POWER ADAPTATION STRATEGIES FOR DELAY CONSTRAINED CHANNELS

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Rohit Negi

September 2000

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John Cioffi
(Principal Adviser)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Andrea Goldsmith

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Benjamin Van Roy

Approved for the University Committee on Graduate Studies:

Abstract

This thesis explores the effect of a delay constraint on optimum transmission strategies in a fading communication channel. Here, the delay constraint is expressed in terms of the fading rate of the channel process rather than in terms of the symbol rate. A block fading channel model is considered, and a delay constraint is imposed on data transmission. Thus, processing of data is assumed to occur in (a small number of) K blocks, each block itself being of (large) length T_0. The idea is to capture the non-ergodicity of the fading process for the duration of the data processing, due to the small K, while at the same time allow limiting arguments to be used to calculate 'capacity', facilitated by the large T_0. A general cost function $\mu(x)$ is considered in solving the delay constrained transmission problem. Two kinds of power constraints are considered, the short term and the long term power constraints. The solution to the long term power constraint is better (results in a higher maximum) because it is a more relaxed constraint. Since the power adaptation that maximizes the cost function has to be causal, hence dynamic programs are found to give the optimum power control solution.

The general cost function is then specialized to the case of expected capacity, by choosing $\mu(x) = x$. Expected capacity is the maximum ensemble-average data rate that can be obtained by optimizing the transmission power. It is observed that

optimizing the transmitted power does not give much benefit at high SNRs, but provides a substantial gain at lower SNRs. At low SNRs, it is proved that the factor increase in capacity, due to power adaptation, is approximately $\frac{\log K}{m}$ if the channel fades according to the χ^2_{2m} statistics.

The case of outage capacity is considered next. Outage capacity is defined as the maximum error-free data rate that can be supported at a given outage probability. Here, outage is the event that a given target rate R_0 cannot be supported by a fading channel over a given time period. It is shown that the optimum power adaptation solution to the long term constraint problem gives a substantial SNR gain at both low and high SNRs. The solution to the short term constraint problem however does not provide any SNR gain at high SNRs. It is shown that whereas the outage probability is inversely related (with a power of m) to the SNR in the short term case, it is related at least exponentially in the long term case. Random coding bounds are derived for the outage capacity case.

A stationary version of the outage probability problem is also considered. The formulation uses an exponential window which weighs the past data rates to approximate a K block window. Stationarity is introduced in the formulation by considering a time-averaged optimization. The solution involves linear programming. An SNR gain is observed when the optimal power control is used rather than a constant power scheme.

Space-time codes are considered as an example of the use of outage probability. In particular, the case of an outdoor wireless multi-antenna transmission system is considered. It is shown that maximum diversity and SNR gain could be obtained by simply combining space-time codes with an appropriate 'beamformer'.

Finally, the problem of blind symbol synchronization in OFDM is considered.

It is shown that the ranks of certain autocorrelation matrices contain information that can be used to blindly synchronize the received signal, even in the presence of multipath. As opposed to previously existing blind synchronization methods, the new algorithm is shown to guarantee correct synchronization asymptotically.

Acknowledgements

As with any major endeavor, a doctoral thesis cannot be completed without the support of colleagues, friends, family and well-wishers. I'm grateful to several folks who saw me through the trial that is the doctoral program, all of whom deserve mention. Prof. John Cioffi, also called 'our leader' or simply 'hey John', as my advisor was a source of inspiration and ideas. His extremely successful professional life has been a strong motivating factor during my Ph.D. From him I've learned that intuition precedes mathematics in any good engineering work. Supporting his research group, our administrative assistant, Joice, has always been very kind and helpful. I'm grateful to my associate advisors, Prof. Tom Kailath and Prof. Andrea Goldsmith for their interest in this thesis. Thanks also to Prof. Ben Van Roy for the interesting discussions we had on dynamic programming. Chapter 5 has borrowed several ideas from these discussions.

The thesis work was supported in various phases by a Stanford Fellowship, an NSF grant and a three year gift by Texas Instruments Inc. I'm grateful to these sponsors for their support.

I would like to thank my co-authors in various publications - Ardavan Maleki, Anand Dabak, Sri Hosur, Sriram Vishwanath, Wonjoon Choi. It was a pleasure working with you. It was also enjoyable working with the 'Cioffi kids', as we in

the research group like to call ourselves. The cooperative, laid-back atmosphere of the group was the key in developing novel ideas for research. In particular, I would like to thank John Fan, Jose Tellado, Kok-Wui Cheong and Acha Leke for the several interesting discussions we had on research issues. I also had the opportunity to participate in internship programs over three summers. I would like to thank my mentors Sanjay Kasturia, Dan Avidor, Debajyoti Pal, Alan Gatherer, and Don Shaver for making these internships both useful and fun.

I would like to thank my friends for adding the element of fun in Ph.D. life. A special note of thanks goes to Moses Charikar, my room-mate of five years and a friend of twice as much. I wish him well in his pursuit of an academic career. Thanks also to Dinkar Singh for advice on the finer (i.e. non-academic) aspects of life. Mayur Joshi, the practical one was always there to keep my feet firmly on the ground (most of the time). Kok-Wui Cheong, Louise Hoo and Jose Tellado have been not only research colleagues but also caring friends. It was a lot of fun 'hanging out' with you. Julia Schmidt, Tanya Logvinenko and Valeria Bertacco were always there to liven up a slow quarter with a fun party. It was a pleasure knowing you.

Last but not the least I would like to thank my family, particularly my parents for the support they've provided during the Ph.D. My father has been a constant source of motivation (He's a physicist, but I suspect he's a better engineer than I am!). My mother has been very understanding about the trials that come with research. Her words of encouragement helped tide me through several rough patches. This thesis is dedicated to my parents.

Contents

Abstract **v**

Acknowledgements **viii**

Notations **xviii**

1 Introduction **1**

1.1 Classical Notion - Ergodic Capacity 2

1.2 Delay Constrained Notions - Expected and Outage Capacities 5

1.3 Time Diversity 9

2 Dynamic Programming for Delay Constrained Channels **13**

2.1 Introduction 13

2.1.1 Fading Channels with Delay Constraint 14

2.1.2 Problem Formulation 16

2.2 Capacity of feedback channels 18

2.3 Power Adaptation for Delay-Constrained Channels 21

2.4 Equal power strategy 25

3 Expected Capacity **29**

3.1 Introduction 29
3.2 Power Adaptation 30
3.2.1 Dynamic programming solution 30
3.2.2 Comments 32
3.2.3 Channels with low SNR 32
3.2.4 Comments 34
3.3 Performance of the Online Algorithms 35
3.4 Low SNR bounds for expected capacity 36
3.5 Long term constraint problem 43

4 Outage Capacity **47**

4.1 Introduction 47
4.2 Power Adaptation for Minimum Outage Probability 49
4.2.1 Power adaptation with short term power constraint 49
4.2.2 Power adaptation with long term power constraint 51
4.3 Performance of the Online Algorithms 53
4.3.1 Simulations 53
4.3.2 A Note on Discretization 54
4.4 Asymptotic behavior 58
4.4.1 Short term constraint policy 58
4.4.2 Long term constraint policy 61
4.5 Random coding bounds 64
4.5.1 Exact computation 66
4.5.2 Small deviation from target rate 66
4.5.3 A weaker upper bound 69

5 Stationary Power Control Strategies **78**

5.1 Problem Formulation 79

5.2 Optimal Solution 80

5.2.1 Optimal scheme 80

5.2.2 Sub-optimal scheme 83

5.3 Simulation and Discussion 83

6 Adaptive Antennas for Space-Time Codes in Outdoor Channels **88**

6.1 Introduction 88

6.2 Problem Formulation 90

6.2.1 MIMO Wireless System 91

6.2.2 Assumptions 92

6.3 ISI-free block time-invariant MIMO system 93

6.4 MIMO Wireless System with ISI 96

6.5 Simulation Results 99

6.5.1 The ISI-free Case 99

6.5.2 The Case with ISI 102

7 Blind OFDM Symbol Synchronization in ISI Channels **107**

7.1 Introduction 107

7.2 Problem Formulation 110

7.2.1 Notations 110

7.2.2 OFDM System 110

7.2.3 Assumptions 111

7.2.4 Problem Formulation 113

7.3 Theoretical Basis For Blind Symbol Synchronization 114

7.4 Practical Considerations 118
7.4.1 Estimation with σ_n^2 known 118
7.4.2 Estimation with σ_n^2 unknown 120
7.4.3 Computational Complexity 120
7.4.4 Consistency of Algorithm 122
7.5 Simulation Results 123

8 Conclusion **127**

A Appendix **130**
A.1 The method of bisection 130
A.2 A refinement of the upper bound in Section 4.5.3 131
A.3 Proof of optimal beamformer for ISI-free case 132
A.4 Proof of lower bound on gain for linear array 133
A.5 Proof of Lemma 4 134
A.6 Proof of Theorem 4 135
A.7 Proof of Corollary 1 138
A.8 Rank Estimation using MDL 139

Bibliography **141**

List of Tables

7.1 Normalized wall energy after synchronization, using the new method and using the Beek-Sandell algorithm, for various values of P, and $SNR = 30$ dB 126

7.2 Normalized wall energy after synchronization, using the new method and using the Beek-Sandell algorithm, for various values of SNR, and $P = 2400$ symbols 126

List of Figures

1.1 A typical outdoor wireless system 3

1.2 Delay constrained block fading channel model 9

2.1 Transmission system for flat block-fading channel 16

2.2 Various possible utility function $\mu(x)$s. $\mu_1(x)$ results in expected capacity, while $\mu_2(x)$ results in outage capacity 19

2.3 Dynamic program used to compute optimal power scheme 22

2.4 Optimal power functions $P_n^*(g,R,P), Q_n^*(g,R,P)$ used when the system is in operation 24

3.1 Capacity improvement ratio for low SNRs and various channel gain distributions 36

3.2 Capacity improvement ratio of optimum power adaptation in a Rayleigh fading channel for various SNRs 37

3.3 Capacity improvement ratio of suboptimum power adaptation in a Rayleigh fading channel for various SNRs 38

3.4 Capacity ratio of optimum power adaptation to suboptimum power adaptation in a Rayleigh fading channel for various SNRs 39

4.1 Short term constraint algorithm in Rayleigh fading for $R_0 = 0.1$ and various K. Solid lines: optimum algorithm; dotted lines: no power adaptation 55

4.2 Long term constraint algorithm in Rayleigh fading for $R_0 = 0.1$ and various K. Solid lines: short term algorithm; dotted lines long term algorithm 56

4.3 Long term constraint algorithm in Rayleigh fading for $R_0 = 3$ and various K. Solid lines: long term constraint algorithm; dotted lines no power adaptation 57

4.4 Trellis to calculate a weak random coding bound 75

4.5 Weak random coding bound for long term constraint algorithm for $K = 5$, $SNR = 20$ dB, $R_0 = 3$. Solid lines are the random coding bound, while dotted line is the outage probability ($T_0 \to \infty$) 76

4.6 Error exponent for long term constraint algorithm for $K = 5$, $SNR = 20$ dB, $R_0 = 3$, for various values of R 77

5.1 Outage probability at high SNRs for $R_0 = 4$, $\alpha = 0.3$. Solid lines refer to the new algorithms, while the dotted line refers to constant power transmission 85

5.2 Outage probability at low SNRs for $R_0 = 0.15$, $\alpha = 0.3$. Solid lines refer to the new algorithms, while the dotted line refer to constant power transmission 86

5.3 Histogram of state R_t for a sample run, at $SNR = 10$ dB, $R_0 = 4$, $\alpha = 0.3$ 87

6.1 Illustration of the physical multipath channel 91

6.2 Histograms $\sqrt{g(8,L)}$, $\sqrt{g(L,L)}$,$\sqrt{g(8,L)/g(L,L)}$, $L=2$ 100
6.3 FER v.s. SNR; ISI-free case; angle pairs $\{20^o, 60^o\}$ 101
6.4 FER v.s. SNR; ISI-free case; angle pairs $\{50^o, 60^o\}$ 102
6.5 Histograms of the SNR gain $P(b)$ for $n=4$ 103
6.6 Histograms of the SNR gain $P(b)$ for $n=8$ 104
6.7 FER v.s. SNR; the ISI case; angle pairs $\{30^o, 75^o\}$ 105
6.8 FER v.s. SNR; the ISI case; angle pairs $\{30^o, 35^o\}$ 106

7.1 Basic OFDM system 109
7.2 Notation for the data and received signal sequences 112
7.3 Theoretical behavior of $\rho[\mathbf{R}_{\mathbf{z}}(i)]$ for some channels: $N=128, \nu=32, L_0=26$ 117
7.4 Impulse response of SFN channels 122
7.5 Decision functions used by Beek-Sandell algorithm. Synchronization position estimate is the maximum of the decision function 123
7.6 Decision function used by the proposed algorithm. Synchronization position estimate is the minimum rank position 124
7.7 Impulse response of channels used for simulation: Channel C_2 has more ISI than channel C_1 125

Notations

$\lceil x \rceil$	The minimum integer $n \geq x$
$\mathbf{E}[x]$	Expected value
$\mathop{\mathbf{E}}_{y}[x]$	Expected value over Y
$f_X(x)$	Probability density function of X
$F_X(x)$	Cumulative distribution function of X
$\mathbf{P}[A]$	Probability of event A
$\mathbf{x}$	Vector
$g^{(n)}$	$\{g_1, g_2, \ldots, g_n\}$
g_m^n	$\{g_m, g_{m+1}, \ldots, g_n\}$
$\mathbf{1}_F(A)$	$= \begin{cases} 1 & \text{if event } A \text{ is true} \\ 0 & \text{else} \end{cases}$
$\mathrm{tr}[A]$	Trace of matrix A
$\mathbf{diag}\,(x, y)$	Diagonal matrix $\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}$

Chapter 1

Introduction

Wireless communications has been ubiquitous ever since radio waves were first used a century back by Guglielmo Marconi to communicate across the Alantic. It has evolved from being a point to point telegraphic system, to being widely used for radio broadcasting, and then finding new life in cellular, mobile telephony. Over the course of evolution of wireless communications, its technology and understanding has evolved hand in hand. Increasing sophistication in radio circuits have been complemented with exponentially rising digital computing power, to such an extent that 'software radios' which are more computing machines than radio circuits, may soon become a reality. This has been made possible by the availability of the digital signal processor (DSP), and the theory of signal and information processing.

Ever since its inception half a century back, information theory has been primarily applied to the field of communications. In fact, wireless communication media (channels) have been studied even in the early decades of information theory. However, the complexity and variety of these channels is such that though our understanding of the subject is considerable, there are several areas where ideas are

only partially understood, and some areas where we've barely scratched the surface. Wireless channels suffer from limitations arising from

- Channel dispersion, which causes a transmitted signal to spread out in time. The theory of equalization and detection deals with this issue.
- Broadcast nature, because the wireless medium is shared by all users, unlike wireline telephony. The problem of interchannel interference is complex, because optimal solutions are usually exponentially complex in the number of users.
- Channel fading, which causes fluctuations in the received signal power. Information theory and coding provide the tools to analyze and solve this problem.

Each of these issues is an area of study in its own right. This thesis will focus on problems involving channel fading. Consider the outdoor wireless communication system in Figure 1.1. The signal received at the mobile receiver is the result of a superposition of several signals. This causes a standing wave pattern in space, and so when the mobile receiver passes through space, the power received by it fluctuates. This fluctuation in power level is called 'fading'.

1.1 Classical Notion - Ergodic Capacity

Information theoretic results on fading channels have been available long enough to be considered classic [53]. However, the more recent past has yielded considerable new insights into fading channels, driven to a large extent by the immediate needs to solve practical 'industrial' problems. Current research on the issue focuses on solving such problems for a variety of interesting practical scenarios. [4] provides an elegant overview of such results, and also a comprehensive list of references.

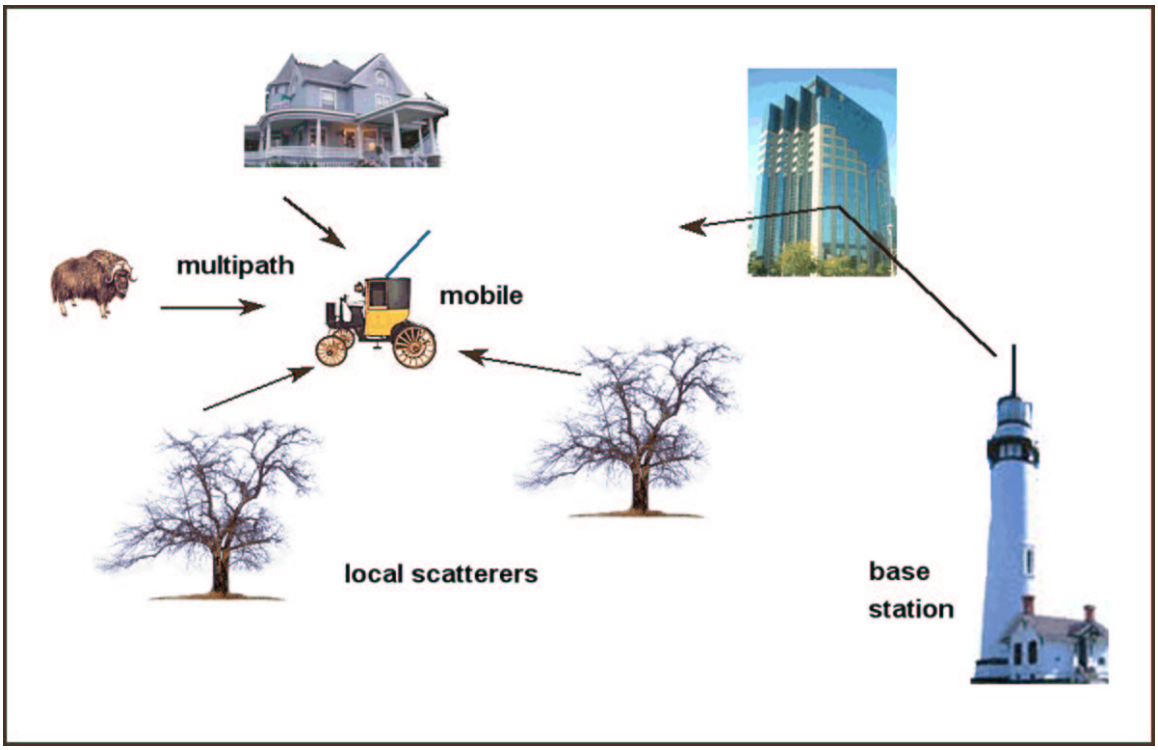

Figure 1.1: A typical outdoor wireless system

The simplest formulation of the problem of communicating over a fading channel is as follows. Let the X, Y and α be the input, output and state of the channel respectively. Define the *ergodic* capacity of the channel as

$$\begin{aligned} C_{erg} &= \max \quad I(X;Y \mid \alpha) \\ &= \int \max \ [\ h(Y|G=\alpha) - h(Y|X,G=\alpha)\ ]\ f_\alpha(\alpha)\ \mathrm{d}\alpha \end{aligned} \tag{1.1}$$

where the maximization is over the conditional input distribution[1] $f_X(x|\alpha)$. $I(\cdot;\cdot|\alpha)$ denotes the conditional mutual information, while $h(\cdot|\alpha)$ denotes the conditional entropy [12]. This definition of capacity is valid under the conditions that,

[1] Refer to the notations defined before Chapter 1.

- The receiver knows the channel state perfectly. This is reflected in the conditioning of the mutual information on α. This thesis will always make this assumption. The effect of imperfect channel information on capacity is hard to analyze, even in the simpler case of ergodic capacity. See [28] for an interesting analysis that shows the extreme sensitivity of the well-known Gaussian coding scheme to channel state information. See also [35],[27] and the references therein for a discussion on mismatched decoding.

- The transmitter knows the channel state information. This is reflected in the fact the the input symbols are chosen from the conditional density $f_X(x|\alpha)$. If the maximization in (1.1) is over $f_X(x)$, then the assumption is that the transmitter does not know the channel state information. However, in our case, we will assume that the transmitter knows the channel state information.

- The codeword length is long enough (ideally, code length $\rightarrow \infty$), so that the ergodic property of the channel state sequence (which is implicitly assumed) kicks in.

It is because of the last requirement that this formula for capacity is called the *ergodic* capacity. Conceptually, this is the most direct extension of the classical Shannon capacity idea for a fixed channel, in which α is constant and known. Thus, if the data rate $R < C_{erg}$, then there exist a sequence of codes $\mathcal{C}_T$ of time length T such that the decoding error rate $P_{err} \rightarrow 0$ as $T \rightarrow \infty$. Also for any $R > C_{erg}$, any code will result in P_{err} being bounded away from 0 (typically $P_{err} \rightarrow 1$ as $T \rightarrow \infty$, a result that is called the 'Strong converse of channel capacity').

Several interesting results exist on ergodic capacity. It can be computed analytically in a number of instances, such as for a Rayleigh fading channel. [20] showed

how this results in a *time-waterfilling* algorithm in a fading AWGN channel. For this channel $y = \alpha x + n$, n being additive white Gaussian noise with variance σ_n^2, and α is the channel gain with probability density $f_\alpha(\alpha)$. The time-waterfilling algorithm controls the power transmitted (variance of X) as

$$\begin{aligned} P(\alpha) &= \left[c - \frac{\sigma_n^2}{|\alpha|^2}\right]^+ \qquad \text{where } c \text{ is chosen so that} \\ \mathop{\mathbf{E}}_{\alpha}[P(\alpha)] &= P_0 \end{aligned} \tag{1.2}$$

c is a positive constant, while P_0 is the average power alloted to the transmitter[2]. The codewords are chosen from a Gaussian codebook. [22] extends this result to the multiuser case, where several users are assumed to share a common transmission medium. It also considers delay-limited capacity [4], which does not require any power adaptation over time (power adaptation over users, for the 'worst-case' channel is required). See also [3], which considers a simple multiple access algorithm, which includes delay considerations. Note the assumption that channel statistics $f_\alpha(\alpha)$ are assumed to exist. In particular, this thesis does not consider the more general, but practically limited case of an 'arbitrarily varying channel' [1],[14] in which no such statistics are presumed to exist.

1.2 Delay Constrained Notions - Expected and Outage Capacities

The calculation of C_{erg} in Section 1.1 emphasizes the use of ergodicity. The expectation over α in (1.1) will translate to a time average only when the codewords are

[2]Refer to the notations defined before Chapter 1.

long enough to capture the ergodicity of the channel fading process. Otherwise the 'capacity' will be a random variable whose value depends on the particular instantiation of the channel fading process over the duration of the codeword. In several practical situations, the codes necessarily have to be short, depending on the latency (delay) that the system can tolerate. A very direct example is voice transmission, in which a high latency would result in unacceptable delays in speech. At the same time, wireless systems require a guaranteed target performance. For example, voice traffic or real-time video traffic require data transmission at a pre-specified target data rate. Such systems require that the target data rate is met with a low probability of outage [4]. However channel fading limits data rate. A deep fade, where the received power is very low, may reduce the data rate that can be supported for the duration of the fade. For systems that can tolerate a high latency, the code length can be made large enough, so that the ergodic capacity of the channel is achieved, such as in [20]. However, [20] does not deal with the case of finite delay constraint. In that case, we would be interested in achieving the specified target criteria, assuming that the codeword is restricted to a specified time span. This gives rise to the important notions of expected and outage capacities. For the most part, this thesis will concentrate on defining and solving problems based on the notions of expected and outage capacities. Therefore, a brief description of these notions follow. For more details, the reader can refer to [4] and [12].

Consider the (discrete time) transmission system with channel inputs $x^{(T)}$, and outputs $y^{(T)}$. Let the channel state for the time period in question $\{1, 2, \ldots, T\}$ be parameterized by $g^{(T)}$. Recall that the notation $g^{(T)}$ denotes the set $\{g_1, g_2, \ldots, g_T\}$. Here, T is the delay constraint that is imposed by the application. Data processing has to occur within these T symbols. Thus, time is divided into segments of length

T. Within each such segment, the application demands a certain performance. Then, a natural quantity to look at from an information theoretic perspective is the average mutual information $\frac{1}{T}I(X^{(T)};Y^{(T)} \mid g^{(T)})$. In the case of ergodic capacity, this quantity reduced to the deterministic quantity $I(X;Y \mid G)$, because $T \to \infty$, and the fading process was assumed to be ergodic. The capital G here denotes averaging over the random variable G, i.e. $I(X;Y \mid G) = \underset{g}{\mathbf{E}}[I(X;Y \mid g)]$. In contrast note that $\frac{1}{T}I(X^{(T)};Y^{(T)} \mid g^{(T)})$ is the mutual information conditioned on a specific instantiation of the fading process $g^{(T)}$. Therefore $\frac{1}{T}I(X^{(T)};Y^{(T)} \mid g^{(T)})$ is a random variable whose value depends on $g^{(T)}$. The notions of expected and outage capacities have been developed as meaningful measures of this random variable.

Expected capacity : Expected capacity is defined as

$$C_{exp} = \max \underset{g^{(T)}}{\mathbf{E}}[\frac{1}{T}I(X^{(T)};Y^{(T)} \mid g^{(T)})]$$

The maximization is over the input distribution $f_{X^{(T)}}(x^{(T)} \mid g^{(T)})$ (assuming that the transmitter knows the channel state information). This notion shows how much rate can be transmitted in each segment of length T, on the average.

Outage capacity : In voice transmission, an 'outage' is said to occur when a data packet containing a voice signal is dropped due to errors. The notion of outage capacity formalizes this idea. Suppose the application needs a constant data rate of R_0 bits/sec, such as in voice telephony. In a given segment of length T, suppose the channel fade is such that the rate R_0 cannot be supported, i.e. $\frac{1}{T}I(X^{(T)};Y^{(T)} \mid g^{(T)}) < R_0$. In that case, an 'outage' is said to occur. Clearly it is desirable to minimize the probability of an outage, by optimizing

the transmission. Therefore, we need to compute the following,

$$P_{out} \;=\; \min \;\; \mathbf{P}[\frac{1}{T} I(X^{(T)}; Y^{(T)} \mid g^{(T)}) \;<\; R_0]$$

Then, the outage probability *at the target rate* R_0 is P_{out}. Another way to say this is that the outage capacity at the outage probability P_{out} is R_0. Clearly then, outage capacity is a function of the allowed outage probability P_{out}. Thus, outage capacity can be thought of as the cumulative distribution function of capacity.

In this thesis, we will assume a block fading channel model [4]. This assumes that the channel power gain g_i for block i is a scalar constant over that block consisting of a large number (T_0) of symbols (slow fading). This makes it possible to apply limiting (large T_0) arguments to $\frac{1}{T} I(X^{(T)}; Y^{(T)} \mid g^{(T)})$, so that it represents a bound on the largest decoding-error-free data rate that can be transmitted. If the delay constraint T is divided into blocks of length T_0, then it covers $K = \lceil T/T_0 \rceil$ blocks, labelled $\{i = 1, 2, \ldots, K\}$. Figure 1.2 shows a delay constrained block fading channel. The dotted line represents the (fading) channel gain $g(t)$ as a function of time t. Each shaded bar represents one block of data. For a block fading channel, in the limiting case of large T_0, $\frac{1}{T} I(X^{(T)}; Y^{(T)} \mid g^{(T)})$ converges to $\frac{1}{K} I(X^{(K)}; Y^{(K)} \mid g^{(K)})$, where (with a slight abuse of notation) $g^{(K)}$ now denotes the channel gains of the K blocks. The thesis will look at this form of the average mutual information, because of the block fading assumption. As stated earlier, the core focus of the thesis is on delay constrained channels, and in particular, on the expected and outage capacity formulations of transmission optimization on such channels.

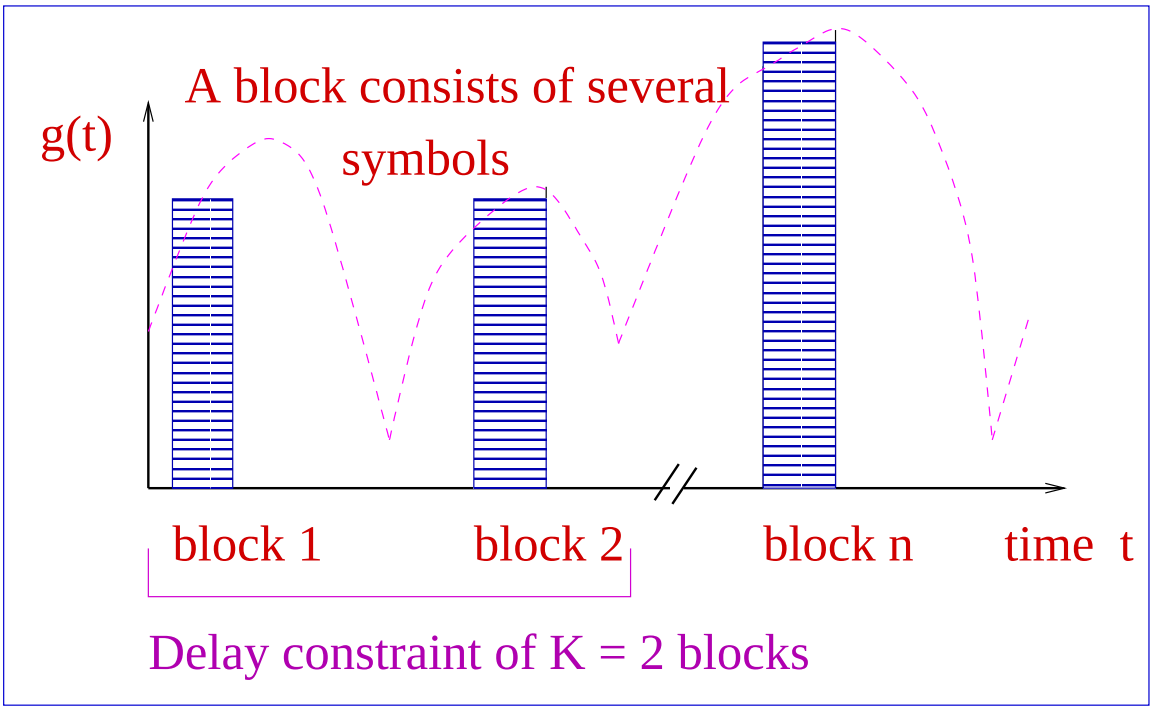

Figure 1.2: Delay constrained block fading channel model

1.3 Time Diversity

Diversity plays a very useful role in fading channels. This is true even in the case of ergodic capacity. The time-waterfilling scheme in [20] exploits time diversity. It may be that this capacity exceeds the capacity of the corresponding AWGN channel having the same average channel gain. Even if the AWGN capacity is not surpassed, knowledge of the channel state information at transmitter may significantly increase the channel capacity over the case where the transmitter does not know the channel state information. The capacity of a fading AWGN channel is always less than or equal to the capacity of a fixed AWGN channel with the same average gain (unless the transmitter exercises power control as mentioned above). However, diversity, which implies several independent channels carrying the same information, can close this

capacity gap, as in [20]. In addition to the time diversity scheme of time-waterfilling, diversity can also be exploited in other forms - space, frequency [49], user [25], modulation [5]. For signal processing or coding approaches to diversity, see [62],[56],[38], [15]. An information theoretic treatment of the case where multiple antennas are used in a Rayleigh fading channel, but neither the transmitter not receiver knows the channel state information appears in [33].

However, exploiting time diversity is harder than frequency, space, etc., because of the causal nature of the channel state information. It is not possible to know the future channel gains, unlike frequency, space, etc. where all the gains can be estimated simultaneously. This is not a problem when ergodic capacity is considered, because the optimal scheme (time-waterfilling) is memoryless. But, when delay constraints are introduced in the problem, the optimal scheme is not necessarily memoryless. [46] discusses the capacity versus outage probability obtained when the transmitter does not have the channel state information. In [6],[7] the K block delay constrained minimum outage problem is solved for the case where the K channel gains are known simultaneously (e.g. Orthogonal Frequency Division Multiplexing). The solution there is an intelligent variation of water-filling. In our case however, the optimal transmission scheme for exploiting time diversity needs to be an 'online' scheme, which makes decisions based solely on past and current channel state information. The delay limited nature of the transmission (K blocks) means that it may not be possible to see all possible channel fades, unlike [20]. This thesis explores the optimum strategies to exploit time diversity in a slow flat fading situation with a K-block delay constraint.

Other authors have dealt with the issue of delay-constrained capacity, but in limited or different contexts. [30], [31] impose a strict delay constraint of $K = 1$, and

solve the problem of minimizing the outage probability. In addition, they demonstrate the effect of peak power constraints. [10] takes a different approach in considering an average delay constraint with a peak power constraint. In their case, the transmitter attempts to clear a queue of data, and thus very directly relates to queueing theory ideas. However, the results have been obtained for the special case of the Gilbert-Elliot channel model [19]. This thesis discusses the delay constrained problem for a general constraint of K blocks (unlike [30], [31]). In particular, it focuses on the expected and outage capacity formulations. We seek to optimize transmission; specifically to maximize either the expected or the outage capacity. The optimization reduces to adapting the transmitted power as a function of the observed channel gain. Since the adaptation has to be a causal function of the channel gains (unlike [6],[7]) , hence dynamic programs are required because they 'provide an estimate of the future'.

The thesis is organized as follows. Chapter 2 defines the problem of delay constrained transmission, and solves it for a general cost function. The solution to the problem involves dynamic programming. Chapter 3 specializes the results to the expected capacity case. Expected capacity [11] considers as cost function the ensemble average of data rates for a fading channel. Analytic results are derived in the chapter, for the case of low signalling power. Chapter 4 solves the minimum outage probability problem. Outage probability is a capacity notion which is commonly used for fading channels. It is equivalent to considering the cumulative distribution function of the capacity random variable. In the area of coding for fading channels [56],[15],[5],[58], codes are designed with the implicit cost function of outage probability. In this chapter, the optimal transmission strategy is presented, and random coding bounds are derived for the strategy. Sections 4.2 and 4.3 in this chapter are joint work with Moses Charikar. Chapter 5 describes the related problem of finding 'stationary' policies for

the delay constraint problem. Transmission strategies are obtained that are *time-invariant* in their specification. Chapter 6 describes an example of the use of outage probability. The problem of optimum transmission in an outdoor mobile channel is considered. The work in this chapter is joint work with Ardavan Maleki. Finally, in Chapter 7, we switch gears and describe the problem of symbol synchronization in Orthogonal Frequency Division Multiplexed (OFDM) transmission. OFDM is a transmission scheme that allows optimization of the transmitted frequency spectrum. Whereas most of the previous chapters consider a flat fading channel (i.e. no inter-symbol interference), this chapter illustrates some of the complications that arise due to dispersive channels. In particular, channel dispersion makes recovery of the OFDM symbol (frame) clock tricky. We present an algorithm that asymptotically guarantees correct recovery of the clock. The main conclusions of the thesis are summarized in Chapter 8.

Chapter 2

Dynamic Programming for Delay Constrained Channels

2.1 Introduction

Ergodic capacity is a single deterministic scalar that is the upper bound on the maximum error-free data rate that can be transmitted over a fading channel. The assumptions required are that the channel fading process is ergodic, and the codeword is long enough to capture the ergodicity of the fading process. However, when delay constraint is introduced, the maximum error-free data rate that can be supported becomes a random variable, that depends on the sequence of channel gains seen for the duration of the codeword. Therefore, various measures of this random variable have been proposed. In particular, 'expected capacity' and 'outage capacity' are often considered by researchers, as being measures of interest. In this chapter, we propose a general transmission problem, with an arbitrarily defined measure of capacity, that captures the general 'utility function' that one expects for the end-user. This general

measure can be specialized to yield the outage and expected capacity measures along with other less common measures. We then solve the problem of optimum transmission under the delay constraint, for fading channels, using this general measure. Optimum transmission will refer to maximizing the chosen utility function of capacity. The delay constraint will be a constraint on the length of the codewords, defined in terms of the channel fading rate. Practically, such a constraint would arise due to the requirements of the application, such as voice transmission.

