, and simulated data for DMMSE reception.

Table 3.1. A and η_{max} for $M = \{1, 3, 5\}$

	Ideal R_X		DMMSE R_X	
M	A	η_{max}	A	η_{max}
1	7.76	5.22	7.40	4.7
3	23.28	15.6	22.23	14.1
5	38.69	26.0	37.14	23.5

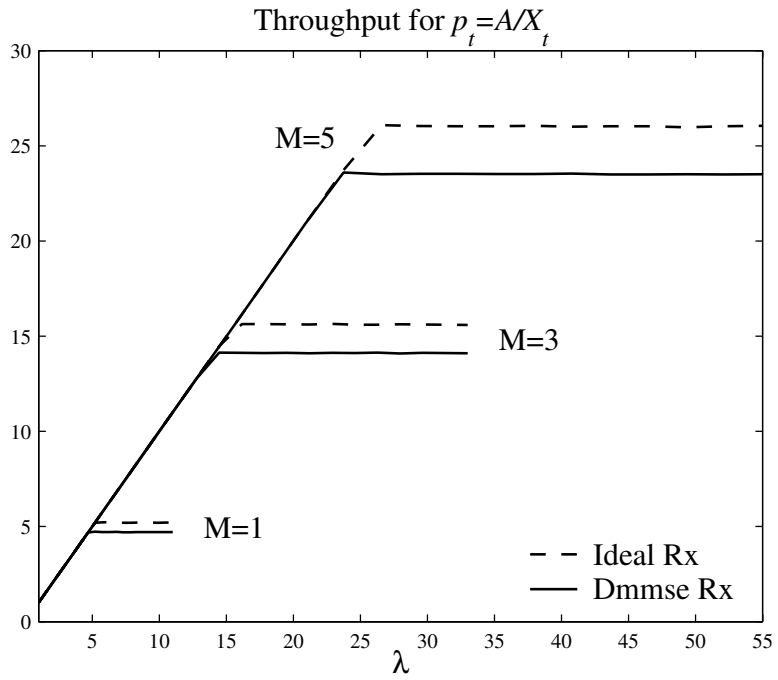

Figure 3.8. Throughput with dynamic control policy $p_t = A/X_t$ for backlog X_t and rate λ Poisson arrivals, simulated for both ideal and DMMSE reception.

pairs of Table 3.1 to be used for stabilizing the system with the dynamic control policy $p_t = A/X_t$. Further, we compare the maximum stable throughput of the DMMSE-based system in Section 3.2 for the case of $M = 1$ to that of $N = 11$ parallel SPR slotted-ALOHA channels by normalizing $\eta_{max} = 4.7$ by the bandwidth expansion factor, N. The resulting normalized maximum stable throughput for the DMMSE-based system is $\tilde{\eta}_{max} = 0.427$, which shows that the use of multiuser detection realizes a multiplexing gain, in this case 16%. It is reasonable to expect these gains to translate to higher normalized maximum throughput bounds when collision resolution techniques similar to those for the SPR channel [20], [65], [74], [100] are applied. As the throughput advantage over N parallel ALOHA is moderate, the gains do not necessarily translate to delay performance. However, the additional capability to detect the number of colliding users can be leveraged to provide delay guarantees to real-time traffic in the presence of delay-insensitive traffic, as shown in Chapter 4.

For the third experiment, the stability of the system is tested for the two receivers

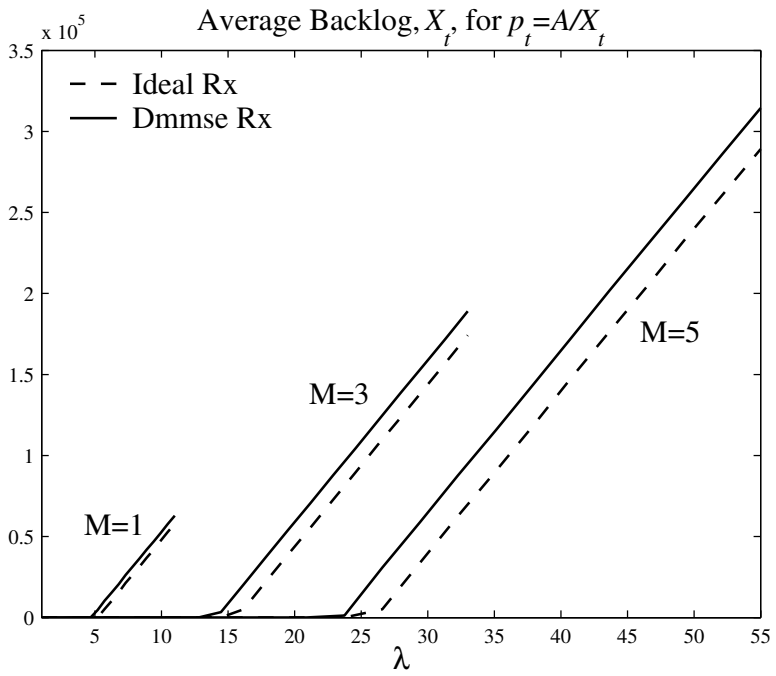

Figure 3.9. Average backlog at the end of 10000 frames with dynamic control policy $p_t = A/X_t$ for backlog X_t and rate λ Poisson arrivals, simulated for both ideal and DMMSE reception.

when there are $M = \{1, 3, 5\}$ contention slots per frame and the dynamic control policy $p_t = \min(A/X_t, 1)$ for backlog X_t is enforced. The values used for A are given in Table 3.1. For this experiment, user arrivals at the AP are Poisson with rate λ, contention follows the DFT model, and users remain in the backlog until successful contention, regardless of the incurred delay. We assume here the presence of a genie providing exact information of the instantaneous backlog, X_t, although this may be replaced by the backlog estimation technique demonstrated next in the fourth experiment. It is clear from Fig. 3.8, which shows the throughput, that the dynamic control policy stabilizes the throughput at the minimum of either the offered load, λ, or the maximum throughput, η_{max}. Note that each data point represents an average over 10 trials of the final backlog after 10000 frames for each value of λ. The resulting maximum throughputs (the values to which the curves flatten horizontally) match the peaks of the curves in Fig. 3.6, as predicted for η_{max} in Table 3.1. Further, as shown in Fig. 3.9, the dynamic control policy also stabilizes the backlog. However, the values for λ at which the backlogs start

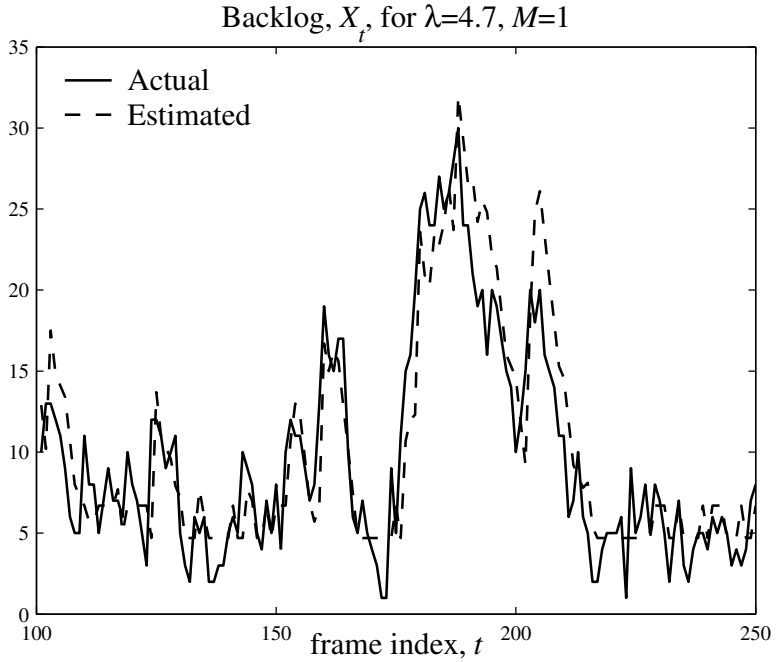

Figure 3.10. Backlog estimation for $M = 1$ contention slots per frame and Poisson arrivals with rate $\lambda = 4.7$, simulated for DMMSE reception.

to grow are slightly less than $\lambda = \eta_{max}$, since, in the neighborhood of the critical point η_{max}, the transmission process of (3.13) has not yet gone to its Poisson limit and thus contains transmissions from the Poisson arrivals as well as a significant non-Poisson retransmission process.

In the fourth experiment, the backlog estimation technique of Section 3.5 is tested using DMMSE reception. Fig. 3.10 shows the actual backlog and the estimate generated using (3.16)-(3.17). The data is plotted from a single trial and shows that the estimate closely tracks the actual value. The estimate "floor" exhibited in Fig. 3.10 is a result of adding the fixed value of the traffic arrival rate λ to the sum of the previous estimate and the evolution factor in (3.14).

3.7 Discussion

Our joint PHY/MAC design can easily support a large volume of random access contention traffic, and meets the design objective of enabling rapid, mobile–centric handoffs

in pseudocellular networks. This is in contrast to standard ALOHA, which can only support one user at a time. The 802.11b standard provides performance little better than standard ALOHA despite the use of spreading, since the spreading code and training sequences used by all users are the same. In contrast, the DMMSE-based receiver, when used with a suitably designed set of training sequences, is shown to enable the successful decoding of multiple simultaneous users without *a priori* carrier synchronization, channel information, or knowledge of the users' virtual subslots and spreading sequences. This permits users to choose spreading sequences randomly without coordinating with the AP, which is key to the multiple-access capability of our contention channel.

The throughput of the new MPR contention channel is shown to be stabilized by using a dynamic control policy when knowledge (or estimates) of the system backlog is available. Further, when the arrival rate of new traffic is below a threshold, the system backlog is also stabilized. This stabilization requires accurate estimates of the backlog, which were shown to be possible using the capability of the PHY to detect the number of transmissions received, regardless of contention success or failure. In the next Chapter, the backlog estimation technique is extended to the case of multiple priority classes. Finally, if information on the traffic characteristics is available, a means for dynamically expanding the capacity of the contention channel by increasing the number of contention slots per contention frame was demonstrated.

While we have focused on pseudocellular networks as the driving application, it is of future interest to explore similar PHY/MAC designs permitting communication with minimal coordination (e.g., in the presence of rapid topology variations) for the *data* channel in peer-to-peer ad hoc networks.

Chapter 4

Quality of Service for Delay-Constrained Traffic

The significant increase in throughput capacity of the system in Chapter 3 over traditional multiaccess channels enables its use in settings with high rates of handoffs and delay constrained traffic. While the stability analysis addressed only stabilization of the backlog or throughput, it does provide a starting point for designing a system that supports QoS guarantees.

4.1 Motivation

The classical slotted-ALOHA model [6], [53], [85] is based on the assumption that exactly one packet can be successful in a given slot. Advances in multiuser detection techniques over the last two decades imply, however, that systems with multipacket reception capabilities are now becoming practically feasible. In Chapter 3, we have shown that joint optimization of the PHY and MAC layers based on a receiver capable of adaptive multiuser detection leads to an ALOHA-like system which is capable of supporting both *Hi*-priority users with deadline constraints, and *Lo*-priority users with best-effort service. The specific application considered in Chapter 3 was the design of a contention

channel for rapid handoffs in a *pseudocellular* network, which supports vehicular mobility, as in cellular networks, using small pseudocells in a WLAN-type infrastructure. For example, a voice connection at vehicular speeds can be supported by ensuring that the mobile can make a timely reservation with the AP whose pseudocell it is entering, thus achieving a mobile-centric handoff. QoS for pseudocellular networks was also considered in [68], where contention policies are derived to provide support for deadlines without MUD capabilities at the receiver. MUD capabilities at the receiver enable us to achieve much higher throughput in the pseudocellular contention channel, as demonstrated in Chapter 3. However, the study of collision channels with MPR is of fundamental importance beyond its application to pseudocellular networks. The purpose of this Chapter is to abstract an analytical model of the system in Chapter 3, in order to obtain, under realistic constraints, a general design framework for QoS for a collision channel with MPR capability. In contrast to prior work on ALOHA with MPR, our modeling assumptions are specifically guided by the capabilities of the adaptive multiuser detection strategy used at the receiver.

