Increased Capacity for TDMA Cellular Systems Through Joint Detection with Diversity Arrays

by

Stephen J. Grant

M.A.Sc., Simon Fraser University, 1996

B.A.Sc., University of British Columbia, 1993

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the School
of
Engineering Science

SIMON FRASER UNIVERSITY
August 2000

APPROVAL

Name: Stephen J. Grant

Degree: Doctor of Philosophy

Title of thesis: Increased Capacity for TDMA Cellular Systems Through Joint Detection with Diversity Arrays

Examining Committee: Dr. Mehrdad Saif
Chair

Dr. James K. Cavers
Senior Supervisor

Dr. Paul Ho
Supervisor

Dr. Jacques Vaisey
Supervisor

Dr. Steve Hardy
Internal Examiner

Dr. Peter McLane
External Examiner, Queen's University

Date Approved: _____

Abstract

Joint detection based on exploiting differences among the channels employed by several users allows a receiver to distinguish cochannel signals without reliance on spectrum spreading. This thesis considers the use of joint detection with diversity antenna arrays in enhancing the capacity of time division multiple access (TDMA) cellular systems by allowing several users in the same cell to share the same time/frequency slot.

For the uplink, a fully-analytical expression for the union bound on average symbol-error rate is provided for joint detection with an arbitrary number of users M and diversity antennas L in a flat-fading environment with symbol-synchronous users, and both perfect and imperfect channel state information (CSI). With joint detection, many more users than the number of antennas may be supported — all enjoying L-fold diversity — with a small degradation in performance with each additional user. This fundamental result, not observed previously, is in stark contrast to classical minimum mean squared error (MMSE) combining of antennas where the number of users is limited to the number of antennas, and the diversity order is only $L - M + 1$ for each user.

For the downlink, a simple structure is devised for supporting multiple users in the same bandwidth as a single user through the use of standard coding and interleaving combined with a transmit antenna array at the base station. An analytical expression for the average bit-error rate is developed based on soft decision joint decoding at the mobile receivers. Unusual behaviour is demonstrated in terms of diversity order: as the number of antennas increases due to an increasing number of users, the diversity order actually decreases. Even with the loss in diversity order, it is possible to obtain good performance for all users with moderate computational load.

Estimation of the channel gains of the multiple cochannel users is a key requirement of joint detection. In this thesis, a pilot-based MMSE technique is developed for jointly

estimating the gains. Two key contributions are made: firstly, the estimator allows for time variation of the channels within and between training sequences, an essential feature in a multiuser environment even at moderate fading rates. Secondly, it addresses the design of training sequences for the multiple users.

A fully-analytical and computationally straightforward technique is developed for the calculation of outage probability in the presence of fading, shadowing, and path loss. Using the technique, system uplink capacity is quantified, while accounting for the increase in interference due to the reuse of channels within each cochannel cell. With joint detection, capacity may be significantly extended: with four antennas and M cochannel users per cell, an M-fold increase may be obtained when compared to a conventional single-user TDMA system with two antennas. In contrast, with MMSE combining, only a modest increase in capacity is demonstrated, as well as a hard limit on the number of user per cell beyond which system capacity actually decreases due to losses in diversity order. This result is surprising and significant, since MMSE combining is frequently advanced as a method to increase capacity.

Acknowledgements

I would like to express my gratitude to Jim Cavers for being the finest mentor I could have hoped to have; I am grateful for all that you have taught me. I would also like to thank my immediate and extended families for their constant interest in my studies, and for putting up with me being a student for so long. Thank-you to to the Natural Sciences and Engineering Research Council of Canada, as well as Ericsson Inc., Research Triangle Park, North Carolina for their support of this work.

Dedication

To Irma for her constant encouragement and unwavering support through our long, and sometimes difficult journey. I thank you sincerely.

Contents

List of Tables

List of Figures

List of Abbreviations

AOA	Angle of arrival
BER	Bit-error rate
BPSK	Binary phase shift keying
CCDF	Complementary cumulative distribution function
CCI	Cochannel interference
CDMA	Code division multiple access
CIR	Channel impulse response
CSI	Channel state information
dB	Decibel
i.i.d.	Independent identically distributed
ISI	Intersymbol interference
ISR	Interference-to-signal ratio
JD	Joint detection
LMS	Least mean squares
LS	Least squares
MAP	Maximum a posteriori
MMSE	Minimum mean squared error
MRC	Maximal ratio combining
MSE	Mean squared error
PDF	Probability density function
PDP	Power-delay profile
PSK	Phase shift keying
RMS	Root mean square
RV	Random variable
RWC	Reuse within cell

SER	Symbol-error rate
SINR	Signal-to-interference-plus-noise ratio
SNR	Signal-to-noise ratio
TDMA	Time division multiple access
VA	Viterbi algorithm
WSSUS	Wide sense stationary uncorrelated scattering
ZF	Zero forcing

Chapter 1

Motivation

Currently, the wireless communications industry is experiencing explosive growth resulting in an ever increasing demand for wireless communication services. A higher demand for services, in turn, calls for higher system capacities. Capacity may always be increased directly by allocating more bandwidth in the electromagnetic spectrum; however, since bandwidth is a scarce resource, this may not be feasible. Thus, to increase capacity, the existing spectrum allocation must be used more efficiently.

Recently, it has been proposed to improve the spectrum efficiency of existing time division multiple access (TDMA) systems by allowing more than one user within a cell to share the same time/frequency slot — a concept known as frequency reuse within cell (RWC). This dictates the use of either multiuser detection or interference suppression techniques that are capable of distinguishing the intracell cochannel signals. Furthermore, it suggests the use of antenna arrays which have been used for many years in the sonar and radar communities for suppressing cochannel interference.

Multiuser detection is considered to be an effective means for reducing interference (and thus increasing system capacity) in code division multiple access (CDMA) systems where all intracell users share the same frequency and transmit continuously. However, the computational complexity of the optimal detector increases exponentially with the number of intracell users rendering it impractical in typical CDMA systems. In contrast, when multiuser — or joint — detection is applied to a TDMA system, the computational complexity of the optimal detector is reasonable, since it is likely that a much smaller number of users will share the same time/frequency slot.

This thesis focuses on extending the capacity of TDMA systems by allowing several

intracell users to share the same time/frequency slot and employing a diversity antenna array at the base station. A joint detection technique is considered that distinguishes the cochannel signals by exploiting differences among the users' channel gains, without relying on spectrum spreading as is the case with multiuser detection for CDMA.

In considering joint detection with diversity arrays applied to TDMA systems, several important questions arise.

- In the uplink, what is the form of the optimal detector, and how does it perform in a fading environment with antenna diversity and imperfect channel estimates?
- In contrast to the uplink, in the downlink the diversity array is located at the transmitter which does not have knowledge of the downlink channel gains with which to perform array processing. Furthermore, the mobile receivers normally have only one antenna at which the multiple cochannel transmit signals combine. Given this more complicated scenario, how should the multiple transmit antennas be used to achieve a capacity increase, as well as provide diversity for the multiple users?
- Given that channel knowledge is the key to distinguishing the cochannel signals with joint detection, how does one obtain accurate channel estimates in a multiuser environment?
- In a multicell system, the cochannel cells also have several users per slot; thus, the overall level of interference increases with RWC, partially diluting the desired capacity increase. What is the actual increase in system capacity, and how may it be further enhanced by the use of a diversity antenna array?

These four questions provide motivation for this thesis. The questions are addressed in Chapters 3–6 respectively, while Chapter 2 provides background to set the topic of joint detection with diversity arrays in context.

Chapter 2

Background

Conventional TDMA systems are based on keeping simultaneous intracell users orthogonal in both time and frequency, while treating user signals received from other cochannel cells as noise. The result is a hard capacity limit, in contrast to CDMA systems, which exhibit a graceful degradation with additional users. The hard capacity limit is defined with reference to Fig. 2.1. In this diagram, each cell within a cluster of C cells has a different set of channels represented by the numbers $1, 2, \cdots, C$. Each channel is one time/frequency slot, and each user requires two channels — one for the uplink and one for the downlink. To use the available spectrum efficiently, the channels are reused in adjacent clusters. For the conceptual hexagonal layout shown in Fig. 2.1, the reuse distance D is related to the cluster size C by $D = R\sqrt{3C}$ where R is the cell radius [2]. The capacity limit is thus $\eta = N_c/2C$ where N_c is the total number of channels available to the whole system. The units of this capacity measure are users/cell.

Clearly, N_c is directly related to the total available bandwidth. For example, in a system that is allocated a bandwidth of 40 MHz and uses 30 kHz frequency channels with 3 time slots per frequency channel, $N_c = 4000$. With $C = 7$ the capacity limit is thus $\eta = 285$, meaning that a maximum of 285 simultaneous calls may occur in the geographic area covered by each cell. With $C = 4$, the capacity grows to 500. The higher the density of users in a geographic area, the smaller the cells have to be in order to obtain an acceptable blocking probability, defined as the probability that all channels are in use, preventing new users from placing calls.

To maximize system capacity, one must choose the cluster size C as small as possible. However, a smaller C means a higher level of cochannel interference (CCI) due to a smaller

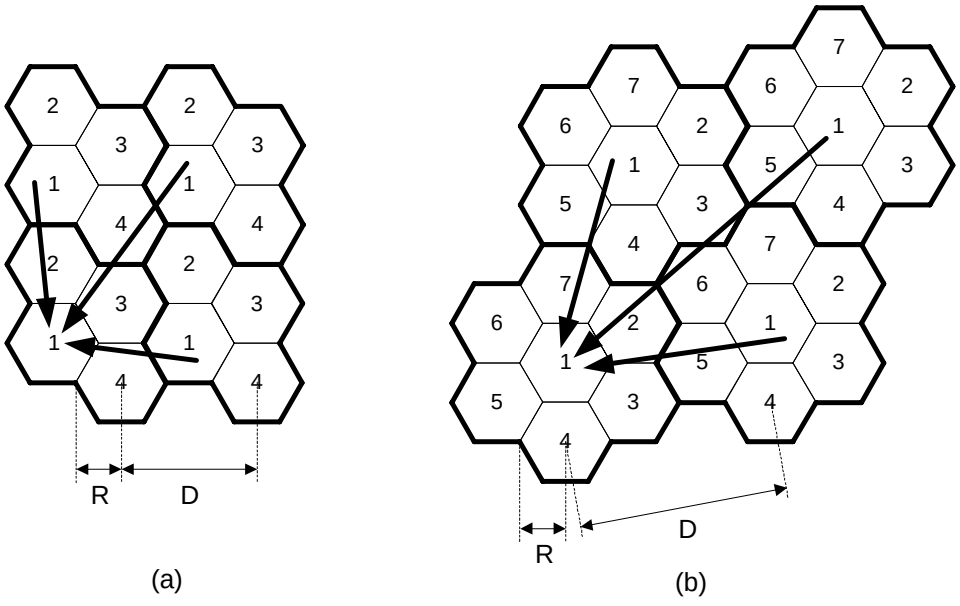

Figure 2.1: Hexagonal cell layouts for a TDMA system with cluster sizes (a) $C = 4$, and (b) $C = 7$. The cell radius is R and the reuse distance is $D = R\sqrt{3C}$. The arrows indicate intercell cochannel interference received in the uplink.

reuse distance D as illustrated in Fig. 2.1. The ability of the modulation, detection, coding, and diversity to withstand cochannel interference sets a lower limit on the cluster size.

Recently, it has been proposed to extend the capacity of conventional TDMA systems by allowing more than one user within a cell to share the same time/frequency slot, i.e., allow RWC. If M users in a cell share the same slot, the capacity is extended to $\eta = MN_c/2C$ users/cell. However, the factor by which system capacity is enhanced may not always be as large as M, since cochannel cells in other clusters also support RWC. As a result, the level of intercell interference increases, possibly requiring a larger cluster size C to compensate. This partially dilutes the desired capacity increase. However, it is demonstrated later in this thesis (see Chapter 6), that the dilution is small using joint detection with diversity arrays.

In addition to the increased intercell interference in systems with RWC, *intracell* interference is present due to the multiple intracell users sharing the same slot. As a result, the receiver has the job of distinguishing the cochannel signals from the intracell users. To accomplish this task, either an interference suppression or joint detection technique may be

implemented in combination with an antenna array at the base station. In this chapter, a survey of existing interference suppression and joint detection techniques is provided in order to place joint detection in context. The survey is focused on the uplink only, and the signal models do not include intercell CCI. A discussion of the downlink and a treatment of intercell CCI are deferred to later in this thesis (see Chapters 4 and 6 respectively).

For illustrative purposes, the system model shown in Fig. 2.2 is used. In this model, the M cochannel signals are synchronized by symbol and are transmitted across frequency-flat fading channels. Most of the surveyed techniques have an asynchronous/frequency-selective fading counterpart, and at various points in the development these are briefly discussed.

The transmitted signal from the mth user is assumed to be linearly modulated and is given by

$$s_m(t) = A_m \sum_n c_m(n)\, p(t - nT) \tag{2.1}$$

where $p(t)$ is a root-Nyquist pulse shape, T is the symbol period, A_m is the transmit amplitude, and $c_m(n)$ is a data symbol drawn from a signal constellation such as PSK or QAM. The base station receiver consists of an array of L antennas followed by a bank of matched filters (MF), each with response $p^*(-t)$. The matched filter outputs are sampled at the symbol rate yielding a vector $\mathbf{r}(k)$ of received samples. The vector $\mathbf{r}(k)$ is then processed yielding the symbol estimate $\widehat{c}_m(k)$ for each user.

In this thesis, the following conventions are used:

- All signals are represented by their complex baseband equivalents. Thus, the power (or variance) of a particular bandpass signal $\widetilde{s}(t)$ with baseband equivalent $s(t)$ is given by $P_s = \frac{1}{2}E\left[|s(t)|^2\right]$, where $E[\cdot]$ is the expectation operator.

- A bold lower case variable such as $\mathbf{x}$ represents a vector, and a bold uppercase variable such as $\mathbf{X}$ represents a matrix.

- The symbol * denotes the complex conjugate of a scalar or vector, the symbol ^{T} denotes the vector or matrix transpose, and the symbol † denotes the complex conjugate transpose, or Hermitian transpose.

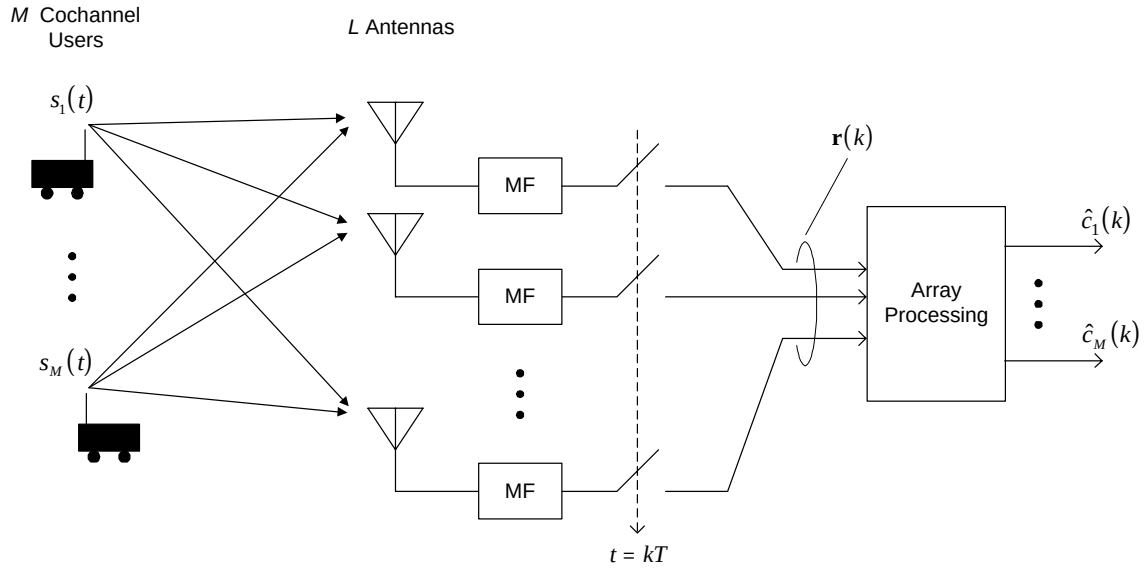

Figure 2.2: Common system model used in the survey of interference suppression and joint detection technqiues using antenna arrays. The M cochannel signals are synchronized by symbol and are transmitted across flat-fading channels.

2.1 Interference Suppression

In mobile communications, the use of antenna arrays for the purposes of interference suppression can be seen as an evolution of classical array signal processing which focuses on beamforming and direction finding [3]. In the following sections, this evolution is discussed, first setting the stage with classical beamforming and then moving on to diversity combining techniques such as maximal ratio combining, zero forcing, and minimum mean squared error (MMSE).

2.1.1 Classical Beamforming

The usual assumption made in classical beamforming is that each cochannel signal originates from a point source in the far field of the antenna array [4]. As a result, the wavefronts impinging on the antenna array are flat. It is important to note that this plane-wave arrival assumption immediately breaks down in a mobile environment where each user is surrounded

by a large number of scatterers. While each individual multipath component may be considered as a plane-wave, the composite signal comprised of all multipath components may not. The result is called angle spread. Fig. 2.3 illustrates the difference between plane-wave and non plane-wave arrivals at a uniform linear antenna array with inter-element spacing d.

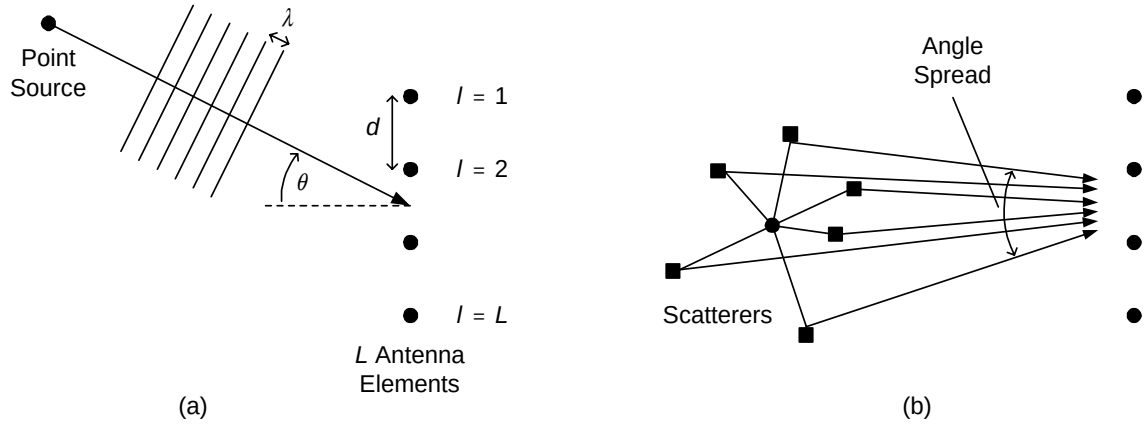

Figure 2.3: (a) Plane-wave arrival from a point source in the far field; (b) Scatterers around a point source causing angle spread.

Referring to the plane-wave arrival scenario in Fig. 2.3(a), θ is the angle of incidence, and λ is the wavelength given by $\lambda = c/f_c$ where c is the speed of light and f_c is the carrier frequency. Using a simple geometric argument, the phase of the carrier at the lth antenna element measured with respect to the first element is $\beta d\,(l-1)\sin\theta$ where $\beta = 2\pi/\lambda$ is the wavenumber. Under the assumption that the fractional bandwidth of the signal $s_m(t)$ is small (the narrowband assumption [4]), the composite received signal vector $\mathbf{r}(k)$ in Fig. 2.2 is given by

$$\mathbf{r}(k) = \sum_{m=1}^{M} A_m c_m(k)\, g_m \mathbf{a}(\theta_m) + \mathbf{n}(k) \tag{2.2}$$

where θ_m is the angle of arrival (AOA) of plane-wave from the mth user and $\mathbf{a}(\theta_m)$ is the steering vector, given by

$$\mathbf{a}(\theta_m) = \begin{pmatrix} 1 & e^{j\beta d\sin\theta_m} & e^{j2\beta d\sin\theta_m} & \cdots & e^{j(L-1)\beta d\sin\theta_m} \end{pmatrix}^T. \tag{2.3}$$

The complex random variable g_m models the gain (attenuation) and phase shift in the mth user's channel, and the vector $\mathbf{n}(k)$ models thermal noise in the receiver. Since each

mobile is modeled as a point source in the far field, the channel gain g_m is the same for all antennas. Although not explicitly shown, g_m and θ_m (and thus the steering vector) are generally functions of time k as the mobile receiver moves.

Providing that the AOAs of all users are different, it is possible to suppress cochannel interference by appropriately weighting and combining the received antenna signals as shown in Fig. 2.4. For the mth user, the weight vector $\mathbf{w}_m = (w_{1m}, w_{2m}, \cdots, w_{Lm})^T$ is chosen such that it is orthogonal to the steering vectors of all other users; that is,

$$\mathbf{w}_m^\dagger \mathbf{a}(\theta_n) = \begin{cases} 1 & , \quad n = m \\ 0 & , \quad n \neq m \end{cases} . \tag{2.4}$$

If this condition is met, the output of the combiner for the mth user is

$$y_m(k) = \mathbf{w}_m^\dagger \mathbf{r}(k) = A_m c_m(k) g_m + \mathbf{w}_m^\dagger \mathbf{n}(k) \tag{2.5}$$

which explicitly shows that the interference from other users has been suppressed. The detection of the users' data sequences may then proceed independently. Clearly, the calculation of the weight vectors requires knowledge of the AOAs of all users. For the case of plane wave arrivals, several blind eigenmethods have been developed for estimating the AOAs. Among these, the most notable are MUSIC [5] and ESPRIT [6].

Further insight is gained by expressing the criterion in (2.4) as the matrix equation $\mathbf{A}\mathbf{w}_m = \mathbf{b}_m$ where the $M \times L$ matrix $\mathbf{A}$ is

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}^\dagger(\theta_1) \\ \mathbf{a}^\dagger(\theta_2) \\ \vdots \\ \mathbf{a}^\dagger(\theta_M) \end{bmatrix} . \tag{2.6}$$

The length-M column vector $\mathbf{b}_m$ has a one in the mth position, and zeros everywhere else. This equation has a well defined solution as long as the number of users M is less than or equal to the number of antennas L, and as long as $\mathbf{A}$ is full rank. The full rank condition is met as long as the AOAs for all users are different. In other words, as long as $M \leq L$ and the users are spatially separated, it is possible to completely suppress interference. If two or more users are not spatially separated, then certain users need to be allocated different channels — an undesirable situation since this reduces overall system capacity. For the case of $M < L$, the above matrix equation is under-determined; thus it has an infinite number of

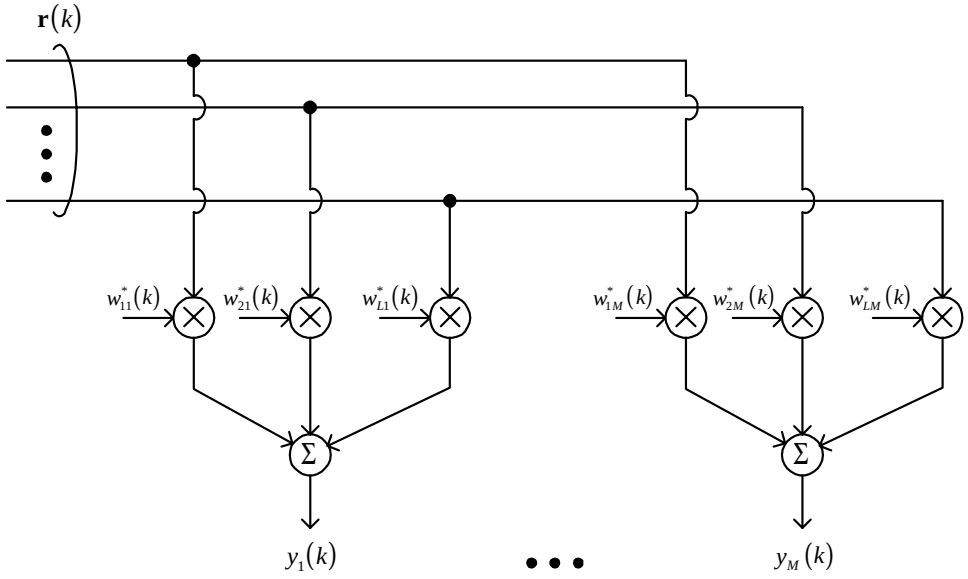

Figure 2.4: Linear antenna combiner for the suppression of cochannel interference.

solutions. A convenient choice is the minimum-norm solution given by $\mathbf{w}_m = \mathbf{A}^{\#}\mathbf{b}_m$ where $\mathbf{A}^{\#}$ is the generalized pseudoinverse or Moore-Penrose generalized inverse of the matrix $\mathbf{A}$ [7]. For the case of $M = L$, the pseudo-inverse becomes the regular inverse $\mathbf{A}^{-1}$.

A graphical representation of the above beamforming approach may be obtained by plotting the array factor for each user as a function of azimuth θ. The array factor for the mth user is defined as $F_m(\theta) = \mathbf{w}_m^{\dagger}\mathbf{a}(\theta)$ [1], and it describes the response of the array to a plane wave arrival at angle θ. According to the criterion in (2.4), $F_m(\theta_m) = 1$ and $F_m(\theta_n) = 0$ for $n \neq m$. At other angles, $F_m(\theta)$ depends on a variety of factors including the number of antennas, the number of interferers, the antenna spacing, as well as which solution of $\mathbf{A}\mathbf{w}_m = \mathbf{b}_m$ is actually chosen. As an example, Fig. 2.5(a) plots $|F_1(\theta)|^2$ using the minimum-norm solution for a two-user system. The desired user is located at an azimuth of $\theta = 30^\circ$, and the interfering user is located at $\theta = -45^\circ$. The uniform linear array has four elements with an inter-element spacing of $d = \lambda/2$. As can be seen, the main lobe of the beam pattern points toward the desired user, and a null is placed in the direction of the interferer. Notice that the beam pattern is symmetric about the axis of the array. This occurs because a signal arriving from angle θ and one from angle $180 - \theta$ produce the same

incremental phase shifts from element to element (see Fig. 2.3).

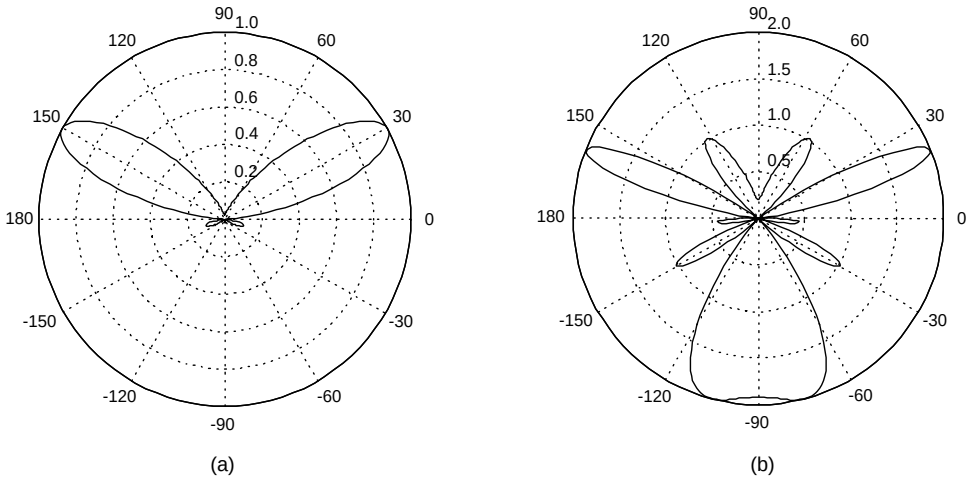

Figure 2.5: Beam patterns for a 4 element uniform linear array with inter-element spacing (a) $d = 0.5\lambda$ and (b) $d = 0.75\lambda$. The desired user is located at $\theta = 30^\circ$ and the cochannel interferer at $\theta = -45^\circ$.

While classical beamforming is a seemingly attractive technique for suppressing cochannel interference, it is not appropriate in a mobile communications environment for several reasons. One is that angle spread makes the number of arrivals exceeds the number of antennas. As a result, it is not possible to place nulls in the direction of all interfering arrivals. Furthermore, the traditional AOA estimation techniques break down.

Another significant problem relates to antenna spacing requirements. In mobile communications, obtaining diversity from a variety of sources is vital in order to combat the effects of fading. The use of multiple antennas is an attractive method of obtaining explicit diversity; however, the antennas must be spaced widely to obtain independent fading across the array. The required separation depends on the degree of angle spread as well as the azimuth of the mobiles [9], but a rule of thumb is that a 10 to 20 wavelength separation decorrelates the fading such that diversity is obtained. Smaller separations lead to higher correlation; however, as found in [9], correlation coefficients as high as about 0.6 do not significantly degrade performance. As shown in Fig. 2.5(b) spacing the antenna elements wider than a half wavelength causes grating lobes [4]. The beam pattern for $d = 0.75\lambda$ shows that while $F_1(30^\circ) = 1$ and $F_1(45^\circ) = 0$ as desired, several other lobes are focused

towards unintended directions. This is undesirable, since intercell cochannel interference may be enhanced. Consequently, interference suppression techniques more suitable to a mobile communications environment have been developed. These are discussed in the next section.

2.1.2 Diversity Combining

In a mobile communications environment, it is necessary to modify the received signal model in order to take into account multipath fading. Because there is no single AOA in angle spread, and because wide antenna spacings are required to achieve diversity, AOA information ceases to be a factor and the model becomes stochastic.

An appropriate model of the received signal vector $\mathbf{r}(k)$ in Fig. 2.2 for the case of frequency-flat Rayleigh fading is

$$\mathbf{r}(k) = \sum_{m=1}^{M} A_m c_m(k)\, \mathbf{g}_m(k) + \mathbf{n}(k). \tag{2.7}$$

The elements of the channel gain vector $\mathbf{g}_m(k)$ are complex Gaussian random variables (RVs) all with the same variance $\sigma_{g_m}^2$. Assuming the antennas are spaced far enough apart to obtain diversity, the elements of $\mathbf{g}_m(k)$ are independent. The elements of the noise vector $\mathbf{n}(k)$ are also independent, identically distributed (i.i.d.) complex Gaussian RVs but with variance N_o. The data symbols are normalized such that $E\left[|c_m(k)|^2\right] = 1$. This expression is similar to that in (2.2); however, AOA information is not explicitly modeled, rather it is lumped in with the random channel gain vectors.

An important fact with regards to the signal model in (2.7) is that even if two mobiles are surrounded by the same set of scatterers, the channel gain vectors at the receiver are independent as long as the mobiles are separated by more than a few wavelengths. This is due to the rapidly varying spatial field surrounding the mobiles: typically the channel gains change in as little as half a wavelength — a distance of only 15 cm for a carrier frequency of 1 GHz. As a result, interference suppression may be accomplished with very little spatial isolation, in contrast to classical beamforming which requires mobiles to be widely separated in azimuth.

Maximal Ratio Combiner

It is instructive to first investigate the special case of a single-user system which leads to the well-known maximal ratio combining (MRC) receiver [10] which is effective for noise suppression. In this case, the received signal in (2.7) is $\mathbf{r}(k) = Ac(k)\,\mathbf{g}(k) + \mathbf{n}(k)$ which is free of interference from other users. To achieve noise suppression, the weight vector $\mathbf{w}(k)$ in Fig. 2.2 is chosen so as to maximize the instantaneous signal-to-noise ratio (SNR) at the combiner output given by

$$\gamma = \frac{\frac{1}{2}E\left[\left|Ac(k)\,\mathbf{w}(k)^{\dagger}\,\mathbf{g}(k)\right|^2\right]}{\frac{1}{2}E\left[\left|\mathbf{w}(k)^{\dagger}\,\mathbf{n}(k)\right|^2\right]} = \frac{A^2}{2N_o}\frac{\left|\mathbf{w}(k)^{\dagger}\,\mathbf{g}(k)\right|^2}{\left|\mathbf{w}(k)^{\dagger}\,\mathbf{w}(k)\right|}. \tag{2.8}$$

Using Schwartz's inequality [7], which states that $\left|\mathbf{x}^{\dagger}\mathbf{y}\right| \leq \|\mathbf{x}\|\,\|\mathbf{y}\|$ where $\|\mathbf{x}\|$ is the Euclidean norm (length) of the vector $\mathbf{x}$,

$$\gamma \leq \frac{A^2}{2N_o}\|\mathbf{g}(k)\|^2. \tag{2.9}$$

Equality is achieved with $\mathbf{w}(k) = K\mathbf{g}(k)$ where K is an arbitrary constant. Choosing $K = 1$ results in the combiner output

$$y(k) = \mathbf{w}^{\dagger}(k)\,\mathbf{r}(k) = A\|\mathbf{g}(k)\|^2 c(k) + \mathbf{g}^{\dagger}(k)\,\mathbf{n}(k) \tag{2.10}$$

which shows that the signal on a particular branch is weighted by the conjugate of the complex gain on that branch. With this weighting, the signals on all branches combine coherently and diversity is obtained.

For the case of binary phase-shift keying (BPSK) modulation, data detection is based on the decision variable $d(k) = \mathrm{Re}[y(k)]$. Assuming the bit-to-symbol mapping $[0,1] \rightarrow [1,-1]$, the detector decides a '0' was transmitted if $d(k) \geq 0$, and a '1' was transmitted if $d(k) < 0$. The resulting average bit-error rate (BER) [8] is

$$P_b = \frac{1}{(1-r)^{2L-1}}\sum_{k=0}^{L-1}\binom{2L-1}{k}(-r)^k \tag{2.11}$$

where

$$r = \frac{1+\sqrt{1+\Gamma^{-1}}}{1-\sqrt{1+\Gamma^{-1}}} \tag{2.12}$$

and Γ is the average per-antenna SNR given by $\Gamma = A^2\sigma_g^2/N_o$.

The diversity effect is observed by looking at the asymptotic BER, obtained by letting Γ in the above expressions become very large. For $\Gamma \gg 1$, the quantity r approaches $r_{asymp} = -4\Gamma$ which increases without bound as Γ increases. Thus, the summation in (2.11) is dominated by the last term, resulting in the asymptotic BER

$$\begin{aligned} P_{b,asymp} &= \binom{2L-1}{L-1}(-r_{asymp})^{-L} \\ &= 4^{-L}\binom{2L-1}{L-1}\Gamma^{-L}. \end{aligned} \tag{2.13}$$

This result shows that the BER varies asymptotically as Γ^{-L}. In other words, L-fold diversity is obtained. The diversity effect is illustrated in Fig. 2.6 which plots both the BER and asymptotic BER. Clearly, diversity reception improves performance dramatically. With L antennas, a 10 dB increase in SNR results in a reduction in BER of L orders of magnitude.

Zero Forcing Combiner

For the case of multiple users in which the general signal model in (2.7) applies, the simplest technique of achieving interference suppression is through the zero forcing (ZF) technique, analogous to ZF equalizers [8]. This method is very similar in spirit to the method of choosing the antenna weights for classical beamforming. The same linear combining structure as in Fig. 2.4 is used, and the weight vector for the mth user is chosen according to the zero-forcing criterion

$$\mathbf{w}_m^\dagger(k)\,\mathbf{g}_n(k) = \begin{cases} 1 & , \quad n = m \\ 0 & , \quad n \neq m \end{cases} \tag{2.14}$$

which results in the combiner output

$$y_m(k) = \mathbf{w}_m^\dagger(k)\,\mathbf{r}(k) = A_m c_m(k) + \mathbf{w}_m^\dagger(k)\,\mathbf{n}(k). \tag{2.15}$$

Clearly, the interference from other users is completely suppressed.

The criterion in (2.14) may also be expressed as $\mathbf{G}(k)\,\mathbf{w}_m(k) = \mathbf{b}_m$ where the mth row of the channel gain matrix $\mathbf{G}(k)$ is given by $\mathbf{g}_m^\dagger(k)$. Like with classical beamforming, the solution is well defined (i.e., the interference from other users is completely suppressed) as long as $M \leq L$ and as long as $\mathbf{G}(k)$ is of full rank. Because $\mathbf{G}(k)$ is random, the full rank condition is probabilistic; however, as long as the channel gains vectors are different

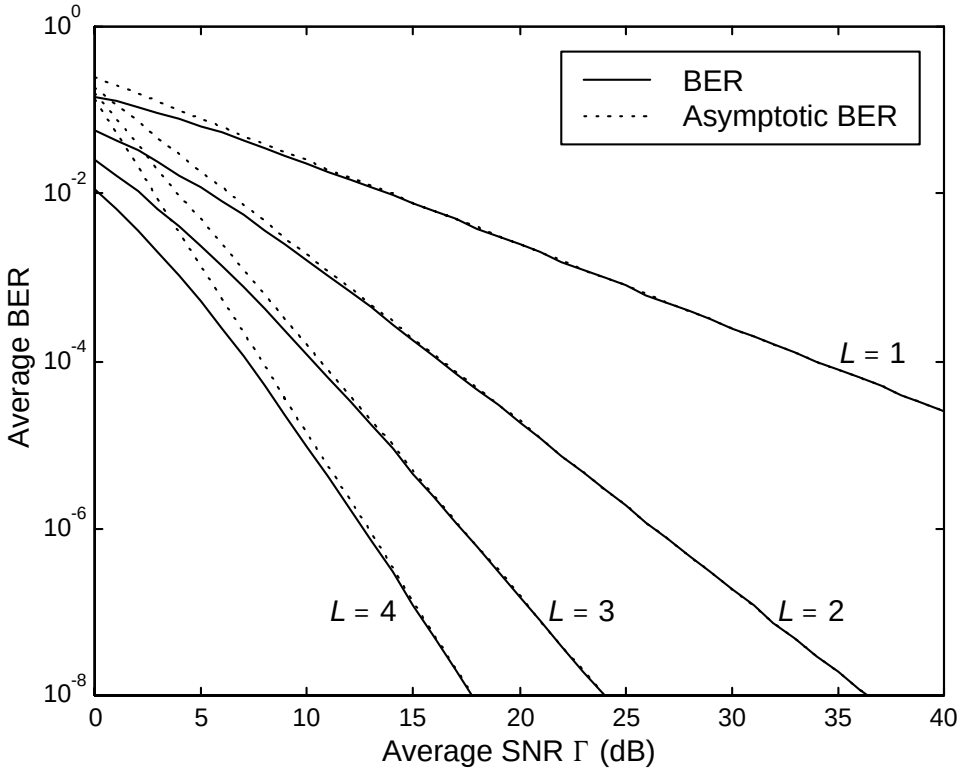

Figure 2.6: Average BER for the MRC receiver with L antennas.

among all users, it occurs with probability one. In this regard, fading is actually a benefit in terms of signal separability: the users need not be separately widely in azimuth in order to suppress interference.

The performance of the ZF combiner may be related to that of the MRC combiner [11] as will now be shown. In this discussion, the time variance of all variables is dropped for compactness of notation. A convenient choice of the weight vector that satisfies (2.14) is $\mathbf{w}_m = \mathbf{U}_m \mathbf{v}_m$ where the matrix $\mathbf{U}_m$ is of dimension $L \times (L - M + 1)$, and the vector $\mathbf{v}_m$ is of dimension $(L - M + 1) \times 1$. The combiner output is thus $y_m = \mathbf{w}_m^\dagger \mathbf{r} = \mathbf{v}_m^\dagger \mathbf{r}'_m$ where $\mathbf{r}'_m = \mathbf{U}_m^\dagger \mathbf{r}$. The length of $\mathbf{r}'_m$ is $(L - M + 1)$. This shows that $\mathbf{r}$ is first projected onto a lower dimensional subspace. The components of the reduced dimensionality vector $\mathbf{r}'_m$ are then weighted and combined using the vector $\mathbf{v}_m$. The function of the matrix $\mathbf{U}_m$ is to suppress interference, and that of $\mathbf{v}_m$ is to perform diversity combining.

In order to suppress interference, the columns of $\mathbf{U}_m$ are chosen to be orthogonal to all but the mth column of the channel gain matrix $\mathbf{G}$. Since the columns of $\mathbf{G}$ (the users' gain vectors) are of dimension L, the columns of $\mathbf{U}_m$ lie in the $L-M+1$ dimensional subspace orthogonal to the space spanned by the set of vectors $\{\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_{m-1}, \mathbf{g}_{m+1}, \cdots, \mathbf{g}_M\}$. Thus,

$$\begin{aligned} \mathbf{r}'_m &= \begin{bmatrix} \mathbf{u}^\dagger_{m,1} \\ \mathbf{u}^\dagger_{m,2} \\ \vdots \\ \mathbf{u}^\dagger_{m,L-M+1} \end{bmatrix} \begin{bmatrix} A_1\mathbf{g}_1 & A_2\mathbf{g}_2 & \cdots & A_M\mathbf{g}_M \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_M \end{bmatrix} + \mathbf{U}^\dagger_m\mathbf{n} \\ &= A_m c_m \mathbf{g}'_m + \mathbf{n}' \end{aligned} \tag{2.16}$$

where the length-L vector $\mathbf{u}_{m,l}$ is the lth column of the matrix $\mathbf{U}_m$. The reduced dimensionality gain and noise vectors are given by $\mathbf{g}'_m = \mathbf{U}^\dagger_m\mathbf{g}_m$ and $\mathbf{n}' = \mathbf{U}^\dagger_m\mathbf{n}$. The latter equality of (2.16) clearly shows that the interference from other users has been suppressed.

If the columns of $\mathbf{U}_m$ are chosen so as to form an orthonormal basis of the orthogonal subspace — that is, $\mathbf{U}^\dagger_m\mathbf{U}_m = \mathbf{I}_{L-M+1}$ where $\mathbf{I}_N$ is the identity matrix of order N — then the statistics of $\mathbf{g}'_m$ and $\mathbf{n}'$ are invariant under the transformation $\mathbf{U}^\dagger_m$. In other words, the elements of both $\mathbf{g}'_m$ and $\mathbf{n}'$ remain independent and identically distributed with unchanged variances $\sigma^2_{g_m}$ and N_o respectively. The only difference is that $\mathbf{g}'_m$ and $\mathbf{n}'$ are of dimension $L-M+1$ instead of L.

In order to satisfy (2.14), the vector $\mathbf{v}_m$ should be chosen as

$$\mathbf{v}_m = \frac{\mathbf{g}'_m}{\|\mathbf{g}'_m\|^2}. \tag{2.17}$$

Interestingly, this choice of $\mathbf{v}_m$ is precisely the same weight vector as one would choose if performing maximal ratio combining with the transformed received signal vector $\mathbf{r}'_m$. Since the components of $\mathbf{r}'_m$ are i.i.d., the performance of the ZF receiver with L antennas and M interfering users is identical to the MRC receiver operating in an interference-free environment with $L-M+1$ antennas. In other words, the price paid for nulling interference is a reduction in diversity order for all users. For example, with one user, the diversity order is L, whereas with M users the diversity order is reduced to unity for all users.

MMSE Combiner

The ZF combiner achieves complete suppression of interference, with no regard to the relative levels of noise and interference. This is undesirable at low SNRs where the noise level is significant. An alternative structure that balances the suppression of noise and interference is the MMSE combiner. With MMSE antenna combining (often called optimum combining [12]) the antenna weights for the mth user are chosen so as to minimize the mean squared error (MSE) between the transmitted symbol $c_m(k)$ and the combiner output $y_m(k)$. In this way, the noise and interference are treated as a combined disturbance to the desired signal. The combiner attempts to minimize the disturbance, irrespective of the relative levels of the noise and interference.

The MSE for a given weight vector $\mathbf{w}_m^\dagger(k)$ is

$$\begin{aligned} \varepsilon_m &= \frac{1}{2}E\left[\left|c_m - \mathbf{w}_m^\dagger \mathbf{r}\right|^2\right] \\ &= \frac{1}{2} - \mathbf{p}_m^\dagger \mathbf{w}_m - \mathbf{w}_m^\dagger \mathbf{p}_m + \mathbf{w}_m^\dagger \mathbf{R}\mathbf{w}_m \end{aligned} \tag{2.18}$$

where $\mathbf{p}_m = \frac{1}{2}E[c_m^*\mathbf{r}]$ is the cross-correlation vector between the transmitted symbol sequence and the vector of received samples, and $\mathbf{R} = \frac{1}{2}E\left[\mathbf{r}\mathbf{r}^\dagger\right]$ is the covariance matrix of the received sample vector. The expectation is taken over the noise ensemble and the symbol ensembles of all users. Again, the time dependence of all variables has been dropped for compactness of notation. Differentiating (2.18) with respect to $\mathbf{w}_m^*$ gives

$$\frac{\partial \varepsilon_m}{\partial \mathbf{w}_m^*} = -\mathbf{p}_m + \mathbf{R}\mathbf{w}_m. \tag{2.19}$$

Setting this result to zero and solving for the optimal weight vector gives $\mathbf{w}_{m,o} = \mathbf{R}^{-1}\mathbf{p}_m$. Substituting this back into (2.18) gives the minimum MSE as $\varepsilon_{m,o} = \frac{1}{2} - \mathbf{p}_m^\dagger \mathbf{R}^{-1}\mathbf{p}_m$.

In more detail, the cross correlation vector is

$$\begin{aligned} \mathbf{p}_m &= \frac{1}{2}E\left[c_m^*\left(\sum_{n=1}^{M} A_n c_n \mathbf{g}_n + \mathbf{n}\right)\right] \\ &= \frac{1}{2}A_m \mathbf{g}_m \end{aligned} \tag{2.20}$$

where it has been assumed that the users' symbol sequences are mutually uncorrelated and

that the symbol sequences and noise sequence are uncorrelated. Under the same assumptions, the covariance matrix is

$$\begin{aligned}\mathbf{R} &= \frac{1}{2}E\left[\sum_{m=1}^{M}\sum_{n=1}^{M}A_m A_n c_m c_n^* \mathbf{g}_m \mathbf{g}_n^\dagger\right] + \frac{1}{2}E\left[\mathbf{n}\mathbf{n}^\dagger\right] \\ &= \frac{1}{2}\sum_{n=1}^{M}A_n^2 \mathbf{g}_n \mathbf{g}_n^\dagger + N_o\mathbf{I}\end{aligned}$$

where $\mathbf{I}$ is the $L \times L$ identity matrix. This gives the optimal weight vector for the MMSE combiner as

$$\mathbf{w}_{m,o}(k) = A_m\left[\sum_{n=1}^{M}A_n^2 \mathbf{g}_n(k)\,\mathbf{g}_n^\dagger(k) + 2N_o\mathbf{I}\right]^{-1}\mathbf{g}_m(k) \tag{2.21}$$

where the time dependence has been reinserted.

An interesting fact is that the asymptotic performance of MMSE and ZF are identical [11]. Normally the MMSE combiner balances interference and noise suppression; however, in the high SNR regime where noise is negligible, it need only suppress interference which is exactly what the ZF combiner does. Thus, for MMSE combining with L antennas and M users, the diversity order experienced by each user is $L - M + 1$, and the maximum number of users is limited to the number of antennas. It will be shown later in this thesis that the loss in diversity order as the number of users increases has an adverse effect on system capacity.

Observing (2.21), one can see that the optimal weight vector is a function of the instantaneous channel gain vectors. Thus, one option for MMSE combining is to explicitly estimate the channel gain vectors through a training-based multiuser channel estimation technique such as the one developed in Chapter 5 of this thesis. Another option is to avoid explicit channel estimation and implement one of a number of adaptive techniques, although these techniques still require some sort of training. For example, [13] considers adaptivity using the well-known least mean squares (LMS) algorithm.

The LMS algorithm in the context of an adaptive array proceeds as follows. Observe that the mean squared error ε_m in (2.18) is a quadratic function of the weight vector $\mathbf{w}_m$. Thus, one may visualize ε_m as a bowl-shaped surface with a unique minimum that occurs at $\mathbf{w}_m = \mathbf{R}^{-1}\mathbf{p}_m$. The adaptation process has the task of continually searching for the minimum point of the error surface. This is accomplished by starting with an arbitrary weight vector and then continually adjusting it in a direction opposite to the gradient of

the error surface $\nabla\varepsilon_m$, i.e., in the direction of steepest descent. The true gradient is given by equation (2.19); however, this expression requires the calculation of $\mathbf{p}_m$ and $\mathbf{R}$ which in turn require explicit knowledge of the channel gain vectors. Thus, at the nth iteration, the LMS algorithm uses an instantaneous estimate of the gradient given by

$$\widehat{\nabla}\varepsilon_m(n) = -\widehat{\mathbf{p}}_m(n) + \widehat{\mathbf{R}}(n)\,\mathbf{w}_m(n) \tag{2.22}$$

where $\widehat{\mathbf{p}}_m(n) = c_m^*(n)\,\mathbf{r}(n)$ and $\widehat{\mathbf{R}}(n) = \mathbf{r}(n)\,\mathbf{r}^\dagger(n)$. The weight vector at the $n+1$th iteration is then

$$\begin{aligned}\mathbf{w}_m(n+1) &= \mathbf{w}_m(n) - \mu\widehat{\nabla}\varepsilon_m(n)\\ &= \mathbf{w}_m(n) + \mu\mathbf{r}(n)\,e_m^*(n),\end{aligned} \tag{2.23}$$

where the error signal $e_m(n)$ is given by

$$e_m(n) = c_m(n) - y_m(n). \tag{2.24}$$

The error signal is the difference between the desired (or reference) signal and the combiner output. It is the mean-square value of the error signal that the algorithm attempts to minimize. The parameter μ is called the step-size, and it controls the speed of adaptation as well as the amount of jitter in the converged value of the weight vector.