Section 2.1.2 formulates the delay constrained optimization problem formally. The formulation assumes that the channel gains are fed back to the transmitter causally. In Section 2.2, we show that this optimization problem includes as its solution the more general problem of output feedback. Section 2.3 then solves the problem using a dynamic programming approach. Several good books exist on dynamic programming, and the reader is referred to any of those for an explanation of the subject. See for example [23]. This thesis utilizes dynamic programming to a large extent, and borrows the concept of 'utility function' from it. In Section 2.4, we show that for a concave, non-decreasing utility function, if the transmitter is not allowed to adapt the powers based on the channel gains, then the best strategy is to simply transmit constant power. This result is useful to provide a comparison with the optimal strategy derived in Section 2.3. The result however does not hold for a general utility function.

2.1.1 Fading Channels with Delay Constraint

We assume a flat fading channel in AWGN. Further, it is assumed that the channel is i.i.d. block-fading [4]. This implies that the channel power gain g_i for block i is a scalar constant over that block consisting of a large number (T_0) of symbols (slow

fading). Also, the gains g_i are i.i.d. with distribution $f_G(g)$. This situation is valid for example in the case of slow frequency-hopped TDMA. In general however, there will be correlation between the blocks. If this correlation is known, the algorithms derived in this chapter can be easily extended by using a state space model of the fading process. However, this thesis will not cover the case of correlated blocks. Without loss of generality, we can absorb the noise power N_0B (which is constant) into the channel gain g_i. Thus, if the power transmitted is P, the received SNR over that block is Pg_i. This situation covers the case of narrowband transmission (possibly using antenna arrays), or even broadband transmission such as wideband CDMA or multicarrier CDMA, where the RAKE structure and tone structure, respectively, allows one to apply the theory developed for flat fading channels. For example, in the case of wideband CDMA, the receiver processes the incoming signal using a RAKE. The channel then looks like a flat (i.e. narrowband) fading channel as far as the symbols are concerned, even though it is a multipath (i.e. wideband) channel at the chip level.

The data transmission system is shown in Figure 2.1. It is assumed that the average power per transmitted symbol is P_0. Both the receiver and transmitter are assumed to have perfect knowledge of the channel gain of the past and current blocks, the latter by use of a feedback channel. However, they obviously do not have any knowledge of future gains, due to the i.i.d. assumption. On this system, we impose a 'K-block' delay constraint, i.e. we require that data transmission occur in groups of K blocks. Within each group of K blocks, the transmitter is allowed to distribute the total power KP_0T_0 as it pleases. Note that on the one extreme, we could require equal power transmission for every block ($K = 1$). On the other extreme, we could have a scheme that adapts power over an infinite number blocks ($K \to \infty$), which

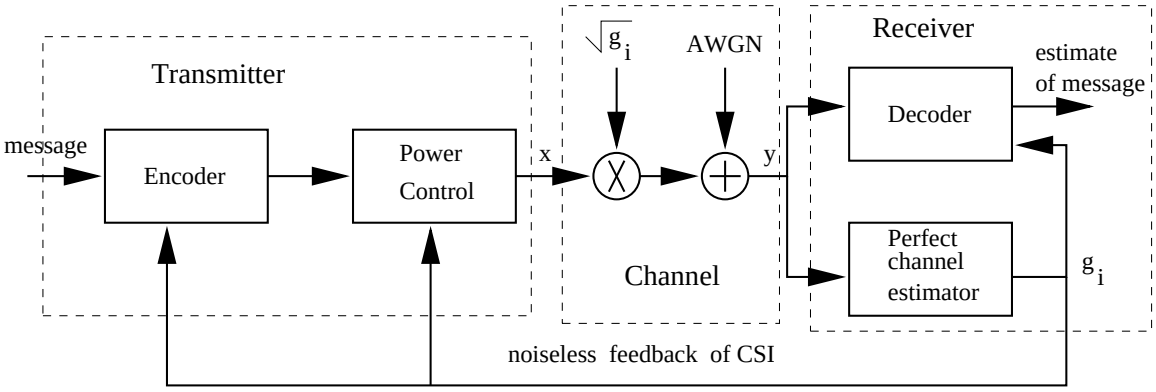

Figure 2.1: Transmission system for flat block-fading channel

provides the maximum flexibility. The assumption throughout is that T_0 is large, whereas K is not. Thus limiting arguments (in T_0) can be used to obtain information theoretic results. On the other hand, the small K means that ergodicity arguments cannot be used for capacity.

We require that the system achieve an optimal performance, defined in terms of a rate measure $\mu(x)$ (see Section 7.2.4). The K block constraint is a practical trade-off for systems that have to be better optimized, but cannot tolerate large latencies. In particular, this chapter includes the expected capacity and the outage probability formulations.

2.1.2 Problem Formulation

Given the channel model of Section 2.1.1, the most general question that could be asked is to design a codebook (a set of messages) that maximizes a rate measure over K blocks, under the total power constraint [1] of KP_0T, with decoding error probability $P_e \to 0$. Clearly, the codeword transmitted will be a function of the channel gains

[1] without loss of generality, we set $T = 1$ in the total power constraint, with the understanding that this gives the per-K-symbol power budget

that are seen. However, for a given average power level in a block, the Gaussian codebook achieves capacity in that block [12]. Therefore, the optimum codebook will be a Gaussian codebook even though the future channel gains are unknown, with the power of transmission varied optimally in the different blocks. Such a power adaptation strategy is called an 'online' strategy because of the causal nature of channel state information. The problem solved in Section 2.3 derives the best online power adaptation, so as to maximize an appropriate rate measure.

Mathematically, given the sequence of K channel power gains $\{g_i, i = 1, \ldots, K\}$ for the K blocks respectively, and power P_0 per transmission

$$\text{maximize} \quad \mathbf{E}[\ \mu(\ \sum_{i=1}^{K} \log(1 + P_i(g^{(i)})g_i)\ )\ ] \quad \text{such that} \tag{2.1}$$

$$P_i(g^{(i)}) \geq 0 \qquad \forall\ i = 1, \ldots, K \qquad \text{and either} \tag{2.2}$$

$$\sum_{i=1}^{K} P_i(g^{(i)}) \leq KP_0 \qquad \text{'Short term power constraint', or} \tag{2.3}$$

$$\mathbf{E}[\sum_{i=1}^{K} P_i(g^{(i)})] \leq KP_0 \qquad \text{'Long term power constraint'} \tag{2.4}$$

where the expectation[2] is over the channel gains $\{g_i, i = 1, \ldots, K\}$, and the maximization is over all online power adaptation strategies $\{P_i(g^{(i)}), i = 1, \ldots, K\}$. Note that transmitted power P_i for the ith block is a function only of the causal channel gains $\{g_1, \ldots, g_i\}$.

The function $\mu(x)$ is chosen based on the practical requirement of the application. In particular, we can choose $\mu(x)$ to solve the delay-constrained expected capacity

[2]Refer to the notations defined before Chapter 1.

problem [40], or the delay-constrained outage capacity problem [39] as follows:

$$\mu(x) = x \qquad \text{solves the } \textit{expected capacity} \text{ problem} \tag{2.5}$$

$$\mu(x) = \mathbf{1}_F(x \geq KR_0) \qquad \text{solves the } \textit{outage capacity} \text{ problem} \tag{2.6}$$

Expected and outage capacities have been defined in Chapter 1. With the expected capacity formulation, one attempts to maximize the expected capacity for the K blocks, while in the outage capacity formulation, one attempts to minimize the probability that the capacity for the K blocks does not meet the target rate of KR_0.

Figure 2.2 shows some of the possible $\mu(x)$ functions. For example, an application that requires a specific data rate 'approximately' may be satisfied with the function $\mu_3(x)$. $\mu_4(x) = \sqrt{x}$ is often used as a utility function in economic planning.

The short term constraint imposes a stricter power constraint than the long term constraint. In the short term constraint, one is allowed to distribute the total power KP_0 among the K blocks optimally. Each sequence of K blocks is allotted exactly power KP_0. However, in the long term constraint, the power allotted to a sequence of K blocks is KP_0 only on the average. Sequences may be allotted higher or lower power than that, as long as the average power constraint is met. Thus, the short term constraint implicitly includes a peak power constraint of KP_0. The choice of which constraint to use depends on the design of the transmitter.

2.2 Capacity of feedback channels

The most general case of feedback would be feedback of the channel gains $\{g^{(K)}\}$ and the channel output $\{y_t, t = 1, 2, \ldots KT_0\}$. One could ask how the solution to the problem posed in Section 2.1.2 differs from the solution to the above feedback

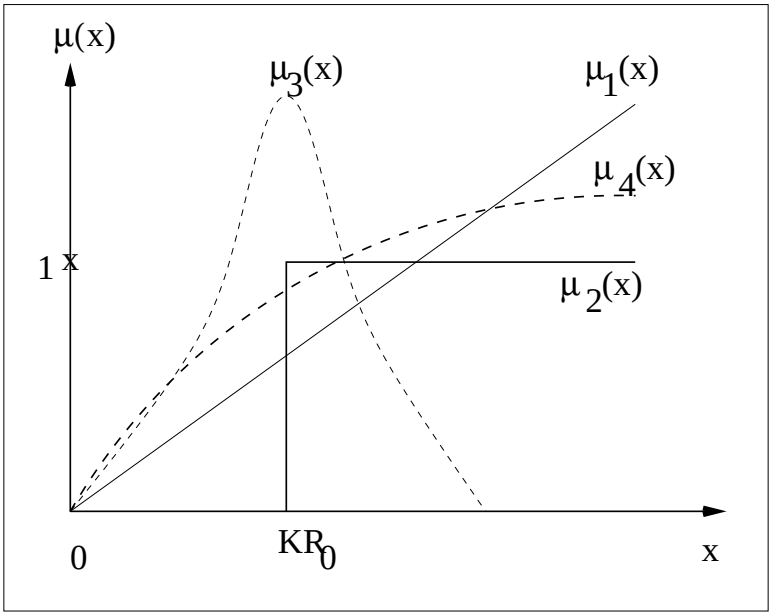

Figure 2.2: Various possible utility function $\mu(x)$s. $\mu_1(x)$ results in expected capacity, while $\mu_2(x)$ results in outage capacity

case. In this section we show that channel capacity is not increased by this general feedback over the case in Section 2.1.2 where only the channel gains $\{g^{(K)}\}$ are fed back. This is not surprising because for a fixed channel, it has been shown [12] that feedback does not increase capacity. The proof basically uses Fano's inequality [12].

Consider the transmission system that encodes message W into KT_0 symbols $x^{(KT_0)}$, which are sent over the fading channel that produces the output symbols $y^{(KT_0)}$. The receiver is assumed to feedback y_t without delay, and so the symbols x_t are the causal functions $x_t = x_t(W, y^{(t-1)})$. The receiver buffers all KT_0 symbols $y^{(KT_0)}$ and decodes them to estimate the message $\hat{W} = \hat{W}(y^{(KT_0)}, g^{(K)})$. The mutual information between W and $y^{(KT_0)}$ is $I(W; Y^{(KT_0)}|G^{(K)})$, where the conditioning on $G^{(K)}$ denotes that the receiver has complete knowledge of all the channel gains prior

to decoding the message. Now we bound the achievable rate R as below,

$$\begin{aligned} KT_0R &\leq H(W|g^{(K)}) \\ &= H(W|Y^{(KT_0)}, G^{(K)}) + I(W; Y^{(KT_0)}|G^{(K)}) \\ &\leq 1 + KT_0RP_{err} + I(W; Y^{(KT_0)}|G^{(K)}) \end{aligned} \tag{2.7}$$

where (2.7) arises due to Fano's inequality. P_{err} is the decoding error probability. Now we can bound the mutual information as,

$$\begin{aligned} I(W; Y^{(KT_0)}|G^{(K)}) &= h(Y^{(KT_0)}|G^{(K)}) - h(Y^{(KT_0)}|W, G^{(K)}) \\ &= h(Y^{(KT_0)}|G^{(K)}) - \sum_{t=1}^{KT_0} h(Y_t|W, G^{(K)}, Y^{(t-1)}) \\ &= h(Y^{(KT_0)}|G^{(K)}) - \sum_{t=1}^{KT_0} h(Y_t|X_t, G^{(K)}) \\ &\leq \sum_{t=1}^{KT_0} h(Y_t|G^{(K)}) - \sum_{t=1}^{KT_0} h(Y_t|X_t, G^{(K)}) \\ &= \sum_{t=1}^{KT_0} I(X_t; Y_t|G^{(K)}) \end{aligned} \tag{2.8}$$

(2.8) arises because X_t is a causal function $x_t = x_t(y^{(t-1)}, W)$. Therefore the data rate is bounded as,

$$\begin{aligned} R &\leq \frac{1}{KT_0} + RP_{err} + \frac{1}{KT_0} \sum_{t=1}^{KT_0} I(X_t; Y_t|G^{(K)}) \\ &= \frac{1}{KT_0} + RP_{err} + \frac{1}{K} \sum_{i=1}^{K} I(X_i; Y_i|G^{(K)}) \end{aligned} \tag{2.9}$$

(2.9) arises because of the block fading assumption. Now for any code that has decoding error probability $P_{err} \to 0$ as $T_0 \to \infty$, we have the bound,

$$R \leq \frac{1}{K} \sum_{i=1}^{K} I(X_i; Y_i | g^{(K)})$$

But R.H.S. in (2.10) can be achieved by codes designed to achieve capacity in a channel with only feedback of channel gains $G^{(K)}$ (and not necessarily feedback of channel outputs ($Y^{(KT_0)}$). In a random coding argument, this would require $X^{(KT_0)}$ to be chosen with the conditionally independent distribution $f_X(x^{(KT_0)}|G^{(K)}, y^{(KT_0)}) = \prod_{t=1}^{KT_0} f_X(x_t|g^{(([t/T_0]))})$. Thus, it is sufficient to feedback only the channel gains, and not the channel outputs. We emphasize again that R is the bound on data rate, and not a fixed target data rate, and P_{err} is the decoding error probability, and not the probability of outage.

2.3 Power Adaptation for Delay-Constrained Channels

The problem posed in Section 2.1.2 is certainly not trivial, because of its 'online' nature. Roughly speaking, on observing the channel gain of the current block, the transmitter must decide whether the gain is 'large enough' for it to be worth transmitting a large power, or whether it is better to conserve power and wait for some future block that *may* have a larger gain. Fortunately, if the statistics of the fades $f_G(g)$ is known, then optimum strategies can be designed.

Theorem 1 *The solution to the maximization problem (2.1) is the power adaptation functions* $\{P_i(g^{(i)}), Q_i^*(g_i, R, P)), i = 1, \ldots, K\}$, *which are calculated using algorithm*

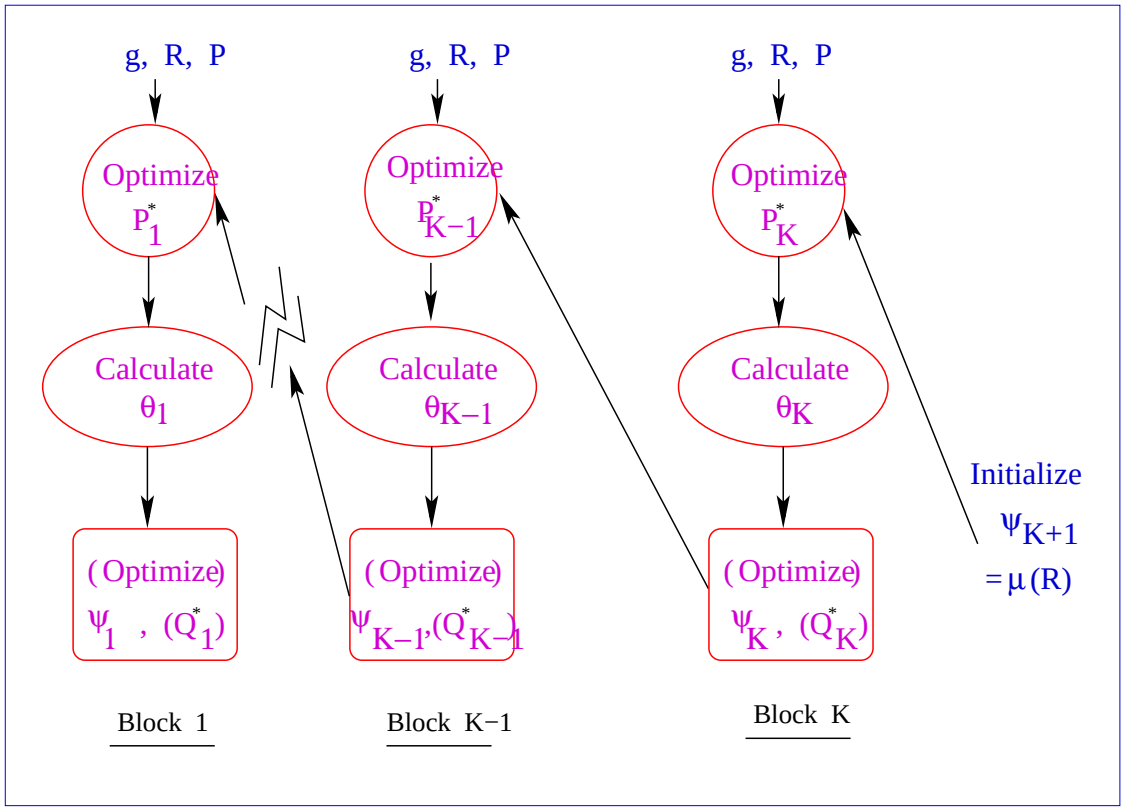

Figure 2.3: Dynamic program used to compute optimal power scheme

Power_adapt(K, P_0, R_0). *The maximum value of the reward function in (2.1) is* $\psi_1(0, KP_0)$.

Algorithm **Power_adapt**(K, P_0, R_0) : For $n = K, K-1, \ldots, 1$

At time n, choose

$$\begin{aligned} P_n^*(g_n, R, P) &= \underset{0 \le P_n \le P}{argmax} \quad \psi_{n+1}(R + \log(1 + P_n g_n), P - P_n) \qquad \text{and} \\ \theta_{n+1}(g_n, R, P) &= \psi_{n+1}(R + \log(1 + P_n^* g_n), P - P_n^*) \quad , \text{compute} \\ \psi_n(R, P) &= \underset{g_n}{\mathbf{E}}[\ \theta_{n+1}(g_n, R, P)\ ] \qquad \text{for } \textit{short-term constraint} \text{ , or} \end{aligned}$$

$$\psi_n(R,P) = max \underset{g_n}{\mathbf{E}}[\ \theta_{n+1}(g_n, R, \tilde{P}(g_n))] \quad \text{for } \textit{long-term constraint} \tag{2.10}$$
$$\psi_{K+1}(R,P) = \mu(R) \qquad \text{initialization}$$

where the maximization in (2.10) is over all functions $\tilde{P}(g_n) \geq 0$ such that $\mathbf{E}[\tilde{P}(g_n)] = P$. Call the maximizing function $\tilde{P}(g_n)$ for each (R,P) pair as $Q_n^*(g_n,R,P)$. This function is identically equal to P in case of the short term constraint algorithm.
In this manner, compute the *functions* $\{\psi_n(R,P), R \geq 0, P \geq 0\}$ by a backward recursion (i.e. proceeding $n = K, K-1, \ldots, 1$). In the process, we also obtain the three-dimensional function pairs $\{P_n^*(g_n,R,P), Q_n^*(g_n,R,P)),\ R \geq 0, P \geq 0\}$. These functions are pre-computed and stored in the system.

Now, when the system is used, then the optimal online power adaptation strategy chooses
For $n = 1, \ldots, K$

$$Q_n(g^{(n)}) = Q_n^*(g_n, R^{(n)}, P^{(n)}) \qquad \text{total power allotted to blocks } n \text{ to } K \tag{2.11}$$
$$P_n(g^{(n)}) = P_n^*(g_n, R^{(n)}, Q_n(g^{(n)})) \qquad \text{where the 'left-over' power} \tag{2.12}$$

$$P^{(n)} = Q_{n-1}(g^{(n-1)}) - P_{n-1}(g^{(n-1)}) \qquad \text{and the achieved rate} \tag{2.13}$$
$$R^{(n)} = \sum_{i=1}^{n-1} \log(1 + P_i(g^{(i)})g_i) \tag{2.14}$$

depend on the fades seen by the earlier blocks. Here, we define $P^{(1)} \equiv KP_0$ and $R^{(1)} \equiv 0$.

Proof: The algorithm is the optimal dynamic program solution [23] to the optimization problem. □

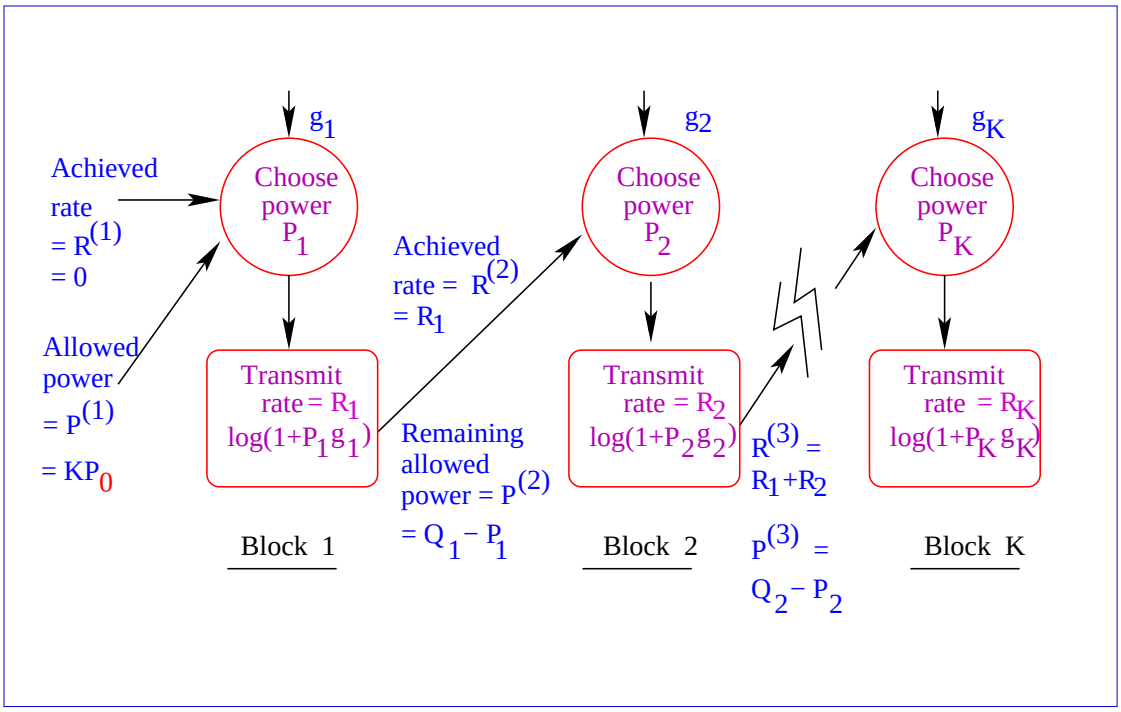

Figure 2.4: Optimal power functions $P_n^*(g, R, P), Q_n^*(g, R, P)$ used when the system is in operation

In the above algorithm, one pre-computes offline and stores the functions $\{P_n^*(g_n, R, P),\ Q_n^*(g_n, R, P)),\ R \geq 0, P \geq 0\}$. When the system is online, depending on the sequence of channel gains seen, one allots power $P_n(g^{(n)})$ to block n.

A key point to note is that whereas power selected for block i should depend on the past and present i gains $g^{(i)}$, it is sufficient to consider the state variables R, P, g for the dynamic program, to obtain the optimal solution. Thus, the complexity of the algorithm increases only linearly with number of blocks K, as opposed to exponentially, if all the past gains were to be explicitly used. Figure 2.3 shows a visual depiction of the dynamic program. It is initialized on the right side of the figure, and calculations progress from right to left (backwards). In the course of the calculations, we obtain the power functions $P_n^*(g, R, P), Q_n^*(g, R, P)$ (three-dimensional

functions), and the value functions $\psi_n(R, P)$ (two-dimensional functions). These functions are computed and stored in the transmission system. When the system is in operation, the power and rate for each block are computed using the power functions $P_n^*(g, R, P), Q_n^*(g, R, P)$ stored in the system, as shown in Figure 2.4. For example, the nth power P_n is selected from $P_n^*(g, R, P)$ by setting $g = g_n$, $P =$ 'Remaining allowed power' and $R =$ 'Achieved rate' coming into block n.

This algorithm can be specialized to the case of expected and outage capacities (among other measures) by choosing $\mu(x)$ appropriately. The next chapters will discuss this issue in detail, and also present simulation results to show the performance of the algorithm for the case of expected and outage capacities.

2.4 Equal power strategy

We would like to compare the solution in Section 2.3 to the best strategy that does not optimize transmission based on the channel gains. To this end, we look at the solution to the optimization problem; given the sequence of K channel power gains $\{g_i, i = 1, \ldots, K\}$ for the K blocks respectively,

$$\text{maximize} \quad \mathop{\mathbf{E}}_{g^{(K)},w}[\ \mu(\ \sum_{i=1}^{K} \log(1 + P_i(w)g_i)\ )\ ] \quad \text{such that} \tag{2.15}$$

$$P_i(w) \geq 0, \quad \forall\ i = 1, \ldots, K$$

$$\sum_{i=1}^{K} P_i(w) \leq KP_0 \quad \text{'Short term power constraint', or} \tag{2.16}$$

$$\mathop{\mathbf{E}}_{w}[\sum_{i=1}^{K} P_i(w)] \leq KP_0 \quad \text{'Long term power constraint'} \tag{2.17}$$

where the random variable w is independent of $\{g^{(K)}\}$. This means that the power control does not make use of the channel gains. We assume that the function $\mu(x)$ is concave and non-decreasing in x, which belongs to an important set of utility functions. We will show that the optimum strategy with the long term power constraint is to transmit equal power. i.e. $P_i(w) \equiv P_0$. Since this strategy also satisfies the short term constraint, this is also the optimum strategy for the short term constraint problem.

The proof basically uses Jensen's inequality. We first show that the function

$$C(P^{(K)}) \;=\; \mu(\sum_{i=1}^{K} \log(1+P_i g_i))$$

is concave in $\{P^{(K)}\}$. Look at the Hessian matrix $\nabla^2 C(P^{(K)}) = [(\nabla^2 C(P^{(K)}))_{ij}]$, where

$$\begin{aligned}(\nabla^2 C(P^{(K)}))_{ij} &= \frac{\partial^2 C(P^{(K)})}{\partial P_i \partial P_j} \\ &= \mu'(\sum_{i=1}^{K} \log(1+P_i g_i)) \frac{-g_i g_j \delta_{ij}}{(1+P_i g_i)(1+P_j g_j)} + \\ &\qquad \mu''(\sum_{i=1}^{K} \log(1+P_i g_i)) \frac{g_i g_j}{(1+P_i g_i)(1+P_j g_j)}\end{aligned}$$

Thus, we can write the Hessian as

$$\begin{aligned}\nabla^2 C(P^{(K)}) \;=\; & -\mu'(\sum_{i=1}^{K} \log(1+P_i g_i)) \;\cdot\; \mathbf{diag}\left(\frac{g_1^2}{(1+P_1 g_1)^2}, \ldots, \frac{g_K^2}{(1+P_K g_K)^2}\right) \;+ \\ & \mu''(\sum_{i=1}^{K} \log(1+P_i g_i)) \cdot \left(\frac{g_1^2}{(1+P_1 g_1)^2}, \ldots, \frac{g_K^2}{(1+P_K g_K)^2}\right)^T \cdot\end{aligned}$$

$$\left(\frac{g_1^2}{(1+P_1g_1)^2},\ldots,\frac{g_K^2}{(1+P_Kg_K)^2}\right)$$
$$\leq \quad 0$$

since both matrices are semi-negative for concave, non-decreasing $\mu(x)$. Thus, $C(P^{(K)})$ is concave in $\{P^{(K)}\}$.

Now, suppose the optimum solution to the above optimization problem is $P_i = P_i^*(w)$. Define $\mathbf{P}^* = \{P_1^*, \ldots, P_K^*\}$. The elements of $\mathbf{P}^*$ can be permuted using a permutation operator $\mathcal{H}$ to produce a new strategy $\mathcal{H}\mathbf{P}^*$. There are $n!$ such permutation operators, which form the set $S_{\mathcal{H}} = \{\mathcal{H}\}$. Due to symmetry, we can assert that

$$\underset{g^{(K)}}{\mathbf{E}}[\ \mu(\ \sum_{i=1}^{K}\log(1+P_i^*(w)g_i)\ )\ ] \quad = \quad \underset{g^{(K)}}{\mathbf{E}}[\ \mu(\ \sum_{i=1}^{K}\log(1+(\mathcal{H}\mathbf{P}^*)_ig_i)\ )\ ] \quad \forall\ \mathcal{H}\epsilon S_{\mathcal{H}}, w$$

Therefore,

$$\underset{g^{(K)},w}{\mathbf{E}}[\ \mu(\ \sum_{i=1}^{K}\log(1+P_i^*(w)g_i)\ )\ ] \quad = \quad \frac{1}{|S_{\mathcal{H}}|}\sum_{\mathcal{H}\epsilon S_{\mathcal{H}}}\left[\ \underset{g^{(K)},w}{\mathbf{E}}[\ \mu(\ \sum_{i=1}^{K}\log(1+(\mathcal{H}\mathbf{P}^*)_ig_i))]\right]$$

$$\leq \quad \underset{g^{(K)},w}{\mathbf{E}}[\ \mu\left(\ \sum_{i=1}^{K}\log\left(1+\frac{1}{|S_{\mathcal{H}}|}\sum_{\mathcal{H}\epsilon S_{\mathcal{H}}}[(\mathcal{H}\mathbf{P}^*)_i]g_i\right)\ \right)\ ] \qquad (2.18)$$

$$\leq \quad \underset{g^{(K)}}{\mathbf{E}}[\ \mu\left(\ \sum_{i=1}^{K}\log\left(1+\underset{w}{\mathbf{E}}[\ \frac{1}{|S_{\mathcal{H}}|}\sum_{\mathcal{H}\epsilon S_{\mathcal{H}}}[(\mathcal{H}\mathbf{P}^*)_i]\ ]g_i\right)\ \right)\ ] \qquad (2.19)$$

$$= \quad \underset{g^{(K)}}{\mathbf{E}}[\ \mu(\ \sum_{i=1}^{K}\log(1+P_0g_i)\ )\ ] \qquad (2.20)$$

(2.18),(2.19) use the concavity of $C(\cdot)$, (2.20) uses symmetry and the long term power constraint. Thus, the optimal solution to the problem (2.15) is bounded above by the equal power strategy. Therefore, we conclude that the equal power strategy

$P_i(w) \equiv P_0$ is indeed an optimum strategy for (2.15), in case of both the short term and the long term constraints. Again, we emphasize that this result has been proved only for the (useful) case of concave, non-decreasing $\mu(x)$.

Numerical results will be presented in the next two chapters to compare the performance of the optimum power adaptation algorithm of Section 2.3 and the constant power scheme.

Chapter 3

Expected Capacity

3.1 Introduction

The concept of expected capacity is the most direct extension of the standard ergodic capacity concept to fading channels. In the light of the general problem posed in Section 2.1.2, it can also be seen as a 'risk-neutral' optimization approach, since $\mu(x)$ is linear. Again, the problem for maximizing expected capacity is; given the sequence of K channel power gains $\{g_i, i = 1, \ldots, K\}$ for the K blocks respectively,

$$\text{maximize} \qquad \mathbf{E}[\sum_{i=1}^{K} \log(1 + P_i(g^{(i)})g_i)] \qquad \text{such that} \tag{3.1}$$

$$\begin{aligned} \sum_{i=1}^{K} P_i(g^{(i)}) &\leq KP_0 \qquad \text{(Short term power constraint), and} \\ P_i(g^{(i)}) &\geq 0, \qquad \forall\, i = 1, \ldots, K \end{aligned} \tag{3.2}$$

where the expectation is over the channel gains $\{g_i, i = 1, \ldots, K\}$, and the maximization is over all online power adaptation strategies $\{P_i(g^{(i)}), i = 1, \ldots, K\}$, which are

functions only of the causal channel gains. Only the short term constraint problem will be considered. Section 3.5 shows that the solution to the long term constraint problem is trivial.

Section 3.2 shows that the optimum power adaptation is specified by a two dimensional function $P_n(g, P)$. The case of low SNR (low values of P_0) is treated separately in Section 3.2.3 because it is interesting in its own right, and is also easier to analyze. Section 3.3 shows some simulation results for the algorithms. Section 3.4 derives some bounds for the performance of the optimum algorithms in the low SNR case. In Section 3.5, we show that the solution to the long term constraint problem in the expected capacity case is simply time waterfill.

3.2 Power Adaptation

3.2.1 Dynamic programming solution

Specializing the solution in Section 2.3, the optimum strategy is summarized by the following lemma.

Lemma 1 *The maximum expected capacity (for any SNR) for Gaussian codes over the K-block fading channel is achieved by the power adaptation functions $\{P_i(g^{(i)}), i = K, K-1, \ldots, 1\}$, which are calculated using algorithm* **Expected_adapt**(K, P_0). *Furthermore, the (maximized) expected capacity is $S_1(KP_0)/K$ per transmission.*

Algorithm **Expected_adapt**(K, P_0) : For $n = K, K-1, \ldots, 1$

At time n, choose

$$\begin{aligned} P_n^*(g_n, P) &= \underset{0 \le P_n(g_n,P) \le P}{argmax} [\log(1 + P_n(g_n, P)g_n) \\ &\quad + S_{n+1}(P - P_n(g_n, P))] \qquad \text{and} \qquad (3.3) \\ S_n(P) &= \underset{g_n}{\mathbf{E}}[\log(1 + P_n^*(g_n, P)g_n) \\ &\quad + S_{n+1}(P - P_n^*(g_n, P))] \qquad \text{with initialization} \qquad (3.4) \\ S_{K+1}(P) &\equiv 0 \end{aligned}$$

end

In this manner, compute the *functions* $\{S_n(P), 0 \le P \le KP_0\}$ by a backward recursion (i.e. proceeding $n = K, K-1, \ldots, 1$). In the process, we also obtain the two-dimensional functions $\{P_n^*(g_n, P), 0 \le P \le KP_0\}$.

Now, the optimal online power adaptation strategy chooses

For $n = 1, \ldots, K$

$$\begin{aligned} P_n(g^{(n)}) &= P_n^*(g_n, P^{(n)}) \qquad \text{where} \qquad (3.5) \\ P^{(n)} &= KP_0 - \sum_{i=1}^{n-1} P_i(g^{(i)}) \qquad (\text{'left-over' power}) \end{aligned}$$

depends on the fades seen by the earlier blocks. Here, we define $P^{(1)} \equiv KP_0$

end

This is the online version of waterfilling in time.