Uncontrolled ALOHA with MPR was studied in [33], where it was shown that the system is stable for arrival rates below a threshold if the number of successes has a nonzero limit as the collision size gets large. The latter condition amounts to requiring some form of capture, and is typically not satisfied when the number of contending users exceeds the capability of the multiuser detector (in which case, most likely, all users fail to decode). There are also other similarly restrictive stability conditions that have been considered in the literature, such as assuming a lower bound on the signal-to-interference ratio [69]. In a wireless network with mobiles contending for communicating with an AP, however, the AP can control the contention policies used by the mobiles based on its knowledge of the network conditions. It is known that classical ALOHA can be stabilized by controlling the transmission probability based on estimates of the backlog [11], [41], [84], [96], and this theory was extended to the MPR channel in [34] as described in Chapter 3. In this chapter, we extend the pseudo-Bayesian backlog

estimation method in [84] and Section 3.5 to multiple priority classes, including both delay-constrained and delay-tolerant traffic. For multiple priority classes, the extension is similar to that for SPR in [26] for mixed priorities. We then employ these estimates to design stable contention policies that provide QoS to *Hi*-priority users (i.e., assuring that they meet a deadline with a specified probability), while maximizing the throughput for *Lo*-priority users. We develop a Markov chain description that specifies the throughput as a function of the traffic arrival rates for each priority class. With the throughput characteristics obtained from the Markov analysis, the backlog estimates are used to generate the contention policies (e.g., to be broadcast by the AP in its beacon prior to each frame). Note that the availability of estimates of collision size from the physical layer simplifies the design, relative to prior work which relies on higher layer information for backlog estimation [112].

We assume only three general properties of the receiver for the analytical model of the MPR channel. First, although users transmit simultaneously in synchronized contention frames, there is a limit, N, to the maximum number of users that can be successfully decoded by the receiver. If the number of transmissions is above this limit, none can be decoded. As an example, this property is analogous to the DS-CDMA processing gain limit on the MUD capabilities of the system in Chapter 2. Second, there is a limit, P, to the number of unique ways, or *virtual subslots*, that users may use to transmit their packets to the AP. When transmitting a packet, a user randomly chooses one of these virtual subslots and, so long as the total number of transmissions is less than N and no other user has chosen the same virtual subslot, that user's packet is successfully decoded. The virtual subslots may be provided by any novel receiver technique, an example of which is the user's choice of training symbol sequence in Chapter 3. The third assumption, which is critical to our backlog estimation algorithms, is that the receiver is able to determine the total number of transmissions that occurred in a given frame, even when some or none of them are successfully decoded. This assumption is approximately satisfied by the DMMSE receiver.

From the analysis developed based on the preceding model, contention policies may be established that provide the required QoS to delay-constrained traffic. However, these policies require estimates of the traffic, and specifically the backlog of users who have not yet successfully contended for resources. We assume that the delay-constrained traffic does not remain in the system indefinitely (i.e., calls are dropped if a timely handoff is not made) and we must, therefore, modify the backlog estimation technique of Section 3.5 to account for these new traffic characteristics.

The remainder of this chapter is organized as follows. In Section 4.2, we begin with a description of a simplified, single-priority class traffic model that can be extended to more general cases. We extend this model in Section 4.3 to the case of mixed-priority traffic, and describe two potential methods for the mixed traffic to jointly utilize the contention channel while maintaining required QoS measures. For both methods, a backlog estimation technique is required for the mixed-traffic contention policies, and we extend our previous technique for this purpose. We demonstrate the contention and backlog estimation techniques in simulations in Section 4.6, and conclude with a brief discussion in Section 4.7.

4.2 Single Priority Class Traffic

For the model, all users are considered to be of the same single class of traffic with a specified QoS requirement. For a specified maximum delay (measured in frames) before successful contention, D, the QoS metric of interest is the minimum expected probability of contention success, R_s. We assume that new users arriving at an AP, via handoffs or initiating new connections, form a single Poisson arrival process of rate λ. Further, if a user does not succeed within the specified delay budget, their call is dropped and they exit the system.

Our goal is to provide the required QoS, while minimizing the system resources used for the contention channel (i.e., maximizing the contention channel utilization under the

QoS constraint). Thus, for a given set of system parameters, $\{M, N, P, R_s, D\}$, we will find the maximum arrival rate, λ_{max}, for which QoS is satisfied. In general, the QoS constraints are tight enough that merely operating in the backlog or throughput stability regimes of Chapter 3 will not provide the required support. In particular, although we certainly must have $\lambda < \eta_{max}$ so as not to exceed the maximum possible throughput, this bound on λ does not guarantee that users will succeed within the given D frames with probability greater than some specified R_s.

The QoS guarantee we aim to provide is $R_s \geq 0.99$ (i.e., a 99% success rate). The analysis for the case when $D = 1$ is straightforward since the transmission process is also Poisson with rate $\nu = \lambda$.[1] Using (3.11), the expected probability of contention failure, θ_ν, for the rate ν Poisson transmission process, is given by

$$\theta_\nu = 1 - \frac{\eta(\nu)}{\nu}. \tag{4.1}$$

Thus, when $D = 1$, $\lambda_{max} = \sup_\lambda\{\lambda : \theta_\nu\big|_{\nu=\lambda} < 1 - R_s\}$. In Fig. 4.1 (reprised here from Chapter 3), the dashed line shows $\eta(\nu)/\nu$, while the horizontal dotted line shows the value $R_s = 0.99$. The maximum arrival rate for which $\theta_\lambda < 0.01$ is shown to be $\lambda_{max} \approx 0.3$. In Section 4.6, plots of simulated and analytical versions of $\eta(\nu)$ yield the stringent value $\lambda_{max} \approx 0.3M$ for which $\theta_\nu < 0.01$ when $D = 1$.

If the delay constraint, D, is relaxed to allow users to contend in multiple successive frames before failure, it is reasonable to expect that a higher traffic arrival rate may be supported while still satisfying the QoS requirements. However, in this case, the total transmission process seen by the AP is no longer strictly Poisson since it consists of an additional retransmission process. To determine λ_{max} for this case, we introduce a Poisson approximation to the transmission process seen by the AP which provides a loose upper bound, followed by a a detailed Markov analysis of this scenario to obtain tighter performance predictions.

[1] The transmission process has the same rate, λ, as the arrival process since users either succeed or fail and exit the contention system after $D = 1$ attempts.

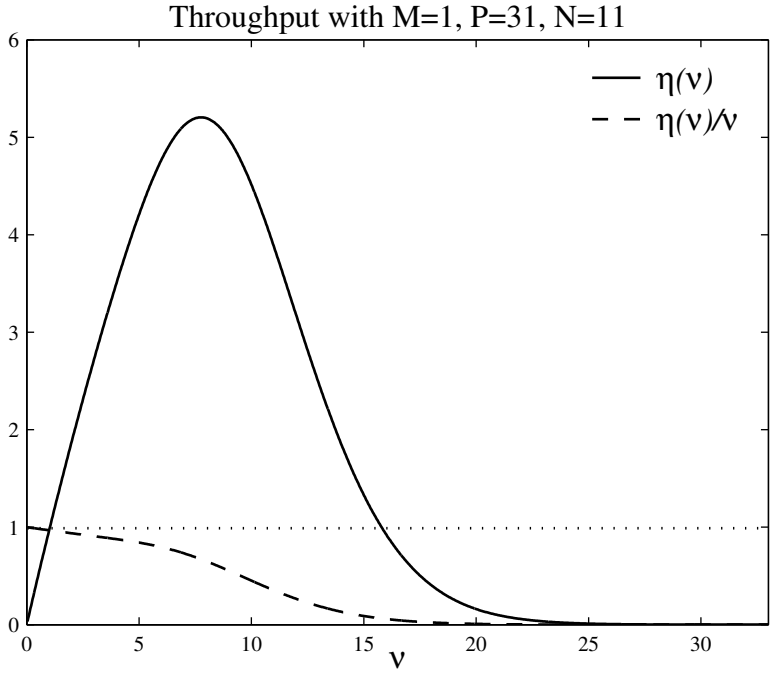

Figure 4.1. Throughput, $\eta(\nu)$, for rate ν Poisson transmissions with $M = 1$, $P = 31$, and $N = 11$.

4.2.1 Poisson Approximation

With $D > 1$, users who fail in a given frame, but whose delay budget is not yet expired, transmit with probability 1 in the next contention frame. The operating point for which we are designing, $R_s \geq 0.99$, yields a low rate transmission process such that failures in a frame are caused primarily by virtual subslot collisions and not overloading of the DMMSE MUD at the receiver. Therefore, in the QoS operating region, failures across consecutive frames are approximately independent.[2] Thus, we assume the retransmission process is "small enough" that we may approximate it as a small fractional function of the arrival process and, therefore, that it is also Poisson. The result is that we approximate the effective transmission process as Poisson with rate

$$\gamma = \lambda(1 + \theta_\gamma + \ldots + \theta_\gamma^{D-1}), \tag{4.2}$$

[2]Once the system is outside the QoS operating region, failure due to overloading the DMMSE MUD becomes significant as well, and failures across consecutive frames becomes highly correlated. This effect is shown by the simulations in Section 4.6.

where $\theta_\gamma = (\theta_\nu)_{\nu=\gamma}$ is the expected probability of failure in a frame for rate γ Poisson transmissions.

Under the assumption that the probability of failure is independent across frames, the probability of failure in D consecutive frames, and thus failure in contention, is given by θ_γ^D. To satisfy the QoS requirement, we must have $\theta_\gamma^D < 1 - R_s$. Then, from (4.1), we can find the maximum transmission rate, γ_{max}, for which $\theta_\gamma < (1 - R_s)^{1/D}$ (i.e., QoS) is satisfied. Finally, using (4.2), we can determine the maximum arrival rate, λ_{max}.

As an example, to satisfy QoS for $D = 3$, the maximum probability of failure in a frame is given by $\theta_\gamma < 0.215$. With $M = 1$ contention slots per contention frame, the analyses of Sections 3.3.1 and 3.4.2 result in $\gamma_{max} \approx 6.19$, and thus $\lambda_{max} \approx 4.9$. The approximation we have used, however, is optimistic. In particular, the λ_{max} resulting from this approach provides a strict, but loose, upper bound since the assumption of independence of failures across consecutive frames is violated by the facts that a minimum of 2 users must collide to cause subslot failure, and N must collide for collision slot failure. In Section 4.6, we evaluate the accuracy of this bound through simulations. In order to more tightly predict the system's performance for delay-constrained traffic and establish system design parameters, we develop a Markov chain in the next section.

4.2.2 Markov Model

We consider here a simple special case of the MPR multiaccess contention channel in [33], based on the properties described in Section 4.1: a user may collide and fail only if either the total number of users exceeds a threshold, N, or another user chooses the same virtual subslot from the P available. For this channel, we extend classical ALOHA theory [11] to write the expression for the throughput as:

$$\eta(\nu) = \sum_{n=1}^{N} \overline{\mathrm{N}}_n(P) P_\nu(n). \tag{4.3}$$

Here, $P_\nu(n) = \frac{1}{n!}\nu^n e^{-\nu}$ is the rate ν Poisson probability of n transmissions. The expected number of successes given n users choosing from P virtual subslots is given by $\overline{\mathrm{N}}_n(P) =$

$\sum_{k=1}^{n} kF(k|n)$, where $F(k|n)$ is the probability of k out of n users choosing unique virtual subslots.