Evidently, the LMS algorithm is very simple; it has only one parameter, and depends only on the value of the error signal and the array output $\mathbf{r}(n)$. However, calculation of the error signal requires knowledge of the symbol sequence of the mth user which is unknown. Thus, periodic training is required, first for initial convergence, then for tracking changes in the channel conditions. In between training periods, a decision directed mode may be used. The LMS algorithm typically suffers from slow convergence, limiting its use to a very slow fading environment [13]. In a faster fading environment, algorithms such as recursive-least-squares (RLS) have been applied, though RLS still requires training and is significantly more complex than LMS.

2.1.3 Diversity Combining in Delay Spread

As mentioned previously, the interference suppression techniques in the foregoing sections have asynchronous user/frequency-selective fading (delay spread) counterparts. For the most part, the receiver structures with antenna diversity are a generalization of well-known

single-channel equalizer structures. In this section, an MMSE linear equalizer/combiner structure and a decision feedback equalizer (DFE) structure are briefly discussed.

In a delay spread environment, an appropriate model for the received signal on the lth antenna after matched filtering and sampling twice per symbol period is

$$r_l(k) = \sum_{m=1}^{M} A_m \mathbf{c}_m^T(k)\, \mathbf{h}_{lm}(k) + n_l(k) \tag{2.25}$$

where $\mathbf{h}_{lm}(k)$ contains $\frac{T}{2}$-spaced samples of the time variant impulse response $h_{lm}(\tau;t)$ of the channel between the mth user and the lth antenna (see the channel model in Chapter 5). The symbol vector $\mathbf{c}_m(k)$ contains the current and past symbols spanning the memory of the channel impulse response (CIR). The CIR includes the effects of the transmit filter $p(t)$, the physical channel $g_{lm}(\tau;t)$, the matched filter $p^*(-t)$, as well as the relative delay τ_m that models the lack of synchronism between users. Note that sampling twice per symbol is required in order to avoid aliasing the CIR.

MMSE combining/equalization in a delay spread environment is accomplished by employing a filter on each diversity branch [14],[15] rather than a single weight as in Fig. 2.4. In this way, the structure is able to suppress inter-symbol interference (ISI) as well as CCI. Fig. 2.7 shows a block diagram of the combiner/equalizer structure. The filter on the lth branch of the mth user's combiner is designed to span the memory of the channel impulse response and has coefficients stored in the length-L_c vector $\mathbf{w}_{lm}(k)$. The input to this filter is the vector $\mathbf{r}_l(k)$ which contains the samples $r_l(k), r_l(k-1), \cdots, r_l(k-L_c-1)$. The output of the combiner for the mth user is then

$$y_m(k) = \sum_{l=1}^{L} \mathbf{w}_{lm}^{\dagger}(k)\, \mathbf{r}_l(k). \tag{2.26}$$

Since the filters on each branch must be jointly optimized, a length-LL_c joint weight vector $\mathbf{u}_m(k)$ is defined which is simply the concatenation of the L different filter vectors $\mathbf{w}_{lm}(k)$ of the mth user's combiner. Similarly, a length-LL_c joint sample vector $\mathbf{x}(k)$ is defined as the concatenation of the L different received sample vectors $\mathbf{r}_l(k)$. The combiner output is then $y_m(k) = \mathbf{u}_m^{\dagger}(k)\, \mathbf{x}(k)$ which is equivalent to (2.26). MMSE optimization then proceeds in a similar manner as described in the previous section. The optimal joint filter vector is $\mathbf{u}_{m,o}(k) = \mathbf{R}^{-1}\mathbf{p}_m$ where $\mathbf{R}$ is the covariance matrix of the vector $\mathbf{x}(k)$, and $\mathbf{p}_m$ is the cross correlation between the mth users symbol $c_m(k)$ and the received sample vector $\mathbf{x}(k)$.

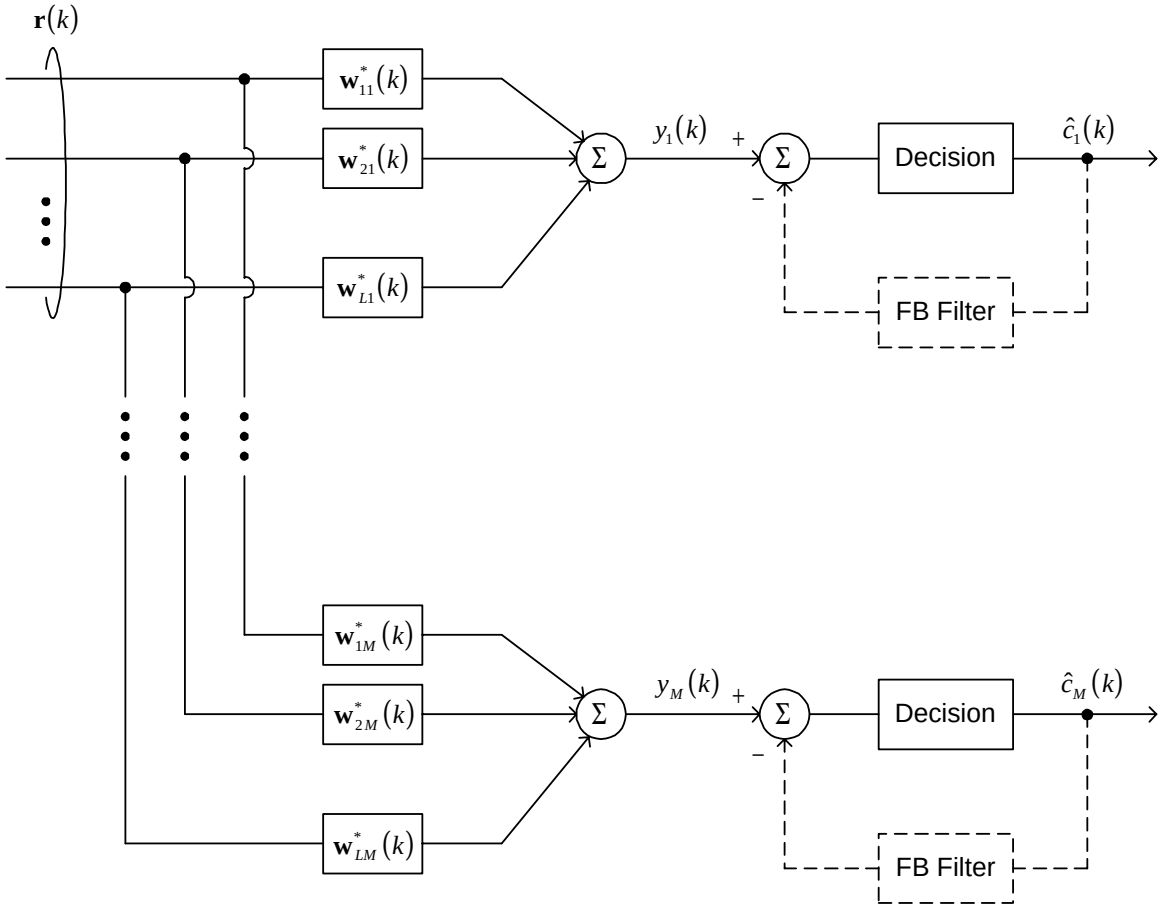

Figure 2.7: Combiner/equalizer structure with or without decision feedback.

As with MMSE combining in flat fading, the MMSE equalizer/combiner here requires training, either to estimate the M users' channel vectors explicitly, or to form an error signal for use in adapting the filter coefficients. Furthermore, it incurs a loss in diversity order as the number of users increases. Precise statements about the actual diversity order are difficult to make, because the equalizer exploits the implicit frequency diversity available in the multipath channels which depends on the actual power-delay profiles [15].

Because linear equalizers are not able to effectively equalize certain channels such as those with spectral nulls, the authors of [16], [17], and [18] propose to improve the performance of the MMSE combiner/equalizer by employing decision feedback to cancel residual ISI due to previously detected symbols. Note that not all the ISI is cancelled; the ISI due to the future undetected symbols remains. In these studies, the decisions of the mth user are fed back to the output of the mth combiner as shown by the dashed lines in Fig. 2.7. However, as with all decision feedback structures, correct decisions must be fed back to avoid error propagation and the associated degradation in performance.

In [19], [20], and [21], the authors propose to improve performance even further by feeding back decisions of other users to cancel the residual ISI due to the previously detected symbols of the cochannel interferers. This cross-feedback structure is reminiscent of joint detection, discussed in the next section, where the structure of the cochannel interference is exploited to the fullest degree in the detection process. Yet another improvement is the use of maximum likelihood sequence estimation (MLSE) in combination with diversity arrays [22]. With this structure, diversity combining is performed within the receiver metric which takes into account the structure of the interference and noise through the interference-plus-noise covariance matrix.

2.2 Joint Detection

With the previous techniques, each signal is separately demodulated after suppressing interference from other cochannel users. In contrast, joint detection (JD) focuses on jointly demodulating the cochannel signals by relying the finite alphabet (FA) property of the interference. With joint detection, differences among the cochannel signals must be exploited in order to distinguish the users. Much attention has been focused on JD in the context of CDMA systems [23] since the cochannel signals may easily be distinguished on the basis of different spreading codes. The interest in JD for CDMA has been motivated primarily

by the poor performance of the conventional detector which ignores multiple access interference (MAI). Early works in this area are [24] and [25], both of which consider optimal joint detection of multiple asynchronous CDMA signals. The former considers an AWGN channel and the latter a Rayleigh fading channel. In [24], it is shown that the optimal detector may be implemented using a bank of matched filters followed by processing using the Viterbi algorithm (VA). Unfortunately, the state set of the VA grows exponentially with the number of users making it prohibitively complex for typical CDMA systems with many users. This has motivated the search for suboptimal structures with reduced complexity. Examples include the decorrelating detector of [26], the parallel interference cancellation (IC) structure of [27], the serial IC structure of [28], and the sequential decoder of [29].

An alternative method of distinguishing the cochannel signals is to exploit differences in the channel impulse responses between each user and the receiver. This method does not rely on spectrum spreading. Thus it may be applied to narrowband systems such as TDMA to achieve the goal of increasing capacity by allowing several users to share the same time/frequency slot. Because only a few users are likely to share the same slot, the computational complexity of the optimal detector is not as significant an issue as for CDMA systems.

Joint detection based solely on channel differences — the topic of this thesis — has received only limited attention in the literature, e.g., [30]–[33]. In both [30] and [31] constant channels and perfect channel state information are assumed, although in [30] multichannel reception is considered. In [32], deterministic channel estimation offsets are investigated, but both the channel and the estimation errors are constant, and the receiver has only one channel available. In [33], fading channels are considered, but antenna diversity is not employed, and only limited simulation results appear.

In this thesis, a number of new contributions to the topic are made which also appear in [34]–[36]: first, both fading channels and antenna diversity are considered; second, a fully-analytical expression for the union bound on symbol average symbol-error rate is developed for an arbitrary number of users and diversity antennas for both perfect and imperfect channel state information; and third, the performance of joint detection is compared with classical MMSE combining.

Although the signal model is simplified by requiring the users to be synchronized by symbol, this work provides motivation for the investigation of asynchronous performance. While it is not demonstrated here, lack of synchronism may improve performance, since the

users' relative delays, once estimated, provide an additional basis upon which the cochannel signals may be distinguished. See, for example, [37] which considers signal separation based on distinct delays.

In the case of asynchronous users and/or frequency selective fading where sequence detection must be employed rather than symbol-by-symbol detection, the optimal detector is implemented using the VA operating on a "joint trellis." The states in the joint trellis are formed simply by concatenating the M users' symbol vectors appearing in the $\frac{T}{2}$-spaced received signal model in equation (2.25). The symbol vector for each user contains the current and past L_c transmitted symbols, where L_c is the memory (in symbol periods) of the channel impulse responses. The number of states is thus Q^{ML_c}, where Q is the number of points in the signal constellation. Clearly the state set grows exponentially with the number of users. The more general case of asynchronous users and/or frequency selective fading is considered in papers that appeared subsequent to [34] and [35], e.g., [38] and [39].

In this thesis, it is implicitly assumed that the users' channels may be estimated through embedded references, i.e., training signals. The availability of embedded references significantly aids the detection process resulting in robust receivers; however, training information is not always available. This has motivated research in the area of blind joint detection, e.g., [40]–[46] and references therein. In [40]–[44], blind joint maximum likelihood techniques are considered. In [45]–[46] the constant modulus (CM) algorithm is considered, which may be thought of as the blind counterpart to interference suppression using MMSE antenna combining.

Chapter 3

Joint Detection in the Uplink

In this chapter the behaviour and performance of joint detection with diversity arrays is established for the case of frequency-flat fading and symbol-synchronous users. To highlight the novel reception behaviour of joint detection, the performance is compared to classical MMSE combining. The material in this chapter — published in [34][1] and [35][2] and to appear in [36] — deals with the uplink, whereas the material in Chapter 4 considers joint detection in the downlink.

3.1 System Model

A diagram of the system model appears in Fig. 2.2 in the previous chapter. The M cochannel signals are assumed to be Q-ary phase-shift keying (PSK) modulated and synchronized by symbol. L-fold antenna diversity is employed at the receiver with the antenna elements spaced far enough apart to ensure independent fading across the array.

For the reader's convenience, the signal models introduced in the previous chapter are repeated here. The mth user's transmitted signal is given by

$$s_m(t) = A_m \sum_n c_m(n)\, p(t - nT) \tag{3.1}$$

where $c_m(n)$ is a PSK data symbol, normalized such that $|c_m(n)|^2 = 1$, $p(t)$ is a root-Nyquist pulse normalized to unity energy, i.e., $\int_{-\infty}^{\infty} p^2(t)\, dt = 1$, T is the symbol period,

[1] © 1998 IEEE. Reprinted, with permission, from *IEEE Transactions on Communications*, vol. 46, pp. 1038–1049, August 1998.

[2] © 1998 IEEE. Reprinted, with permission, from *Proc. IEEE VTC'98*, Ottawa, Canada, May 1998, pp. 1039–1043.

and A_m is related to the average power P_m in the bandpass signal $\widetilde{s}_m(t)$ by $A_m = \sqrt{2P_m}$.

The channels between the multiple users and multiple antennas are described by the $L \times M$ channel gain matrix

$$\mathbf{G}(t) = \begin{bmatrix} g_{11}(t) & g_{12}(t) & \cdots & g_{1M}(t) \\ g_{21}(t) & g_{22}(t) & \cdots & g_{2M}(t) \\ \vdots & \vdots & \ddots & \vdots \\ g_{L1}(t) & g_{L2}(t) & \cdots & g_{LM}(t) \end{bmatrix} \tag{3.2}$$

where the zero-mean complex-Gaussian random variable $g_{lm}(t)$ models flat Rayleigh fading of the channel between the mth user and the lth antenna. Due to independent fading across the antenna array, the elements of the mth column of $\mathbf{G}(t)$, denoted by the channel gain vector $\mathbf{g}_m(t)$, are independent and are assumed to have equal variance $\sigma^2_{g_m}$. Furthermore, the users are assumed to be spaced far enough apart (a few wavelengths) that the users' channel gain vectors are mutually independent.

After matched filtering with the filter $p^*(-t)$ and symbol-rate sampling at times $t = kT$, the vector of received samples is

$$\mathbf{r}(k) = \sum_{m=1}^{M} A_m \mathbf{g}_m(k) c_m(k) + \mathbf{n}(k) \tag{3.3}$$

where the elements of the noise vector $\mathbf{n}(k)$ are i.i.d. complex Gaussian RVs each with variance N_o. The per-branch SNR for user m at the input to the detector is

$$\Gamma_m = \frac{\frac{1}{2}E\left[\left|A_m g_{lm}(k) c_m(k)\right|^2\right]}{\frac{1}{2}E\left[\left|n_l(k)\right|^2\right]} = \frac{A_m^2 \sigma^2_{g_m}}{N_o}. \tag{3.4}$$

To enable detection, a channel estimator provides estimates, denoted $\mathbf{V}(k)$, of the channel gain matrix. Consistent with the previous notation, the channel estimate vector $\mathbf{v}_m(k)$ denotes the mth column of $\mathbf{V}(k)$. The elements of $\mathbf{v}_m(k)$ are assumed to be i.i.d. complex Gaussian RVs all with the same variance $\sigma^2_{v_m}$. To keep the treatment general at the moment, a specific channel estimation scheme is not prescribed; however, in Chapter 5, a practical pilot-based multiuser channel estimation scheme is developed. The detector uses $\mathbf{V}(k)$ along with the received sample vector $\mathbf{r}(k)$ to make a joint decision on all users' symbols using the metric derived in the next section. The output of the detector is the vector of symbol decisions $\widehat{\mathbf{c}}(k)$, which is an estimate of the transmitted joint symbol vector $\mathbf{c}(k) = (c_1(k), c_2(k), \cdots, c_M(k))$.

3.2 Receiver Metric

Let $\{\mathbf{c}_i = (c_{i1}, c_{i2,} \cdots, c_{iM})\}$ be the set of all possible transmitted data vectors where the time dependence of all variables has been dropped for convenience in the subsequent analysis. For M users and a PSK constellation size of Q, the set $\{\mathbf{c}_i\}$ contains Q^M vectors. The joint detection metric is derived using the maximum *a posteriori* (MAP) criterion whereby the detector selects that vector $\mathbf{c}_i$ from the set $\{\mathbf{c}_i\}$ for which the *a posteriori* probability $p(\mathbf{c}_i|\mathbf{r}, \mathbf{V})$ is maximum. Under the assumption of equiprobable data vectors, this is equivalent to maximizing the probability $p(\mathbf{r}|\mathbf{c}_i, \mathbf{V})$.

Since the channel estimates and the true channel gains are jointly Gaussian, $\mathbf{r}$ conditioned on $\mathbf{c}_i$ and $\mathbf{V}$ is Gaussian with conditional mean vector $\boldsymbol{\mu}_{\mathbf{r}|\mathbf{c}_i,\mathbf{V}}$ and conditional covariance matrix $\mathbf{R}_{\mathbf{r}|\mathbf{c}_i,\mathbf{V}}$. Consequently the conditional probability density function (PDF) of $\mathbf{r}$ is

$$p(\mathbf{r}|\mathbf{c}_i, \mathbf{V}) = \frac{1}{(2\pi)^L \left|\mathbf{R}_{\mathbf{r}|\mathbf{c}_i,\mathbf{V}}\right|} \exp\left[-\frac{1}{2}\left(\mathbf{r} - \boldsymbol{\mu}_{\mathbf{r}|\mathbf{c}_i,\mathbf{V}}\right)^{\dagger} \mathbf{R}^{-1}_{\mathbf{r}|\mathbf{c}_i,\mathbf{V}} \left(\mathbf{r} - \boldsymbol{\mu}_{\mathbf{r}|\mathbf{c}_i,\mathbf{V}}\right)\right]. \tag{3.5}$$

The conditional mean vector and covariance matrix are determined through the joint statistics of $\mathbf{g}_m$ and $\mathbf{v}_m$ as follows. First define the vector

$$\mathbf{x}_m = \begin{bmatrix} \mathbf{g}_m \\ \mathbf{v}_m \end{bmatrix} \tag{3.6}$$

which has covariance matrix

$$\mathbf{R}_{\mathbf{x}_m} = \frac{1}{2} E\left[\mathbf{x}_m \mathbf{x}_m^{\dagger}\right] = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^{\dagger} & \mathbf{D} \end{bmatrix} \tag{3.7}$$

where $\mathbf{A} = \frac{1}{2} E\left[\mathbf{g}_m \mathbf{g}_m^{\dagger}\right]$, $\mathbf{B} = \frac{1}{2} E\left[\mathbf{g}_m \mathbf{v}_m^{\dagger}\right]$, and $\mathbf{D} = \frac{1}{2} E\left[\mathbf{v}_m \mathbf{v}_m^{\dagger}\right]$. Due to independent fading across the antenna array, $\mathbf{A}$ and $\mathbf{D}$ are given simply by $\mathbf{A} = \sigma^2_{g_m}\mathbf{I}$ and $\mathbf{D} = \sigma^2_{v_m}\mathbf{I}$ where $\mathbf{I}$ is the $L \times L$ identity matrix. Denoting ρ_m as the correlation coefficient between the channel estimate and the true channel gain on a particular antenna (assumed the same for all antennas), the matrix $\mathbf{B}$ is given simply by $\mathbf{B} = \rho_m \sigma_{g_m} \sigma_{v_m} \mathbf{I}$. Here it is further assumed that the channel estimate on one antenna is uncorrelated with the true gain on another antenna.

It is important to emphasize that the ρ_m's (generally complex quantities with $|\rho_m| \leq 1$) collectively reflect the quality of channel estimation. For the limiting case of perfect channel state information (CSI), ρ_m is equal to unity for all users. Although it is not explicitly

noted, with pilot-based channel estimation schemes such as the one discussed in Chapter 5, ρ_m varies with SNR, tending to unity with very large SNR. This will be discussed in more detail later in this chapter.

Since $\mathbf{g}_m$ and $\mathbf{v}_m$ are jointly Gaussian, it is well known that the MMSE estimate of $\mathbf{g}_m$ from $\mathbf{v}_m$ is given by the conditional mean

$$\boldsymbol{\mu}_{\mathbf{g}_m|\mathbf{v}_m} = \mathbf{B}\mathbf{D}^{-1}\mathbf{v}_m = \beta_m \mathbf{v}_m \tag{3.8}$$

where $\beta_m = \rho_m \sigma_{g_m}/\sigma_{v_m}$. The quality of the MMSE estimator is measured through the conditional covariance matrix

$$\mathbf{R}_{\mathbf{g}_m|\mathbf{v}_m} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{B}^{\dagger} = \sigma^2_{e_m}\mathbf{I} \tag{3.9}$$

where

$$\sigma^2_{e_m} = \left(1 - |\rho_m|^2\right)\sigma^2_{g_m} \tag{3.10}$$

is the variance of each element of the estimation error vector

$$\mathbf{e}_m = \mathbf{g}_m - \beta_m \mathbf{v}_m. \tag{3.11}$$

By the principle of orthogonality [7], the estimation error $\mathbf{e}_m$ and the basis of the estimate $\mathbf{v}_m$ are uncorrelated. Note that for the special case of perfect CSI from the channel estimator, $\sigma^2_{e_m} = 0$ and $\mathbf{v}_m = \mathbf{g}_m$.

The conditional mean $\boldsymbol{\mu}_{\mathbf{r}|\mathbf{c}_i,\mathbf{V}}$ and the conditional covariance matrix $\mathbf{R}_{\mathbf{r}|\mathbf{c}_i,\mathbf{V}}$ in (3.5) may be now be found in terms of (3.8) and (3.9). Using (3.3), the results are

$$\begin{aligned}
\boldsymbol{\mu}_{\mathbf{r}|\mathbf{c}_i,\mathbf{V}} &= E\left[\mathbf{r}|\mathbf{c}_i,\mathbf{V}\right] \\
&= \sum_{m=1}^{M} A_m E\left[\mathbf{g}_m|\mathbf{v}_m\right] c_{im} \\
&= \sum_{m=1}^{M} A_m \boldsymbol{\mu}_{\mathbf{g}_m|\mathbf{v}_m} c_{im}
\end{aligned} \tag{3.12}$$

and

$$\begin{aligned}
\mathbf{R}_{\mathbf{r}|\mathbf{c}_i,\mathbf{V}} &= \frac{1}{2}E\left[\left(\mathbf{r} - \boldsymbol{\mu}_{\mathbf{r}|\mathbf{c}_i,\mathbf{V}}\right)\left(\mathbf{r} - \boldsymbol{\mu}_{\mathbf{r}|\mathbf{c}_i,\mathbf{V}}\right)^{\dagger}|\mathbf{c}_i,\mathbf{V}\right] \\
&= \sum_{m=1}^{M}\sum_{n=1}^{M} A_m A_n \frac{1}{2}E\left[\left(\mathbf{g}_m - \boldsymbol{\mu}_{\mathbf{g}_m|\mathbf{v}_m}\right)\left(\mathbf{g}_n - \boldsymbol{\mu}_{\mathbf{g}_n|\mathbf{v}_n}\right)^{\dagger}|\mathbf{V}\right] c_{im}c^*_{in} + \frac{1}{2}E\left[\mathbf{n}\mathbf{n}^{\dagger}\right] \\
&= \sum_{m=1}^{M} A_m^2 \mathbf{R}_{\mathbf{g}_m|\mathbf{v}_m} + N_o\mathbf{I}.
\end{aligned} \tag{3.13}$$

In the latter equality of (3.13), it has been assumed that the users' channel estimation error vectors (each given by $\mathbf{e}_m = \mathbf{g}_m - \boldsymbol{\mu}_{\mathbf{g}_m|\mathbf{v}_m}$) are mutually uncorrelated. As shown in Chapter 5, this assumption is reasonable as long as the users' training sequences have low cross correlation. For the special case of $M = 2^n$ where n is an integer, suitable training sequences are columns of the Hadamard matrix of order n which are known to be orthogonal, resulting in completely uncorrelated estimation error vectors.

Finally, substituting (3.8)–(3.13) into (3.5) gives the desired PDF as

$$\begin{aligned} p\left(\mathbf{r}|\mathbf{c}_i, \mathbf{V}\right) &= \frac{1}{(2\pi)^L \left(\sum_{m=1}^{M} A_m^2 \sigma_{e_m}^2 + N_o\right)^L} \\ &\quad \cdot \exp\left[-\frac{1}{2} \frac{\sum_{l=1}^{L} \left| r_l - \sum_{m=1}^{M} A_m \beta_m v_{lm} c_{im} \right|}{\sum_{m=1}^{M} A_m^2 \sigma_{e_m}^2 + N_o}\right]. \end{aligned} \tag{3.14}$$

Neglecting hypothesis-independent terms, the appropriate receiver metric is thus

$$\Lambda_i = \sum_{l=1}^{L} \left| r_l - \sum_{m=1}^{M} A_m \beta_m v_{lm} c_{im} \right|^2 \tag{3.15}$$

which must be minimized over the Q^M data vectors in the set $\{\mathbf{c}_i\}$. Clearly, the computational complexity per unit time increases exponentially with the number of users; however, M is likely to be small for a TDMA system, keeping the complexity reasonable.

As can be seen in (3.15), the receiver implicitly performs diversity combining within the metric. To accomplish this, the receiver requires knowledge of the product $A_m v_{lm}$ for every user. Fortunately, this quantity is generated explicitly in a pilot-based channel estimator. The other required parameter, β_m, is determined at design time; however, if the true channel statistics differ from the design statistics, a bias is introduced. Here the bias is assumed to be zero, and the analysis is focused on the random channel estimation errors.

It is interesting to note that for a single user, the metric defined in (3.15) leads to the well known MRC receiver. For $M = 1$, after neglecting hypothesis independent terms, (3.15) reduces to

$$\Lambda_i = \mathrm{Re}\left[c_i^* \sum_{l=1}^{L} r_l v_l^*\right]. \tag{3.16}$$

Evidently, the receiver weights each antenna signal r_l by v_l^*, combines the L signals, and derotates the sum by the various symbol hypotheses to determine the most likely transmitted symbol.

Further insight into the joint detection process may be gained by considering the case of perfect CSI ($\beta_m = 1$ for all users and $\mathbf{V}(k) = \mathbf{G}(k)$). In this case, the metric for flat-fading and symbol-synchronous users in (3.15) may be expressed as

$$\Lambda_i(k) = \left\| \mathbf{r}(k) - \mathbf{G}'(k)\, \mathbf{c}_i^T(k) \right\|^2 \tag{3.17}$$

where $\|\mathbf{x}\|$ is the Euclidean norm (length) of the vector $\mathbf{x}$. The mth column of the matrix $\mathbf{G}'(k)$ is simply the channel gain vector $\mathbf{g}_m(k)$ weighted by A_m. Evidently, the receiver must calculate the distance between the actual received vector $\mathbf{r}(k)$ and the hypothesis $\mathbf{G}'(k)\, \mathbf{c}_i^T(k)$.

For the special case of a single receive antenna ($L = 1$), the distance in (3.17) is equivalent to the Euclidean distance between the received sample $r(k)$ and the ith constellation point $\sum_{m=1}^{M} A_m g_m(k)\, c_{im}(k)$. In Fig. 3.1, the Q^M constellation points are plotted for $A_m = 1$ and arbitrarily selected channel gains. Note that for BPSK, $Q = 2$, and for QPSK, $Q = 4$. The operation of the receiver is to choose the joint symbol vector $\mathbf{c}_i(k)$ corresponding to the constellation point closest in Euclidean distance to the received sample $r(k)$. This process is directly analogous to vector quantization used in source coding for speech and images.

Clearly, the constellation points depend on the set of channel gains which must be estimated. As long as all of the gains are different, it is guaranteed that there will be Q^M distinct points, meaning that all the users may be distinguished. If some of the gains are the same, several points collapse into one another, resulting in decision ambiguities. However, this occurs with zero probability as long as the users are separated by several wavelengths.

Based on the above, intuition suggests that no constraint exists on the number of users that may be supported with respect to the number of antennas. As either M or Q increases, the constellation simply becomes more dense meaning that higher SNR is required to maintain a fixed symbol-error probability, but no orders of diversity are sacrificed. This intuition is confirmed in the next section.

3.3 Performance Analysis

3.3.1 Joint Detection

In this section, a fully-analytical expression for the union bound on the symbol error rate (SER) for the mth user is derived based on the metric in (3.15). Let the transmitted data

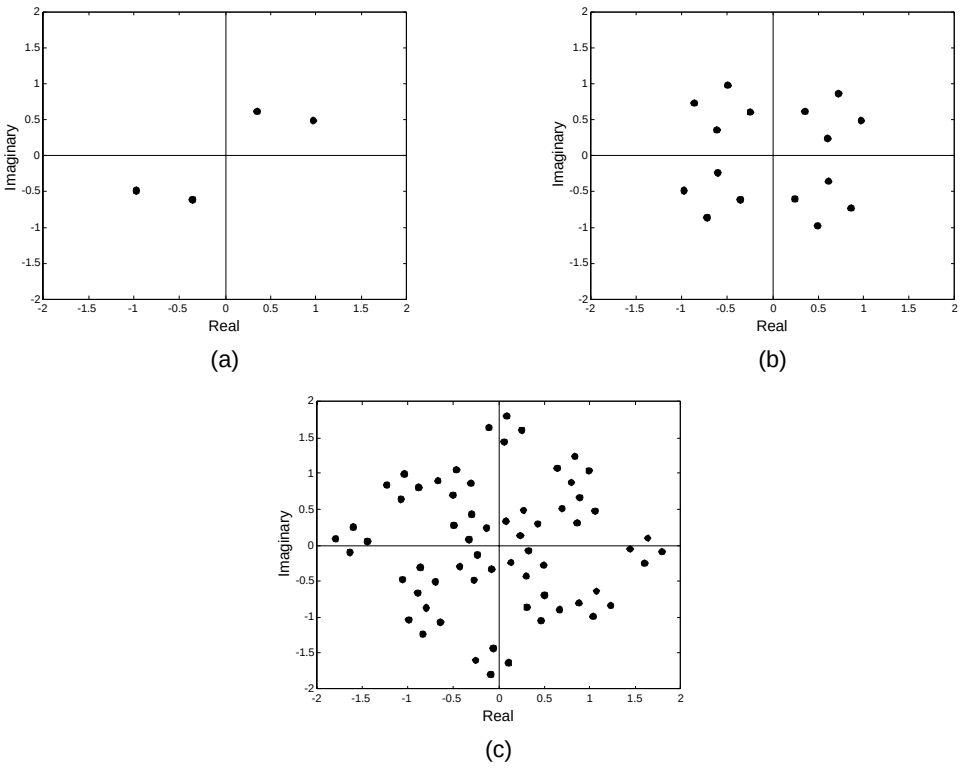

Figure 3.1: Typical signal constellations for a single receive antenna and (a) 2 users/BPSK, (b) 2 users/QPSK, and (c) 3 users/QPSK.

vector be $\mathbf{c}_j = (c_{j1}, c_{j2,} \cdots, c_{jM})$. According to (3.15), the detector chooses the erroneous data vector $\mathbf{c}_i$ over $\mathbf{c}_j$ if $\Lambda_i < \Lambda_j$. The probability of this pairwise error event is

$$P_{2_{ij}} = P\left(D_{ij} < 0 | \mathbf{c}_j\right) \tag{3.18}$$

where $D_{ij} = \Lambda_i - \Lambda_j$. The union bound on the SER for the mth user, given that $\mathbf{c}_j$ is transmitted, is then

$$P_{s_{mj}} \leq \sum_{i \in C_{mj}} P_{2_{ij}} \tag{3.19}$$

where C_{mj} is the subset of data vectors in $\{\mathbf{c}_i\}$ that differ in their mth position from $\mathbf{c}_j$.

The pairwise error probability may be found by expressing the RV D_{ij} as the sum $D_{ij} = \sum_{l=1}^{L} d_{ijl}$ where $d_{ijl} = \mathbf{z}_l^{\dagger} \mathbf{F}_{ij} \mathbf{z}_l$ is a Hermitian quadratic form in zero-mean complex

Gaussian RVs. The vector $\mathbf{z}_l$ is defined as

$$\mathbf{z}_l = (r_l, v_{l1}, v_{l2}, \cdots, v_{lM})^T \tag{3.20}$$

and the Hermitian matrix $\mathbf{F}_{ij}$ as

$$\mathbf{F}_{ij} = \mathbf{u}_i^* \mathbf{u}_i^T - \mathbf{u}_j^* \mathbf{u}_j^T \tag{3.21}$$

where the vectors $\mathbf{u}_i$ and $\mathbf{u}_j$ are given by

$$\begin{aligned} \mathbf{u}_i &= (1, -A_1\beta_1 c_{i1}, -A_2\beta_2 c_{i2}, \cdots, -A_M\beta_M c_{iM})^T \\ \mathbf{u}_j &= (1, -A_1\beta_1 c_{j1}, -A_2\beta_2 c_{j2}, \cdots, -A_M\beta_M c_{jM})^T . \end{aligned} \tag{3.22}$$

Since the characteristic function of the d_{ijl}'s is well known [10], $P_{2_{ij}}$ may be evaluated easily through the characteristic function of D_{ij}. Denoting $f_{D_{ij}}(D)$ as the PDF of D_{ij} and $\Phi_{D_{ij}}(s)$ as the two-sided Laplace transform of $f_{D_{ij}}(D)$ (i.e., the characteristic function), the pairwise error probability is

$$P_{2_{ij}} = \int_{-\infty}^{0} f_{D_{ij}}(D)\, dD = L_{II}^{-1}\left\{\frac{1}{s}\Phi_{D_{ij}}(s)\right\}\Big|_{D=\mathbf{0}} \tag{3.23}$$

where $L_{II}^{-1}\{\bullet\}$ denotes the inverse two-sided Laplace transform.

Due to independent fading across the antenna array, the $\mathbf{z}_l$'s are independent. Consequently, the d_{ijl}'s are independent, and the characteristic function of D_{ij} factors as the product $\Phi_{D_{ij}}(s) = \prod_{l=1}^{L} \phi_{d_{ijl}}(s)$ where $\phi_{d_{ijl}}(s)$ is the characteristic function of d_{ijl}. According to [10, eq. (B-3-21)],

$$\phi_{d_{ijl}}(s) = \frac{1}{\det\left[\mathbf{I} + 2s\mathbf{R}_j\mathbf{F}_{ij}\right]} \tag{3.24}$$

where $\mathbf{I}$ is the $(M+1)\times(M+1)$ identity matrix and $\mathbf{R}_j = \frac{1}{2}E\left[\mathbf{z}_l\mathbf{z}_l^\dagger|\mathbf{c}_j\right]$ is the covariance matrix of $\mathbf{z}_l$ conditioned on the transmit vector $\mathbf{c}_j$ given by

$$\mathbf{R}_j = \begin{bmatrix} \sum_{m=1}^{M} A_m^2\sigma_{g_m}^2 + N_o & A_1\beta_1\sigma_{v1}^2 c_{j1} & A_2\beta_2\sigma_{v2}^2 c_{j2} & \cdots & A_M\beta_M\sigma_{vM}^2 c_{jM} \\ A_1\beta_1^*\sigma_{v1}^2 c_{j1}^* & \sigma_{v1}^2 & 0 & \cdots & 0 \\ A_2\beta_2^*\sigma_{v2}^2 c_{j2}^* & 0 & \sigma_{v2}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_M\beta_M^*\sigma_{vM}^2 c_{jM}^* & 0 & 0 & \cdots & \sigma_{vM}^2 \end{bmatrix}. \tag{3.25}$$

The region of convergence of $\phi_{d_{ijl}}(s)$ is the vertical strip enclosing the $j\omega$ axis bounded by the closest pole on either side. Because the fading statistics on each antenna are identical, $\mathbf{R}_j$ is independent of l, and the characteristic function of D_{ij} is given simply by $\Phi_{D_{ij}}(s) = \left[\phi_{d_{ijl}}(s)\right]^L$.

Using the fact the eigenvectors of a Hermitian matrix are orthonormal, $\Phi_{D_{ij}}(s)$ may be written in such a way to expose the eigenvalues of the Hermitian matrix $2\mathbf{R}_j\mathbf{F}_{ij}$ in (3.24). Let $\mathbf{S}_{ij}$ be a diagonal matrix containing the $M+1$ eigenvalues, with the kth eigenvalue denoted by λ_{ijk}. Also, let $\mathbf{M}_{ij}$ be the modal matrix, i.e., the matrix with the eigenvectors of $2\mathbf{R}_j\mathbf{F}_{ij}$ arranged as columns. Since the eigenvectors are orthonormal, $\mathbf{M}_{ij}\mathbf{M}_{ij}^\dagger = \mathbf{I}$ where $\mathbf{I}$ is the $(M+1)\times(M+1)$ identity matrix. Furthermore, since $2\mathbf{R}_j\mathbf{F}_{ij} = \mathbf{M}_{ij}\mathbf{S}_{ij}\mathbf{M}_{ij}^\dagger$, the characteristic function of D_{ij} may be written as

$$\begin{aligned}\Phi_{D_{ij}}(s) &= \left[\frac{1}{\det\left[\mathbf{M}_{ij}\mathbf{M}_{ij}^\dagger + s\mathbf{M}_{ij}\mathbf{S}_{ij}\mathbf{M}_{ij}^\dagger\right]}\right]^L \\ &= \left[\frac{1}{\det\left[\mathbf{M}_{ij}\left(\mathbf{I}+s\mathbf{S}_{ij}\right)\mathbf{M}_{ij}^\dagger\right]}\right]^L. \end{aligned} \tag{3.26}$$

Since the determinant of a matrix product is equal to the product of the determinants, and since $\det(\mathbf{M}_{ij}) = \det\left(\mathbf{M}_{ij}^\dagger\right) = 1$, the above expression reduces to

$$\Phi_{D_{ij}}(s) = \left[\frac{1}{\prod_{k=1}^{M+1}(1+s\lambda_{ijk})}\right]^L \tag{3.27}$$

which exposes the eigenvalues as desired.

The inversion of (3.23) is made easier by the fact that the $(M+1)\times(M+1)$ matrix $\mathbf{R}_j\mathbf{F}_{ij}$ (and thus $2\mathbf{R}_j\mathbf{F}_{ij}$) has only two nonzero eigenvalues, regardless of M. This may be established under the following argument. Since the covariance matrix $\mathbf{R}_j$ is generally full rank, any limit on the rank of the product $\mathbf{R}_j\mathbf{F}_{ij}$ is imposed by that of $\mathbf{F}_{ij}$. The rank of $\mathbf{F}_{ij}$ can be determined by looking at its null space, i.e., the solutions to the equation $\mathbf{F}_{ij}\mathbf{x} = \mathbf{0}$. Substituting (3.21) into this equation and rearranging gives

$$\mathbf{u}_i\left(\mathbf{u}_i^\dagger\mathbf{x}\right) - \mathbf{u}_j\left(\mathbf{u}_j^\dagger\mathbf{x}\right) = \mathbf{0}. \tag{3.28}$$

For $\mathbf{u}_i$ and $\mathbf{u}_j$ linearly independent, which is indeed the case for different $\mathbf{c}_i$ and $\mathbf{c}_j$, (3.28) is satisfied only for both $\mathbf{u}_i^\dagger\mathbf{x} = 0$ and $\mathbf{u}_j^\dagger\mathbf{x} = 0$. In other words, solutions to (3.28) lie in the

subspace orthogonal to both $\mathbf{u}_i$ and $\mathbf{u}_j$. The dimension of this subspace, i.e., the nullity of $\mathbf{F}_{ij}$, is $M-1$. Since the dimension of $\mathbf{F}_{ij}$ is $M+1$, the rank of $\mathbf{F}_{ij}$ is $(M+1)-(M-1)=2$. Consequently, the rank of the product $\mathbf{R}_j\mathbf{F}_{ij}$ is a maximum of two for all M, implying that $\mathbf{R}_j\mathbf{F}_{ij}$ has only two non zero eigenvalues. Since the PDF of D_{ij} is nonzero for both $D_{ij}<0$ and $D_{ij}\geq 0$, one of the two nonzero eigenvalues is positive, and the other is negative. The convention used here is that λ_{ij1} is positive, and λ_{ij2} is negative.

Modifying (3.27) to take into account the zero eigenvalues gives the desired characteristic function as

$$\Phi_{D_{ij}}(s)=\left[\frac{1}{(1+s\lambda_{ij1})(1+s\lambda_{ij2})}\right]^L. \tag{3.29}$$

This expression may be now be substituted into (3.23) which gives the desired pairwise error probability after inversion. The result, obtained by suitably modifying [8, eq. (4B.7)] which considers a similar inversion, is

$$P_{2_{ij}}=\frac{1}{(1-r_{ij})^{2L-1}}\sum_{k=0}^{L-1}\binom{2L-1}{k}(-r_{ij})^k \tag{3.30}$$

where $r_{ij}=\lambda_{ij1}/\lambda_{ij2}$ is the ratio of the two nonzero eigenvalues of $\mathbf{R}_j\mathbf{F}_{ij}$ (equivalent to the ratio of eigenvalues of $2\mathbf{R}_j\mathbf{F}_{ij}$).

The eigenvalue ratio is found as follows. For convenience, the covariance matrix $\mathbf{R}_j$ in (3.25) is written in an alternate form by expressing the SNR $A_m^2\sigma_{g_m}^2/N_o$ of the mth user as $K_m\Gamma$, where K_m is a scale factor and Γ is a reference SNR. The reference SNR is arbitrary, but logical choices might be the arithmetic or geometric average of the M users' SNRs, or one user's SNR in particular — perhaps that of the strongest user. Making use of the relation $\beta_m=\rho_m\sigma_{g_m}/\sigma_{v_m}$, the alternate form for $\mathbf{R}_j$ in terms of the reference SNR is

$$\mathbf{R}_j=\mathbf{N}_o\begin{bmatrix} 1+\sum_{m=1}^{M}K_m\Gamma & \frac{|\rho_1|^2}{A_1\beta_1^*}c_{j1}^*K_1\Gamma & \frac{|\rho_2|^2}{A_2\beta_2^*}c_{j2}^*K_2\Gamma & \cdots & \frac{|\rho_M|^2}{A_M\beta_M^*}c_{jM}^*K_M\Gamma \\ \frac{|\rho_1|^2}{A_1\beta_1}c_{j1}^*K_1\Gamma & \frac{|\rho_1|^2}{A_1^2|\beta_1|^2}K_1\Gamma & 0 & \cdots & 0 \\ \frac{|\rho_2|^2}{A_2\beta_2}c_{j2}^*K_2\Gamma & 0 & \frac{|\rho_2|^2}{A_2^2|\beta_2|^2}K_2\Gamma & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{|\rho_M|^2}{A_M\beta_M}c_{jM}^*K_M\Gamma & 0 & 0 & 0 & \frac{|\rho_M|^2}{A_M^2|\beta_M|^2}K_M\Gamma \end{bmatrix}. \tag{3.31}$$

Furthermore, the expression for the matrix $\mathbf{F}_{ij}$ in (3.21) is expanded and written in the

alternate form $\mathbf{F}_{ij} = \mathbf{U}_{ij} + \mathbf{U}_{ij}^{\dagger}$ where

$$\mathbf{U}_{ij} = \begin{bmatrix} 0 & A_1\beta_1\left(c_{j1} - c_{i1}\right) & A_2\beta_2\left(c_{j2} - c_{i2}\right) & \cdots & A_M\beta_M\left(c_{jM} - c_{iM}\right) \\ 0 & 0 & \begin{array}{c} A_1A_2\beta_1^*\beta_2 \\ \cdot\left(c_{i1}^*c_{i2} - c_{j1}^*c_{j2}\right) \end{array} & \cdots & \begin{array}{c} A_1A_M\beta_1^*\beta_M \\ \cdot\left(c_{i1}^*c_{iM} - c_{j1}^*c_{jM}\right) \end{array} \\ 0 & 0 & 0 & \cdots & \begin{array}{c} A_2A_M\beta_2^*\beta_M \\ \cdot\left(c_{i2}^*c_{iM} - c_{j2}^*c_{jM}\right) \end{array} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{3.32}$$

In the following, it is assumed that the transmitted data vector is the all-ones vector $\mathbf{c}_j = (1, 1, \cdots, 1)$. Consequently, the subscript j is dropped from all variables. No loss in generality is incurred, since the SER does not depend on which data vector is actually transmitted. Since the rank of the matrix $\mathbf{RF}_i$ is only two, the characteristic polynomial of the matrix product may be expressed in the following form

$$\det\left[\mathbf{RF}_i - \lambda\mathbf{I}\right] = \lambda^{M-1}\left(\lambda^2 + \alpha_{i1}\lambda + \alpha_{i2}\right) = 0. \tag{3.33}$$

According to Bôcher's formula [47], the coefficients of the quadratic polynomial are given by $\alpha_{i1} = -T_{i1}$ and $\alpha_{i2} = -\frac{1}{2}\left(\alpha_{i1}T_{i1} + T_{i2}\right)$ where $T_{ik} = \text{trace}\left[\left(\mathbf{RF}_i\right)^k\right]$. Knowing the coefficients allows one to calculate the ratio of the two nonzero eigenvalues as

$$r_i = \frac{T_{i1} + \sqrt{2T_{i2} - T_{i1}^2}}{T_{i1} - \sqrt{2T_{i2} - T_{i1}^2}}. \tag{3.34}$$

Evaluation of the various traces is made easier if the matrix $\mathbf{R}$ is written as the sum of two simpler matrices $\mathbf{R}_1$ and $\mathbf{R}_2$, where $\mathbf{R}_1$ contains the first column and first row of $\mathbf{R}$ with the remainder of the elements set to zero, and $\mathbf{R}_2$ contains the main diagonal of $\mathbf{R}$ (excluding the top left element) with the remainder of the elements set to zero. T_{i1} and T_{i2} are then given by

$$\begin{aligned} T_{i1} &= \text{trace}\left[\left(\mathbf{R}_1 + \mathbf{R}_2\right)\left(\mathbf{U}_i + \mathbf{U}_i^{\dagger}\right)\right] \\ T_{i2} &= \text{trace}\left[\left(\mathbf{R}_1 + \mathbf{R}_2\right)\left(\mathbf{U}_i + \mathbf{U}_i^{\dagger}\right)\left(\mathbf{R}_1 + \mathbf{R}_2\right)\left(\mathbf{U}_i + \mathbf{U}_i^{\dagger}\right)\right]. \end{aligned} \tag{3.35}$$

The above expressions require evaluation of the traces of several matrix products: four two-fold products in the former, and 16 four-fold products in the latter. Fortunately, the

sparse nature and special form of the matrices make some of the traces zero and limit the complexity of others. The intermediate calculations are not shown because of their size.

After laborious evaluation of the nonzero traces using $c_{jm} = 1$ for all m in (3.31) and (3.32), the eigenvalue ratio is

$$r_i = \frac{a_i\Gamma + \sqrt{b_i\Gamma^2 + 2a_i\Gamma}}{a_i\Gamma - \sqrt{b_i\Gamma^2 + 2a_i\Gamma}} \tag{3.36}$$

where

$$\begin{aligned} a_i &= \sum_{m=1}^{M} |\rho_m|^2 K_m (1 - \mathrm{Re}\,[c_{im}]) \\ b_i &= 2\sum_{n=1}^{M}\sum_{m=1}^{M} |\rho_m|^2 K_n K_m (1 - \mathrm{Re}[c_{im}]) \\ &\quad + \left(\sum_{m=1}^{M} |\rho_m|^2 K_m \mathrm{Re}[c_{im}]\right)^2 - \left(\sum_{m=1}^{M} |\rho_m|^2 K_m\right)^2 . \end{aligned} \tag{3.37}$$

In the derivation, explicit use is made of the relation $\mathrm{Re}[c_{im}]^2 + \mathrm{Im}[c_{im}]^2 = 1$ for PSK. Substituting the above results into (3.30) and then (3.19) gives the following closed-form analytical solution for the union bound on the SER for the mth user:

$$P_{s_m} \leq \sum_{i \in C_m} \frac{1}{(1 - r_i)^{2L-1}} \sum_{k=0}^{L-1} \binom{2L-1}{k} (-r_i)^k \tag{3.38}$$

where C_m is the subset of data vectors that differ in their mth position from the length-M all-ones vector.