3.2.2 Comments

Note that the linearity of $\mu(x)$ makes the state variable R in the dynamic program described in Theorem 1 superfluous. Thus, the optimum power functions P_n^* reduce by one dimension. As $K \to \infty$, the online algorithm should converge to the 'time-waterfilling' algorithm described in [20].

The algorithm described in Section 3.2.1 is complicated because of the need to store K two-dimensional power adaptation functions $\{P_n^*(g_n, P), 0 \leq P \leq KP_0\}$. Therefore, it is interesting to look at an online algorithm that transmits all the power KP_0 in one block, just like the optimum low SNR algorithm described in Section 3.2.3. Section 3.3 will compare the performance of such an algorithm against the optimum.

3.2.3 Channels with low SNR

A special case of interest is the low SNR case. i.e. $SNR = KP_0\mathbf{E}[g] \ll 1$. For one, the optimum strategy is simpler for this case, and for another, the optimum strategy translates directly to the multiple access scenario when all users have low SNRs. Since the assumption $\log(1 + SNR_1 + SNR_2 + \cdots) \approx SNR_1 + SNR_2 + \cdots$ can be made in the low SNR case, hence the capacity maximization will hold due to linearity, even in the multiple access scenario.

For the low SNR case, Lemma 1 reduces to the following.

Lemma 2 *For low values of SNR, the maximum expected capacity for Gaussian codes over the K-block fading channel is achieved by the choosing the power adaptation functions* $\{P_i(g^{(i)}), i = K, K-1, \ldots, 1\}$*, which are calculated using algorithm* **Expected_low**(K, P_0)*. Furthermore, the (maximized) expected capacity is* S_1/K *per*

transmission.

Algorithm **Expected_low**(K, P_0) : For $n = K, K-1, \ldots, 1$

At time n, choose

$$P_n(g^{(n)}) = \begin{cases} KP_0 & \text{if } KP_0 g_n \geq S_{n+1} \\ 0 & \text{if } KP_0 g_n < S_{n+1} \end{cases} \tag{3.6}$$

If $P_n(g^{(n)}) = KP_0$, set $P_{n+1} = \cdots = 0$ and exit loop.

end

S_n is defined by the backward recursion

$$\begin{aligned} S_n - S_{n+1} &= \mathbf{P}[KP_0 g_n > S_{n+1}] \cdot \\ &\quad (\mathbf{E}[KP_0 g_n \mid KP_0 g_n > S_{n+1}] - S_{n+1}) \end{aligned} \tag{3.7}$$

$$S_{K+1} \equiv 0 \tag{3.8}$$

As an example, the performance of this strategy can be computed for the Rayleigh fading channel. In this case $f_G(g) = \frac{1}{g_0}\exp(-g/g_0)$, so that the online algorithm results in the following recursion for S_n

$$S_n - S_{n+1} = KP_0 g_0 \times \exp(-S_{n+1}/KP_0 g_0)$$

Approximate S_1 for large K by approximating the recursion by the equation

$$\begin{aligned} \frac{\partial S(x)}{\partial x} &= KP_0 g_0 \times \exp(-S(x)/KP_0 g_0) \\ S(0) &\equiv 0 \end{aligned}$$

which solves to $S(x) = KP_0g_0 \log(x+1)$. In the approximation, x was the time variable, and so the expected capacity is $S_1/K \approx S(K)/K = P_0g_0 \log(K+1)$. It is clear that the online power adaptation increases the capacity by a factor of approximately $\log(K+1)$, over the unknown channel state information case. Though the analysis was straight-forward in the Rayleigh fading case, it is seldom mathematically tractable in general. In Section 3.4, we show approximate bounds of performance in the important case of an χ^2_{2m} (Nakagami) fading channel.

3.2.4 Comments

In the low SNR case, we see that the online power adaptation basically picks just one block to transmit the entire power, similar to the waterfill algorithm [12], which picks just one subchannel to transmit on when the SNR is low. Since $S_n \geq S_{n+1}, \ \forall \ n$, hence the adaptation scheme keeps lowering the transmission thresholds for subsequent blocks. In practice, when K becomes large enough, the 'low SNR' assumption will break down, since the transmission will occur with high probability in a block with large gain.

It is interesting to compare the capacity gain obtained here to the case where the gains of all K blocks are available apriori. If selection diversity is used in those cases (i.e. transmitting only on the largest gain channel), then for i.i.d. Rayleigh fading on the channels, we see that the capacity gain over the equal-power case is again a factor of approximately $\log(K)$. It is interesting to observe that in our online algorithm, the causal nature of the channel state information at the transmitter does not significantly hurt the expected capacity! Section 3.4 shows that this result also holds for χ^2_{2m} distributed gains [48], in general.

On the other hand, the capacity distribution in the case of the online algorithm

is poor (high probability of taking low values) as compared to the apriori knowledge case. Thus, this is not a good algorithm if one intends to guarantee a certain outage probability. The results in [39] are relevant in that case, which will be described in Chapter 4.

3.3 Performance of the Online Algorithms

We demonstrate the capacity improvement afforded by the online power adaptation scheme over a constant power scheme ($K = 1$), by calculating the ratio of the respective capacities for various cases. In Figure 3.1, we have the case of low SNR (Section 3.2.3). The channels shown have g distributed as χ^2 random variables with degrees of freedom 2,4,8,16 (corresponding to Rayleigh fading, and Nakagami fading of diversity of 2,4,8 respectively). We plot the capacity ratio against the delay requirements (K). It is clear that for the Rayleigh fading channel, the capacity ratio is roughly $\log(1 + K)$, as promised. For higher degrees of diversity, the capacity ratio falls because the channel already has sufficient diversity.

Figure 3.2 shows the capacity ratio of the optimum algorithm of Section 3.2.1 over the constant power case against the delay requirements (K), for various values of SNR. A Rayleigh fading channel is assumed in this case (though a similar calculation could be made for any gain distribution). As the SNR increases, the capacity improvement decreases. This agrees with the results in [20] ($K \to \infty$). Figure 3.3 shows the performance of a one-block transmission scheme, similar to the one in Section 3.2.3 but without using the log approximation, used for larger SNRs. A Rayleigh fading channel is assumed. It is interesting to note that as K increases, the capacity ratio begins to drop, even falling below one! This shows that a waterfill-like solution

that assigns power to more than one block is essential in those cases. Finally, in Figure 3.4, we compare the optimum power adaptation strategy of Section 3.2.1 to the above one-block suboptimal strategy. In the range of SNRs and K where the two strategies agree well, one would prefer to use the simpler suboptimal strategy.

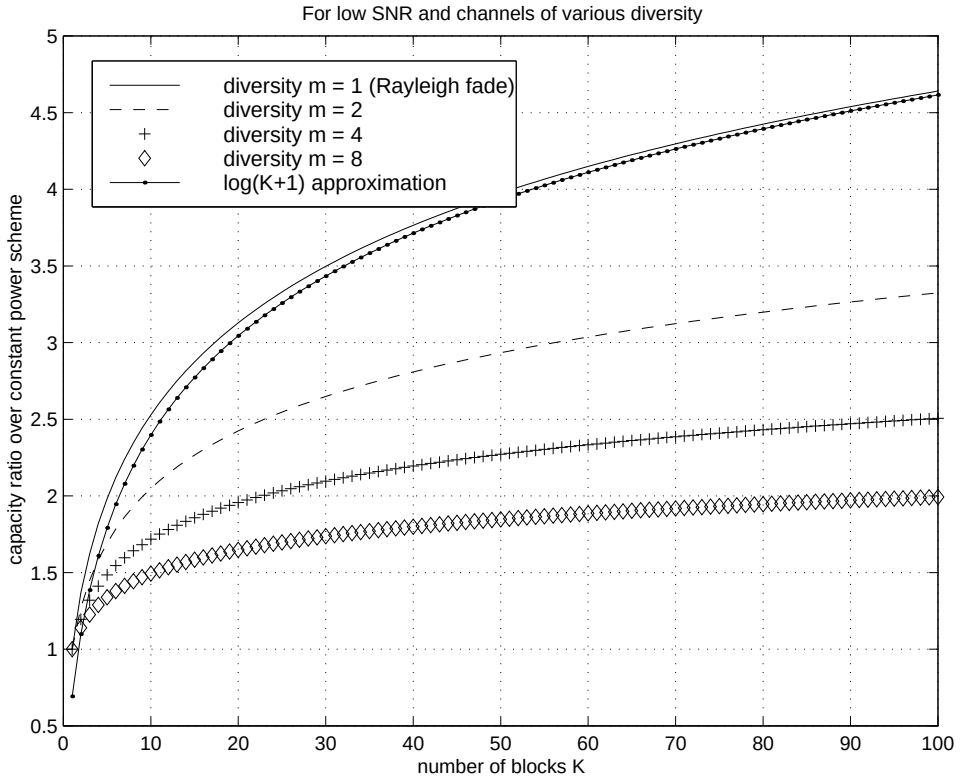

Figure 3.1: Capacity improvement ratio for low SNRs and various channel gain distributions

3.4 Low SNR bounds for expected capacity

The expected capacity case shows substantial gain in the low SNR regime. We will derive bounds to quantify this gain in capacity in the case of χ^2_{2m} distributed gains, which can represent a channel with diversity m, such as an m-multipath fading

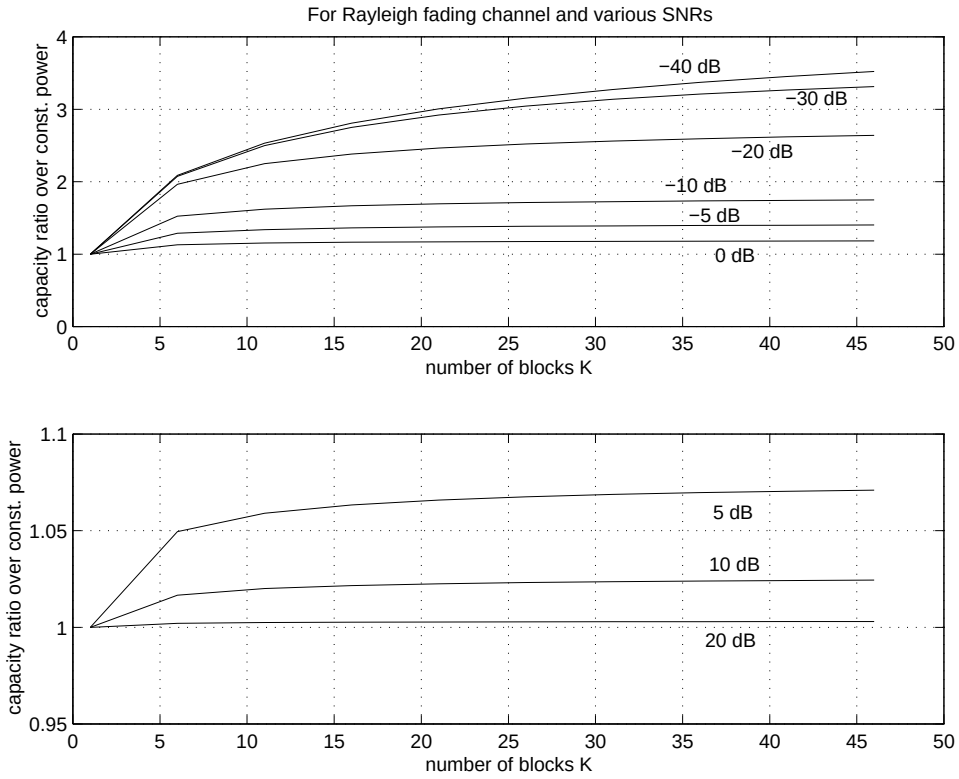

Figure 3.2: Capacity improvement ratio of optimum power adaptation in a Rayleigh fading channel for various SNRs

channel - this would arise for example, when using antenna arrays or in broadband CDMA after RAKE-combining. This channel has density [48],

$$f_G(g) = me^{-mg}\frac{(mg)^{m-1}}{(m-1)!} \tag{3.9}$$

$$F_G(g) = 1 - e^{-mg}\left[1 + (mg) + \cdots + \frac{(mg)^{m-2}}{(m-2)!} + \frac{(mg)^{m-1}}{(m-1)!}\right] \quad \text{so that} \tag{3.10}$$

$$\mathop{\mathbf{E}}_{g}[g] = 1 \tag{3.11}$$

Without loss of generality, the average channel gain $\mathop{\mathbf{E}}_{g}[g]$ has been normalized to one. We first derive a lower bound for the expected capacity, when the optimum power control strategy for K blocks (3.6) is used. For convenience, we relabel the blocks

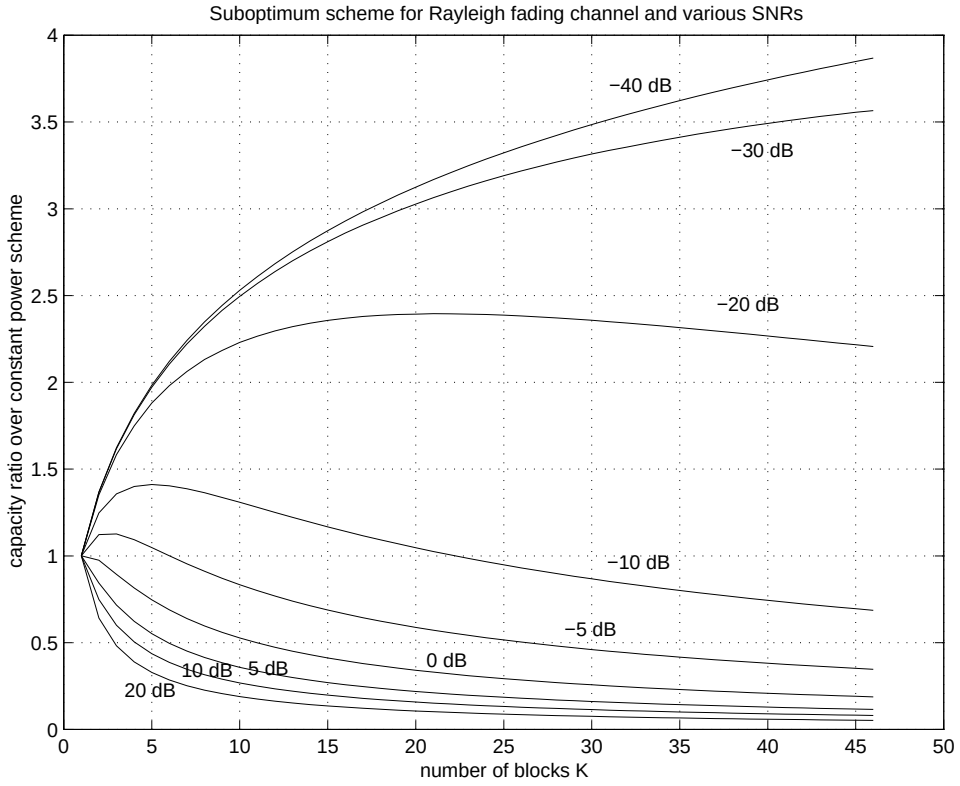

Figure 3.3: Capacity improvement ratio of suboptimum power adaptation in a Rayleigh fading channel for various SNRs

$1, 2, \ldots, K$ as $K, K-1, \ldots, 1$. Then the dynamic program to calculate capacity (3.8) is,

$$S_0 \equiv 0 \tag{3.12}$$

$$S_{n+1} - S_n = \mathbf{P}[KP_0 g_{n+1} > S_n] \cdot \tag{3.13}$$

$$(\mathbf{E}[KP_0 g_{n+1} \mid KP_0 g_{n+1} > S_n] - S_n)$$

$$= \int_{S_n/KP_0}^{\infty} KP_0 g f_G(g) \mathrm{d}g \; - \; S_n \int_{S_n/KP_0}^{\infty} f_G(g) \mathrm{d}g \tag{3.14}$$

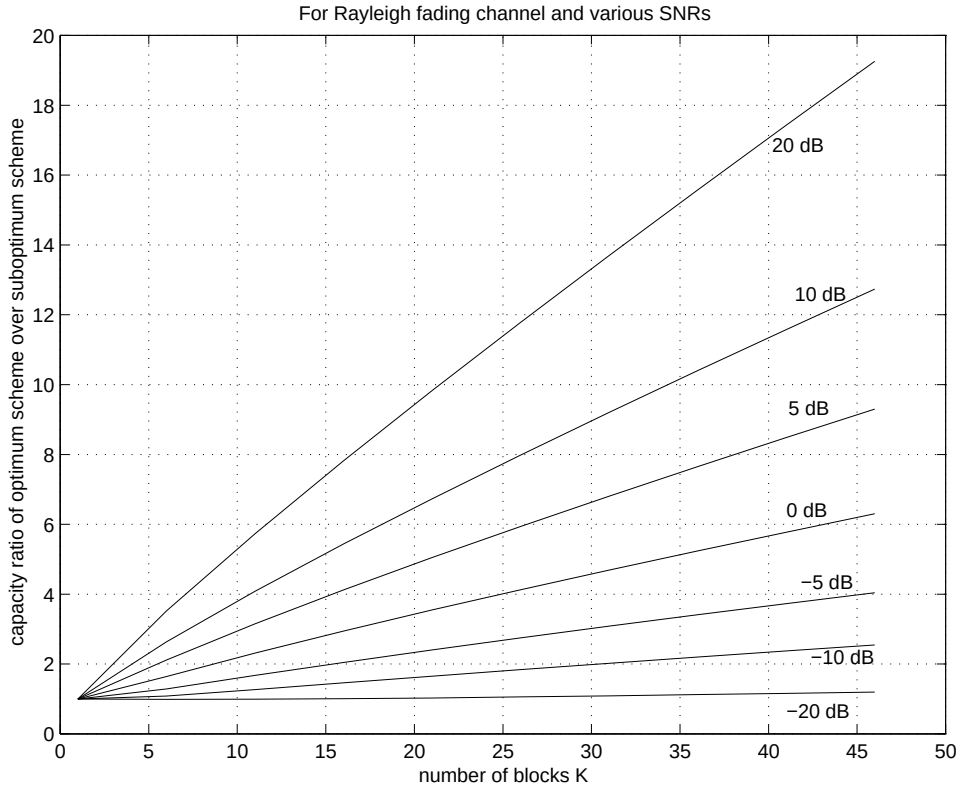

Figure 3.4: Capacity ratio of optimum power adaptation to suboptimum power adaptation in a Rayleigh fading channel for various SNRs

$$= KP_0 \int_{S_n/KP_0}^{\infty} (1 - F_G(g))\mathrm{d}g \tag{3.15}$$

$$= KP_0 G\left(\frac{S_n}{KP_0}\right) \qquad \text{where} \tag{3.16}$$

$$G(y) \doteq \int_{y}^{\infty} (1 - F_G(g))\mathrm{d}g \tag{3.17}$$

In particular, we look at χ^2_{2m} distributed gains (3.9), for which we calculate

$$G(y) = \frac{1}{m} e^{-my}\left[m + (m-1)(my) + \cdots + 2\frac{(my)^{m-2}}{(m-2)!} + \frac{(my)^{m-1}}{(m-1)!}\right]$$

$$\geq e^{-my} \qquad \forall \;\; y \geq 0$$

$$S_{n+1} - S_n \geq KP_0 e^{-mS_n/KP_0}$$

We compare this to the solution of the differential equation,

$$\begin{aligned} \frac{\partial P(x)}{\partial x} &= KP_0 e^{-mP(x)/KP_0} \\ P(0) &= 0 \qquad \text{which has the solution} \\ P(x) &= \frac{KP_0}{m}\log(1+mx) \end{aligned}$$

Since both S_n and $P(x)$ are always positive, it is clear that $P(n) \leq S_n, \quad \forall\ n \geq 0$. Therefore, we get the lower bound,

$$\begin{aligned} \frac{S_K}{K} &\geq \frac{P(K)}{K} \\ &= \frac{P_0}{m}\log(1+mK) \end{aligned}$$

This shows that the capacity per transmission roughly increases at least as fast as $\frac{1}{m}\log(K)$.

We now derive an upper bound on the optimum power control capacity, by bounding the capacity per transmission when apriori knowledge of all the K gains is available. Note again that in the low SNR case, if all the K gains were available apriori (rather than in a causal fashion, as in the optimum power control), then the optimum strategy chooses to transmit on the maximum gain only (selection diversity). Selection diversity therefore upper bounds the optimum power control capacity in the low SNR case. To bound the capacity obtained by selection diversity in a channel with χ^2_{2m} distributed gains, we proceed as below,

$$\begin{aligned} T_K &\doteq \mathop{\mathbf{E}}_{g^{(K)}}[\ KP_0 \max[g_1, g_2, \ldots, g_K]\ ] \\ V_{n+1} &\doteq \max[g_1, g_2, \ldots, g_{n+1}] \end{aligned}$$

$$= \max[g_{n+1}, V_n]$$

We first indicate how $\mathbf{E}[V_n]$ grows in comparison with S_n.

$$\begin{aligned}
\mathbf{E}[V_{n+1}] &= \mathop{\mathbf{E}}_{g_{n+1},V_n}[g_{n+1}\ ;\ g_{n+1} \geq V_n] + \mathop{\mathbf{E}}_{V_n}[V_n\ \mathbf{P}[g_{n+1} < V_n | V_n]] \\
&= \mathop{\mathbf{E}}_{V_n}[V_n(1 - F_G(V_n)) + \int_{V_n}^{\infty} (1 - F_G(g))\mathrm{d}g] + \mathop{\mathbf{E}}_{V_n}[V_n F_G(V_n)] \\
\mathbf{E}[V_{n+1}] - \mathbf{E}[V_n] &= \mathop{\mathbf{E}}_{V_n}[G(V_n)] \qquad \text{where } G(y) \text{ is as in (3.17)}
\end{aligned}$$

It can be directly seen that $G(y)$ is a non-increasing, convex function with $G(0) = \mathbf{E}[g] = 1$ and $G(\infty) = 0$. Thus, by Jensen's inequality,

$$\begin{aligned}
\mathbf{E}[V_{n+1}] - \mathbf{E}[V_n] &\geq G(\mathop{\mathbf{E}}_{V_n}[V_n]) \\
\mathbf{E}[V_0] &= 0
\end{aligned}$$

Compare this to (3.16),

$$\begin{aligned}
\frac{S_{n+1}}{KP_0} - \frac{S_n}{KP_0} &= G\left(\frac{S_n}{KP_0}\right) \\
\frac{S_0}{KP_0} &= 0
\end{aligned}$$

As n becomes large, we can expect the relation in (3.18) to be close to an equality. Then, since $\frac{S_n}{KP_0} \leq \mathop{\mathbf{E}}_{V_n}[V_n]$ which implies that $G(\frac{S_n}{KP_0}) \geq G(\mathop{\mathbf{E}}_{V_n}[V_n])$, we can expect $\frac{S_n}{KP_0}$ to 'catch-up' with $\mathop{\mathbf{E}}_{V_n}[V_n]$ as n increases. To make this statement more precise, we continue with the upper bound, assuming now that the channel gains are χ^2_{2m}

distributed.

$$\begin{aligned}
w_i &\doteq mg_i \\
y_i &\doteq e^{(\sqrt{w_i}-1)^2} \\
w_i &= w(y_i) = \left(1+\sqrt{\log y_i}\right)^2 \qquad \text{so that} \\
f_W(w_i) &= e^{-w_i}\frac{w_i^{m-1}}{(m-1)!} \\
\mathbf{E}[y_i] &= \underset{w_i}{\mathbf{E}}[e^{(\sqrt{w_i}-1)^2}] \\
&= \int_0^\infty e^{(w-2\sqrt{w}+1)}e^{-w}\frac{w^{m-1}}{(m-1)!}\mathrm{d}w \\
&= \frac{2e}{(m-1)!}\frac{(2m-1)!}{2^{2m}}
\end{aligned}$$

$w(y_i)$ is a concave function of y. Now, we bound

$$\begin{aligned}
\max[w_1, w_2, \ldots, w_K] &\leq \left(1+\sqrt{\log(\sum_{i=1}^{K} y_i)}\right)^2 \\
\mathbf{E}[\ \max[w_1, w_2, \ldots, w_K]\ ] &\leq \left(1+\sqrt{\log(\mathbf{E}[\sum_{i=1}^{K} y_i])}\right)^2 \qquad \text{because } w(y) \text{ is concave} \\
\frac{T_K}{K} &= \frac{P_0}{m}\mathbf{E}[\ \max[w_1, w_2, \ldots, w_K]\ ] \\
&\leq \frac{P_0}{m}\left(1+\sqrt{\log(K)+\log\left(\frac{2e}{(m-1)!}\frac{(2m-1)!}{2^{2m}}\right)}\right)^2
\end{aligned}$$

Since $\frac{S_K}{K} \leq \frac{T_K}{K}$, hence we have the upper bound,

$$\begin{aligned}
\frac{S_K}{K} &\leq \frac{P_0}{m}\mathbf{E}[\ \max[w_1, w_2, \ldots, w_K]\ ] &(3.18)\\
&\leq \frac{P_0}{m}\left(1+\sqrt{\log(K)+\log\left(\frac{2e}{(m-1)!}\frac{(2m-1)!}{2^{2m}}\right)}\right)^2 &(3.19)
\end{aligned}$$

(3.18) and (3.19) together show that the expected capacity increases approximately as $\frac{1}{m}\log(K)$. In fact, for a Rayleigh fading channel ($m = 1$), we can produce an even tighter upper bound, because $\frac{T_K}{K}$ can be evaluated exactly as

$$\begin{aligned}
\frac{T_K}{K} &= P_0\mathbf{E}[\ \max[g_1, g_2, \ldots, g_K]\ ] && (3.20)\\
&= P_0\int_0^\infty (1 - F_G^K(g))\mathrm{d}g && (3.21)\\
&= P_0\int_0^\infty e^{-g}(F_G^{K-1}(g) + F_G^{K-2}(g) + \cdots + 1)\mathrm{d}g && (3.22)\\
&= P_0\int_0^\infty (F_G^{K-1}(g) + F_G^{K-2}(g) + \cdots + 1)\mathrm{d}F_G(g) && (3.23)\\
&= P_0 \cdot \left(1 + \frac{1}{2} + \cdots \frac{1}{K}\right) && (3.24)\\
&\leq P_0 \cdot (1 + \log K) && (3.25)
\end{aligned}$$

Thus, for a Rayleigh fading channel, we get the tight bounds,

$$P_0\log(1+K) \quad \leq \quad \frac{S_K}{K} \quad \leq \quad P_0(1+\log K)$$

This shows that for a Rayleigh fading channel, causality causes negligible loss with respect to the optimum acausal strategy (selection diversity). For higher diversity channels, the bounds are not as tight ((3.18), (3.19)), but they're sufficiently close.

3.5 Long term constraint problem

We consider the solution to the optimization problem (2.1) with the long term power constraint (2.4), in the expected capacity case. We will show that the optimum power control strategy for this case reduces to choosing $P_i = P_i(g_i)$. i.e. power depends only on current channel gain. In fact, the optimum solution is time waterfilling. To prove

this, we upper bound the capacity for the case that all K channel gains are known apriori (the 'acausal capacity'). This capacity in turn upper bounds the capacity in (3.1).

Therefore, we will upper bound the capacity,

$$\text{maximize} \quad \mathop{\mathbf{E}}_{g^{(K)}}[\ \sum_{i=1}^{K} \log(1 + P_i(g^{(K)})g_i)\ ] \quad \text{such that} \tag{3.26}$$

$$\begin{aligned} P_i(g^{(K)}) &\geq 0, \qquad \forall\ i = 1, \ldots, K \qquad \text{and} \\ \mathop{\mathbf{E}}_{g^{(K)}}[\sum_{i=1}^{K} P_i(g^{(K)})] &\leq KP_0 \qquad \text{Long term constraint} \end{aligned} \tag{3.27}$$

Note the dependence of the power on all of the gains $g^{(K)}$. Let the optimum power control strategy for this case be $\{P_i = P_i^*(g^{(K)})\}$. Define a new strategy,

$$Q_i(g_i) = \mathop{\mathbf{E}}_{\{g_j, j \neq i\}}[P_i^*(g^{(K)})] \tag{3.28}$$

Then we bound the capacity as,

$$\begin{aligned} \text{maximum} \quad \mathop{\mathbf{E}}_{g^{(K)}}[\ \sum_{i=1}^{K} \log(1 + P_i(g^{(K)})g_i)\ ] &= \mathop{\mathbf{E}}_{g^{(K)}}[\ \sum_{i=1}^{K} \log(1 + P_i^*(g^{(K)})g_i)\ ] \\ &= \sum_{i=1}^{K} \mathop{\mathbf{E}}_{g_i}[\mathop{\mathbf{E}}_{\{g_j, j \neq i\}}[\log(1 + P_i^*(g^{(K)})g_i)]] \\ &\leq \sum_{i=1}^{K} \mathop{\mathbf{E}}_{g_i}[\log(1 + \mathop{\mathbf{E}}_{\{g_j, j \neq i\}}[P_i^*(g^{(K)})]g_i)] \\ &= \sum_{i=1}^{K} \mathop{\mathbf{E}}_{g_i}[\log(1 + Q_i(g_i)g_i)] \end{aligned} \tag{3.29}$$

Thus, it is clear that the optimum strategy is of the form $P_i = Q_i(g_i)$. Continuing, we can see that the optimum $Q_i(g_i)$ is obtained by time waterfilling,

$$\begin{aligned} Q_i^*(g_i) &= \left[K_i - \frac{1}{g_i}\right]^+ \qquad \text{such that} && (3.30)\\ \mathop{\mathbf{E}}_{g_i}[Q_i^*(g_i)] &= \beta_i && (3.31)\\ \sum_{i=1}^{K} \beta_i &\leq KP_0 && (3.32) \end{aligned}$$

Here $[x]^+ \doteq \max[x, 0]$, and $\beta_i \geq 0, \forall\ i$. All that remains is to find the best set of $\{\beta_i\}$ to maximize the upper bound. This can be obtained by noting that that the function $U(\beta) = \mathop{\mathbf{E}}_{g}[\ \log(1 + Q(g)g)\ ]$ is concave in β.

Lemma 3 *The function* $U(\beta) = \mathop{\mathbf{E}}_{g}[\ \log(1 + Q(g)g)\ ]$, *where* $Q(g) = [K - \frac{1}{g}]^+$ *such that* $\mathop{\mathbf{E}}_{g}[Q(g)] = \beta$, *is concave in* $\beta \geq 0$.

Proof: Let $0 \leq \lambda \leq 1$, $\overline{\lambda} = 1 - \lambda$, $\beta_1, \beta_2 \geq 0$. Then,

$$\begin{aligned} \lambda U(\beta_1) + \overline{\lambda} U(\beta_2) &= \lambda \mathop{\mathbf{E}}_{g}[\ \log(1 + Q_1(g)g)\ ] + \overline{\lambda} \mathop{\mathbf{E}}_{g}[\ \log(1 + Q_2(g)g)\ ] \\ &\leq \mathop{\mathbf{E}}_{g}[\log(1 + (\lambda Q_1(g) + \overline{\lambda} Q_2(g))g)] \text{and} && (3.33)\\ \mathop{\mathbf{E}}_{g}[\lambda Q_1(g) + \overline{\lambda} Q_2(g)] &= \lambda\beta_1 + \overline{\lambda}\beta_2 \qquad \text{So,} \\ \lambda U(\beta_1) + \overline{\lambda} U(\beta_2) &\leq U(\lambda\beta_1 + \overline{\lambda}\beta_2) && (3.34) \end{aligned}$$

(3.34) holds because (3.33) is maximized when $\lambda Q_1(g) + \overline{\lambda} Q_2(g)$ is the waterfill solution. □

Thus, continuing from (3.29),

$$\text{maximum} \quad \mathop{\mathbf{E}}_{g^{(K)}}[\ \sum_{i=1}^{K} \log(1 + P_i(g^{(K)})g_i)\ ] \quad \leq \quad \sum_{i=1}^{K} U(\beta_i)$$

$$\begin{aligned} &= K\frac{1}{K}\sum_{i=1}^{K} U(\beta_i) \\ &\leq KU\left(\frac{1}{K}\sum_{i=1}^{K}\beta_i\right) \qquad \text{due to concavity of } U(\beta) \\ &\leq KU(P_0) \qquad \text{due to (3.32)} \end{aligned} \tag{3.35}$$

The R.H.S. in (3.35) is achieved when $P_i(g^{(K)})$ is chosen as the i.i.d. waterfill strategy,

$$\begin{aligned} P_i^*(g^{(K)}) &= P_i^*(g_i) \\ &= \left[K - \frac{1}{g_i}\right]^+ \qquad \text{such that} \\ \mathop{\mathbf{E}}_{g_i}[P_i^*(g_i)] &= P_0 \end{aligned}$$

But since this is a memoryless strategy, hence even the casual system (3.1) (with the long term constraint) can achieve this bound. Therefore the optimum solution to (3.1) with the long term constraint is the memoryless strategy (3.36).

Chapter 4

Outage Capacity

4.1 Introduction

The concept of outage capacity is useful for systems that require a guaranteed data rate. For example, in voice communications, unless a raw data rate of 8 kilobits/sec is communicated (perhaps a little lower with better source coding), it may be difficult to understand the message. However, a delay of around 20 millisec is considered tolerable in wireless voice transmission. If the 20 millisec long voice packet is corrupted, it has to be dropped, a phenomenon called 'outage'. Clearly, too many outages will result in unintelligible speech. Hence, there is a need to minimize the probability of outage at a target data rate of 8 kilobits/sec (in the case of voice). Suppose the minimum probability of outage that can be achieved when transmitting at a rate of 8 kilobits/sec is P_{out} (clearly, P_{out} will be an increasing function of transmission rate). We say then that the outage capacity is 8 kilobits/sec, at the outage probability P_{out}. Thus in general, outage capacity is R_0 at an error probability of P_{out}. The function $P_{out}(R_0)$ then represents the cumulative distribution function of the capacity random variable.

Note that if a data rate of R_0 is transmitted, then the decoding error probability is,

$$\begin{aligned} P_{err} &= \mathbf{P}[\text{Decoding error, given } I(X,Y|G) \geq R_0] \cdot \mathbf{P}[I(X,Y|G) \geq R_0] \; + \\ &\quad \mathbf{P}[\text{Decoding error, given } I(X,Y|G) < R_0] \cdot \mathbf{P}[I(X,Y|G) < R_0] \end{aligned}$$

If capacity achieving codes can be used, then $\mathbf{P}[\text{Decoding error, given } I(X,Y|G) \geq R_0] \rightarrow 0$, while $\mathbf{P}[\text{Decoding error, given } I(X,Y|G) < R_0] \rightarrow 1$. This shows that $P_{err} \rightarrow P_{out}$ in the limit.

Mathematically, the outage capacity problem is; given the sequence of K channel power gains $\{g_i, i = 1, \ldots, K\}$ for the K blocks respectively, and given the target rate R_0 per transmission, and power P_0 per transmission

$$\begin{aligned} \text{minimize} \qquad & \mathbf{P}[\sum_{i=1}^{K} \log(1 + P_i(g^{(i)})g_i) < KR_0] \\ P_i(g^{(i)}) &\geq 0 \qquad \forall\, i = 1, \ldots, K \qquad \text{and either} \\ \sum_{i=1}^{K} P_i(g^{(i)}) &\leq KP_0 \qquad \text{'Short term',} \quad \text{or} \qquad (4.1) \\ \mathbf{E}[\sum_{i=1}^{K} P_i(g^{(i)})] &\leq KP_0 \qquad \text{'Long term'} \qquad (4.2) \end{aligned}$$

where the expectation is over the channel gains $\{g_i, i = 1, \ldots, K\}$, and the minimization is over all online power adaptation strategies $\{P_i(g^{(i)}), i = 1, \ldots, K\}$. Note that transmitted power P_i for the ith block is a function only of the causal channel gains $\{g_1, \ldots, g_i\}$. The cost function $\mathbf{P}[(\cdot)]$ is the probability of outage at the target transmission rate of R_0 (KR_0 for K blocks).