To determine $F(k|n)$, we first find the total number of ways for *at least* k out of n users to choose unique subslots, defined as $G(k|n)$. Given that the first user chooses from P subslots, followed by the next user choosing from $P-1$, and so on until the kth user chooses from $P-k+1$ subslots, the remaining $n-k$ users may choose any combination from the remaining $P-k$ subslots. Using the binomial coefficient to count the number of ways of making this set of selections, we have

$$G(k|n) = \begin{cases} \binom{n}{k} P(P-1)(P-2)\dots(P-k+1)(P-k)^{n-k} & \text{for } k \leq n, \\ 0 & \text{for } k > n. \end{cases} \tag{4.4}$$

Alternatively, since there are P^n ways to assign all n users to the P subslots, $P^n F(k|n)$ gives the number of different ways of uniquely assigning a particular set of exactly k users among the P subslots (with $n-k$ having non-unique assignments). Since these exact sets have been multiply counted by the overlaps in (4.4), we sum over $j = k, \dots, n$ the $\binom{j}{k}$ extra ways they have been counted in (4.4) to obtain the total number of ways of at least k users choosing uniquely, $G(k|n)$:

$$G(k|n) = \sum_{j=k}^{n} \binom{j}{k} P^n F(j|n). \tag{4.5}$$

Combining (4.4)-(4.5) and rearranging, we obtain the result

$$\sum_{j=k}^{n} \binom{j}{k} F(j|n) = \binom{n}{k} \frac{P!(P-k)^{n-k}}{P^n (P-k)!}. \tag{4.6}$$

In order to solve for the specific probabilities $F(k|n)$, (4.6) is first evaluated for $k = n$ to obtain $F(n|n)$. Then, in recursive fashion, $F(n-1|n)$ is found using the result $F(n|n)$, $F(n-2|n)$ is found using the results $\{F(n|n), F(n-1|n)\}$, etc.

Since, for $N < P$, $\eta(\nu)$ and $\nu_{opt} \triangleq \arg\max_{\nu} \eta(\nu)$ are increasing functions of N, and $\nu_{opt} \leq N$, rates higher than N will produce decreasing throughput. Therefore, we restrict the following analysis to the range $\nu \in [0, N]$.

We consider now the problem of providing QoS support for a single class of delay-constrained users where each needs to make a successful transmission within a delay deadline of D reservation frames with probability of at least R_s. The expectation is that the supported Poisson transmission rates will be less than ν_{opt} in order to satisfy this QoS. For the simple case of $D = 1$, each user either succeeds on its first attempt, or its deadline expires. Thus, there are no retransmissions and the total transmission process is the same as the arrival process of new users, which is Poisson with rate λ. Further, the probability of the deadline expiring is $P_{exp} = 1 - \frac{\eta(\nu)}{\nu}$ and is completely determined by (4.3). As an example, for $N = 11$ and $P = 31$, $\lambda = 0.311$ is the maximum sustainable arrival rate for which $P_{exp} \leq 1 - R_s = 0.01$. This is much smaller than the corresponding $\nu_{opt} = 7.757$ with throughput $\eta(7.757) = 5.20$ (note that this matches the results of Chapter 3). Intuitively, if the QoS requirement is relaxed by increasing the delay deadline to $D = 3$, a higher arrival rate might be supported. However, as discussed in the previous section, the transmission process then consists of a combination of a Poisson arrival process and a retransmission process, and the total process is no longer Poisson. Therefore, we must develop a Markov model for analysis. Before deriving the model, we state the assumptions and some notation.

- Users contend with a single packet in the MPR channel, and new arrivals form a Poisson process with rate λ. Without loss of generality, we assume that $N = 11$, $P = 31$, $D = 3$, and $R_s = 0.99$.
- A user with a contention packet transmits their packet in the contention frame with probability $p_{Tx} = 1$.
- For $D = 3$, the contention channel backlog is modeled as a two-dimensional Markov process where the i^{th} state $\mathbf{s}_i = (\mathrm{N}_1, \mathrm{N}_2)$ represents the number of backlogged users with N_j having already collided j times. Since $p_{Tx} = 1$, all packets are lost for any state where $\sum_j \mathrm{N}_j > N$. Therefore, the state space is truncated by restricting N_j to the range $[0, N + 1]$. This effectively neglects the probability of arrival bursts larger than $N+1$, however, for the range of λ satisfying $R_s \geq 0.99$, this assumption is justified. The total number of states in the Markov chain is $L = (N + 2)^{D-1}$.
- For a system in state $\mathbf{s}_i = (\mathrm{N}_1, \mathrm{N}_2)$ at the beginning of a frame where there are k

new arrivals, let n_1, n_2, and n_k denote the number of successes from the N_1, N_2, and k users, respectively. Further, the general probability of getting the a-tuple of $(n_1, n_2, \ldots, n_a)$ successes from the corresponding a-tuple of $(\mathrm{N}_1, \mathrm{N}_2, \ldots, \mathrm{N}_a)$ users in a frame is denoted by $f(n_1, n_2, \ldots, n_a|\mathrm{N}_1, \mathrm{N}_2, \ldots, \mathrm{N}_a)$.

In order to find the probability of deadline expiration, P_{exp}, we compute the stationary probability distribution $\mathbf{p} = [p_1, p_2, \ldots, p_L]^T$, where $p_i = p(\mathbf{s}_i)$, the stationary probability of state $\mathbf{s}_i$. The probability of transition from state $\mathbf{s}_i = (\mathrm{N}_1, \mathrm{N}_2)$ to state $\mathbf{s}_j = (k - n_k, \mathrm{N}_1 - n_1)$ is given by

$$Q_{i,j} = \sum_{k=0}^{N+1} P_\lambda(k) Q_{i,j|k}, \tag{4.7}$$

where $P_\lambda(k)$ is the (Poisson) probability of k new arrivals. The conditional probability of transition from state $\mathbf{s}_i$ to state $\mathbf{s}_j$ given k new arrivals, $Q_{i,j|k}$, is given by

$$\begin{aligned} Q_{i,j|k} &= \sum_{n_2=0}^{\mathrm{N}_2} f(n_k, n_1, n_2|k, \mathrm{N}_1, \mathrm{N}_2) \\ &= \sum_{n_2=0}^{\mathrm{N}_2} \frac{\binom{k}{n_k}\binom{\mathrm{N}_1}{n_1}\binom{\mathrm{N}_2}{n_2}}{\binom{k+\mathrm{N}_1+\mathrm{N}_2}{n_k+n_1+n_2}} F(n_k + n_1 + n_2|k + \mathrm{N}_1 + \mathrm{N}_2). \end{aligned} \tag{4.8}$$

Using (4.7), (4.8), and the results from (4.6), we can compute the stationary probability distribution $\mathbf{p}$, and similarly $E[\mathrm{N}_{exp,i} = \mathrm{N}_2 - n_2]$, the expected number of expired users in state $\mathbf{s}_i$. Averaging $E[\mathrm{N}_{exp,i}]$ over $\mathbf{p}$ yields the overall expected number of expirations, $E[\mathrm{N}_{exp}]$. Finally, $P_{exp} = \frac{1}{\lambda} E[\mathrm{N}_{exp}]$. The state space in Fig. 4.2 illustrates the contribution of each state to the overall expiration rate.

4.3 Multiple Priority Class Traffic

The previous section showed that it is possible to provide QoS to a single class of delay-constrained traffic. However, since it is possible for the QoS requirements to be "overly" satisfied, the analysis also suggests that there exists extra capacity that could be allotted to delay-tolerant traffic without violating the QoS for the delay-constrained traffic. Therefore, we consider now the case of supporting best effort traffic, using the

"spare capacity" from when the *Hi*-priority users' arrival rate, λ_{Hi}, is less than the maximum sustainable for QoS. In this situation, the resulting channel utilization should be higher than when all users have QoS constraints, but lower than $\eta(\nu_{opt})$ for no QoS constraints on any users. We describe two methods for enabling the additional service.

4.3.1 Sharing the channel

We extend the previous model by including the extra *Lo*-priority transmission process, which changes the transition probabilities of the Markov chain. Equations (4.7) and (4.8) can be rewritten for this case as:

$$Q_{i,j} = \sum_{k_{Lo},k_{Hi}=0}^{N+1} P_{\lambda_{Lo}}(k_{Lo}) P_{\lambda_{Hi}}(k_{Hi}) Q_{i,j|k_{Lo},k_{Hi}}, \tag{4.9}$$

$$Q_{i,j|k_{Lo},k_{Hi}} = \sum_{n_2=0}^{\mathrm{N}_2} f(n_{Lo}, n_{Hi}, n_1, n_2 | k_{Lo}, k_{Hi}, \mathrm{N}_1, \mathrm{N}_2). \tag{4.10}$$

As before, we compute $\mathbf{p}$ and P_{exp}. Results for the cases $\lambda_{Hi} = \{2.0, 2.5, 3.0\}$ are shown in Fig. 4.3. From these results, the maximum sustainable η_{Lo} corresponding to a given value of λ_{Hi}, i.e., the *Lo* capacity, $C_{Lo}\big(\lambda_{Hi}\big)$, is determined. After describing an alternative method for providing *Lo*-priority service next, we consider the design of contention policies that maintain these throughputs and QoS for the multiple priority class traffic.

4.3.2 Reserving the Channel

In Section 4.3.1, we discussed the possibility of sharing the contention channel between multiple classes of users. *Lo*-priority users were allowed to contend in the same channel, as long as the QoS constraint for the *Hi*-priority users was satisfied. In the "shared channel" scheme, the *Lo*-priority users have the same collision rates as the *Hi*-priority users, effectively forcing them to have the same effective maximum rate of failures per contention slot as the *Hi* users, despite not requiring the same QoS. In this section, we consider an alternative scheme, called "reserved channel", wherein we reserve a number,

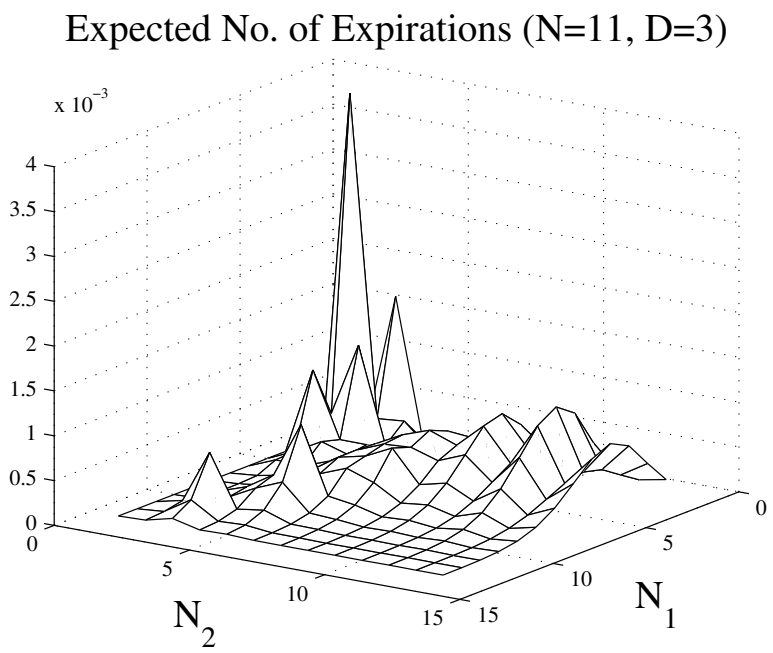

Figure 4.2. Expected number of deadline expirations in each state.