Two immediate inferences can be drawn from this result: first, the pairwise error probability depends only on the erroneous data vector $\mathbf{c}_i$ and each user's own SNR $\Gamma_m = K_m\Gamma$ and channel estimate correlation coefficient ρ_m. It does not depend separately on A_m or β_m.

Second, in the special case of equipower users and perfect CSI, an interesting interpretation of SNR may be obtained. With perfect CSI, $\rho_m = 1$ for all users, and the parameter b_i in (3.37) simplifies to $b_i = a_i^2$. If the cochannel signals arrive at the base station with equal average powers, perhaps as a result of power control, then $K_m = 1$ for all users, and the parameter a_i in (3.37) simplifies to $a_i = \sum_{m=1}^{M}(1 - \mathrm{Re}[c_{im}])$. Consequently, from (3.36) one can see that the eigenvalue ratio, and therefore the SER, depend only on an effective

SNR

$$\Gamma_{i,eff} = \Gamma \sum_{m=1}^{M} (1 - \mathrm{Re}[c_{im}]). \tag{3.39}$$

Further, if the modulation is BPSK, then a_i becomes twice the Hamming distance $d_{i,H}$ between $\mathbf{c}_j$ and $\mathbf{c}_i$ (the number of positions in which the transmitted and erroneous data vectors differ), since the transmitted vector $\mathbf{c}_j$ has been taken to be all ones. Thus, the pairwise error probability depends only on the effective SNR $\Gamma_{i,eff} = 2d_{i,H}\Gamma$: the larger the Hamming distance, the larger the effective SNR, and the smaller the pairwise error probability.

As expected, for the special case of a single user, BPSK, and perfect CSI, (3.38) reduces to the result for the exact BER (rather than the union bound) for the MRC receiver. In this case, the set of error vectors has only one member ($c_{i1} = -1$) with corresponding eigenvalue ratio

$$r = \frac{1 + \sqrt{1 + \Gamma^{-1}}}{1 - \sqrt{1 + \Gamma^{-1}}}. \tag{3.40}$$

The resulting BER is thus

$$P_b = \frac{1}{(1-r)^{2L-1}} \sum_{k=0}^{L-1} \binom{2L-1}{k} (-r)^k \tag{3.41}$$

which is equivalent to [8, eq. (7.4.15)].

Much insight into the behaviour of joint detection may be gained by considering its asymptotic performance, that is, the performance as the common SNR Γ becomes large. One would hope to see the eigenvalue ratio (3.36) continue to increase with Γ in order to drive the pairwise error probability (3.30) toward zero. The questions are whether the ratio does increase without limit and, if so, how quickly. These questions are answered in the following two sections which consider two different models for the channel estimation correlation coefficients: fixed and variable with SNR.

Fixed Correlation Coefficient Error Floor

In one model of imperfect CSI, the correlation coefficients ρ_m of channel estimation are fixed, and do not improve with increasing SNR. In this case, joint detection exhibits an error floor (an irreducible SER for large SNR). The floor may be easily quantified by noting

that a_i and b_i remain constant, so that the eigenvalue ratio (3.36) does not continue to increase with increasing Γ, but reaches a finite limit

$$r_{i,floor} = \frac{a_i + \sqrt{b_i}}{a_i - \sqrt{b_i}}. \tag{3.42}$$

Substitution of $r_{i,floor}$ into (3.38) gives the union bound on the SER floor. As a check, recall that perfect CSI with all $\rho_m = 1$ results in $b_i = a_i^2$ (even for nonequipower users). Consequently, $r_{i,floor}$ becomes infinite, and the SER floor becomes zero, as expected.

One can use (3.42) to illustrate and quantify the interdependence of detection of data from different mobiles. Suppose that the first $M-1$ mobile signals are received with perfect CSI, so $\rho_m = 1$ for $m = 1 \ldots M-1$, but that channel estimation for mobile M is flawed, so $|\rho_M| < 1$. For simplicity, consider equipower signals, so that all $K_m = 1$. Define $d'_{i,H} = \frac{1}{2}\sum_{m=1}^{M-1}(1 - \mathrm{Re}[c_{im}])$ and $d'_{M,H} = \frac{1}{2}(1 - \mathrm{Re}[c_{iM}])$, although these definitions can no longer be interpreted as Hamming distances for general PSK. Then, from (3.37)

$$\begin{aligned} a_i &= 2d'_{i,H} + 2|\rho_M|^2 d'_{M,H} \\ b_i &= a_i^2 + 4\left(d'_{i,H} - |\rho_M|^2 (d'_{i,H} - d'_{M,H}) - |\rho_M|^4 d'_{M,H}\right). \end{aligned} \tag{3.43}$$

Now consider the error rate of one of the mobiles with perfect CSI: the most likely error event is the one in which all c_{im} are 1 except for that of the designated mobile, so $0 < d'_{i,H} \leq 1$ and $d'_{M,H} = 0$. Clearly, the denominator of (3.42) is not zero and $r_{i,floor}$ remains finite at $\left(1 + \sqrt{1 - |\rho_M|^2 / d'_{i,H}}\right) / \left(1 - \sqrt{1 - |\rho_M|^2 / d'_{i,H}}\right)$. Consequently, the imperfect channel estimation for a single user's signal produces an error floor for all users, even though there is effectively no receiver noise and most users enjoy perfect channel estimation. Further, their error floor is no less than that experienced by the user with imperfect estimation (obtained by $d'_{i,H} = 0$ and $0 < d'_{M,H} \leq 1$).

Pilot-Based Channel Estimation

An alternative model of imperfect CSI allows the channel estimation correlation coefficients ρ_m to improve as the SNR increases, thereby eliminating the error floor. For example, with the pilot-based multiuser channel estimation scheme developed in Chapter 5, it is shown that the mth user's normalized channel estimation error variance $\sigma^2_{e_m}/\sigma^2_{g_m} = 1 - |\rho_m|^2$ is inversely proportional to that user's own SNR $\Gamma_m = K_m\Gamma$ when Γ becomes large. The constant of proportionality, denoted here as k_m, depends on the number of users M as well

as the parameters of the channel estimation scheme such as interpolator order, frame length, Doppler fade rate, and the choice of training sequences. Typical values for k_m are on the order of 10^{-1}. With this model of variation, $|\rho_m|^2 = 1 - (k_m/K_m)\,\Gamma^{-1}$. Substitution into (3.37) and (3.36) gives the eigenvalue ratio in the following alternative form after collecting terms with like powers of Γ:

$$r_i = \frac{\alpha_i\Gamma + \beta_i + \sqrt{\alpha_i^2\Gamma^2 + \gamma_i\Gamma + \delta_i}}{\alpha_i\Gamma + \beta_i - \sqrt{\alpha_i^2\Gamma^2 + \gamma_i\Gamma + \delta_i}} \tag{3.44}$$

where the coefficients are

$$\begin{aligned}
\alpha_i &= \sum_{m=1}^{M} K_m\left(1 - \mathrm{Re}\left[c_{im}\right]\right) \\
\beta_i &= -\sum_{m=1}^{M} k_m\left(1 - \mathrm{Re}\left[c_{im}\right]\right) \\
\gamma_i &= 2\left(1 + \sum_{m=1}^{M} k_m\,\mathrm{Re}\left[c_{im}\right]\right)\sum_{m=1}^{M} K_m\left(1 - \mathrm{Re}\left[c_{im}\right]\right) \\
\delta_i &= \sum_{m=1}^{M} k_m\,\mathrm{Re}\left[c_{im}\right]\left(2 + \sum_{m=1}^{M} k_m\,\mathrm{Re}\left[c_{im}\right]\right) - \sum_{m=1}^{M} k_m\left(2 + \sum_{m=1}^{M} k_m\right).
\end{aligned} \tag{3.45}$$

To obtain the asymptotic error rate, let the SNR Γ in (3.44) become very large. The eigenvalue ratio approaches

$$\begin{aligned}
r_{i,asymp} &= \frac{4\alpha_i^2}{2\alpha_i\beta_i - \gamma_i}\Gamma \\
&= \frac{-2\sum_{m=1}^{M} K_m\left(1 - \mathrm{Re}\left[c_{im}\right]\right)}{1 + \sum_{m=1}^{M} k_m}\Gamma
\end{aligned} \tag{3.46}$$

which increases without bound as Γ increases. As a check, the special case of equipower users, perfect CSI, and BPSK modulation produces $r_{i,asymp} = -4d_{i,H}\Gamma$, which is identical to the result obtained by setting $a_i = 2d_{i,H}$ and $b_i = a_i^2$ in (3.36) and allowing Γ to approach infinity.

The pairwise error probability is obtained by substituting (3.46) into (3.30) and noting that the asymptotic form of the result is determined by the last term in the summation, giving $P_{2_i,asymp} = \binom{2L-1}{L-1}\left(-r_{i,asymp}\right)^{-L}$. Substitution of (3.46) and summation over all error

events yields, finally,

$$P_{s_m,asymp} \leq \left[2^{-L} \binom{2L-1}{L-1} \left(1 + \sum_{m=1}^{M} k_m \right)^L \sum_{i \in C_m} \left(\sum_{m=1}^{M} K_m \left(1 - \mathrm{Re}\left[c_{im} \right] \right) \right)^{-L} \right] \Gamma^{-L}. \tag{3.47}$$

This analytical result shows clearly that the union bound on SER varies asymptotically as Γ^{-L}, regardless of the number of users. That is, all users enjoy L-fold diversity with no error floor. Evidently, the cost of supporting additional users at a fixed error rate is only an increase in required SNR (quantified below), rather than losses in diversity order. This behaviour occurs for perfect CSI as well where $k_m = 0$ for all m. These fundamental results, not observed previously, are extremely valuable in a fading environment and lead to large system capacity gains as shown in Chapter 6. The new mode of reception behaviour observed here is in stark contrast to MMSE combining where the maximum number of users is limited to the number of antennas and the diversity order is only $L - M + 1$ for each user.

Recall that the quality of multiuser pilot-based channel estimation is expressed by the coefficients k_m. It is clear from (3.47), that imperfect channel estimation degrades the SNR by an asymptotic factor of $1 + \sum_{m=1}^{M} k_m$ for every error pairwise error event. This is another illustration of the interdependence of data detection for different users, since any user with poor channel estimation (a large value of k_m) degrades the performance of all users, even those with perfect CSI ($k_m = 0$).

The performance degradation with additional users is quantified by the additional SNR required to support M users at a fixed symbol-error rate in the asymptotic region, compared with that required to support a single user at the same error rate. Here an expression is derived for the SNR penalty experienced by a system with equipower users and perfect CSI ($K_m = 1$ and $k_m = 0$). In this case, the asymptotic SER is

$$P_{s,asymp} = \left[2^{-L} \binom{2L-1}{L-1} \sum_{i \in C_1} \left(\sum_{m=1}^{M} \left(1 - \mathrm{Re}\left[c_{im} \right] \right) \right)^{-L} \right] \Gamma_M^{-L}. \tag{3.48}$$

The subscript m is dropped since all users have identical SER which allows one to consider only C_1, i.e., the set of vectors that differ from unity in the first position. The SNR has been denoted Γ_M to emphasize the number of users. For the reference single-user system with Q points in the PSK constellation, $i \in \{1, 2, \cdots, Q-1\}$ and $\mathrm{Re}[c_{i1}] = \cos[2\pi i/Q]$, resulting in the simple expression $\left[2^{-L} \binom{2L-1}{L-1} \sum_{i=1}^{Q-1} \left(1 - \cos\left[2\pi i/Q \right] \right)^{-L} \right] \Gamma_1^{-L}$ for asymptotic SER.

Define $U_M = \Gamma_M/\Gamma_1$ as the penalty factor by which the SNR must be increased to maintain the same SER as a single-user system. From (3.48), the SNR penalty is

$$U_M = \left[\frac{\sum_{i \in C_1} \left(\sum_{m=1}^{M} \left(1 - \mathrm{Re}\left[c_{im}\right]\right) \right)^{-L}}{\sum_{i=1}^{Q-1} \left(1 - \cos\left[2\pi i/Q\right]\right)^{-L}} \right]^{1/L} . \tag{3.49}$$

This expression will be used to investigate the degradation with increasing M and Q in Section 3.4.

3.3.2 MMSE Combining

For comparison purposes, a quasi-analytical expression for the BER of the MMSE combining receiver is derived here. The combiner structure is shown in Fig. 2.4 in the previous chapter, but unlike in that chapter, here the weight vectors are derived and the performance is evaluated for the general case of imperfect CSI.

For MMSE combining with channel estimates, the antenna weights for the mth user are chosen so as to minimize the mean squared error between the transmitted symbol $c_m(k)$ and the combiner output $y_m(k)$ given the matrix of channel estimates $\mathbf{V}(k)$. The MSE for a given weight vector $\mathbf{w}_m(k)$ is

$$\varepsilon_m(k) = \frac{1}{2} E\left[\left| c_m(k) - \mathbf{w}_m^{\dagger}(k)\, \mathbf{r}(k) \right|^2 | \mathbf{V}(k) \right] \tag{3.50}$$

where the expectation is taken over the noise ensemble, the ensemble of channel estimation errors, and the symbol ensembles of all users. In a similar fashion to Section 2.1.2 of the previous chapter, the optimal weight vector with channel estimates is given by $\mathbf{w}_{m,o} = \mathbf{R}^{-1}\mathbf{p}_m$, where $\mathbf{p}_m = \frac{1}{2}E\left[c_m^* \mathbf{r} | \mathbf{V}\right]$ and $\mathbf{R} = \frac{1}{2}E\left[\mathbf{r}\mathbf{r}^{\dagger} | \mathbf{V}\right]$. The difference here is that $\mathbf{p}_m$ and $\mathbf{R}$ are conditioned on the channel estimate matrix $\mathbf{V}$. Note that the time variation of all quantities has been dropped for compactness of notation.

In more detail, the cross correlation vector $\mathbf{p}_m$, using (3.8), is

$$\begin{aligned} \mathbf{p}_m &= \frac{1}{2} E\left[c_m^* \left(\sum_{n=1}^{M} A_n c_n \mathbf{g}_n + \mathbf{n} \right) | \mathbf{V} \right] \\ &= \frac{1}{2} A_m E\left[\mathbf{g}_m | \mathbf{v}_m\right] \\ &= \frac{1}{2} A_m \beta_m \mathbf{v}_m \end{aligned} \tag{3.51}$$

where it has been assumed that the users' symbols are mutually uncorrelated and that the symbols and noise are uncorrelated. Under the same assumptions, the covariance matrix $\mathbf{R}$, using (3.10) and (3.11), is

$$\begin{aligned}\mathbf{R} &= \frac{1}{2}E\left[\sum_{n=1}^{M}\sum_{m=1}^{M}A_nA_mc_mc_n^*\mathbf{g}_n\mathbf{g}_m^\dagger|\mathbf{V}\right]+\frac{1}{2}E\left[\mathbf{n}\mathbf{n}^\dagger\right]\\ &= \sum_{n=1}^{M}A_n^2\frac{1}{2}E\left[\left(\beta_n\mathbf{v}_n+\mathbf{e}_n\right)\left(\beta_n\mathbf{v}_n+\mathbf{e}_n\right)^\dagger|\mathbf{v}_n\right]+N_o\mathbf{I}\\ &= \sum_{n=1}^{M}A_n^2\left(\frac{1}{2}\left|\beta_n\right|^2\mathbf{v}_n\mathbf{v}_n^\dagger+\sigma_{e_n}^2\mathbf{I}\right)+N_o\mathbf{I}\end{aligned} \tag{3.52}$$

where $\mathbf{I}$ is the $L\times L$ identity matrix. In turn, the optimal weight vector is

$$\mathbf{w}_{m,o}=A_m\beta_m\left[\sum_{n=1}^{M}A_n^2\left(\left|\beta_n\right|^2\mathbf{v}_n\mathbf{v}_n^\dagger+2\sigma_{e_n}^2\mathbf{I}\right)+2N_o\mathbf{I}\right]^{-1}\mathbf{v}_m. \tag{3.53}$$

Equation (3.53) generalizes the weight equation derived in [12] to include the case of imperfect channel estimation. For the special case of perfect channel estimation ($\sigma_{e_m}^2=0$ and $\mathbf{v}_m=\mathbf{g}_m$) (3.53) and [12, eq. (9)] are equivalent.

For the case of BPSK modulation, the probability of bit error for the m-th user, denoted P_{b_m}, is

$$P_{b_m}=P\left[\mathrm{Re}\left[y_m\right]<0|c_m=+1\right]. \tag{3.54}$$

Unfortunately, the PDF of the combiner output y_m is difficult to obtain for $M>1$ due to the matrix inversion in (3.53). Thus, simulation is required in order to determine bit error rates for arbitrary M. For $M=1$, however, it can be shown that by using the matrix inversion lemma [4], (3.53) reduces to the weight vector for the MRC receiver with bit error probability given by (3.41).

Before resorting to simulation, some progress can be made towards a quasi-analytical bit-error rate expression by recognizing that the PDF of the random variable y_m is Gaussian when conditioned on both the channel estimate matrix $\mathbf{V}$ and the joint symbol vector $\mathbf{c}=(c_1,c_2,\cdots,c_M)$. To see this, let $y_m=\mathbf{w}_m^\dagger\mathbf{r}=x+z$ where, using (3.3) and (3.11),

$$x=\mathbf{w}_m^\dagger\sum_{n=1}^{M}A_n\beta_nc_n\mathbf{v}_n \tag{3.55}$$

and

$$z = \mathbf{w}_m^\dagger \left(\sum_{n=1}^{M} A_n c_n \mathbf{e}_n + \mathbf{n} \right). \tag{3.56}$$

Observing (3.53), $\mathbf{w}_m$ is deterministic for a given $\mathbf{V}$. Thus, for a given $\mathbf{c}$, x is deterministic as well, and z is Gaussian since both $\mathbf{e}_m$ and $\mathbf{n}$ are Gaussian. Furthermore, z is zero-mean. Therefore, given $\mathbf{V}$ and $\mathbf{c}$ with $c_m = +1$, the conditional BER is the probability that the real part of y_m goes negative, given by

$$P\left[\mathrm{Re}\left[y_m\right] < 0 | \mathbf{c}, \mathbf{V}\right] = Q\left(\frac{\mathrm{Re}\left[x\right]}{\sigma_z}\right) \tag{3.57}$$

where $Q(\cdot)$ is the Gaussian Q-function, and the conditional variance of z is

$$\sigma_z^2 = \left(\sum_{n=1}^{M} A_n^2 \sigma_{e_n}^2 + N_o \right) \mathbf{w}_m^\dagger \mathbf{w}_m. \tag{3.58}$$

The average probability of bit error for user m follows by substituting (3.55) and (3.58) in (3.57) and taking the expectation over the ensemble of interfering users' symbols and the ensemble of channel estimates:

$$P_{b_m} = E\left[Q\left(\frac{\mathrm{Re}\left[\mathbf{w}_m^\dagger \sum_{n=1}^{M} A_n \beta_n c_n \mathbf{v}_n\right]}{\sqrt{\left(\sum_{n=1}^{M} A_n^2 \sigma_{e_n}^2 + N_o\right) \mathbf{w}_m^\dagger \mathbf{w}_m}} \right) \right] \tag{3.59}$$

where it is understood that $c_m = +1$. In this way, the simulation accuracy is increased by performing the average over the noise and channel estimation error ensembles analytically.

Using (3.59), the BER for user m is determined through Monte Carlo simulation by the following method: at each iteration generate an $L \times M$ matrix $\mathbf{V}$ of independent zero-mean complex Gaussian random samples, with the elements in column m having variance $\sigma_{v_m}^2$; calculate the weight vector $\mathbf{w}_m$ using (3.53) along with (3.10); then average the Q-function in (3.59) over all possible $\mathbf{c}$'s with $c_m = +1$.

3.4 Performance Results

In this section, numerical results are provided that follow from the analysis in Section 3.3. Specifically, the results highlight the following: 1) joint detection of equipower signals with perfect CSI; 2) joint detection of equipower signals with imperfect CSI; 3) comparison

of joint detection and MMSE combining; and 4) joint detection of nonequipower signals. As discussed before, perfect CSI is modeled by setting the channel estimation correlation coefficient ρ_m to unity for all users; values of ρ_m less than unity imply imperfect CSI with relative estimation error variance $\sigma^2_{e_m}/\sigma^2_{g_m} = 1 - |\rho_m|^2$.

For the case of equipower users, the SNR for all users is the same; i.e., $K_m = 1$ and $\Gamma_m = \Gamma$ for all m. For all results concerning equipower users, the quality of channel estimation (measured by ρ_m) is assumed to be the same for all users. Thus, according to (3.38), all users have the same symbol-error rate, and only the SER of user 1 will be plotted.

3.4.1 Equipower Signals and Perfect CSI

Fig. 3.2 shows the performance of joint detection of equipower signals with perfect CSI. As can be seen, each additional user causes an incremental degradation in performance with no loss in diversity order. In other words, each user experiences L-fold diversity, regardless of the number of users. Comparing the two sets of curves reveals that the incremental degradation in performance is reduced by the use of additional antennas. This will be investigated more fully in a later graph.

The above behaviour is consistent with that predicted earlier; namely, as the number of users increases, the signal constellations shown in Fig. 3.1 become more dense requiring a larger SNR to maintain a fixed error rate. Furthermore, the number of users that may be supported is not constrained by the number of antennas. Clearly, joint detection can be applied successfully even with only a single antenna.

The diversity performance is further confirmed in Fig. 3.3 which compares the true union bound, obtained analytically using (3.38), with the asymptotic form, obtained using (3.47). Note that in (3.47), perfect CSI is modeled by setting all k_m to zero. Again, regardless of the number of users, all users experience L-fold diversity and the degradation in performance with additional users is reduced by the use of additional antennas.

The tightness of the union bound in (3.38) is investigated in Fig. 3.4 which compares (3.38) with the BER obtained by Monte Carlo simulation. Evidently, the bound is asymptotically tight with increasing SNR. As the number of users increases, a given accuracy of the bound requires somewhat higher SNR; however, the accuracy is quite satisfactory for normal values of BER. As for increasing the number of PSK constellation points Q, the bound becomes loose, like the union bound for Q-ary PSK in conventional single-user operation. In such cases, a better approximation can be obtained by considering only the

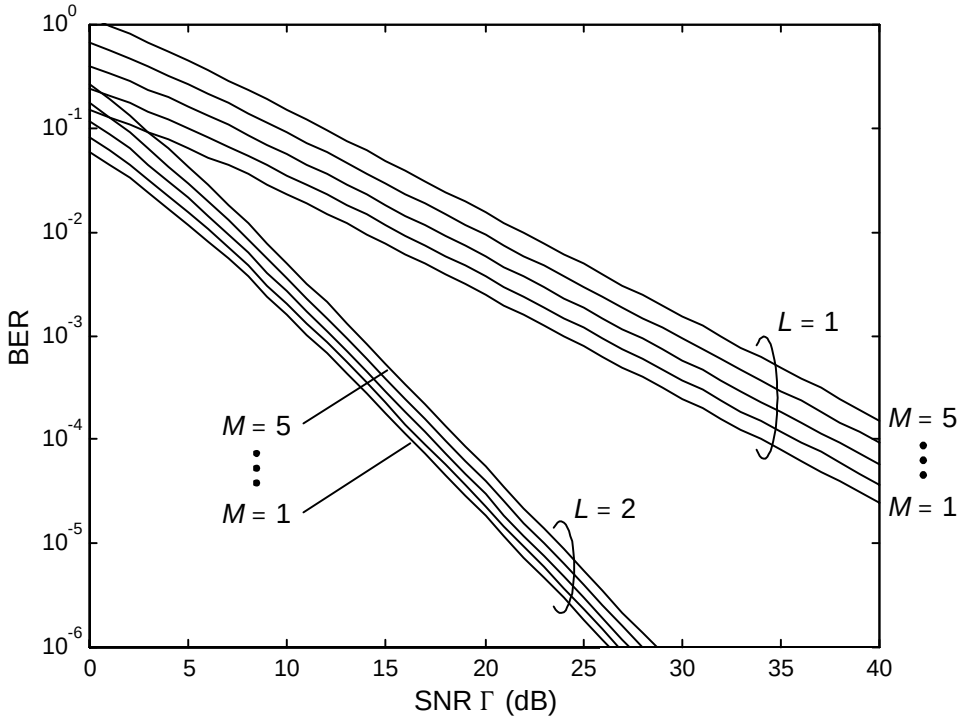

Figure 3.2: Performance of joint detection of M equipower BPSK signals using L antennas with perfect CSI.

dominant error events.

The simulation results in Fig. 3.4 are obtained as follows: at each iteration, generate an $M \times 1$ vector $\mathbf{g}$ of independent, zero-mean, complex Gaussian random samples each with variance σ_g^2, as well a single zero-mean, complex Gaussian random sample n with variance N_o; calculate the value of the received sample r using (3.3), assuming the transmitted symbols are all $+1$ and $A_m = \sqrt{2P_m}$ where $P_m = 1$ for all users; calculate the metric in (3.15) for all possible values of the hypothesized symbol vector $\mathbf{c}_i$ for the case of a single antenna ($L = 1$) and perfect CSI ($\beta_m = 1$ and $\mathbf{v} = \mathbf{g}$); choose the symbol vector corresponding to the smallest calculated metric; and finally, count an error if $\mathbf{c}_{i1} = -1$. After N iterations, the estimated BER (the same for all equipower users) is simply the number of counted errors divided by N.

The performance degradation with additional users is quantified by the SNR penalty in the asymptotic BER region using (3.49). A plot of (3.49) in Fig. 3.5 shows that for

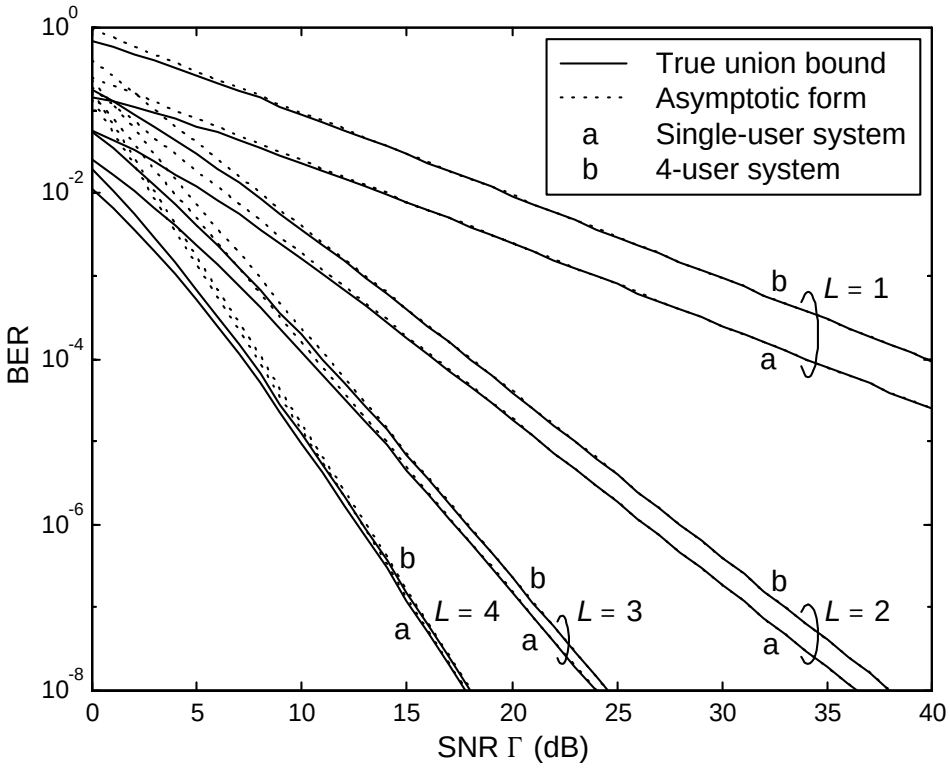

Figure 3.3: Comparison of the true union bound on BER and its asymptotic form for equipower BPSK signals and perfect CSI.

a given number of antennas L, there is a constant penalty in dB for each additional user. Moreover, as the number of antennas increases, the degradation per additional user decreases dramatically. For example, with a single antenna, the degradation is approximately 2 dB per additional user, whereas it is less than 0.1 dB per additional user with 4 antennas. Thus, not only do multiple antennas give better performance for a single user, they maintain that performance better than a single antenna in the face of an increasing user population.

Fig. 3.6 plots the SNR penalty for three different modulation formats, namely BPSK, QPSK, and 8-PSK. Note that for a given number of constellation points Q, the SNR penalty is always referenced to a single-user system with the same Q. As can be seen, increasing the constellation size results in a significant degradation in performance, especially for the 4-user system. More importantly though, the degradation is quickly reduced as the number of antennas is increased. In fact, with 3 antennas, the performance degradation is less than

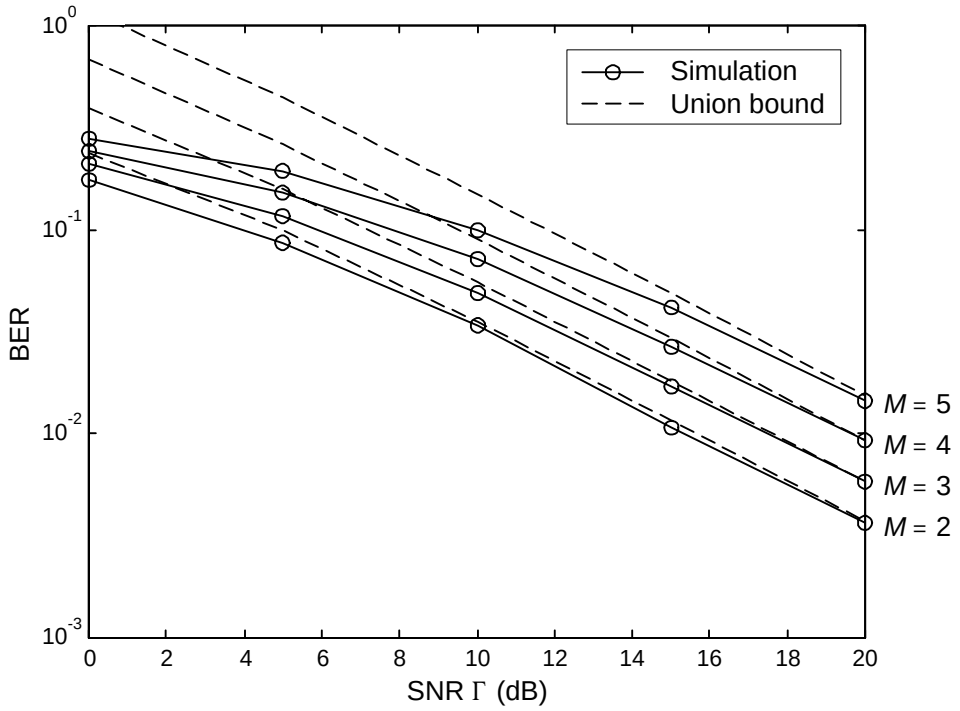

Figure 3.4: Tightness of the union bound on BER for the joint detection of equipower BPSK signals with a single antenna and perfect CSI.

2 dB for all three modulation formats and for both the 2- and 4-user systems.

3.4.2 Equipower Signals and Imperfect CSI

In this section, the effect of imperfect CSI on joint detection is investigated using two different models for the channel estimation correlation coefficient ρ_m. The first model considers ρ_m to be fixed which introduces an error floor. The second model considers ρ_m to vary with SNR as is the case with pilot-based channel estimation.

Fixed Correlation Coefficient

Fig. 3.7 demonstrates the effect of imperfect CSI on joint detection when the channel estimation correlation coefficient is fixed at $\rho_m = 0.995$. This value results in a relative channel estimation error variance of approximately 1%. In the low SNR region, the imperfect CSI

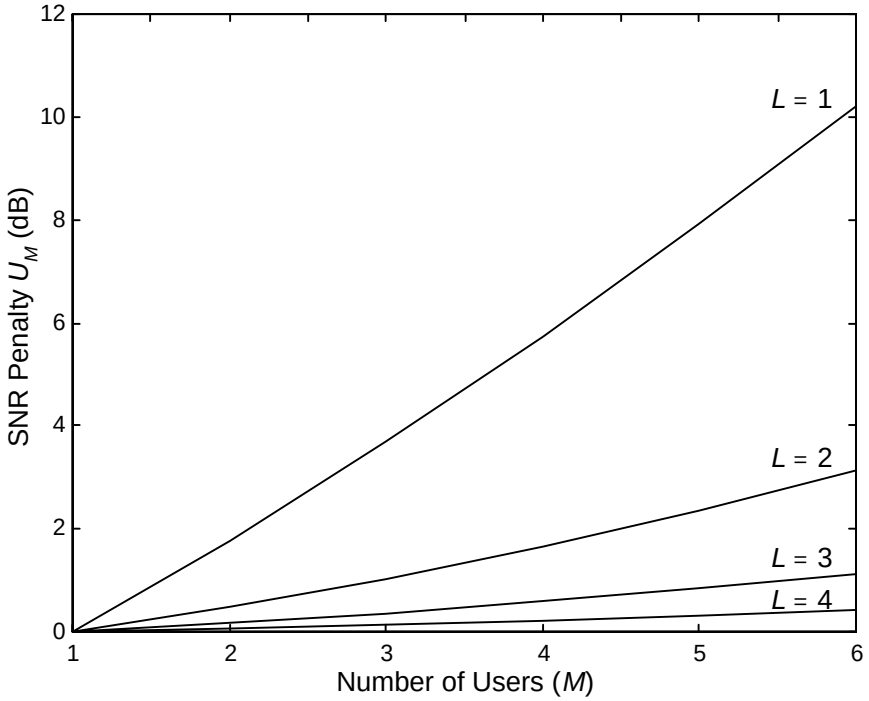

Figure 3.5: SNR penalty in the asymptotic region for equipower BPSK signals and perfect CSI.

curves are coincident with those for perfect CSI indicating that noise, rather than channel estimation error, is the dominant effect determining error rate; that is, $N_o \gg \sigma^2_{\hat{e}_m}$ in this region. However, in the high SNR region, channel estimation error dominates the performance producing an irreducible error rate — or error floor — similar to that observed in systems employing differential detection. The floor value is given by substituting (3.42) into (3.38). Evidently, the floor value increases with the number of users M, but is decreased quickly by use of additional antennas.

The latter effect implies that the use of antenna diversity relaxes the required accuracy of channel estimation in order to achieve a given performance level. This is illustrated in Fig. 3.8 which plots the channel estimation correlation coefficient required to achieve a floor value of 10^{-3}. For example, with two antennas and four users, the correlation coefficient must be no less than about 0.995, i.e., 1% relative estimation error variance. With four antennas and the same number of users, the correlation coefficient must be no less than

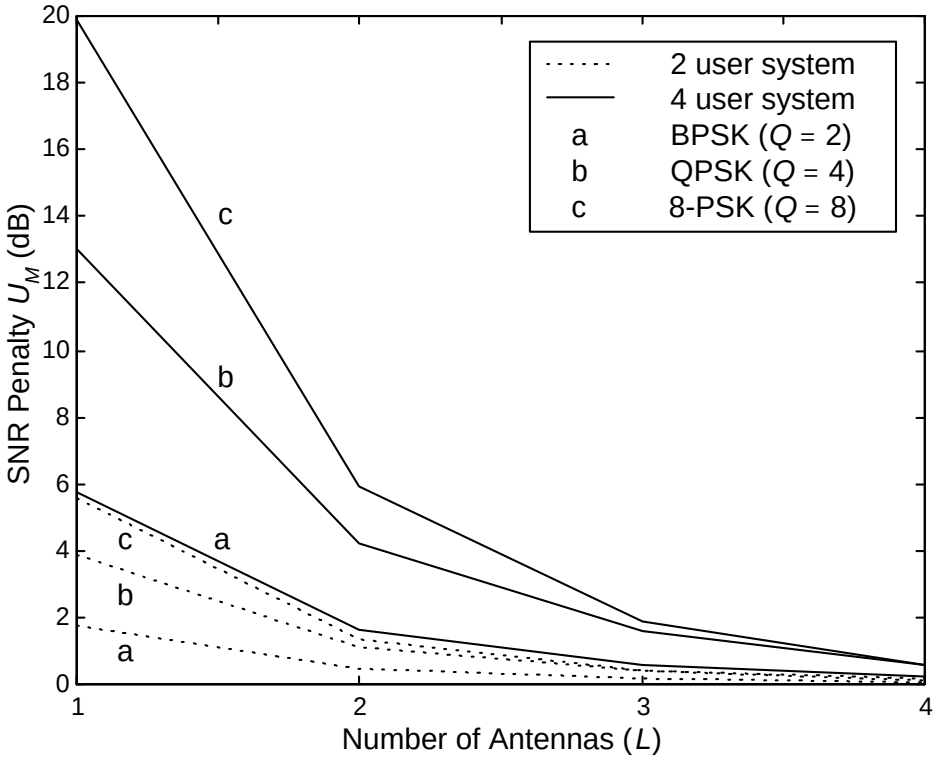

Figure 3.6: SNR penalty in the asymptotic region for several different modulation formats with equipower signals and perfect CSI.

about 0.95, resulting in a relative estimation error variance of approximately 10%.

Variable Correlation Coefficient

With pilot-based channel estimation, the channel estimation correlation coefficient typically varies with SNR, tending to unity with increasing SNR. As discussed previously, the model of variation in the asymptotic (high-SNR) region using pilot-based multiuser channel estimation is $|\rho_m|^2 = 1 - k_m\Gamma^{-1}$ for the case of equipower users. As shown in Chapter 5, the constant k_m depends on the number of users M as well as the parameters of the channel estimation scheme such as interpolator order, frame length, Doppler fade rate, and the choice of training sequences. Clearly, for large SNR, $\rho_m \approx 1$.

For illustrative purposes, BER results are presented in Fig. 3.9 using representative values

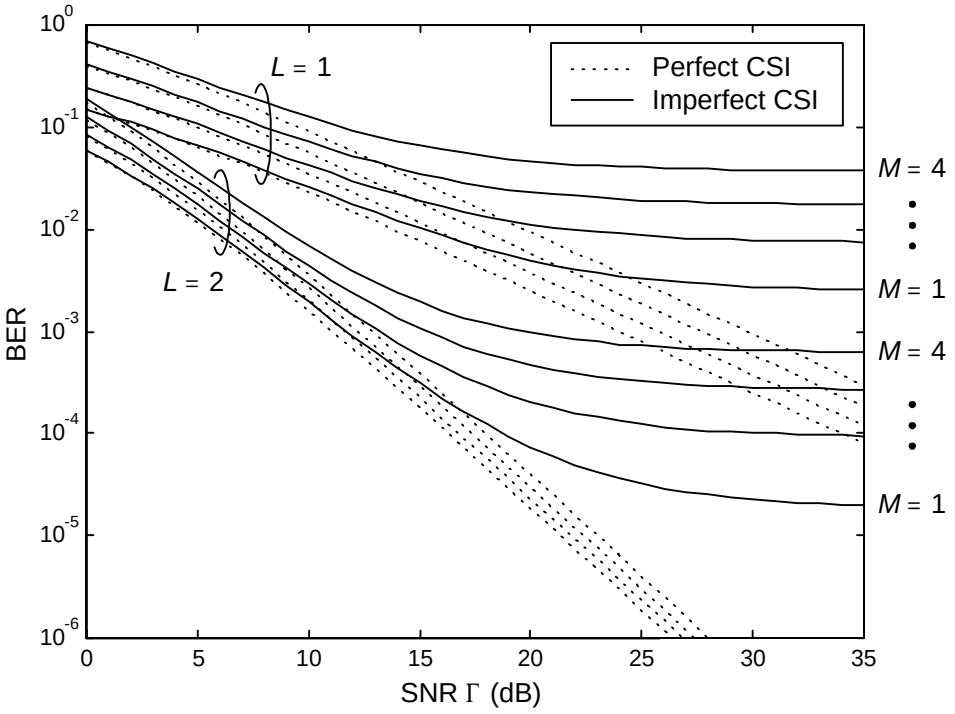

Figure 3.7: Performance of joint detection of equipower BPSK signals with imperfect CSI and fixed channel estimation correlation coefficient $\rho_m = 0.995$ for all users.

of the parameters. In this plot, the channel estimator (interpolator) is optimized for each SNR point. Using the optimal interpolator coefficients, ρ_m is calculated explicitly and used in (3.36) and (3.38) to calculate BER. Note that since the data throughput is reduced by the training overhead, the BER is plotted versus the SNR per bit defined as $\Gamma_b = \eta\Gamma$. The throughput η is defined as $\eta = (N - M)/N$ where N is the frame length (30 in this case).

The most striking difference compared to the fixed ρ_m model is that the error floor disappears since $\rho_m \approx 1$ in the high-SNR region. At low SNR, the performance is degraded from the perfect CSI case since the high noise level reduces the quality of the channel estimates. Moreover, the imperfect CSI curve runs essentially parallel to the perfect CSI curve with a performance degradation that depends on the number of users and antennas. In other words, asymptotic diversity order L is maintained for all users. This behaviour is consistent with that observed in equation (3.47).

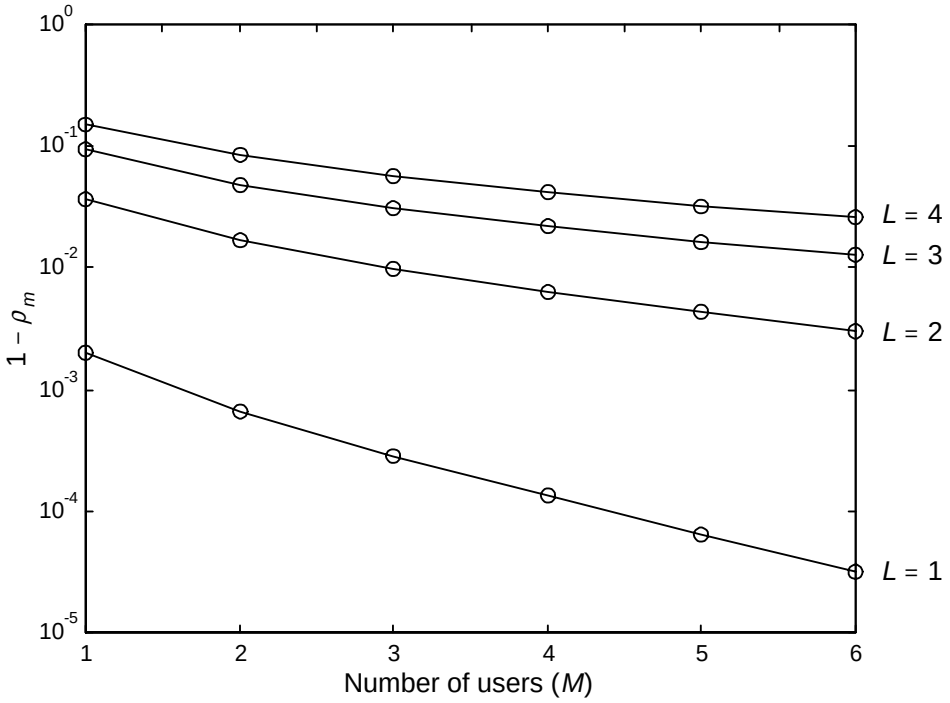

Figure 3.8: Channel estimation accuracy required to acieve an error floor of 10^{-3} for equipower BPSK signals.

3.4.3 Comparison of Joint Detection and MMSE Combining

Fig. 3.10 compares the performance of joint detection and MMSE combining. The BER for MMSE is obtained through Monte Carlo simulation as explained in Section 3.3.2. Note that for $M = 1$, the performance of JD and MMSE is the same. Clearly, JD outperforms MMSE by a large margin. Moreover, with JD the diversity order is two, regardless of the number of users. In contrast, with MMSE the order of diversity is reduced to one with $M = 2$. When $M > 2$, MMSE combining collapses. In general, with L antennas and M users, all users enjoy L-fold diversity with JD, whereas with MMSE, each user experiences diversity order $L - M + 1$.

The performance gains of JD compared to MMSE are further illustrated in Fig. 3.11 where the BER is plotted for a fixed SNR of 12 dB. Clearly, the performance of JD degrades gracefully with each additional user, whereas with MMSE it degrades very quickly and saturates at an unacceptably high error-rate when the number of users exceeds the number

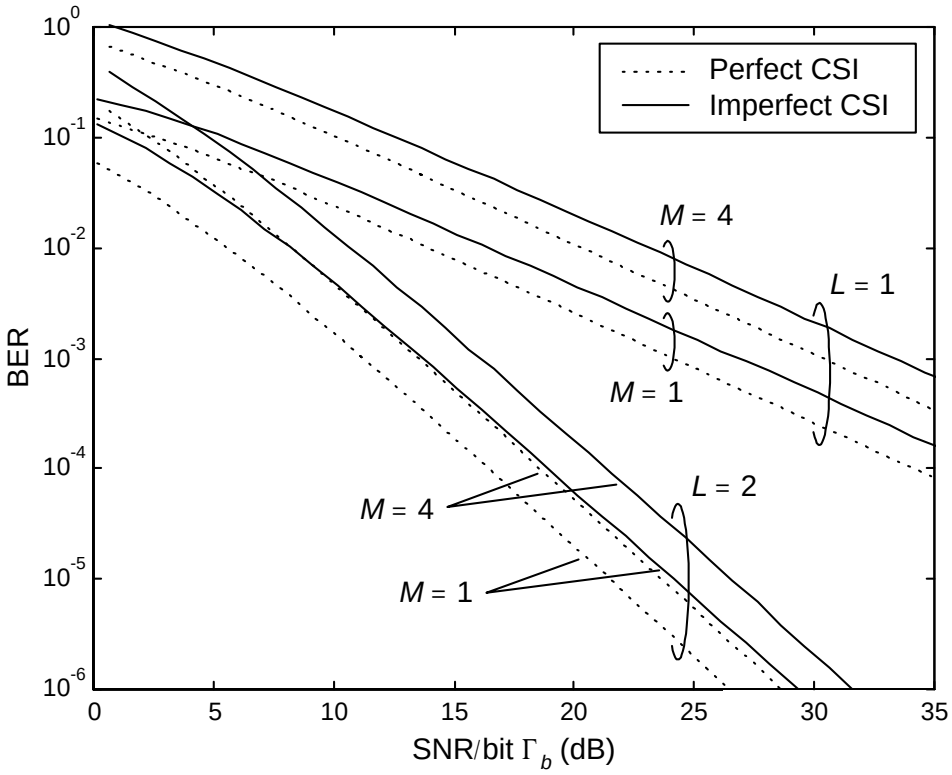

Figure 3.9: Performance of joint detection of equipower BPSK signals using pilot-based multiuser channel estimation with a 1% Doppler fade rate, interpolator order 7, frame length 30, and training sequence length M.

of antennas.

3.4.4 Nonequipower Signals

Thus far, the performance results have applied to the case of equipower signals. It is interesting to investigate the case of nonequipower signals to see how an unequal distribution of SNRs affects the performance of both the weak and strong users. Two different distributions are examined, each for the detection of four users: case 1 corresponds to three strong users and one weak user; case 2 corresponds to one strong user and three weak users. The difference in SNR between the weak and strong users in both cases is 10 dB. Recall that with nonequipower users, the SNR for the mth user is $\Gamma_m = K_m\Gamma$ where Γ is a reference SNR,

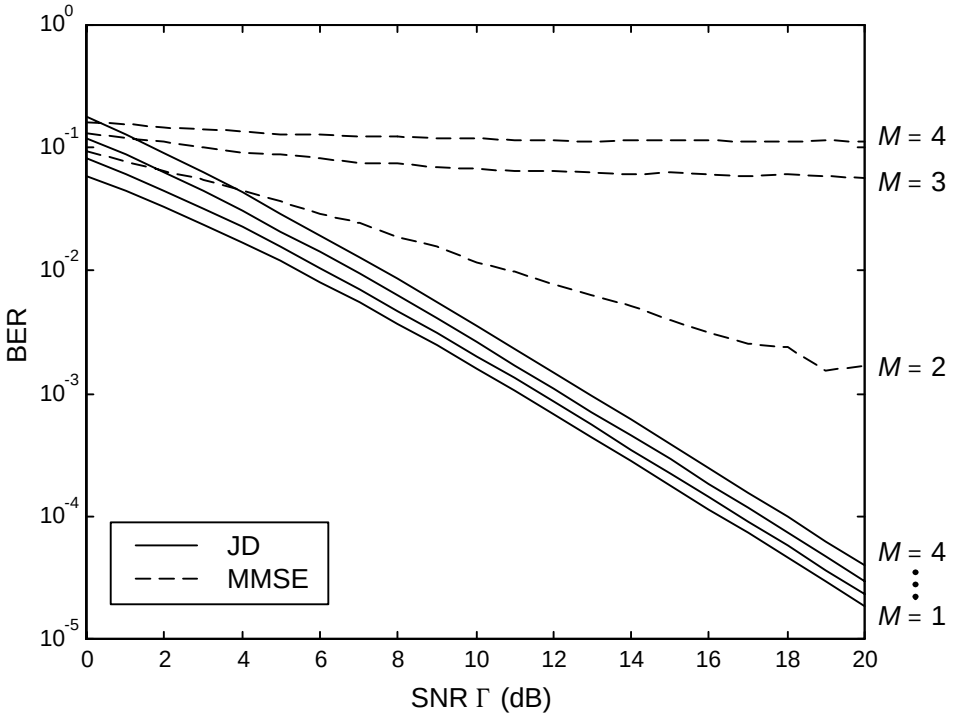

Figure 3.10: Comparison of joint detection and MMSE combining with two anntenas, equipower BPSK signals, and perfect CSI.

arbitrarily taken to be that of the strongest user. Thus, for case 1 we have $K_1, K_2, K_3 = 1$ and $K_4 = 0.1$; for case 2 we have $K_1 = 1$ and $K_2, K_3, K_4 = 0.1$.