Section 4.2 describes the solution based on dynamic programming. The notation

used is slightly different from that in Section 2.3, to make it more intuitive in the context of outage capacity. Section 4.3 shows some simulation results for the algorithms. Section 4.4 shows bounds that explain the difference between the performance of the short term and long term power constraint solutions. Specifically, we show that the short term constraint solution results in an outage probability that falls as an inverse power of the SNR, while the long term constraint solution results in an outage probability that falls off at least exponentially with SNR. Often, error exponents are used to show how quickly the error probability decays with code length. This provides more detail on possible data rates than simply the capacity figure, because the smaller the data rate transmitted as compared to the capacity, the faster the error probability falls to 0. Outage capacity based systems potentially have a higher error exponent because the codeword can span the entire K blocks. Section 4.5 presents 'error exponents' based on a random coding argument. The single quotes indicate that the exponents are not true error exponents; these don't exist in the delay constrained fading case because the error probability converges to P_{out} and not to 0 as code length $\to \infty$.

4.2 Power Adaptation for Minimum Outage Probability

4.2.1 Power adaptation with short term power constraint

Theorem 2 *The probability of outage for rate R_0 and the short term power constraint (4.1) is minimized for Gaussian codes over the K-block fading channel by choosing the power adaptation functions $\{P_i(g^{(i)}), i = 1, \ldots, K\}$, which are calculated using*

algorithm **Outage_short**(K, P_0, R_0). *The minimum probability of outage for rate* R_0 *is* $\psi_1(KR_0, KP_0)$.

Algorithm **Outage_short**(K, P_0, R_0) : For $n = K, K-1, \ldots, 1$

At time n, choose

$$P_n^*(g_n, R, P) = \underset{0 \le P_n \le P}{argmin}\ \psi_{n+1}(R - \log(1 + P_n g_n), P - P_n) \qquad \text{and} \tag{4.3}$$

$$\psi_n(R, P) = \underset{g_n}{\mathbf{E}}[\psi_{n+1}(R - \log(1 + P_n^*(g_n, R, P) g_n), P - P_n^*(g_n, R, P))] \tag{4.4}$$

$$\psi_{K+1}(R, P) = \begin{cases} 1 & \text{if} \quad R > 0 \\ 0 & \text{if} \quad R \le 0 \end{cases} \qquad \text{initialization} \tag{4.5}$$

end

In this manner, compute the *functions* $\{\psi_n(R, P), 0 \le R \le KR_0, 0 \le P \le KP_0\}$ by a backward recursion (i.e. proceeding $n = K, K-1, \ldots, 1$). In the process, we also obtain the three-dimensional functions $\{P_n^*(g_n, R, P),\ 0 \le R \le KR_0, 0 \le P \le KP_0\}$. $\psi_n(R, P)$ is interpreted as the minimum outage probability if the power allotted is P and the target rate is R.

Now, the optimal power adaptation strategy chooses

For $n = 1, \ldots, K$

$$P_n(g^{(n)}) = P_n^*(g_n, R^{(n)}, P^{(n)}) \qquad \text{where power} \tag{4.6}$$

$$P^{(n)} = KP_0 - \sum_{i=1}^{n-1} P_i(g^{(i)}) \qquad \text{and rate} \tag{4.7}$$

$$R^{(n)} = KR_0 - \sum_{i=1}^{n-1} \log(1 + P_i(g^{(i)}) g_i) \tag{4.8}$$

depend on the fades seen by the earlier blocks. Here, we define $P^{(1)} \equiv KP_0$ and

$R^{(1)} \equiv KR_0$. $P^{(n)}$ and $R^{(n)}$ are the remaining power and required rate seen for the nth block.

Note that in the algorithm above, $\psi_K(R,P) = F_G(\frac{1}{P}(e^R - 1))$, where $F_G(g)$ is the cumulative distribution function of the channel gain. In the algorithm, one precomputes offline and stores the functions $\{P_n^*(g_n, R, P),\ 0 \leq R \leq KR_0, 0 \leq P \leq KP_0\}$. When the system is online, depending on the sequence of channel gains seen, one allots power $P_n(g^{(n)})$ to block n.

As a corollary, it is easy to show that the optimum strategy simplifies to the following strategy when $SNR \ll 1$. For such SNR, we can approximate $\log(1+Pg) \approx Pg$, and so (4.3) implies that given a target rate R_0 and power P_0, the transmitter should not transmit until the channel gain in the current block supports that rate. i.e. until $KP_0 g_n \geq KR_0$. Then, the minimum outage probability $\psi_1(KR_0, KP_0) = [F(R_0/P_0)]^K$. This is interesting because at low SNR, the optimum (waterfill) strategy when all the K channel gains *are known in advance* is selection diversity (select the maximum gain channel), which results in the same outage probability as that for the online strategy!

4.2.2 Power adaptation with long term power constraint

Theorem 3 *The probability of outage for rate R_0 and the long term power constraint (4.2) is minimized for Gaussian codes over the K-block fading channel by choosing the power adaptation function pairs $\{(P_i(g^{(i)}), Q_i(g^{(i)})), i = 1, \ldots, K\}$, which are calculated using algorithm* **Outage_long**(K, P_0, R_0). *The minimum probability of outage for rate R_0 is $\psi_1(KR_0, KP_0)$.*

Algorithm **Outage_long**(K, P_0, R_0) : For $n = K, K-1, \ldots, 1$

At time n, choose $P_n^*(g_n, R, P)$ as in (4.3). Then choose,

$$\begin{aligned} \theta_{n+1}(g_n, R, P) &= \psi_{n+1}(R - \log(1 + P_n^*(g_n, R, P)g_n), \\ &\quad P - P_n^*(g_n, R, P)) \quad \text{, and compute} \qquad (4.9) \\ \psi_n(R, P) &= min \ \underset{g_n}{\mathbf{E}}[\theta_{n+1}(g_n, R, \tilde{P}(g_n))] \qquad (4.10) \end{aligned}$$

Initialize $\psi_{K+1}(R, P)$ as in (4.5). The minimization in (4.10) is over all non-negative functions $\tilde{P}(g_n) \geq 0$ such that $\mathbf{E}[\tilde{P}(g_n)] = P$. To compute this minimum, one can use the calculus of variations. Alternately, we can use another dynamic program, which runs over the index of channel gains. Call the minimizing function $\tilde{P}(g_n)$ for each (R, P) pair as $Q_n^*(g_n, R, P)$.

In this manner, compute the *functions* $\{\psi_n(R, P), 0 \leq R \leq KR_0, P \geq 0\}$ by a backward recursion (i.e. proceeding $n = K, K-1, \ldots, 1$). In the process, we also obtain the three-dimensional function pairs $\{(P_n^*(g_n, R, P), Q_n^*(g_n, R, P)), \ 0 \leq R \leq KR_0, P \geq 0\}$.

Now, the optimal power adaptation strategy chooses

For $n = 1, \ldots, K$

$$\begin{aligned} Q_n(g^{(n)}) &= Q_n^*(g_n, R^{(n)}, P^{(n)}) \qquad \text{a power function} \\ P_n(g^{(n)}) &= P_n^*(g_n, R^{(n)}, Q_n(g^{(n)})) \qquad \text{where} \qquad (4.11) \\ P^{(n)} &= Q_{n-1}(g^{(n-1)}) - P_{n-1}(g^{(n-1)}) \qquad \text{and} \\ R^{(n)} &= KR_0 - \sum_{i=1}^{n-1} \log(1 + P_i(g^{(i)})g_i) \end{aligned}$$

depend on the fades seen by the earlier blocks. Here, we define $P^{(1)} \equiv KP_0$ and $R^{(1)} \equiv KR_0$. $Q_n(g^{(n)})$ is the total power allotted to blocks n to K, $P^{(n)}$ and $R^{(n)}$ are

the remaining power and required rate seen for the nth block.

In the above algorithm, one pre-computes offline and stores the function pairs $\{(P_n^*(g_n, R, P), Q_n^*(g_n, R, P)), 0 \leq R \leq KR_0, P \geq 0\}$. When the system is online, depending on the sequence of channel gains seen, one allots power $P_n(g^{(n)})$ to block n.

4.3 Performance of the Online Algorithms

We demonstrate the capacity improvement afforded by the online power adaptation scheme over a constant power scheme (which would occur, say when there is a stringent delay constraint of $K = 1$, or when the CSI is unknown at the transmitter).

4.3.1 Simulations

In all the simulations, the channel is assumed to be Rayleigh fading, though the algorithms could very well be applied to other kinds of channel distributions too. In Figure 4.1, we have the case of low SNR. The figure shows the outage probability achieved by using the online algorithm with short term power constraint (Section 4.2.1), at various SNRs and $K = 1, 2, 3, 5$. The target data rate is $R_0 = 0.1$ nats per transmission, which is the capacity of a fixed gain AWGN channel operating at -9.8 dB. The dotted lines refer to the outage probability obtained in the absence of channel knowledge, with a delay constraint of K blocks. In that case, the transmitter would transmit at a constant rate of R_0 and power P_0 per transmission. It is clear that a substantial SNR gain is obtained over this by adapting the power. Figure 4.2 shows the performance of the online algorithm the with long term power constraint (Section 4.2.2) for rate $R_0 = 0.1$. The performance of the short term algorithm is also plotted

for comparison. It is clear that the long term constraint algorithm outperforms the short term constraint algorithm. The improved performance may be worth the higher complexity of the long term constraint algorithm for some cases.

Unlike the low SNR case, the high SNR case shows negligible gain when the short term constraint power adaptation is used. But Figure 4.3 shows that the long term constraint algorithm produces a significant SNR gain even at large SNRs. In this figure, the target data rate is $R_0 = 3$ nats per transmission, corresponding to an AWGN SNR of 12.8 dB.

It is clear from the plots that in the case of the short term constraint, a larger K results in graph with a larger slope, reflecting the higher diversity effect. What is not clear is that the long term constraint algorithm should ideally produce at least an exponentially decaying error graph (See Section 4.4 for a proof). The plots do not show such an exponential decay in the case of the long term constraint algorithm. This discrepancy occurs because of the 'flooring effect'; the power adaptation functions were calculated sampled at a finite number of points, and that decides the outage probability floor (or the rate at which it falls). This can be reduced by increasing the discretization.

4.3.2 A Note on Discretization

Since the optimal algorithms are obtained by dynamic programming, hence it becomes necessary to study the effect of discretization of the state variables on the optimality of the solution. In fact in several cases, such as the long term constraint case, discretization severely limits optimal performance at high SNRs. The standard analysis (see e.g. [23]) uses Lipschitz conditions on the utility function $\mu(x)$, among other requirements. This will not be satisfied by (2.6). Our analysis of the effect

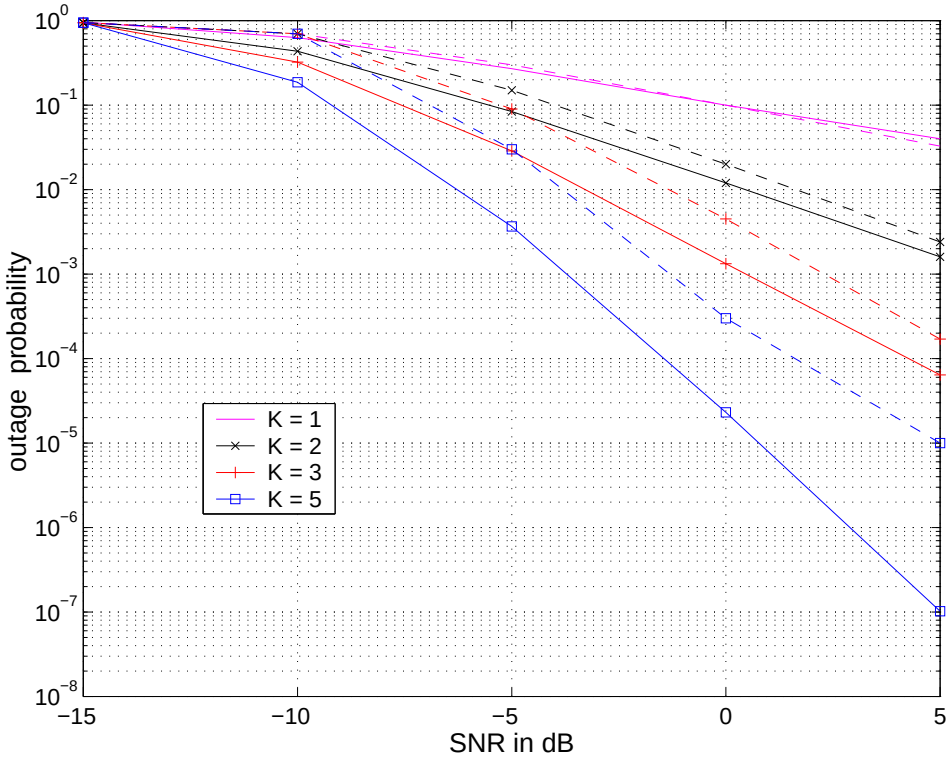

Figure 4.1: Short term constraint algorithm in Rayleigh fading for $R_0 = 0.1$ and various K. Solid lines: optimum algorithm; dotted lines: no power adaptation

of discretization shown below, is different from the standard analysis and we believe more elegant.

We assume a linear discretization of the state variables rate R, and power P, and a geometric discretization of the channel gain g. The motivation for the linear discretization of R is that the *sum* of rates in the K blocks is important. The motivation for the linear discretization of P is that the *sum* of powers in the K blocks is constrained. On the other hand, g is discretized geometrically because it appears as a product with P in $\log(1 + Pg)$. Thus, the discrete values for R are $S_R = \{0, \varepsilon_R, 2\varepsilon_R, \ldots\}$, the discrete values for P are $S_P = \{0, \varepsilon_P, 2\varepsilon_P, \ldots\}$, and those for g are $S_g = \{0, g_0, \varepsilon_g g_0, \varepsilon_g^2 g_0, \ldots, g_\infty\}$, where $\varepsilon_g > 1$.

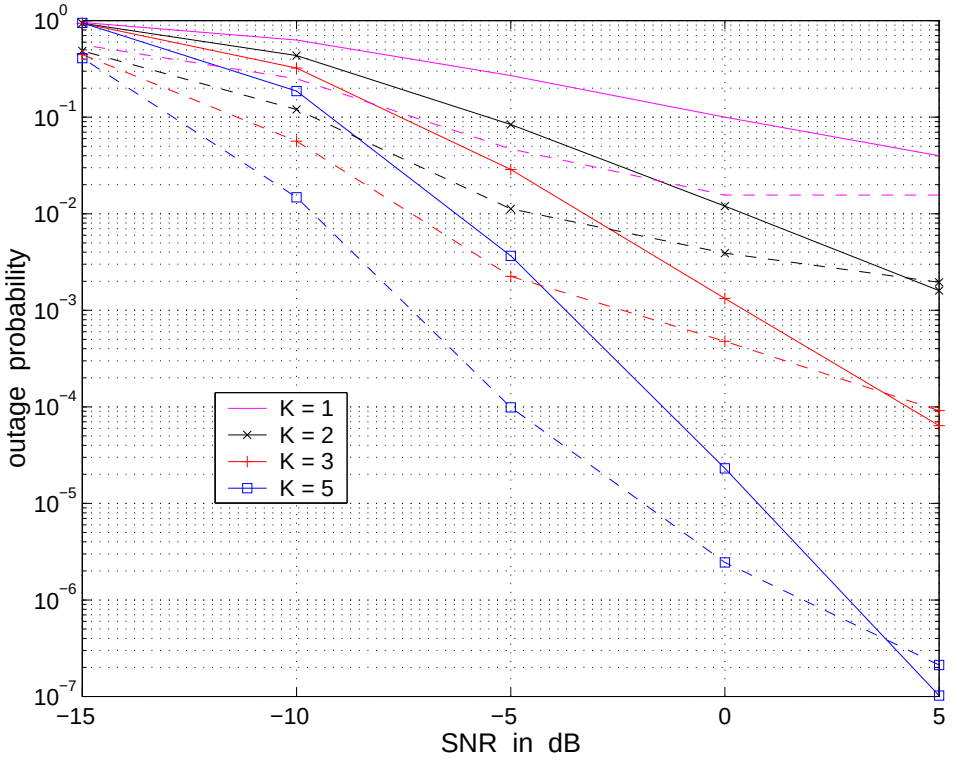

Figure 4.2: Long term constraint algorithm in Rayleigh fading for $R_0 = 0.1$ and various K. Solid lines: short term algorithm; dotted lines long term algorithm

We first analyze the effect of discretizing each of the three variables separately,

Discretization of g: The values of g_i are rounded down to the highest value in S_g smaller than g_i. Call the discretized versions of g_i as $\tilde{g}_i$ respectively. The event that any of the $\tilde{g}_i$s are either 0 or g_∞ occurs with probability $P_{err} \leq (1-(1-F_G(g_0))^K) + (1-F_G^K(g_\infty))$. Call the set of such events S_{err}. Then, for any event in the complementary set $\overline{S}_{err}$, we can achieve the same or better performance as the continuous g (optimal) scheme, by transmitting power $\varepsilon_g P$ instead of power P. Due to the geometric discretization of g, we can compare the *discrete scheme* with a continuous variable scheme *that has fading gains* g/ε_g *but power levels* $\varepsilon_g P$. The latter scheme clearly has the same performance as the

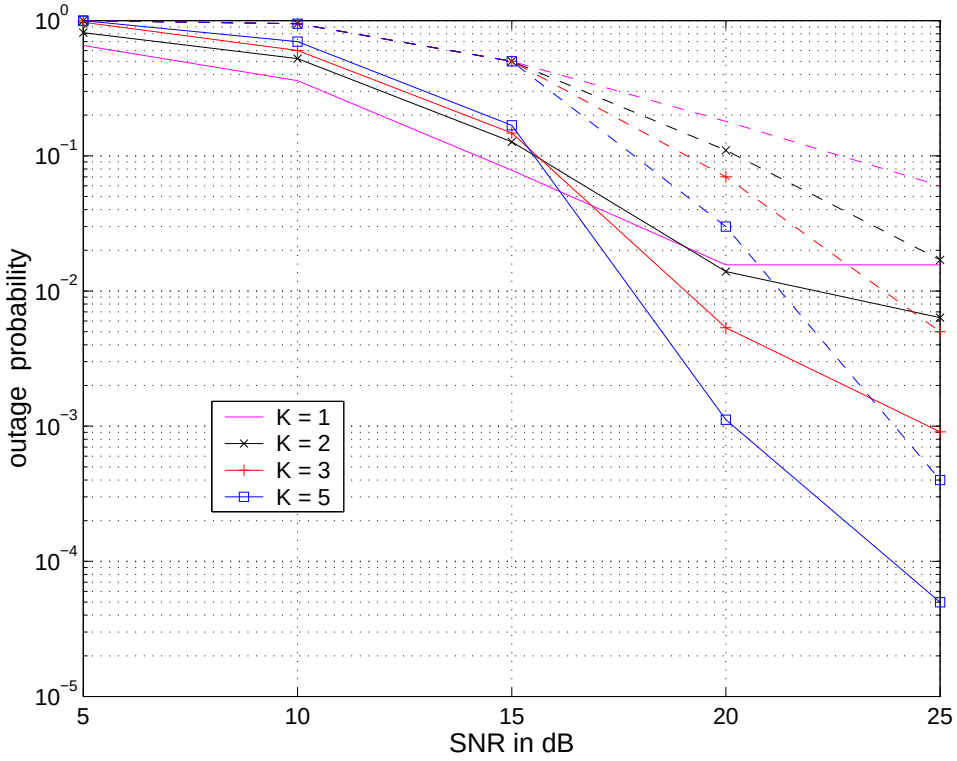

Figure 4.3: Long term constraint algorithm in Rayleigh fading for $R_0 = 3$ and various K. Solid lines: long term constraint algorithm; dotted lines no power adaptation

(optimal) continuous variable scheme. Thus, by allowing the discretized scheme to use a power constraint of $\varepsilon_g P_0$ instead of P_0, we can guarantee performance at least as good as the optimal continuous variable scheme, for all events in $\overline{S}_{err}$. This occurs with probability $1 - P_{err}$.

Discretization of P: Since ε_P is the maximum quantization error in P, hence by increasing the power allotted to each stage by ε_P, we can make the discrete scheme perform at least as well as the optimum continuous variable scheme. Therefore, the discrete scheme with power constraint $KP_0 + K\varepsilon_P$ performs at least as well as the optimum scheme with power KP_0.

Discretization of R: Since ε_R is the maximum discretization error in R at any stage, hence the outage probability will be smaller for a discrete scheme that has a target data rate of $KR_0 - K\varepsilon_R$, than for the optimal continuous scheme that has a target rate of KR_0.

Combining these results, we can assert that if the optimal scheme with discretized g, R, P, a power constraint of $KP_0 + K\varepsilon_P$, and a target data rate of $KR_0 - K\varepsilon_R$, is compared with the optimal scheme with continuous variables, but power constraint of KP_0 and target rate KR_0, then the outage probability of the discrete scheme is higher than the outage probability of the continuous scheme by at most P_{err}, where $P_{err} \leq (1 - (1 - F_G(g_0))^K) + (1 - F_G^K(g_\infty))$.

4.4 Asymptotic behavior

Whereas outage probability cannot be found analytically for the optimum policies, we can attempt to bound their asymptotic behavior for large SNR. In this section, we will present some such bounds for the Rayleigh fading channel, which is commonly encountered in narrowband communications. These show dramatically different behavior for policies with the long term constraint as opposed to policies with the short term constraint. In this Section, without loss of generality, we can normalize $\mathbf{E}[g] = 1$.

4.4.1 Short term constraint policy

First we show that the short term constraint results in a policy (4.6) that produces a diversity-K outage probability graph. Formally, if $P_{short}(P_0)$ is the minimum outage probability at $SNR = P_0$ for the short term constraint policy, then there exist

constants c_1, c_2, P_{th} such that for $P_0 > P_{th}$

$$c_1 - K\log(P_0) \quad \leq \quad \log(P_{short}) \quad \leq \quad c_2 - K\log(P_0) \tag{4.12}$$

To prove this, we first show that in a Rayleigh fading channel, the constant power policy (which assigns power P_0 per block) results in a diversity-K outage probability graph. Whereas this result is well-known in uncoded (and AWGN coded) systems, we've not come across a similar analysis for systems at capacity. To that end, we show that if

$$P_{const} = \mathbf{P}[\sum_{i=1}^{K} \log(1 + P_0 g_i) < KR_0]$$

then

$$c - K\log(P_0) \;\leq\; \log(P_{const}) \;\leq\; c - K\log(P_0) + K(e^{KR_0} - 1)\left(\frac{1}{P_{th}} - \frac{1}{P_0}\right) \tag{4.13}$$

It is clear from these inequalities that $P_{const} \sim P_0^{-K}$ for large P_0 (SNR).

This result can be easily derived because P_{const} is a continuous function of P_0 when g is Rayleigh fading. Look at,

$$\frac{\partial \log P_{const}}{\partial \log P_0} \quad = \quad \frac{P_0}{P_{short}} \frac{\partial P_{const}}{\partial P_0}$$

Define

$$\begin{aligned} w &= \sum_{i=1}^{K} \log(1 + P_0 g_i) \\ y_i &= \log(1 + P_0 g_i) \end{aligned}$$

Now denoting the probability density function of W by $f_W(w)$,

$$\begin{aligned}
\frac{\partial P_{const}}{\partial P_0} &= \int_0^{KR_0} \frac{\partial f_W(w)}{\partial P_0} \cdot \mathrm{d}w \\
&= \int_0^{KR_0} \frac{\partial(\ \underset{K}{*} f_Y(w)\ )}{\partial P_0} \cdot \mathrm{d}w \\
&= \int_0^{KR_0} \cdots \int_0^{KR_0} \frac{\partial(f_Y(\tau_1) \cdot f_Y(\tau_2 - \tau_1) \cdots f_Y(w - \tau_{K-1}))}{\partial P_0} \cdot \\
&\qquad \mathrm{d}\tau_1 \cdot \mathrm{d}\tau_2 \cdots \mathrm{d}\tau_{K-1} \cdot \mathrm{d}w \\
&= K \cdot \int_0^{KR_0} \cdots \int_0^{KR_0} \frac{\partial f_Y(\tau_1)}{\partial P_0} \cdot f_Y(\tau_2 - \tau_1) \cdots f_Y(w - \tau_{K-1}) \cdot \\
&\qquad \mathrm{d}\tau_1 \cdot \mathrm{d}\tau_2 \cdots \mathrm{d}\tau_{K-1} \cdot \mathrm{d}w
\end{aligned} \tag{4.14}$$

where '$\underset{K}{*}$' denotes the K-fold convolution. Since

$$\begin{aligned}
f_Y(y) &= \frac{e^y}{P_0} f_G\left(\frac{e^y - 1}{P_0}\right) \qquad \text{and} \\
f_G(g) &= e^{-g} \qquad \text{hence} \\
\frac{\partial f_Y(y)}{\partial P_0} &= -\frac{f_Y(y)}{P_0} \cdot \left(1 - \frac{e^y - 1}{P_0}\right)
\end{aligned}$$

Within the limits of integration of the dummy variables $0 \le \tau_1, \tau_2, \ldots, w \le KR_0$,

$$-\frac{f_Y(y)}{P_0} \le \frac{\partial f_Y(y)}{\partial P_0} \le -\frac{f_Y(y)}{P_0} \cdot \left(1 - \frac{e^{KR_0} - 1}{P_0}\right)$$

Substituting in (4.14),

$$-K\frac{P_{const}}{P_0} \le \frac{\partial P_{const}}{\partial P_0} \le -K\frac{P_{const}}{P_0}\left(1 - \frac{e^{KR_0} - 1}{P_0}\right)$$

Integrating these inequalities for $P_0 > P_{th}$ gives the result (4.13).

With these bounds we can now look at the asymptotic behavior of the short term

constraint policy. In particular, we will derive (4.12). It is clear that $P_{short}(P_0) \leq P_{const}(P_0)$. Also, if we allot power KP_0 *to each block*, we will clearly outperform the short term constraint policy, which is limited to a total power of KP_0. Therefore, $P_{short}(P_0) \geq P_{const}(KP_0)$. Combining these facts with (4.13), we conclude that for $P_0 > P_{th}$,

$$c - K\log K - K\log(P_0) \;\leq\; \log(P_{short}) \;\leq\; c - K\log(P_0) + K(e^{KR_0} - 1)\left(\frac{1}{P_{th}} - \frac{1}{P_0}\right)$$

and so

$$c_1 - K\log(P_0) \;\leq\; \log(P_{short}) \;\leq\; c_2 - K\log(P_0)\ , \qquad \text{with}$$

$$\begin{aligned} c_1 &\doteq c - K\log K \\ c_2 &\doteq c + \frac{K}{P_{th}}(e^{KR_0} - 1) \end{aligned}$$

Whereas these bounds are not tight, they will suffice to illustrate the main difference between the performance of the short term and long term constraint policies.

4.4.2 Long term constraint policy

In contrast to the short term constraint policy, the long term constraint policy (4.11) produces an outage probability graph that is at least exponentially decreasing with SNR. Formally, if $P_{long}(P_0)$ is the minimum outage probability at $SNR = P_0$ for the long term constraint policy, then there exist constants c_3, P_{th} such that for $P_0 > P_{th}$

$$\log(P_{long}) \;\leq\; c_3 - \frac{1}{e^{R_0} - 1}P_0 \tag{4.15}$$

To prove this, we first show that in a Rayleigh fading channel, a power inversion policy (which optimally distributes power P_0 per block) results in an exponential graph. Define

$$\begin{aligned} P_{inv} &= \mathbf{P}[\log(1+P(g)g) < R_0] \qquad \text{where} \\ P(g) &= \begin{cases} \frac{e^{R_0}-1}{g} & \text{if } g \geq g_0 \\ 0 & \text{else} \end{cases} \\ \mathop{\mathbf{E}}_{g}[P(g)] &= P_0 \qquad \text{solve for } g_0 \end{aligned}$$

We also define

$$\begin{aligned} \phi(g_0) &= \mathop{\mathbf{E}}_{g}[\frac{1}{g} \; ; \; g \geq g_0] \\ &= \int_{g_0}^{\infty} \frac{1}{g} f_G(g) \mathrm{d}g \end{aligned}$$

It is clear that

$$\begin{aligned} \phi(g_0) &= \frac{P_0}{e^{R_0}-1} \\ P_{inv} &= F_G(g_0) \end{aligned}$$

where $F_G(g)$ is the cumulative distribution function of g. So

$$\begin{aligned} \frac{\partial P_{inv}}{\partial d g_0} &= f_G(g_0) \\ \phi'(g_0) &= \frac{1}{e^{R_0}-1} \cdot \frac{\partial P_0}{\partial g_0} \qquad \text{, so} \\ -\frac{1}{g_0} \cdot f_G(g_0) &= \frac{1}{e^{R_0}-1} \cdot \frac{\partial P_0}{\partial g_0} \qquad \text{and} \\ \frac{\partial \log P_{inv}}{\partial P_0} &= \frac{1}{P_{inv}} \cdot \frac{\partial P_{inv}}{\partial g_0} \cdot \frac{\partial g_0}{\partial P_0} \end{aligned}$$

$$\begin{aligned} &= \frac{1}{P_{inv}} \cdot f_G(g_0) \cdot \left(-\frac{1}{f_G(g_0)} \cdot \frac{g_0}{e^{R_0}-1} \right) \\ &= -\frac{1}{e^{R_0}-1} \cdot \frac{g_0}{F_G(g_0)} \end{aligned}$$

For a Rayleigh fading channel $F_G(g) = 1 - e^{-g}$, therefore $F_G(g_0) \leq g_0$. Thus

$$\frac{\partial \log P_{inv}}{\partial P_0} \leq -\frac{1}{e^{R_0}-1}$$

Integrating the inequality, we can assert that for $P_0 > P_{th}$,

$$\log P_{inv} \leq c - \frac{P_0}{e^{R_0}-1}$$

Now, to upper bound P_{long}, we note that a suboptimal solution to the long term constraint problem is to allot average power P_0 to each block, and to use this power optimally in each block to achieve a rate of exactly R_0. In that case, the probability that any block will not achieve the rate R_0 is simply P_{inv}. Therefore, P_{long} is upper bounded by the probability that any one of the K blocks fails to achieve rate R_0 in the suboptimal scheme.

$$\begin{aligned} \log P_{long} &\leq \log(KP_{inv}) \\ &\leq c + \log K - \frac{P_0}{e^{R_0}-1} \end{aligned}$$

which is the bound (4.15).

It is clear on comparing (4.12) and (4.15) that whereas the short term policy results in an outage probability that is inversely proportional to the Kth power of SNR, the long term policy results in an outage probability that is exponential in the SNR. Therefore, we can expect the difference in their performance to grow as SNR

increases. Practically speaking though, the discretization required in the dynamic programs reduces this advantage.

4.5 Random coding bounds

The random coding bound for the optimal power control algorithms can be calculated following the development in [18] (see also [32]). Random coding bounds specify an upper bound (which is usually tight in the exponent) on the decoding error, of the form,

$$P_{err} \quad \leq \quad \exp(-E_r(R, q(\mathbf{X})) \cdot T)$$

where R is the rate of transmission, $q(\mathbf{X})$ is the input symbol density, and T is the code length. The error exponent E_r is a decreasing function of R. Thus if R is further (smaller) away from capacity, P_{err} decays faster. The random coding bound therefore contains more information than just the capacity bound. For fading channels, E_r also depends on the channel gains, as described below.

In the subsequent development, we assume that the optimal power control algorithms were designed for a target data rate of R_0, but we transmit at a data rate R (typically $R < R_0$). Thus, the Gaussian codebook is assumed to have e^{KRT_0} codewords.

$$P_{err} \quad \leq \quad \underset{g^{(K)}}{\mathbf{E}} [\ \exp(-KTE_r(q(\mathbf{X}), g^{(K)}, R)) \ ] \tag{4.16}$$

with random coding exponent

$$E_r(q(\mathbf{X}), g^{(K)}, R) = \max_{0\le\rho\le 1} \frac{1}{K}(\sum_{n=1}^{K} E_0(q_n(X), \rho, g_n) - \rho RK) \quad (4.17)$$

and input symbol density

$$q(\mathbf{X}) = \Pi_{n=1}^{K}\, q_n(X_n) \qquad \text{(independent but not identical)}$$

$$E_0(q(X), \rho, g) = -\log\left(\int_y \left(\int_x q(x)p(y/x,\alpha)^{\frac{1}{1+\rho}}\mathrm{d}x\right)^{1+\rho}\mathrm{d}y\right)$$

Here, $q(X)$ is the probability density of the input X to the channel, while $p(Y/X,\alpha)$ is the channel transition density function. Note that for a fading channel, $g = |\alpha|^2$. Strictly speaking, $E_0(q(X),\rho,g)$ should depend on α, but since a Gaussian codebook is assumed, it effectively depends on only g (see below)

Since we use a Gaussian codebook,

$$q_n(x) = \frac{1}{\pi P_n(g^{(n)})}\exp(-|x|^2/P_n(g^{(n)}))$$

while since the channel is assumed to be AWGN, with unit noise variance within each block[1], hence

$$p(y/x,\alpha) = \frac{1}{\pi}\exp(-|y-\alpha x|^2)$$

Simplifying, we calculate

$$E_0(q(X),\rho,g) = \rho\log\left(1+\frac{gP}{1+\rho}\right) \qquad \text{where } P \text{ is the variance of } X. \text{ i.e.}$$

$$q(x) = \frac{1}{\pi P}\exp(-|x|^2/P)$$

[1] again, the noise power has been absorbed into channel gain g

The random coding exponent 4.17 simplifies to

$$E_r(q(\mathbf{X}), g^{(K)}, R) = \max_{0 \le \rho \le 1} \frac{\rho}{K} \left(\sum_{n=1}^{K} \log \left(1 + \frac{g_n P_n(g^{(n)})}{1+\rho} \right) - RK \right)$$

Summarizing, the random coding bound on the error probability, with data rate is R and power control strategy $\{P_n(g^{(n)}), n = 1, \ldots, K\}$, is

$$P_{err} = \mathop{\mathbf{E}}_{g^{(K)}} [\ \exp(-KTE_r(q(\mathbf{X}), g^{(K)}, R))\ ] \tag{4.18}$$

$$E_r(q(\mathbf{X}), g^{(K)}, R) = \max_{0 \le \rho \le 1} \frac{\rho}{K} \left(\sum_{n=1}^{K} \log \left(1 + \frac{g_n P_n(g^{(n)})}{1+\rho} \right) - RK \right) \tag{4.19}$$

4.5.1 Exact computation

The difficulty in calculating the error bound is, of course, the maximization in (4.19). Each K-tuple $g^{(K)}$ requires a separate maximization over ρ. Monte Carlo calculations will therefore require number of calculations exponential in K. This can be done for small values of K. A simplifying property that can be used is that

$$E_r(\rho, q(\mathbf{X}), g^{(K)}, R) = \frac{\rho}{K} \left(\sum_{n=1}^{K} \log \left(1 + \frac{g_n P_n(g^{(n)})}{1+\rho} \right) - RK \right) \tag{4.20}$$

is concave in ρ.