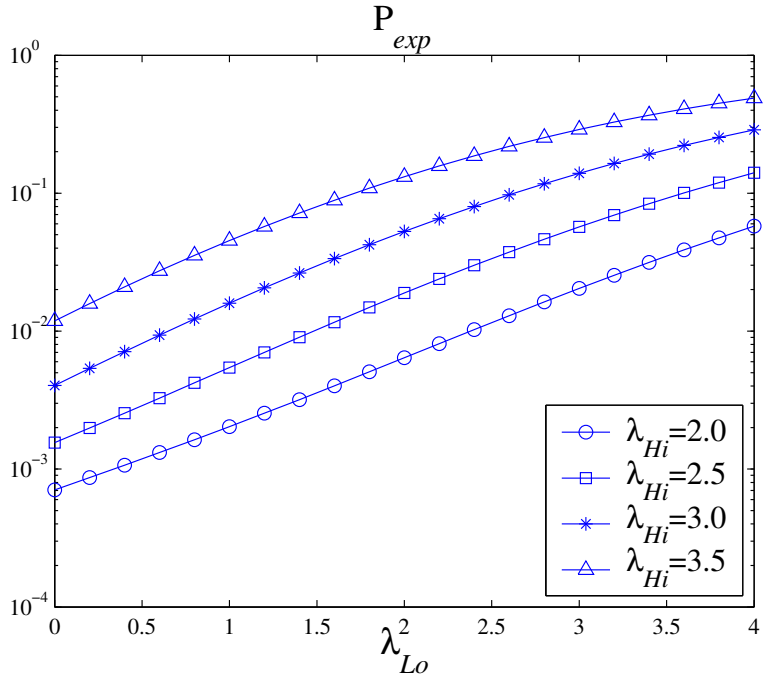

Figure 4.3. QoS for *Hi* priority users as a function of *Lo* priority traffic in an uncontrolled system.

P_{Hi}, of virtual subslots out of the total P available for only *Hi*-priority users, and constrain the *Lo*-priority users to contend only in the remaining $P_{Lo} = P - P_{Hi}$ subslots.

The intuition behind the reserved channel scheme is that, by isolating the *Lo*-priority users in a subset of virtual subslots, we can eliminate the possibility of *Lo*-priority users colliding with *Hi*-priority users. As a result, it might be possible to support higher transmission rates for the *Lo*-priority users without impacting the QoS of the *Hi*-priority users. However, while there is still a limit, N, to the *total* number of transmissions, there is an additional tradeoff for the reserved channel scheme. Namely, in the case of a fixed set of reserved subslots (i.e. P_{Hi} fixed ahead of time), there is a loss of multiplexing gain from dividing the available subslots into two smaller subsets. Further, as a result of the limit, N, even with subslot reservations, it is not possible to *completely* isolate the *Hi*-priority users from the effect of *Lo*-priority user traffic.

From this discussion, we expect that for a stricter the QoS constraint (e.g., a smaller deadline D, or a higher value of R_s), the reserved channel model becomes more attractive because the penalty suffered by the *Lo*-priority users increases. In order to evaluate these ideas quantitatively, we now extend the analytical model of Section 4.2.2 to the reserved channel case.

For the reserved scheme, we rewrite (4.7) and (4.8) to get:

$$Q_{i,j} = \sum_{k_{Hi}}^{N+1} P_{\lambda_{Hi}}(k_{Hi}) Q_{i,j|k_{Hi}}, \tag{4.11}$$

$$Q_{i,j|k_{Hi}} = \sum_{n_2=0}^{\mathrm{N}_2} f(n_{Hi}, n_1, n_2 | k_{Hi}, \mathrm{N}_1, \mathrm{N}_2). \tag{4.12}$$

We then compute P_{exp} as before, and by varying η_{Lo}, we can determine the maximum levels of *Lo*-priority transmissions while sustaining specified *Hi*-priority QoS requirements using subslot reservation P_{Hi}. The process is then iterated for different values of P_{Hi} to determine the optimum subslot allocation that corresponds to maximum value $\eta_{Lo}$and the corresponding P_{Hi}.

In Fig. 4.4, we have plotted the maximum throughput for *Lo*-priority traffic (which

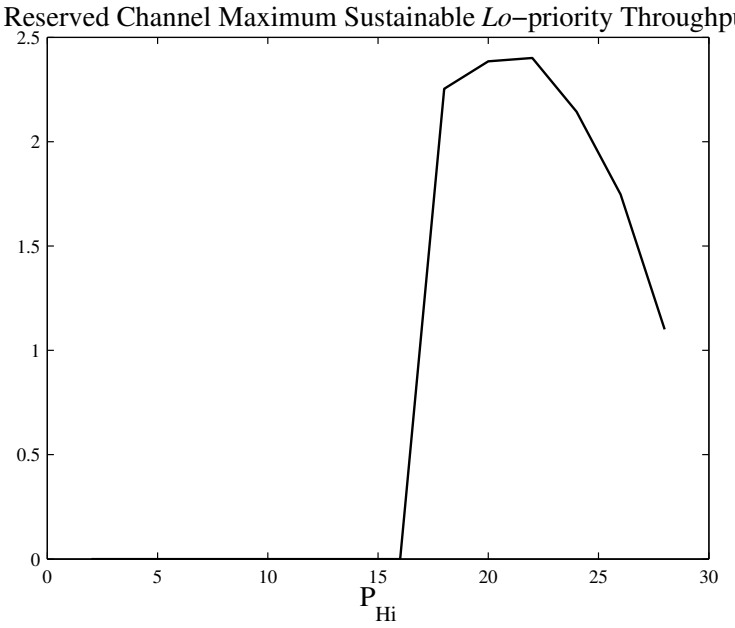

Figure 4.4. Maximum sustainable *Lo*-priority throughput while providing QoS to *Hi* users with parameters $\lambda_{Hi} = 1.0,\ R_s = 0.999,\ D = 3$ for varying values of P_{Hi}.

has a corresponding maximum *Lo*-priority transmission rate, η_{Lo}) for rate $\lambda_{Hi} = 1.0$ *Hi*-priority arrivals requiring QoS metrics ($R_s = 0.999, D = 3$), against different values of P_{Hi}. We observe that we can support a throughput for the *Lo*-priority users of approximately 2.4 successes per slot. In comparison, for the same *Hi*-priority traffic parameters in the shared channel scheme, the maximum *Lo*-priority throughput is approximately 1.6 successes per slot.

Next we consider the design of contention policies that maintain these throughputs and QoS for the multiple priority class traffic.

4.4 Contention Policies

In the preceding sections, we have established techniques for providing service to mixed classes of traffic in the contention channel. For a given QoS level, these techniques result in maximum average (Poisson) transmission rates for *Lo*-priority traffic. To satisfy the transmission level constraints for *Lo*-priority traffic, we apply the analysis in Section

3.4.2: for estimated *Lo* backlog $\hat{X}_{Lo,t}$ at the beginning of frame t, the transmission policy, $p_{Lo,t}$, that achieves the desired average number of *Lo*-priority transmissions $C_{Lo}(\lambda_{Hi})$, is given by

$$p_{Lo,t} = \min\Big(C_{Lo}(\lambda_{Hi})/\hat{X}_{Lo,t}\,,\,1\Big). \tag{4.13}$$

Recall that $C_{Lo}(\lambda_{Hi})$ represents the transmission capacity of the channel for the *Lo*-priority users. It is a function of the *Hi*-priority arrival rate, λ_{Hi}, which is specified differently for the two channel use schemes in Section 4.3. Using the policy in (4.13), if a reasonably accurate estimate of $\hat{X}_{Lo,t}$ is available, the average throughput for the *Lo* users can be maximized without violating the QoS requirements of the *Hi* users.

Remark 4.1: Note that for large $\hat{X}_{Lo,t}$, the transmission process is well approximated as a Poisson with rate $C_{Lo}(\lambda_{Hi})$ [90]. In our experiments, a slightly different form of (4.13) has shown to offer the same or better performance. In the alternate form, the "instantaneous" *Lo* capacity, $C_{Lo}(X_{Hi,t})$, is used in the numerator of (4.13) instead of the average capacity, $C_{Lo}(\lambda_{Hi})$. There are two potential advantages to this modification. First, if $X_{Hi,t} < \lambda_{Hi}$, there is additional capacity available for use by *Lo* users, enhancing their throughput. Second, when $X_{Hi,t} > \lambda_{Hi}$, $p_{Lo,t}$ can be scaled back to allow additional capacity to serve the burst of *Hi* users at the highest possible QoS.

4.5 Backlog Estimation

To enable the use of the contention policies prescribed in the preceding sections, a technique is required for generating separate backlog estimates for the different priority classes. To do this, we first extend the framework in Section 3.5 to apply to homogeneous, delay-constrained traffic. We then apply the results to the mixed-priority traffic cases, in similar fashion to [26] for the SPR channel. In our model, we assume that the receiver is capable of determining how many users transmitted packets in a given frame, regardless of the number who succeeded or failed.

4.5.1 Homogeneous, Delay-Constrained Traffic

Similar to the method in Section 3.5, for estimated backlog $\hat{X}_t$ at the beginning of frame t, the backlog estimate in frame $t+1$ is given by

$$\hat{X}_{t+1} = \max\Big(\hat{X}_t + \lambda + c\big(f[t]\big), 0\Big), \tag{4.14}$$

where the evolution factor, $c\big(f[t]\big)$, is based on the feedback, $f[t] = (Y_t, S_t)$, on the outcome in frame t. The max operation is taken to ensure that the estimate is nonnegative since, as explained next, the evolution factor may cause the estimate to be less than 0. As in Section 3.5, Y_t and S_t are defined as the number of received transmissions and the number of successes in frame t, respectively. The evolution factor, however, now takes a different form than (3.16) for two reasons:

R1: As described in Section 4.2, users with delay constraints transmit with probability 1 until either success or expiration of their delay budget.

R2: A user remains in the system for up to a maximum of D frames, before either success or exiting the system due to failure and dropping of their call.

Rule *R1* results in an expected number of transmissions equal to the current backlog estimate, $E[Y_t] = \hat{X}_t$. The effect is that, if a frame occurs where no transmissions are received, i.e., $Y_t = 0$, or only successes occur, i.e., $Y_t = S_t$, then the backlog should be estimated to be empty except for the expected number of new arrivals (assuming the DFT policy). Rule *R2* means that users who succeed do so within D frames, and the information on their number of attempts can be conveyed to the AP to retroactively adjust its backlog estimate. Thus, the evolution factor takes the following form:

$$c\big(f[t]\big) = \begin{cases} -\hat{X}_t & \text{if } Y_t = \{0, S_t\}, \\ -\xi_t - S_t & \text{otherwise.} \end{cases} \tag{4.15}$$

With $V_t = Y_t - S_t$ representing the number of collisions in frame t, and V_t' representing the number of those collisions "claimed" by successful users in frames $[t+1, \ldots, t+D-1]$, we may define the correction factor as

$$\xi_t = V_{t-D+1} + \sum_{j=t-D+2}^{t-1} V_j'. \tag{4.16}$$

Thus, ξ_t counts the collisions in frames $[t-D+1, t-1]$ of users who have just succeeded in frame t, and any "unclaimed" collisions from frame $t-D+1$ and, therefore, represent users whose delay budgets have expired. Note that the definition of V_t also allows for the case when, in an RF implementation, users might individually fail because their packet BER is too high due to interference.

4.5.2 Heterogeneous Traffic

The backlog estimation technique of (4.14)-(4.16) requires some modification to extend to the mixed-priority traffic of Section 4.3. Since the contention policies are based on the backlog of each priority class, estimates must be maintained separately for each class (thus, any cost associated with fine-tuning the algorithm as in Remark 4.1 is already mandated). The backlog estimate for priority class i at the beginning of frame $t+1$ is given by

$$\hat{X}_{i,t+1} = \max\Big(\hat{X}_{i,t} + \lambda_i + c_i\big(f[t]\big), 0\Big). \tag{4.17}$$

The function of the evolution factor, $c_i\big(f[t]\big)$, is similar to that of Section 4.5.1, however it must be computed differently for the case of multiple priority class traffic.

To compute the evolution factor, $c_i\big(f[t]\big)$, we again represent the total number of failures in frame t as $V_t = Y_t - S_t$. In addition, we also define the quantity $H_t = \max\big(E[Y_t] - Y_t, 0\big)$ to be the number of "holes" as defined in [84]. Holes represent the "missing" expected transmissions given transmission policies $p_{i,t}$ and backlog estimates $\hat{X}_{i,t}$ in frame t. Although we assume the receiver detects Y_t, if there are transmissions that fail (i.e., $V_t > 0$), or there are holes (i.e., $H_t > 0$),[3] we do not assume it can distinguish the priority classes of the users who collided or did not transmit as expected. Therefore, we apply an approach analogous to the pseudo-Bayesian technique for mixed priority traffic in the SPR contention channel in [26], where holes and failures are proportionally distributed to the different classes based on the expected number of transmissions from each traffic class. To do this, we compute a proportion factor,

[3]Note that both collisions and holes may occur in the same frame.