Fig. 3.12 compares the performance of cases 1 and 2 with the case of equipower users. In this graph, the BER of each user is plotted against its *own* SNR Γ_m given by (3.4). This convention makes it appear that the weak users have a lower BER than the strong users; however, this is not true. In a typical operating scenario, all users are detected at a common receiver noise level N_o. Consequently, the weak and strong users operate at SNR values that are 10 dB apart, and the horizontal axis must be interpreted in this light. For example, if the strong users are at 20 dB SNR, then the weak users are at 10 dB with a BER greater than that of the strong users as expected. Interpreting the graph in this way reveals that the performance of the strong users is better than that of the weak users by about 7–8 dB. Furthermore, the performance of all users is degraded from the equipower case, indicating that in an operational system, some degree of power control may be desirable to keep the

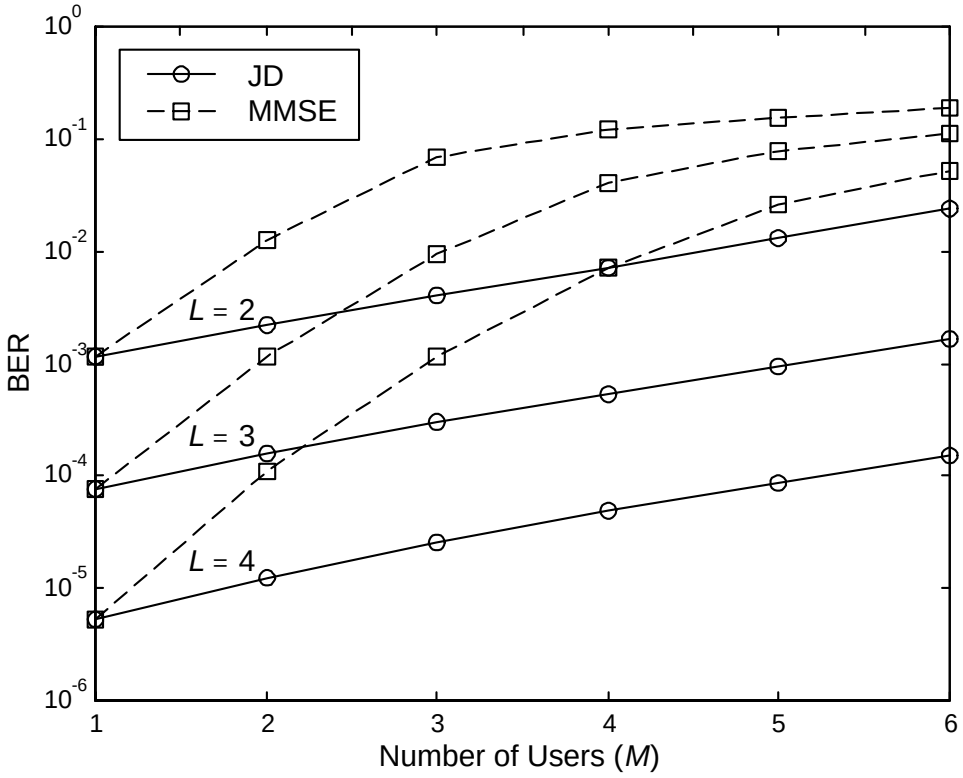

Figure 3.11: Comparison of joint detection and MMSE combining at a fixed SNR of 12 dB with equipower BPSK signals and perfect CSI.

distribution of SNRs more or less uniform. Fig. 3.12 also reveals that the performance of case 1 is approximately 2 dB better for both strong and weak users than case 2, indicating that the larger the ratio of number of strong users to weak users, the better the performance.

For the case of imperfect CSI with pilot-based multiuser channel estimation, the strong users are expected to have a larger ρ_m than the weak users. However, as mentioned previously with respect to equation (3.47), any user with poor CSI degrades the performance of all users, even those with better CSI. As a result, all the curves in Fig. 3.12 shift to the right by an amount that depends on the accuracy of channel estimates, but the relative position of the case 1 and 2 curves with respect to the equipower curve remain the same.

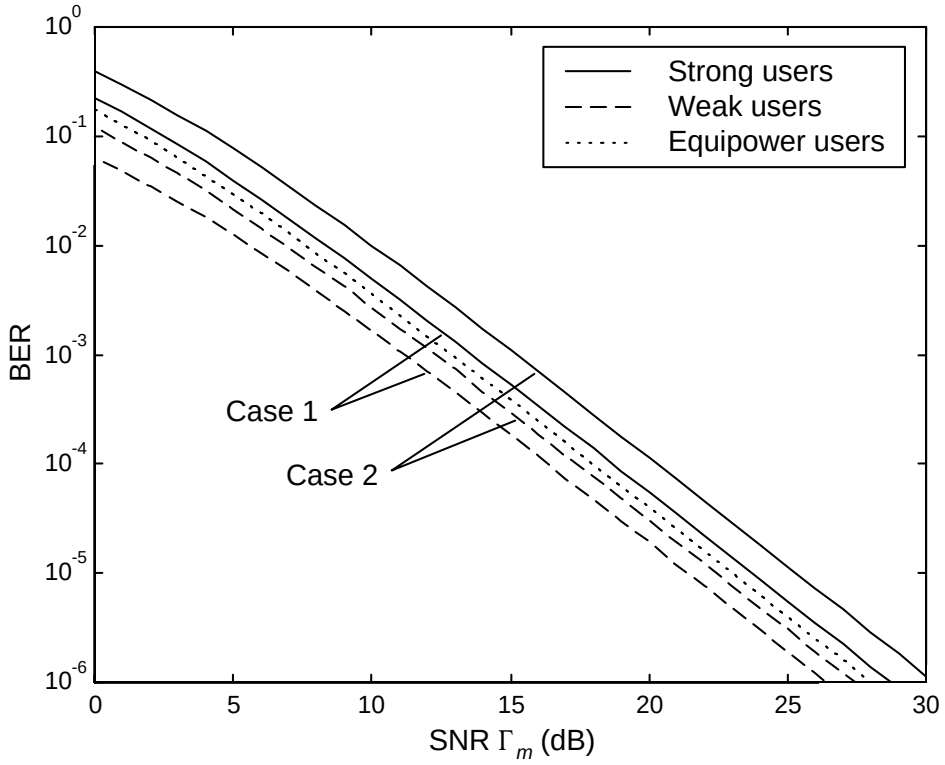

Figure 3.12: Performance of joint detection of four nonequipower BPSK signals with two antennas and perfect CSI. The strong and weak signals have a difference in SNR of 10 dB. Case 1: three strong/one weak. Case 2: one strong/three weak.

3.5 Conclusions

In this chapter, the joint detection of multiple cochannel symbol-synchronous PSK signals using a diversity array with channel estimates is investigated. A fully-analytical expression is derived that gives the union bound on average symbol-error rate for an arbitrary number of users M and antennas L for both perfect and imperfect CSI. The analysis is general in the sense that it may be applied to a variety of channel estimation schemes.

It is demonstrated that with joint detection, many more users than the number of antennas may be supported — all enjoying L-fold diversity — with a small degradation in performance with each additional user. Further, it is demonstrated that even with only a single antenna, with accurate CSI, several users can experience good performance. This new

mode of reception behaviour starkly contrasts that of classical MMSE combining where the maximum number of users is limited to the number of antennas, and the diversity order is only $L - M + 1$ for each user.

It is shown that the cost of supporting additional users at a fixed error-rate is a small increase in the required per-user SNR. The SNR penalty relative to a single user BPSK system is quantified as a function of the number of users, constellation density, and number of antennas. For example, with BPSK modulation, a single antenna, and perfect CSI, only a 2 dB penalty is incurred for each additional user. With four antennas, this penalty is reduced to less than 0.1 dB.

The union bound on BER for a single antenna and perfect CSI is compared with results from simulation and it is found to be asymptotically tight with increasing SNR. As the number of users increases, a given accuracy of the bound requires a somewhat higher SNR; however, the accuracy is quite satisfactory in the useful BER range.

Two models for the channel estimation correlation coefficient are considered: one in which the coefficient is fixed, and another in which it is variable with SNR, as in pilot-based channel estimation. It is shown that the fixed correlation coefficient produces an error floor that increases with the number of users, but is reduced quickly by the use of additional antennas. Alternatively, the channel estimation accuracy required to maintain a fixed floor value may be relaxed by use of additional antennas. With pilot-based channel estimation where the correlation coefficient improves with SNR, the error floor is eliminated. The cost of imperfect CSI is simply an increase in the required SNR in order to maintain a fixed error-rate. The interdependence of all users participating in the joint detection process is also demonstrated: the channel estimation of one user affects the performance of all users, even those with perfect CSI.

Unequal SNR distributions are investigated, and it is found that both the weak and strong users' performance is degraded from the equipower case, indicating that power control may be desirable in a practical system. Furthermore, it is found that performance depends on the ratio of number of strong users to weak users and improves as this ratio increases.

Chapter 4

Joint Detection in the Downlink

In the downlink, the transmission scenario is different from the uplink: the antenna array is located at the base station transmitter, rather than receiver, and the mobile receivers typically have only one antenna due to space constraints. Consequently, any array processing must performed at the base station which usually has no knowledge of the downlink channel gains. This configuration makes diversity reception at the mobiles more difficult to obtain than in the uplink. Due in part to this more complicated scenario, methods of achieving diversity in the downlink for multiple cochannel users has received little attention in the literature. In the following, several existing methods are discussed in order to place the current work in context. Within this discussion, one must keep in mind that the goal is to allow several users to occupy the same bandwidth as a single user, i.e., allow RWC, in order to obtain a system capacity increase.

One conceptually simple method of achieving diversity in the downlink for multiple cochannel users is through adaptive transmission [48]. With adaptive transmission, a different weight vector is applied to each of the M cochannel signals, and the sum of the resulting vector signals is transmitted from the diversity array. The weight vectors are jointly optimized, using estimates of the downlink channel gains for all users. The optimization is designed such that the transmitted signals due to the mth user combine coherently at the mth mobile — thus providing diversity — and those due to other users combine destructively. This is analogous to ZF combining in the uplink where the weight vector for the mth user is chosen to be orthogonal to the uplink channel gain vectors for the other users. With adaptive transmission, each additional user reduces the order of diversity by one for all users, similar to ZF and MMSE combining in the uplink.

Adaptive transmission requires estimates of the downlink channel gains for all users. If the uplink and downlink gains are correlated to some degree, the downlink gains may be estimated from the uplink gains at the base station. Sufficient correlation occurs in time-division duplex (TDD) systems with short bursts, or frequency-division duplex (FDD) systems with narrow separation between the uplink and downlink frequencies. Alternatively, the base may obtain estimates of the downlink gains via feedback from the mobiles. However, in typical TDMA cellular systems, there is no mobile-to-base feedback, and the uplink and downlink frequencies are widely separated. Consequently, adaptive transmission is impractical in all but specialized circumstances.

Another method of achieving diversity in the downlink is through space-time codes [49],[50] which do not require the base station to have knowledge of the downlink channels. However, space-time codes have been designed for single-user systems only. In a multiuser system, the only option is to multiplex the users' bit streams together into a single high-rate stream, and then apply space-time coding principles. However, in a system with M users, this increases the symbol-rate by a factor of M, requiring M times the transmission bandwidth of a single user. Clearly this negates the system capacity improvement sought in the first place.

In this thesis the following alternative structure is proposed: the users' bit sequences are multiplexed together as just suggested, but standard convolutional coding and interleaving of the composite bit sequence is employed in order to provide temporal diversity to the mobiles. Like space-time codes, the transmitter does not require estimates of the downlink channels. Using an array of L antennas, L successive code-symbols are transmitted simultaneously on the different antennas to reduce the bandwidth required to transmit the high-rate coded sequence. The simultaneous transmissions are then jointly detected at each mobile receiver, and the data from other users is discarded. If the number of transmit antennas is equal to the number of users, it is possible to support multiple users in the same bandwidth as a single user, as desired.

The proposed structure is similar in spirit to BLAST [51], which is intended for fixed wireless systems as a method of obtaining a rate increase for a single user. With BLAST, a single high rate bit stream is coded using a standard block code in order to obtain temporal diversity. In order to reduce the required transmission bandwidth, each codeword is split into L sub-blocks which are transmitted simultaneously on the L diversity antennas. At the receiver, an interference suppression technique such as ZF or MMSE combining is

used to separate the simultaneous transmissions. An elegant layered space-time processing technique is employed in which the interference suppression is improved by pre-canceling the interference from previously decoded sub-blocks with the use of channel estimates. In normal operation, the codeword error rate is small, thus the pre-canceling operation is reliable. In contrast to the structure proposed in this thesis, BLAST requires as many receive antennas as transmit antennas due to the use of ZF or MMSE. Consequently, it is unsuitable for a mobile environment where the handsets have only one antenna.

Based on the proposed structure, a number of new contributions are made in this thesis which also appear in [52]. Since several successive code symbols are transmitted simultaneously from different antennas, a single vector channel usage may span several trellis transitions. To handle this unusual situation in both analysis and in decoder implementation, the concept of a "merged trellis" is introduced in which the trellis is modified such that a single channel usage spans only one transition. Using the merged trellis, the optimal receiver structure is derived based on soft-decision joint decoding. A special case of the resulting structure appears in [53], in which a rate increase is obtained for a single user; however, the problem of multiple trellis transitions per channel usage is not addressed, and the performance results are obtained by simulation. In contrast to [53], a fully-analytical expression is for the bit-error rate of the optimal receiver is derived here. From this, novel behaviour is demonstrated in terms of diversity order: as the number of antennas increases due to an increasing number of users, the diversity order actually decreases due to the simultaneous transmission of successive code symbols. However, with several users and a code with moderate constraint length, good performance for all users may be obtained.

4.1 System and Signal Models

Fig. 4.1(a) shows the proposed structure of the base station transmitter. The M users' bit streams are first multiplexed into a common bit stream of rate $R_b = \sum_{m=1}^{M} R_{b_m}$ where R_{b_m} is the rate of the mth user's bit stream in bits per second. The common bit stream is then encoded using an (n_c, k_c, L_c) convolutional code, where k_c is the number of input bits per encoding interval, n_c is the number of output bits, and L_c is the constraint length of the code (measured in blocks of k_c bits). The code rate is $R_c = k_c/n_c$. The output bits are then mapped to PSK symbols using Gray coding. The number of points in the PSK constellation is denoted as Q. The resulting symbol sequence is interleaved keeping blocks

of L symbols together, where L is the number of transmit antennas. The symbols within each length-L block are then transmitted simultaneously on the L different antennas. The vector of transmitted symbols is denoted $\mathbf{c}(k) = (c_1(k), c_2(k), \cdots, c_L(k))^T$.

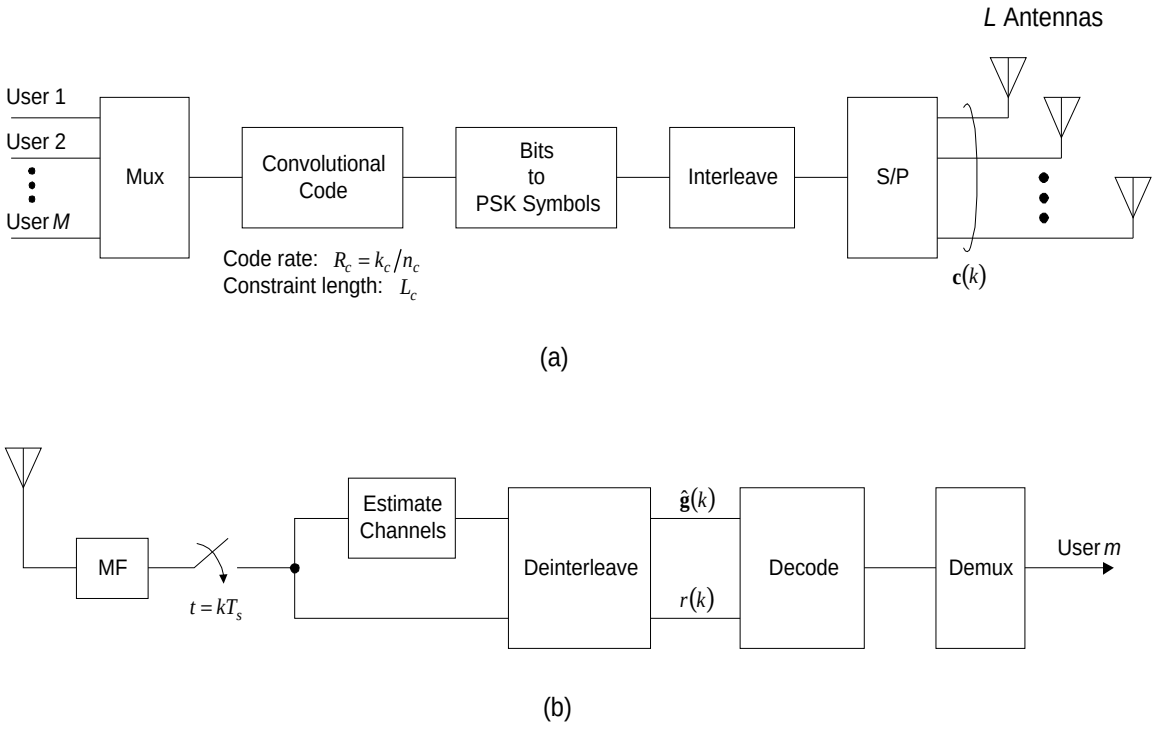

Figure 4.1: Structure of (a) base station transmitter, and (b) mobile receiver for user m.

The simultaneous transmission of successive code symbols slows the symbol rate on each antenna by a factor of L, thus reducing the bandwidth required to transmit the multiplexed bit stream. In terms of the various parameters, the symbol rate on each antenna is $R_s = R_b/N_b$ where

$$N_b = R_c L \log_2 Q \tag{4.1}$$

is the number of information bits transmitted per vector channel usage. A vector channel usage is simply the transmission of one symbol vector $\mathbf{c}(k)$.

If all M users have the same bit rate R_{b_1}, then $R_s = (M/N_b) R_{b_1}$. The quantity M/N_b may be interpreted as the bandwidth expansion factor. As a check, for the usual situation of

a single user and one transmit antenna, the bandwidth expansion factor is $1/R_c \log_2 Q$. As expected, this is equal to two for the example of a rate one half code and BPSK modulation ($Q = 2$). For a multiuser system, if the number of antennas equals the number of users, the bandwidth expansion factor is also $1/R_c \log_2 Q$, implying that M users may be supported in the same bandwidth as a single user (assuming the same code rate and constellation density). If the number of antennas exceeds the number of users, a rate increase for all users may be supported in the same bandwidth.

In terms of decoding complexity, the proposed structure has advantages over assigning one user per antenna and independently coding the users' bits sequences — an alternative for supporting M users in the same bandwidth as a single user. The proposed use of a common code operating on the multiplexed bit sequence implies that the soft decision decoder operates on a single trellis, rather than a joint trellis as would be the case with independently coding the users. With an (n_c, k_c, L_c) convolutional code, the single trellis has $2^{k_c(L_c-1)}$ states, whereas the joint trellis has $2^{Mk_c(L_c-1)}$ states. Clearly, the proposed structure has the advantage that the size of the state set does not increase exponentially with the number of users.

At the receiver shown in Fig. 4.1(b), the deinterleaved received sample sequence is

$$r(k) = A\mathbf{g}^T(k)\,\mathbf{c}(k) + n(k). \tag{4.2}$$

The elements of the vector $\mathbf{g}(k)$ are i.i.d. complex Gaussian RVs with variance σ_g^2 that model the independent flat Rayleigh fading channels between the L transmit antennas and the single receive antenna. The scalar $n(k)$ is a complex Gaussian noise sample with variance N_o, and the constant A is the amplitude of the transmitted signal on each antenna. It is assumed that the total transmit power P is divided equally among the transmit antennas, giving the transmit amplitude as $A = \sqrt{2(P/L)}$. Recall that the symbols in the transmit symbol vector $\mathbf{c}(k)$ are drawn from a PSK constellation, which is assumed to have unit-radius.

As explained in Section 4.3, the deinterleaved symbol sequence $r(k)$ is passed onto the soft decision decoder which implicitly performs joint detection of the multiple cochannel signals within the branch metric. The joint detection requires estimates of the channel gain vectors, which may be obtained using the multiuser (multi-source) channel estimation technique developed in Chapter 5. After decoding, the composite bit sequence is demultiplexed, and the bits corresponding to other users are discarded.

Since the received signal consists of a sum of cochannel signals from L different sources, the receive SNR is defined on a per-antenna source as

$$\Gamma = \frac{\frac{1}{2}E\left[\left|Ag_l(k)\,c_l(k)\right|^2\right]}{\frac{1}{2}E\left[\left|n(k)\right|^2\right]} = \frac{A^2\sigma_g^2}{N_o}. \tag{4.3}$$

However, the receive SNR per information bit is the standard measure used when coding is employed, and it is given by $\Gamma_b = \Gamma/N_b$. Also of interest is the per-user *transmit* SNR defined as $\Gamma_T = \left(\frac{P}{N_o}\right)/M$. This quantity is useful when comparing the transmit power requirements of systems with different numbers of users and antennas. Using the expression $A = \sqrt{2\,(P/L)}$, the per-user transmit SNR is given by

$$\Gamma_T = \frac{N_b}{2\sigma_g^2}\left(\frac{L}{M}\right)\Gamma_b. \tag{4.4}$$

4.2 Trellis Merges

Because several successive code symbols are transmitted simultaneously from different antennas, a single vector channel usage may span several trellis transitions. For example, consider a system with two users, two antennas, a rate one half convolutional code, and QPSK modulation, i.e., $L = M = 2$, $R_c = 1/2$, and $Q = 4$. For illustrative purposes in this example, the simple constraint length three convolutional code with generator polynomials $g_1(D) = D^2 + 1$ and $g_2(D) = D^2 + D + 1$ is used. The trellis diagram for this code is shown in Fig. 4.2(a), in which a solid line indicates an input '0' and a dashed line indicates an input '1'.

Now consider two successive encoder input bits b_0 and b_1, and assume the encoder is initially in state 0. The two input bits correspond to two transitions in the trellis. After two transitions, the encoder may end up in states 0, 1, 2, or 3 depending on the values of b_0 and b_1. The encoder output bits along the paths to each of these four possible states are shown on the trellis diagram. Pairs of the output bits are mapped to QPSK symbols using Gray coding, i.e., $00 \rightarrow +1$, $01 \rightarrow +j$, $10 \rightarrow -j$, and $11 \rightarrow -1$. With two antennas, two successive code symbols are transmitted simultaneously resulting in the following four possible QPSK transmit vectors corresponding to the final states 0, 1, 2, and 3:

$$\mathbf{c}(k) = \begin{bmatrix} +1 \\ +1 \end{bmatrix}, \begin{bmatrix} -1 \\ +j \end{bmatrix}, \begin{bmatrix} +1 \\ -1 \end{bmatrix}, \text{ or } \begin{bmatrix} -1 \\ -j \end{bmatrix}. \tag{4.5}$$

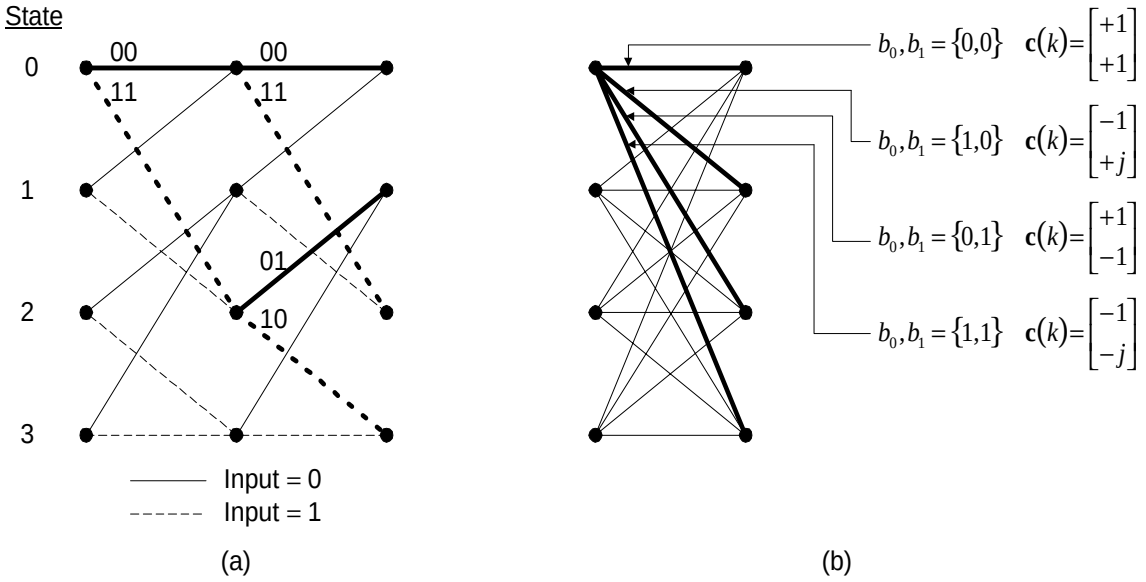

Figure 4.2: (a) Trellis diagram for a $(2, 1, 3)$ convolutional code, and (b) merged trellis for the example of a system with two users, two antennas, and QPSK modulation.

Clearly, a single vector channel usage spans two trellis transitions. This is unusual since we are used to one or more channel usages in a single transition. To handle this unusual situation in both analysis and decoder implementation, the concept of a "merged trellis" is introduced in which the trellis is modified such that a single vector channel usage spans only one transition. Fig. 4.2(b) illustrates the merged trellis for the above example. As can be seen, one transition in the merged trellis enumerates all possible successor states after two transitions in the original trellis. Now, instead of two branches leading into and out of each state, there are four. Associated with each branch of the merged trellis are two input bits but only one transmit symbol vector, as desired.

In order to generalize the above example, it is convenient to define the parameter N_u as the number of vector channel usages per transition in the original (un-merged) trellis, given by

$$N_u = \frac{n_c}{L \log_2 Q}. \tag{4.6}$$

In the above example, $N_u = 1/2$, meaning that one vector channel usage spans two trellis

transitions. The parameter N_u signals whether or not the trellis must be merged. If $N_u < 1$, then merges are required; if $N_u \geq 1$ (the conventional situation) then no merges are required.

For the case of $N_u < 1$ where merges are required, define the parameter $t = 1/N_u$. If t is an integer, then t transitions in the original trellis must be merged into one. In the merged trellis, there are thus $2^{k_c t}$ branches entering and leaving each state. Since the number of states in the merged trellis remains the same, the total number of branches per transition becomes $2^{k_c(L_c+t-1)}$ where L_c is the constraint length of the code. Furthermore, $k_c t$ input bits and only one transmit vector $\mathbf{c}(k)$ are associated with each branch.

If t is not an integer, then find the smallest integer n such that $t' = nt$ is an integer. The merged trellis may then be defined by merging t' transitions into one. Analogous to the case of integer t, the total number of branches per transition becomes $2^{k_c(L_c+t'-1)}$ with $k_c t'$ input bits associated with each branch. The only difference is that n transmit vectors are now associated with each branch in the merged trellis instead of just one.

For simplicity, it is assumed in this work that if trellis merges are required ($N_u < 1$), then N_u is a reciprocal integer ($\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, etc.), meaning that only one transmit vector $\mathbf{c}(k)$ is associated with each branch in the merged trellis. If no merges are required ($N_u \geq 1$), then N_u is assumed to be an integer. These assumptions place constraints on the allowable constellation density and number of antennas. Table 4.1 lists various allowed combinations of Q and L for several common code rates along with the corresponding parameters N_u and N_b. As can be seen, increasing either the constellation density or the number of antennas leads to a larger number of bits per channel usage, i.e., better spectrum efficiency. On the other hand, it leads to a smaller number of usages per transition, eventually requiring trellis merges when N_u becomes fractional.

An important property of the proposed transmitter structure is that as the parameter N_u decreases below unity, the length of the shortest error event in the merged trellis decreases. This property is illustrated in Fig. 4.3, which shows several error events in the merged trellis for the same (2,1,3) code considered above. The cases of $N_u = 1$, $\frac{1}{2}$, and $\frac{1}{3}$ are shown. For example, this could correspond to the use of QPSK modulation with one, two, and three antennas respectively.

The shortest error event in the original trellis is of length three (measured in transitions). For $N_u = 1/2$, the merged trellis is fully connected, implying that the shortest error event is of length two. For $N_u = 1/3$, the merged trellis contains parallel transitions, implying that the shortest error event is only of length one. In general, for a code with constraint

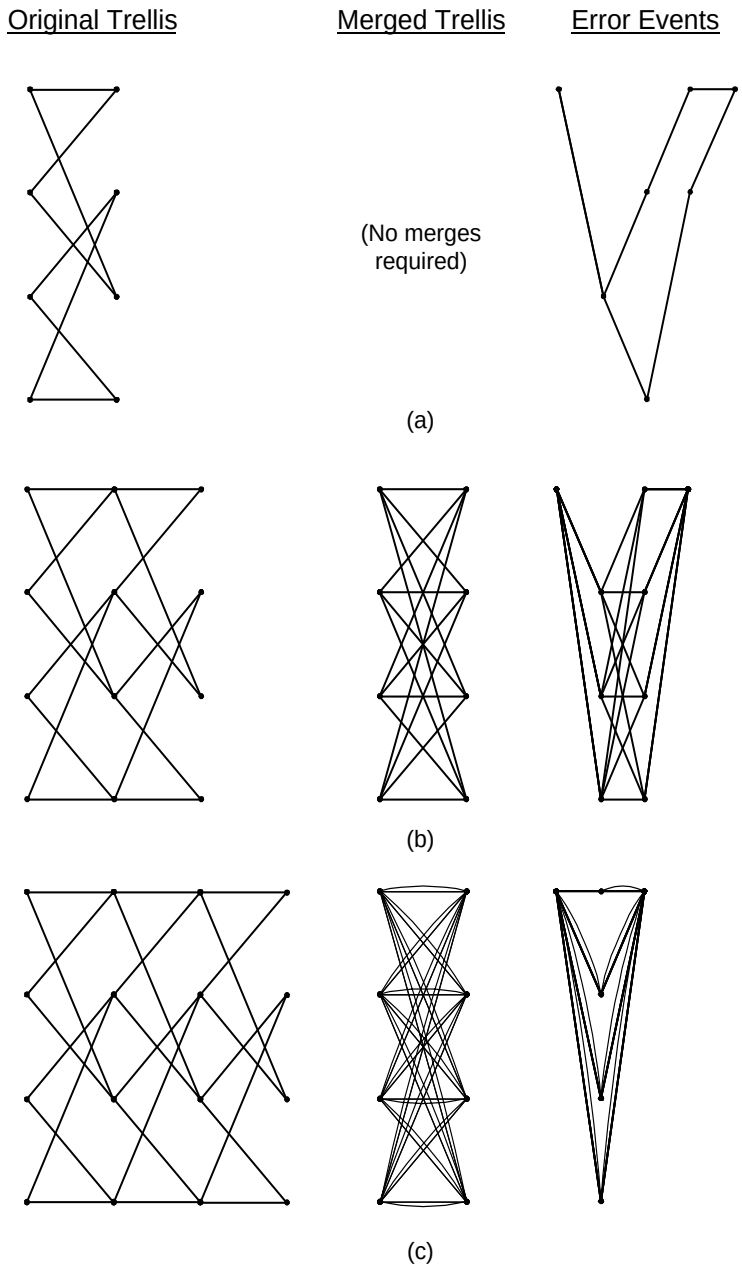

Figure 4.3: Some error events in the merged trellises of the $(2, 1, 3)$ code for (a) $N_u = 1$, (b) $N_u = \frac{1}{2}$, and (c) $N_u = \frac{1}{3}$.

Table 4.1: Number of bits per usage N_b and usages per transition N_u for several allowed combinations of the PSK constellation size Q, the number of antennas L, and the code rate R_c.

		$R_c = 1/3$		$R_c = 1/2$		$R_c = 2/3$	
Q	L	N_b	N_u	N_b	N_u	N_b	N_u
2	1	1/3	3	1/2	2	2/3	3
	2	–	–	1	1	–	–
	3	1	1	–	–	2	1
	4	–	–	2	1/2	–	–
4	1	–	–	1	1	–	–
	2	–	–	2	1/2	–	–
	3	1/2	1/2	3	1/3	4	1/2
	4	–	–	4	1/4	–	–
8	1	1	1	–	–	2	1
	2	2	1/2	3	1/3	4	1/2
	3	3	1/3	–	–	6	1/3
	4	4	1/4	6	1/6	8	1/4

length L_c, a fully connected merged trellis occurs if $1/N_u = L_c - 1$, and a merged trellis with parallel transitions occurs if $1/N_u = L_c$. For a large constraint length, N_u has to be quite small (large L and/or Q) before the merged trellis contains parallel transitions and the length of the shortest error event is reduced to one.

The above behaviour has important implications for performance. In a fading environment, the diversity order obtained through coding and interleaving is related to the length of the shortest error event: the longer this error event, the higher the number of independent channel usages across the event, and thus the higher the order of diversity. Thus, for a fixed constraint length code, as N_u decreases (due to increasing L or Q), the diversity order decreases due to the error events in the merged trellis becoming shorter. The decrease with the number of antennas is unusual, since in the uplink we are used to seeing the opposite. However, this behaviour may be attributed to the fact the antenna array is being used for a different purpose. Rather than for diversity combining, the antenna array is used to provide independent channels on which to transmit successive code symbols in order to reduce the transmission bandwidth. Thus, as N_u decreases, a larger constraint length code must be used to ensure that the length of the shortest error event in the merged trellis, and hence the diversity order, is sufficient. The cost is increased decoding complexity due to the larger state set.

It is interesting to note that a special case of the proposed structure appears in [53] where a rate increase is obtained for a single user by use of multiple transmit antennas. The special case considers a rate one half code with BPSK modulation and two transmit antennas. This results in $N_b = 1$ bit/usage and $N_u = 1$ vector channel usage per transition. Because of the chosen parameters, the problem of a single channel usage spanning several trellis transitions does not appear, and is thus not addressed. Furthermore, the performance results are obtained by simulation, whereas in this work, a fully-analytical expression for average BER is derived, from which the above novel diversity behaviour may be ascertained.

4.3 Decoder Implementation

According to the MAP decision rule, the decoder chooses the state sequence $\mathbf{x}_i$ that maximizes the *a posteriori* probability $P(\mathbf{x}_i \,|\, \mathbf{r}, \mathbf{G})$ where $\mathbf{r}$ is the sequence of deinterleaved received samples, and $\mathbf{G}$ is the deinterleaved sequence of channel gain vectors. In this work, it is assumed that the receiver has perfect knowledge of $\mathbf{G}$. If $N_u < 1$, $\mathbf{x}_i$ is a sequence of states through the merged trellis. If $N_u \geq 1$, $\mathbf{x}_i$ is a sequence of states through the original (un-merged) trellis. Assuming all state sequences are equiprobable, the receiver may maximize the forward probability $P(\mathbf{r} \,|\, \mathbf{x}_i, \mathbf{G})$ instead. Equivalently it may minimize $-\ln P(\mathbf{r} \,|\, \mathbf{x}_i, \mathbf{G})$. Because the sequence of noise samples is white, the desired probability factors as a product, and the log-likelihood function may be expressed as the summation

$$-\ln P(\mathbf{r} \,|\, \mathbf{x}_i, \mathbf{G}) = -\sum_k \ln P\left[r(k) \,|\, \mathbf{x}_i, \mathbf{g}(k)\right] \tag{4.7}$$

Since $r(k)$ is conditionally Gaussian with mean $A\mathbf{g}^T(k)\,\mathbf{c}(k)$ and variance N_o,

$$\ln P\left[r(k) \,|\, \mathbf{x}_i, \mathbf{g}(k)\right] = -\frac{1}{2N_o}\left|r(k) - A\mathbf{g}^T(k)\,\mathbf{c}_i(k)\right|^2 - 2\pi N_o \tag{4.8}$$

where $\mathbf{c}_i(k)$ is the transmit vector corresponding to the appropriate transition in the hypothesized state sequence $\mathbf{x}_i$.

Neglecting hypothesis independent terms, the cumulative metric to be minimized is thus $\Lambda_i = \sum_k \mu_i(k)$ where the branch metric $\mu_i(k)$ is given by

$$\mu_i(k) = \left| r(k) - A\sum_{l=1}^{L} g_l(k)\, c_{il}(k) \right|^2 . \tag{4.9}$$

Note that the inner product $\mathbf{g}^T(k)\,\mathbf{c}_i(k)$ in (4.8) has been expressed as a summation over the L transmit antennas. The branch metric in (4.9) has precisely the same form as the joint detection metric for the uplink derived in Chapter 3. Consequently, the mobile implicitly performs joint detection of the cochannel signals from the L different transmit antennas within the receiver metric. An attractive feature of the joint detection is that the cochannel signals are synchronized by symbol, since the base station controls the simultaneous transmissions from the different antennas. Furthermore, the cochannel signals are of equal average power. These observations allow the reuse of selected analytical results from Chapter 3.

Using the cumulative metric Λ_i, decoding may be performed using the standard Viterbi algorithm (VA). The VA operates on the merged trellis in the case of $N_u < 1$, or the original trellis in the case of $N_u \geq 1$. In both cases, k indexes channel usages. For the former there is only one vector channel usage per transition in the merged trellis (by design), and in the latter, there may be one or more usages.

4.4 Performance Analysis

As in other studies that consider trellis decoding, e.g., [54], the BER is estimated as

$$P_b \approx \frac{1}{\alpha} \sum_i \beta_i P_{2_i} \tag{4.10}$$

where β_i is the number of bit-errors associated with the ith error event, and P_{2_i} is the pairwise error probability corresponding to that event, i.e., the probability that the cumulative metric for the ith error event is more favourable than that for the all-zero state sequence. The variable α is the number of encoder input bits per trellis transition: for the case of $N_u < 1$, $\alpha = k_c/N_u$; for the case of $N_u \geq 1$, $\alpha = k_c$. Typically the infinite summation in (4.10) is truncated such that only error events with length less than or equal to a certain threshold are included.

The BER estimate in (4.10) is based on the assumption that the pairwise error events are mutually exclusive. This assumption is invalid for low SNRs, leading to an overestimate of BER in this region. However, for high SNRs, the estimate is asymptotically accurate if a sufficient number of error events are included in the summation. A reasonable rule of thumb is to include all error events of length less than or equal to $\lfloor 2L_c N_u \rfloor$ for the case of $N_u < 1$, and length less than or equal to $2L_c$ for the case of $N_u \geq 1$.

The pairwise error probability in (4.10) is given by $P_{2_i} = P[D_i < 0]$ where $D_i = \Lambda_i - \Lambda_\mathbf{0}$ and $\Lambda_\mathbf{0}$ is the cumulative metric for that portion of the all-zero state sequence of length equal to the ith error event. Noting that for the all-zero state sequence $\mathbf{c_0}(k) = (1, 1, \cdots, 1)$, the random variable D_i is given by

$$D_i = \sum_k \left(\left| r(k) - A \sum_{l=1}^{L} g_l(k) c_{il}(k) \right|^2 - \left| r(k) - A \sum_{l=1}^{L} g_l(k) \right|^2 \right). \tag{4.11}$$

where k indexes those channel usages of the ith error event for which $\mathbf{c}_i(k)$ is not equal to the all-ones vector. In a similar fashion to the previous chapter, D_i may be expressed as the sum $D_i = \sum_k d_{ik}$ where $d_{ik} = \mathbf{z}^\dagger(k) \mathbf{Q}_i(k) \mathbf{z}(k)$ is a Hermitian quadratic form in zero-mean complex Gaussian random variables. The vector $\mathbf{z}(k)$ is defined as

$$\mathbf{z}(k) = (r(k), g_1(k), g_2(k), \cdots, g_L(k))^T \tag{4.12}$$

and the Hermitian matrix $\mathbf{Q}_i(k)$ as

$$\mathbf{Q}_i(k) = \mathbf{u}_i^*(k) \mathbf{u}_i^T(k) - \mathbf{u}_\mathbf{0}^*(k) \mathbf{u}_\mathbf{0}^T(k) \tag{4.13}$$

where the vectors $\mathbf{u}_i(k)$ and $\mathbf{u_0}(k)$ are given by

$$\begin{aligned} \mathbf{u}_i(k) &= (1, -Ac_{i1}(k), -Ac_{i2}(k), \cdots, -Ac_{iL}(k))^T \\ \mathbf{u_0}(k) &= (1, -A, -A, \cdots, -A)^T. \end{aligned} \tag{4.14}$$

Substituting (4.14) into (4.13) and expanding gives $\mathbf{Q}_i(k) = \mathbf{F}_i(k) + \mathbf{F}_i^\dagger(k)$ where

$$\mathbf{F}_i(k) = \begin{bmatrix} 0 & A(1 - c_{i1}(k)) & A(1 - c_{i2}(k)) & \cdots & A(1 - c_{iL}(k)) \\ 0 & 0 & A^2(c_{i1}^*(k) c_{i2}(k) - 1) & \cdots & A^2(c_{i1}^*(k) c_{iL}(k) - 1) \\ 0 & 0 & 0 & \cdots & A^2(c_{i2}^*(k) c_{iL}(k) - 1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{4.15}$$

The pairwise error probability may be evaluated easily using

$$P_{2_i} = L_{II}^{-1} \left\{ \frac{1}{s} \Phi_{D_i}(s) \right\} \Big|_{D=\mathbf{0}} \tag{4.16}$$

where $\Phi_{D_i}(s)$ is the two-sided Laplace transform of the PDF of D_i (i.e., the characteristic function of D_i), and $L_{II}^{-1}\{\bullet\}$ denotes the inverse two-sided Laplace transform.

In this work, perfect interleaving is assumed, implying that the channel gain vectors $\mathbf{g}(k)$ from one channel usage to the next are independent. Consequently, the d_{ik}'s are independent, and the characteristic function of D_i factors as the product $\Phi_{D_i}(s) = \prod_k \phi_{d_{ik}}(s)$ where $\phi_{d_{ik}}(s)$ is the characteristic function of d_{ik} [10] given by

$$\begin{aligned} \phi_{d_{ik}}(s) &= \frac{1}{\det\left[\mathbf{I} + 2s\mathbf{R}\mathbf{Q}_i(k)\right]} \\ &= \frac{1}{(1+s\lambda_{ik1})(1+s\lambda_{ik2})} \end{aligned} \tag{4.17}$$

where

$$\begin{aligned} \mathbf{R} &= \frac{1}{2}E\left[\mathbf{z}(k)\mathbf{z}^{\dagger}(k)\,|\mathbf{c_0}(k)\right] \\ &= N_o \begin{bmatrix} 1+L\Gamma & \Gamma/A & \Gamma/A & \cdots & \Gamma/A \\ \Gamma/A & \Gamma/A^2 & 0 & \cdots & 0 \\ \Gamma/A & 0 & \Gamma/A^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Gamma/A & 0 & 0 & \cdots & \Gamma/A^2 \end{bmatrix} \end{aligned} \tag{4.18}$$

and $\mathbf{I}$ is the $(L+1)\times(L+1)$ identity matrix. Recall that the SNR Γ is related to the SNR per information bit Γ_b by $\Gamma = N_b\Gamma_b$. The latter equality of (4.17) follows from the former since the rank of the matrix $2\mathbf{R}\mathbf{Q}_i(k)$ is only two; thus, $2\mathbf{R}\mathbf{Q}_i(k)$ has only two non-zero eigenvalues λ_{ik1} and λ_{ik2}, one positive and one negative. This fact is proved in the previous chapter which considers matrices of a similar form.

The eigenvalues are be found by expressing the characteristic polynomial of the matrix product $2\mathbf{R}\mathbf{Q}_i(k)$ as

$$\det\left[2\mathbf{R}\mathbf{Q}_i(k) - \lambda\mathbf{I}\right] = \lambda^{L-1}\left(\lambda^2 + \beta_{ik1}\lambda + \beta_{ik2}\right) = 0. \tag{4.19}$$

As explained in the previous chapter, the coefficients of the quadratic polynomial are given by $\beta_{ik1} = -T_{ik1}$ and $\beta_{ik2} = -\frac{1}{2}\left(\beta_{ik1}T_{ik1} + T_{ik2}\right)$ where $T_{ikn} = \text{trace}\{[2\mathbf{R}\mathbf{Q}_i(k)]^n\}$. Knowing the coefficients allows one to calculate the eigenvalues as

$$\begin{aligned} \lambda_{ik1} &= \frac{1}{2}\left[T_{ik1} + \sqrt{2T_{ik2} - T_{ik1}^2}\right] \\ &= 2N_o\left[a_{ik}\Gamma + \sqrt{a_{ik}^2\Gamma + 2a_{ik}\Gamma}\right] \end{aligned} \tag{4.20}$$

and

$$\begin{aligned}\lambda_{ik2} &= \frac{1}{2}\left[T_{ik1} - \sqrt{2T_{ik2} - T_{ik1}^2}\right] \\ &= 2N_o\left[a_{ik}\Gamma - \sqrt{a_{ik}^2\Gamma + 2a_{ik}\Gamma}\right] \end{aligned} \tag{4.21}$$

where

$$a_{ik} = \sum_{l=1}^{L}\left(1 - \mathrm{Re}\left[c_{il}\left(k\right)\right]\right). \tag{4.22}$$

The latter equalities in (4.20) and (4.21) are found by labourious evaluation of the matrix traces while using the fact that $\left(\mathrm{Re}\left[c_{il}\left(k\right)\right]\right)^2 + \left(\mathrm{Im}\left[c_{il}\left(k\right)\right]\right)^2 = 1$ for PSK. Since the intermediate results are large, they are not presented here.

Using (4.17), the characteristic function of D_i may now be expressed in terms of the above eigenvalues as

$$\Phi_{D_i}\left(s\right) = \prod_k \frac{p_{ik1}p_{ik2}}{\left(s - p_{ik1}\right)\left(s - p_{ik2}\right)} \tag{4.23}$$

where $p_{ik1} = -1/\lambda_{ik1}$ and $p_{ik2} = -1/\lambda_{ik2}$ denote the pole locations. Since λ_{ik1} is positive and λ_{ik2} is negative, p_{ik1} lies in the left half plane (LHP), and p_{ik2} lies in the RHP for all k. Substituting (4.23) into (4.16) and inverting gives the pairwise error probability as desired. The inversion may always be accomplished using numerical contour integration; however, in this work, a novel analytical technique based on residues is used [55]. This technique considers characteristic functions of the form in (4.23), and applies to the general case of a combination of both distinct poles and poles with arbitrary multiplicity, giving a solution that is both exact and numerically stable.

Recall that in (4.23), k indexes those channel usages of the ith error event for which $\mathbf{c}_i\left(k\right)$ is not equal to the all-ones vector. Denote the number of such channel usages as K_i. Analogous to the results of the previous chapter, the pairwise error probability for the ith error event varies asymptotically as Γ^{-K_i}. Since the BER is a sum of the pairwise error probabilities, the asymptotic BER varies as $\Gamma^{-K_{i,min}}$ where $K_{i,min}$ is the smallest K_i in the set of all error events (typically corresponding to the shortest error event). In other words, the asymptotic order of diversity is equal to $K_{i,min}$. Clearly, the diversity order is strongly dependent on the constraint length of the code: larger constraint lengths result in longer error events which result in a larger $K_{i,min}$.

4.5 Results

In this section the proposed downlink scheme is demonstrated using rate one half maximum free distance convolutional codes [8]. Such codes may not be optimal in terms of BER; they are used simply for illustrative purposes. Although not considered here, trellis codes may be a good choice. With trellis codes, the same analysis as presented in this chapter applies, except that N_u is always less than or equal to unity for all combinations of Q and L. In all the results presented here, the number of antennas L is equal to the number of users M such that the multiple users occupy the same bandwidth as a single user. The use of more antennas than the number of users allows for an increase in capacity for each user, but the cost is a loss in diversity order compared to the $L = M$ case due to a smaller N_u.

Fig. 4.4 compares the estimated BER using (4.10) to the BER obtained by simulation for a system with $M = 1, 2$, and 4 users, a constraint length five code, and BPSK modulation ($Q = 2$). Clearly, the estimated BER is accurate for all BERs of interest. Note that a three-user system is not considered, since $L = 3$, $Q = 2$ is not an allowed combination in Table 4.1.