4.5.2 Small deviation from target rate

As explained in the earlier section, exact computation of the random coding bound would be very complex for large K. However, if the power control algorithm is designed to achieve rate R_0 (as in the algorithms obtained by dynamic programming for outage probability problems), we can compute an approximation to the error

bound (4.18) for the rates $R = R_0 - \varepsilon, \varepsilon \ll R_0$ (small deviation from target rate). These bounds will be valid if the power control is designed for rate R_0, but in fact the codebook has rate R.

To compute the bound for the small deviation case, the ensemble of $g^{(K)}$ can be divided into those for which the power control is able to achieve the target rate R_0 (no outage) and those $g^{(K)}$ for which it is unable to meet the target rate (outage). For the outage event, the optimum $\rho = 0$. For the event where there is no outage, we can approximate the calculation of the random coding bound by approximating (4.20) by its truncated Taylor series

$$\begin{aligned} E_r(\rho, q(\mathbf{X}), g^{(K)}, R) &= E_r(\rho = 0) + \rho \left.\frac{\partial E_r(\rho)}{\partial \rho}\right|_{\rho=0} + \\ &\quad \frac{1}{2}\rho^2 \left.\frac{\partial^2 E_r(\rho)}{\partial \rho^2}\right|_{\rho=0} + o(\rho^3) \end{aligned} \tag{4.21}$$

The reasoning is that if $R = \text{capacity} - \varepsilon$, then for all such $g^{(K)}$ the optimum ρ will be approximately 0. Simplifying (4.21) and using the fact that $R = R_0 - \varepsilon$, $R_0 = \frac{1}{K}\sum_{n=1}^{K} \log(1 + g_n P_n)$ (since we are considering $g^{(K)}$ for which there is no outage), hence

$$E_r(\rho, q(\mathbf{X}), g^{(K)}, R) = \varepsilon\rho - \frac{\rho^2}{K} \sum_{n=1}^{K} \left(\frac{P_n g_n}{1 + P_n g_n} \right) + o(\rho^3)$$

Therefore,

$$E_r(q(\mathbf{X}), g^{(K)}, R) = \max_{0 \le \rho \le 1} E_r(\rho, q(\mathbf{X}), g^{(K)}, R)$$

is achieved approximately by

$$\begin{aligned} \rho^{(opt)} &= \frac{\varepsilon}{2\mathcal{H}}, \qquad \text{where} \\ \mathcal{H} &= \frac{1}{K}\sum_{n=1}^{K}\left(\frac{P_n g_n}{1+P_n g_n}\right) \end{aligned}$$

which yields the maximized $E_r = \varepsilon^2/(4\mathcal{H})$. Recollect again that $\rho^{(opt)}$ is the optimum ρ only for the case that there is no outage. i.e. for the case that the sequence of channel gains $g^{(K)}$ is such that the power control algorithm succeeds in meeting the target rate. Combining the error bound for such $g^{(K)}$ which do not result in an outage, with the outage event case, we can bound the error probability approximately by

$$P_{err} \leq \underset{g^{(K)}:\text{no outage}}{\mathbf{E}}[\ \exp(-KT\varepsilon^2/(4\mathcal{H}))\ ] + \mathbf{P}[\text{outage}]$$

To avoid Monte Carlo calculations, a weaker upper bound can be obtained by bounding $\mathcal{H}$ as

$$\begin{aligned} \mathcal{H} &= 1 - \frac{1}{K}\sum_{n=1}^{K}\left(\frac{1}{1+P_n g_n}\right) \\ &\leq 1 - \left(\Pi_{n=1}^{K}\ \frac{1}{1+P_n g_n}\right)^{1/K} \qquad (4.22) \\ &= 1 - e^{-R_0} \qquad (4.23) \end{aligned}$$

where (4.22) uses the Arithmetic-Geometric means inequality, while (4.22) uses the fact that in the absence of outage, the optimum power control achieves $R_0 = \frac{1}{K}\sum_{n=1}^{K} \log(1+g_n P_n)$. This gives the random coding bound

$$P_{err} \leq \exp\left(-KT\frac{\varepsilon^2}{4(1-\exp(-R_0))}\right) + \mathbf{P}[\text{outage}]$$

Again, this approximate bound is only valid for the rates $R = R_0 - \varepsilon, \varepsilon \ll R_0$ (small deviation from target rate).

4.5.3 A weaker upper bound

If an error bound is desired for an R that is very different from R_0, then the approximation in Section 4.5.2 cannot be used. At the same time, if the K is large, the exact computation in Section 4.5.1 may be too complex. However, an upper bound is yet possible if we are willing to settle for a weaker bound. This is obtained by optimizing ρ over a set of $g^{(K)}$, in contrast to optimizing ρ for each $g^{(K)}$. We can then use the fact that the optimum power control solution results in a Markov chain $\{P^{(1)}, R^{(1)}\} \rightarrow \{P^{(2)}, R^{(2)}\} \rightarrow \cdots \rightarrow \{P^{(K+1)}, R^{(K+1)}\}$ (see equation (4.8) for the definitions of $\{P^{(n)}, R^{(n)}\}$). Due to this, (another!) dynamic program can be used to calculate the random coding bound. Thus, a weaker bound is calculated in the general R case, using a dynamic program that utilizes the Markov property of $\{P^{(n)}, R^{(n)}\}$). Again, this Markov property arises because the *optimal* power control is used.

To this end, we proceed as follows. For the moment, assume that ρ is a constant over all $g^{(K)}$. In that case, (4.16) can be rewritten as

$$P_{err} \leq \min_{0 \leq \rho \leq 1} P_{err}(\rho) \ , \qquad \text{where} \tag{4.24}$$

$$P_{err}(\rho) = \mathop{\mathbf{E}}_{g^{(K)}} [\ \exp(-KTE_r(\rho, q(\mathbf{X}), g^{(K)}, R))\ ] \ , \qquad \text{and} \tag{4.25}$$

$$E_r(\rho, q(\mathbf{X}), g^{(K)}, R) = \frac{\rho}{K} \left(\sum_{n=1}^{K} \log \left(1 + \frac{g_n P_n(g^{(n)})}{1+\rho} \right) - RK \right) \ , \qquad (\text{as in } (4.20))$$

This bound is weaker than (4.18) because it assumes a constant ρ over *all* $g^{(K)}$ (which

is then optimized), as opposed to the exact bound which optimizes ρ for *each* $g^{(K)}$. The strategy we use is to first find an algorithm to calculate $P_{err}(\rho)$. We can show easily that $P_{err}(\rho)$ is convex in ρ. Therefore, it can be minimized efficiently over $\{0 \leq \rho \leq 1\}$ using bisection (Section A.1).

The next few pages show how $P_{err}(\rho)$ can be computed using a dynamic program. If desired, the reader can skip the derivation and go directly to the result, which is summarized by algorithm **Calculate_error(ρ)**.

The state space of the dynamic program used to compute the optimal power control scheme in Section 4.2 was $Z_n = \{P^{(n)}, R^{(n)}\}$. Note that $Z_1 = \{KP_0, KR_0\}$ is a constant. Since the optimal strategy is Markovian in this state space, hence $Z_1 \to Z_2 \to \cdots \to Z_{K+1}$ forms a Markov chain, when the optimal power control strategy is used. In terms of this state, the optimal power functions chosen for the nth block are (see (4.6) and (4.11))

$$P_n(g^{(n)}) \;=\; P_n^*(g_n, Z_n) \tag{4.26}$$

Thus, $P_{err}(\rho)$ can be computed as below

$$P_{err}(\rho) \;=\; \mathop{\mathbf{E}}_{g^{(K)}}\left[\ \exp\left(-KT\frac{\rho}{K}\left(\sum_{n=1}^{K}\log\left(1+\frac{g_nP_n(g^{(n)})}{1+\rho}\right)-RK\right)\right)\ \right] \tag{4.27}$$

$$=\; \mathop{\mathbf{E}}_{g^{(K)}}\left[\ \exp\left(-KT\frac{\rho}{K}\left(\sum_{n=1}^{K}\log\left(1+\frac{g_nP_n^*(g_n,Z_n)}{1+\rho}\right)-RK\right)\right)\ \right]$$

$$=\; \mathop{\mathbf{E}}_{g^{(K)}}\left[\ \exp(-KT(\sum_{n=1}^{K}\phi(Z_n,g_n)))\ \right], \qquad \text{where we define} \tag{4.28}$$

$$\phi(Z_n,g_n) \;=\; \frac{\rho}{K}\left(\log\left(1+\frac{g_nP_n^*(g_n,Z_n)}{1+\rho}\right)-R\right)$$

Now, we can make use of the Markov property of $\{Z_n\}$.

$$\begin{aligned}
P_{err}(\rho) &= \mathop{\mathbf{E}}_{Z_2, g_2^K}[\ \exp\left(-KT\sum_{n=2}^{K}\phi(Z_n, g_n)\right)\ \cdot\ \mathop{\mathbf{E}}_{g_1}[\exp(-KT\phi(Z_1, g_1))\mid Z_2, g_2^K]\ ] \\
&= \mathop{\mathbf{E}}_{Z_2, g_2^K}[\exp\left(-KT\sum_{n=2}^{K}\phi(Z_n, g_n)\right)\ \cdot\ \gamma_1(Z_1, Z_2)] \qquad \text{where} \\
\gamma_1(Z_1, Z_2) &\doteq \mathop{\mathbf{E}}_{g_1}[\exp(-KT\phi(Z_1, g_1))\mid Z_2, g_2^K] \\
&= \mathop{\mathbf{E}}_{g_1}[\exp(-KT\phi(Z_1, g_1))\mid Z_2] \qquad (4.29)
\end{aligned}$$

(4.29) is valid because $Z_2 = Z_2(Z_1, g_1)$, and $\{g_n\}$ are assumed to be i.i.d., and hence $\{Z_1, g_1\}$ is independent of g_2^K, conditional on Z_2. Continuing,

$$\begin{aligned}
P_{err}(\rho) &= \mathop{\mathbf{E}}_{Z_2^3, g_3^K}[\ \gamma_1(Z_1, Z_2)\ \cdot\ \exp\left(-KT\sum_{n=3}^{K}\phi(Z_n, g_n)\right)\ \cdot \\
&\qquad \mathop{\mathbf{E}}_{g_2}[\ \exp(-KT\phi(Z_2, g_2))\mid Z_2^3, g_3^K\ ]\ ] \\
&= \mathop{\mathbf{E}}_{Z_2^3, g_3^K}[\ \gamma_1(Z_1, Z_2)\ \cdot\ \gamma_2(Z_2, Z_3)\ \cdot \\
&\qquad \exp\left(-KT\sum_{n=3}^{K}\phi(Z_n, g_n)\right)\ ] \qquad \text{where} \\
\gamma_2(Z_2, Z_3) &= \mathop{\mathbf{E}}_{g_2}[\ \exp(-KT\phi(Z_2, g_2))\mid Z_2^3, g_3^K\ ] \\
&= \mathop{\mathbf{E}}_{g_2}[\ \exp(-KT\phi(Z_2, g_2))\mid Z_2, Z_3]
\end{aligned}$$

since g_2 is independent of g_3^K conditional on Z_2, Z_3. Continuing in this manner, we conclude that

$$\begin{aligned}
P_{err}(\rho) &= \mathop{\mathbf{E}}_{Z_2^{K+1}}[\ \Pi_{n=1}^{K}\ \gamma_n(Z_n, Z_{n+1})\ ] \qquad \text{where} \qquad (4.30) \\
\gamma_n(Z_n, Z_{n+1}) &= \mathop{\mathbf{E}}_{g_n}[\ \exp(-KT\phi(Z_n, g_n))\mid Z_n, Z_{n+1}]
\end{aligned}$$

Note that for the optimal power control, Z_{n+1} is the deterministic function $Z_{n+1} = Z_{n+1}(Z_n, g_n)$. Also, $Z_1 = \{KP_0, KR_0\}$. Therefore $f_G(g_n|Z_n, Z_{n+1})$, which is required in the calculation of $\gamma_n(Z_n, Z_{n+1})$, can be found as

$$f_G(g_n|Z_n, Z_{n+1}) = \frac{f(g_n) \cdot \mathbf{P}[Z_{n+1} \mid Z_n, g_n]}{\mathbf{P}[Z_{n+1} \mid Z_n]}$$

which can be calculated from the optimal power control solution.

The expectation in (4.30) is easy to compute as below

$$\begin{aligned} P_{err}(\rho) &= \mathop{\mathbf{E}}_{Z_2^K}[\ \Pi_{n=1}^{K-1}\gamma_n(Z_n, Z_{n+1}) \ \cdot \mathop{\mathbf{E}}_{Z_{K+1}}[\gamma_K(Z_K, Z_{K+1}) \mid Z_2^K]\ ] \\ &= \mathop{\mathbf{E}}_{Z_2^K}[\ \Pi_{n=1}^{K-1}\gamma_n(Z_n, Z_{n+1}) \ \cdot \mathop{\mathbf{E}}_{Z_{K+1}}[\gamma_K(Z_K, Z_{K+1}) \mid Z_K]\ ] \qquad (4.31) \\ &= \mathop{\mathbf{E}}_{Z_2^K}[\ \Pi_{n=1}^{K-1}\gamma_n(Z_n, Z_{n+1}) \ \cdot \nu_K(Z_K)\ ], \qquad \text{where} \\ \nu_K(Z_K) &= \mathop{\mathbf{E}}_{Z_{K+1}}[\gamma_K(Z_K, Z_{K+1}) \mid Z_K] \end{aligned}$$

and (4.31) occurs because Z_n is Markovian. Continuing the nested conditionals,

$$\begin{aligned} P_{err}(\rho) &= \mathop{\mathbf{E}}_{Z_2^{K-1}}[\ \Pi_{n=1}^{K-2}\gamma_n(Z_n, Z_{n+1}) \ \cdot \mathop{\mathbf{E}}_{Z_K}[\nu_K(Z_K) \cdot \gamma_{K-1}(Z_{K-1}, Z_K) \mid Z_{K-1}]\ ] \\ &= \mathop{\mathbf{E}}_{Z_2^{K-1}}[\ \Pi_{n=1}^{K-2}\gamma_n(Z_n, Z_{n+1}) \cdot \nu_{K-1}(Z_{K-1})] \\ &= \cdots \quad \cdots \quad \cdots \\ &= \mathop{\mathbf{E}}_{Z_2}[\ \gamma_1(Z_1, Z_2) \cdot \nu_2(Z_2)\ ], \qquad \text{where} \qquad (4.32) \\ \nu_n(Z_n) &= \mathop{\mathbf{E}}_{Z_{n+1}}[\ \gamma_n(Z_n, Z_{n+1}) \cdot \nu_{n+1}(Z_{n+1}) \mid Z_n\ ] \end{aligned}$$

$\nu_n(Z_n)$ requires calculation of $\mathbf{P}[Z_{n+1}|Z_n]$, which can be found from the optimal power control. (4.32) is the required result.

Summarizing, the calculation of $P_{err}(\rho)$ (defined in (4.27)) proceeds as,

Algorithm Calculate_error(ρ) :

1. Based on the optimal power control solution to either the short term or the long term constraint problem (Section 4.2) which results in a Markovian $Z_n = \{P^{(n)}, R^{(n)}\}$, calculate

$$\begin{aligned} \mathbf{P}[Z_{n+1}|Z_n] &= \int \mathbf{P}[Z_{n+1} \mid Z_n, g_n] \cdot f_G(g_n)\, \mathrm{d}g_n \qquad \text{and the densities} \\ f_G(g_n \mid Z_n, Z_{n+1}) &= \frac{f_G(g_n) \cdot \mathbf{P}[Z_{n+1} \mid Z_n, g_n]}{\mathbf{P}[Z_{n+1} \mid Z_n]} \end{aligned}$$

2. Set up the trellis of Markovian states $\{Z_n, n = 1, \ldots, K+1\}$ (see Figure 4.4). Calculate the branch metrics for the trellis stages $n = 1, \ldots, K$ as,

$$\gamma_n(Z_n, Z_{n+1}) = \mathop{\mathbf{E}}_{g_n}[\exp(-KT\phi(Z_n, g_n)) \mid Z_n, Z_{n+1}]$$

3. Calculate the node values of the trellis states for stages $n = K, \ldots, 1$ (in a backwards manner) using the dynamic program

$$\nu_n(Z_n) = \mathop{\mathbf{E}}_{Z_{n+1}}[\gamma_n(Z_n, Z_{n+1}) \cdot \nu_{n+1}(Z_{n+1}) \mid Z_n]$$

   The dynamic program is initialized as $\nu_{K+1}(Z_{K+1}) \equiv 1$.

4. The error probability desired to be calculated is

$$\begin{aligned} P_{err}(\rho) &= \nu_1(Z_1) \\ &= \nu_1(KP_0, KR_0) \end{aligned}$$

Since $P_{err}(\rho)$ is convex in ρ, it is minimized over $\{0 \leq \rho \leq 1\}$ using bisection to give the optimal value $\rho^{(opt)}$. See Section A.1 for a short description of the method of bisection. This gives the desired bound (4.24) on P_{err}.

Each calculation of **Calculate_error(ρ)** is effected in $O(K|Z||G|)$ operations, and the bisection requires $O(\log(1/\varepsilon_\rho))$ calls to the algorithm, where ε_ρ is the tolerance in ρ. This is usually much less than the $O(|G|^K)$ optimizations required by a Monte Carlo calculation of (4.18).

The error bound (4.24) may be weak (or even useless) for small T. In particular, notice that if $\frac{1}{K}\sum_{n=1}^{K}\log(1+g_nP_n) < R$, then the optimal ρ for this set of $g^{(K)}$ is $\rho = 0$. Using $\rho^{(opt)} > 0$ found by the bisection will result in an error bound for that set of $g^{(K)}$ that is greater than one! This is clearly a gross over-estimation. Appendix A.2 shows one method to reduce this over-estimation at the expense of increased complexity.

The weak random coding bound was calculated for the long term power constraint solution for the Rayleigh fading channel in the case, $R_0 = 3$ nats per transmission, $K = 5$, $SNR = 20$ dB. From Figure 4.3, the outage probability for this case is 10^{-3}. Two sets $\{\mathcal{S}_1, \mathcal{S}_2\}$ were used with $R_{th,0} = 0, R_{th,1} = KR_0 - KR, R_{th,2} = KR_0$ (See Appendix A.2). Figure 4.5 shows the random coding bound for various values of block length T_0. The dotted line is the outage probability obtained when the long term constraint solution is used with different code rates R ($R < R_0$). Outage probability assumes $T_0 \to \infty$, and is therefore the limit to which the random coding bound must converge. Note that for $R = 2.9$, the outage probability is close to 10^{-3}, as expected.

It is clear that the error bound increases with increasing R, in keeping with the increasing outage probability. Another point to note is that a smaller T_0 results in

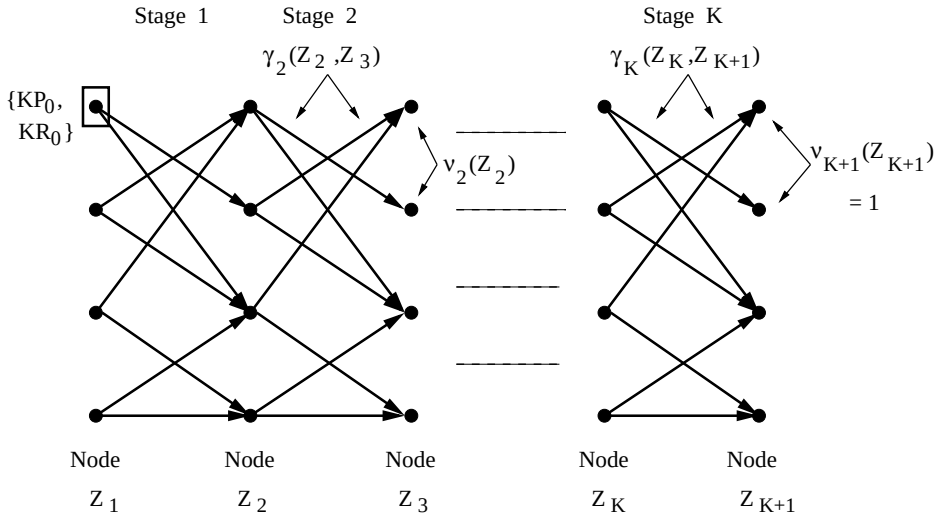

Figure 4.4: Trellis to calculate a weak random coding bound

a larger error bound, due to the higher decoding error. Figure 4.6 plots the 'error exponent', defined as $-\log(P_{err})/(KT_0)$, for various R. Larger T_0 result in a smaller exponent, because the exponent converges to 0 when the system has outage. Also, a larger R results in a larger exponent, reflecting the smaller decoding error.

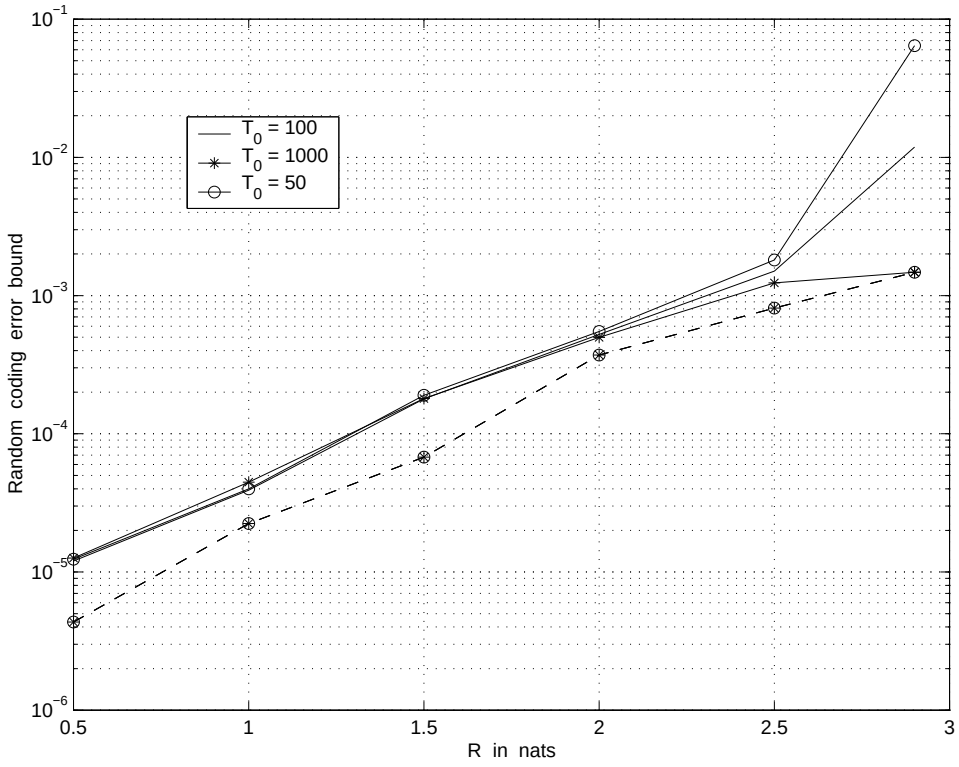

Figure 4.5: Weak random coding bound for long term constraint algorithm for $K = 5$, $SNR = 20$ dB, $R_0 = 3$. Solid lines are the random coding bound, while dotted line is the outage probability ($T_0 \to \infty$)

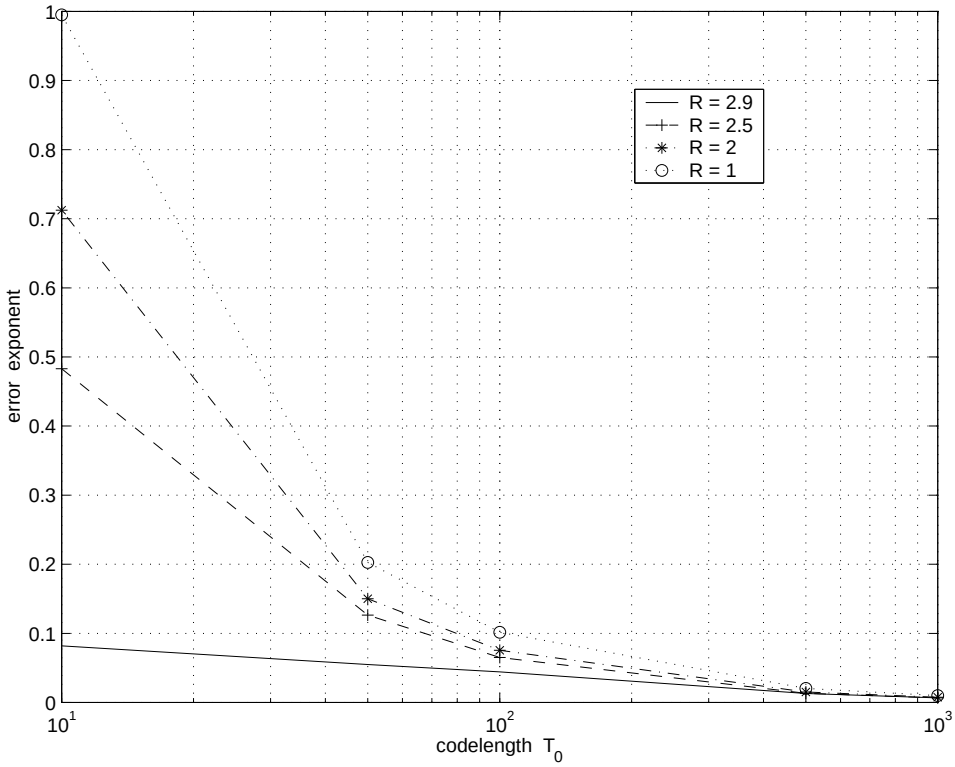

Figure 4.6: Error exponent for long term constraint algorithm for $K = 5$, $SNR = 20$ dB, $R_0 = 3$, for various values of R

Chapter 5

Stationary Power Control Strategies

The problem formulation in Chapter 4 requires formation of a 'super-frame' of K frames of data, and the power control scheme has to be different for each of the K frames within the super-frame. Some systems may not be able to introduce this additional structure (such as existing communication protocols). Also for large K, the memory requirements increase at least linearly. In this chapter, we propose a power control scheme for the outage capacity problem that does not need such a super-frame structure. The new scheme tries to achieve a target rate within the delay constraint, for any position of the delay window. Thus such a scheme is a 'stationary scheme'.

[10] also deals with a stationary power control strategy, However, the approach there is to minimize the average power transmitted, under an average delay constraint. It obtain results for the Gilbert-Elliott channel only, which limits their applicability to more general situations. However it introduces the interesting concept of a queue

of data, in the context of a fading channel.

This chapter is organized as follows. Section 5.1 introduces the problem formally. The optimum stationary power adaptation strategies are derived in Section 5.2. Two kinds of power constraints are considered, which result in different policies. Section 5.3 demonstrates the performance of these techniques under various scenarios.

5.1 Problem Formulation

As in the previous chapters, we assume a flat fading channel in AWGN, that is i.i.d. block-fading. It is assumed that the average power per transmitted symbol is P_0. Both the receiver and transmitter are assumed to have perfect knowledge of the channel gain of the past and current blocks, but not the future. We require that the system have a low probability of outage. The stationary version of the K-block outage probability problem should ideally be posed as below.

Given the sequence of channel power gains $\{g_i, i = \ldots, -1, 0, 1, \ldots\}$, and given the target rate R_0, and power P_0 per transmission

$$\begin{aligned} \text{minimize} \qquad & average\{\ \mathbf{P}[\ \sum_{i=t-K+1}^{t} \log(1+P_i g_i) < KR_0]\ \}\ \text{ such that } \quad (5.1) \\ P_i \ \geq \ & 0 \qquad \forall\ i = \ldots, -1, 0, 1, \ldots \qquad \text{and} \\ average\{P_t\} \ \leq \ & P_0 \qquad \text{(power constraint)} \end{aligned}$$

where the minimization is over the powers P_i transmitted during the i^{th} block (a causal function $P_i(g^i_{-\infty})$), and *average* refers to the time average $\lim_{s\to\infty} \frac{1}{2s+1} \sum_{t=-s}^{t=s}(\cdot)(t)$. The quantity within the $\mathbf{P}[\,\cdot\,]$ denotes an outage event. The optimum solution to this problem requires storing K state variables, which makes the solution very complex.

Therefore we deal with a modified problem, where the K-block window is approximately captured by an exponential window, as shown below.

Mathematically the modified problem is,

$$\begin{aligned} \text{minimize} \qquad & average\{\ \mathbf{P}[\ \sum_{i=-\infty}^{t} e^{-\alpha(t-i)} \log(1+P_i g_i) < R_0]\ \} \qquad \text{such that} \\ P_i \ &\geq\ 0 \qquad \forall\ \ i = \ldots, -1, 0, 1, \ldots \qquad \text{and} \\ average\{P_t\} \ &\leq\ P_0 \qquad \text{(power constraint)} \end{aligned} \tag{5.2}$$

where the various variables are as defined above, and α is a positive scalar. With this problem formulation, the delay-constraint is captured by the 'exponential window' $e^{-\alpha(\cdot)}$ used to discount the transmission rates $\log(1+Pg)$ of the past blocks. The smaller α is, the larger the allowable delay. The stationarity of the scheme is captured by the fact that the *average* probability (averaged over t) is minimized.

5.2 Optimal Solution

The optimum strategies can be found using dynamic programming [51].

5.2.1 Optimal scheme

The problem posed earlier is called an average reward problem in dynamic program theory. The solution of the problem is obtained by solving a specific linear program. The optimal solution is a *randomized* solution. i.e. the power transmitted at a given state is not deterministic. The solution can be precomputed and stored in the transmitter and receiver. When transmission actually occurs, the transmitter simply chooses the right power using a table lookup.

Whereas the theoretical basis for using a linear program to obtain the solution can be found in [51], it can also be understood intuitively. We begin by defining

$$\begin{aligned} R_t &= \sum_{i=-\infty}^{t} e^{-\alpha(t-i)} \log(1 + P_i g_i) \\ s_t &= \{R_{t-1}, g_t\} \\ c(s_t, P_t) &= \mathbf{1}_F(e^{-\alpha} R_{t-1} + \log(1 + P_t g_t) \geq R_0) \end{aligned}$$

R_t is the discounted rate up to time t, s_t is the state variable at time t (Conversely, its two elements are written as $R(s_t)$ and $g(s_t)$), $c(s_t, P_t)$ is a 'reward function', and $\mathbf{1}_F(\cdot)$ is the indicator function that is 1 if the event $(\cdot)$ is true and 0 otherwise. Note that

$$R_t = e^{-\alpha} R_{t-1} + \log(1 + P_t g_t)$$

Thus, we can find an optimal policy P_t^* that depends only on s_t. i.e. $P_t^* = P_t^*(s_t)$. We discretize the variables R_t, P_t, g_t, so that numerical methods can be used. Denote the sets of discretized R_t, P_t, g_t as B_R, B_P, B_g respectively. Denote $B_s = B_R \times B_g$. The optimization can now be expressed by the linear program,

$$\begin{aligned} \text{Maximize} \quad & \sum_{s \epsilon B_s} \sum_{P \epsilon B_P} c(s, P) x(s, P) \qquad \text{such that} \\ \sum_{P \epsilon B_P} x(j, P) &= \sum_{s \epsilon B_s} \sum_{P \epsilon B_P} \mathbf{P}[j \mid s, P] \cdot x(s, P) \qquad \forall \ \ j \epsilon B_s \end{aligned}$$

$$\begin{aligned} x(s,P) &\geq 0 \qquad \forall \ s\epsilon B_s \ , \ P\epsilon B_P \\ \sum_{s\epsilon B_s}\sum_{P\epsilon B_P} x(s,P) &= 1 \\ \sum_{s\epsilon B_s}\sum_{P\epsilon B_P} P\cdot x(s,P) &\leq P_0 \end{aligned} \tag{5.3}$$

$x(s,P)$ can be thought of as the probability that the Markov chain will be in state s and choose power P to transmit in that block. Then, the linear program maximizes the probability that the discounted rate R_t exceeds R_0. The constraints are easy to understand. Note in particular that the power constraint is expressed as in (5.3). The transition probability that the Markov chain in state s and choosing power P will transition to state j is given by

$$\begin{aligned} \mathbf{P}[j|s,P] &= \mathbf{P}[R(j),g(j) \mid R(s),g(s),P] \\ &= \begin{cases} \mathbf{P}[g=g(j)] & \text{if} \quad \log(1+Pg(s))+e^{-\alpha}R(s)=R(j) \\ 0 & \text{else} \end{cases} \end{aligned}$$

Solving the linear program for the unknown variables $x(s,P)$ results in the optimum power control strategy. To obtain the strategy $P_t^* = P_t^*(s)$, we note that if for a fixed s the optimum $x(s,P)=0$, then for those P whenever $s_t = s$, we never transmit $P_t = P$. In general, the optimum power control will be randomized. It will be deterministic for those s, for which only one P results in a non-zero $x(s,P)$. So, the power control can be defined as follows. If $s_t = s$, then transmit $P_t = Q$ with probability $x(s,Q)/(\sum_P x(s,P))$. As the discretization increases, we would expect the optimum strategy to become completely deterministic. The linear program has $O(|B_R|\cdot|B_g|\cdot|B_P|)$ variables, and $O(|B_R|\cdot|B_g|)$ equations.

5.2.2 Sub-optimal scheme

A simpler scheme can be obtained by artificially adding more power constraints. This results in a scheme with less computational complexity, but with worse performance than the optimum scheme. In this scheme, the power constraint (5.2) is modified to the constraint

$$\begin{aligned} \underset{g_t \epsilon B_g}{\mathbf{E}} [P_t(s_t)] &= \underset{g_t \epsilon B_g}{\mathbf{E}} [P_t(R_{t-1}, g_t)] \\ &\leq P_0 \qquad \forall \;\; R_{t-1} \,\epsilon\, B_R \end{aligned}$$

With this modification, the average reward problem can be solved using the method of policy iteration [51]. This involves finding a value function $V(R)$ and scalar c such that

$$\begin{aligned} V(R) + c &= \max \;\; \mathbf{P}[\; e^{-\alpha} R + \log(1 + P(g)g) \geq R_0 \;] \\ &\quad + \underset{g \epsilon B_g}{\mathbf{E}} [\; V(e^{-\alpha} R + \log(1 + P(g)g) \;] \qquad \forall \;\; R \,\epsilon\, B_R \end{aligned}$$

where the maximization is over all functions $P(g) \geq 0$ such that $\underset{g \epsilon B_g}{\mathbf{E}} [P(g)] \leq P_0$. Policy iteration involves optimization and solving linear equations, done alternately. These have a complexity of $O(|B_R| \cdot |B_g| \cdot |B_P|^2)$, and $O(|B_R|^3)$ respectively.

5.3 Simulation and Discussion

We demonstrate the capacity improvement afforded by the online power adaptation scheme over a constant power scheme in a Rayleigh fading channel. Figure 5.1 shows the outage probability achieved at various SNRs, $R_0 = 4$ nats, and $\alpha = 0.3$. An SNR gain of around $2 - 3$ dB is obtained over the constant power scheme, by adapting the

power.

Further, the sub-optimal scheme has a small penalty over the optimal scheme. A similar improvement is seen at low SNRs, such as those occurring in CDMA (Figure 5.2). Increasing the discretization for the various parameters increases the SNR gain achieved to some extent. An interesting observation comes from Figure 5.3 which shows the histograms of the state R_t for a sample run. The upper histogram is for the constant power scheme, while the lower histogram is obtained by using the (suboptimal) power adaptation scheme. It can be observed that the power adaptation tends to bunch the discounted rate R_t into discontinuous sets, unlike the constant power scheme. Also, the power adaptation reduces the variance of R_t.