$\hat{\beta}_{i,t} = p_{i,t}\hat{X}_{i,t} / \sum_k p_{k,t}\hat{X}_{k,t}$, representing the expected proportion of class i users from all users contending in frame t, based on the backlog estimates, $\hat{X}_{k,t}$, for each class. For the *Hi* and *Lo*-priority classes used here, these factors are given by

$$\hat{\beta}_{Hi,t} = \frac{p_{Hi,t}\hat{X}_{Hi,t}}{p_{Hi,t}\hat{X}_{Hi,t} + p_{Lo,t}\hat{X}_{Lo,t}}, \tag{4.18}$$

$$\hat{\beta}_{Lo}[n] = 1 - \hat{\beta}_{Hi}[n]. \tag{4.19}$$

The proportion factors are then applied to generate the estimated hole and failure measures for class i:

$$\hat{H}_{i,t} = \hat{\beta}_{i,t}H_t, \tag{4.20}$$

$$\hat{V}_{i,t} = \hat{\beta}_{i,t}V_t. \tag{4.21}$$

With these definitions, the general form of the evolution factor in (4.17) for priority class i becomes:

$$c_i\big(f[t]\big) = -S_{i,t} - \hat{H}_{i,t} + \hat{V}_{i,t} + \xi_{i,t}, \tag{4.22}$$

where $\xi_{i,t}$ is a correction factor for class i, similar to the correction factor in Section 4.5.1, and explained next in detail.

In (4.22), as in (4.16), the primary function of $\xi_{i,t}$ is to account for the *Hi* users' effect on both $\hat{X}_{Hi,t}$ and $\hat{X}_{Lo,t}$, resulting from Rules *R1* and *R2*. If a *Hi* user succeeds after having collided at least once, these extra collisions are removed from the backlog estimates $\hat{X}_{Hi,t}$ and $\hat{X}_{Lo,t}$ retroactively, where they were included in (4.22). Also, *Hi* users whose delay budgets have been exceeded and are exiting the system are removed from the *Hi* backlog estimate. Finally, since *Hi* users contend with $p_{Hi} = 1$, if $Y_t = \{0, S_t\}$, $\hat{X}_{Hi,t+1} = \lambda_{Hi}$, and any *Hi* failure measures for the previous $D - 1$ frames are retroactively attributed instead to the *Lo* backlog estimate. The result is that the

Table 4.1. γ_{max} and λ_{max} for $D = 3$ and $R_s = 0.99$

	Ideal R_X			DMMSE R_X		
M	γ_{max}	$\hat{\lambda}_{max}$	λ_{max}	γ_{max}	$\hat{\lambda}_{max}$	λ_{max}
1	6.19	4.9	3.46	5.40	4.28	3.14
3	18.56	14.7	12.12	16.2	12.9	11.1
5	30.94	24.5	21.9	27.31	21.6	20.1

correction factors for the *Hi* and *Lo*-priority classes take the following form:

$$\xi_{Hi,t} = \begin{cases} -\hat{X}_{Hi,t} + S_{Hi,t} + \hat{H}_{Hi,t} - \hat{V}_{Hi,t} & \text{if } Y_t = \{0, S_t\}, \\ -\hat{V}_{Hi,t-D+1} - \sum_{j=t-D+2}^{t-1} \hat{V}'_{Hi,j} & \text{otherwise,} \end{cases} \tag{4.23}$$

$$\xi_{Lo,t} = \begin{cases} \hat{X}_{Hi,t} - S_{Hi,t} - \hat{H}_{i,t} + \hat{V}_{Hi,t} & \text{if } Y_t = \{0, S_t\}, \\ \sum_{j=t-D+1}^{t-1} \frac{\hat{\beta}_{Lo,j}}{\hat{\beta}_{Hi,j}} \hat{V}'_{Hi,j} & \text{otherwise.} \end{cases} \tag{4.24}$$

4.6 Simulations

We now compare the results of our analysis with simulation results for our joint PHY/MAC design in Chapter 3 based on the DMMSE receiver Chapter 2. The simulations are performed with the same parameters as in the results in Chapter 3: the DS-CDMA system has processing gain $N = 11$, and 124 training symbols, values chosen to be comparable to 1-2 Mbps 802.11b WLANs [3] (however, the 802.11b format is not amenable to multipacket reception); the number of virtual subslots, $P = 31$, equals the number of distinct training sequences that transmitting users can choose from; each user experiences flat Rayleigh fading.

The first comparison we make is for the Poisson retransmission approximation of Section 4.2.1. For the comparison, the results for the probability of success given rate ν transmissions, shown in Fig. 3.7 of Chapter 3, allow us to find the estimates for γ_{max} and λ_{max} that upper bound the transmission and arrival processes, respectively, under

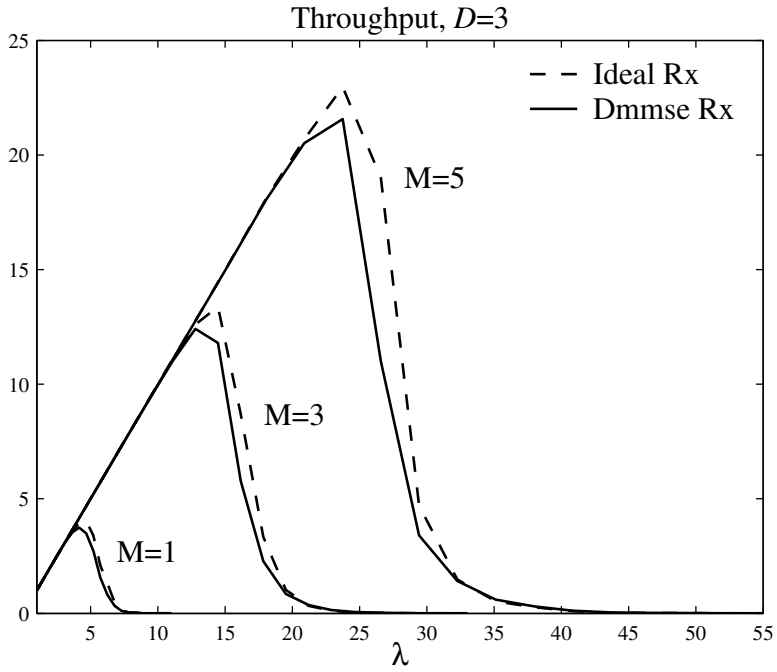

Figure 4.5. Throughput with delay budget $D = 3$ attempts and Poisson arrivals with rate λ, simulated for both ideal and DMMSE reception.

QoS requirement $\theta_\gamma < (1 - R_s)^{1/D}$. These values are summarized in Table 4.1, where $\hat{\lambda}_{max}$ is the predicted value of the actual λ_{max}, which is determined by simulations next.

For the first experiment of this section, we test the system under a simple model for the contention channel traffic. The combination of users desiring a handoff into the pseudocell and existing users generating new traffic is simulated as a single aggregate Poisson arrival process with rate λ per contention frame. The users are assumed to have a tight delay budget: they may contend in up to $D = 3$ consecutive contention frames before their connection attempt fails. The throughput for the system is shown in Fig. 4.5, where the sharper roll-off than in Fig. 3.6 is a result of users exiting the system and being counted as having failed in contention once the average delay grows just large enough that it exceeds the delay budget, $D = 3$. Fig. 4.6 shows the probability of failure to successfully contend before expiration of the delay budget. The actual maximum arrival rates while still satisfying the QoS constraint of $R_s = 0.99$ are listed in Table 4.1 as λ_{max}. These limits are tighter than those predicted by the analysis in Section 4.2, but

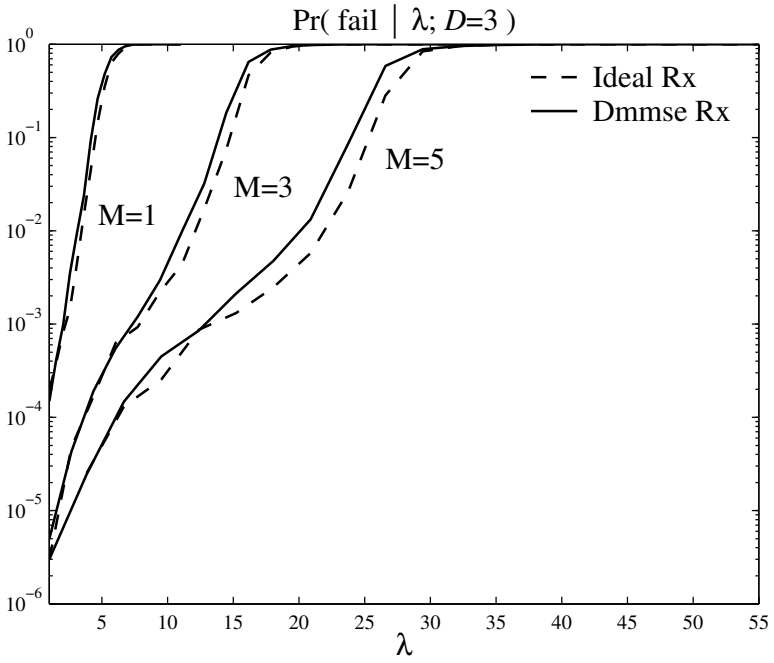

Figure 4.6. Probability of failure with delay budget $D = 3$ attempts and Poisson arrivals with rate λ, simulated for both ideal and DMMSE reception.

that is expected since the transmission process is not Poisson due to the correlation of failures across consecutive frames, where a detailed Markov analysis would be required to find more accurate bounds. The results do, however, show the system's generous capability to handle heavy loads of delay-constrained traffic. Also note the significant performance improvement from increasing the number of contention slots per contention frame (also evident from Figs. 3.4 and 3.5). This suggests that an access point designed such that it can dynamically increase or decrease the number of contention slots based on observed contention traffic loads [111] may serve a large, fluctuating population of users while preserving the maximum possible amount of wireless resources for user data.

In the second experiment, the analytical Markov model for the throughput for the case where only *Hi* users contend with $D = 3$ is compared with the actual system. Fig. 4.7 shows the analytical and simulated curves for the expiration rates in the QoS region of interest. The close match observed therein validates the analytical model. Further, the system is shown to be capable of supporting up to $\lambda_{Hi} = 3.5$ arrivals per

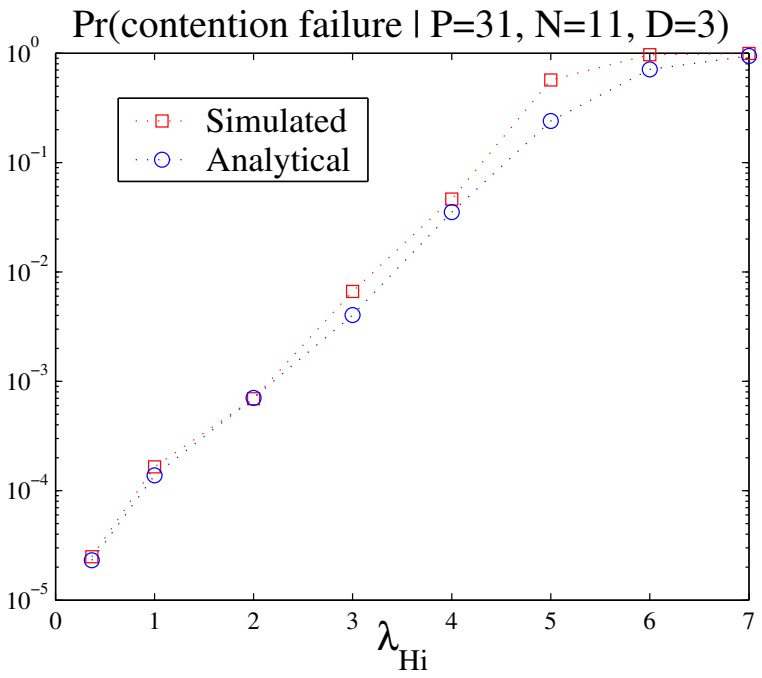

Figure 4.7. Probability of contention failure for a delay budget of $D = 3$ frames as a function of λ_{Hi}, with only *Hi* priority traffic contending.

frame at an expiration rate of $P_{exp} = 1 - R_s \leq 0.01$. As expected, this rate is between the achievable $\lambda_{Hi} = 0.311$ for the stringent constraint of $D = 1$ and the maximum achievable $\eta(\nu_{opt}) = 5.20$ from (4.3) without QoS.