As is common practice, the BER in this graph is plotted against the receive SNR per information bit $\Gamma_b = \Gamma/N_b$. The number of bits per vector channel usage N_b is proportional to the number of antennas L — and thus the number of users, since $L = M$. This convention makes it appear that the four-user system has a lower BER than both the two-user and single-user systems. However, systems with different numbers of users should be compared on the basis of the same number of information bits per user N_b/M. Thus, the BER of the single-user system at $\Gamma_b = 6$ dB should be compared with the BER of the two-user system at $\Gamma_b = 3$ dB and that of the four-user system at $\Gamma_b = 0$ dB. Equivalently, the BER could be plotted vs. $M\Gamma_b$ instead of Γ_b. Such a plot appears in Fig. 4.5 which shows that the curve for M users is simply the corresponding curve in Fig. 4.4 shifted to the right by $10\log_{10} M$ dB — a 3 dB shift for two users, and a 6 dB shift for four users.

From either Fig. 4.4 or Fig. 4.5, it can be seen that the performance degrades as the number of users increases. However, it is still possible to obtain good performance for all users at reasonably low SNRs using a short constraint length code. The cost of adding additional users is a loss in diversity order. As the number of users increases, N_u decreases which decreases the diversity order as explained in Section 4.2. For the example code used here, the diversity order for $M = 1$ is seven. Because BPSK modulation is used, this is

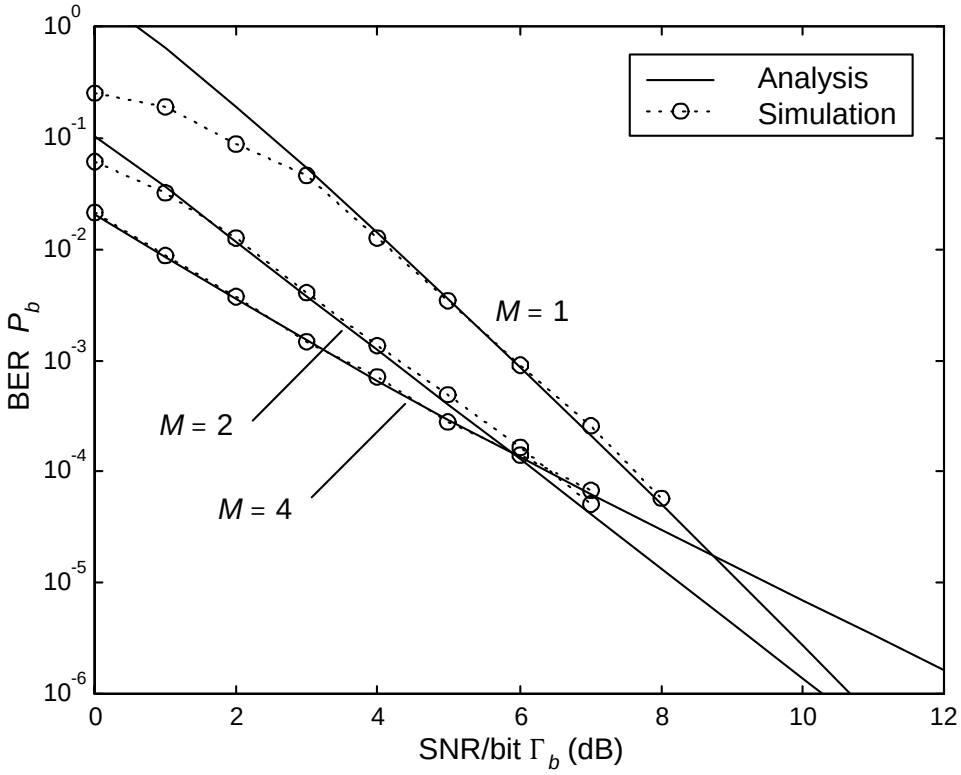

Figure 4.4: Estimated and simulated BER with a (2,1,5) code and BPSK modulation.

equal to the free distance of the code. For $M = 2$, the diversity order is five, and for $M = 4$ where trellis merging is required ($N_u = \frac{1}{2}$), the diversity order is three.

Fig. 4.6 shows the variation in BER with the constellation density Q for both a two-user and a four-user system with a constraint length seven code. As predicted, the decrease in N_u with increasing constellation density causes a reduction in diversity order. With a dense constellation and a large number of users, the reduction can be significant. For the two-user system, the diversity order is equal five, four, and three for BPSK ($Q = 2$), QPSK ($Q = 4$), and 8PSK ($Q = 8$) respectively. For the four-user system, the diversity order is three, two, and two respectively. Order two diversity for the four-user/8PSK system is consistent with the fact that the merged trellis is fully connected, which occurs when $1/N_u = L_c - 1 = 6$.

Fig. 4.7 illustrates the cost in total transmit power to support M users with a constellation density Q. This graph is generated by plotting the per-user transmit SNR Γ_T in (4.4),

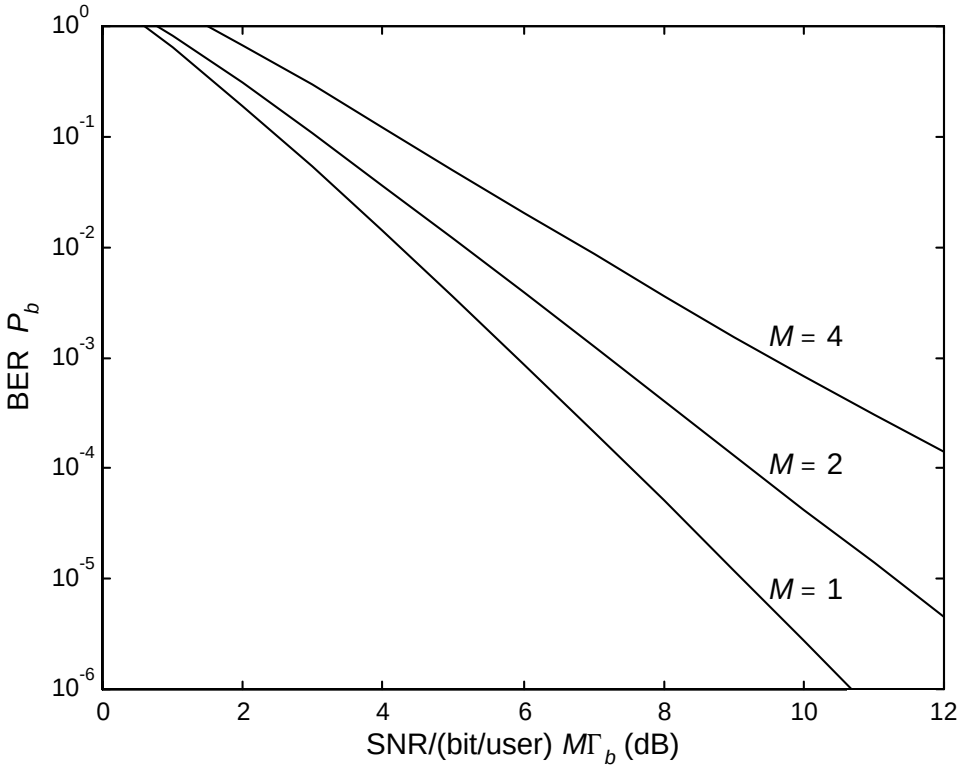

Figure 4.5: BER plotted against the SNR/(bit/user) $M\Gamma_b$ for a (2,1,5) code and BPSK modulation.

calculated using the receive SNR/bit Γ_b corresponding to a BER of 10^{-3}. Evidently, the required transmit power per user increases with the number of users, and as the constellation density grows, the increase becomes faster. However, observation of the BPSK curve indicates that four users may be supported with a transmit power only 2.5 dB per user greater than in a single user system — not a large cost considering the system capacity gain that may be realized.

4.6 Conclusions

In this chapter, a simple transmitter structure is proposed for supporting multiple intracell users in the same time/frequency slot in the downlink using a transmit antenna array at the base station and joint detection at the single-antenna mobile receivers. The structure is

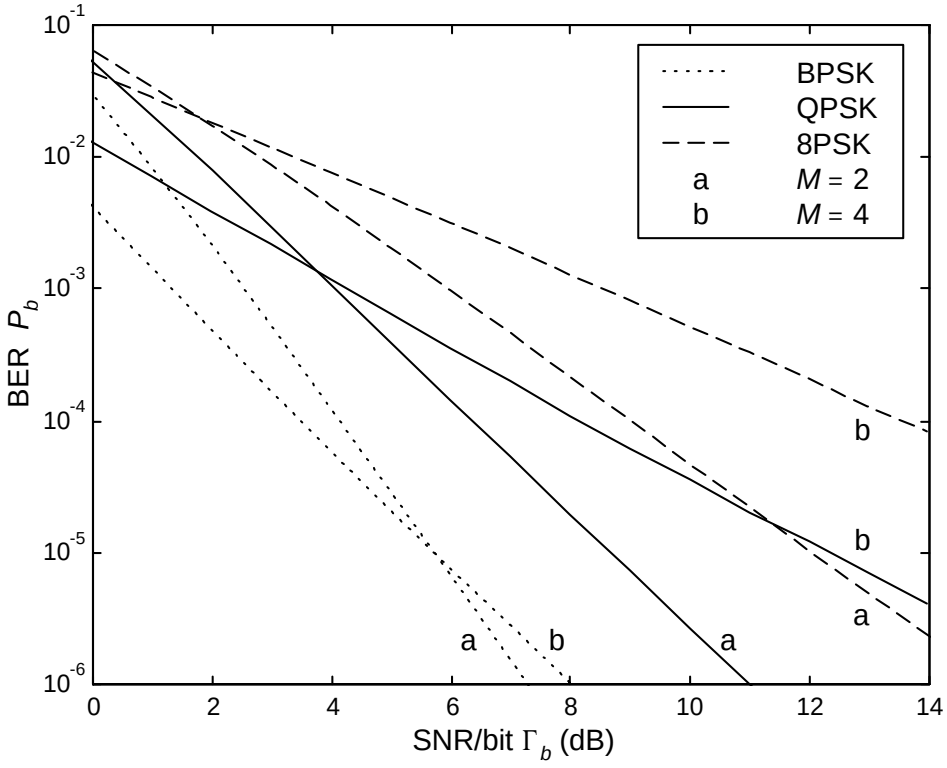

Figure 4.6: BER with a (2,1,7) code and various modulation types.

based upon multiplexing the users' bit sequences together, then employing standard coding and interleaving of the composite sequence in order to supply temporal diversity. The bandwidth required to transmit the high-rate coded sequence is reduced by simultaneously transmitting several successive code symbols on different antennas. At the mobile receivers, soft decision decoding is employed using the Viterbi algorithm with a metric that implicitly performs joint detection of the cochannel signals transmitted from the multiple antennas.

The simultaneous transmission of successive code symbols causes a single channel usage to span several trellis transitions. To handle this unusual situation in both analysis and decoder implementation, the concept of a merged trellis is introduced in which the trellis is modified such that a single channel usage spans only one transition. Using the merged trellis, the optimal decoder is identified, and an analytical expression for the average bit-error rate is derived. Unusual behaviour is demonstrated in terms of diversity order: as

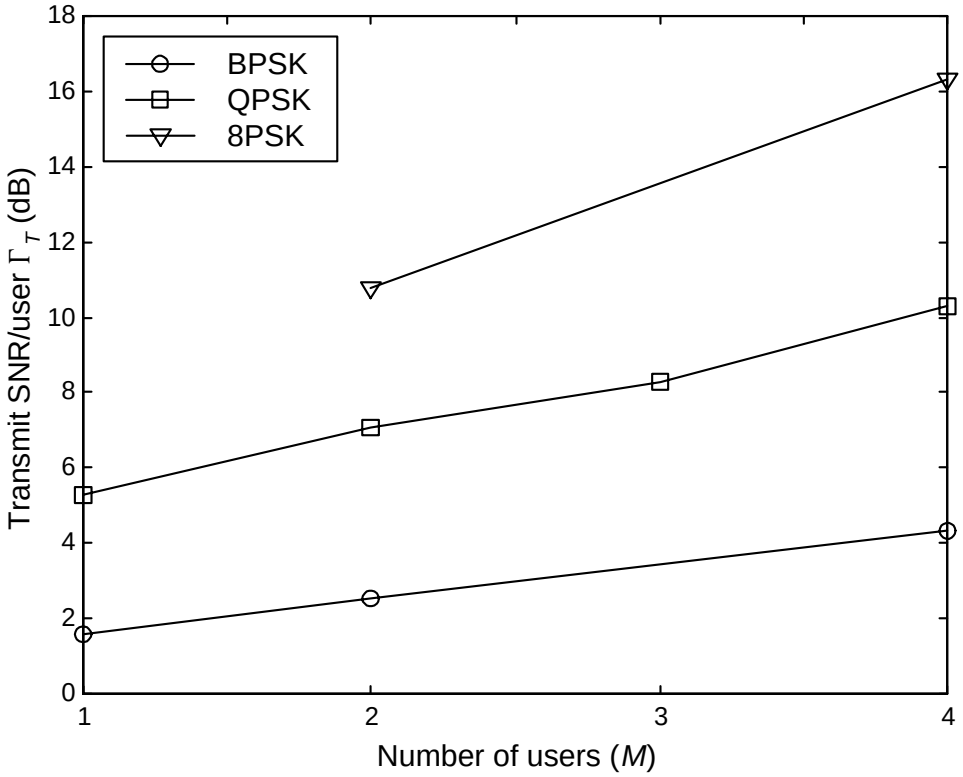

Figure 4.7: Per-user transmit SNR Γ_T required to achieve a BER of 10^{-3} using a (2,1,7) code.

the number of antennas increases, due to an increasing number of users, the diversity order actually decreases due to the simultaneous transmission of successive code symbols.

Even with the loss in diversity, the method provides a major increase in downlink capacity while maintaining good performance for all users at low signal-to-noise ratios with moderate computational load. For example, using a rate one half, constraint length seven code with BPSK modulation and four antennas, four users may be supported in the same bandwidth as a single-user at a BER of 10^{-3} with only 2.5 dB/user more transmit power than a single-user system.

Chapter 5

Multiuser Channel Estimation

In this chapter, a practical multiuser channel estimation technique that may be used with joint detection is developed. While the technique applies to the most general situation of frequency-selective fading and asynchronous users, it includes the flat-fading/synchronous user situation considered in Chapters 3 and 4 as a special case.

The use of pilot symbols is a well-known method for obtaining good channel impulse response estimates in single-user systems, e.g., [57]–[61]. For the case of multiuser systems, pilot-based channel estimation has been studied extensively only for CDMA, e.g., [62] and [63], where processing gain suppresses interference in the channel estimator. However, little is known about pilot-based multiuser channel estimation in TDMA systems. The sole prior study appears to be [33], in which the channels are treated as time-invariant.

In this thesis, a pilot-based MMSE technique for estimating the channel impulse responses of multiple cochannel users in a TDMA system is adopted. This work makes two key contributions which also appear in [65][1] and [64]: first, the analysis accounts for time variation of the channels within and between the training sequences. This feature, previously not addressed, is essential in a multiuser environment where the training sequences are necessarily longer than in a single-user environment [33], resulting in time variation during the training periods.

Second, this chapter addresses the design of appropriate training sequences. In contrast to the single-user case, for which it is often possible to select a single training sequence with perfect autocorrelation properties (as done in [58] and [59]), it is difficult to select

[1]

multiple training sequences of arbitrary length with both perfect auto- and crosscorrelation properties. To overcome these difficulties, reasonable selection criteria are presented for designing good, suboptimal training sequences with training symbols constrained to lie within the modulation alphabet.

5.1 Signal and Channel Models

Fig. 5.1 shows a diagram of the transmission of M cochannel signals through independently fading, dispersive channels. Each channel is represented by its time-variant channel impulse response $g_m(\tau;t)$ where τ represents the memory of the impulse response, and t represents time variation. The mth user's transmitted signal is given by

$$s_m(t) = A_m \sum_n c_m(n)\, u(t - nT - \tau_m) \tag{5.1}$$

where $c_m(n)$ is a data or training symbol for the mth user normalized such that $E\left[|c_m(n)|^2\right] = 1$, $u(t)$ is a root-Nyquist pulse with autocorrelation function $x(\alpha) = \int u(t)\, u^*(t-\alpha)\, dt$ and energy $x(0) = 1$, and A_m is related to the average transmit power P_m by $A_m = \sqrt{2P_m}$. The relative delay τ_m appears in (5.1) since the signals are considered to be asynchronous.

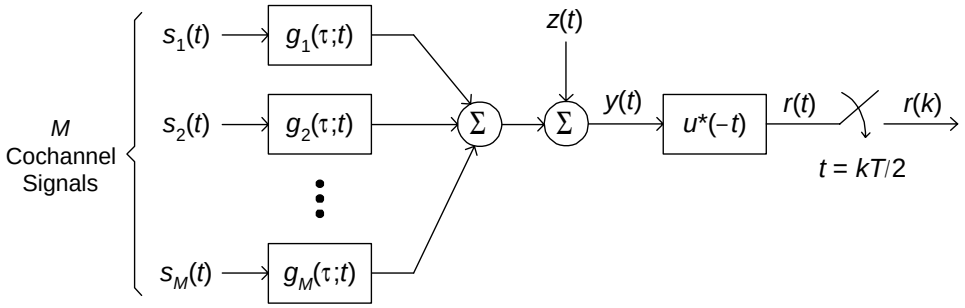

Figure 5.1: Cochannel signal model. The M time-variant, dispersive channels are assumed to fade independently.

The received signal $y(t)$ consists of the sum of the M filtered cochannel signals and an additive white Gaussian noise component $z(t)$ with double-sided power spectral density N_o. Assuming that the time variations of the channels are slow enough such that $g_m(\tau;t)$ does not vary significantly over the memory of the transmit pulse (a few symbols), the output of

the matched filter $u^*(-t)$ is

$$\begin{aligned} r(t) &= \sum_{m=1}^{M} A_m \sum_n c_m(n) \int_{-\infty}^{\infty} g_m(\tau;t)\, x(t-nT-\tau-\tau_m)\, d\tau + n(t) \\ &= \sum_{m=1}^{M} \sum_n c_m(n)\, h_m(t-nT;t) + n(t) \end{aligned} \tag{5.2}$$

where $h_m(\tau;t)$ is the mth user's composite impulse response given by

$$h_m(\tau;t) = A_m g_m(\tau-\tau_m;t) \circledast x(\tau). \tag{5.3}$$

The operator $\circledast$ denotes convolution. Notice that the relative delay τ_m is now considered to be part of the channel impulse response. In (5.2), $n(t)$ is the noise component of the matched filter output. Since $z(t)$ is white, the autocorrelation function of $n(t)$ is $\phi_n(\alpha) = N_o x(\alpha)$.

5.1.1 Channel Statistics

In this thesis, an MMSE estimation technique is adopted which leads naturally to considering time variation of the channels within and between training periods. MMSE estimation of the users' channels requires knowledge of the second order statistics of $h_m(\tau;t)$ summarized by the correlation function

$$\begin{aligned} R_{h_m}(\tau_1,\tau_2,\alpha) &= \frac{1}{2}E\left[h_m(\tau_1;t)\, h_m^*(\tau_2;t-\alpha)\right] \\ &= A_m^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2}E\left[g_m(\lambda_1-\tau_m;t)\, g_m^*(\lambda_2-\tau_m;t-\alpha)\right] \\ &\quad \cdot x(\tau_1-\lambda_1)\, x(\tau_2-\lambda_2)\, d\lambda_1 d\lambda_2. \end{aligned} \tag{5.4}$$

Assuming a wide sense stationary uncorrelated scattering (WSSUS) channel as well as a separable scattering function, the correlation function reduces to

$$R_{h_m}(\tau_1,\tau_2,\alpha) = A_m^2 \frac{R_{g_m}(\alpha)}{\sigma_{g_m}^2} \cdot \int_{-\infty}^{\infty} G_m(\tau)\, x(\tau_1-\tau)\, x^*(\tau_2-\tau)\, d\tau. \tag{5.5}$$

In this expression, $R_{g_m}(\alpha)$ is the temporal autocorrelation function of the channel. For example, with isotropic scattering, $R_{g_m}(\alpha) = \sigma_{g_m}^2 J_0(2\pi f_{D_m}\alpha)$, where $J_0(\cdot)$ is the Bessel function of the first kind and f_{D_m} is the maximum Doppler shift. The function $G_m(\tau)$ is given by

$$\begin{aligned} G_m(\tau) &= \frac{1}{2}E\left[\left|g_m(\tau-\tau_m;t)\right|^2\right] \\ &= \int_{-\infty}^{\infty} P_{g_m}(\tau-\tau_m)\, f_m(\tau_m)\, d\tau_m \end{aligned} \tag{5.6}$$

where $f_m(\tau_m)$ is the PDF of the relative delay τ_m (often uniform over $[-T/2, T/2]$) and $P_{g_m}(\tau)$ is the power-delay profile (PDP) of the channel. One common example is the exponential PDP $P_{g_m}(\tau) = \left(\sigma_{g_m}^2/\tau_{rms_m}\right)\exp\left[-\tau/\tau_{rms_m}\right]$, where τ_{rms_m} is the root-mean-square (RMS) delay spread. Another is the single-spike profile for flat fading given by $P_{g_m}(\tau) = \sigma_{g_m}^2\delta(\tau)$ where $\delta(\tau)$ is the Dirac delta function. Observing (5.6), $G_m(\tau)$ is given by the convolution of the PDP and PDF of τ_m. Because the uncertainty in timing is often much larger than the RMS delay spread, the shape of $G_m(\tau)$, and thus the performance of the MMSE estimator, is not very sensitive to τ_{rms_m}.

Evidently, $R_{h_m}(\tau_1, \tau_2, \alpha)$ does not depend on τ_m itself — only on its PDF. In other words, explicit timing recovery is unnecessary; the relative delays are simply estimated as part of the channels. However, if the relative delays happen to be known, the PDF of τ_m becomes and impulse and $G_m(\tau) = P_{g_m}(\tau - \tau_m)$.

5.1.2 SNR Definition

The SNR of the mth user is defined as $\Gamma_m = E_{s_m}/N_o$, where E_{s_m} is the average received energy per symbol from that user, given by

$$\begin{aligned} E_{s_m} &= \int_{-\frac{T}{2}}^{\frac{T}{2}} \frac{1}{2} E\left[\left|g_m(\tau;t) \circledast s_m(t)\right|^2\right] dt \\ &= \int_{-\infty}^{\infty} P_{g_m}(\tau) \int_{-T/2}^{T/2} E\left[\left|s_m(t-\tau)\right|^2\right] dt d\tau. \end{aligned} \tag{5.7}$$

Again, a WSSUS channel has been assumed. The inner integral of the latter expression is simply equal to twice the average power in $\widetilde{s}_m(t)$. Furthermore, since the area under $P_{g_m}(\tau)$ is $\sigma_{g_m}^2$, the mth user's SNR is simply

$$\Gamma_m = \frac{2P_m\sigma_{g_m}^2}{N_o} = \frac{A_m^2\sigma_{g_m}^2}{N_o}. \tag{5.8}$$

5.1.3 Channel Vectors

Samples of the matched filter output $r(t)$ are taken at times $t = kT/2$ yielding the discrete-time sequence

$$r(k) = \sum_{m=1}^{M} \mathbf{c}_m^T(k)\,\mathbf{h}_m(k) + n(k), \qquad k = \cdots, -2, -1, 0, 1, 2, \cdots \tag{5.9}$$

where the symbol vector $\mathbf{c}_m(k)$ is given by

$$\mathbf{c}_m(k) = \begin{cases} \left[\begin{array}{ccccccccc} c_m\left(\lfloor\frac{k}{2}\rfloor - L_1\right) & 0 & \cdots & c_m\left(\lfloor\frac{k}{2}\rfloor\right) & 0 & \cdots & c_m\left(\lfloor\frac{k}{2}\rfloor - L_2\right) & 0 \end{array}\right]^T, & k \text{ even} \\ \left[\begin{array}{ccccccccc} 0 & c_m\left(\lfloor\frac{k}{2}\rfloor - L_1\right) & \cdots & 0 & c_m\left(\lfloor\frac{k}{2}\rfloor\right) & \cdots & 0 & c_m\left(\lfloor\frac{k}{2}\rfloor - L_2\right) \end{array}\right]^T, & k \text{ odd} \end{cases} \tag{5.10}$$

and mth user's channel vector — to be estimated — is

$$\mathbf{h}_m(k) = \begin{bmatrix} h_m\left(L_1 T; \frac{kT}{2}\right) \\ h_m\left(\left(L_1 + \frac{1}{2}\right)T; \frac{kT}{2}\right) \\ \vdots \\ h_m\left(0; \frac{kT}{2}\right) \\ h_m\left(\frac{T}{2}; \frac{kT}{2}\right) \\ \vdots \\ h_m\left(L_2 T; \frac{kT}{2}\right) \\ h_m\left(\left(L_2 + \frac{1}{2}\right)T; \frac{kT}{2}\right) \end{bmatrix} \tag{5.11}$$

where L_1 and L_2 are integers. Clearly, $\mathbf{h}_m(k)$ consists of samples of $h_m(\tau;t)$ at $T/2$-spaced delays evaluated at time $t = \frac{kT}{2}$. It is assumed that $h_m(\tau;t)$ is generally non-causal such that $L_1 \leq 0$ and $L_2 \geq 0$.

The second order statistics of $\mathbf{h}_m(k)$ are summarized by the autocorrelation matrix $\mathbf{R}_{\mathbf{h}_m}(j) = \frac{1}{2}E\left[\mathbf{h}_m(k)\,\mathbf{h}_m^{\dagger}(k-j)\right]$. Using (5.5), the u,vth element of this matrix is

$$\left\{\mathbf{R}_{\mathbf{h}_m}(j)\right\}_{u,v} = R_{h_m}\left(\frac{uT}{2}, \frac{vT}{2}, \frac{jT}{2}\right) \tag{5.12}$$

where $u, v \in \{2L_1, \cdots, 0, \cdots 2L_2 + 1\}$. Evidently, the tap gains (elements of $\mathbf{h}_m(k)$) are correlated even though a WSSUS channel has been assumed [60]. This is due to the convolution of $g_m(\tau - \tau_m; t)$ with the pulse autocorrelation function $x(\tau)$ as shown in (5.3).

Observing (5.11), the length of $\mathbf{h}_m(k)$ is $2(L_c + 1)$ where $L_c = L_2 - L_1$. To maintain computational complexity as low as possible, it is desirable to choose L_1 and L_2 as small as possible, keeping only those channel taps with significant variance. Since the impulse response $h_m(\tau;t)$ decays to zero for large $|\tau|$, the sum of the tap variances, given by trace$[\mathbf{R}_{\mathbf{h}_m}(0)]$, saturates for large L_c.

As a guide for the selection of the minimum L_c required, Fig. 5.2 shows a plot of trace$[\mathbf{R}_{\mathbf{h}_m}(0)]$ vs. L_c for several combinations of RMS delay spread and pulse rolloff — the two factors that directly influence the duration of $h_m(\tau;t)$. In this graph, $G_m(\tau)$ in

(5.6) is calculated using the exponential PDP and a uniform distribution of τ_m over the interval $[-T/2, T/2]$. Evidently, $L_c = 4$ is sufficient for capturing most of the energy in the impulse response for $\tau_{rms} \leq 0.4T$.

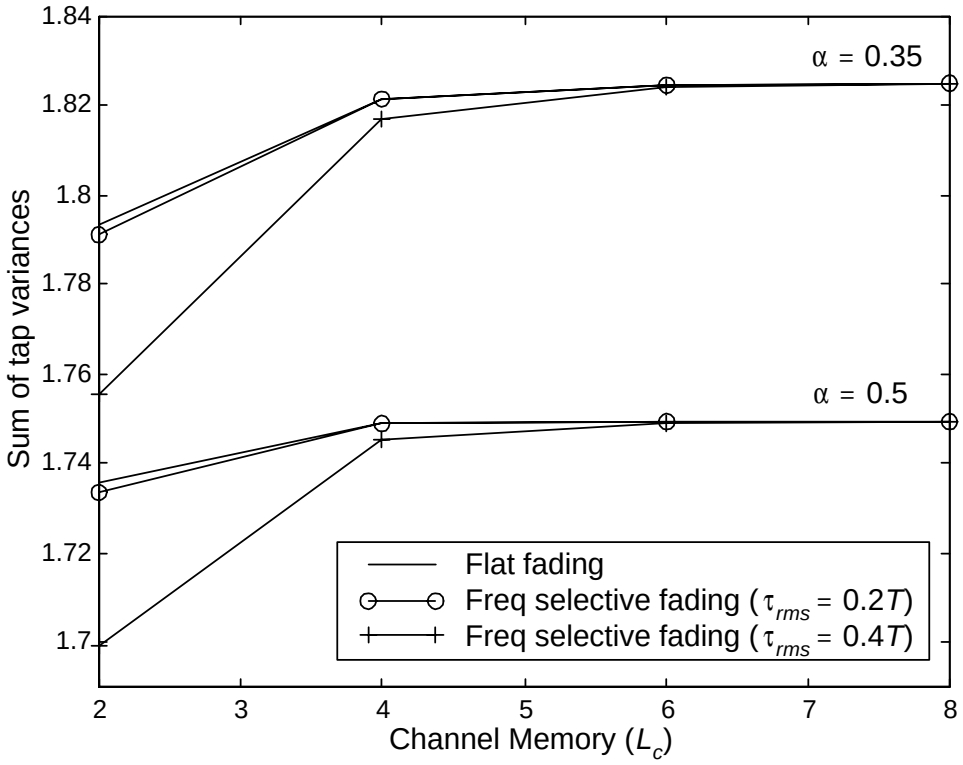

Figure 5.2: Sum of channel tap variances vs. the channel memory length L_c for flat and frequency-selective fading conditions. The variable α is the rolloff parameter of the root-Nyquist transmit pulse $u(t)$.

5.2 Joint Channel Estimation

5.2.1 General Structure of Estimator

MMSE estimation of the users' channels relies upon the periodic insertion of a unique training sequence into each user's data sequence. The design of the training sequences is discussed in Section 5.3. Unique training sequences are required for each user so that the

cochannel signals may be distinguished (e.g., see [33]). This is in contrast to CDMA systems where spreading codes are used to distinguish users and suppress interference in the channel estimator.

It is assumed that the asynchronous users are slot-synchronous such that their training sequences are inserted at the same time, although different propagation delays make their arrivals symbol-asynchronous as discussed previously. The received samples during the training periods are then used to derive estimates of the channels which are interpolated between training periods. In this way, time variations of the channels are tracked. The frame structure, along with the symbol and frame indexing conventions used throughout this chapter, is shown in Fig. 5.3. In this structure, the length of each frame is N symbols, and the length of each training sequence is N_t symbols. Note that n indexes symbols, and k indexes samples; thus, $n = \lfloor k/2 \rfloor$. To provide detail, a single training period and the two adjacent data blocks are shown in exploded view. The exploded view shows the zeroth frame which starts at the beginning of the training period and extends to the end of the subsequent data block. The estimation interval extends from mid-frame to mid-frame either side of the training period.

Since the users' channels are to be estimated jointly, the length-$2M(L_c+1)$ vector $\mathbf{h}(k)$ is defined as the concatenation of the M users' individual channel vectors:

$$\mathbf{h}(k) = \begin{bmatrix} \mathbf{h}_1^T(k) & \mathbf{h}_2^T(k) & \cdots & \mathbf{h}_M^T(k) \end{bmatrix}^T. \tag{5.13}$$

The autocorrelation matrix of the joint channel vector is $\mathbf{R}_{\mathbf{h}}(j) = \frac{1}{2}E\left[\mathbf{h}(k)\,\mathbf{h}^{\dagger}(k+j)\right]$. Because the users' channels fade independently, $\mathbf{R}_{\mathbf{h}}(j)$ is block diagonal and is given by

$$\mathbf{R}_{\mathbf{h}}(j) = \begin{bmatrix} \mathbf{R}_{\mathbf{h}_1}(j) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{h}_2}(j) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}_{\mathbf{h}_M}(j) \end{bmatrix} \tag{5.14}$$

where $\mathbf{R}_{\mathbf{h}_m}(j)$ is defined in (5.12).

Consider the MMSE estimation of $\mathbf{h}(k)$ in the estimation interval $\lfloor -N/2 \rfloor - L_2 + 1 \leq n \leq \lfloor N/2 \rfloor - L_2$ shown in Fig. 5.3. The channel estimator uses the received samples from the training blocks of each of the $2Q+1$ frames centered about frame-0 to form its estimate. These samples are contained in the vector

$$\mathbf{z} = \begin{bmatrix} \mathbf{r}^T(-Q) & \cdots & \mathbf{r}^T(0) & \cdots & \mathbf{r}^T(Q) \end{bmatrix}^T \tag{5.15}$$

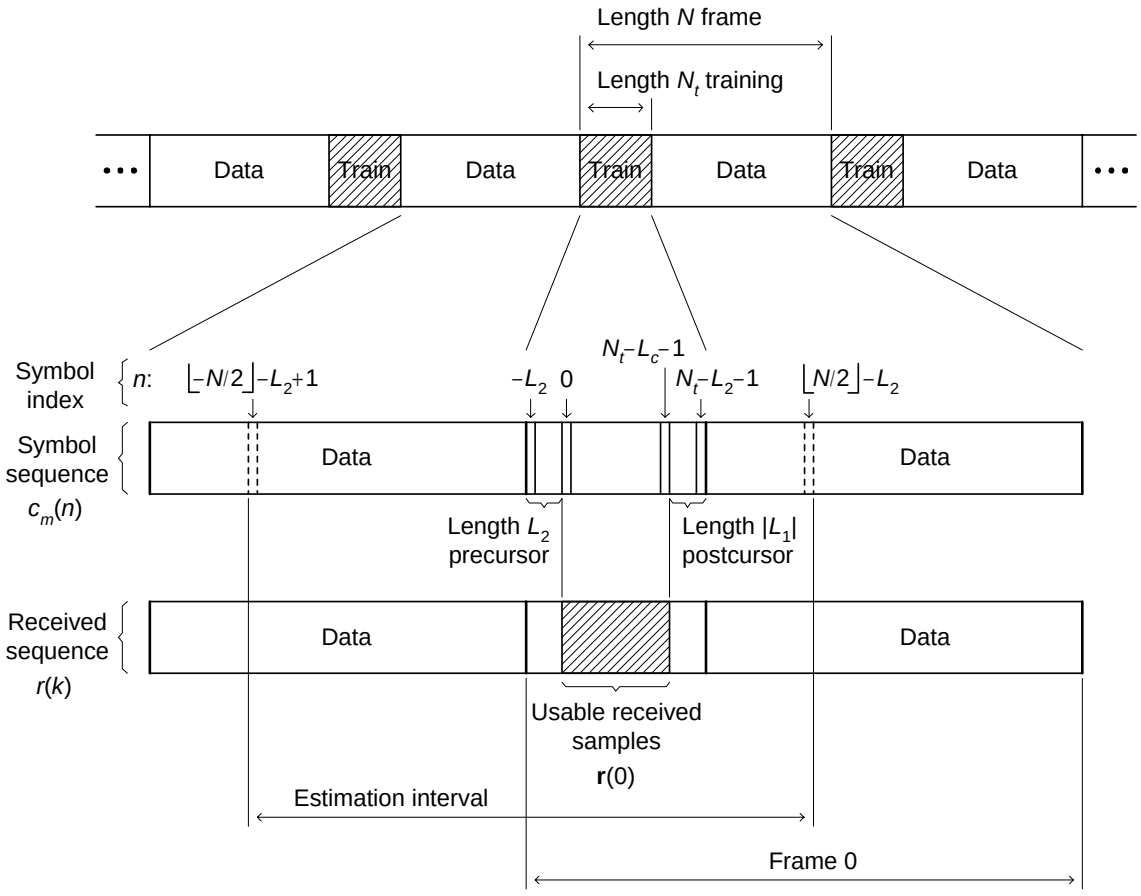

Figure 5.3: Frame structure and indexing conventions.

where

$$\mathbf{r}\,(q) = \left[\begin{array}{cccc} r\,(2qN) & r\,(2qN+1) & \cdots & r\,(2qN+2\,(N_t - L_c) - 1) \end{array} \right]^T. \tag{5.16}$$

The length-$2\,(N_t - L_c)$ vector $\mathbf{r}\,(q)$ contains a subset of the received samples during the qth training block called the "usable samples." For example, Fig. 5.3 shows the usable samples for the zeroth frame ($q = 0$). With use of this subset, $\mathbf{r}\,(q)$ depends only on training symbols — not on unknown data symbols — due to the length-L_2 precursor and the length-$|L_1|$ postcursor inserted in each training sequence.

Since $\mathbf{z}$ contains samples of a bandlimited process sampled at a rate greater than the

Nyquist rate, the covariance matrix of $\mathbf{z}$, given by $\mathbf{R_z} = \frac{1}{2}E\left[\mathbf{z}\mathbf{z}^\dagger\right]$, becomes ill-conditioned as N_t increases (due to an increasing number of users). This suggests the use of rank reduction to remove dependencies in $\mathbf{z}$ as well as to avoid explicit inversion of $\mathbf{R_z}$. Accordingly, eigendecomposition is used to write the covariance matrix of $\mathbf{z}$ as

$$\mathbf{R_z} = \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & \mathbf{0} \\ \mathbf{0} & \Lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{M}_1^\dagger \\ \mathbf{M}_2^\dagger \end{bmatrix}. \tag{5.17}$$

The diagonal matrix Λ_1 contains the dominant eigenvalues of $\mathbf{R_z}$, and Λ_2 contains those eigenvalues that fall below some very small threshold, e.g., $10^{-9}\lambda_{\max}$, where $\lambda_{\max}$ is the maximum eigenvalue of $\mathbf{R_z}$. The non-square matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ contain the normalized eigenvectors, arranged as columns, corresponding to the eigenvalues in Λ_1 and Λ_2 respectively. Now, base the estimate of $\mathbf{h}(k)$ on the reduced dimensionality vector $\mathbf{w} = \mathbf{M}_1^\dagger\mathbf{z}$ (instead of $\mathbf{z}$ itself) which has covariance matrix $\mathbf{R_w} = \mathbf{M}_1^\dagger\mathbf{R_z}\mathbf{M}_1 = \Lambda_1$.

The optimal (MMSE) estimate of $\mathbf{h}(k)$ based on $\mathbf{w}$ is given by the conditional mean $\mathbf{v}(k) = E[\mathbf{h}(k)|\mathbf{w}]$. Since $\mathbf{h}(k)$ and $\mathbf{w}$ are jointly Gaussian, the conditional mean is linear in $\mathbf{w}$, and is given by

$$\begin{aligned} \mathbf{v}(k) &= \left(\frac{1}{2}E\left[\mathbf{h}(k)\,\mathbf{w}^\dagger\right]\right)\mathbf{R_w}^{-1}\mathbf{w} \\ &= \mathbf{P}(k)\,\mathbf{R_z}^{\#}\mathbf{z} \end{aligned} \tag{5.18}$$

where $\mathbf{P}(k) = \frac{1}{2}E\left[\mathbf{h}(k)\,\mathbf{z}^\dagger\right]$ and $\mathbf{R_z}^{\#} = \mathbf{M}_1\Lambda_1^{-1}\mathbf{M}_1^\dagger$. The latter quantity is recognized as the pseudoinverse, or Moore-Penrose generalized inverse, of $\mathbf{R_z}$ [7]. Note that for short training sequences (small N_t), $\mathbf{R_z}$ may not be ill-conditioned; in this case $\mathbf{R_z}^{\#} = \mathbf{R_z}^{-1}$. Evidently, the conditional mean depends upon k, giving a different interpolation matrix $\mathbf{P}(k)\,\mathbf{R_z}^{\#}$ for each position within the estimation interval. However, recall that $\mathbf{h}(k)$ is WSS and $r(k)$ is cyclostationary; thus, it is sufficient to calculate the interpolation matrix for each position in only the estimation interval shown in Fig. 5.3. The same matrix repeats in subsequent frames.

5.2.2 Quality of Channel Estimates

Let $\widehat{\mathbf{v}}(k) = \mathbf{U}(k)\,\mathbf{z}$ denote an arbitrary, not necessarily optimal, estimate of $\mathbf{h}(k)$ with associated channel estimation error

$$\mathbf{e}(k) = \mathbf{h}(k) - \widehat{\mathbf{v}}(k). \tag{5.19}$$

The estimation error covariance matrix is then

$$\begin{aligned}\mathbf{R}_{\mathbf{e}}(k) &= \mathbf{R}_{\mathbf{h}}(0) - \mathbf{R}^{\dagger}_{\hat{\mathbf{v}},\mathbf{h}}(k) - \mathbf{R}_{\hat{\mathbf{v}},\mathbf{h}}(k) + \mathbf{R}_{\hat{\mathbf{v}}}(k) \\ &= \mathbf{R}_{\mathbf{h}}(0) - \mathbf{U}(k)\mathbf{P}^{\dagger}(k) - \mathbf{P}(k)\mathbf{U}^{\dagger}(k) + \mathbf{U}(k)\mathbf{R}_{\mathbf{z}}\mathbf{U}^{\dagger}(k) \end{aligned} \tag{5.20}$$

where $\mathbf{R}_{\hat{\mathbf{v}}}(k)$ is the covariance matrix of $\widehat{\mathbf{v}}(k)$ and $\mathbf{R}_{\hat{\mathbf{v}},\mathbf{h}}(k)$ is the cross-covariance matrix of $\widehat{\mathbf{v}}(k)$ and $\mathbf{h}(k)$. For the optimal channel estimate, $\widehat{\mathbf{v}}(k) = \mathbf{v}(k)$, $\mathbf{U}(k) = \mathbf{P}(k)\mathbf{R}^{\#}_{\mathbf{z}}$ and the error covariance matrix is

$$\mathbf{R}^{opt}_{\mathbf{e}}(k) = \mathbf{R}_{\mathbf{h}}(0) - \mathbf{P}(k)\mathbf{R}^{\#}_{\mathbf{z}}\mathbf{P}^{\dagger}(k). \tag{5.21}$$

In this work, one measure of channel estimation quality for the mth user is the sum of tap error variances, normalized by the sum of the tap variances, that is,

$$\sigma^2_{e_m}(k) = \frac{\text{trace}\left[\mathbf{R}_{\mathbf{e}_m}(k)\right]}{\text{trace}\left[\mathbf{R}_{\mathbf{h}_m}(0)\right]} \tag{5.22}$$

where $\mathbf{R}_{\mathbf{e}_m}(k)$ is the mth block along the main diagonal of $\mathbf{R}_{\mathbf{e}}(k)$.

Another measure of channel estimation quality is the set of correlation coefficients between corresponding taps of the estimated channel vector $\widehat{\mathbf{h}}_m(k)$ and the true channel vector $\mathbf{h}_m(k)$. As shown in the analysis of joint detection in Chapter 3, the channel estimation correlation coefficients are a major determinate of performance. Since each of the $2(L_c+1)$ channel taps has its own correlation coefficient, the following average is defined:

$$\overline{\rho_m}(k) = \frac{1}{2(L_c+1)} \sum_j \frac{\left\{\mathbf{R}_{\hat{\mathbf{v}}_m,\mathbf{h}_m}(k)\right\}_{j,j}}{\sqrt{\left\{\mathbf{R}_{\hat{\mathbf{v}}_m}(k)\right\}_{j,j}}\sqrt{\left\{\mathbf{R}_{\mathbf{h}_m}(0)\right\}_{j,j}}} \tag{5.23}$$

where j indexes the individual channel taps, and $\{\cdot\}_{j,j}$ denotes the diagonal elements of the bracketed matrix. In this expression, $\mathbf{R}_{\hat{\mathbf{v}}_m}(k)$ and $\mathbf{R}_{\hat{\mathbf{v}}_m,\mathbf{h}_m}(k)$ are the mth blocks along the main diagonals of $\mathbf{R}_{\hat{\mathbf{v}}}(k)$ and $\mathbf{R}_{\hat{\mathbf{v}},\mathbf{h}}(k)$ respectively. Although both $\sigma^2_{e_m}(k)$ and $\overline{\rho_m}(k)$ depend on k, it is found that very little variation occurs across the frame. Note that for perfect channel estimation, $\sigma^2_{e_m}(k) = 0$ and $\overline{\rho_m}(k) = 1$.

5.2.3 Details of Optimal Estimator

In this section the optimal estimator in (5.18) is examined in more detail and the required matrices are obtained. It is convenient to first introduce the following data matrix:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_M \end{bmatrix} \tag{5.24}$$

where the mth submatrix of $\mathbf{A}$ is given by

$$\mathbf{A}_m = \begin{bmatrix} \mathbf{c}_m^T(0) \\ \mathbf{c}_m^T(1) \\ \vdots \\ \mathbf{c}_m^T(2(N_t - L_c) - 1) \end{bmatrix}. \tag{5.25}$$

and $\mathbf{c}_m(\cdot)$ is given in (5.10). Due to the precursor and postcursor inserted during each training period, $\mathbf{A}_m$ consists only of symbols from the mth user's training sequence, and no unknown symbols from the adjacent data sequences. Using (5.9), (5.13), (5.24), and (5.25), the jth component of $\mathbf{r}(q)$ can be written as

$$r(2qN + j) = \mathbf{a}_j \mathbf{h}(2qN + j) + n(2qN + j) \tag{5.26}$$

where $\mathbf{a}_j$ is the jth row of the data matrix $\mathbf{A}$ and $j \in \{0, 1, \cdots, 2(N_t - L_c) - 1\}$. With this expression in hand, the elements of the matrices $\mathbf{P}(k)$ and $\mathbf{R_z}$ in (5.18) may be easily determined.

Using (5.15),

$$\mathbf{P}(k) = \begin{bmatrix} \frac{1}{2}E\left[\mathbf{h}(k)\mathbf{r}^\dagger(-Q)\right] & \cdots & \frac{1}{2}E\left[\mathbf{h}(k)\mathbf{r}^\dagger(Q)\right] \end{bmatrix} \tag{5.27}$$

where $\mathbf{P}(k)$ is of dimension $2M(L_c + 1) \times 2(2Q + 1)(N_t - L_c)$. Now using (5.16) and (5.26), and assuming the noise and channel fading process are uncorrelated, the jth column of the qth submatrix of $\mathbf{P}(k)$ is

$$\left\{\frac{1}{2}E\left[\mathbf{h}(k)\mathbf{r}^\dagger(q)\right]\right\}_j = \mathbf{R_h}(k - 2qN - j)\mathbf{a}_j^\dagger \tag{5.28}$$

where $q \in \{-Q, \cdots, Q\}$.

Using (5.15) again,

$$\mathbf{R_z} = \begin{bmatrix} \frac{1}{2}E\left[\mathbf{r}(-Q)\mathbf{r}^\dagger(-Q)\right] & \cdots & \frac{1}{2}E\left[\mathbf{r}(-Q)\mathbf{r}^\dagger(Q)\right] \\ \vdots & \ddots & \vdots \\ \frac{1}{2}E\left[\mathbf{r}(Q)\mathbf{r}^\dagger(-Q)\right] & \cdots & \frac{1}{2}E\left[\mathbf{r}(Q)\mathbf{r}^\dagger(Q)\right] \end{bmatrix} \tag{5.29}$$

where $\mathbf{R_z}$ is of dimension $2(2Q + 1)(N_t - L_c) \times 2(2Q + 1)(N_t - L_c)$. Using (5.16) and (5.26) again, the i, jth element of the q, pth submatrix of $\mathbf{R_z}$ is

$$\left\{\frac{1}{2}E\left[\mathbf{r}(q)\mathbf{r}^\dagger(p)\right]\right\}_{i,j} = \mathbf{a}_i \mathbf{R_h}(2(q - p)N + i - j)\mathbf{a}_j^\dagger + \phi_n(2(q - p)N + i - j) \tag{5.30}$$

where $q,p \in \{-Q,\cdots,Q\}$ and $i,j \in \{0,1,\cdots,2(N_t-L_c)-1\}$. $\phi_n(j) = N_o x\left(\frac{jT}{2}\right)$ is the autocorrelation function of the (coloured) noise sequence.

Observing (5.27)–(5.30), one can see that the optimal interpolation matrix $\mathbf{P}(k)\,\mathbf{R}_{\mathbf{z}}^{\#}$ depends only on the data matrix $\mathbf{A}$, the channel autocorrelation matrix $\mathbf{R}_{\mathbf{h}}(j)$, and the noise autocorrelation function $\phi_n(j)$. For a given scattering environment, the latter two depend on the Doppler fade rate f_{D_m}, the RMS delay spread τ_{rms_m}, and the SNR Γ_m for each user. These parameters may not be known at design time, and may be different for each user. However, in a similar fashion to [57] which considers the case of a single user and flat fading, a worst-case design methodology may be adopted whereby the interpolator is designed assuming worst-case fading conditions and a typical operating SNR. This is discussed further in Section 5.4.4.

5.2.4 Special Cases

For the special case of a single user, the more general results of the preceding sections provide an extension to those contained in [58], since the colouration of the sampled noise sequence and the correlation between channel taps has been considered which is ignored in [58]. As will be shown in Section 5.4.4, the consideration of these effects leads to a significant reduction in estimation error.