A true 'K-block sliding window' approach would require K state variables, instead of the $\{R_t, g_t\}$ variables in our formulation. That is because the power would depend on the past K gains. By using an exponential window, rather than the rectangular window that the K-block approach demands, we have succeeded in reducing the complexity of the algorithms. A better approximation to the K-block formulation would be to use two exponentials for the window,

$$\begin{aligned} \text{minimize} \qquad & average\{\ \mathbf{P}[\ \sum_{i=-\infty}^{t} \beta_1 e^{-\alpha_1 (t-i)} \log(1 + P_i g_i) + \\ & \beta_2 e^{-\alpha_2 (t-i)} \log(1 + P_i g_i) < R_0]\ \} \end{aligned}$$

By choosing $\alpha_1, \alpha_2, \beta_1, \beta_2$ appropriately, we can approximate a rectangular window.

This formulation requires two state variables besides g_t as below

$$R_t^{(1)} = \sum_{i=-\infty}^{t} e^{-\alpha_1(t-i)} \log(1 + P_i g_i)$$

$$R_t^{(2)} = \sum_{i=-\infty}^{t} e^{-\alpha_2(t-i)} \log(1 + P_i g_i)$$

We can approximate the K-block formulation better, using even more exponentials to design the window, at the cost of increased complexity.

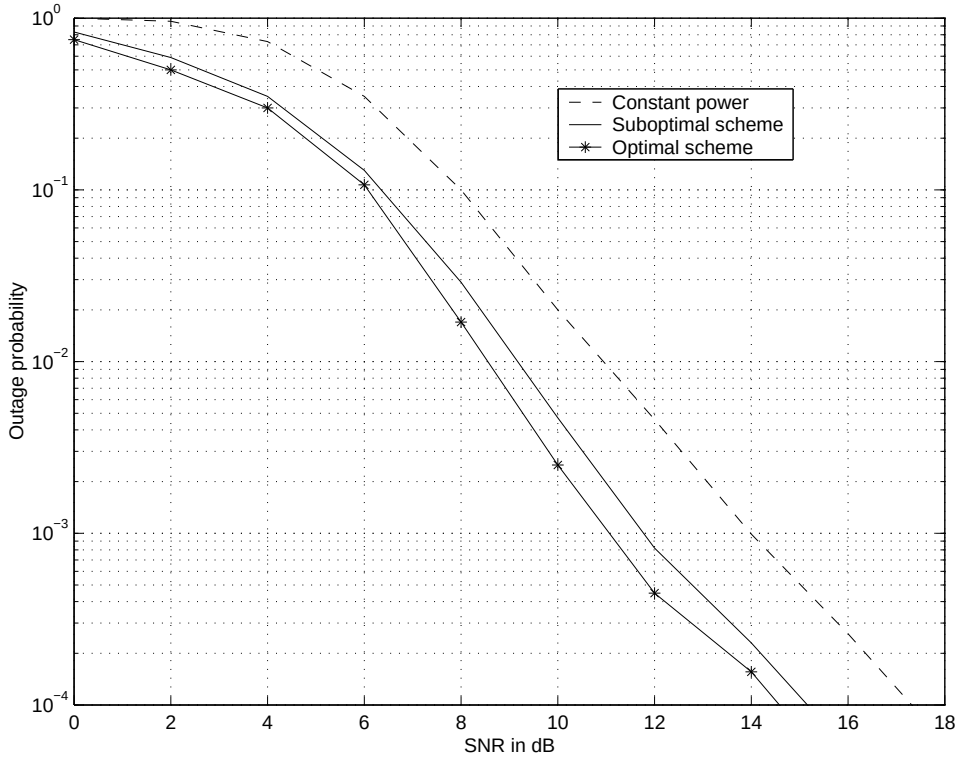

Figure 5.1: Outage probability at high SNRs for $R_0 = 4$, $\alpha = 0.3$. Solid lines refer to the new algorithms, while the dotted line refers to constant power transmission

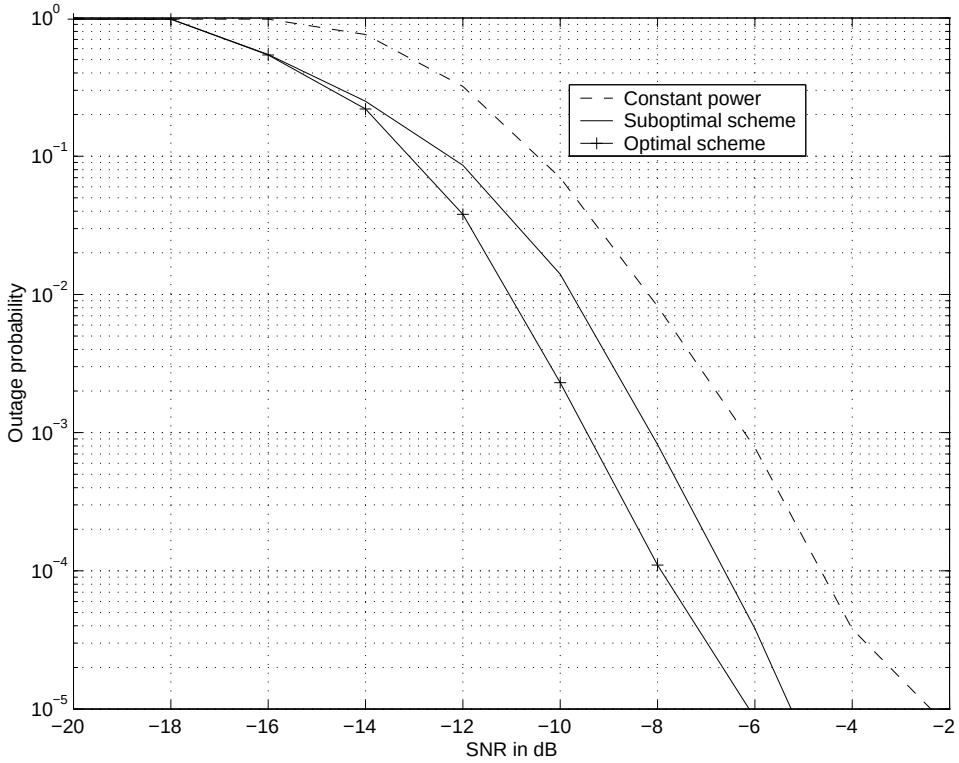

Figure 5.2: Outage probability at low SNRs for $R_0 = 0.15$, $\alpha = 0.3$. Solid lines refer to the new algorithms, while the dotted line refer to constant power transmission

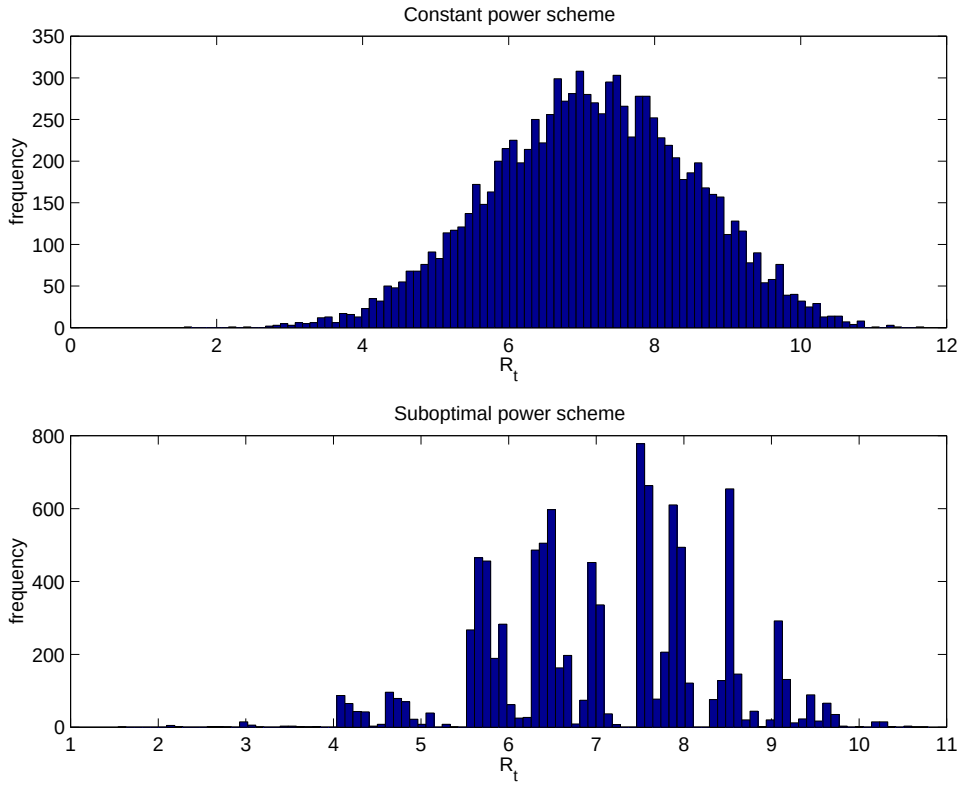

Figure 5.3: Histogram of state R_t for a sample run, at $SNR = 10$ dB, $R_0 = 4$, $\alpha = 0.3$

Chapter 6

Adaptive Antennas for Space-Time Codes in Outdoor Channels

6.1 Introduction

Space-time coding is a direct example of the use of the outage capacity measure. Space-time codes can be used in transmission systems that have multiple transmit antennas, in environments where the channel undergoes Rayleigh fading. The design of these codes involves using the minimum error probability performance metric, which should converge to the outage probability as the codes become more powerful. However a union bound argument is used for the error probability, and hence the design is really based on minimizing the (union) error bound. Thus, strictly speaking the error probability may not converge to the outage probability as the codes become more powerful. Yet the the design of these codes retains an outage capacity flavor.

Space-time coding for the general case of Multiple-Input Multiple-Output (MIMO)

channels has been studied in [61, 52, 64, 63, 60] and space-time codes have been recently introduced in multipath fading environments [8, 44, 37, 55, 13] to improve mobile system performance. In the block time-invariant environment (where channel is time-invariant during transmission of one block of data) [56], it has been shown that using multiple antennas at the base station and mobile allows one to achieve a maximum diversity of mn, where m and n are the number of mobile and base station antennas respectively. Delay diversity codes were shown to be special cases of space-time codes, that were capable of achieving maximum diversity. However, all the work to date in this subject assumes a multipath channel model in which the fading from each base station antenna to any mobile antenna is independent, or at least non-degenerate. i.e. If we collect the mn fading gains into a vector, then the auto-covariance matrix of the vector is full rank. This assumption requires that both the base station and the mobile be surrounded by local scatterers, or that the base station antennas must be spaced far apart, so that their signals are uncorrelated. In several situations, the base station is placed high above the ground, and practicalities dictate that its antennas are spaced close together. In this situation, the base antenna array will be a phased array. i.e. have completely correlated signals. In this case, an L-multipath channel model such as the one proposed in [52] would be valid. In this chapter, we will consider the above scenario (Figure 6.1). The mobile, being at ground level, is assumed to have a diversity array. Thus, the fades of each of the L-multipaths at the base are completely correlated, but those at the mobile are uncorrelated. The case that assumes complete correlation at the mobile also, would give essentially the same results as those derived in this chapter. We will discuss schemes that would achieve diversity as well as SNR gain in this scenario. We will only consider the downlink case (Base station to mobile link) in this chapter. The

uplink would require a different approach, which we believe would be more along the lines of classical beamforming. We have considered two distinct cases. The first is the case in which all the the multipaths have the same delay, which results in a channel that is free of Inter-Symbol Interference (ISI). In the second case, the multipaths have different delays, causing ISI in the receiver. We have published some of these results in in [43]. The results of this chapter are joint work with Ardavan Maleki-Tehrani.

The chapter is organized as follows. The problem is formulated for the ISI-free case in Section 6.2. In Section 6.3 we show that for the ISI-free case, an appropriate scheme for achieving the maximum diversity of mL, and obtaining an SNR gain, is to combine beamforming with a space-time code such as a delay diversity code. The optimum beamformer is derived for this case and is seen to be very different from the classical beamformer. In section 6.4 we formulate the problem for the case with ISI, and show that whereas using more than one base antenna does not increase the diversity, it does provide a large SNR gain. We will provide suboptimal results for the SNR gain solution. Section 6.5 illustrates these ideas with simulations.

6.2 Problem Formulation

The problem is to design 'space-time coding schemes' [54] that achieve low frame error rates. In various papers [8, 16, 17, 54, 56, 55], space-time coding schemes have been discussed that essentially assume that the elements of the channel matrix **FA** (see below) are non-degenerate, even if they are dependent. In the scenario we consider however, this is not true. Further, for our case, we show explicitly how the concepts of beamforming and space-time coding can be combined. It will be shown that such a combination can provide diversity as well as SNR gain.

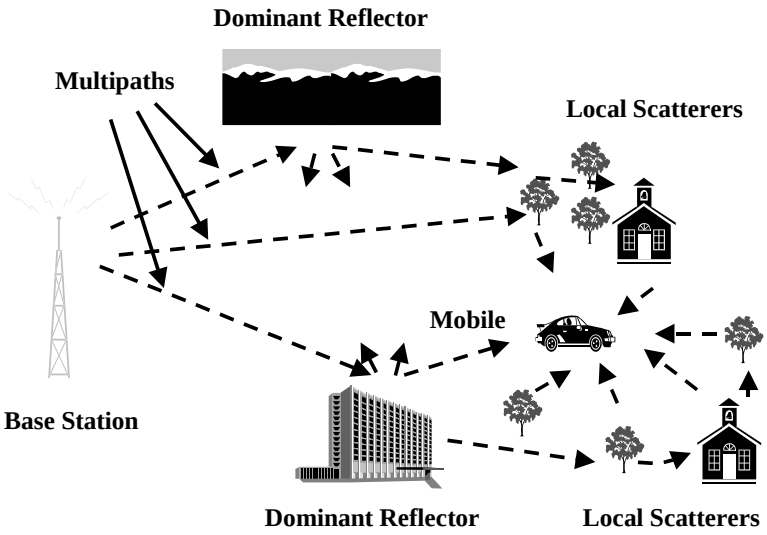

Figure 6.1: Illustration of the physical multipath channel

6.2.1 MIMO Wireless System

The downlink system equation in the time domain can be written as

$$\mathbf{y}_k = \sum_{p=1}^{L} \boldsymbol{\alpha}_p \cdot \mathbf{a}(\theta_p) \cdot \mathbf{s}(kT - \tau_p) + \mathbf{v}_k \tag{6.1}$$

where T is the symbol period, L is the number of multipaths, τ_p is the time delay associated with the pth multipath. i, k denote discrete time, while t denotes continuous time. $\mathbf{s}(t)$ is the $n \times 1$ signal vector waveform transmitted, using the n base station antennas, $\mathbf{y}_k$ and $\mathbf{v}_k$ are the $m \times 1$ vector of the signal and noise respectively, received using the m mobile antennas. $\mathbf{s}(t)$ can be written in terms of the data symbol vectors

$\mathbf{x}_i$ as

$$\mathbf{s}(t) = \sum_{i=-\infty}^{\infty} \mathbf{x}_i \cdot g(t - iT) \tag{6.2}$$

$g(t)$ is the pulse response of the transmit signal. Assume that $g(t)$ is normalized as $g(0) = 1$.

$\boldsymbol{\alpha}_p$ is the $m \times 1$ vector of the fading channel gains at the m mobile antennas, for the pth multipath, while $\mathbf{a}(\theta_p)$ is the $1 \times n$ vector of the base station antenna array response to the pth multipath, that is incident on the base array at an angle θ_p. For example, for a linear array: $\mathbf{a}(\theta) = (1 \quad e^{j2\pi\delta sin(\theta)} \quad \dots \quad e^{j2\pi\delta(n-1)sin(\theta)})$

6.2.2 Assumptions

The system model (6.1) is valid because it is assumed that the base station does not have any local scatterers and that the base antennas are spaced close together, and hence the base array is a phased array (i.e. completely correlated response at all antennas for each multipath). Also, the mobile is assumed to be at ground level, so that the local scatterers cause uncorrelated fading at its antennas. Further, the L multipaths are assumed uncorrelated and having equal power, since they're presumably reflected by objects well separated in space. Thus, the fading channel gains $\boldsymbol{\alpha}_p$s have i.i.d. elements, which are assumed complex Gaussian random variables of variance σ_f^2 each. The angles $\{\theta_p\}$ are assumed to be distinct. Note that if the power of the L multipaths is measured and unequal, this can be easily incorporated in the theory that follows, by absorbing the powers in the $\mathbf{a}(\theta_p)$s . This trivial case is not considered for simplicity of presentation. The noise vector $\mathbf{v}_k$ is assumed to consist of zero mean i.i.d. complex Gaussian random variables (white noise) of variance σ_n^2 each. The transmitter is assumed to have knowledge of the multipath angles $\{\theta_p\}$,

since these are expected to change slowly, and so can be estimated. However, it does not know the fading channel gains $\boldsymbol{\alpha}_p$s due to the high doppler rate. The mobile is assumed to have knowledge of the entire channel state information (CSI). The α_ps are assumed to remain constant for several symbol intervals (block-fading model).

6.3 ISI-free block time-invariant MIMO system

Consider a Multiple-Input Multiple-Output (MIMO) system with L multipaths, all of which have the same delay $\tau = 0$. In that case, there is no ISI, and $\mathbf{s}(t) = \sum_{k=-\infty}^{\infty} \mathbf{x}_k \cdot \delta(t - kT)$. The downlink system equation in the time domain can be written as

$$\mathbf{y}_k = \sum_{p=1}^{L} \boldsymbol{\alpha}_p \mathbf{a}(\theta_p)\mathbf{x}_k + \mathbf{v}_k \tag{6.3}$$

All notations are as defined in Section 6.2.1. Consider a block transmission scheme, wherein a frame of data consisting of a sequence of l vector symbols $\{\mathbf{x}_k, k = 0, 1, \ldots, l-1\}$ is transmitted. The channel is assumed to be constant during this interval. We can write the block transmission in matrix form as below

$$\mathbf{Y} = \mathbf{FAX} + \mathbf{V} \tag{6.4}$$

where $\mathbf{Y} \doteq (\mathbf{y}_0 \ \cdots \ \mathbf{y}_{l-1})$, $\mathbf{X} \doteq (\mathbf{x}_0 \ \cdots \ \mathbf{x}_{l-1})$, $\mathbf{F} \doteq (\boldsymbol{\alpha}_1 \ \cdots \ \boldsymbol{\alpha}_L)$, and $\mathbf{A} \doteq \left(\mathbf{a}(\theta_1)^T \ \cdots \ \mathbf{a}(\theta_L)^T\right)^T$. Since the fading is assumed constant over a frame, system equation (6.4) is valid. The performance criterion is obtained in a manner similar to [54]. Since the mobile knows the ideal CSI, and the noise is assumed white Gaussian, hence the probability of the decoder deciding in favor of code matrix $\mathbf{X} = \mathbf{X}_e$

when in fact the code matrix $\mathbf{X} = \mathbf{X}_0$ was transmitted (i.e. matrix $\mathbf{X}$ in (6.4)) is approximated by

$$P(\mathbf{X}_0 \rightarrow \mathbf{X}_e|\mathbf{FA}) \leq \exp(-||\mathbf{FA}(\mathbf{X}_0 - \mathbf{X}_e)||_F^2/4\sigma_n^2) \tag{6.5}$$

Now, if $n = 1$ (Single-Input Multiple-Output case), then $\mathbf{FA}$ collapses into a column vector of m independent random variables, and it is clear that a diversity of m is achievable. Whenever $n \geq L$, a maximum diversity of mL is achievable. This is because we can write

$$||\mathbf{FA}(\mathbf{X}_0 - \mathbf{X}_e)||_F^2 \quad = \quad \mathrm{tr}[\mathbf{FV\Lambda V}^*\mathbf{F}^*]$$

where $\mathbf{A}(\mathbf{X}_0 - \mathbf{X}_e)(\mathbf{X}_0 - \mathbf{X}_e)^*\mathbf{A}^* = \mathbf{V\Lambda V}^*$ is the singular value decomposition. $\mathbf{FV}$ is an $m \times L$ matrix with i.i.d. complex Gaussian elements. If one designs the codebook such that for every pair of codewords, $\mathbf{V\Lambda V}^*$ has full rank L, then the exponent in (6.5) is a χ^2 random variable with $2mL$ degrees of freedom ([48]) and hence the diversity gain is mL. This is possible if and only if $n \geq L$.

We will assume subsequently that a diversity gain of mL is always the target. In that case, it is easy to show that $n = L$ and even a simple delay-diversity code achieves the diversity of mL.

Now the question arises as to the benefit of using more than L antennas at the base station. We can show that by using an appropriate concept of beamforming at the base, we can get an SNR gain over the system that uses only L antennas. Both systems however, have the same diversity gain of mL, since that's the maximum achievable. However when using more than L transmit antennas the SNR advantage is not insignificant, especially when the target frame error rate is high. Hence, the

advantage of using $n > L$. In the following, we derive the beamforming concept that is applicable in this case.

When $n > L$, to achieve a diversity of mL, we begin with a diversity-L achieving space-time code (such as a delay-diversity code) as the core code. Call the $L \times l$ Toeplitz code matrix of this code as $\mathbf{C}$. Now we map the $L \times 1$ vector symbol at each transmission (i.e. a column of $\mathbf{C}$) into an $n \times 1$ transmit vector using a linear transform represented by the $n \times L$ matrix $\mathbf{W}$. Thus, the final code matrix transmitted from the n antennas is $\mathbf{X} = \mathbf{WC}$. If (and only if) $\mathbf{W}$ is chosen such that the product $\mathbf{AW}$ is full rank, then this coding scheme will achieve a diversity of mL. Now we optimize $\mathbf{W}$ so as to get the largest SNR gain possible. To this end, note that the SNR gain is maximized by maximizing the determinant [54]

$$\det[\mathbf{AW}(\mathbf{C}_0 - \mathbf{C}_e)(\mathbf{C}_0 - \mathbf{C}_e)^*\mathbf{W}^*\mathbf{A}^*] \tag{6.6}$$

for any pair of code matrices $\{\mathbf{C}_0, \mathbf{C}_e\}$. Since the codewords have already been chosen, we can take $\det[(\mathbf{C}_0 - \mathbf{C}_e)(\mathbf{C}_0 - \mathbf{C}_e)^*]$ as a constant. Therefore, the optimization problem reduces to

$$\begin{aligned} \underset{\mathbf{W}}{\text{maximize}} \quad & \det[\mathbf{AWW}^*\mathbf{A}^*] \\ \text{subject to} \quad & \text{tr}[\mathbf{WW}^*] = L \end{aligned} \tag{6.7}$$

The constraint (6.7) arises due to total transmitted power constraint
$\text{tr}[\mathbf{W}\ E[\mathbf{c}_k\mathbf{c}_k^*]\ \mathbf{W}^*] = P$, which occurs because the space-time code used by the L base antenna system allots power P/L to each antenna.

The solution to the maximization problem in (6.7) is found using standard linear algebra (see Appendix A.3) to be $\mathbf{W} = \mathbf{Q}_+$ where $\mathbf{Q}_+$ is found from the SVD of $\mathbf{A}$

as:

$$\mathbf{A} \;=\; \mathbf{T}\boldsymbol{\Sigma}\boldsymbol{Q}^* \;=\; \mathbf{T}\;\;(\boldsymbol{\Sigma}_+ \mid \mathbf{0})\;\;(\underbrace{\mathbf{Q}_+}_{n\times L} \mid \square\,)^*$$

and the value of the maximum is $g(n,L) = (\det[\boldsymbol{\Sigma}_+])^2$. The SNR gain when using an n antenna system, over an L antenna one, is given by $\{g(n,L)/g(L,L)\}^{1/L}$. It is clear that this is more than 1.

In fact, for the linear array case where n/L is an integer, we can (somewhat) easily show the lower bound to be

$$\{g(n,L)/g(L,L)\}^{1/L} \;\geq\; \frac{n}{L} \tag{6.8}$$

See Appendix A.4 for the proof.

6.4 MIMO Wireless System with ISI

We now develop the case where the MIMO channel has ISI. We shall see that the issues that come up in this case are qualitatively different from the case without ISI. Using (6.1), the multipath channel with ISI can be modeled as below:

$$\mathbf{y}_k = \sum_{p=1}^{L}\sum_{i=-l_h}^{l_h} \boldsymbol{\alpha}_p \cdot \mathbf{a}(\theta_p) \cdot \mathbf{x}_{k-i} \cdot g(iT - \tau_p) + \mathbf{v}_k$$

where the pulse response is assumed to be of finite length less than $2l_h + 1 + max_p\ 2\tau_p$. All notations are as defined in Section 6.2.1. Except for the introduction of ISI, all other assumptions in Section 6.2.2 still hold. As in Section 6.3, if we consider a block of l vector symbols $\{\mathbf{x}_k, k = 0, 1, \ldots, l-1\}$ transmitted, then the channel equation

can be written in matrix form as below:

$$\mathbf{y}_k = \mathbf{F} \cdot \begin{pmatrix} g^1_{-l_h}\ \mathbf{a}(\theta_1) & \cdots & g^1_0\ \mathbf{a}(\theta_1) & \cdots & g^1_{l_h}\ \mathbf{a}(\theta_1) \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ g^L_{-l_h}\ \mathbf{a}(\theta_L) & \cdots & g^L_0\ \mathbf{a}(\theta_L) & \cdots & g^L_{l_h}\ \mathbf{a}(\theta_L) \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x}_{k+l_h} \\ \vdots \\ \mathbf{x}_k \\ \vdots \\ \mathbf{x}_{k-l_h} \end{pmatrix} + \mathbf{v}_k \quad (6.9)$$

where for ease of notation, we define $g^p_i = g(iT - \tau_p)$. The $\{x_i\}$ are assumed to be zero outside the interval $\{k = 0, 1 \ldots, l-1\}$. As earlier, $\mathbf{F} \doteq (\boldsymbol{\alpha}_1 \ \cdots \ \boldsymbol{\alpha}_L)$. If $\{\tau_1, \tau_2, \ldots, \tau_L\}$are all distinct, both SIMO and MIMO will have the same diversity mL, but MIMO potentially will have higher SNR gain, which depends upon $\{\theta_1, \theta_2, \ldots, \theta_L\}$. e.g. If $\{\theta_1 = \theta_2 = \cdots = \theta_L\}$, then the SNR gain would be n. This is the case we consider here. Assume that a diversity of mL is the target here.

Now suppose we use a powerful code for the SIMO system that achieves diversity mL (this implies that $\mathbf{U}_{SISO}$ defined below has full rank) that has the $L \times l$ code matrix

$$\mathbf{U}_{SISO} = \begin{pmatrix} x_{l_h} & x_{l_h+1} & \cdots & x_{l_h+l-1} \\ \vdots & \vdots & \cdots & \vdots \\ x_{-l_h} & x_{-l_h+1} & \cdots & x_{-l_h+l-1} \end{pmatrix}$$

Then we can use the same code for the MIMO case, by mapping the scalar input x_k to the $n \times 1$ vector $\mathbf{x}_k = \mathbf{b}\ x_k$ using the beamforming vector $\mathbf{b}$. This results in the following simplification of (6.9):

$$\mathbf{Y} = \mathbf{FBGU}_{SISO} + \mathbf{V} \quad (6.10)$$

where, as earlier $\mathbf{Y} \doteq (\mathbf{y}_0 \ \cdots \ \mathbf{y}_{l-1})$, $\mathbf{B} \doteq \mathbf{diag}\,(\mathbf{a}(\theta_1)\mathbf{b}, \cdots, \mathbf{a}(\theta_L)\mathbf{b})$, and $\mathbf{G}$ is the

pulse shaping matrix

$$\mathbf{G} = \begin{pmatrix} g^1_{-l_h} & \cdots & g^1_0 & \cdots & g^1_{l_h} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ g^L_{-l_h} & \cdots & g^L_0 & \cdots & g^L_{l_h} \end{pmatrix}$$

Equation (6.10) allows us to maximize the SNR gain by maximizing the determinant [54]

$$\det[\mathbf{BG}(\mathbf{U}^0_{SISO} - \mathbf{U}^e_{SISO})(\mathbf{U}^0_{SISO} - \mathbf{U}^e_{SISO})^*\mathbf{G}^*\mathbf{B}^*] \tag{6.11}$$

$\mathbf{U}^0_{SISO}$ and $\mathbf{U}^e_{SISO}$ refer to a pair of distinct codeword matrices of the SISO code. Assuming that $\mathbf{G}(\mathbf{U}^0_{SISO} - \mathbf{U}^e_{SISO})(\mathbf{U}^0_{SISO} - \mathbf{U}^e_{SISO})^*\mathbf{G}^*$ is full rank (otherwise even the SISO code will not achieve diversity mL), and noting that only $\mathbf{B}$ depends on the weight vector $\mathbf{b}$, we see that we need to maximize $\det(\mathbf{BB}^*)$. Thus, for optimum performance, we need to solve the following

$$\begin{aligned} \underset{\mathbf{b}}{\text{maximize}} \quad P(\mathbf{b}) &\doteq \left(\prod_{p=1}^{L} \mid \mathbf{a}(\theta_p)\mathbf{b} \mid^2 \right)^{1/L} \\ \text{subject to} \quad & \parallel \mathbf{b} \parallel^2 = 1 \end{aligned} \tag{6.12}$$

The normalization $\parallel \mathbf{b} \parallel^2 = 1$ ensures that the transmitted power is the same as the SIMO case. The SNR gain over the SIMO case is simply $P(\mathbf{b}_{opt})$.

The maximization problem in (6.12) has a non-convex cost function, which seems difficult to solve exactly. However, a sub-optimal solution for $\mathbf{b}$ is given by:

$$\mathbf{b} = \frac{\sum_{p=1}^{L} \pm \mathbf{a}^*(\theta_p)}{\parallel \sum_{p=1}^{L} \pm \mathbf{a}^*(\theta_p) \parallel} \tag{6.13}$$

where the $\pm$ sign indicates that we choose that sign for each $\mathbf{a}^*(\theta_p)$, such that $P(\mathbf{b})$

is maximized. In simulations, we search over all 2^L sign combinations of $\mathbf{a}^*(\theta_p)$s, and choose the one that leads to the maximum gain $P(\mathbf{b})$.
It is clear that the pulse response does not affect the SNR gain. If the multipaths are well-separated in time, then $\mathbf{G}$ is 'more orthogonal', and the SNR gain is larger for *both* the SIMO and the MIMO cases.

Note the interesting feature of our scheme: we are able to separate the signal processing and the coding aspects of the problem. In other words, one can choose a fixed SISO code (over time only), and adapt it to the multiple transmit antenna environment (over space only), using the weight vector $\mathbf{b}$.

6.5 Simulation Results

Since the results obtained for the ISI and ISI-free cases are so different qualitatively, we report the simulation results for each case separately.

6.5.1 The ISI-free Case

The ideas described in Section 6.3 were tested by simulating a MIMO system with $m = 2, L = 2$ and different values of n. The base station was assumed to use a linear array with antenna spacing $d = 0.5$ wavelengths.

In Figure 6.2, we show the histograms of the SNR gains $g(n, L)^{1/L}$, $g(L, L)^{1/L}$ and $\{g(n, L)/g(L, L)\}^{1/L}$ for $n = 8$, when the two multipath arrive randomly within an angle of 60^o of each other. In the histograms, the X-axis represents gain in dB units, whereas the Y-axis represents the frequency with which the gain occurs. It is clear from the figure that $\{g(n, L)/g(L, L)\}^{1/L} \geq n/L$. In fact, it is usually much greater than the lower bound. The factor of n/L can be understood as the gain

in classical beamforming due to better directivity of the antenna array. However, a second effect occurs wherein the n antenna array is better able to resolve the L multipaths than the L antenna array (i.e. the $\mathbf{A}$ matrix is 'more' orthogonal). This further increases the SNR gain. In fact, the histogram of $g(n,L)^{1/L}$ is seen to have a lower variance than that of $g(L,L)^{1/L}$, due to the same reason. Thus the n antenna array 'better guarantees' a given (large) SNR gain than the L antenna array.

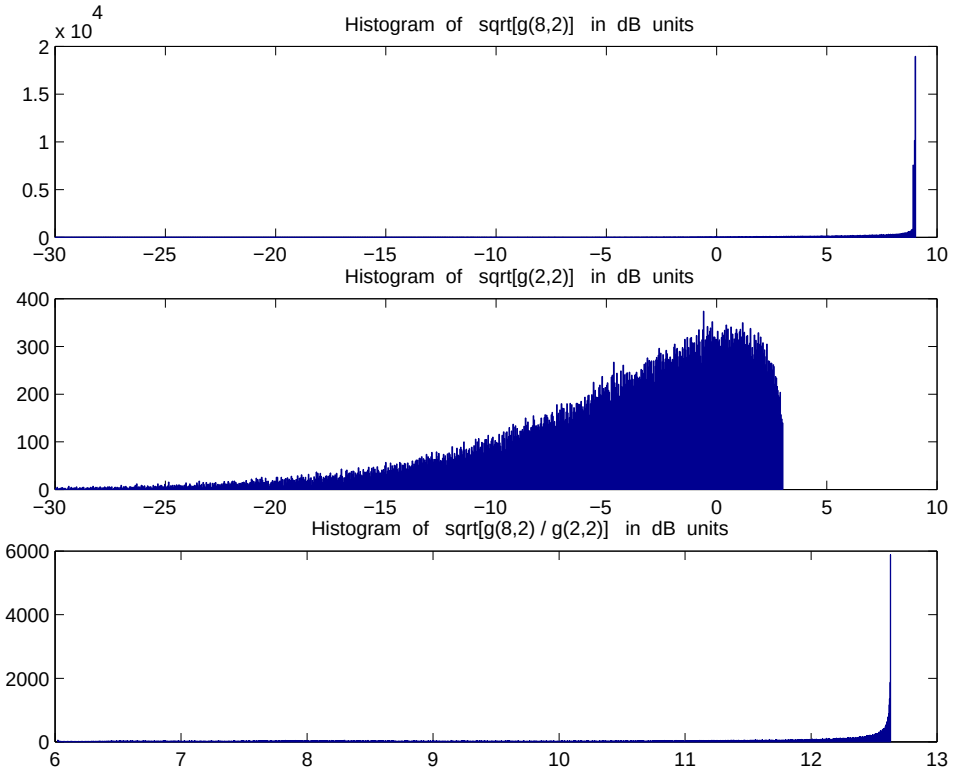

Figure 6.2: Histograms $\sqrt{g(8,L)}$, $\sqrt{g(L,L)}$, $\sqrt{g(8,L)/g(L,L)}$, $L=2$

The entire MIMO transmission system was simulated by transmitting and maximum-likelihood (Viterbi) decoding, using the delay diversity code with a QPSK signal constellation. The frame length was chosen as 100, and $10,000$ frames were transmitted.

Two different multipath angle pairs, $\{50^o, 60^o\}$ and $\{20^o, 60^o\}$, were chosen to illustrate the performance. Figures 6.3 and 6.4 show the frame error rate (FER) as a function of the SNR for the cases $n = 2, 4, 8$. SNR is defined as $SNR = L\mathcal{E}_x\sigma_f^2/\sigma_n^2$, where $\mathcal{E}_x$ is the transmitted signal energy per transmission. Thus SNR is the SNR per mobile antenna per symbol, assuming omnidirectional transmission.