Finally, Fig. 4.8 shows a trace over 1000 frames of the backlog estimates for *Lo* priority users compared with the actual backlog values. For these curves, the value of the estimate was used when generating the contention policy from (4.13). The close match of the curves show that the technique provides accurate estimates that are well suited for dynamic control of the contention policy.

4.7 Discussion

While earlier work had shown that the performance of ALOHA can be improved with MPR capability provided by multiuser detection, a novel feature of our work is the exploitation of collision size estimates provided by the receiver. The latter is what

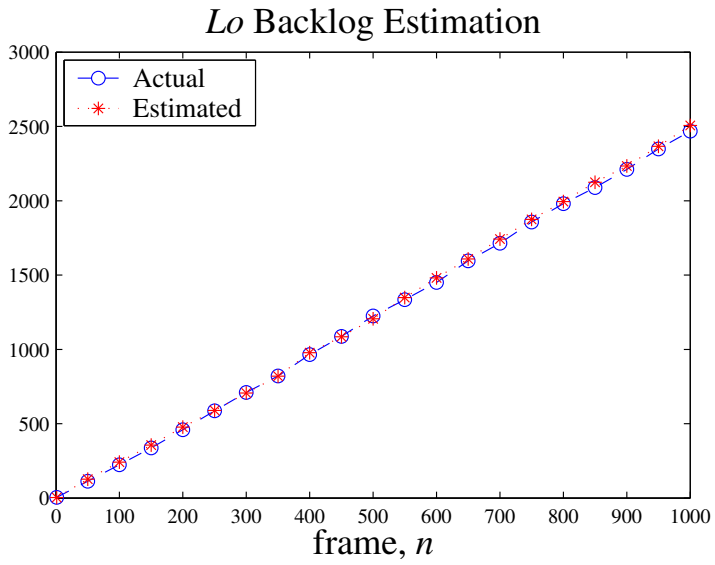

Figure 4.8. Estimated and actual *Lo* backlog with both *Hi* and *Lo* priority traffic contending.

enables us to obtain accurate backlog estimates for the different priority classes, which in turn are used to control the system to operate in the desired QoS/stability regime. Our results show the large performance gains to be obtained from cross-layer optimization. Moreover, the close match between our analytical model and the simulation model with the DMMSE receiver shows that our analysis is accurate and tractable, and offers a useful abstraction of the cross-layer interactions in the architecture.

Chapter 5

Conclusion

This chapter provides an overview of the research and results presented in this dissertation. In addition, some remarks on open questions and possible future directions of research are given.

5.1 Summary of Research

We have presented a new cross-layer design for efficient wireless medium access capable of multipacket reception and multiplicity feedback. The resulting contention channel can easily support a large volume of random access contention traffic, and meets the design objective of enabling rapid, mobile-centric handoffs in pseudocellular networks. The DMMSE-based receiver, when used with a suitably designed set of training sequences, is shown to enable the successful decoding of multiple simultaneous users without *a priori* carrier synchronization, channel information, or knowledge of the users' virtual subslots and spreading sequences. This permits users to choose spreading sequences randomly without coordinating with the AP, which is key to the multiple-access capability of our contention channel.

The DMMSE receiver [113] we use in our design provides adaptive linear receivers which are robust to rapid channel time variations. In this dissertation, we have refined

the theoretical results that show the DMMSE receiver to be a scalar multiple of the standard MMSE receiver, and thus inherit its robustness to multiple access interference and the near–far problem. We have also provided a detailed analysis of the convergence properties of adaptive, single-correlator implementations of the DMMSE receiver, which shows that careful choice of scaling factors leads to low-noise adaptations that avoid the zero-correlator. For frequency-selective fading, a new theorem has been given to characterize the structure of the DMMSE *eigenrake* receiver. The analysis leads to the new *selective eigenrake* receiver, which is capable of blindly detecting the multipaths from a desired user for implicit timing acquisition, diversity, and interference suppression. The capabilities of the DMMSE selective eigenrake for blindly detecting user transmissions and suppressing MAI are essential features that enable it to form the foundation of our joint PHY/MAC design.

For the new MPR contention channel based on the DMMSE receiver, the throughput is shown to be stabilized by using a dynamic control policy when knowledge (or estimates) of the system backlog is available. Further, when the arrival rate of new traffic is below a threshold, the system backlog is also stabilized. This stabilization requires accurate estimates of the backlog, which, instead of relying on (possibly less accurate) higher layer information [112], were shown to be possible using the capability of the PHY to detect the number of transmissions received, regardless of contention success or failure. The stable throughput thus attained is better than that of N narrowband ALOHA systems operating in parallel, where N is the processing gain of our system. This shows that, in addition to its superiority in supporting delay constraints, our system also provides better throughput by virtue of the statistical multiplexing enabled by the use of multiuser detection. Further, if information on the traffic characteristics is available, a means for dynamically expanding the capacity of the contention channel by increasing the number of contention slots per contention frame was demonstrated. The methods and analysis developed for the cross-layer design were extended to multiple priority classes. The design was shown to be capable of guaranteeing QoS requirements

for delay-constrained traffic in the presence of best-effort, delay-tolerant traffic through the use of dynamic stabilization policies. The backlog estimates for the separate traffic classes, necessary for the control policies, were provided by the extension of pseudo-Bayesian techniques to mixed traffic in the MPR contention channel with multiplicity feedback.

In the current state, the results of the work herein may be directly applied for enhancements of the 1 and 2 Mbps modes in 802.11-based WLANs [3]. In these modes, even with a DS-CDMA processing gain of $N = 11$, the channel remains an SPR channel. With moderate firmware or hardware modifications, the APs could be made DMMSE-capable, enabling an MPR contention channel and drastically reducing contention delay. At the higher data rates of the 802.11a standard [2], the large bandwidth provides for rapid access to resources even for delay-constrained traffic, however their PHYs are not capable of supporting high-mobility users experiencing the fast-fading channels induced by high-mobility. These high-mobility users would be supported with a DMMSE-based implementation. A particular challenge to the implementation of any multiuser detection scheme, particularly under near-far conditions, is the dynamic range of the required analog-to-digital converters. Advancements in integrated circuit technology, however, are rapidly making such tools available. In combination with techniques such as bandpass sampling, the sampling of wide bandwidths with strong dynamic range is becoming possible. On a broader scale, the capability to detect users in the presence of interfering transmissions at neighboring APs may reduce the frequency planning required for deployment of these WLANs in hot-spot areas. Even for high-rate systems with no spreading, spatial interference suppression for MIMO systems [17], [31], [32], [89], [110], may be an alternative application for DMMSE techniques. In either case, the robustness against interference would permit higher degrees of coverage overlap between APs and frequency reuse among them. The application of these techniques to full-system design, as opposed to the "retrofitting" of current 802.11-based WLANs, presents many open questions for future research.

5.2 Future Directions

While we have focused on a "reservation channel" for pseudocellular networks as the driving application, it is of future interest to explore similar PHY/MAC designs permitting communication with minimal coordination (e.g., in the presence of rapid topology variations) for the *data* channel in peer-to-peer ad hoc networks [25]. Even for pseudocellular networks, exploiting multiuser detection to allow multiple simultaneous data transmissions may be appropriate when the data rate for an individual connection is significantly smaller (e.g., for a voice connection) than the maximum available date rate on the shared channel. Further, the throughput on the data channel may be enhanced from that demonstrated in Chapter 3 through the application of collision resolution techniques such as in [14], [20], [27], [28], [64], [97]–[99], particularly when multiplicity feedback is available [30], [38], [74], or traffic priorities may be enforced [71].

Since the lack of structure of a pseudocellular wireless network precludes traditional centralized control mechanisms for channel planning and arbitration, cochannel interference mitigation must be achieved in a distributed or possibly inherent fashion. A further premium is placed on resources by the wireless interconnectivity between APs and the need for an efficient, high-speed wireless backhaul. The cross-layer design of the preceding sections presents several potential approaches for addressing these problems, from simply increasing the speed and efficiency of distributed dynamic resource allocation, to full uplink, downlink, and backhaul implementation.

5.2.1 Inter- and Intra-cell Interference Suppression

The limited availability of wireless spectrum necessitates re-using channel resources throughout networks and allocating channels for reuse requires that nearby links do not cause cochannel interference. Existing approaches to cochannel interference mitigation rely on orthogonal channels despite the requirement of significant network planning with fixed channel allocations and centralized control. More recent approaches to cen-

tralized [23] or distributed [18], [19], [70], [76] channel allocation have been proposed to increase channel reuse efficiency by allowing the system to dynamically adapt to temporal and spatial channel demand (load) by redistributing available channels to the high-demand regions. Under all of these schemes, the base stations, in coordination with neighboring cells, are the channel arbiters within their own cells and orthogonal channels are required. Additionally, when the base stations are not interconnected via wireline or are themselves mobile, valuable wireless resources must be expended on the induced overhead for channel reallocations [70].

Alternative techniques for cochannel interference mitigation, and thereby dynamic channel allocation, will be required from the perspective of the pseudocellular design concept. Deployment of a pseudocellular network makes any centralized control of channels and handoffs impractical. Further, the availability of a usable set of orthogonal channels can not be assured since there may be several pseudocells overlapping with each other. Instead, multiuser detection might allow for neighboring cells to use the same channels in a reliable fashion without the need for tight control (e.g., power control) of the mobiles. We conjecture that this is possible using a channel design based on multiuser detection, similar to that in Chapter 3 where the DS-CDMA spreading code and the fading channel in combination[1] provide diversity for a DMMSE receiver that suppresses the effects of severe near-far interference. This capability relaxes the need for tight feedback power control. However, it may be possible or even advantageous to use loose power control, for example, in situations at the perimeter of a network, or in areas where coverage is sparse, to bring more distant mobiles "into range". Further, if the impact on complexity can be minimized, the results of [39] suggest that overall system capacity could improved.

[1]Since the DMMSE tracks the ratio between successively faded user symbols, the channel rotations are implicitly tracked.

5.2.2 Wireless Backhaul

Since the APs themselves are not connected via wireline, a high capacity wireless backhaul must be available to transport data across the network. Several recent proposals [15], [21], [70], [88] share similar design goals, but lack one or more key aspects to yield the pseudocellular wireless backhaul. In [15], MMSE detection was shown to improve link reliability, however centralized control and planning was required. Distributed synthesis of backhaul links (using orthogonal channels) is considered in [70], [88] where highly mobile backhaul nodes induce continual reconfiguration of network topologies. While the technique in [70] focused strictly on eliminating cochannel interference, it was shown in [10], [88] that inclusion of capacity, QoS and MAC constraints on topological synthesis could improve operational efficiency. Routing, link scheduling and power control were jointly optimized in [21], and the proposed methods explicitly incorporated multiaccess interference. However, optimal results from [21] require significant centralized computation and network-wide scheduling.