For the special case of flat-Rayleigh fading and symbol synchronous users considered in Chapters 3 and 4, the preceding analysis simplifies considerably. In this case, only one sample per symbol is required, which means that the symbol and sample indices (n and k in Fig. 5.3) are the same. With symbol synchronous users, $\tau_m = 0$ in (5.1). Thus, with flat-fading, the composite channel impulse response in (5.3) becomes $h_m(\tau;t) = A_m g_m(t)\,x(\tau)$ where $g_m(t)$ is the complex gain of the channel. After symbol-rate sampling, the joint channel vector to be estimated becomes

$$\mathbf{h}(k) = \begin{bmatrix} A_1 g_1(k) & A_2 g_2(k) & \cdots & A_M g_M(k) \end{bmatrix}^T \tag{5.31}$$

which consists of only one tap per user, i.e., the users complex channel gains. The autocorrelation matrix of the joint channel vector is thus

$$\mathbf{R_h}(j) = \begin{bmatrix} R_{h_1}(j) & 0 & \cdots & 0 \\ 0 & R_{h_2}(j) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{h_M}(j) \end{bmatrix}. \tag{5.32}$$

Since the PDP of the channel is simply $P_{g_m}(\tau) = \sigma_{g_m}^2 \delta(\tau)$, the scaler-valued correlation function $R_{h_m}(j)$ in (5.32) is given by $R_{h_m}(j) = A_m^2 R_{g_m}(jT)$ where $R_{g_m}(\alpha)$ is the temporal autocorrelation function of the channel.

Since the sampled channel is memoryless, no precursor or postcursor is needed in the training sequences, i.e., $L_1 = L_2 = 0$, resulting in a minimum training sequence length of $N_t = M$. The data matrix in (5.24) then becomes

$$\mathbf{A} = \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 & \cdots & \mathbf{t}_M \end{bmatrix} \tag{5.33}$$

where the length-N_t vector $\mathbf{t}_m$ is the training sequence for the mth user. As shown in Section 5.3, the optimal data matrix for the case of $M = 1, 2$, or 4 is given by the Hadamard matrix of order M resulting in orthogonal training sequences.

The optimal (MMSE) estimate of the joint channel vector is still $\mathbf{v}(k) = \mathbf{P}(k)\mathbf{R}_\mathbf{z}^{\#}\mathbf{z}$ where the basis vector $\mathbf{z}$ is given by (5.15). However, because sampling occurs at the symbol rate, the usable received samples during the qth training period (formerly given by (5.16)) are now

$$\mathbf{r}(q) = \begin{bmatrix} r(qN) & r(qN+1) & \cdots & r(qN+N_t-1) \end{bmatrix}^T. \tag{5.34}$$

Thus, the jth column of the qth submatrix of $\mathbf{P}(k)$ becomes

$$\left\{\frac{1}{2}E\left[\mathbf{h}(k)\mathbf{r}^\dagger(q)\right]\right\}_j = \mathbf{R_h}(k - qN - j)\mathbf{a}_j^\dagger \tag{5.35}$$

where $j \in \{0, 1, \cdots, N_t - 1\}$. The i,jth element of the q,pth submatrix of $\mathbf{R_z}$ becomes

$$\left\{\frac{1}{2}E\left[\mathbf{r}(q)\mathbf{r}^\dagger(p)\right]\right\}_{i,j} = \mathbf{a}_i\mathbf{R_h}((q-p)N + i - j)\mathbf{a}_j^\dagger + N_o\delta(q,p,i,j) \tag{5.36}$$

where $q,p \in \{-Q, \cdots, Q\}$ and $i,j \in \{0, 1, \cdots, N_t - 1\}$, and

$$\delta(q,p,i,j) = \begin{cases} 1, & q = p \text{ and } i = j \\ 0, & \text{otherwise} \end{cases}. \tag{5.37}$$

Note that for the special case of a single user ($M = 1$) with a single pilot symbol ($N_t = 1$), (5.35) and (5.36) give identical interpolator coefficients to those derived in [57].

For the optimal channel estimate, the first measure of channel estimation quality defined in (5.22) becomes

$$\sigma^2_{e_m}(k) = \frac{\left\{\mathbf{R}_{\mathbf{e}}^{opt}(k)\right\}_{m,m}}{R_{h_m}(0)} \tag{5.38}$$

where the numerator is the m,mth element of the optimal estimation error covariance matrix defined in (5.21). For the case of $M = 1, 2$, or 4 users where the training sequences are orthogonal, $\mathbf{R}_{\mathbf{e}}^{opt}(k)$ is diagonal, meaning that the users' channel estimation errors are mutually uncorrelated. With non-orthogonal training sequences, $\mathbf{R}_{\mathbf{e}}^{opt}(k)$ is strongly diagonal providing that the training sequences are designed with good crosscorrelation properties. The second quality measure, namely the channel estimation correlation coefficient defined in (5.23), becomes

$$\rho_m(k) = \sqrt{\frac{\left\{\mathbf{P}(k)\,\mathbf{R}_{\mathbf{z}}^{\#}\,\mathbf{P}^{\dagger}(k)\right\}_{m,m}}{R_{h_m}(0)}} \tag{5.39}$$

for the optimal channel estimate. Notice that an average is no longer required since the channel for each user consists of only one tap.

5.3 Training Sequence Design

Optimal selection of the users' training sequences requires testing all possible combinations of M length-N_t symbol sequences in order to minimize each user's channel estimation error $\sigma^2_{e_m}(k)$ defined in (5.22). For several users and practical training sequence lengths, the resulting search space is prohibitively large; furthermore, the amount of computation required to test each candidate sequence is high. In order to overcome these difficulties, a simplified, suboptimal search strategy is developed below which not only yields good training sequences, but offers more insight than an exhaustive computer search.

In the development of this suboptimal search strategy, several assumptions are made: first, the users' channels are assumed to vary slowly enough that they may be considered constant over the duration of each training period; second, the matrix $\mathbf{R}_{\mathbf{z}}$ in (5.18) is assumed to be non-singular, so that $\mathbf{R}_{\mathbf{z}}^{\#} = \mathbf{R}_{\mathbf{z}}^{-1}$; and third, the noise sequence $n(k)$ is

assumed to be white. Under these assumptions, (5.18) may be expressed in an alternate form allowing for a simplified selection criterion. It must be emphasized, however, that these assumptions are made for the purposes of training sequence design only. The resulting sequences are then used to calculate the optimal channel estimate vector $\mathbf{v}(k)$ using (5.18), (5.27), and (5.29) which do not depend on the simplifying assumptions.

Using the slow time variation assumption, the channel vector $\mathbf{h}(2qN+j)$ in (5.26) may be approximated by $\mathbf{h}(2qN)$ for all j. Thus, the received vector $\mathbf{r}(q)$ may be written as

$$\mathbf{r}(q) = \mathbf{A}\mathbf{h}(2qN) + \mathbf{n}(q) \tag{5.40}$$

where $\mathbf{n}(q)$ is the vector of noise samples in the qth training period which has covariance matrix $\mathbf{R_n} = N_o\mathbf{I}$ under the assumption of white noise.

Using the above expression for $\mathbf{r}(q)$, the qth submatrix of $\mathbf{P}(k)$ may be rewritten as

$$\frac{1}{2}E\left[\mathbf{h}(k)\,\mathbf{r}^{\dagger}(-Q)\right] = \mathbf{R_h}(k-2qN)\,\mathbf{A}^{\dagger}. \tag{5.41}$$

The q,pth submatrix of $\mathbf{R_z}$ may be rewritten as

$$\frac{1}{2}E\left[\mathbf{r}(q)\,\mathbf{r}^{\dagger}(p)\right] = \mathbf{A}\mathbf{R_h}(2(q-p)N)\,\mathbf{A}^{\dagger} + N_o\mathbf{I}\delta(q-p) \tag{5.42}$$

where $\delta(l) = 1$ if $l = 0$ and zero otherwise. By using these simplified expressions to form the matrices $\mathbf{P}(k)$ and $\mathbf{R_z}$, and then by using the matrix inversion lemma twice to rewrite the product $\mathbf{P}(k)\,\mathbf{R_z^{-1}}\mathbf{z}$ in (5.18) (see appendix in [58] for the single-user case), the channel estimate vector $\mathbf{v}(k)$ may be expressed as

$$\mathbf{v}(k) = \mathbf{W}(k)\begin{bmatrix} \left(\mathbf{A}^{\dagger}\mathbf{A}\right)^{-1}\mathbf{A}^{\dagger}\mathbf{r}(-Q) \\ \vdots \\ \left(\mathbf{A}^{\dagger}\mathbf{A}\right)^{-1}\mathbf{A}^{\dagger}\mathbf{r}(Q) \end{bmatrix} \tag{5.43}$$

where

$$\mathbf{W}(k) = \left[\mathbf{R_h}(k+2QN)\ \cdots\ \mathbf{R_h}(k-2QN)\right]\cdot\left(\begin{bmatrix} \mathbf{R_h}(0) & \cdots & \mathbf{R_h}(-4QN) \\ \vdots & \ddots & \vdots \\ \mathbf{R_h}(4QN) & \cdots & \mathbf{R_h}(0) \end{bmatrix} + \begin{bmatrix} N_o\left(\mathbf{A}^{\dagger}\mathbf{A}\right)^{-1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & N_o\left(\mathbf{A}^{\dagger}\mathbf{A}\right)^{-1} \end{bmatrix}\right)^{-1}. \tag{5.44}$$

Notice that the term $\left(\mathbf{A}^{\dagger}\mathbf{A}\right)^{-1}\mathbf{A}^{\dagger}\mathbf{r}\left(q\right) \triangleq \widehat{\mathbf{h}}_{LS}\left(q\right)$ in (5.43) is the least-squares (LS) estimate of the channel vector $\mathbf{h}\left(2qN\right)$ during the qth training period, with estimation error $\mathbf{e}_{LS}\left(q\right) = \left(\mathbf{A}^{\dagger}\mathbf{A}\right)^{-1}\mathbf{A}^{\dagger}\mathbf{n}\left(q\right)$, and associated error covariance matrix

$$\mathbf{R}_{\mathbf{e}}^{LS} = N_o\left(\mathbf{A}^{\dagger}\mathbf{A}\right)^{-1}. \tag{5.45}$$

Clearly, $\mathbf{R}_{\mathbf{e}}^{LS}$ is the same for each training block. The channel estimate $\mathbf{v}\left(k\right)$ is then an interpolation (using $\mathbf{W}\left(k\right)$) of the LS estimates made during the $2Q+1$ training periods centered about frame-0. In [33], LS estimation of the users' channels in each of the training periods is also performed; however, since the channels are assumed to be time-invariant, no interpolation between training periods is performed. Furthermore, [33] does not address the design of appropriate training sequences.

Equations (5.43) and (5.44) immediately suggest that the training sequences have a minimum required length. In order to form the LS estimates, the matrix $\mathbf{A}^{\dagger}\mathbf{A}$ must be non-singular. This occurs if the $2\left(N_t - L_c\right) \times 2M\left(L_c+1\right)$ matrix $\mathbf{A}$ is of full column rank, which can only occur if the number of rows of $\mathbf{A}$ is greater than or equal to the number of columns. Consequently, the minimum training sequence length is $N_t = M\left(L_c+1\right) + L_c$.

Equation (5.43) also suggests a simplified criterion for designing good training sequences. Rather than choosing the sequences to minimize $\sigma_{\tilde{e}_m}^2\left(k\right)$ in (5.22) for each user (the optimal criterion), in this work, the sequences are chosen to minimize $\mathrm{trace}\left[\mathbf{R}_{\mathbf{e}}^{LS}\right]$ — an easier task. This is reasonable, since one would expect that minimizing the error variance of the acquired LS estimates during each training block would also lead to a low interpolation error between training blocks. Note, however, that the same minimum sequence length applies to both the optimal and simplified criteria, except that for the optimal criterion, a rank deficiency in $\mathbf{A}$ causes $\sigma_{\tilde{e}_m}^2\left(k\right)$ to be excessively high rather than causing an explicit singularity as for the simplified criterion.

Minimization of $\mathrm{trace}\left[\mathbf{R}_{\mathbf{e}}^{LS}\right]$ is made easier by defining the $\left(N_t - L_c\right) \times M\left(L_c+1\right)$ matrix $\mathbf{B}$ formed by deleting the odd numbered rows of $\mathbf{A}$ and removing the zeros from the even numbered rows. No information is lost here, since, observing (5.10) and (5.25), one can see that an even numbered row of $\mathbf{A}$ and its associated odd numbered row contain the same symbols interspersed with zeros. The odd numbered row is just a right shift by one position of the even numbered row. As a consequence, the even and odd numbered rows are linearly independent such that if $\mathbf{B}$ is of full column rank, then so is $\mathbf{A}$. Furthermore, due to the sparse nature of $\mathbf{A}$, one can show that $\mathrm{trace}\left[\mathbf{R}_{\mathbf{e}}^{LS}\right] = 2N_o\mathrm{trace}\left[\mathbf{G}^{-1}\right]$, where the Gram

matrix $\mathbf{G}$ is given by

$$\mathbf{G} = \mathbf{B}^\dagger\mathbf{B} = \begin{bmatrix} \mathbf{B}_1^\dagger\mathbf{B}_1 & \mathbf{B}_1^\dagger\mathbf{B}_2 & \cdots & \mathbf{B}_1^\dagger\mathbf{B}_M \\ \mathbf{B}_2^\dagger\mathbf{B}_1 & \mathbf{B}_2^\dagger\mathbf{B}_2 & \cdots & \mathbf{B}_2^\dagger\mathbf{B}_M \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_M^\dagger\mathbf{B}_1 & \mathbf{B}_M^\dagger\mathbf{B}_2 & \cdots & \mathbf{B}_M^\dagger\mathbf{B}_M \end{bmatrix}. \tag{5.46}$$

Evidently, selection of the users' training sequences is accomplished simply by minimizing $\text{trace}\left[\mathbf{G}^{-1}\right]$. Although the preceding analysis has been simplified by assuming that the sampled noise sequence is white, it is found that the training sequences that minimize $\text{trace}\left[\mathbf{G}^{-1}\right]$ also minimize the LS estimation error variance for the case of noise colouration due to the matched filter.

In [66] for the case of a single user ($M = 1$), it is shown that $\text{trace}\left[\mathbf{G}^{-1}\right]$ is minimized by choosing a single training sequence such that $\mathbf{G} = \mathbf{B}_1^\dagger\mathbf{B}_1$ is diagonal. This implies that the training sequence must have perfect autocorrelation properties, that is, zero autocorrelation for all lags except zero. The design of such a sequence of arbitrary length is not difficult; for example, see [67].

In the multiuser case, on the other hand, a diagonal $\mathbf{G}$ matrix implies that the M different sequences have not only perfect autocorrelation properties, but perfect crosscorrelation properties as well — that is, zero crosscorrelation for all lags. This is generally very difficult to achieve for arbitrary M and L_c if the training symbols are constrained to lie within the modulation alphabet. In this work, BPSK training sequences are selected such that the off-diagonal elements of $\mathbf{G}$ (autocorrelation values for nonzero lags and crosscorrelation values for all lags) all fall below a certain threshold, which is chosen to be as low as possible for a given M and L_c. Since the diagonal elements of $\mathbf{G}$ are all equal to $N_t - L_c$, this procedure makes $\mathbf{G}$ strongly diagonal.

Training sequence design is made somewhat easier if the first L_c symbols of each user's sequence are constrained to be the same as the last L_c symbols. With this constraint, the mth user's (modified) data matrix is

$$\mathbf{B}_m = \begin{bmatrix} \mathbf{b}_m & T^{-1}\mathbf{b}_m & \cdots & T^{-L_c}\mathbf{b}_m \end{bmatrix} \tag{5.47}$$

where $\mathbf{b}_m$ is a length-$(N_t - L_c)$ column vector, and the operator $T^l\mathbf{b}_m$ denotes a circular shift of $\mathbf{b}_m$ by l positions. The shift is up if l is positive, and down if l is negative. Note that the mth user's training sequence is the concatenation of the last L_c symbols of $\mathbf{b}_m$ and

$\mathbf{b}_m$ itself. The m,nth submatrix of $\mathbf{G}$ can be now be written as

$$\mathbf{B}_m^\dagger \mathbf{B}_n = \begin{bmatrix} \theta_{\mathbf{b}_n,\mathbf{b}_m}(0) & \theta_{\mathbf{b}_n,\mathbf{b}_m}(1) & \cdots & \theta_{\mathbf{b}_n,\mathbf{b}_m}(L_c) \\ \theta_{\mathbf{b}_n,\mathbf{b}_m}(-1) & \theta_{\mathbf{b}_n,\mathbf{b}_m}(0) & \cdots & \theta_{\mathbf{b}_n,\mathbf{b}_m}(L_c-1) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{\mathbf{b}_n,\mathbf{b}_m}(-L_c) & \theta_{\mathbf{b}_n,\mathbf{b}_m}(-L_c+1) & \cdots & \theta_{\mathbf{b}_n,\mathbf{b}_m}(0) \end{bmatrix} \tag{5.48}$$

where $\theta_{\mathbf{x},\mathbf{y}}(l)$ is the periodic crosscorrelation function of the column vectors $\mathbf{x}$ and $\mathbf{y}$ [68] defined as

$$\theta_{\mathbf{x},\mathbf{y}}(l) = \left\langle \mathbf{x}, T^l\mathbf{y} \right\rangle = \left(T^l\mathbf{y}\right)^\dagger \mathbf{x}. \tag{5.49}$$

The design of the training sequences now involves selecting a set of M different $\mathbf{b}_m$'s that satisfy three criteria: i) $|\theta_{\mathbf{b}_m,\mathbf{b}_m}(l)|$ is less than the threshold for $l \in \{1,2,\cdots,L_c\}$ for all m; ii) $|\theta_{\mathbf{b}_n,\mathbf{b}_m}(l)|$ is less than the threshold for $l \in \{-L_c,\cdots,0,\cdots,L_c\}$ for all $m \neq n$; and iii) $\mathbf{B}$ is full column rank. Since the number of combinations of M different $\mathbf{b}_m$'s is huge for several users and typical channel memory lengths, a sequential search is used, rather than an exhaustive one, to build up a set of M training sequences one-by-one that satisfy the three criteria.

Table 5.1 shows the results of a computer search for minimum length binary training sequences that meet these criteria. For compactness, the training sequences are listed in hexadecimal form. The most significant bit corresponds to the first symbol to be transmitted in the training sequence, that is, $c_m(-L_2)$, and the least significant bit to the last symbol, that is, $c_m(N_t - L_2 - 1)$. BPSK symbols are derived from the bits of the training sequences using the mapping $\{0,1\} \rightarrow \{-1,+1\}$. Note that for the case of $M = 1$, $L_c = 1$ where the best training sequence is 001, the constraint of first and last symbols being equal is lifted in order to avoid a singular Gram matrix.

The table also lists the threshold used for the off-diagonal elements of $\mathbf{G}$. For each case, the threshold is first set to zero, and then increased until a full set of training sequences is found. A value of zero indicates that $\mathbf{G}$ is diagonal, implying that the M training sequences have perfect auto and crosscorrelation properties. This occurs when $N_t - L_c = 1, 2$, or 4. It is interesting to note that the resulting data matrix $\mathbf{B}$ is equivalent to the Hadamard matrix of order 1,2, or 4 respectively which is known to have orthogonal columns resulting in a diagonal Gram matrix.

Although results are shown only for binary training sequences, the search technique presented in this work applies to the more general case of nonbinary sequences, e.g., QPSK,

Table 5.1: Minimum length binary training sequences (including pre and postcursors) found by sequential search. The sequences are in hexadecimal form and must be zero-padded to the left to make-up the full length N_t.

M	L_c	N_t	Threshold	Training Sequence(s)
1	0	1	0	1
	1	3	0	1
	2	5	1	9
	3	7	0	11
	4	9	1	21
2	0	2	0	0,1
	1	5	0	11,4
	2	8	2	41,C7
	3	11	2	30B,111
	4	14	2	2C0B,461
3	0	3	1	0,1,2
	1	7	2	41,4,4B
	2	11	3	209,62F,69F
	3	15	4	3053,5065,30BB
	4	19	5	18053,5818B,8361
4	0	4	0	0,3,5,6
	1	9	4	101,4,10B,20
	2	14	4	1009,3047,206A,312F
	3	19	4	30053,50335,10581,70BDF
	4	24	4	B0032B,302DC3,D0657D,E1506E

8-PSK, and 16-QAM. However, experimentation with QPSK sequences showed that while the search space is larger, the auto- and crosscorrelation properties of the resulting sequences are no better than for the binary sequences listed in Table 5.1.

5.4 Design Issues and Performance

In this section, several design issues are treated, namely the choice of interpolator order, choice of frame length, and efficiency. The performance of the joint channel estimation scheme is then investigated, using the channel estimation quality measures $\sigma^2_{\hat{e}_m}(k)$ and $\overline{\rho_m}(k)$ defined in (5.22) and (5.23) respectively. Unless otherwise specified, the optimal interpolator $\mathbf{U}_{opt}(k) = \mathbf{P}(k)\,\mathbf{R}_{\mathbf{z}}^{\#}$ is used so that $\mathbf{R}_{\mathbf{e}}(k)$ in (5.20) is equal to $\mathbf{R}_{\mathbf{e}}^{opt}(k)$.

Frequency selective fading is considered using various values of Doppler spread f_{D_m} and RMS delay spread τ_{rms_m}. The power delay profile is assumed to be exponential, and the

relative delay τ_m is assumed to be distributed uniformly on the interval $[-T/2, T/2]$. Also, $\sigma_{g_m}^2$ is set to 1/2 resulting in the mth user's SNR being $\Gamma_m = P_m/N_o$. Unless otherwise specified, the root-raised cosine transmit pulse $u(t)$ has 50% excess bandwidth. Furthermore, according to Fig. 5.2, the parameter L_c is set to 4 with $L_1 = -2$ and $L_2 = 2$. The training sequences listed in Table 5.1 are used in all cases. Recall that these sequences are of the minimum length $N_t = M(L_c + 1) + L_c$.

5.4.1 Interpolator Order

Fig. 5.4 shows a plot of channel estimation error vs. interpolator order, defined as $2Q+1$. As can be seen, the use of more than about nine training blocks ($Q = 4$) to form the estimate of $\mathbf{h}(k)$ does not significantly decrease the channel estimation error. This behaviour was found to be representative of a large variety of fading and SNR conditions. It is also consistent with that observed in [57] for the case of a single user and flat fading. Note that user 2 experiences slightly better performance than user 1 since the training sequence for user 2 happens to have slightly better autocorrelation properties.

5.4.2 Frame Length

Fig. 5.5 shows a plot of channel estimation error vs. frame length N. For a fixed Doppler spread, as N is increased beyond a critical value, the channel estimation error increases sharply, due to the fact that the fading channels are not sampled often enough to allow proper interpolation. Clearly, as the Doppler spread increases, the fading channels must be sampled at a higher rate (shorter frame length): for $f_{D_m}T = 0.0025, 0.005$, and 0.01, the critical frame lengths are approximately 180, 90, and 45 symbols respectively. These values correspond closely to the inverse of the Nyquist rate $2f_{D_m}T$. Again, this behaviour is consistent with that observed in [57].

5.4.3 Efficiency

The transmission efficiency — or throughput — experienced by any user is given by the ratio of the number of data symbols per frame $(N - N_t)$ to the frame length N. As the number of users increases, so does the required length of training sequence, causing the user efficiency to drop. Using the minimum training sequence length found earlier, the user efficiency is $\eta_u = (N - M(L_c + 1) - L_c)/N$. Fig. 5.6 shows a plot of user efficiency vs. number of

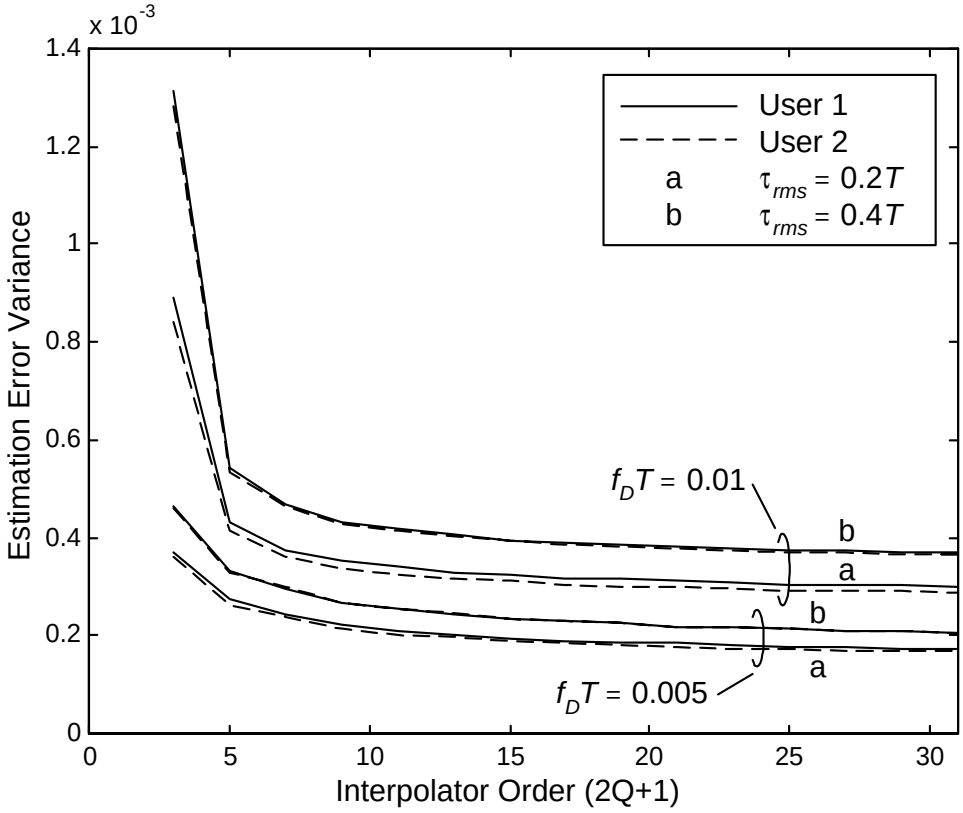

Figure 5.4: Channel estimation error vs. interpolator order for two equipower users at an SNR of 30 dB. The frame length is $N = 2N_t = 28$ symbols.

users for the critical values of N found above. This plot illustrates significantly reduced efficiency for short frame lengths and a large number of users. In the extreme of fast fading ($f_{D_m}T = 0.01$) with four users and a frame length of $N = 45$, the user efficiency drops from its value of 80% corresponding to a single user to a value near 50%. However, with RWC, system capacity is enhanced through joint detection by allowing 4 users to share the same frequency/time slot which offsets this reduction in user efficiency. Therefore, it is useful to define system efficiency as $\eta_s = M\eta_u$ which is also plotted on Fig. 5.6. Evidently, an optimal value of M exists. This optimal value and the corresponding optimal η_s both increase for slower fading where the frame length can be much greater than for fast fading.

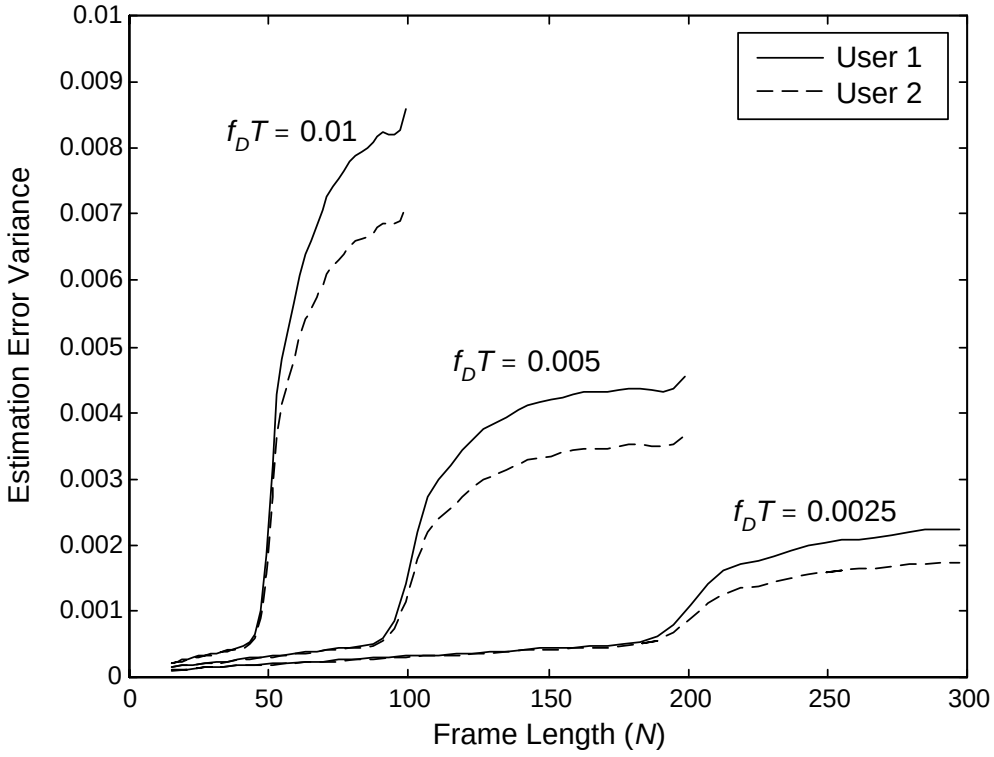

Figure 5.5: Channel estimation error vs. frame length for two equipower users at an SNR of 30 dB with interpolator order nine. The fading is frequency selective with $\tau_{rms_m}/T = 0.2$ for each user.

5.4.4 Performance

Fig. 5.7 shows a plot of channel estimation error vs. number of users. The estimation error actually decreases with each additional user, which is due to longer training sequences as each additional user is added. Again, each user has a slightly different estimation error variance due to the fact that users' training sequences have slightly different autocorrelation properties.

In addition to the consideration of multiple users, the analysis presented in this thesis provides an extension to the single-user results of [58] by accounting for colouration of the sampled noise sequence and correlation between the channel taps. Fig. 5.8 shows the estimation gain achieved by considering these two effects. This graph plots the estimation

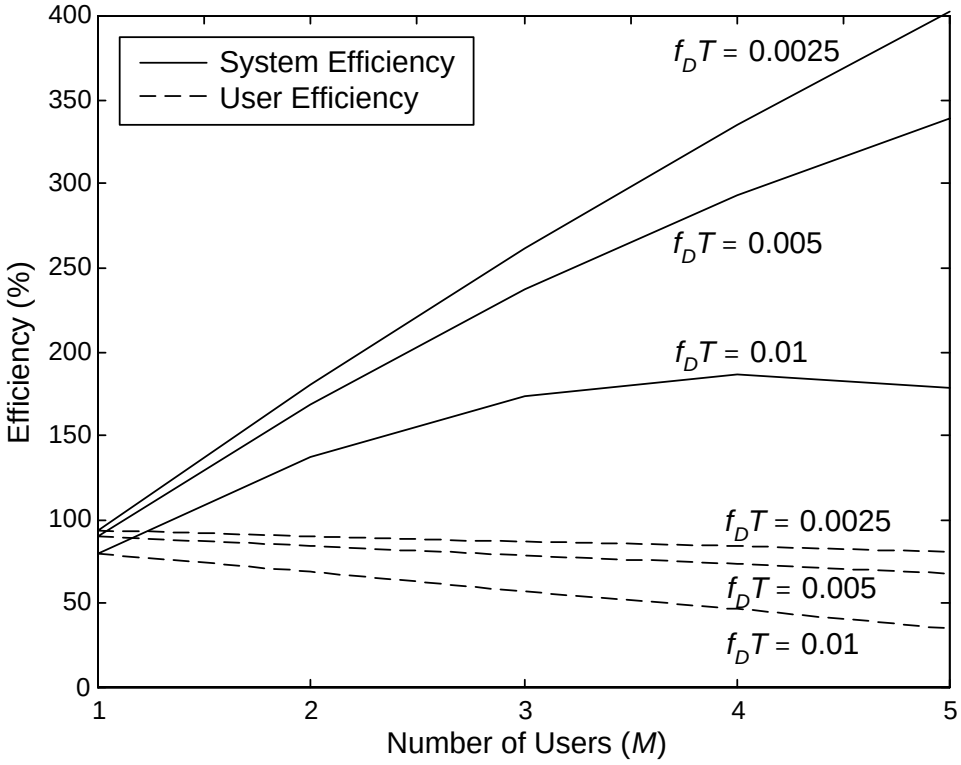

Figure 5.6: User and system efficiency for the critical frame lengths found in the previous figure.

error variance $\sigma^2_{\hat{e}_m}(k)$ and one minus the average correlation coefficient $(1 - \overline{\rho_m}(k))$ for both the optimal interpolator $\mathbf{U}_{opt}(k) = \mathbf{P}(k)\mathbf{R}_{\mathbf{z}}^{\#}$ and the suboptimal interpolator $\mathbf{U}(k) = \mathbf{P}'(k)\mathbf{R}_{\mathbf{z}}^{'\#}$. The matrices $\mathbf{P}'(k)$ and $\mathbf{R}_{\mathbf{z}}^{'}$ are obtained by modifying $\mathbf{P}(k)$ and $\mathbf{R}_{\mathbf{z}}$ as follows: the off-diagonal elements of $\mathbf{R}_{\mathbf{h}}(\cdot)$ in (5.28) and (5.30) are set to zero, and the autocorrelation function of the sampled noise sequence in (5.30) is redefined so that $\phi_n(j) = N_o$ for $j = 0$ and zero otherwise. Observing the estimation error curves, one can see that for moderate SNR, up to a 6 dB gain in estimation error may be achieved by considering the noise colouration and the inter-tap correlations. An even a larger improvement is observed in the correlation coefficient. It should be noted that this benefit comes at little or no cost in computational load when performing the channel estimation.

This plot also demonstrates that the asymptotic channel estimation error varies inversely

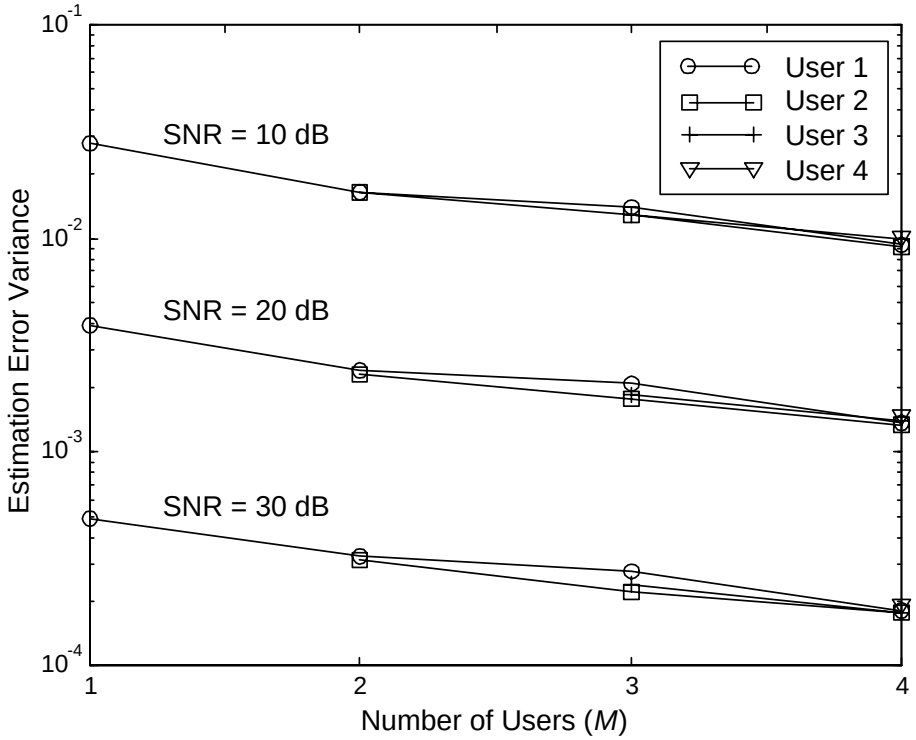

Figure 5.7: Channel estimation error vs. number of equipower users for frame length $N = 50$ and interpolator order nine. The fading is frequency selective with $\tau_{rms_m}/T = 0.2$ and $f_{D_m}T = 0.005$ for each user.

with SNR, as expected. Similarly, the channel estimate and the true channel become increasingly correlated with increasing SNR. For example, at an SNR of 40 dB, use of the optimal interpolator results in an average correlation coefficient of approximately $\overline{\rho_m}(k) = 0.9997$ — a high degree of correlation. This behaviour is responsible for the elimination of the error floor for joint detection observed in Chapter 3 when using pilot-based channel estimation.

The above graphs have all illustrated the performance for equipower signals. However, in normal system operation, one can expect power differences among the users. Even in this case, the *absolute* estimation error variance, given by trace$[\mathbf{R}_{\mathbf{e}_m}(k)]$ (see (5.21)), is approximately equal for all users, since the estimation error is produced primarily by the receiver noise $n(t)$. On the other hand, the *relative* estimation error variance $\sigma^2_{e_m}(k)$, defined in (5.22), is normalized by the sum of channel tap variances. The tap variances, in turn,

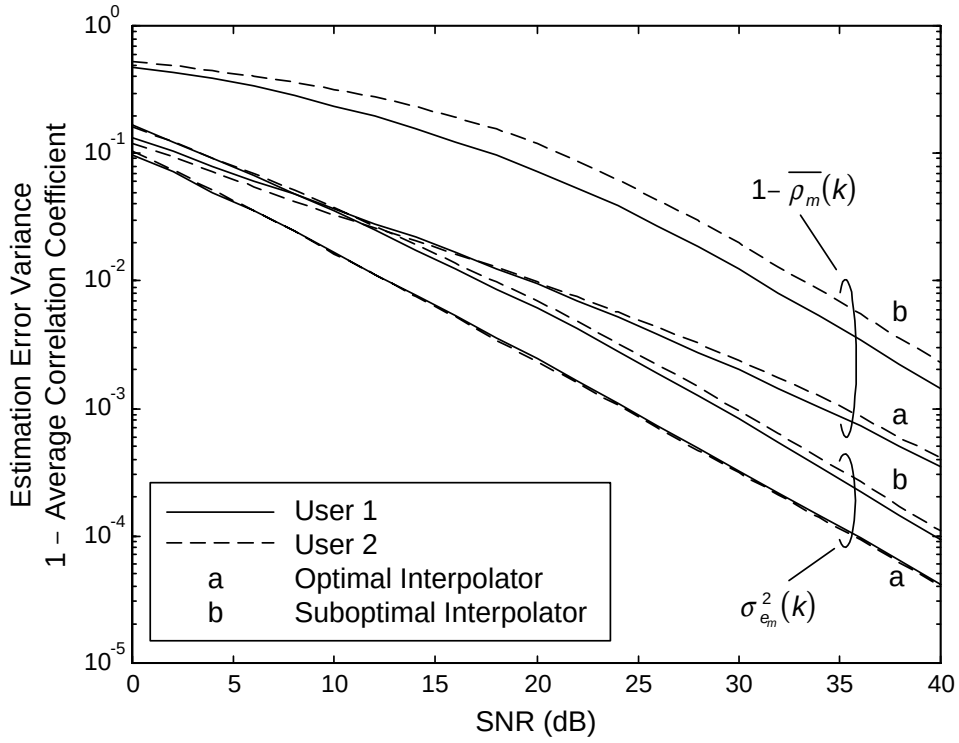

Figure 5.8: Channel estimation error and average correlation coefficient for optimal and suboptimal interpolators with two equipower users, frame length $N = 50$, and interpolator order nine. The fading is frequency selective with $\tau_{rms_m}/T = 0.2$ and $f_{D_m}T = 0.005$ for each user.

are proportional to $A_m^2 = 2P_m$ (see (5.5)). Therefore, at a given noise level, the relative error variance is close to inversely proportional to signal power, giving stronger users better relative error results than weaker users.

Fig. 5.9 shows the relative estimation error variance and average correlation coefficient for the case of two users with a power difference of 10 dB, i.e. $P_1 = 1$ and $P_2 = 0.1$. For reference, the performance of the equipower case ($P_1 = P_2 = 1$) is shown on the same graph. Note that each user's curves are plotted against the user's *own* SNR, defined in (5.8). As a result, the curves are closely spaced which illustrates the approximate inverse dependence on signal power, as discussed above. In a typical operating scenario, all users are detected at a common receiver noise level N_o. Consequently, the weak and strong users operate at SNR values that are 10 dB apart, and the horizontal axis must be interpreted in this light. For

example, if the strong user is at 30 dB SNR, then the weak user is at 20 dB with a relative estimation error variance roughly 10 dB greater than that of the strong user. Similarly, the average correlation coefficient for the stronger user is significantly better (closer to unity) than for the weaker user.

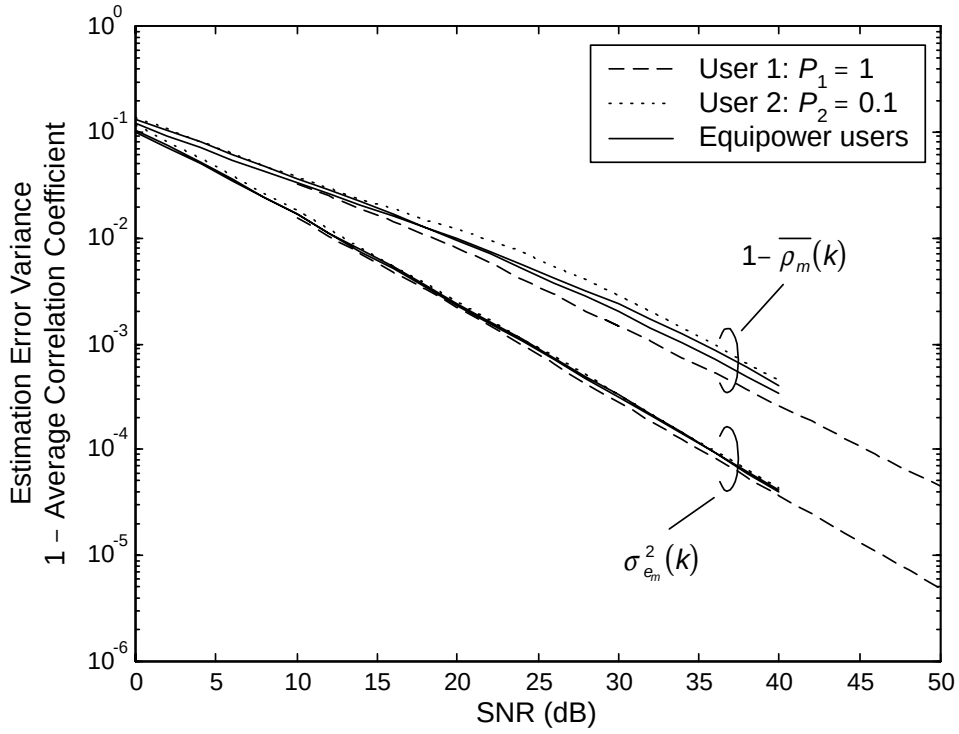

Figure 5.9: Channel estimation error and average correlation coefficient with two nonequipower users, frame length $N = 50$, and interpolator order nine. The fading is frequency selective with $\tau_{rms_m}/T = 0.2$ and $f_{D_m}T = 0.005$ for each user.

In all of the results presented here, it is assumed that the interpolator is designed assuming perfect knowledge of the Doppler fade rate f_{D_m}, the RMS delay spread τ_{rms_m}, and the SNR Γ_m for each user. In a practical situation, these parameters may not be known at design time and may be different for each user. Furthermore, they may change as the scattering environment changes. As mentioned previously, this may be handled by adopting a worst-case design methodology. This approach was investigated here by designing an interpolator using worst-case values of f_{D_m} and τ_{rms_m} (equal for all users) and a typical

operating SNR Γ_m (again, equal for all users). The performance of this fixed interpolator was then investigated in an environment with different channel and SNR conditions.

Not surprisingly, it was found that better performance may be obtained by optimizing the interpolator to match the actual fading and SNR conditions. More importantly, however, it was found that the performance is neither degraded nor improved if the actual channel conditions are better than those designed for (lower f_{D_m} and τ_{rms_m} than the design values) and if the actual SNRs are different from the design value. If the actual channel conditions are worse than those designed for, significant degradation in performance occurs.

5.5 Conclusions

In this chapter, a pilot-based MMSE technique has been developed for jointly estimating the channel impulse responses of multiple cochannel signals. This work makes two key contributions: first, it accounts for time variation of the channels both within and between training periods. The former is essential in a multiuser environment where the training sequences are necessarily longer than in a single-user environment, resulting in significant variation during training. Second, a simple strategy for the selection of appropriate training sequences for the multiple cochannel users is developed. The selection strategy is demonstrated for the special case of BPSK sequences.

Several design issues are considered including the choice of interpolator order, the choice of frame length, and efficiency. Results show that the per-user throughput decreases with each additional user, since the minimum length of the training sequences grows linearly with the number of users. However, system efficiency increases, due to multiple users sharing the same time/frequency slot.

Performance results are presented, and it is shown that the channel estimation error decreases with each additional user due to increasing training sequence lengths. Furthermore, it is shown that for nonequipower users, the absolute channel estimation error variance for all users is approximately equal, irrespective of power differences, but the degree of correlation between the channel estimate and the true channel is higher for the stronger users than the weaker users.

Chapter 6

System Capacity with Joint Detection

In previous chapters, joint detection in both the uplink and downlink has been investigated for single-cell systems only. In other words, the effects of CCI from cochannel cells in a multi-cell system has been ignored. In this chapter, the intercell CCI is accounted for and overall system capacity using joint detection is evaluated.

The goal in using joint detection is to achieve a system capacity increase by allowing several intracell users to share the same frequency/time slot. However, each cochannel cell also allows reuse within cell, causing an increase in the overall level of intercell CCI compared to conventional TDMA systems with only one user/slot. This necessitates an increase in reuse distance and/or an increase in the number of antennas to combat the additional CCI. The natural questions that arise are the following: how do the number of users per sector, the reuse distance, and the number of antennas trade off against each other in determining system capacity, and how much is capacity actually increased using joint detection? Furthermore, how does the capacity increase with joint detection compare to that using classical MMSE antenna combining?

These interesting questions are answered in this part of the thesis. In doing so, the first analysis and the first comparison of the effects of joint detection and MMSE combining on system capacity is presented. The analysis accounts for shadowing and the random location of users in cochannel cells. For the latter, a novel, fully-analytical technique is developed for calculating the spatial average of user locations necessary for the computation

of outage probability. This method avoids time-consuming Monte-Carlo simulation used in past capacity studies, e.g., [69]–[72]. Furthermore, it provides more insight than that obtainable from a simulation-based study. An attractive feature of the technique is that it does not rely upon a rigid hexagonal cell layout, thus enabling the study of more generalized systems in which the cochannel cells are of arbitrary size and location, and the cell sectors are of arbitrary width and orientation.

The analysis shows that joint detection produces a significant increase in capacity, and that capacity is soft-limited, similar to CDMA. In striking contrast, it is shown that the alternative of MMSE combining produces only a modest increase in capacity and puts a hard limit on the number of users per sector, beyond which capacity surprisingly decreases. The above novel contributions also appear in part in [74][1] and [75][2], and in full in [76] and [77].

6.1 System Model

It must be stated at the outset that the evaluation of system capacity in this thesis is focused on the uplink; however, similar models and analysis may be applied to the downlink with suitable modifications. In assessing uplink capacity, the common cellular layout shown in Fig. 6.1 is used. The layout consists of equi-size circular cells with the cell centres located in the same positions as those in a standard hexagonal layout with unit-radius hexagons, cluster size C, and reuse distance $D = \sqrt{3C}$. The allowed cluster sizes are given by $C = j^2 + jk + k^2$ where the integers j and k are the shift parameters of the hexagonal layout [2]. The radius of each circular cell is scaled such that its area is the same as a unit-radius hexagon, resulting in $R = \sqrt{3\sqrt{3}/2\pi}$.

In Fig. 6.1, the hatched areas represent cochannel sectors. The first three tiers of cochannel cells are considered, resulting in a total of 36 cochannel sectors. However, assuming the antennas at the base station in the central cell (cell 0) have an ideal 120° beam pattern, interference is received at the base from only 13 of the cochannel sectors — 2 in the first tier (numbered 1–2), 5 in the second tier (numbered 3–7), and 6 in the third tier (numbered 8–13). The number of desired users in the central cell is M, and it is assumed that

[1] © 1999 IEEE. Reprinted, with permission, from *Proc. IEEE VTC'99-Fall*, Amsterdam, The Netherlands, September 1999.

[2] © 1999 IEEE. Reprinted, with permission, from *Proc. IEEE VTC'99-Fall*, Amsterdam, The Netherlands, September 1999.