It can be seen that the SNR gain allows better performance at a given SNR when n increases. The SNR gain is significant, especially in the case of the angle pair $\{50^o, 60^o\}$, where the multipaths are not well separated in space. This again illustrates the idea that using a higher n helps to resolve multipath better.

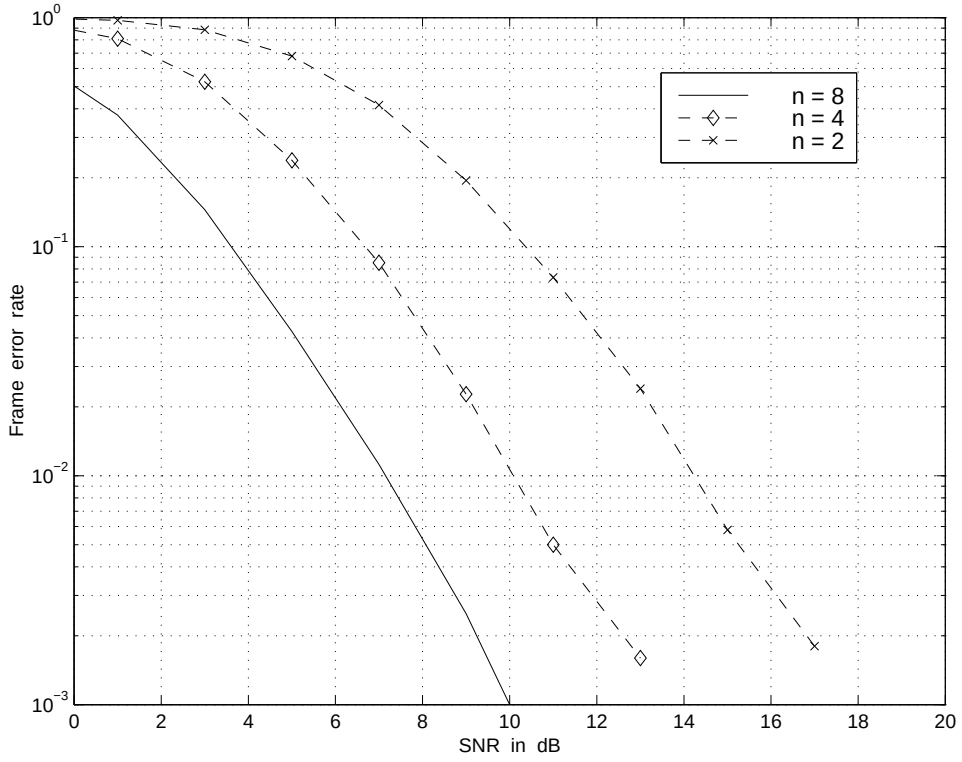

Figure 6.3: FER v.s. SNR; ISI-free case; angle pairs $\{20^o, 60^o\}$

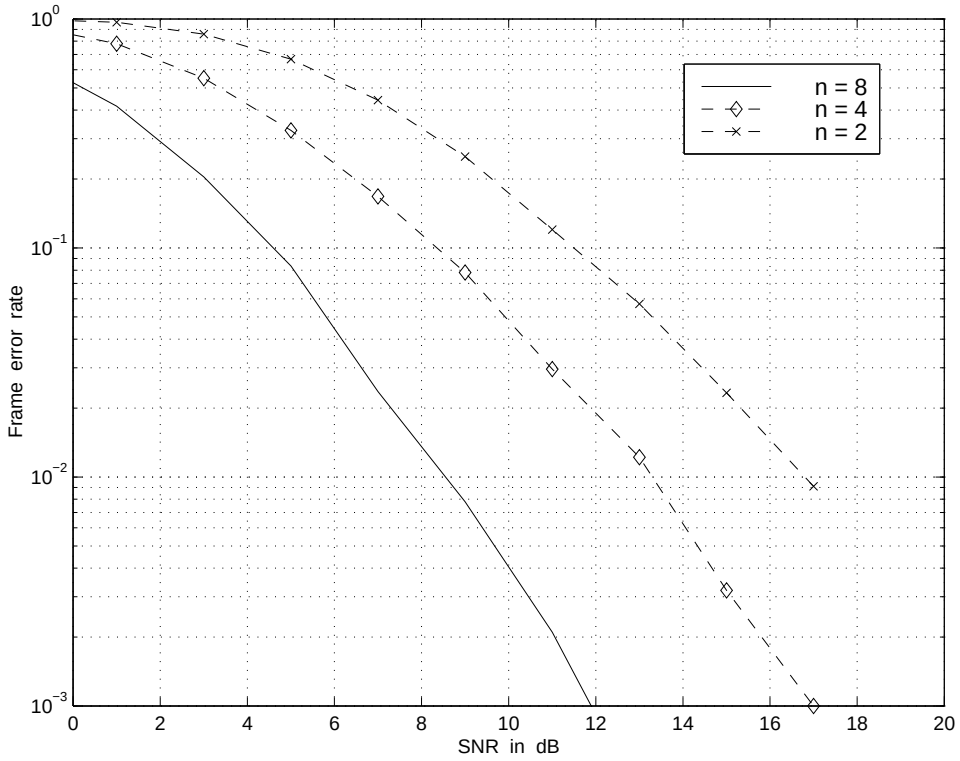

Figure 6.4: FER v.s. SNR; ISI-free case; angle pairs $\{50^o, 60^o\}$

6.5.2 The Case with ISI

As in Section 6.5.1, a MIMO system is simulated with $m = 2, L = 2$ and different values of n. The base station is assumed to use a linear array with antenna spacing $d = 0.5$ wavelengths. However it is assumed that one of the multipaths arrives with a one symbol delay with respect to the other, thus causing ISI.

In Section 6.4, we proposed a sub-optimal solution to the beamforming vector $\mathbf{b}$, that would maximize the SNR gain over the SIMO case. In Figures 6.5 and 6.6, we show the histograms of the SNR gains $P(\mathbf{b})$ for $n = 4$ and $n = 8$, when the two multipath arrive randomly within an angle of 60^o of each other. As in the ISI-free case, the gain is $\geq n/L$. The gain is usually much greater than the lower bound. The

gain upper bound of n is achieved when the multipaths are spatially close together.

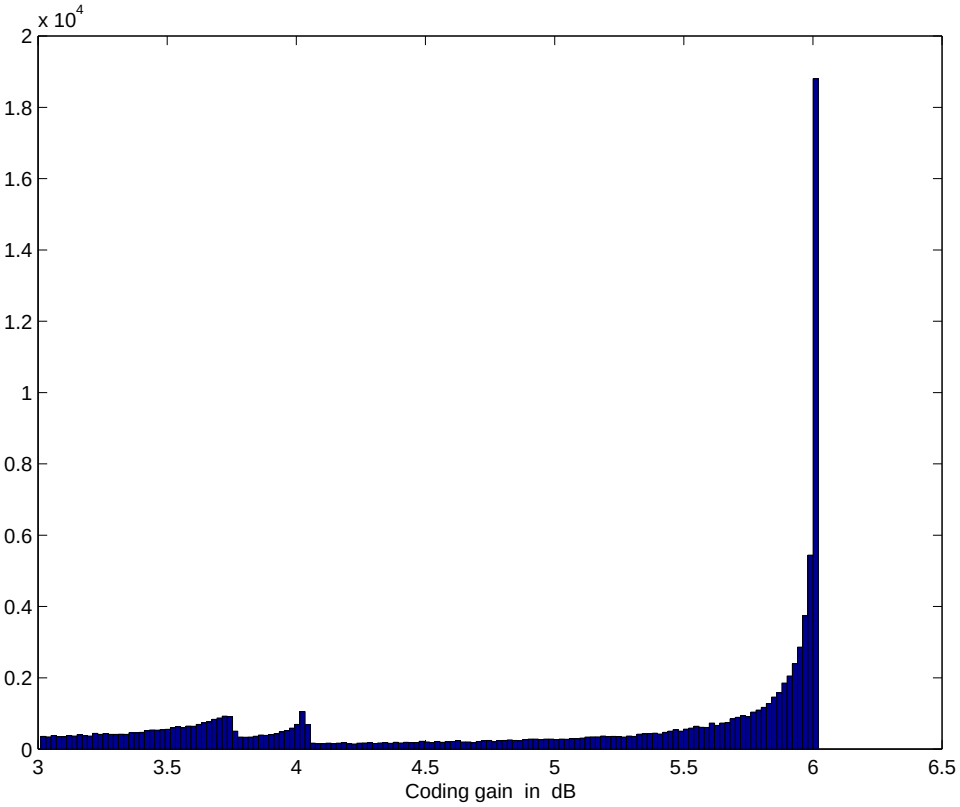

Figure 6.5: Histograms of the SNR gain $P(b)$ for $n = 4$

The entire SIMO and MIMO transmission systems were simulated by transmitting and maximum-likelihood (Viterbi) decoding, using the delay diversity code with a QPSK signal constellation. The frame length was chosen as 100, and 10,000 frames were transmitted. Here we considered two multipath angle pairs, different from the ones used in previous section, $\{30^o, 35^o\}$ and $\{30^o, 75^o\}$, to illustrate the performance of the system. Figures 6.7 and 6.8 show the frame error rate as a function of the SNR for the above cases, for $n = 1, 4, 8$. SNR is defined as in Section 6.5.1. It can be seen that when n increases, the SNR gain allows better performance at a given SNR, even though the diversity is the same in all cases. And as discussed in Section 6.4,

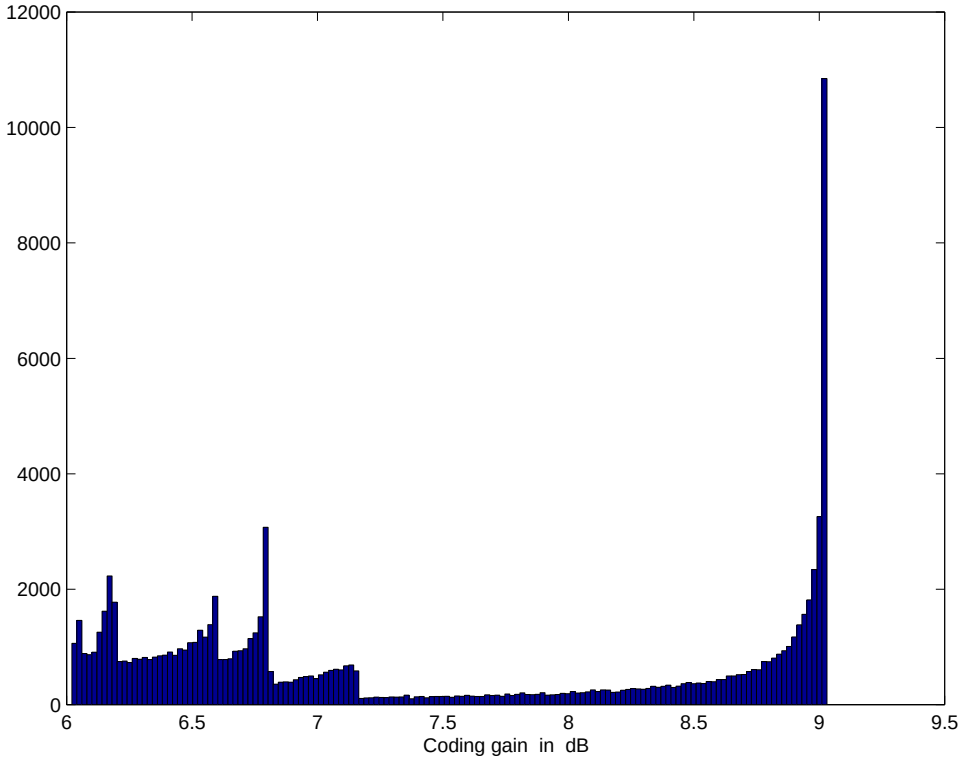

Figure 6.6: Histograms of the SNR gain $P(b)$ for $n = 8$

the SNR gain is significant, especially in the case of the angle pair $\{30^o, 35^o\}$,where the multipaths are spaced close together.

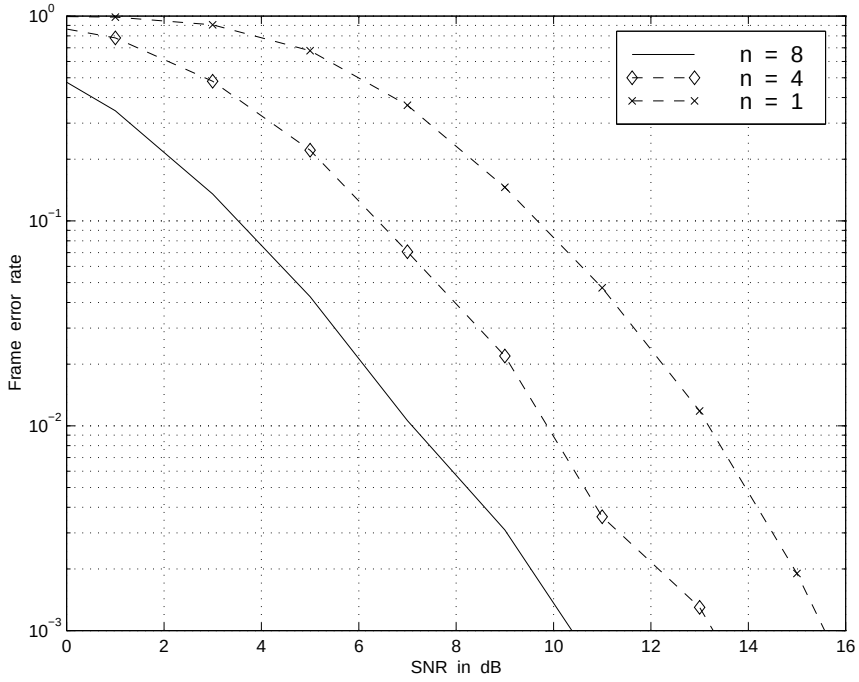

Figure 6.7: FER v.s. SNR; the ISI case; angle pairs $\{30^o, 75^o\}$

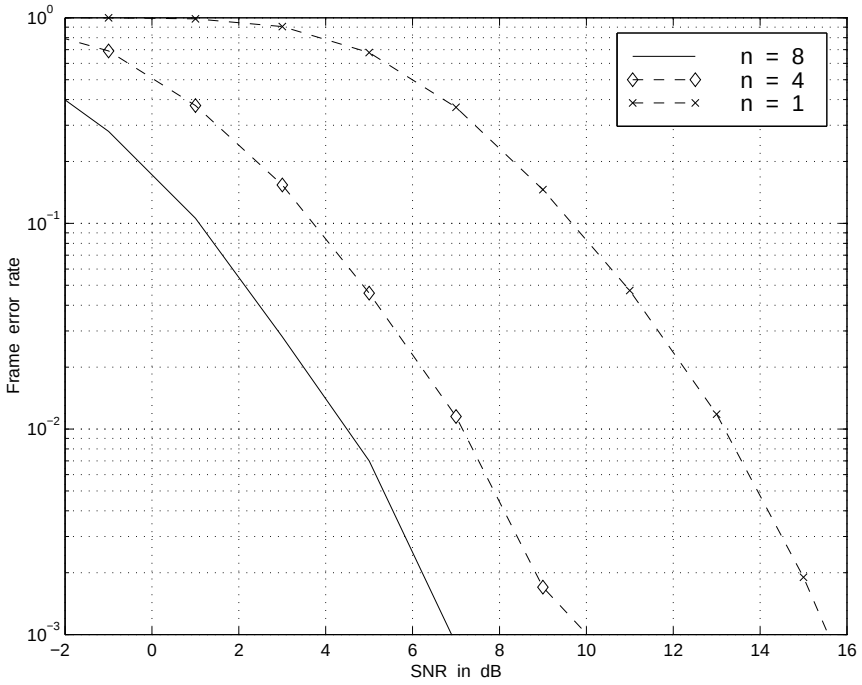

Figure 6.8: FER v.s. SNR; the ISI case; angle pairs $\{30^o, 35^o\}$

Chapter 7

Blind OFDM Symbol Synchronization in ISI Channels

7.1 Introduction

Recently, there has been considerable interest in using Orthogonal frequency division multiplexing (OFDM) systems for wireless transmission [65], such as in Digital Audio Broadcasting (DAB) and digital television. OFDM is a transmission scheme in which data is simultaneously explicitly sent over several frequency bands. The resilience of OFDM systems to frequency selective fading can be attributed to the cyclic prefix inserted between symbols, that allows decomposition of the channel into independent subchannels by use of the Fast Fourier Transform (FFT). However, a consequence of this 'frame' structure of an OFDM symbol, is that it becomes important for the receiver to identify the beginning of each new symbol. This is the problem of symbol synchronization. Usually, once the correct symbol synchronization has been achieved,

tracking this position is a simpler problem. The previous chapters dealt with flat fading channels (frequency selective fading was considered in Section 6.4, but detection issues were glossed over). Frequency selective fading introduces an additional complication, because of intersymbol interference. However, OFDM is an example of a wideband system to which the results of Chapter 2, 3, 4 can be applied with only minor modifications. $\log(1 + P_n g_n)$ is replaced with the sum of capacities over the N frequency slots $\sum_{k=1}^{N} \log(1 + P_{k,n} g_{k,n})$. In this chapter however, we switch topics and show how frequency selectivity causes complications in the detection of the signal; in particular we look at the problem of symbol synchronization.

Symbol synchronization can be achieved by transmitting pilot symbols. However, this is an unnecessary waste of bandwidth, especially in broadcast systems, where the transmitter would have to keep transmitting pilot symbols periodically to allow new users to synchronize. Therefore various schemes have been proposed [57],[29],[2],[47] that use only the transmitted symbol statistics for symbol synchronization. These blind methods essentially exploit the redundancy in the cyclic prefix, and therefore do not require additional pilot symbols.

However, the blind synchronization methods proposed to date assume in their analysis that the channel is free of Intersymbol Interference (ISI). Whereas they are extremely efficient in their use of symbol statistics, they do not guarantee correct synchronization when the channel has ISI. Also, they assume that the various tones in a symbol carry data that is i.i.d. In particular, if the symbol is shaped by a non constant power profile, special techniques may be required by these algorithms to handle the pulse shape [26]. See [36] for a comparison of some of these methods.

This chapter presents a new algorithm for blind symbol synchronization, which

given sufficient signal statistics, is guaranteed to achieve correct symbol synchronization even in the presence of ISI. The algorithm is necessarily more complex than any of the algorithms proposed earlier, and requires more statistics. However, the guarantee of correctness is an attractive feature, especially in fixed receiver broadcast systems, where a particular user may not have much choice in the (nearly time-invariant) channel it sees. We have presented some of the results of the chapter in [41].

The chapter is organized as follows. The problem is formulated in Section 7.2. In Section 7.3 we present the theoretical basis on which the new algorithm is based. It will be shown that the ranks of certain autocorrelation matrices convey information about the correct symbol synchronization position. In Section 7.4, we show how these ideas can be used in a realistic scenario, where we only have a noisy estimate of the autocorrelation matrices. Section 7.5 illustrates these ideas with simulations.

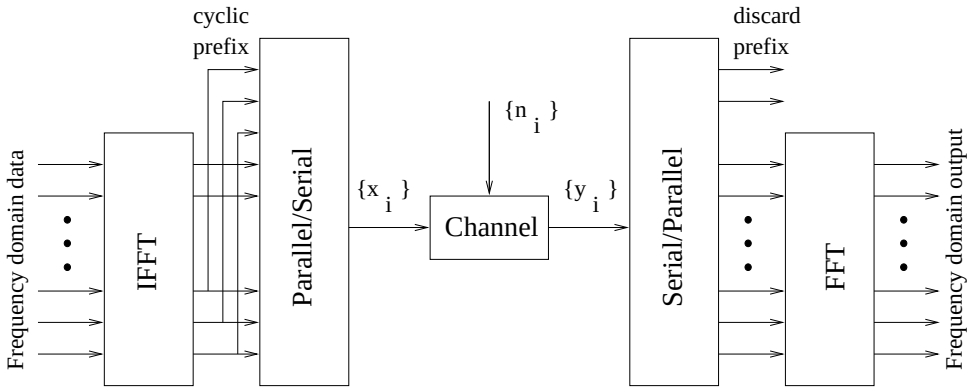

Figure 7.1: Basic OFDM system

7.2 Problem Formulation

7.2.1 Notations

Standard notations are used in this chapter. Bold letters denote vectors and matrices. Other notation are as follows.

$(\cdot)^t$	Transpose
$(\cdot)^*$	Hermitian
$(\cdot)^{1/2}$	(any) Square root of positive definite matrix
$\mathbf{I}_n$	$n \times n$ Identity matrix
$\widehat{(\cdot)}$	Estimate of $(\cdot)$
$\mathbf{R}_{\mathbf{z}}(i)$	$\doteq E\{\mathbf{z}_i \mathbf{z}_i^*\}$
$\rho[(\cdot)]$	Rank of matrix $(\cdot)$

7.2.2 OFDM System

The basic baseband-equivalent OFDM system is shown in Figure 7.1. The system equations in the time domain can be written as

$$\mathbf{z}_i = \mathbf{H}\mathbf{x}_i \tag{7.1}$$

$$\mathbf{y}_i = \mathbf{z}_i + \mathbf{n}_i$$

$$\mathbf{x}_i \doteq (x_i, x_{i+1}, \cdots, x_{N+\nu+i-1})^t \quad \text{, vector of length } N+\nu$$

$$\mathbf{n}_i \doteq (n_i, n_{i+1}, \cdots, n_{N+\nu-L_0+i-1})^t \quad \text{, similarly } \mathbf{z}_i, \mathbf{y}_i \text{ are of length } N+\nu-L_0$$

$$\mathbf{H} \doteq \begin{pmatrix} h_{L_0} & \cdots & h_0 & 0 & 0 \\ 0 & \ddots & & \ddots & 0 \\ 0 & 0 & h_{L_0} & \cdots & h_0 \end{pmatrix} = (\mathbf{h}_0, \ldots, \mathbf{h}_{N+\nu-1})$$

where

N	Number of OFDM tones
ν	Cyclic prefix length
L_0	Maximum length of channel allowed
L	Length of specific channel ($\Delta \doteq L_0 - L \geq 0$)
i	Time subscript
$\{h_i, i = 0, \ldots, L\}$	Channel impulse response
$\{x_i, i = \ldots, 0, \ldots\}$	Time domain transmitted data
$\{n_i, i = \ldots, 0, \ldots\}$	Time domain noise
$\{z_i, i = \ldots, 0, \ldots\}$	Time domain noiseless output (unmeasurable)
$\{y_i, i = \ldots, 0, \ldots\}$	Time domain noisy output (measured)

Note that $\{\mathbf{x}_0, \mathbf{x}_{N+\nu}, \mathbf{x}_{2N+2\nu}, \ldots\}$ are the successive transmitted symbols, each a vector of length $N + \nu$. The cyclic prefix in the symbols implies $x_{i+p(N+\nu)} = x_{i+N+p(N+\nu)},\ i = 0, \ldots, \nu - 1,\ \forall\ p$.

Figure 7.2 illustrates the notation defined above.

7.2.3 Assumptions

It is assumed that the transmitted symbol vectors $\{\mathbf{x}_0, \mathbf{x}_{N+\nu}, \ldots\}$ are identically distributed. Thus, the sequences $\{x_i\}, \{z_i\}, \{y_i\}$ are cyclo-stationary with period $N+\nu$.

The channel is assumed to be *at least one tap shorter* than the cyclic prefix (i.e. maximum channel length $L_0 < \nu$). This assumption is slightly stronger than the assumption normally used, i.e. $L_0 \leq \nu$. This assumption is usually satisfied in OFDM systems by appropriate choice of parameters N, ν.

The channel will be assumed to be time-invariant (or at least slowly time-varying)

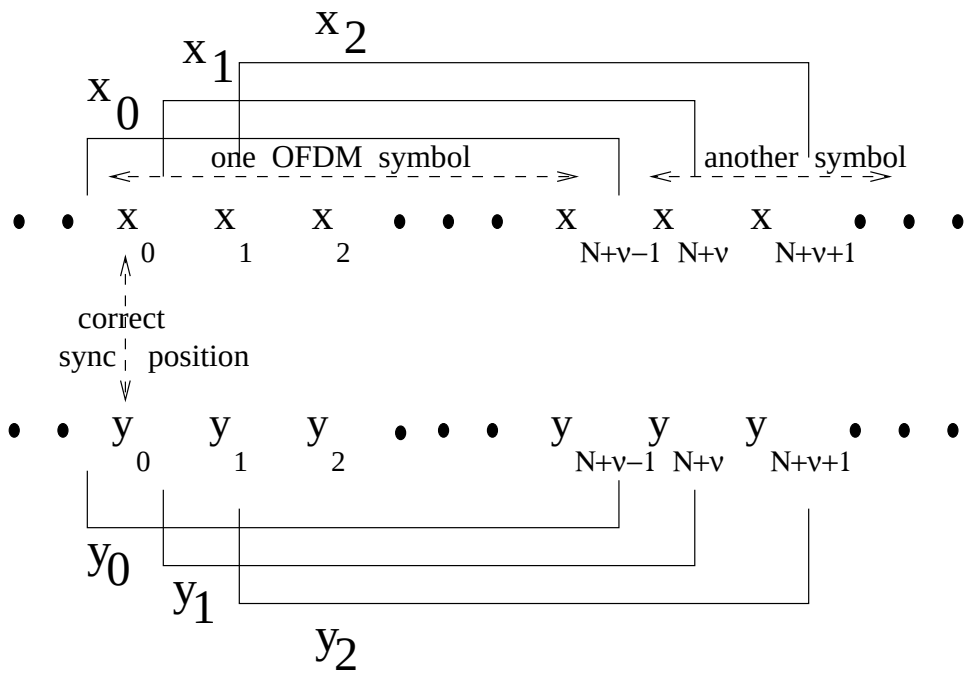

Figure 7.2: Notation for the data and received signal sequences

so that second order statistics of the output signal can be collected. This assumption is valid for fixed receiver broadcast OFDM systems.

The noise will be assumed to be additive, white and Gaussian (AWGN). The transmitted data $\{x_i\}$ will also be assumed to be Gaussian. This is a valid assumption for OFDM, when N is large. The assumption is required so that the MDL criterion can be used to estimate the ranks of certain matrices [59].

It is further assumed that the transmitted data (except for the cyclic prefix) is *statistically non-degenerate*, i.e. if we strip the cyclic prefixes from the transmitted data, then with probability one, no data element in the resulting sequence can be expressed as a linear combination of other elements in the sequence. This assumption will be weakened in Section 7.4.3, where a modified algorithm will be described. Note that no assumption is made on the resulting sequence being i.i.d., unlike [57],[29]. In

particular, we allow for the transmitted symbols to bear any pulse shape.

7.2.4 Problem Formulation

The problem of blind symbol synchronization in OFDM is to use only the statistics (in our case, second-order statistics) of the received signal, to identify the correct positions i where the OFDM symbols begin. For a channel of length L, this amounts to identifying any one of the positions $i = -\Delta, \ldots, -1, 0$. For these positions, due to the cyclic prefix, an N-point FFT can be used to decompose the channel into N parallel independent subchannels [50].

Earlier approaches to this problem (see [36] for a survey) have concentrated on using the autocorrelation function of the sequence $\{y_i\}$. They use the information in the correlation function

$$\gamma(i) \quad = \quad \sum_{k=i}^{i+\nu-1} y(k)y^*(k+N) \tag{7.2}$$

Though these algorithms are derived for flat-fading channels, yet they work remarkably well in mild-ISI channels. However, they breakdown in the presence of stronger ISI.

Our goal is to derive algorithms that explicitly account for ISI in channels. Specifically, we would like to use the autocorrelation matrices $\{\mathbf{R}_{\mathbf{y}}(i)\}$ of the noisy output vectors $\{\mathbf{y}_i\}$. Note that these matrices are well defined because the sequence $\{y_i\}$ is cyclo-stationary with a period $N + \nu$. Also, note that only $\{i = 0, 1, \ldots, N + \nu - 1\}$ need be considered, because of the cyclo-stationarity.

However, in practice, only an estimate $\widehat{\mathbf{R}}_{\mathbf{y}}(i)$ of these matrices is known, say by time-averaging over P vectors; $\widehat{\mathbf{R}}_{\mathbf{y}}(i) = \frac{1}{P}\sum_{p=0}^{P-1} \mathbf{y}_{i+p(N+\nu)}\mathbf{y}^*_{i+p(N+\nu)}$.

The problem is therefore, given the estimates $\{\widehat{\mathbf{R}}_{\mathbf{y}}(i), i = 0, \ldots, N+\nu-1\}$ (these matrices are known in the correct order, but without knowing which one corresponds to $i = 0$), to identify any one of the positions $\{-\Delta, \ldots, -1, 0\}$.

7.3 Theoretical Basis For Blind Symbol Synchronization

In this section, we describe the basic theorems that will be used for blind symbol synchronization. We first assume knowledge of the exact autocorrelation matrices $\{\mathbf{R}_{\mathbf{z}}(i), i = 0, \ldots, N+\nu-1\}$, and make some fundamental observations about the behavior of their ranks. It will be shown that the ranks can be used to identify the correct synchronization positions.

We first begin by bounding the ranks.

Lemma 4 *The rank of the autocorrelation of the noiseless output vector can be bounded as*

$$g_l(i) \leq \rho[\mathbf{R}_{\mathbf{z}}(i)] \leq g_u(i) \qquad \textit{where the functions}$$

$$g_l(i) = \begin{cases} N + |i| - L_0 & i = -\nu, \ldots, 0, \ldots, \nu \\ N + \nu - L_0 & i = \nu, \ldots, N \end{cases}$$

$$g_u(i) = \begin{cases} N + |i| & i = -(\nu - L_0), \ldots, 0, \ldots, (\nu - L_0) \\ N + \nu - L_0 & i = (\nu - L_0), \ldots, N + L_0 \end{cases}$$

Proof: See Appendix A.5 for the proof. □

Lemma 4 indicates that the rank of the $\mathbf{R_z}(i)$ matrices may be of use in determining the correct symbol synchronization. The bounds depend on the position i selected. They are minimum for the correct synchronization position $i = 0$. More importantly, the lemma shows that the $\mathbf{R_z}(i)$ matrices have a rank of at least $N - L_0$, a fact that will be used in Section 7.4 to simplify the computations.

Next, we state the main theorem, that describes the behavior of $\rho[\mathbf{R_z}(i)]$ in detail.

Theorem 4 *For a channel where $L = L_0$, $\rho[\mathbf{R_z}(i)]$ has the following behavior*

$$\rho[\mathbf{R_z}(i)] = \begin{cases} \rho[\mathbf{R_z}(i-1)] - \mu_i & i = -\nu, \ldots, -(\nu - L_0) \\ \rho[\mathbf{R_z}(i-1)] - 1 & i = -(\nu - L_0 - 1), \ldots, 0 \\ \rho[\mathbf{R_z}(i+1)] - 1 & i = 0, \ldots, (\nu - L_0 - 1) \\ \rho[\mathbf{R_z}(i+1)] - \mu_i & i = (\nu - L_0), \ldots, \nu \\ N + \nu - L_0 & i = \nu, \ldots, N \end{cases}$$

where $\mu_i \in \{0, 1\}$ are unknown integers.

Proof: See Appendix A.6 for the proof. □

Theorem 4 provides a clear idea of the synchronization method that can be adopted. It shows that as the chosen position i gets closer to the correct synchronization position ($i = 0$) from either side, the rank of $\mathbf{R_z}(i)$ is non-increasing, and in fact decreases near $i = 0$. This shows that the correct position $i = 0$ is the position of minimum rank for the matrices $\{\mathbf{R_z}(i), i = 0, \ldots, N + \nu - 1\}$. It is also clear that the rank cannot decrease by more than ν.

Theorem 4 is restrictive because it assumes a channel of length exactly L_0. In most realistic scenarios, the channel will be shorter than the maximum allowed. The next corollary extends Theorem 4 to this general case.

Corollary 1 *For a channel where $L \leq L_0$, $\rho[\mathbf{R}_{\mathbf{z}}(i)]$ has the following behavior*

$$\rho[\mathbf{R}_{\mathbf{z}}(i)] = \begin{cases} \rho[\mathbf{R}_{\mathbf{z}}(i-1)] - \mu_i & i = -\nu, \ldots, -(\nu - L) \\ \rho[\mathbf{R}_{\mathbf{z}}(i-1)] - 1 & i = -(\nu - L - 1), \ldots, -\Delta \\ \rho[\mathbf{R}_{\mathbf{z}}(0)] & i = -\Delta, \ldots, 0 \\ \rho[\mathbf{R}_{\mathbf{z}}(i+1)] - 1 & i = 0, \ldots, (\nu - L_0 - 1) \\ \rho[\mathbf{R}_{\mathbf{z}}(i+1)] - \mu_i & i = (\nu - L_0), \ldots, (\nu - \Delta) \\ N + \nu - L_0 & i = (\nu - \Delta), \ldots, N \end{cases}$$

where $\mu_i \in \{0, 1\}$ are unknown integers.

Proof: Δ was defined earlier as $\Delta = L - L_0$. See Appendix A.7 for the proof. □

The Corollary shows that the effect of a short channel $L < L_0$ is to increase the number of positions of minimum $\rho[\mathbf{R}_{\mathbf{z}}(i)]$ from $\{i = 0\}$ only, to the set $S_c = \{i | i = -\Delta, \ldots, -1, 0\}$. Noting that the positions of correct symbol synchronization are indeed the set S_c for a channel of length L, we conclude that the minimum rank criterion can yet be used for symbol synchronization.

The behavior of $\rho[\mathbf{R}_{\mathbf{z}}(i)]$ is illustrated in Figure 7.3 for the case $N = 128, \nu = 32, L_0 = 26$ for channels of various lengths. It is clear that the positions of the minima can be used to identify the correct synchronization position.

All the previous observations assumed knowledge of $\mathbf{R}_{\mathbf{z}}(i)$, which can only be obtained in the absence of noise. The next theorem (trivially) extends these results to the case of AWGN.