Given the capacity increase in the contention channel in Chapter 3, and in the data channel that may result from results on the discussion in Section 5.2.1, we conjecture that backhaul links could be dynamically established between APs using the same contention and data channels as the mobiles. We note that the benefits of a full MUD implementation over a high bandwidth-efficient scheme with orthogonal channels are not obvious, and discussion of the efficiency of differing systems with and without MUD have been ongoing [42], [43], [55], [95]. In general, these works consider primarily QoS based on average received SINR levels, while not satisfactorily addressing the issue of traffic dynamics and resource scheduling to support delay deadlines. We expect, however, that, minimally, the utility of MUD would still increase the efficiency of distributed resource allocation in the control/contention channels even when using orthogonal data channels. Effectively, the APs themselves would form an ad-hoc backhaul network with routing performed by application of existing techniques such as in [73] with MUD-enhanced link scheduling, or joint routing and scheduling performed using a suboptimal, but dis-

tributed, adaptation of [21].

Bibliography

[1] "Digital cellular telecommunications system (phase 2+); general packet radio service (GPRS); service description; stage 2," European Telecommunications Standards Institute), Tech. Rep. ETSI EN 301 344 v7.4.1 (2000-09), 2000, available for download at www.etsi.org.

[2] "Supplement to IEEE standard for information technology – telecommunications and information exchange between systems – local and metropolitan area networks – specific requirements. part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications; high-speed physical layer in the 5 GHz band," IEEE, Tech. Rep. IEEE Std. 802.11a-1999, 2000.

[3] "Supplement to IEEE standard for information technology - telecommunications and information exchange between systems - local and metropolitan area networks - specific requirements. part 11: wireless LAN medium access control (MAC) and physical layer (PHY)," IEEE, Tech. Rep. IEEE Std. 802.11b-1999, 2000.

[4] "Universal mobile telecommunications system (UMTS); physical channels and mapping of transport channels onto physical channels (FDD)," European Telecommunications Standards Institute, Tech. Rep. ETSI TS 125 211 v3.3.0 (2000-06), 2000, available for download at www.etsi.org.

[5] M. Abdulrahman, A. U. H. Sheikh, and D. D. Falconer, "Decision feedback equalization for CDMA in indoor wireless communication," *IEEE J. Select. Areas Commun.*, vol. 12, no. 4, pp. 698–704, May 1994.

[6] N. Abramson, "The ALOHA system - another alternative for computer communications," in *Proc. Fall 1970 AFIPS Computer Conf.*, vol. 37, Montvale, NJ, 1970, pp. 281–285.

[7] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Trans. Commun.*, vol. 25, no. 1, pp. 117–128, Jan. 1977.

[8] N. Abramson, "Development of the ALOHANET," *IEEE Trans. Inform. Theory*, vol. 31, no. 2, pp. 119–123, Mar. 1985.

[9] N. Abramson, "Multiple access in wireless digital networks," *Proc. IEEE*, vol. 82, no. 9, pp. 1360–1370, Sept. 1994.

[10] A. Behzad and I. Rubin, "Multiple access protocol for power-controlled wireless access nets," *IEEE Trans. Mobile Comput.*, vol. 3, no. 4, pp. 307–316, Oct.–Dec. 2004.

[11] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1992.

[12] D. G. Brennan, "Linear diversity combining techniques," *Proc. IRE*, vol. 47, pp. 1075–1102, June 1959.

[13] D. G. Brennan, "Linear diversity combining techniques," *Proc. IEEE*, vol. 91, no. 2, pp. 331–356, Feb. 2003.

[14] J. I. Capetanakis, "Tree algorithms for packet broadcast channels," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 505–515, Sept. 1979.

[15] K. Chawla, X. Qiu, and M. V. Clark, "Design of a wireless backhaul network for microcells," in *Proc. IEEE Wireless Commun. and Networking Conf. (WCNC'99)*, vol. 1, New Orleans, LA, Sept. 21–24, 1999, pp. 428–432.

[16] R.-R. Chen, R. Koetter, U. Madhow, and D. Agrawal, "Joint noncoherent demodulation and decoding for the block fading channel: A practical framework for approaching shannon capacity," *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1676–1689, Oct. 2003.

[17] C.-N. Chuah, D. N. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[18] L. J. Cimini, Jr. and G. J. Foschini, "Distributed algorithms for dynamic channel allocation in microcellular systems," in *Proc. 42nd IEEE Veh. Technol. Conf. (VTC'92)*, vol. 2, Denver, CO, May 10–13, 1992, pp. 641–644.

[19] L. J. Cimini, Jr., G. J. Foschini, C.-L. I, and Z. Miljanic, "Call blocking performance of distributed algorithms for dynamic channel allocation in microcells," *IEEE Trans. Commun.*, vol. 42, no. 8, pp. 2600–2607, Aug. 1994.

[20] R. L. Cruz and B. Hajek, "A new upper bound to the throughput of a multi-access broadcast channel," *IEEE Trans. Inform. Theory*, vol. 28, no. 3, pp. 402–405, May 1982.

[21] R. L. Cruz and A. V. Santhanam, "Optimal routing, link scheduling and power control in multi-hop wireless networks," in *Proc. 22nd Joint Conf. of IEEE Computer and Commun. Societies (INFOCOM'03)*, vol. 1, San Francisco, CA, Mar. 30–Apr. 3, 2003, pp. 702–711.

[22] D. H. Davis and S. A. Gronemeyer, "Performance of slotted ALOHA random access with delay capture and randomized time of arrival," *IEEE Trans. Commun.*, vol. 28, no. 5, pp. 703–710, May 1980.

[23] E. Del Re, R. Fantacci, and G. Giambene, "Handover and dynamic channel allocation techniques in mobile cellular networks," *IEEE Trans. Veh. Technol.*, vol. 44, no. 2, pp. 229–237, May 1995.

[24] D. Divsalar and M. K. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 38, no. 3, pp. 300–308, Mar. 1990.

[25] A. Ephremides and B. Hajek, "Information theory and communication networks: An unconsummated union," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2416–2434, Oct. 1998.

[26] J.-F. Frigon and V. C. M. Leung, "A pseduo-bayesian ALOHA algorithm with mixed priorities," *Wireless Networks*, vol. 7, no. 1, pp. 55–63, Jan. 2001.

[27] R. G. Gallager, "Conflict resolution in random access broadcast networks," in *Proc. AFOSR Workshop in Commun. Theory and Applic.*, Provincetown, MA, Sept. 17–20, 1978, pp. 74–76.

[28] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory*, vol. 31, no. 2, pp. 124–142, Mar. 1985.

[29] A. Ganz, C. M. Krishna, D. Tang, and Z. J. Haas, "On optimal design of multitier wireless cellular systems," *IEEE Commun. Mag.*, vol. 35, no. 2, pp. 88–93, Feb. 1997.

[30] L. Georgiadis and P. Papantoni-Kazakos, "A collision resolution protocol for random access channels with energy detectors," *IEEE Trans. Commun.*, vol. 30, no. 11, pp. 2413–2420, Nov. 1982.

[31] D. Gesbert, H. Bölcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[32] D. Gesbert, M. Shafi, D.-S. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.

[33] S. Ghez, S. Verdú, and S. C. Schwartz, "Stability properties of slotted aloha with multipacket reception capability," *IEEE Trans. Automat. Contr.*, vol. 33, no. 7, pp. 640–648, July 1988.

[34] S. Ghez, S. Verdú, and S. C. Schwartz, "Optimal decentralized control in the random access multipacket channel," *IEEE Trans. Automat. Contr.*, vol. 34, no. 11, pp. 1153–1163, Nov. 1989.

[35] R. Gold, "Optimal binary sequences for spread spectrum multiplexing," *IEEE Trans. Inform. Theory*, vol. 13, no. 4, pp. 619–621, Oct. 1967.

[36] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins University Press, 1996.

[37] D. J. Goodman and A. A. M. Saleh, "The near/far effect in local ALOHA radio communications," *IEEE Trans. Veh. Technol.*, vol. 36, no. 1, pp. 19–27, Feb. 1987.

[38] E. Gulko, "Tree-based multiaccess protocols where collision mulitiplicities are known," *IEEE Trans. Commun.*, vol. 33, no. 9, pp. 999–1001, Sept. 1985.

[39] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[40] B. Hajek, A. Krishna, and R. O. LaMaire, "On the capture probability for a large number of stations," *IEEE Trans. Commun.*, vol. 45, no. 2, pp. 254–260, 1997.

[41] B. Hajek and T. van Loon, "Decentralized dynamic control of a multiaccess broadcast channel," *IEEE Trans. Automat. Contr.*, vol. 27, no. 3, pp. 559–569, 1982.

[42] S. V. Hanly and D. N. Tse, "The multi-access fading channel: Shannon and delay limited capacities," in *Proc. 33rd Annual Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, Oct. 4–6, 1995, pp. 786–795.

[43] S. V. Hanly and D. N. C. Tse, "Multiaccess fading channels–Part II: Delay-limited capacities," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2816–2831, Nov. 1998.

[44] S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1991.

[45] M. Honig, U. Madhow, and S. Verdú, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, no. 4, pp. 944–960, July 1995.

[46] M. L. Honig, S. L. Miller, M. J. Shensa, and L. B. Milstein, "Performance of adaptive linear interference suppression in the presence of dynamic fading," *IEEE Trans. Commun.*, vol. 49, no. 4, pp. 635–645, Apr. 2001.

[47] M. L. Honig, M. J. Shensa, S. L. Miller, and L. B. Milstein, "Performance of adaptive linear interference suppression for DS-CDMA in the presence of flat Rayleigh fading," in *Proc. IEEE Veh. Technol. Conf. (VTC'97)*, vol. 3, Phoenix, AZ, May 4–7, 1997, pp. 2191–2195.

[48] L.-R. Hu and S. S. Rappaport, "Personal communication systems using multiple hierarchical cellular overlays," *IEEE J. Select. Areas Commun.*, vol. 13, no. 2, pp. 406–415, Feb. 1995.

[49] C.-L. I, L. J. Greenstein, and R. D. Gitlin, "A microcell/macrocell cellular architecture for low- and high-mobility wireless users," *IEEE J. Select. Areas Commun.*, vol. 11, no. 6, pp. 885–891, Aug. 1993.

[50] B. Jabbari and W. F. Fuhrmann, "Teletraffic modeling and analysis of flexible hierarchical cellular networks with speed-sensitive handoff strategy," *IEEE J. Select. Areas Commun.*, vol. 15, no. 8, pp. 1539–1548, Oct. 1997.

[51] W. C. Jakes, *Microwave Mobile Communications*. New York: IEEE Press, 1994.

[52] G. J. M. Janssen, "Receiver structure for simultaneous reception of two BPSK modulated cochannel signals," *Electron. Lett.*, vol. 29, no. 12, pp. 1095–1097, June 1993.

[53] L. Kleinrock and S. S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation," *IEEE Trans. Commun.*, vol. 23, no. 4, pp. 410–423, Apr. 1975.

[54] L. Kleinrock and F. A. Tobagi, "Packet switching in radio channels: Part I–carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE Trans. Commun.*, vol. 23, no. 12, pp. 1400–1416, Dec. 1975.

[55] S. V. Krishnamurthy, A. S. Acampora, and M. Zorzi, "On the radio capacity of TDMA and CDMA for broadband wireless packet communications," *IEEE Trans. Veh. Technol.*, vol. 52, no. 1, pp. 60–70, Jan. 2003.

[56] S. S. Lam and L. Kleinrock, "Packet switching in a multiaccess broadcast channel: Dynamic control procedures," *IEEE Trans. Commun.*, vol. 23, no. 9, pp. 891–904, Sept. 1975.

[57] M. Latva-aho and M. J. Juntti, "LMMSE detection for DS-CDMA systems in fading channels," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 194–199, Feb. 2000.

[58] C. T. Lau and C. Leung, "Capture models for mobile packet radio networks," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 917–925, May 1992.

[59] U. Madhow, "Blind adaptive interference suppression for the near-far resistant acquisition and demodulation of direct-sequence CDMA signals," *IEEE Trans. Signal Processing*, vol. 45, no. 1, pp. 124–136, Jan. 1997.