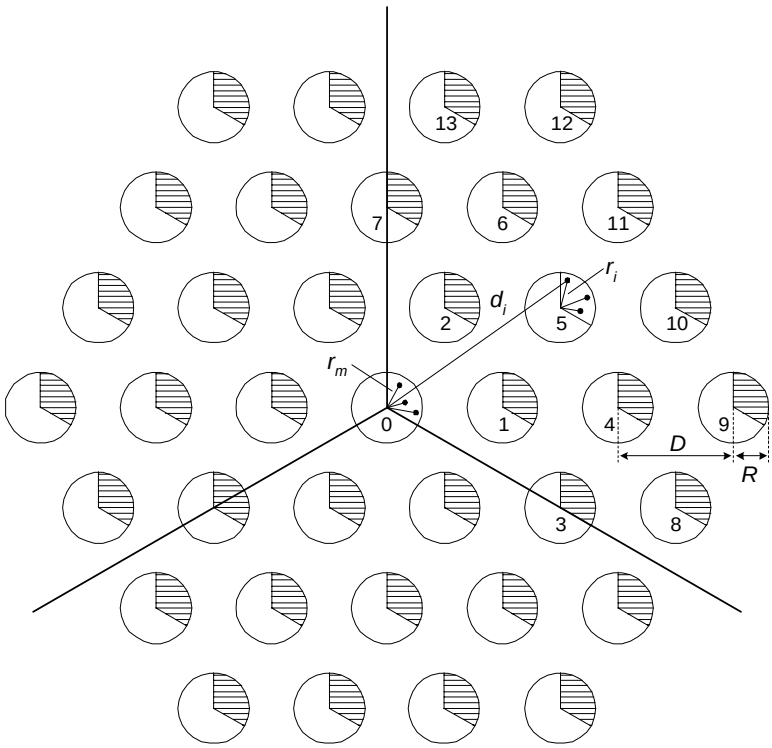

Figure 6.1: Three-tier hexagonal layout of equi-size circular cells with reuse distance $D = \sqrt{3C}$ and ideal 120° sectorization. All cochannel sectors are populated with M active users, resulting in a total of $13M$ interferers.

each cochannel sector contains the same number of active users, resulting in a total of $13M$ interferers. Clearly, allowing several intracell cochannel users increases the overall level of interference compared to a standard TDMA system in which $M = 1$.

The hexagonal cell layout in Fig. 6.1 is a special case of the more general layout shown in Fig. 6.2 in which the central cell and one representative cochannel cell is shown. The following parameters that describe the model are all arbitrary: the inter-base station distance D_i, the cell radii R_m and R_i, the sector widths Δ_m and Δ_i, and the sector orientations θ_m and θ_i. By convention, the sector orientations are defined in a counter-clockwise sense; thus, θ_m and θ_i in Fig. 6.2 are both negative. Note that the desired users are indexed by m, and the interferers by i. This model goes beyond the rigid hexagonal cell layout used in previous capacity studies, e.g., [69], [70], and [73], by modeling additional realism through the above

arbitrary parameters.

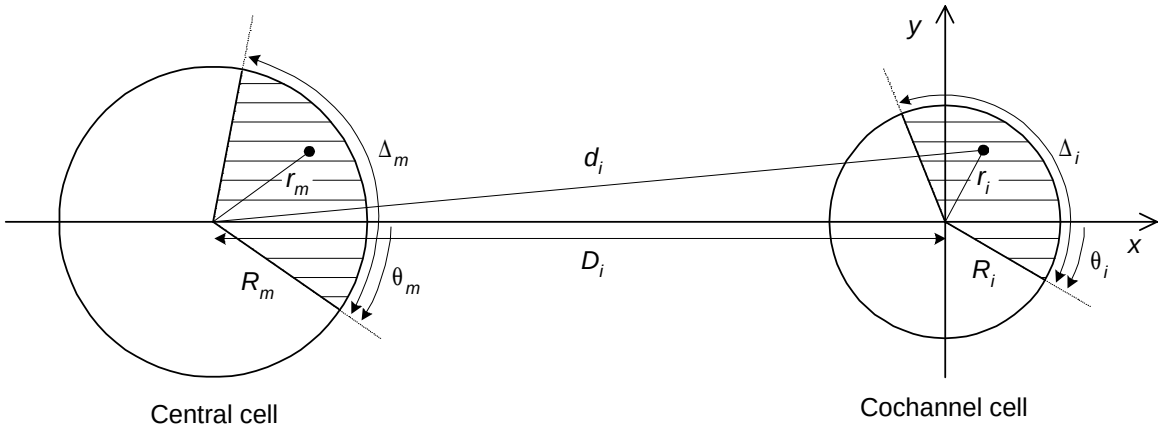

Figure 6.2: Geometry of generalized cell layout.

In Fig.6.2, the desired user is located at a random distance r_m from its own base. The interfering user is also located randomly with a distance of r_i to its own base and a distance of d_i to the base in the central cell. As in other capacity studies, e.g., [70] and [71], each mobile is assumed to communicate with the geographically closest base station. If a mobile ventures outside the circular cell boundaries, it is handed off to another base station and is no longer tracked; thus, at all times, $r_m \leq R_m$ and $r_i \leq R_i$.

The analysis presented in this thesis corresponds to the above generalized layout; as a result, it may be applied to a wide variety of systems. However, the capacity results presented later are tailored to the hexagonal layout in Fig. 6.1 by letting $R_m = R_i = \sqrt{3\sqrt{3}/2\pi}$ and $\Delta_m = \Delta_i = 120^\circ$ for all desired and interfering users, letting $\theta_m = -30^\circ$ for all desired users, and appropriately setting the parameters D_i and θ_i depending on the cell in which the ith interferer is located. Table 6.1 lists the values for these two parameters for each of the 13 cochannel sectors in the hexagonal layout. The distance D_i is expressed in terms of the reuse distance $D = \sqrt{3C}$.

Table 6.1: Cell distance and sector orientation parameters for the standard hexagonal cell layout with 120 degree sectorization.

Cell index	Tier	Distance D_i	Orientation θ_i
1	1	D	-30°
2	1	D	-90°
3	2	$\sqrt{3}D$	0°
4	2	$2D$	-30°
5	2	$\sqrt{3}D$	-60°
6	2	$2D$	-90°
7	2	$\sqrt{3}D$	-120°
8	3	$\sqrt{7}D$	-10°
9	3	$3D$	-30°
10	3	$\sqrt{7}D$	-50°
11	3	$\sqrt{7}D$	-70°
12	3	$3D$	-90°
13	3	$\sqrt{7}D$	-110°

6.2 Signal Models

The signal transmitted from the mth desired user in the central cell is

$$s_m(t) = \sqrt{2E_{s_m}} \sum_n c_m(n)\, p(t - nT) \tag{6.1}$$

where $c_m(n)$ is a symbol from the PSK symbol sequence $\mathbf{c}_m$, E_{s_m} is the average transmitted energy per symbol, $p(t)$ is a root-Nyquist pulse normalized to unit energy, with rolloff parameter β and autocorrelation function $q(t) = p(t) \circledast p^*(-t)$ ($\circledast$ denotes convolution), and T is the symbol period. For convenience, the data symbols are normalized such that $E\left[|c_m(k)|^2\right] = 1$. The signal $s_i(t)$ from the ith interferer has the same form except the subscript 'm' in the above is replaced by the subscript 'i'.

The transmitted signals from both the desired and interfering mobiles undergo three effects, namely log-normal shadowing, path loss, and frequency-flat Rayleigh fading. The effects of log-normal shadowing and path loss are modeled by a random attenuation (gain) of x dB superimposed on a path loss of $\gamma 10 \log_{10} d$ dB. Here x is a zero mean Gaussian RV with standard deviation σ_x (in units of dB), d is the distance from the mobile to the base in the central cell, and γ is the path loss coefficient. For typical cellular systems, σ_x is usually between 6 and 8 dB [69], and γ is between 3 and 4 [70]. Frequency-flat Rayleigh fading is

modeled by the zero mean complex Gaussian random gain process $g(t)$. Accordingly, the received signal on the lth antenna from the mth desired user is

$$y_{lm}(t) = 10^{x_m/20} r_m^{-\gamma_m/2} g_{lm}(t)\, s_m(t) \tag{6.2}$$

and that from the ith interfering user is

$$y_{li}(t) = 10^{x_i/20} d_i^{-\gamma_i/2} g_{li}(t)\, s_i(t-\tau_i)\,. \tag{6.3}$$

The distances r_m and d_i are shown in Fig. 6.2. Unlike the desired users' signals, the interfering users' signals are more generally assumed to arrive asynchronously. The asynchronous nature of the interference is modeled by the random relative delay τ_i in (6.3).

As before, the receiver consists of an array of L antenna elements followed by a bank of matched filters, the outputs of which are sampled synchronously at the symbol rate (see Fig. 2.2 in Chapter 2). Using (6.2), the sampled output due to the mth desired user is described by the length-L vector

$$\mathbf{r}_m(k) = \sqrt{2E_{s_m}}\, 10^{x_m/20} r_m^{-\gamma_m/2} \mathbf{g}_m(k)\, c_m(k) \tag{6.4}$$

where the elements of the channel gain vector $\mathbf{g}_m(k)$ are i.i.d. complex Gaussian RVs with variance $\sigma_{g_m}^2$. Using (6.3), the sampled output due to the ith interfering user is

$$\mathbf{r}_i(k) = \sqrt{2E_{s_i}}\, 10^{x_i/20} d_i^{-\gamma_i/2} \mathbf{g}_i(k) \sum_n c_i(n)\, q((k-n)T - \tau_i) \tag{6.5}$$

where the elements of the channel gain vector $\mathbf{g}_i(k)$ are also i.i.d. complex Gaussian RVs, except with variance $\sigma_{g_i}^2$.

In this work, it is assumed that the base station in the central cell controls the transmit power of the desired users in such a way to exactly compensate for shadowing and path loss. Consequently, the average power of each received signal is constant regardless of shadowing and position, that is,

$$\int_{-T/2}^{T/2} \frac{1}{2} E\left[|y_{lm}(t)|^2\right] dt = (2E_{s_m}/T)\, 10^{x_m/10} r_m^{-\gamma_m} \sigma_{g_m}^2 = P_o \tag{6.6}$$

where P_o is the resulting constant received power (the same for all m). Here statistical averaging is performed across the fading ensemble which is followed by time averaging over one symbol period since the received signal is a cyclostationary process.

The base station in each cochannel cell performs the same power control function for users in those cells, except that the path loss is determined by the distance r_i from the interfering user to its *own* base rather than the distance d_i to the base in the central cell. Furthermore, the shadowing is described by a different zero mean Gaussian random variable y_i (independent of x_i) with standard deviation σ_{y_i}. Consequently, the power control law for interfering mobiles is

$$(2E_{s_i}/T)\, 10^{y_i/10} r_i^{-\gamma_i} \sigma_{g_i}^2 = P_o. \tag{6.7}$$

Substituting the power control laws into (6.4) and (6.5) gives

$$\mathbf{r}_m(k) = \frac{\sqrt{P_o T}}{\sigma_{g_m}} \mathbf{g}_m(k)\, c_m(k) \tag{6.8}$$

and

$$\mathbf{r}_i(k) = \frac{\sqrt{P_o T}}{\sigma_{g_i}} 10^{w_i/20} \left(\frac{r_i}{d_i}\right)^{\gamma_i/2} \mathbf{g}_i(k) \sum_n c_i(n)\, q\left((k-n)T - \tau_i\right) \tag{6.9}$$

where the shadowing variable in (6.9) is given by $w_i = x_i - y_i$, which has variance $\sigma_{w_i}^2 = \sigma_{x_i}^2 + \sigma_{y_i}^2$.

The composite signal at the output of the matched filter bank is now

$$\mathbf{r}(k) = \sum_{m=1}^{M} \mathbf{r}_m(k) + \mathbf{z}(k). \tag{6.10}$$

The interference-plus-noise vector $\mathbf{z}(k)$ is given by the sum

$$\mathbf{z}(k) = \sum_{i=1}^{N} \mathbf{r}_i(k) + \mathbf{n}(k). \tag{6.11}$$

where N is the number of interferers. The elements of the noise vector $\mathbf{n}(k)$ are i.i.d. complex Gaussian RVs with variance N_o. It is important to realize that since the noise vector $\mathbf{n}(k)$ and the channel gain vectors $\mathbf{g}_i(k)$ are Gaussian, $\mathbf{z}(k)$ is conditionally Gaussian — that is, Gaussian when conditioned on the random shadowing, location, symbol sequences, and relative delays of the interfering users described by the RVs w_i, r_i, d_i, $\mathbf{c}_i$, and τ_i. Furthermore, due to independent fading across the antenna array, the elements of $\mathbf{z}(k)$ are i.i.d. with conditional variance σ_z^2.

In Chapter 3, the performance of joint detection and MMSE combining is investigated using a signal model similar in form to (6.10), namely a sum of desired Rayleigh faded

cochannel signals plus a Gaussian disturbance vector. In that analysis, a single-cell system is considered meaning that the disturbance vector $\mathbf{z}(k)$ is comprised of noise only, i.e., no interference. The resulting symbol-error rate is thus a function of the per-antenna SNR $\Gamma_m = P_m/\sigma_z^2$, where $P_m = \frac{1}{2}E\left[|r_{lm}(k)|^2\right]$ is the variance of the lth component of $\mathbf{r}_m(k)$ and $\sigma_z^2 = N_o$.

In the multi-cell system considered here, the disturbance vector $\mathbf{z}(k)$ is comprised of noise and interference, but it is still Gaussian when conditioned on w_i, r_i, d_i, $\mathbf{c}_i$, and τ_i. Consequently, the analytical expression for average symbol-error rate (SER) derived in Chapter 3 remains the same in the context of a multi-cell system; however, the SER becomes a *conditional* SER. The conditional SER is still a function of $\Gamma_m = P_m/\sigma_z^2$ except that the variance of the disturbance is now

$$\sigma_z^2 = P_o T \sum_{i=1}^{N} 10^{w_i/10} \left(\frac{r_i}{d_i}\right)^{\gamma_i} \left|\sum_n c_i(n)\, q\left((k-n)T - \tau_i\right)\right|^2 + N_o. \tag{6.12}$$

The ratio Γ_m is now the per-antenna signal-to-interference-plus-noise ratio (SINR), instead of just SNR.

Due to power control, the received power $P_m = P_o T$ is constant and the same for all desired users in the central cell, resulting in the same SINR for all desired users. Using (6.12), the common SINR is thus

$$\Gamma = \frac{P_o T}{\sigma_z^2} = \frac{1}{\sum_{i=1}^{N} 10^{w_i/10} \left(\frac{r_i}{d_i}\right)^{\gamma_i} \left|\sum_n c_i(n)\, q\left((k-n)T - \tau_i\right)\right|^2 + \Gamma_N^{-1}} \tag{6.13}$$

where $\Gamma_N = P_o T/N_o$ is the constant received SNR.

6.3 Outage Probability

Clearly, the SINR — and thus the conditional symbol-error rate $P_s(\Gamma)$ — is a random variable that depends on the RVs w_i, r_i, d_i, $\mathbf{c}_i$, and τ_i. For a given system realization, if the symbol-error rate is above a certain threshold, an outage event is said to have occurred. The probability of this event is denoted $P_{out} = P\left[P_s(\Gamma) > P_{s,t}\right]$ where $P_{s,t}$ is the threshold — typically 10^{-2} or 10^{-3}. System capacity is usually defined in terms of the outage probability P_{out}. Since $P_s(\Gamma)$ decreases monotonically with increasing Γ, outage probability is equivalent

to

$$\begin{aligned} P_{out} &= P\left[\Gamma < \Gamma_t\right] \\ &= P\left[\Gamma_I^{-1} > \Gamma_t^{-1} - \Gamma_N^{-1}\right] \end{aligned} \tag{6.14}$$

where Γ_t is the SINR that corresponds to the threshold $P_{s,t}$. In the latter equality, the quantity Γ_I^{-1} is the interference-to-signal ratio (ISR) given by

$$\Gamma_I^{-1} = \sum_{i=1}^{N} 10^{w_i/10} \left(\frac{r_i}{d_i}\right)^{\gamma_i} \left| \sum_n c_i(n)\, q\left((k-n)\,T - \tau_i\right) \right|^2 . \tag{6.15}$$

The threshold SINR Γ_t depends on the detection technique (e.g., joint detection or MMSE combining), the number of antennas, and the number of users. For joint detection, it may be found by numerically solving the nonlinear equation $P_s(\Gamma_t) = P_{s,t}$ for Γ_t. For MMSE combing, it must be determined through simulation. Table 6.2 shows the threshold SINRs for JD and MMSE found using the results of Chapter 3 for the case of equipower users (due to perfect power control), BPSK modulation, and perfect CSI. The threshold SER is $P_{s,t} = 10^{-3}$. Several entries are missing from the MMSE side since the number of users is limited to the number of antennas. For both JD and MMSE, Γ_t decreases as additional antennas are used, indicating an increasingly higher tolerance to interference. As the number of users increases, Γ_t increases slowly for JD, but very quickly for MMSE. This is due to the fact that L-fold diversity is always maintained with JD, whereas with MMSE, each additional user reduces the order of diversity by one for all users. As will be shown, these differences have profound effects on system capacity.

Table 6.2: Threshold SINRs (in dB) for JD and MMSE combining with M users and L antennas.

	L (JD)			L (MMSE)		
M	2	3	4	2	3	4
1	11.1	6.55	4.03	11.1	6.55	4.03
2	11.6	6.80	4.18	21.1	10.0	6.03
3	12.2	7.08	4.34	–	18.7	9.14
4	12.9	7.38	4.51	–	–	16.6
5	13.6	7.72	4.70	–	–	–
6	14.4	8.08	4.90	–	–	–
7	15.3	8.49	5.12	–	–	–
8	16.2	8.93	5.36	–	–	–

Since Γ_t and Γ_N are constants, one can see from (6.14) that an exact assessment of outage probability requires knowledge of the complementary cumulative distribution function (CCDF) of the ISR Γ_I^{-1}. Unfortunately, Γ_I^{-1} is sum of randomly weighted log-normal RVs for which the distribution is unknown. In many capacity studies, however, the distribution is approximated by a convenient two-parameter distribution, e.g., log-normal [69],[78], and [79] or Gaussian [70]. In this work, the distribution is approximated as log-normal which leads to a closed-form expression for outage probability. As will be shown in Section 6.5, the accuracy of this approximation is very good.

The closed-form expression for outage probability is obtained as follows. First define a log-normal random variable $v = 10^{x/10}$, where x is Gaussian with mean μ_x and standard deviation σ_x. Respectively, the PDF and the CCDF of v are

$$f_V(v) = \frac{1}{\sqrt{2\pi}\sigma v} e^{-\frac{1}{2\sigma^2}\ln^2\left(\frac{v}{\mu}\right)} \tag{6.16}$$

and

$$F_V(v) = \int_v^\infty f_V(t)\,dt = Q\left[\frac{1}{\sigma}\ln\left(\frac{v}{\mu}\right)\right] \tag{6.17}$$

where $\mu = 10^{\mu_x/10}$ and $\sigma = [\ln(10)/10]\,\sigma_x$ [79], and $Q(v)$ is the Gaussian Q-function. Furthermore, the first two moments of v are $E[v] = \mu e^{\sigma^2/2}$ and $E\left[v^2\right] = \mu^2 e^{2\sigma^2}$.

Next approximate Γ_I^{-1} as log-normal, i.e., $\Gamma_I^{-1} \approx v$, which results in

$$P_{out} \approx Q\left[\frac{1}{\sigma}\ln\left(\frac{\Gamma_t^{-1} - \Gamma_N^{-1}}{\mu}\right)\right]. \tag{6.18}$$

The two parameters μ and σ may be obtained by equating the first two moments of Γ_I^{-1} with the first two moments of v and then solving for μ and σ. This is known as Wilkinson's method [78]. The result is

$$\begin{aligned} \mu &= \frac{\left(E\left[\Gamma_I^{-1}\right]\right)^2}{\sqrt{E\left[\Gamma_I^{-2}\right]}} \\ \sigma &= \sqrt{\ln\left(\frac{E\left[\Gamma_I^{-2}\right]}{\left(E\left[\Gamma_I^{-1}\right]\right)^2}\right)}. \end{aligned} \tag{6.19}$$

Clearly, the analytical calculation of P_{out} requires computation of the first two moments of the ISR. In the next section, fully-analytical expressions are derived for the moments.

6.4 Moments of ISR

Observing (6.15), one can see that Γ_I^{-1} is a sum of N independent random variables. Consequently, its first two moments are

$$\begin{aligned} E\left[\Gamma_I^{-1}\right] &= \sum_{i=1}^{N} E\left[z_i\right] \\ E\left[\Gamma_I^{-2}\right] &= \sum_{i=1}^{N} E\left[z_i^2\right] + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} E\left[z_i\right] E\left[z_j\right] \end{aligned} \tag{6.20}$$

where z_i is the ith term in the summation of (6.15). Evidently, the calculation of the nth moment of Γ_I^{-1} requires the calculation of $E\left[z_i^n\right]$ for $n = 1$ and 2 for each interferer.

Due to the fact that the individual random variables in z_i are uncorrelated, $E\left[z_i^n\right]$ factors into three separate expectations as follows

$$E\left[z_i^n\right] = E\left[10^{nw_i/10}\right] E\left[\left(\frac{r_i}{d_i}\right)^{n\gamma_i}\right] E\left[\left|\sum_j c_i\left(j\right) q\left(\left(k-j\right)T - \tau_i\right)\right|^{2n}\right]. \tag{6.21}$$

The expectation over the shadowing ensemble is easily performed using the moments of a log-normal random variable discussed previously. The result is

$$E\left[10^{nw_i/10}\right] = \exp\left[\frac{1}{2}n^2\left(\frac{\ln 10}{10}\sigma_{w_i}\right)^2\right]. \tag{6.22}$$

The two remaining averages are a spatial average and an average over the symbol and delay ensembles. These are calculated in the following two subsections.

6.4.1 Spatial Average

Since spatial averaging is key to capacity studies, the following function called the "position moment" is defined:

$$\Pi_p\left(R_i', \Delta_i, \theta_i\right) \triangleq E\left[\left(\frac{r_i}{d_i}\right)^p\right] \tag{6.23}$$

where $R_i' \triangleq R_i/D_i$. Observing Fig. 6.2, the expectation is calculated by weighting the quantity $(r_i/d_i)^p$ by the probability that a mobile occupies an infinitesimal area $dxdy$ at (x, y) and then integrating the result over the sector defined by the angular width Δ_i and orientation θ_i. Assuming a uniform distribution of location, the probability that the user occupies

the infinitesimal area is simply $dxdy \diagup \left(\frac{1}{2}\Delta_i R_i^2\right)$ where the bracketed quantity is the area of the sector.

In the coordinate system of Fig. 6.2, the distance between the interferer and its own base is $r_i = \sqrt{x^2+y^2}$, and the distance between the interferer and the central base is $d_i = \sqrt{(D_i+x)^2+y^2}$. Consequently, the position moment is given by

$$\Pi_p\left(R_i', \Delta_i, \theta_i\right) = \frac{1}{\frac{1}{2}\Delta_i R_i'^2} \iint_{S_i} \left(\frac{u^2+v^2}{(1+u)^2+v^2}\right)^{p/2} dudv \tag{6.24}$$

where the transformation $u = x/D_i$ and $v = y/D_i$ is employed to make all variables dimensionless quantities. The region of integration is denoted S_i and is defined by the transformed sector boundaries. The transformation highlights the fact the position moment is a function of the ratio $R_i' = R_i/D_i$ and not R_i and D_i individually. The reciprocal of R_i' is interpreted as the normalized reuse distance.

To the author's knowledge, no closed form solution of (6.24) exists. This explains why in many capacity studies, e.g., [69] and [70], spatial averaging is performed by time-consuming Monte Carlo simulation. In this work, a simple, and virtually harmless, geometric approximation is made that allows the derivation of a closed-form analytical solution of the double integral, thereby avoiding long simulation runs. The geometric approximation is based on the observation that for sufficiently small R_i' (distant cochannel cells), $d_i \approx D_i + x$ which results in the following expression for the approximate position moment after transformation to dimensionless quantities:

$$\Pi_p\left(R_i', \Delta_i, \theta_i\right) \approx \frac{1}{\frac{1}{2}\Delta_i R_i'^2} \iint_{S_i} \frac{\left(u^2+v^2\right)^{p/2}}{(1+u)^p} dudv \tag{6.25}$$

To establish the limits of the integral, the sector S_i is expressed as the union of several "sub-sectors" for which the limits are more easily determined. An example is shown in Fig. 6.3(a). Due to symmetries in the integrand of (6.25), a relatively simple generalization of this decomposition process exists for sectors of arbitrary width and orientation. Notice that the integrand is an even function of v; consequently, the integral over any region below the u-axis (e.g., region S_{i3}) may be evaluated over the reflection of that region through the u-axis. Thus, in general, the integral in (6.25) may be written as a sum of integrals over only two basic sub-sectors $S_F\left(R, \theta\right)$ and $S_N\left(R, \theta\right)$ shown in Fig. 6.3(b). For example, the integral over the sector S_i in Fig. 6.3(a) may be written as the sum of the integrals over

$S_F\left(R_i', \pi/2\right)$, $S_N\left(R_i', \theta_i + \Delta_i\right)$, and $S_F\left(R_i', |\theta_i|\right)$. The absolute value of θ_i is used in the latter since, in this case, θ_i is negative.

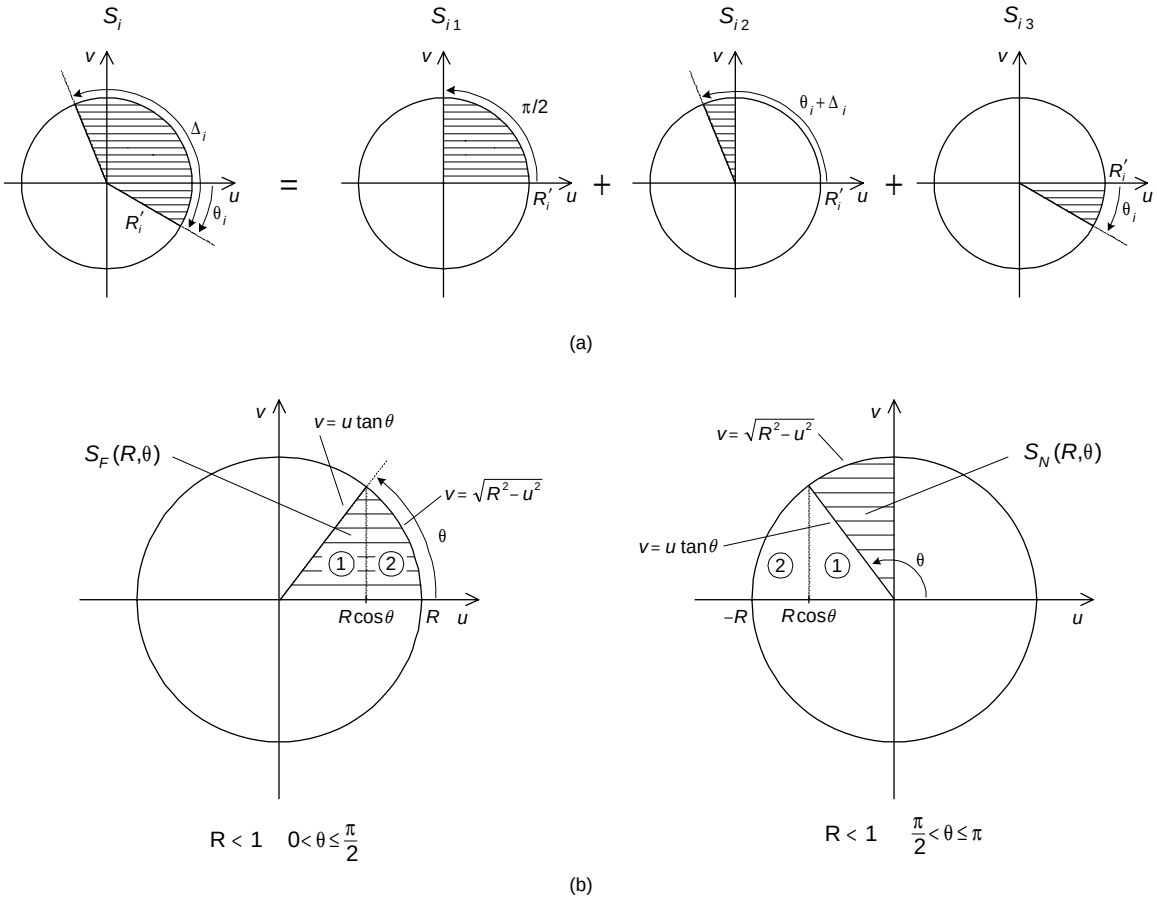

Figure 6.3: (a) Example of the decomposition of a sector into sub-sectors. (b) Definition of the two basic sub-sectors $S_F\left(R, \theta\right)$ and $S_N\left(R, \theta\right)$ used for the computation of the position moment for a sector of arbitary width and orientation.

The integral over $S_F\left(R_i', \theta\right)$ is denoted $F_p\left(R_i', \theta\right)$ and is termed the "far-side integral," since the sub-sector is located on the far-side of the cochannel cell with respect to the central cell. The integral over $S_N\left(R_i', \theta\right)$ is denoted $N_p\left(R_i', \theta\right)$ and is termed the "near-side integral." Using this notation, the position moment for the sector S_i shown in Fig. 6.3(a) is

given by

$$\Pi_p\left(R_i', \Delta_i, \theta_i\right) = \frac{1}{\frac{1}{2}\Delta_i R_i'^2}\left[F_p\left(R_i', \pi/2\right) + N_p\left(R_i', \theta_i + \Delta_i\right) + F_p\left(R_i', |\theta_i|\right)\right]. \tag{6.26}$$

Closed-form analytical solutions for $F_p\left(R_i', \theta\right)$ and $N_p\left(R_i', \theta\right)$ for integer values of p are provided in Appendices A and B, respectively. Note that the restriction to integer values of p implies that the path loss coefficient γ_i takes on integer values. For non-integer values of γ_i, the position moment may be obtained by interpolation of the nearest integer-p values of $\Pi_p\left(R_i', \Delta_i, \theta_i\right)$.

In order to demonstrate the accuracy of the geometric approximation, the exact and approximate position moments are compared in Fig. 6.4, where the exact position moment is determined through numerical integration of (6.24). As expected, the accuracy of the approximation improves as the reuse distance increases. More importantly, for typical reuse distances (e.g., $D_i/R_i > 3$ or 4), the approximation results in very good accuracy.

Fig. 6.5 shows the dependence of the position moment on sector width and orientation. Evidently, the position moment achieves its minimum value (corresponding to minimum interference) when the sector points directly away from the central base station. This occurs for an orientation of $\theta_i = -60^\circ$ for a sector of width $\Delta_i = 120^\circ$, and for an orientation of $\theta_i = -30^\circ$ for a sector of width $\Delta_i = 60^\circ$. As expected, the position moment does not depend on orientation for 360 degree sectors. Furthermore, the position moment becomes less sensitive to sector orientation as the reuse distance increases.

As discussed in Section 6.1, a standard hexagonal cell layout is used to study the uplink capacity with joint detection. Careful observation of the hexagonal layout in Fig. 6.1 reveals that some sectors reside completely on the far-side of the cell, while portions of others reside on both the far-side and near side. Consequently, with this layout, a different combination of the far-side and near-side integrals is required to calculate the position moment $\Pi_p\left(R_i', \Delta_i, \theta_i\right)$ for each cell. For cells 3 and 8 the position moment is

$$\Pi_p\left(R_i', \Delta_i, \theta_i\right) = \frac{1}{\frac{1}{2}\Delta_i R_i'^2}\left[F_p\left(R_i', |\theta_i|\right) + F_p\left(R_i', \frac{\pi}{2}\right) + N_p\left(R_i', \theta_i + \Delta_i\right)\right]. \tag{6.27}$$

For cells 1–2, 4–6, and 9–12 the position moment is

$$\Pi_p\left(R_i', \Delta_i, \theta_i\right) = \frac{1}{\frac{1}{2}\Delta_i R_i'^2}\left[F_p\left(R_i', |\theta_i|\right) + F_p\left(R_i', \theta_i + \Delta_i\right)\right], \tag{6.28}$$

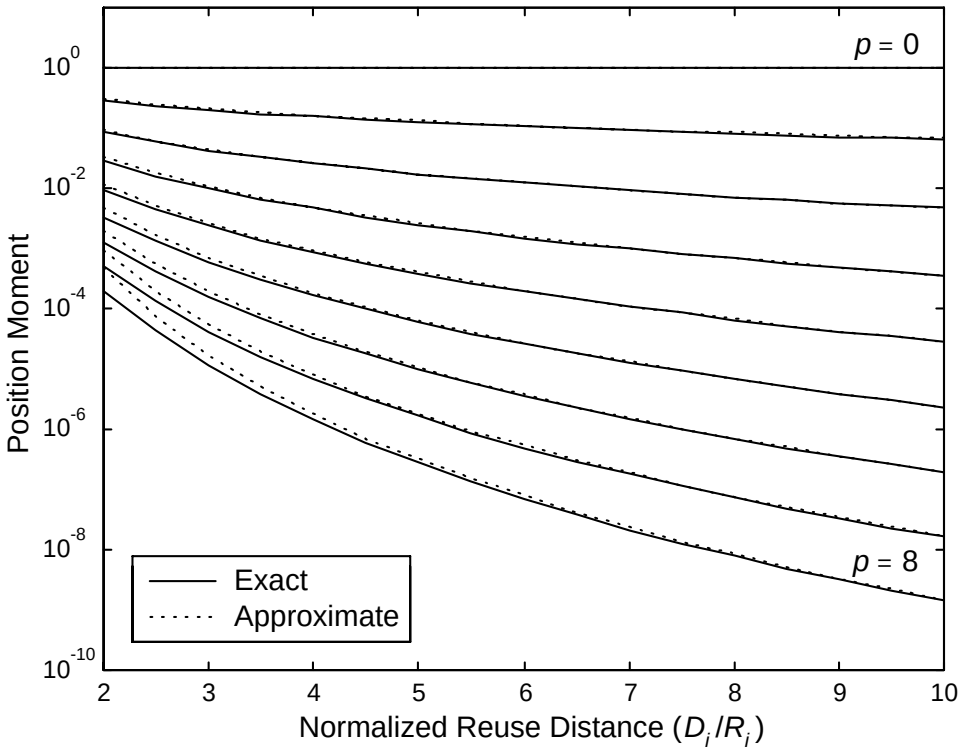

Figure 6.4: Comparison of exact and approximate position moments for a sector of width $\Delta_i = 120^\circ$ and orientation $\theta_i = -10^\circ$.

and for cells 7 and 13 it is

$$\Pi_p\left(R_i', \Delta_i, \theta_i\right) = \frac{1}{\frac{1}{2}\Delta_i R_i'^2}\left[N_p\left(R_i', |\theta_i|\right) + F_p\left(R_i', \frac{\pi}{2}\right) + F_p\left(R_i', \theta_i + \Delta_i\right)\right]. \tag{6.29}$$

In all cases, $\Delta_i = 120^\circ$ and $R_i' = R_i/D_i$ where $R_i = \sqrt{3\sqrt{3}/2\pi}$. The parameters D_i and θ_i for each cell are listed in Table 6.1.

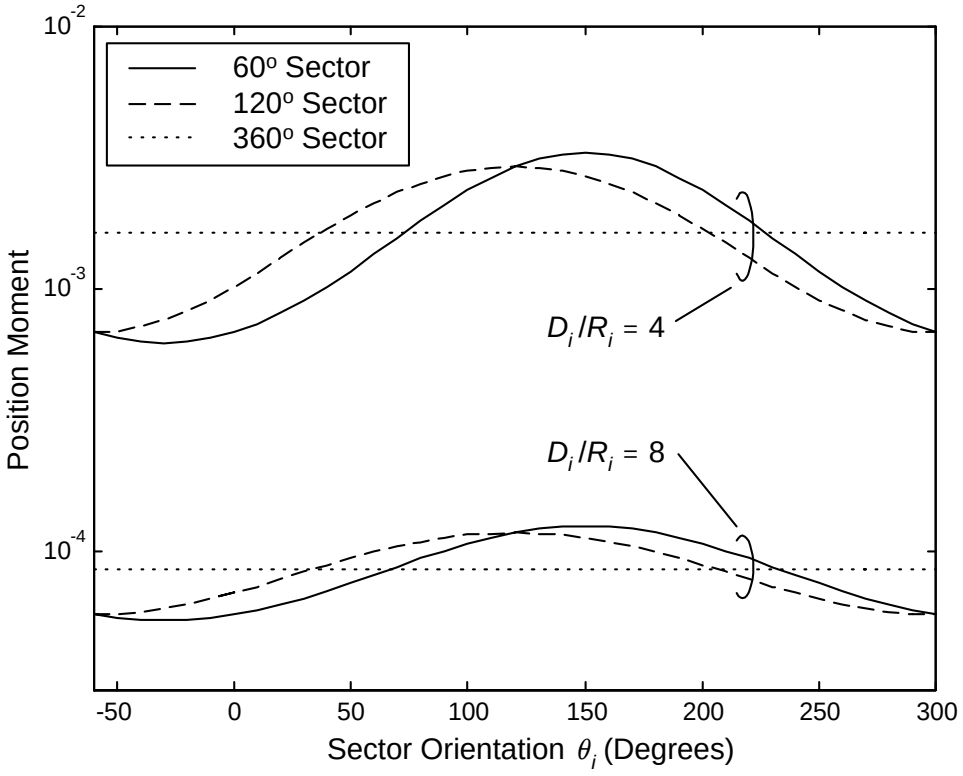

Figure 6.5: Dependence of position moment on sector width and orientation.

6.4.2 Data/Delay Average

The remaining average to be calculated is the average over the data and delay ensembles, namely

$$\begin{aligned} \Upsilon_i(n) &= E\left[\left|\sum_j c_i(j)\, q\left((k-j)\,T-\tau_i\right)\right|^{2n}\right] \\ &= E\left[\left|\sum_k c_i(k)\, q\left(kT+\tau_i\right)\right|^{2n}\right]. \end{aligned} \tag{6.30}$$

The latter equality follows from the former since a uniform distribution of delay τ_i is assumed which makes $\Upsilon_i(n)$ independent of time. The index k in the latter is understood to span $(-\infty,\infty)$. Recall that the data symbol $c_i(k)$ is drawn from a unit-radius PSK constellation with $Q=2^{n_b}$ points where n_b is the number of bits per symbol.

For $n = 1$, (6.30) is expanded as a double summation noting that $c_i(k)$ and τ_i are uncorrelated:

$$\Upsilon_i(1) = \sum_k \sum_l E\left[c_i(k)\, c_i^*(l)\right] E\left[q(kT+\tau_i)\, q(lT+\tau_i)\right]. \tag{6.31}$$

Since the data sequence $\mathbf{c}_i$ is assumed to be white, the only nonzero terms occur when $k = l$. Thus, assuming the PDF of τ_i is uniform over $\left[-\frac{T}{2}, \frac{T}{2}\right]$, $\Upsilon_i(1)$ is given by

$$\begin{aligned}\Upsilon_i(1) &= \frac{1}{T}\sum_k \int_{-\frac{T}{2}}^{\frac{T}{2}} q^2(kT+\tau)\, d\tau \\ &= \frac{1}{T}\int_{-\infty}^{\infty} q^2(\tau)\, d\tau \end{aligned} \tag{6.32}$$

The integral in the latter equality may be evaluated analytically in the frequency domain using Parseval's theorem. The result is a function of the pulse rolloff parameter β.

For $n = 2$, (6.30) is expanded as a quadruple summation as follows

$$\begin{aligned}\Upsilon_i(2) &= \sum_k \sum_l \sum_m \sum_n E\left[c_i(k)\, c_i^*(l)\, c_i(m)\, c_i^*(n)\right] \\ &\quad \cdot E\left[q(kT+\tau_i)\, q(lT+\tau_i)\, q(mT+\tau_i)\, q(nT+\tau_i)\right]. \end{aligned} \tag{6.33}$$

For BPSK ($n_b = 1$) where the data symbols are real, the only nonzero terms in the summation occur for the following four mutually exclusive cases: (a) $k = l = m = n$, (b) $k = l$ and $m = n$ and $m \neq k$, (c) $k = m$ and $l = n$ and $l \neq k$, and (d) $k = n$ and $l = m$ and $l \neq k$. For PSK with $n_b \geq 2$ where the data symbols are complex, the terms corresponding to case (c) are zero since $E\left[c_i^2(k)\right] = E\left[c_i^{*2}(k)\right] = 0$. Accounting for the four cases, $\Upsilon_i(2)$ may be reduced to the following

$$\Upsilon_i(2) = \begin{cases} \frac{1}{T}\int_{-\infty}^{\infty} q^4(\tau)\, d\tau + \frac{3}{T}\sum_k \sum_{\substack{m \\ m\neq k}} \int_{-\frac{T}{2}}^{\frac{T}{2}} q^2(kT+\tau)\, q^2(mT+\tau)\, d\tau, & n_b = 1 \\ \frac{1}{T}\int_{-\infty}^{\infty} q^4(\tau)\, d\tau + \frac{2}{T}\sum_k \sum_{\substack{m \\ m\neq k}} \int_{-\frac{T}{2}}^{\frac{T}{2}} q^2(kT+\tau)\, q^2(mT+\tau)\, d\tau, & n_b \geq 2 \end{cases}. \tag{6.34}$$

Again, the PDF of τ_i is assumed to be uniform over $\left[-\frac{T}{2}, \frac{T}{2}\right]$. The first term in each expression accounts for case (a), and the second term collectively accounts for cases (b), (c), and (d). Notice that for PSK with $n_b \geq 2$, the factor multiplying the double summation is 2 rather than 3, since the terms corresponding to case (c) are zero.

The integrals in (6.34) are most easily calculated numerically. Resorting to numerical integration is not unattractive since $\Upsilon_i(n)$ may be precomputed and stored for later use.

Moreover, since all interferers use the same modulation format, $\Upsilon_i(n)$ need only be computed once and the same value used for all i. Values of the data/delay averages for two different pulse rolloff values are listed in Table 6.3.

Table 6.3: Data/Delay average $\Upsilon_i(n)$ for PSK modulation with n_b bits per symbol.

	$\beta = 0.35$		$\beta = 0.5$	
n	$n_b = 1$	$n_b \geq 2$	$n_b = 1$	$n_b \geq 2$
1	0.9125	0.9125	0.8750	0.8750
2	1.2170	1.0267	1.0657	0.9195

6.4.3 Summary of Technique

In summary, outage probability is calculated through the moments of ISR as follows:

1. For $n = 1, 2$ and $i = 1 \ldots N$ calculate and store $E[z_i^n]$ using (6.21). This requires

   (a) calculation of the shadowing average using (6.22)

   (b) calculation of the position moment $\Pi_p(R_i', \Delta_i, \theta_i)$ with $p = n\gamma_i$. This requires expressing the sector S_i as the union of the two basic subsectors $S_F(R, \theta)$ and $S_N(R, \theta)$ as discussed in Section 6.4.1. For the standard hexagonal layout in Fig. 6.1, the appropriate union of subsectors for each cell is given at the end of Section 6.4.1. Closed-form analytical expressions for the integral over $S_F(R, \theta)$ (the far-side integral) and over $S_N(R, \theta)$ (the near-side integral) are provided in Appendix A and B, respectively.

   (c) calculation of the data/delay average $\Upsilon_i(n)$. Precalculated values for two different pulse rolloff values are listed in Table 6.3.

2. Calculate the first two moments of the ISR using (6.20) with the stored values of $E[z_i^n]$.

3. Calculate and store μ and σ using (6.19).

4. Calculate outage probability using (6.18) with the stored values of μ and σ.

6.5 Results

In all of the results presented in this section, the standard hexagonal layout with cluster size C shown in Fig. 6.1 is used. For all interferers, the path loss coefficient is $\gamma_i = 4$, the shadowing standard deviation is $\sigma_{w_i} = \sqrt{2} \cdot 6$ dB, the modulation is BPSK, and the pulse rolloff is $\beta = 0.5$. Since perfect power control is assumed, the results do not depend on the shadowing and path loss of the desired users.

Before presenting capacity results, the accuracy of the various approximations made in the preceding analysis is investigated. Recall that the ISR Γ_I^{-1} is approximated as log-normal. Consequently, outage probability is approximated as

$$\begin{aligned} P_{out} &= P\left[\Gamma_I^{-1} > \Gamma_{I,t}^{-1}\right] \\ &\approx F_V\left(\Gamma_{I,t}^{-1}\right) \end{aligned} \tag{6.35}$$

where

$$\Gamma_{I,t}^{-1} = \Gamma_t^{-1} - \Gamma_N^{-1} \tag{6.36}$$

is the threshold ISR, and $F_V(v)$ is the log-normal CCDF given in (6.17). The geometric approximation and the approximation implicit in Wilkinson's moment matching method are embodied in the analytical calculation of $F_V(v)$ through the parameters μ and σ in (6.19).

To illustrate the combined accuracy of the above approximations, Fig. 6.6 compares the log-normal CCDF $F_V\left(\Gamma_{I,t}^{-1}\right)$ to the empirical CCDF $P\left[\Gamma_I^{-1} > \Gamma_{I,t}^{-1}\right]$ obtained by Monte Carlo simulation with 5×10^6 iterations. These results are for a single-user system ($M = 1$); multiuser results appear below. The accuracy in outage probability is reflected in the vertical separation of the corresponding analytical and simulation curves. Evidently, the log-normal CCDF is a very good approximation of the empirical CCDF for outage probabilities less than about 10%. The same accuracy was observed for several different combinations of the path loss coefficient in the range 3–4 and the shadowing standard deviation in the range 6–8 dB. This result indicates that even though the ISR Γ_I^{-1} is a sum of randomly weighted log-normal random variables, rather than a sum of pure log-normal RVs as studied in [78], it is still closely log-normal.

The simulation results in Fig. 6.6 are obtained as follows: at each iteration, for each interferer, generate the shadowing random variable w_i, the mobile location random variables r_i and d_i, and the data and delay random variables $\mathbf{c}_i$ and τ_i, where $\mathbf{c}_i$ is a finite length,

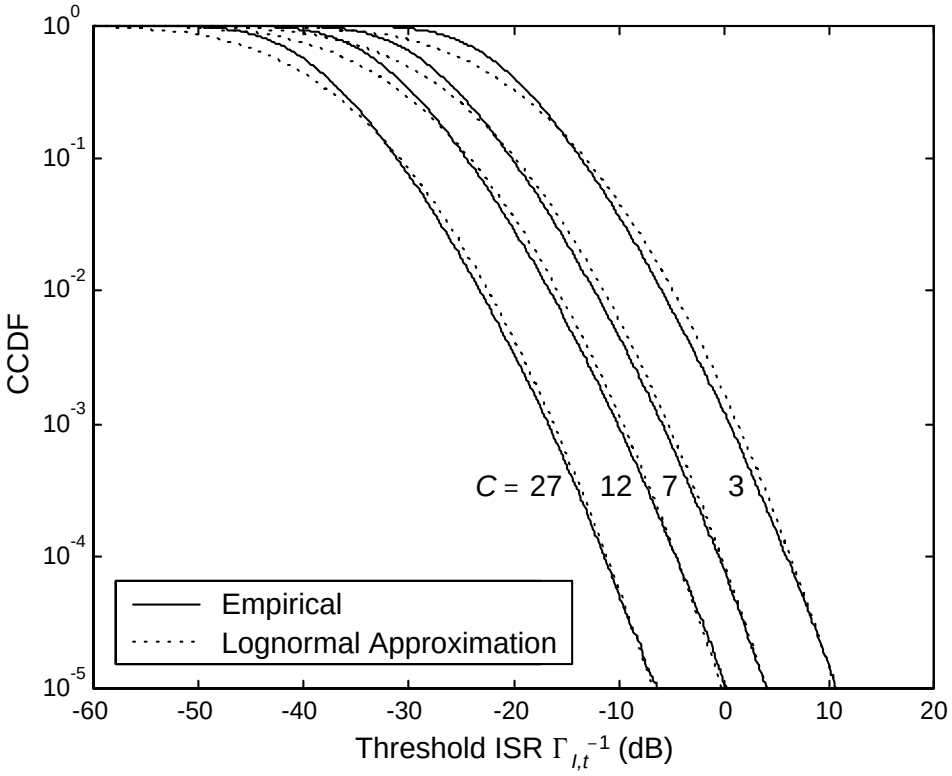

Figure 6.6: Outage probability vs. threshold ISR for the standard hexagonal layout with cluster size C and a single user per slot ($M = 1$).

random sequence of BPSK symbols; calculate the ISR Γ_I^{-1} using (6.15) for time $k = 0$; and store the computed ISR value in a vector. After a large number of iterations, generate the empirical CCDF from the vector of ISR points.

The generation of the RVs w_i, $\mathbf{c}_i$, and τ_i is straightforward; however, the generation of r_i and d_i requires some explanation. Recall that the interferer is located in a sector of radius R_i, width Δ_i, and orientation θ_i, and that the interferer's own base station is located at a distance of D_i from the base station in the central cell (see Fig. 6.2). Using these parameters, the RVs r_i and d_i are calculated by first generating the independent RVs u_i and ϕ_i, where u_i is uniform on $[0, 1]$, and ϕ_i is uniform on $[\theta_i, \theta_i + \Delta_i]$. Assuming a uniform distribution of location within the sector, the distance r_i is then given by $r_i = R_i\sqrt{u_i}$. This shows that users are more likely to be located near the cell edge than near the cell centre.