[60] U. Madhow, "Blind adaptive interference suppression for direct-sequence CDMA," *Proc. IEEE*, vol. 86, no. 10, pp. 2049–2069, Oct. 1998.

[61] U. Madhow, K. Bruvold, and L. J. Zhu, "Differential MMSE: A framework for robust adaptive interference suppression for DS-CDMA over fading channels," *IEEE Trans. Commun.*, vol. 53, no. 8, pp. 1377–1390, Aug. 2005.

[62] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3178–3188, Dec. 1994.

[63] U. Madhow, L. J. Zhu, and L. Galup, "Differential MMSE: new adaptive algorithms for equalization, interference suppression, and beamforming," in *Proc. 32nd Asilomar Conf. on Signals, Systems and Computers (Asilomar'98)*, Pacific Grove, CA, Nov. 1–4, 1998, pp. 640–644.

[64] N. Mehravari, "Random-access communication with multiple reception," *IEEE Trans. Inform. Theory*, vol. 36, no. 3, pp. 614–622, May 1990.

[65] V. A. Mikhailov and B. S. Tsybakov, "Upper bound for the capacity of a random multiple access system," *Problem Inform. Transm.*, vol. 17, no. 1, pp. 90–95, Jan.–Mar. 1981.

[66] S. L. Miller, "An adaptive direct-sequence code-division multiple-access receiver for multiuser interference rejection," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 1746–1755, Feb./Mar./Apr. 1995.

[67] S. L. Miller, M. L. Honig, and L. B. Milstein, "Performance analysis of MMSE receivers for DS–CDMA in frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 48, no. 11, pp. 1919–1929, Nov. 2000.

[68] R. Mudumbai, G. Barriac, and U. Madhow, "Optimizing medium access control for rapid handoffs in pseudocellular networks," in *Proc. 60th IEEE Veh. Technol. Conf. (VTC Fall'04)*, Los Angeles, CA, Sept. 26–29, 2004.

[69] V. Naware and L. Tong, "Stability of slotted ALOHA with spatial diversity," in *Proc. 2003 IEEE Internat. Conf. on Commun. (ICC'03)*, vol. 1, Anchorage, AK, May 11–15, 2003, pp. 463–467.

[70] S. Nesargi and R. Prakash, "Distributed wireless channel allocation in networks with mobile base stations," *IEEE Trans. Veh. Technol.*, vol. 51, no. 6, pp. 1407–1421, Nov. 2002.

[71] T. Papantoni-Kazakos, N. B. Likhanov, and B. S. Tsybakov, "A protocol for random multiple access of packets with mixed priorities in wireless networks," *IEEE J. Select. Areas Commun.*, vol. 13, no. 7, pp. 1324–1331, Sept. 1995.

[72] S. Parkvall, E. G. Ström, L. B. Milstein, and B. E. Ottersten, "Asynchronous near–far resistant DS–CDMA receivers without a priori synchronization," *IEEE Trans. Commun.*, vol. 47, no. 1, pp. 78–88, Jan. 1999.

[73] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," in *Proc. 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'99)*, New Orleans, LA, Feb. 25–26, 1999, pp. 90–100.

[74] N. Pippenger, "Bounds on the perfomance of protocols for a multiple-access broadcast channel," *IEEE Trans. Inform. Theory*, vol. 27, no. 2, pp. 145–151, Mar. 1981.

[75] M. F. Pop and N. C. Beaulieu, "Limitations of sum-of-sinusoids fading channel simulators," *IEEE Trans. Commun.*, vol. 49, no. 4, pp. 699–708, Apr. 2001.

[76] R. Prakash, N. G. Shivaratri, and M. Singhal, "Distributed dynamic fault-tolerant channel allocation for cellular networks," *IEEE Trans. Veh. Technol.*, vol. 48, no. 6, pp. 1874–1888, Nov. 1999.

[77] M. B. Pursley, "The role of spread spectrum in packet radio networks," *Proc. IEEE*, vol. 75, no. 1, pp. 116–134, Jan. 1987.

[78] M. B. Pursley and D. V. Sarwate, "Evaluation of correlation parameters for periodic sequences," *IEEE Trans. Inform. Theory*, vol. 23, no. 4, pp. 508–513, July 1977.

[79] B. Ramamurthi, A. A. M. Saleh, and D. J. Goodman, "Perfect-capture ALOHA for local radio communications," *IEEE J. Select. Areas Commun.*, vol. 5, no. 5, pp. 806–814, June 1987.

[80] P. B. Rapajic, "Performance analysis of slotted ALOHA/CDMA system with adaptive MMSE receivers," *IEICE Trans. Fund. Electron. Commun. Comput. Sci.*, vol. E80-A, no. 12, pp. 2485–2492, Dec. 1997.

[81] P. B. Rapajic and B. S. Vucetic, "Adaptive receiver structures for asynchronous CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 12, no. 4, pp. 685–697, May 1994.

[82] S. S. Rappaport and L.-R. Hu, "Microcellular communication systems with hierarchical macrocell overlays: Traffic performance models and analysis," *Proc. IEEE*, vol. 82, no. 9, pp. 1383–1397, Sept. 1994.

[83] D. Raychaudhuri, "Performance analysis of random access packet-switched code division multiple access systems," *IEEE Trans. Commun.*, vol. 29, no. 6, pp. 895–901, June 1981.

[84] R. L. Rivest, "Network control by bayesian broadcast," *IEEE Trans. Inform. Theory*, vol. 33, no. 3, pp. 323–328, May 1987.

[85] L. G. Roberts, "ALOHA packet system with and without slots and capture," in *ASS Note 8*, Stanford Res. Inst., Adv. Res. Projects Agency, Network Inform. Ctr., Stanford, CA, Jan. 1972.

[86] W. A. Rosenkrantz and D. Towsley, "On the instability of the slotted ALOHA multiaccess algorithm," *IEEE Trans. Automat. Contr.*, vol. 28, no. 10, pp. 994–996, Oct. 1983.

[87] S. Roy, "Subspace blind adaptive detection for multiuser CDMA," *IEEE Trans. Commun.*, vol. 48, no. 1, pp. 169–175, Jan. 2000.

[88] I. Rubin, A. Behzad, R. Zhang, H. Luo, and E. Caballero, "TBONE: A mobile-backbone protocol for ad hoc wireless networks," in *Proc. IEEE Aerospace Conf.*, vol. 6, Big Sky, MT, Mar. 9–16, 2002, pp. 2727–2740.

[89] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2198–2206, Dec. 2001.

[90] H. Stark and J. W. Woods, *Probability, Random Processes, And Estimation Theory For Engineers*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1994.

[91] H.-J. Su and E. Geraniotis, "Low-complexity joint channel estimation and decoding for pilot symbol-assisted modulation and multiple differential detection systems with correlated rayleigh fading," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 249–261, Feb. 2002.

[92] H. Takagi and L. Kleinrock, "Throughput analysis for persistant CSMA systems," *IEEE Trans. Commun.*, vol. 33, no. 7, pp. 627–638, July 1985.

[93] L. Tong, Q. Zhao, and G. Mergen, "Multipacket reception in random access wireless networks: From signal processing to optimal medium access control," *IEEE Commun. Mag.*, vol. 39, no. 11, pp. 108–112, Nov. 2001.

[94] N. D. Tripathi, J. H. Reed, and H. F. VanLandingham, "Handoff in cellular systems," *IEEE Personal Commun. Mag.*, vol. 5, iss. 6, pp. 26–37, Dec. 1998.

[95] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 641–657, Mar. 1999.

[96] J. N. Tsitsiklis, "Analysis of a multiaccess control scheme," *IEEE Trans. Automat. Contr.*, vol. 32, no. 11, pp. 1017–1020, Nov. 1987.

[97] B. S. Tsybakov, "Survey of USSR contributions to random multiple-access communications," *IEEE Trans. Inform. Theory*, vol. 31, no. 2, pp. 143–165, Mar. 1985.

[98] B. S. Tsybakov, "Packet multiple access for channel with binary feedback, capture, and multiple reception," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1073–1085, June 2004.

[99] B. S. Tsybakov and V. A. Mikhailov, "Random multiple packet access: Part-and-try algorithm," *Problem Inform. Transm.*, vol. 16, no. 4, pp. 65–79, Oct.–Dec. 1980.

[100] B. S. Tsybakov, V. A. Mikhailov, and N. B. Likhanov, "Bounds for packet transmission rate in a random-multiple-access system," *Problem Inform. Transm.*, vol. 19, no. 1, pp. 61–81, Jan.–Mar. 1983.

[101] S. Verdú, *Multiuser Detection.* New York: Cambridge University Press, 1998.

[102] L.-C. Wang, G. L. Stüber, and C.-T. Lea, "Architecture design, frequency planning, and performance analysis for microcell/macrocell overlaying system," *IEEE Trans. Veh. Technol.*, vol. 46, no. 4, pp. 836–848, Nov. 1997.

[103] X. Wang and H. V. Poor, "Blind adaptive multiuser detection in multipath CDMA channels based on subspace tracking," *IEEE Trans. Signal Processing*, vol. 46, no. 11, pp. 3030–3044, Nov. 1998.

[104] X. Wang and H. V. Poor, "Blind multiuser detection: A subspace approach," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 677–690, Mar. 1998.

[105] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, July 1999.

[106] J. Ward and R. T. Compton, Jr., "High throughput slotted ALOHA packet radio networks with adaptive arrays," *IEEE Trans. Commun.*, vol. 41, no. 3, pp. 460–470, Mar. 1993.

[107] D. Warrier and U. Madhow, "Noncoherent communication in space and time," in *Proc. 33rd Conf. on Inform. Science and Systems (CISS'99)*, Johns Hopkins University, Baltimore, MD, Mar. 17–19, 1999.

[108] C. Xiao, Y. R. Zheng, and N. C. Beaulieu, "Statistical simulation models for rayleigh and rician fading," in *Proc. 2003 IEEE Internat. Conf. on Commun. (ICC'03)*, vol. 5, Anchorage, AK, May 11–15, 2003, pp. 3524–3529.

[109] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[110] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input–multiple-output (MIMO) transmission systems," *IEEE Trans. Commun.*, vol. 42, no. 12, pp. 3221–3231, Dec. 1994.

[111] I.-S. Yoon and B. G. Lee, "DDR: A distributed dynamic reservation scheme that supports mobility in wireles multimedia commnunications," *IEEE J. Select. Areas Commun.*, vol. 19, no. 11, pp. 699–708, Nov. 2001.

[112] Q. Zhao and L. Tong, "A multiqueue service room MAC protocol for wireless networks with multipacket reception," *IEEE/ACM Trans. Networking*, vol. 11, no. 1, pp. 125–137, Feb. 2003.

[113] L. J. Zhu, "Differential MMSE: A new approach for adaptive interference suppression over time-varying channels," Ph.D. dissertation, Univ. of Illinois, Urbana-Champaign, Urbana, IL, 1999.

[114] L. J. Zhu and U. Madhow, "Adaptive interference suppression for direct sequence CDMA over severely time-varying channels," in *Proc. IEEE Global Telecom. Conf. (GLOBECOM'97)*, vol. 2, Phoenix, AZ, Nov. 3–8, 1997, pp. 917–922.

[115] L. J. Zhu and U. Madhow, "Adaptive interference suppression for DS-CDMA over a Rayleigh fading channel," in *Proc. 31st Conf. on Inform. Science and Systems (CISS'97)*, Johns Hopkins University, Baltimore, MD, Mar. 19–21, 1997.

[116] M. Zorzi, "On the capture performance of smart antennas in a multicellular environment," *IEEE Trans. Commun.*, vol. 50, no. 4, pp. 536–539, Apr. 2002.

[117] M. Zorzi and R. R. Rao, "Capture and retransmission control in mobile radio," *IEEE J. Select. Areas Commun.*, vol. 12, no. 8, pp. 1289–1298, Oct. 1994.