Using Pythagoras' theorem along with the computed distance r_i, the distance d_i is given by $d_i = \sqrt{D_i^2 + r_i^2 + 2D_i r_i \cos\phi_i}$.

Fig. 6.7 shows the log-normal CCDF $F_V\left(\Gamma_{I,t}^{-1}\right)$ for multiuser systems ($M > 1$). Although simulation results are not shown on this graph, the same accuracy as above was observed. Clearly, for a fixed threshold ISR, outage probability increases as the number of users increases and/or the cluster size decreases.

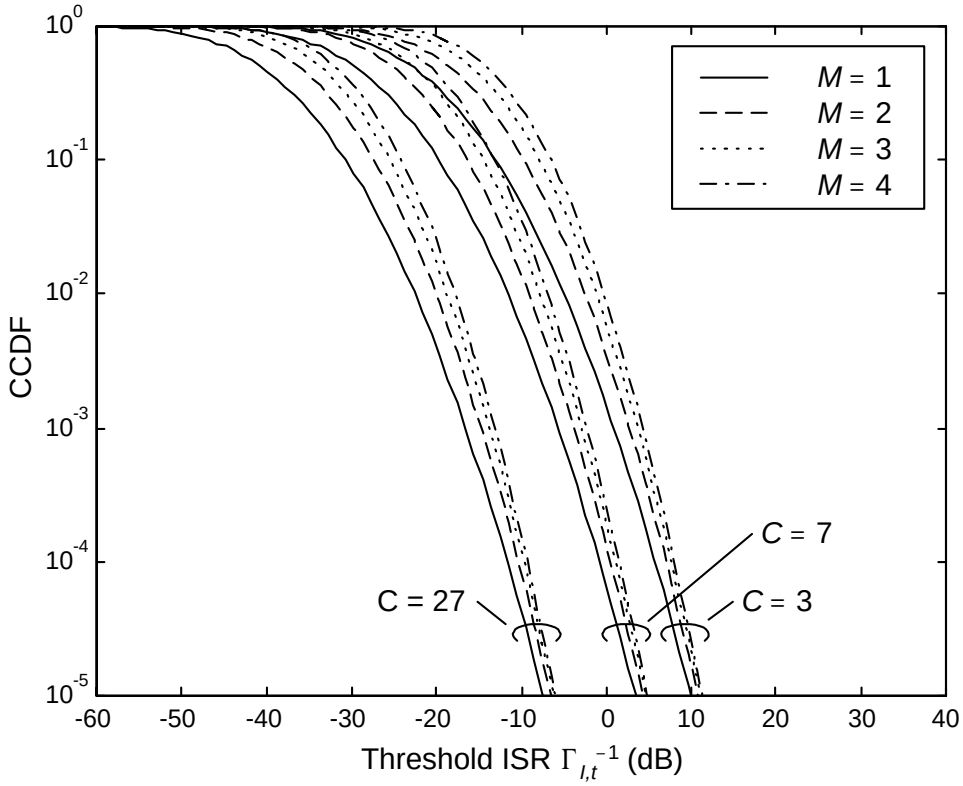

Figure 6.7: Outage probability vs. threshold ISR for the standard hexagonal layout with cluster size C and multiple users per slot.

Recall from Chapter 2, that system capacity is defined as $\eta = M \cdot (N_c/C)$ where N_c is the total number of channels available to the entire system and C is the cluster size. The bracketed quantity is the number of channels available to each cell. Since capacity scales directly with the total number of channels, N_c is set to unity, resulting in $\eta = M/C$. The cluster size C is chosen to be as small as possible with the constraint that outage probability

must be less than or equal to 1%. The resulting minimum cluster size is denoted $C_{min}(M, L)$ to emphasize the fact that it is a function of the number of users per sector and the number of antennas (as well as the detection technique). The maximum system capacity is thus

$$\eta_{max} = \frac{M}{C_{min}(M, L)}. \tag{6.37}$$

Typically, as M increases, $C_{min}(M, L)$ must increase to mitigate the effect of additional interference. Consequently, depending on the number of antennas, η_{max} may not increase linearly with M. Note that with this capacity measure, if M and L are such that $\eta_{max} = 1$, an effective reuse factor of unity is obtained, i.e., the number of users per cell equals the cluster size C.

For a given number of users and antennas, the minimum cluster size $C_{min}(M, L)$ is found as follows. First, the nonlinear equation

$$P_{out} = Q\left[\frac{1}{\sigma}\ln\left(\frac{\Gamma_{I,t}^{-1}}{\mu}\right)\right] = 0.01 \tag{6.38}$$

is solved for the threshold ISR $\Gamma_{I,t}^{-1}$. Recall that the parameters μ and σ depend on M and C; thus, this is done for each allowed cluster size $C \in \{1, 3, 4, 7, 9, 12, \cdots\}$. Graphically, this corresponds to drawing a horizontal line on the graph in Fig. 6.7 at 10^{-2} then finding the point on the abscissa that corresponds to the intersection of this line with the appropriate curve. Next, the appropriate threshold SINR Γ_t is found from Table 6.2 and is used along with $\Gamma_{I,t}^{-1}$ in (6.36) to solve for the SNR parameter. The result is

$$\Gamma_{N,r} = \frac{1}{\Gamma_t^{-1} - \Gamma_{I,t}^{-1}}. \tag{6.39}$$

The subscript 'r' is used to signify that this is the minimum *required* SNR to achieve 1% outage probability. For cluster sizes too small to guarantee 1% outage, $\Gamma_{N,r}$ turns out negative — clearly an invalid result. $C_{min}(M, L)$ is simply the smallest C for which $\Gamma_{N,r}$ is positive.

Table 6.4 shows the minimum cluster sizes for JD and MMSE combining. In both cases $C_{min}(M, L)$ increases with M in order to account for the additional interference. For JD, however, the increase is much slower than for MMSE since JD maintains diversity order L regardless of the number of users, whereas MMSE loses one order of diversity with each additional user. This has profound effects on capacity as will now be shown.

Table 6.4: Minimum cluster sizes $C_{min}(M, L)$ for JD and MMSE combining.

	L (JD)			L (MMSE)		
M	2	3	4	2	3	4
1	7	4	3	7	4	3
2	12	7	4	31	9	7
3	13	7	7	-	28	9
4	16	9	7	-	-	25
5	19	9	7	-	-	-
6	21	12	7	-	-	-
7	25	12	7	-	-	-
8	31	12	9	-	-	-

Fig. 6.8 plots the maximum capacity η_{max} in (6.37) vs. the number of user/sector. With joint detection, as more intracell cochannel users are allowed, capacity continues to increase towards a soft limit with a rate governed by the number of antennas. With four antennas and seven users, $\eta_{max} = 1$, meaning that an effective reuse factor of unity is achieved — a striking fact usually associated only with CDMA systems. Note that capacity does not always increase monotonically with M, since the allowed cluster sizes are irregularly spaced.

In stark contrast, MMSE combining exhibits a hard capacity limit: an optimum number of users exists, above which capacity actually decreases. For $L = 2$ and 3, the optimum number of users is one. For $L = 4$ the optimum is three, but the capacity is no higher than with a single user. Thus, with four or fewer antennas, a performance penalty is incurred by allowing more than one cochannel user per sector. This result is surprising and significant, since MMSE combining is frequently advanced as a method to increase capacity.

It is interesting to compare the capacity of a system using either JD or MMSE combining to that of a standard TDMA system with a single user per sector and maximal ratio combining of two antennas (the $M = 1$, $L = 2$ point in Fig. 6.8). Making this comparison reveals that with the joint detection of $M \leq 7$ users, at least an M-fold increase in capacity may be obtained with four antennas. The best that can be hoped for with MMSE combining with four antennas is just over a two-fold increase.

Fig. 6.9 shows the minimum SNR $\Gamma_{N,r}$ in (6.39) required to achieve an outage probability of 1%. Again, the behaviour is slightly erratic due to the irregularly spaced cluster sizes, but several observations can be made. The required SNR for both JD and MMSE exhibits an increasing trend, albeit non-monotonic, as the number of users increases; however, the increase is much slower with JD. Moreover, for virtually all antenna/user combinations, the

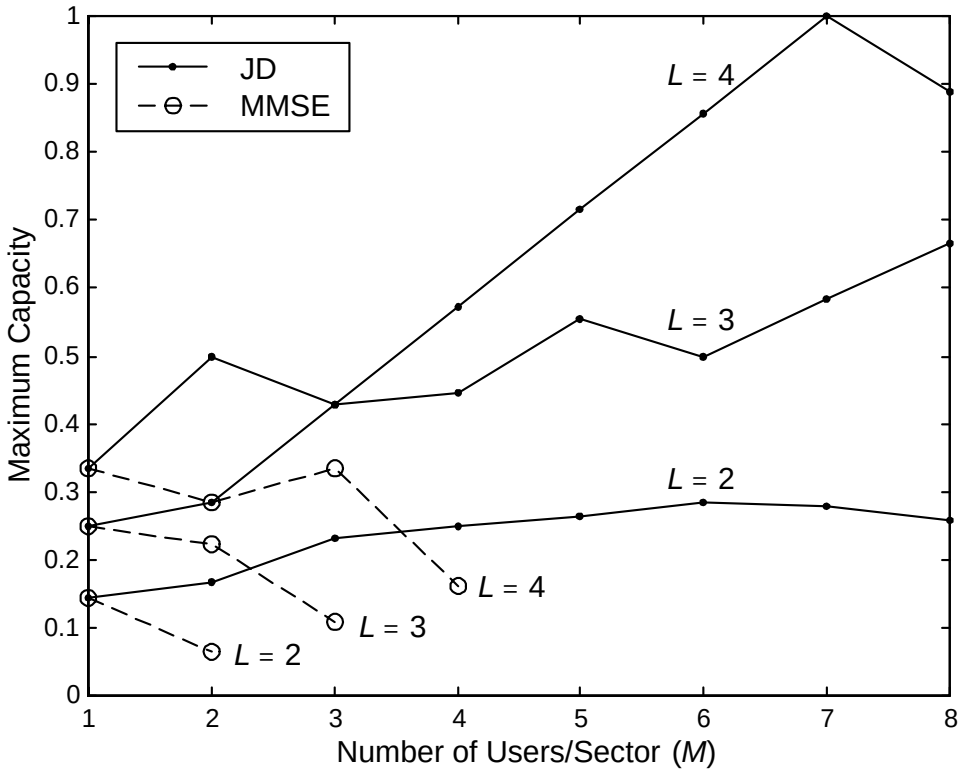

Figure 6.8: Maximum capacity η_{max} of joint detection and MMSE combining systems.

required SNR for JD is significantly less than for MMSE. For example, for $M = 3$ and $L = 4$ where JD achieves a three-fold increase in capacity and MMSE just over a two-fold increase, JD requires a striking 10 dB less SNR than MMSE.

The consequences of the above SNR behaviour are that one may use a cheaper base-station amplifier with JD than with MMSE and/or require a lower mobile transmit power, thus reducing battery drain. Alternatively, with JD one may utilize larger cell sizes than with MMSE implying reduced infrastructure costs for a fixed coverage area. The effect on cell size may be seen by considering the power control law introduced in (6.6), which assumes that the mobile units have infinite dynamic range. Taking into account the mobiles' limited transmit power places an upper limit on the cell radius R_m, so that power control may effectively compensate for path loss and shadowing at all locations within the cell. A lower required SNR implies a lower required received power P_o which enables the use of

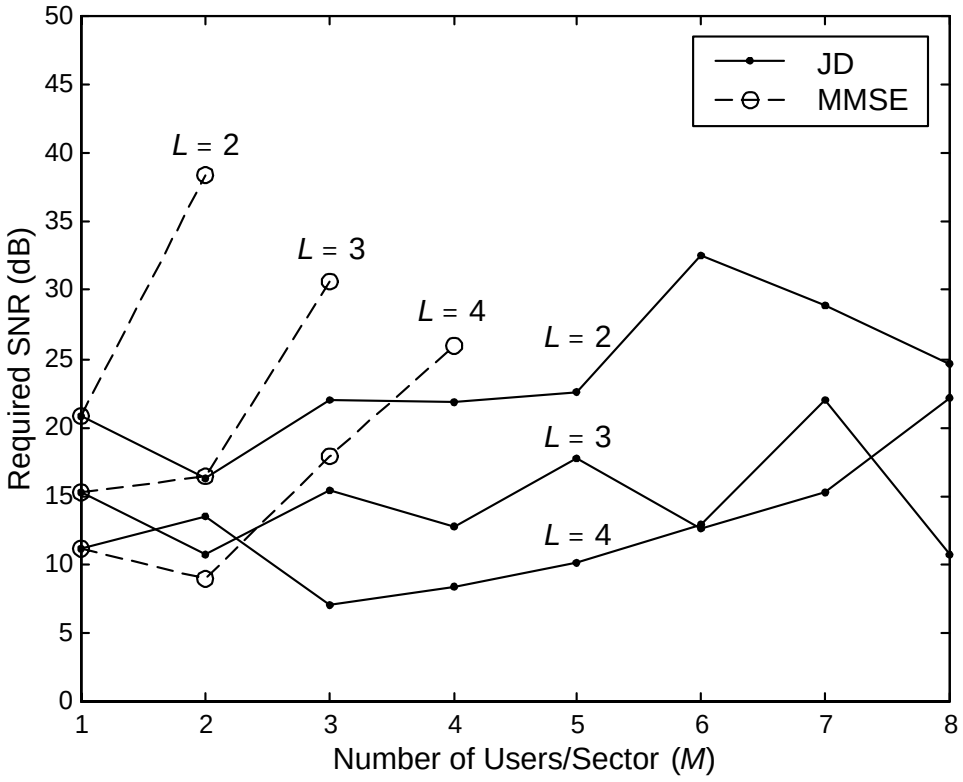

Figure 6.9: Required SNR $\Gamma_{N,r}$ to achieve 1% outage probability for joint detection and MMSE combining.

larger cells. A disadvantage of larger cells, however, is a system with increased blocking probability.

6.6 Conclusions

In this chapter, the uplink capacity of a TDMA cellular system employing joint detection with a diversity antenna array at the base station receiver has been analyzed in the presence of fading, shadowing, path loss, and cochannel interference. For reference, the analysis is compared to MMSE combining, a common technique for configuring diversity receivers. Both techniques allow several intracell users to share the same time/frequency slot in order to increase system capacity.

With JD, a soft limit on capacity is demonstrated, but before that limit is reached, capacity may be significantly extended. With four antennas and $M \leq 7$ cochannel users per sector, an M-fold increase in capacity may be obtained when compared to a conventional single-user TDMA system employing maximal ratio combining and two antennas. Furthermore, with JD and four or more antennas, it is possible to obtain an effective reuse factor of unity — a configuration usually associated only with CDMA systems.

In stark contrast, with MMSE combining only a modest increase in capacity may be obtained, and a hard limit on the number of user per sector exists beyond which system capacity actually decreases due to losses in diversity order. With four antennas, just over a two-fold increase in capacity may be obtained when compared to a single-user TDMA system with two antennas.

For both JD and MMSE combining, the required receive SNR to achieve a fixed outage rate increases with the number of users per sector; however, with JD the increase is much slower, and the required SNR significantly less, than for MMSE. Consequently, for a given capacity improvement, JD allows for less expensive base station amplifiers, lower mobile battery drain, and lower infrastructure costs due to increased cell coverage.

Also in this chapter, a new fully-analytical technique has been presented for the computation of outage probability used for the assessment of uplink capacity. The technique may be used for very general cellular layouts, since the cochannel cells are considered to be circular, rather than hexagonal, and may be of arbitrary size and location. Furthermore, the cells may contain sectors of arbitrary width and orientation.

The core of the technique is the analytical calculation of the moments of the interference-to-signal ratio, and the key contribution is a closed-form geometric approximation of the spatial average of the mobiles' position-dependent path loss. The moments of ISR are used with Wilkinson's method to approximate the complementary cumulative distribution function of the ISR, which is then used to compute outage probability directly. The closed-form expression for outage probability may be used to calculate system capacity much faster than with traditional Monte Carlo simulation. Results comparing analysis and simulation show that the analytical technique achieves very good accuracy.

Chapter 7

Conclusions

This thesis has focused on extending the capacity of TDMA cellular systems by allowing several users in the same cell to share the same time/frequency slot and employing joint detection with diversity arrays to distinguish the cochannel signals. In conclusion, the novel contributions and significant findings of the thesis are summarized below.

1. Performance of joint detection with diversity arrays in the uplink:

    - A fully-analytical expression for the union bound on average symbol-error rate is provided for arbitrary number of users M and antennas L in a flat fading environment with symbol-synchronous users. Both perfect and imperfect channel state information (CSI) are considered.
    - It is demonstrated that with joint detection, many more users than the number of antennas may be supported — all enjoying L-fold diversity — with a small degradation in performance with each additional user. This fundamental result, not observed previously, is extremely valuable in a fading environment and leads to large system capacity gains. With accurate CSI, it is demonstrated that even with only a single antenna several users can experience good performance. This new mode of reception behaviour starkly contrasts that of classical MMSE combining of antennas where the maximum number of users is limited to the number of antennas, and the diversity order is only $L - M + 1$ for each user.
    - It is demonstrated that the cost of supporting additional users at a fixed error rate is a small increase in the required per-user SNR. The SNR penalty relative to a

single-user system is quantified as a function of the number of users, constellation density, and number of antennas. For example, with BPSK modulation, a single antenna, and perfect CSI, only a 2 dB penalty is incurred for each additional user. With four antennas, this penalty is reduced to less than 0.1 dB.

- With imperfect channel state information from a pilot-based multiuser channel estimator, it is found that no error floor appears. The cost of imperfect CSI is simply an increase in the required SNR in order to maintain a fixed error-rate. Further, it is found that an interdependence exists between all users participating in the joint detection: the channel estimation error of one user affects the performance of all users, even those with perfect CSI.
- Unequal SNR distributions are investigated, and it is demonstrated that both the weak and strong users' performance is degraded from the equipower case. This indicates that power control may be desirable in a practical system.

2. Use of joint detection and multiple transmit antennas in the downlink:

   - A simple structure is proposed for supporting multiple users in the downlink through the use of a transmit antenna array. The users' bit sequences are multiplexed together, and the resulting high-rate bit sequence is coded and interleaved to provide temporal diversity to the mobiles. The bandwidth required to transmit the high-rate coded sequence is reduced by simultaneously transmitting several successive code symbols from different antennas. At the mobile receivers, soft decision decoding is employed using a metric that implicitly performs joint detection of the cochannel signals transmitted from the multiple antennas.
   - It is demonstrated that the simultaneous transmission of successive code symbols causes a single channel usage to span several trellis transitions. To handle this unusual situation in both analysis and decoder implementation, the concept of a merged trellis is introduced in which the trellis is modified such that a single channel usage spans only one transition. Using the merged trellis, the optimal decoder is identified, and a fully-analytical expression for the average bit-error rate is derived. Unusual behaviour is demonstrated in terms of diversity order: as the number of antennas increases due to an increasing number of users, the diversity order actually decreases due to the simultaneous transmission of successive code

symbols.

- Even with the loss in diversity, it is demonstrated that good performance may be obtained for all users at reasonably low SNRs, provided that a code with sufficient constraint length is used. For example, using a rate one half, constraint length seven code with BPSK modulation and four antennas, four users may be supported in the same bandwidth as a single-user at a bit-error rate of 10^{-3} with only 2.5 dB/user more transmit power than a single-user system.

3. Multiuser channel estimation for joint detection:

   - A pilot-based MMSE technique is developed for jointly estimating the channel impulse responses of multiple cochannel signals. The technique is derived for the general case of frequency selective fading and asynchronous users; however, it includes the situation of flat-fading and symbol synchronous users considered in other parts of the thesis as a special case. A novel feature of the estimator is that it allows for time variation of the channels within the training sequences, which is essential in a multiuser environment even at moderate fading rates.
   - The design of good training sequences for the multiple cochannel users is addressed. It is found that the per-user data throughput decreases with each additional user, since the minimum length of the training sequences grows linearly with the number of users. However, system efficiency increases due to multiple users sharing the same time/frequency slot.

4. System capacity using joint detection:

   - A fully-analytical and computationally straightforward technique is developed for the computation of outage probability with joint detection in the presence of fading, shadowing, and path loss. In contrast to previous studies, the spatial average of the mobiles' position-dependent path loss is calculated analytically, rather than by Monte Carlo simulation, through a simple, yet accurate, geometric approximation. The technique does not rely upon a hexagonal cell layout, thus enabling the study of more general systems in which the cochannel cells are of arbitrary size and location and have sectors of arbitrary width and orientation.

- System capacity is quantified through the analytical computation of outage probability, while accounting for the increased interference from cochannel cells due to reuse within cell.
- A soft limit on capacity is demonstrated, similar to CDMA, but before that limit is reached, capacity may be significantly extended. With four antennas and $M \leq 7$ cochannel users, an M-fold increase in capacity may be obtained when compared to a conventional single-user TDMA system employing maximal ratio combining and two antennas. Furthermore, with joint detection and four or more antennas is possible to obtain an effective reuse factor of unity — a configuration usually associated only with CDMA systems.
- In stark contrast to joint detection, it is demonstrated that MMSE combining produces only a modest increase in capacity, as well as a hard limit on the number of users per cell, beyond which system capacity actually decreases due to losses in diversity order. This conclusion is surprising and significant, since MMSE combining is frequently advanced as a method to increase capacity.

In view of the above, the author has confidence that joint detection with diversity arrays is an attractive technique for significantly extending TDMA system capacity.

Appendix A

Solution of Far-side Integral

In this appendix, a closed-form analytical solution is provided for the far-side integral

$$F_p(R,\theta) = \int\int_{S_F(R,\theta)} \frac{\left(u^2+v^2\right)^{p/2}}{(1+u)^p} du dv \tag{A.1}$$

where $R < 1$ and p is an integer. The region $S_F(R,\theta)$ is shown in Fig. 6.3(b) which consists of two sub-regions, labeled 1 and 2, each with different upper limits on v. Consequently, $F_p(R,\theta)$ is given by the sum

$$F_p(R,\theta) = F_p^{(1)}(R,\theta) + F_p^{(2)}(R,\theta) \tag{A.2}$$

where

$$F_p^{(1)}(R,\theta) = \int_0^{R\cos\theta} \frac{1}{(1+u)^p} \left(\int_0^{u\tan\theta} \left(u^2+v^2\right)^{p/2} dv \right) du \tag{A.3}$$

and

$$F_p^{(2)}(R,\theta) = \int_{R\cos\theta}^{R} \frac{1}{(1+u)^p} \left(\int_0^{\sqrt{R^2-u^2}} \left(u^2+v^2\right)^{p/2} dv \right) du \tag{A.4}$$

correspond to sub-regions 1 and 2 respectively.

For the case of p even, both (A.3) and (A.4) may be solved analytically. For the case of p odd, on the other hand, the solution of the inner integrals involve complicated logarithmic functions (see [80, eq. (2.260)]) leading to analytically intractable single integrals in u. Fortunately, numerical evaluation of (A.1) reveals that the far-side integral for odd

p is approximated very accurately by the geometric mean of the nearest even-p values of $F_p(R,\theta)$, that is,

$$F_p(R,\theta) \approx \sqrt{F_{p-1}(R,\theta)\, F_{p+1}(R,\theta)}, \quad p = 1,3,5,\ldots \tag{A.5}$$

Thus, (A.3) and (A.4) need only be solved for the case of p even. Fig. 6.4 confirms the accuracy of (A.5).

When the closed-form expression for $F_p^{(1)}(R,\theta)$ (derived below) is calculated, numerical instabilities occur for θ near $\pi/2$, even when using double precision arithmetic. This is due to the $\tan\theta$ term in the upper limit of the v integral of (A.3) which appears directly in the closed-form expression. As θ approaches $\pi/2$, $\tan\theta$ becomes very large causing $F_p^{(1)}(R,\theta)$ to become unstable. However, the instability may be avoided by approximating the far-side integral as $F_p(R,\theta) \approx F_p^{(2)}(R,\theta)$ for θ near $\pi/2$. This approximation is reasonable since, for θ near $\pi/2$, region 1 is much smaller than region 2 resulting in $F_p^{(2)}(R,\theta)$ being several orders of magnitude larger than $F_p^{(1)}(R,\theta)$.

A.1 Solution of $F_p^{(1)}(R,\theta)$

Denote the inner integral of (A.3) as $y(u)$. This integral is solved by first expanding the term $\left(u^2+v^2\right)^{p/2}$ in a binomial series and then integrating with respect to v. The result is

$$y(u) = \left[\sum_{k=0}^{p/2} \frac{\binom{p/2}{k}}{2k+1} (\tan\theta)^{2k+1}\right] u^{p+1}. \tag{A.6}$$

Substituting $y(u)$ back into (A.3) gives

$$\begin{aligned} F_p^{(1)}(R,\theta) &= \left[\sum_{k=0}^{p/2} \frac{\binom{p/2}{k}}{2k+1} (\tan\theta)^{2k+1}\right] \int_0^{R\cos\theta} \frac{u^{p+1}}{(1+u)^p} du \\ &= \sum_{k=0}^{p/2} \frac{\binom{p/2}{k}}{2k+1} (\tan\theta)^{2k+1} \sum_{j=0}^{p+1} (-1)^j \binom{p+1}{j} \int_1^{1+R\cos\theta} w^{1-j} dw \end{aligned} \tag{A.7}$$

where the latter equality is found by making the substitution $w = 1+u$ and then expanding the term $u^{p+1} = (w-1)^{p+1}$ in a binomial series.

Observing (A.7), one can see that the following single-integral in w must be solved for the case of $-1 \le m \le p$

$$f_m(R,\theta) = \int_1^{1+R\cos\theta} \frac{dw}{w^m}. \tag{A.8}$$

This integral has the solution

$$f_m(R,\theta) = \begin{cases} \frac{1}{2}\left[(1+R\cos\theta)^2 - 1\right] & , \quad m = -1 \\ R\cos\theta & , \quad m = 0 \\ \ln(1+R\cos\theta) & , \quad m = 1 \\ \frac{1}{m-1}\left[1-(1+R\cos\theta)^{1-m}\right] & , \quad m \geq 2 \end{cases} . \tag{A.9}$$

In terms of this result, the closed-form analytical solution of $F_p^{(1)}(R,\theta)$ is

$$F_p^{(1)}(R,\theta) = \sum_{k=0}^{p/2} \frac{\binom{p/2}{k}}{2k+1} (\tan\theta)^{2k+1} \sum_{j=0}^{p+1} (-1)^j \binom{p+1}{j} f_{j-1}(R,\theta), \quad p = 0,2,4,6,\ldots \tag{A.10}$$

A.2 Solution of $F_p^{(2)}(R,\theta)$

Denote the inner integral of (A.4) as $y(u)$. This integral is solved by first expanding the term $\left(u^2+v^2\right)^{p/2}$ in a binomial series and then integrating with respect to v. The result is

$$y(u) = \sum_{k=0}^{p/2} \frac{\binom{p/2}{k}}{2k+1} u^{p-2k} \left(R^2-u^2\right)^k \sqrt{R^2-u^2} \tag{A.11}$$

Substituting $y(u)$ back into (A.4) gives

$$\begin{aligned} F_p^{(2)}(R,\theta) &= \sum_{k=0}^{p/2} \frac{\binom{p/2}{k}}{2k+1} \int_{R\cos\theta}^{R} \frac{u^{p-2k}\left(R^2-u^2\right)^k \sqrt{R^2-u^2}}{(1+u)^p} du \\ &= \sum_{k=0}^{p/2} \frac{\binom{p/2}{k}}{2k+1} \sum_{j=0}^{k} (-1)^j \binom{k}{j} R^{2(k-j)} \\ &\quad \cdot \int_{R\cos\theta}^{R} \frac{u^{p-2(k-j)}\sqrt{R^2-u^2}}{(1+u)^p} du \end{aligned} \tag{A.12}$$

where the latter equality is found by expanding the term $\left(R^2-u^2\right)^k$ in a binomial series.

Observing (A.12), one can see that the following single-integral in u must be solved for the case of $0 \leq m \leq p$

$$g_m(R,\theta) = \int_{R\cos\theta}^{R} \frac{u^m\sqrt{R^2-u^2}}{(1+u)^p} du. \tag{A.13}$$

By making the substitution $w = 1 + u$ and then expanding the term $u^m = (w-1)^m$ in a binomial series, $g_m(R,\theta)$ may be written as

$$\begin{aligned} g_m(R,\theta) &= \int_{1+R\cos\theta}^{1+R} \frac{\sum_{l=0}^{m}(-1)^l \binom{m}{l} w^{m-l}\sqrt{R^2-(w-1)^2}}{w^p}dw \\ &= \sum_{l=0}^{m}(-1)^l \binom{m}{l} I_{p-m+l}(R,\theta) \end{aligned} \tag{A.14}$$

where

$$I_n(R,\theta) = \int_{1+R\cos\theta}^{1+R} \frac{\sqrt{R^2-(w-1)^2}}{w^n}dw. \tag{A.15}$$

Since $m \le p$ in (A.14), $I_n(R,\theta)$ need only be evaluated for $n \ge 0$. Suitably modifying [80, eq. (2.265)] gives the solution of $I_n(R,\theta)$ for $n \ge 0$ as the difference equation

$$\begin{aligned} I_n(R,\theta) &= \frac{1}{(n-1)(1-R^2)}\left[(2n-5)I_{n-1}(R,\theta)\right. \\ &\quad \left. +(4-n)I_{n-2}(R,\theta) - \frac{R^3\sin^3\theta}{(1+R\cos\theta)^{n-1}}\right] \end{aligned} \tag{A.16}$$

which has initial conditions

$$I_0(R,\theta) = \frac{1}{2}R^2\left(\theta - \frac{1}{2}\sin 2\theta\right) \tag{A.17}$$

and

$$I_1(R,\theta) = \theta - R\sin\theta + \sqrt{1-R^2}\left[\arcsin\left(\frac{R+\cos\theta}{1+R\cos\theta}\right) - \frac{\pi}{2}\right]. \tag{A.18}$$

Thus, starting with $I_0(R,\theta)$ and $I_1(R,\theta)$, $I_n(R,\theta)$ may be found in a recursive manner for any $n \ge 0$. In terms of the above results, the closed-form analytical solution of $F_p^{(2)}(R,\theta)$ is

$$F_p^{(2)}(R,\theta) = \sum_{k=0}^{p/2}\frac{\binom{p/2}{k}}{2k+1}\sum_{j=0}^{k}(-1)^j\binom{k}{j}R^{2(k-j)}g_{p-2(k-j)}(R,\theta), \quad p = 0,2,4,6,\ldots \tag{A.19}$$

Appendix B

Solution of Near-side Integral

In this appendix, a closed-form analytical solution is provided for the near-side integral

$$N_p(R,\theta) = \int\int_{S_N(R,\theta)} \frac{\left(u^2+v^2\right)^{p/2}}{(1+u)^p} du dv \tag{B.1}$$

where $R < 1$ and p is an integer. The region $S_N(R,\theta)$ is shown in Fig. 6.3(b). In order to avoid numerical instabilities when θ is near $\pi/2$, $N_p(R,\theta)$ is written as the following sum

$$N_p(R,\theta) = N_p^{(2)}\left(R,\frac{\pi}{2}\right) - \left[N_p^{(1)}(R,\theta) + N_p^{(2)}(R,\theta)\right] \tag{B.2}$$

where

$$N_p^{(1)}(R,\theta) = \int_{R\cos\theta}^{0} \frac{1}{(1+u)^p}\left(\int_0^{u\tan\theta}\left(u^2+v^2\right)^{p/2} dv\right) du \tag{B.3}$$

and

$$N_p^{(2)}(R,\theta) = \int_{-R}^{R\cos\theta} \frac{1}{(1+u)^p}\left(\int_0^{\sqrt{R^2-u^2}}\left(u^2+v^2\right)^{p/2} dv\right) du \tag{B.4}$$

correspond to sub-regions 1 and 2 respectively. In effect, the integral over $S_N(R,\theta)$ is calculated by integrating over the whole quadrant, then subtracting the integral over regions 1 and 2 from the result. Numerical instabilities are avoided in the same way as for the far-side integral by making the approximation $N_p^{(1)}(R,\theta) + N_p^{(2)}(R,\theta) \approx N_p^{(2)}(R,\theta)$ when θ is near $\pi/2$. Like before, (B.3) and (B.4) may be solved analytically for the case of p even. For the case of p odd, the following accurate approximation is used

$$N_p(R,\theta) \approx \sqrt{N_{p-1}(R,\theta)\, N_{p+1}(R,\theta)}, \quad p = 1,3,5,\ldots \tag{B.5}$$

B.1 Solution of $N_p^{(1)}(R,\theta)$

The integral in (B.3) is identical to that in (A.3) except that the limits of the u integral are reversed. Consequently,

$$N_p^{(1)}(R,\theta) = -F_p^{(1)}(R,\theta) \tag{B.6}$$

where the solution to $F_p^{(1)}(R,\theta)$ is given in (A.10). As expected, $N_p^{(1)}(R,\theta)$ is still a positive quantity since $F_p^{(1)}(R,\theta)$ evaluated at any θ in the interval $(\pi/2,\pi]$ is negative.

B.2 Solution of $N_p^{(2)}(R,\theta)$

The innter integral of (B.4) is identical to that of (A.4), but the limits of the outer integral are slightly different. Thus, $N_p^{(2)}(R,\theta)$ is obtained simply by changing the limits of the single integral in (A.12) giving

$$N_p^{(2)}(R,\theta) = \sum_{k=0}^{p/2} \frac{\binom{p/2}{k}}{2k+1} \sum_{j=0}^{k} (-1)^j \binom{k}{j} R^{2(k-j)} \int_{-R}^{R\cos\theta} \frac{u^{p-2(k-j)}\sqrt{R^2-u^2}}{(1+u)^p} du. \tag{B.7}$$

Following a similar development as in the previous appendix, define the integral

$$\begin{aligned} h_m(R,\theta) &= \int_{-R}^{R\cos\theta} \frac{u^m\sqrt{R^2-u^2}}{(1+u)^p} du \\ &= \sum_{l=0}^{m} (-1)^l \binom{m}{l} J_{p-m+l}(R,\theta) \end{aligned} \tag{B.8}$$

where

$$J_n(R,\theta) = \int_{1-R}^{1+R\cos\theta} \frac{\sqrt{R^2-(w-1)^2}}{w^n} dw. \tag{B.9}$$

Suitably modifying [80, eq. (2.265)], gives the solution of $J_n(R,\theta)$ for $n \geq 0$ as the difference equation

$$\begin{aligned} J_n(R,\theta) &= \frac{1}{(n-1)(1-R^2)} \left[(2n-5) I_{n-1}(R,\theta) \right. \\ &\quad \left. + (4-n) I_{n-2}(R,\theta) + \frac{R^3 \sin^3\theta}{(1+R\cos\theta)^{n-1}} \right] \end{aligned} \tag{B.10}$$

with has initial conditions

$$J_0(R,\theta) = \frac{1}{2}R^2\left(\pi - \theta + \frac{1}{2}\sin 2\theta\right) \tag{B.11}$$

and

$$J_1(R,\theta) = \pi - \theta + R\sin\theta - \sqrt{1-R^2}\left[\arcsin\left(\frac{R+\cos\theta}{1+R\cos\theta}\right) + \frac{\pi}{2}\right]. \tag{B.12}$$

In terms of the above results, the closed-form analytical solution of $N_p^{(2)}(R,\theta)$ is

$$N_p^{(2)}(R,\theta) = \sum_{k=0}^{p/2} \frac{\binom{p/2}{k}}{2k+1} \sum_{j=0}^{k} (-1)^j \binom{k}{j} R^{2(k-j)} h_{p-2(k-j)}(R,\theta), \quad p = 0,2,4,6,\ldots \tag{B.13}$$

Bibliography

[1] J. Litva and T.K. Lo, *Digital Beamforming in Wireless Communications*. Boston: Artech House, 1996.

[2] V.H. MacDonald, "The Cellular Concept," *Bell Systems Technical Journal*, vol. 58, pp. 15–41, January 1979.

[3] A.J. Paulraj and C.B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, pp. 49–83, November 1997.

[4] R.A. Monzingo and T.W. Miller, *Introduction to Adaptive Arrays*. New York: Wiley, 1980.

[5] R.O. Schmidt, *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*, Ph.D. Thesis, Stanford University, November 1981.

[6] R. Roy and T. Kailaith, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 984–995, July 1989.

[7] S. Haykin, *Adaptive Filter Theory,* 3rd ed. Upper Saddle River, NJ: Prentice Hall, 1996.

[8] J.G. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1989.

[9] J. Salz and J.H. Winters, "Effect of fading correlation on adaptive arrays in digital mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 1048–1057, November 1994.

[10] M. Schwartz, W. Bennett, and S. Stein, *Communications Systems and Techniques*. New York: McGraw Hill, 1966.

[11] J.H. Winters, J. Salz, and R.D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Transactions on Communications*, vol. 42, pp. 1740–1751, February/March/April 1994.

[12] J.H. Winters, "Optimum combining in digital mobile radio with cochannel interference," *IEEE Journal on Selected Areas of Communications*, vol. 2, pp. 528–539, July 1984.

[13] J.H. Winters, "Signal acquisition and tracking with adaptive arrays in the digital mobile radio system IS-54 with flat fading," *IEEE Transactions on Vehicular Technology*, vol. 42, pp. 377–384, November 1993.

[14] M.G. Hebley, W.K. Kennedy, and D.P. Taylor, "The performance of a space diversity receiver with finite-length fractionally-spaced equalisers with frequency selective Rayleigh fading and co-channel interference," in *Proc. IEEE GLOBECOM'94*, San Francisco, CA, November 1994.

[15] M.V. Clark, L.J. Greenstein, W.K. Kennedy, and M.Shafi, "Optimum linear diversity receivers for mobile communications," *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 47–56, February 1994.

[16] P. Balaban and J. Salz, "Optimum diversity combining and equalization in digital data transmission with applications to cellular mobile radio—Part I: Theoretical considerations," *IEEE Transactions on Communications*, vol. 40, pp. 885–894, May 1992.

[17] N.W.K. Lo, D.D. Falconer, and A.U.H. Sheikh, "Adaptive equalization for a multipath fading environment with interference and noise," in *Proc. IEEE VTC'94*, Stockholm, Sweden, June 1994, pp. 252–256.

[18] C.L.B. Despins, D.D. Falconer, and S.A. Mahmoud, "Compound strategies of coding, equalization, and space diversity for wide-band TDMA indoor wireless channels," *IEEE Transactions on Vehicular Technology*, vol. 41, pp. 369–379, November 1992.

[19] P. Monsen, "MMSE Equalization of interference on fading diversity channels," *IEEE Transactions on Communications*, vol. COM-32, pp. 5–12, January 1984.

[20] T. Fulghum and A. Duel-Hallen, "Adaptive multielement DFE receivers for narrowband multiuser detection," in *Proc. IEEE VTC'98*, Ottawa, Canada, May 1998, pp. 41–45.

[21] C. Tidestav, M. Sternad and A. Ahlén, "Reuse within a cell—interference rejection or multiuser detection?," *IEEE Transactions on Communications*, vol. 47, pp. 1511–1522, October 1999.

[22] G.E. Bottomley, K.J. Molnar, and S. Chennakeshu, "Interference cancellation with an array processing MLSE receiver," *IEEE Transactions on Vehicular Technology*, vol. 48, pp. 1321–1331, September 1999.

[23] S. Verdú, *Multiuser Detection*, Cambridge University Press, 1998.

[24] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Transactions on Information Theory*, vol. IT-32, pp. 85–96, January 1986.

[25] Z. Zvonar and D. Brady, "Multiuser detection in single-path fading channels," *IEEE Transactions on Communications*, vol. 42, pp. 1729–1739, February/March/April 1994.

[26] R. Lupus and S. Verdú, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Transactions on Communications*, vol. 38, pp. 496–508, April 1990.

[27] M.K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple-access communications," *IEEE Transactions on Communications*, vol. 38, pp. 509–519, April 1990.

[28] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a DS/CDMA system," *IEEE Journal on Selected Areas in Communications*, vol. 12, pp. 896–807, June 1994.

[29] Z. Xie, C.K. Rushforth, and R.T. Short, "Multiuser signal detection using sequential decoding," *IEEE Transactions on Communications*, vol. 38, pp. 578–583, May 1990.

[30] W. Van Etten, "Maximum likelihood receiver for multiple channel transmission systems," *IEEE Transactions on Communications*, vol. COM-24, pp. 276–283, February 1976.

[31] K. Giridhar et al., "Nonlinear techniques for the joint estimation of cochannel signals," *IEEE Transactions on Communications*, vol. 45, pp. 473–484, April 1997.

[32] S.D. Gray, M. Kocic, and D. Brady, "Multiuser detection in mismatched multiple-access channels," *IEEE Transactions on Communications*, vol. 43, pp. 3080–3089, December 1995.

[33] P.A. Ranta, A. Hottinen, and Z. Honkasalo, "Co-channel interference cancelling receiver for TDMA mobile systems," in *Proc. IEEE ICC'95*, Seattle, WA, June 1995, pp. 17–21.

[34] S.J. Grant and J.K. Cavers, "Performance enhancement through joint detection of cochannel signals using diversity arrays," *IEEE Transactions on Communications*, vol. 46, pp. 1038–1049, August 1998.

[35] S.J. Grant and J.K. Cavers, "Performance enhancement through joint detection of cochannel signals using diversity arrays," in *Proc. IEEE VTC'98*, Ottawa, Canada, May 1998, pp. 1039–1043.

[36] S.J. Grant and J.K. Cavers, "Further analytical results on the joint detection of cochannel signals using diversity arrays," to appear in *IEEE Transactions on Communications*, vol. 48, November 2000.

[37] M. Brandt-Pearce, "Signal separation using fractional sampling in multiuser communications," *IEEE Transactions on Communications*, vol. 48, pp. 242–251, February 2000.

[38] J. Joung and G.L. Stüber, "Performance of truncated co-channel interference canceling MLSE for TDMA systems," in *Proc. IEEE VTC'98*, Ottawa, Canada, May 1998, pp. 1710–1714.

[39] B.C.W. Lo and K.B. Lataief, "Adaptive equalization and interference cancellation for wireless communication systems," *IEEE Transactions on Communications*, vol. 47, pp. 538–545, April 1999.

[40] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array—Part I: Algorithms," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1184–1197, May 1996.

[41] S. Talwar and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array—Part II: Performance," *IEEE Transactions on Signal Processing*, vol. 45, pp. 706–718, May 1997.

[42] J. Wang and J.K. Cavers, "Block based blind joint detection of co-channel signals over Rayleigh flat-fading channels with an antenna array," in *Proc. IEEE VTC'99-Spring*, Houston, TX, May 1999.

[43] J. Wang and J.K. Cavers, "Blind joint detection of cochannel CPFSK signals over Rayleigh flat-fading channels with a diversity antenna array," in *Proc. IEEE VTC'00-Spring*, Tokyo, Japan, May 2000.

[44] H. Liu and G. Xu, "Smart antennas in wireless systems: Uplink multiuser blind channel and sequence detection," *IEEE Transactions on Communications*, vol. 45, pp. 187–189, February 1997.

[45] A. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1136–1155, May 1996.

[46] T. E. Biedka, W.H. Tranter, and J.H. Reed, "Convergence analysis of the least mean squares constant modulus algorithm in interference cancellation applications," *IEEE Transactions on Communications*, vol. 48, pp. 491–501, March 2000.

[47] P. DeRusso, R. Roy, and C. Close, *State Variables for Engineers*, New York: John Wiley and Sons, 1965.

[48] J.K. Cavers, "Single user and multiuser adaptive maximal ratio transmission for Rayleigh channels," accepted for publication in *IEEE Transactions on Vehicular Technology*, April 2000.

[49] J.C. Guey, M.P. Fitz, M.R. Bell, and W.Y. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 47, pp. 527–537, April 1999.

[50] V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.

[51] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, pp. 41–59, September-October 1996.

[52] S.J. Grant and J.K. Cavers, "Transmit arrays with joint detection for increasing downlink capacity," submitted to *IEEE Transactions on Communications*, July 2000.

[53] B.A. Bjerke and J.G. Proakis, "Multiple-antenna diversity techniques for transmission over fading channels," in *Proc. IEEE WCNC'99*, New Orleans, LA, September 1999.

[54] J.K. Cavers and P. Ho, "Analysis of the performance of trellis-coded modulations in Rayleigh-fading channels," *IEEE Transactions on Communications*, vol. 40, pp. 74–83, January 1992.

[55] L. Welburn and J.K. Cavers, "Accurate error rate calculations through the inversion of characteristic functions with both mulitple and discrete poles," SFU Internal Report, February, 2000.

[56] G.D. Forney Jr., "The Viterbi algorithm," *IEEE Proceedings*, vol. 61, pp. 268–278, March 1973.

[57] J.K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 686–693, November 1991.

[58] S.A. Fechtel and H. Meyer, "Optimal parametric feedforward estimation of frequency-selective fading radio channels," *IEEE Transactions on Communications*, vol. 42, pp. 1639–1650, February/March/April 1994.

[59] N.W.K. Lo, D.D. Falconer, and A.U.H. Sheikh, "Adaptive equalization and diversity combining for mobile radio using interpolated channel estimates," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 636–645, August 1991.

[60] B.C. Ng, M. Cedervall, and A. Paulraj, "A structured channel estimator for maximum-likelihood sequence detection," *IEEE Communication Letters*, vol. 1, pp. 52–55, March 1997.

[61] M.L. Moher and J.H. Lodge, "TCMP-a modulation and coding strategy for Rician fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 1347–1355, December 1989.

[62] F. Ling, "Coherent detection with reference symbol based channel estimation for direct sequence CDMA uplink communications," in *Proc. IEEE VTC'93*, Secaucus, NJ, May 1993, pp. 335–344.

[63] O. Nesper and P. Ho, "A pilot symbol assisted interference cancellation scheme for an asynchronous DS/CDMA system," in *Proc. IEEE Globecom'96*, London, UK, November 1996, pp.1447–1451.

[64] S.J. Grant and J.K. Cavers, "Multiuser channel estimation for detection of cochannel signals," submitted to *IEEE Transactions on Communications*, October 1998.

[65] S.J. Grant and J.K. Cavers, "Multiuser channel estimation for detection of cochannel signals," in *Proc. IEEE ICC'99*, Vancouver, Canada, June 6-10, 1999.

[66] S.N. Crozier, D.D. Falconer, and S.A. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *IEE Proceedings-F*, vol. 138, pp. 371–378, August 1991.

[67] J.C.L. Ng, K.B. Letaief, and R.D. Murch, "Complex optimal sequences with constant magnitude for fast channel estimation initialization," *IEEE Transactions on Communications*, vol. 46, pp. 305–308, March 1998.

[68] D.V. Sarwate and M.B. Pursley, "Crosscorrelation properties of pseudorandom and related sequences," *IEEE Proceedings*, vol. 68, pp. 593–619, May 1990.

[69] Y.Yeh and S.C. Schwartz, "Outage probability in mobile telephony due to multiple log-normal interferers," *IEEE Transactions on Communications*, vol. 32, pp. 380–388, April 1984.

[70] K.S. Gilhousen et al., "On the capacity of a cellular CDMA system," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 303–312, May 1991.

[71] S. Chennakeshu, A.A. Hassan, J.B. Anderson, and B. Gudmundson, "Capacity analysis of a TDMA-based slow-frequency-hopped cellular system," *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 531–542, August 1996.

[72] P.A. Ranta, Z. Honkasalo, and J. Tapaninen, "TDMA cellular network application of an interference cancellation technique," in *Proc. IEEE VTC'95*, Chicago, IL, May 1995.

[73] A.F. Naguib, A. Paulraj, and T. Kailath, "Capacity improvement with base-station antenna arrays in cellular CDMA," *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 691-698, August 1994.

[74] S.J. Grant and J.K. Cavers, "Analytical calculation of outage probability for a general cellular mobile radio system," in *Proc. IEEE VTC'99-Fall*, Amsterdam, The Netherlands, September 1999.

[75] S.J. Grant and J.K. Cavers, "Increased uplink capacity for TDMA systems through joint detection with diversity arrays," in *Proc. IEEE VTC'99-Fall*, Amsterdam, The Netherlands, September 1999.

[76] S.J. Grant and J.K. Cavers, "Analytical evaluation of the uplink capacity of a general cellular mobile radio system," submitted to *IEEE Transactions on Communications*, January 2000.

[77] S.J. Grant and J.K. Cavers, "Increased uplink capacity for TDMA systems through joint detection with diversity arrays," submitted to *IEEE Transactions on Communications*, February 2000.

[78] N.C. Beaulieu, A.A. Abu-Dayya, and P.J. McLane, "Estimating the distribution of a sum of independent log-normal random variables," *IEEE Transactions on Communications*, vol. 43, pp. 2869–2873, December 1995.

[79] D.C. Schleher, "Generalized Gram-Charlier series with application to the sum of log-normal variates," *IEEE Transactions on Information Theory*, pp. 275–280, March 1977.

[80] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*. San Diego: Academic Press, 1994.