Theorem 5 *In the case of a frequency-selective AWGN channel for which $L < \nu$, and the noise variance σ_n^2 is known, the correct symbol synchronization positions $S_c = (i = -\Delta, \ldots, -1, 0)$ can be identified by seeking those i, for which the number of singular values of $\mathbf{R}_{\mathbf{y}}(i)$ that are equal to σ_n^2, is maximum.*

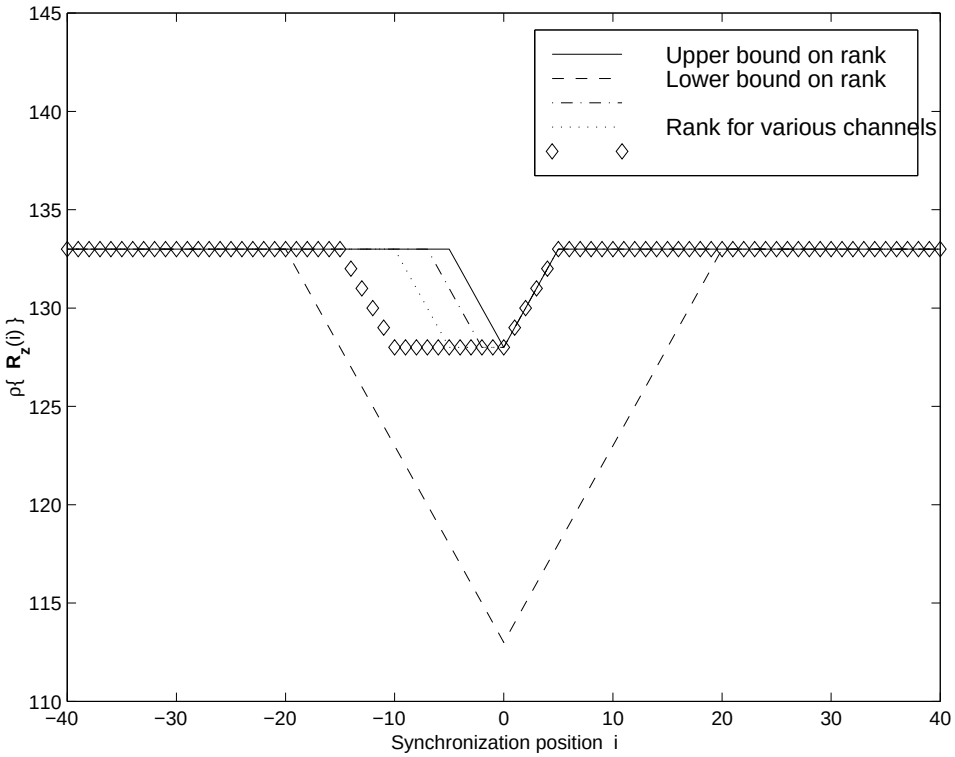

Figure 7.3: Theoretical behavior of $\rho[\mathbf{R}_\mathbf{z}(i)]$ for some channels: $N = 128, \nu = 32, L_0 = 26$

Proof: Corollary 1 shows that $\rho[\mathbf{R}_\mathbf{z}(i)]$ is minimum for all $i = -\Delta, \ldots, -1, 0$, and only for these positions. These are the positions of correct symbol synchronization.

In the AWGN case with noise variance σ_n^2, we can write

$$\mathbf{R}_\mathbf{y}(i) \;=\; \mathbf{R}_\mathbf{z}(i) + \sigma_n^2 \mathbf{I}_{N+\nu-L_0}$$

By the spectral shift theorem, we see that

$$\rho[\mathbf{R}_\mathbf{z}(i)] \;=\; (N+\nu-L_0) - (\text{number of singular values of } \rho[\mathbf{R}_\mathbf{y}(i)] \text{ that equal } \sigma_n^2)$$

Therefore, the positions of correct synchronization can be identified as specified in

Theorem 5. □

7.4 Practical Considerations

The ideal method to estimate the correct position(s) of symbol synchronization would be to jointly estimate the ranks of the matrices $\{\mathbf{R}_{\mathbf{z}}(i), i = 0, \ldots, N+\nu-1\}$ (from matrices $\{\mathbf{R}_{\mathbf{y}}(i)\}$, making use of the spectral shift theorem), and then use Corollary 1 to estimate the position. The optimum solution to this is non-obvious.

As a first attempt, we could simply find the minimum rank $\mathbf{R}_{\mathbf{z}}(i)$, and thus the correct synchronization position, as specified in Theorem 5. Therefore, the problem reduces to estimating the rank of matrices $\{\mathbf{R}_{\mathbf{z}}(i), i = 0, \ldots, N+\nu-1\}$ *separately*, and searching for the minimum. In practice, however, we only have estimates $\{\widehat{\mathbf{R}}_{\mathbf{y}}(i)\}$ of $\{\mathbf{R}_{\mathbf{y}}(i)\}$. Since this is invariably noisy, it becomes important to use a mathematically plausible method for the rank estimation. We use a modification of the well-known rank estimation method proposed in [59], that's based on Rissanen's Minimum Description Length (MDL) criterion.

7.4.1 Estimation with σ_n^2 known

It is shown in Appendix A.8 that the MDL criterion to compute an estimate $\widehat{\rho}[\mathbf{R}_{\mathbf{z}}(i)]$ of the rank of $\mathbf{R}_{\mathbf{z}}(i)$, applied to the case of known σ_n^2, reduces to

$$\widehat{\rho}[\mathbf{R}_{\mathbf{z}}(i)] = \underset{k\in\{N-L_0,\ldots,N+\nu-L_0\}}{\operatorname{argmin}} \left\{ \sum_{p=N-L_0+1}^{N+\nu-L_0} \left(\log(\frac{\lambda_p}{\hat{\lambda}_p}) + (\frac{\hat{\lambda}_p}{\lambda_p}) \right) + \frac{\log P}{2P} \left(k(2(N+\nu-L_0)-k+1)+1\right) \right\} \quad \text{where}$$

$$\lambda_p \;=\; \begin{cases} \hat{\lambda}_p & p = 1, \ldots, k \\ \sigma_n^2 & p = k+1, \ldots, N+\nu-L_0 \end{cases}$$

and $\{\hat{\lambda}_1, \ldots, \hat{\lambda}_{N+\nu-L_0}\}$ are the singular values of $\widehat{\mathbf{R}}_{\mathbf{y}}(i)$, arranged in decreasing order of magnitude. Note that only the smallest ν singular values need to be computed. Note again that P is the number of vectors averaged to calculate $\widehat{\mathbf{R}}_{\mathbf{y}}(i)$.

Therefore, the algorithm for symbol synchronization is as below:

Algorithm Symbol_synch() :

1. Compute $\{\widehat{\mathbf{R}}_{\mathbf{y}}(i), i = 0, 1, \ldots, N+\nu-1\}$ by time-averaging over P received signal vectors; $\widehat{\mathbf{R}}_{\mathbf{y}}(i) = \frac{1}{P}\sum_{p=0}^{P-1} \mathbf{y}_{i+p(N+\nu)}\mathbf{y}^*_{i+p(N+\nu)}$.

2. Use the $\{\widehat{\mathbf{R}}_{\mathbf{y}}(i)\}$ to compute $\{\widehat{\rho}[\mathbf{R}_{\mathbf{z}}(i)], i = 0, 1, \ldots, N+\nu-1\}$, using the MDL criterion described above.

3. Estimate the position(s) of correct synchronization as the set $\widehat{S}_c$

$$S_c \;=\; \{i \mid i = \underset{i \in \{0,\ldots,N+\nu-1\}}{\operatorname{argmin}} \{\widehat{\rho}[\mathbf{R}_{\mathbf{z}}(i)]\}\ \}$$

4. Since the algorithm works with noisy data, we check for the validity of the result using Corollary 1. i.e. We check that the function $\widehat{\rho}[\mathbf{R}_{\mathbf{z}}(i)]$ (as a function of i) behaves as specified in Corollary 1. If not, we have the option of applying heuristics (such as checking for a 'reasonable match'). Alternately, we could declare a mis-estimation due to lack of statistics, and repeat from step 1 again, using a larger value of P. This is not necessarily a lot more computation, because the recomputed singular values will be close to the previous set, and so the previous set can be used as a good initialization point for the re-computation.

7.4.2 Estimation with σ_n^2 unknown

When σ_n^2 is unknown, one could think of using the method described in [59] (Wax-Kailath method), which jointly estimates rank and σ_n^2. The problem however, in applying the method directly to our case, is that the method essentially estimates rank based solely on the clustering of the smallest singular values - it does not take into account their absolute magnitude. Therefore, it is possible that it would declare a low rank for $\widehat{\mathbf{R}}_{\mathbf{z}}(i)$ even though it is 'supposed to be full rank'.

Therefore, our approach is to use the Wax-Kailath method to compute $\{\widehat{\rho}[\mathbf{R}_{\mathbf{z}}(i)], i = 0, \ldots, N+\nu-1\}$. This then gives us an estimate of the 'noise variance' $\hat{\sigma}_n^2$ for each i. Since ideally, the smallest singular value is equal to σ_n^2, hence we choose the lowest of the computed noise variances, and declare it as σ_n^2. Then we can use the algorithm in Section 7.4.1 to estimate the correct symbol synchronization position.

7.4.3 Computational Complexity

The new algorithm seems to be prohibitively computation intensive, since it requires $N+\nu$ singular value decompositions, each on a $(N+\nu-L) \times (N+\nu-L)$ matrix. However, a closer look at the proof of Theorem 4 and Corollary 1 shows that it is sufficient to consider matrices $\mathbf{R}_{\mathbf{u}}(i)$ in place of the matrices $\mathbf{R}_{\mathbf{z}}(i)$ respectively, where the vectors $\mathbf{u}_i$ are

$$\mathbf{u}_i \;=\; (z_i \; z_{i+1} \; \cdots \; z_{i+\nu-1} \; z_{N-L_0-1} \; z_{N-L_0} \; \cdots \; z_{N+\nu-L0-1})^t \tag{7.3}$$

i.e. the vectors $\mathbf{u}_i$ are formed from the top and bottom ν elements of the vectors $\mathbf{z}_i$ respectively.

$\mathbf{R}_{\mathbf{u}}(i)$ satisfies Theorem 4 and Corollary 1, except that the maximum rank is

now 2ν instead of $N + \nu - L_0$. Therefore, the correct symbol synchronization can be identified by applying Theorem 5 on matrices $\mathbf{R_q}(i)$ in place of the matrices $\mathbf{R_y}(i)$ respectively, where the vectors $\mathbf{q}_i$ are defined as

$$\mathbf{q}_i \;=\; (y_i \; y_{i+1} \; \cdots \; y_{i+\nu-1} \; y_{N-L_0-1} \; y_{N-L_0} \; \cdots \; y_{N+\nu-L0-1})^t \tag{7.4}$$

Thus, the modified algorithm requires $(N + \nu)$ SVDs, each on a $2\nu \times 2\nu$ matrix only. Since ν is usually much smaller than N, hence this offers substantial savings in computation.

Another direct benefit of this modified algorithm is that it is no longer necessary to insist on non-degeneracy of the transmitted signal (Section 7.2.3). In fact, often transmission is turned off in some OFDM tones, which would result in a degenerate signal. However, as long as the number of active tones is at least 4ν, the various $\mathbf{R_w}(i)$ matrices (Appendix A.6) are always full rank, and thus the various proofs hold. Therefore, even in such a case, the algorithm achieves the correct synchronization.

Further savings in computation can be achieved by using the following two observations

1. The MDL rank estimation procedure described in Section 7.4.1 shows that only the smallest ν singular values need to be computed for each $\widehat{\mathbf{R}}_\mathbf{y}(i)$. This is also true for $\widehat{\mathbf{R}}_\mathbf{q}(i)$. Various specialized algorithms can do this quickly [21].

2. The matrix $\widehat{\mathbf{R}}_\mathbf{q}(i+1)$ can be obtained from $\widehat{\mathbf{R}}_\mathbf{q}(i)$ by replacing a two rows and two columns. To recompute the singular values therefore takes only $O((2\nu)^2)$ operations [21].

Either of these observations can be used to gain significant computational savings.

7.4.4 Consistency of Algorithm

It is shown in [59] that the rank estimation method presented there is consistent. i.e. The estimator yields the correct rank with probability one, as the sample size P increases to infinity. A similar argument can be made to show that the algorithms proposed in this chapter are also consistent. Thus, we will always identify the symbol synchronization positions correctly, provided we are willing to wait and collect sufficient statistics. This is in contrast to the methods presented in [57],[29],[2], where no guarantees can be made for ISI channels.

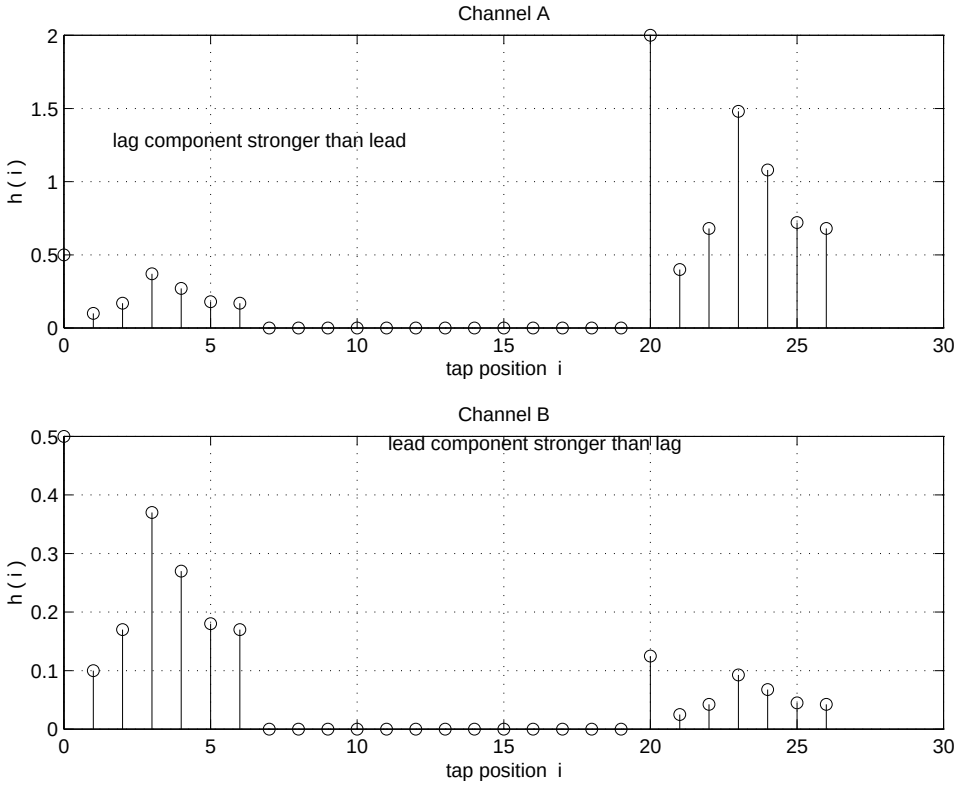

Figure 7.4: Impulse response of SFN channels

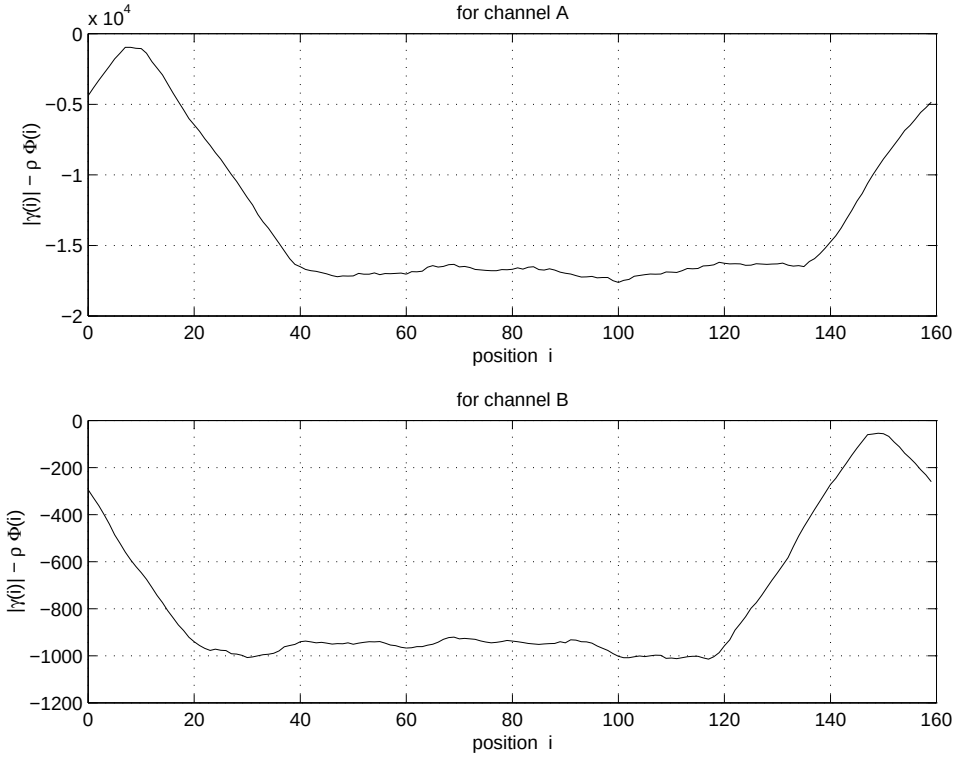

Figure 7.5: Decision functions used by Beek-Sandell algorithm. Synchronization position estimate is the maximum of the decision function

7.5 Simulation Results

To demonstrate the performance of our method , we choose the case $N = 128, \nu = 32, L_0 = 26$. We compare the performance to that of the maximum likelihood based synchronization algorithm proposed in [57] (Beek-Sandell), which is typical of the auto-correlation function based algorithms (see also [29],[2]). We first illustrate how the proposed algorithm handles ISI better. We choose a strong ISI case, as in Figure 7.4, at an SNR of 30 dB. Such channels occur in Single Frequency Networks ([47]). In all cases, σ_n^2 is assumed known. Frequency offset is assumed to be negligible (or corrected by an appropriate algorithm).

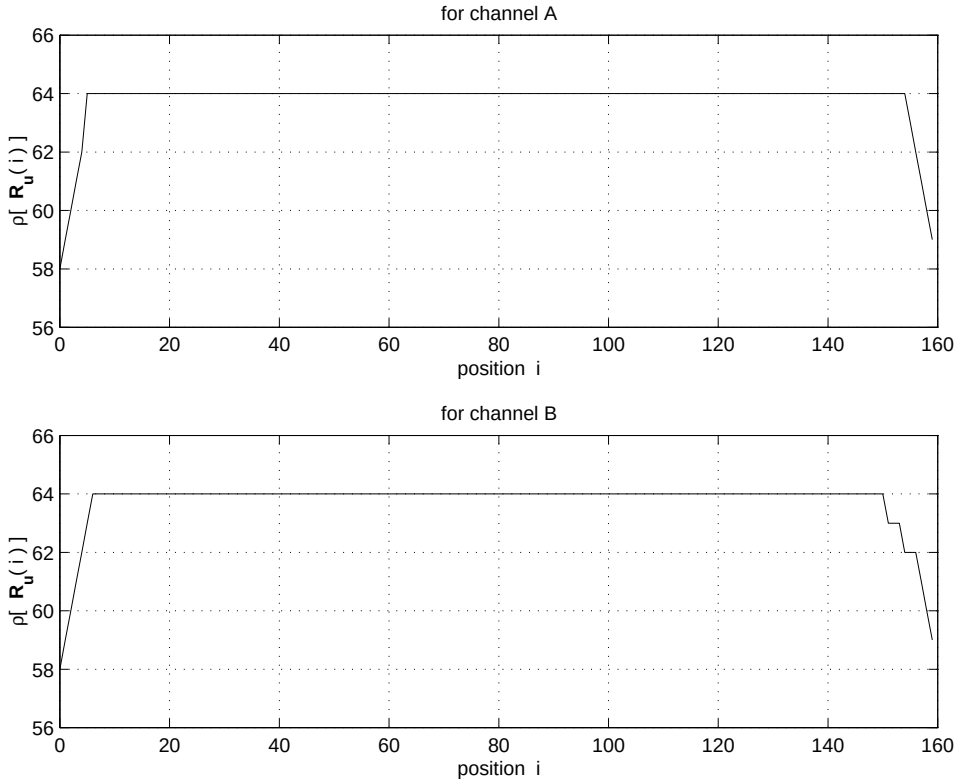

Figure 7.6: Decision function used by the proposed algorithm. Synchronization position estimate is the minimum rank position

Both algorithms were run on these channels, and the estimated position of correct synchronization was found by collecting data from 600 symbols. In the case of the Beek-Sandell algorithm, the estimated position was considered to represent the center of the channel, as suggested in [57]. In the case of our method (we use the modified method described in Section 7.4.3), the estimated position was found using the algorithm described in Section 7.4.1 (without step 4, so as to illustrate some limitations). Figure 7.5 shows the decision functions used by the Beek-Sandell algorithm. For both channels A and B, the algorithm determines the position of synchronization, as the maximum of the decision function, erroneously shifted from the correct position $i = 0$. In contrast, Figure 7.6 shows that the proposed algorithm correctly

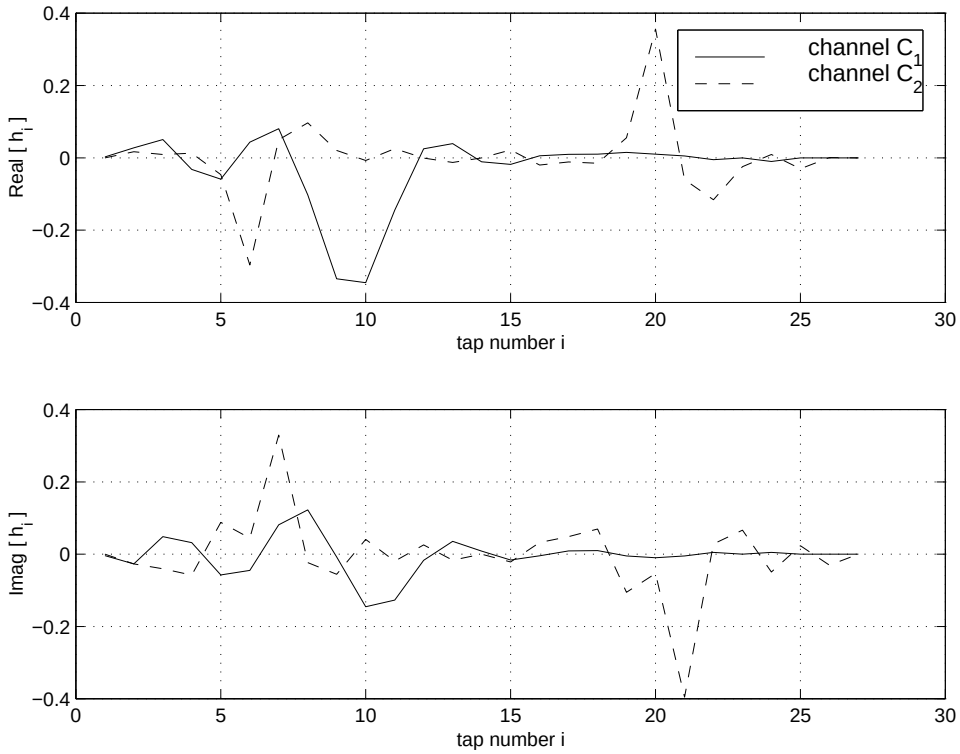

Figure 7.7: Impulse response of channels used for simulation: Channel C_2 has more ISI than channel C_1

points out the synchronization position (minimum rank position) as $i = 0$.

We next choose three cases that will further illustrate the differences between the two methods. The first two use channels C_1 and C_2 (Figure 7.7) respectively and i.i.d. data symbols. The third case is that of channel C_1 but dependent data symbols (i.e. non-diagonal $\mathbf{R_w}(0)$). Tables 7.1 and 7.2 show the results for these cases. Table 7.1 lists the normalized 'wall energy' (i.e. the energy of the channel taps that fall outside the cyclic prefix, once the method has been used to synchronize the symbols) for an SNR of 30 dB, but using different number of training symbols P. Table 7.2 lists the normalized wall energy for $P = 2400$ symbols, but at different values of SNR.

The tables show that whereas the Beek-Sandell algorithm works well on channels

with mild ISI (case 1), it can be misled when the channel has a reasonable amount of ISI (case 2) or when the input data symbols are not i.i.d. (case 3). Higher values of SNR or P do not help in alleviating this problem.

On the other hand, the new method is seen to work reasonably well, and more importantly, always gives the correct result, provided either SNR or P is large enough. This is because the method explicitly allows for ISI and colored input data.

P	C_1, i.i.d. data		C_2, i.i.d. data		C_1, colored data	
	New	B-S	New	B-S	New	B-S
150	0.9957	0	1	0.1296	0.9957	1
300	0.0028	0	1	0.1296	0.9957	1
600	0.0005	0	0	0.1296	0.9957	1
1200	0	0	0	0.1296	0.9957	1
2400	0	0	0	0.1296	0.0005	1
4800	0	0	0	0.1296	0	1

Table 7.1: Normalized wall energy after synchronization, using the new method and using the Beek-Sandell algorithm, for various values of P, and $SNR = 30$ dB

SNR (dB)	C_1, i.i.d. data		C_2, i.i.d. data		C_1, colored data	
	New	B-S	New	B-S	New	B-S
10	0.9957	0	0.0130	0.1296	0.9957	1
18	0.9957	0	0	0.1296	0.9957	1
26	0	0	0	0.1296	0.9957	1
34	0	0	0	0.1296	0	1
42	0	0	0	0.1296	0	1
50	0	0	0	0.1296	0	1

Table 7.2: Normalized wall energy after synchronization, using the new method and using the Beek-Sandell algorithm, for various values of SNR, and $P = 2400$ symbols

Chapter 8

Conclusion

This thesis explored the effect of a delay constraint on optimum transmission strategies. Here, the delay constraint was in terms of the fading rate of the channel process rather than in terms of the symbol rate. A block fading channel model was considered, and a delay constraint was imposed on data transmission. Thus, processing of data was assumed to occur in (a small number of) K blocks, each block itself being of (large) length T_0. The idea was to capture the non-ergodicity of the fading process for the duration of the data processing, due to the small K, while at the same time allow limiting arguments to be used to calculate 'capacity', facilitated by the large T_0. A general cost function $\mu(x)$ was considered in solving the delay constrained transmission problem. Two kinds of power constraints were considered, the short term and the long term power constraints. The solution to the long term power constraint is better (results in a higher maximum) because it is a more relaxed constraint. Since the power adaptation that maximized the cost function had to be causal, hence dynamic programs were found to provide the optimum power adaptation solution.

The general cost function was then specialized to the case of expected capacity,

by choosing $\mu(x) = x$. Expected capacity is the maximum ensemble-average data rate that can be obtained by optimizing the transmission power. It was observed that optimizing the transmitted power does not give much benefit at high SNRs, but provides a substantial gain at lower SNRs. At low SNRs, it was proved that the factor increase in capacity, due to power adaptation, was approximately $\frac{\log K}{m}$ if the channel faded according to the χ^2_{2m} (Nakagami) statistics. It was also shown that the long term power constraint solution in the expected capacity case was simply the memoryless strategy of 'time-waterfilling' [20].

The case of outage capacity was considered next. Outage capacity is defined as the maximum error-free data rate that can be supported at a given outage probability. Here, outage is the event that a given target rate R_0 cannot be supported by a fading channel over a given time period. It was shown that the optimum power adaptation solution to the long term constraint problem gives a substantial SNR gain at both low and high SNRs. The solution to the short term constraint problem however does not provide any SNR gain at high SNRs. It was shown that whereas the outage probability is inversely related (with a power of m) to the SNR in the short term case, it is related at least exponentially in the long term case. Thus we can expect the SNR gain of the long term constraint solution over no power adaptation, to increase with decreasing outage probability. However, the effect of discretization of the dynamic program state space reduces the SNR gain. Random coding bounds were also considered for the outage capacity case. These are non-trivial because in this case, the codewords span K fading blocks. It was seen that $T_0 = 100$ may be enough to get the decoding error probability close to the outage probability.

A stationary version of the outage probability problem was also considered. The

formulation used an exponential window which weighed the past data rates to approximate a K block window. Stationarity was introduced in the formulation by considering a time-averaged optimization. The solution involved linear programming. An SNR gain was observed when the optimal power control was used rather than a constant power scheme.

Space-time codes were considered as an example of the use of outage probability. In particular, the case of an outdoor wireless multi-antenna transmission system was considered. It was shown that maximum diversity and SNR gain could be obtained by simply combining space-time codes with an appropriate 'beamformer'. Thus, we could separate the signal processing and the coding aspects of the transmitter and optimize each separately.

Finally, the problem of blind symbol synchronization in OFDM was considered. It was shown that the ranks of certain autocorrelation matrices contain information that can be used to blindly synchronize the received signal, even in the presence of multipath. As opposed to previously existing blind synchronization methods, the new algorithm was shown to guarantee correct synchronization asymptotically.

Appendix A

A.1 The method of bisection

The method of bisection can be used to minimize a one-dimensional convex function $\Phi(x)$ over the domain $0 \leq x \leq 1$ (clearly the domain can be generalized to any bounded interval). One such variation of bisection is described below,

Algorithm Bisection() :

Initialize $x_1 = 0$, $x_2 = 0.5$, $x_3 = 1$.

Do,

1. If $x_2 - x_1 \geq x_3 - x_2$, choose $x_0 = \frac{x_1+x_2}{2}$ and reorder the points in increasing order as, $x_3 \rightarrow x_4, x_2 \rightarrow x_3, x_0 \rightarrow x_2, x_1 \rightarrow x_1$.

   If $x_2 - x_1 < x_3 - x_2$, choose $x_0 = \frac{x_2+x_3}{2}$ and reorder the points in increasing order as, $x_3 \rightarrow x_4, x_0 \rightarrow x_3, x_2 \rightarrow x_2, x_1 \rightarrow x_1$.

   This step bisects the larger of the two intervals.

2. If $\Phi(x_3) \leq \Phi(x_2)$, then reorder the points as $x_2 \rightarrow x_1, x_3 \rightarrow x_2, x_4 \rightarrow x_3$.

   If $\Phi(x_3) > \Phi(x_2)$, then do not reorder the points (simply throw away x_4).

   This step rejects the subinterval that clearly does not contain the minimum.

until $x_3 - x_1 < \epsilon_x$, where $\epsilon_x > 0$ is some preset threshold.

Return $\Phi(x_2)$ as the solution.

End Algorithm

Some thought will show that this algorithm should find the minimum of any convex (or even quasi-convex) function.

A.2 A refinement of the upper bound in Section 4.5.3

Whereas the Monte Carlo calculation in Section 4.5.1 is very complex, the dynamic program in Section 4.5.3 results in a looser upper bound. One way to retain the dynamic programming flavor of calculating the bound and yet yield a tighter bound is to partition the calculation of the bound into several sums, each of which are optimized separately as below,

$$P_{err} \leq \sum_j \{ \min_{0 \leq \rho_j \leq 1} \mathop{\mathbf{E}}_{g^{(K)} \epsilon \mathcal{S}_j} [\exp(-KTE_r(\rho, q(\mathbf{X}), g^{(K)}, R))] \} \qquad \text{(A.1)}$$

This is tighter than (4.24) which uses only a single set $\mathcal{S}_1 = g^{(K)}$. The sets $\{\mathcal{S}_j\}$ can be any convenient partition of the set $\{g^{(K)}\}$. In particular, we choose them as $\mathcal{S}_j = \{g^{(K)} : R_{th,j-1} \leq R^{(K+1)} \leq R_{th,j}\}$, where $\{R_{th,0} = 0, R_{th,1}, R_{th,2}, \ldots\}$ is any partition of the interval $[0, KR_0]$. Note again that $R^{(K+1)}$ is the target rate remaining after K transmissions, i.e. KR_0- achieved rate. The jth summand in (A.1) can be calculated by algorithm **Calculate_error(ρ)**, with the modification

that the initialization in step (3) is done using the indicator function,

$$\nu_{K+1}(Z_{K+1}) \;=\; \begin{cases} 1 & R_{th,j-1} \le R^{(K+1)} \le R_{th,j} \\ 0 & \text{else} \end{cases}$$

This increases the complexity of the bound calculation by a factor equal to the number of sets $\mathcal{S}_j$ in the partition. At the very least, we could choose two sets by choosing $R_{th,0} = 0, R_{th,1} = KR_0 - KR, R_{th,2} = KR_0$. In this case we can see that the 2nd summand is optimized by $\rho_2 = 0$, because all $g^{(K)} \epsilon \mathcal{S}_2$ result in achieving a rate smaller than the code rate R. This modification itself should significantly tighten the error bound (A.1).

A.3 Proof of optimal beamformer for ISI-free case

Proof: Let the SVD of $\mathbf{A}$ be:

$$\mathbf{A} \;=\; \mathbf{T}\boldsymbol{\Sigma}\boldsymbol{Q}^* \;=\; \mathbf{T}\;\; (\boldsymbol{\Sigma}_+ \mid \mathbf{0})\;\; (\underbrace{\mathbf{Q}_+}_{n\times L} \mid \square\,)^*$$

The maximization criterion can be written as $\det[\boldsymbol{\Sigma}\mathbf{V}\mathbf{V}^*\boldsymbol{\Sigma}^*]$, where $\mathbf{V}{=}\mathbf{Q}^*\mathbf{W}$. The criterion further simplifies to $\det[\boldsymbol{\Sigma}_+\mathbf{V}_+\mathbf{V}_+^*\boldsymbol{\Sigma}_+^*] = \det[\boldsymbol{\Sigma}_+\boldsymbol{\Sigma}_+^*]\det[\mathbf{V}_+^*\mathbf{V}_+]$, where $\mathbf{V}^* = (\mathbf{V}_+^* \mid \mathbf{V}_-^*)$. Meanwhile, the constraint $\mathrm{tr}[\mathbf{W}\mathbf{W}^*] = L$ simplifies to $\mathrm{tr}[\mathbf{V}\mathbf{V}^*] = \mathrm{tr}[\mathbf{V}^*\mathbf{V}] = \mathrm{tr}[\mathbf{V}_+^*\mathbf{V}_+] + \mathrm{tr}[\mathbf{V}_-^*\mathbf{V}_-] = L$. Since $\mathbf{V}_-$ does not appear in the maximization criterion, we choose $\mathbf{V}_- = 0$. Thus the problem is now to

$$\begin{aligned} \underset{\mathbf{V}_+}{\text{maximize}} \quad & \det[\mathbf{V}_+^*\mathbf{V}_+] \\ \text{subject to} \quad & \mathrm{tr}[\mathbf{V}_+^*\mathbf{V}_+] = L \end{aligned} \tag{A.2}$$

The solution to this is to choose $\mathbf{V}_+ = \mathbf{I}$, which implies that $\mathbf{W} = \mathbf{QV} = \mathbf{Q}_+$. □

A.4 Proof of lower bound on gain for linear array

Proof: The special structure of the linear array response allows us to write for the n antenna case

$$\mathbf{A}_n \mathbf{A}_n^* = (\mathbf{Z}_1 \mid \mathbf{Z}_2) \tag{A.3}$$

where $\mathbf{Z}_1 = \mathbf{A}_L \mathbf{A}_L^*$ and $\mathbf{Z}_2 = (\mathbf{D}_2 \mathbf{A}_L \mathbf{A}_L^* \mathbf{D}_2^* \mid \cdots \mid \mathbf{D}_{n/L} \mathbf{A}_L \mathbf{A}_L^* \mathbf{D}_{n/L}^*)$. Here, $\mathbf{A}_n, \mathbf{A}_L$ are the $\mathbf{A}$ matrices for the n and L antenna case respectively, while the $\mathbf{D}$s are diagonal matrices with unit magnitude diagonal entries.

Now the concavity of the $\log\det(\mathbf{Z})$ function for positive definite matrices $\mathbf{Z}$ allows us to conclude that ([24])

$$\det[\beta \mathbf{Z}_1 + (1-\beta)\mathbf{Z}_2] \;\geq\; (\det[\mathbf{Z}_1])^{\beta} (\det[\mathbf{Z}_2])^{1-\beta} \tag{A.4}$$

for positive definite $\mathbf{Z}_1, \mathbf{Z}_2$ and $0 \leq \beta \leq 1$. So, by induction,

$$\begin{aligned} \det[\mathbf{A}_n \mathbf{A}_n^*] &\geq \left(\det\left[\frac{n}{L}\mathbf{Z}_1\right]\right)^{L/n} \left(\det\left[\frac{n}{n-L}\mathbf{Z}_2\right]\right)^{(n-L)/n} \\ &\geq \cdots \quad \cdots \quad \cdots \\ &= \left(\frac{n}{L}\right)^{L} \det[\mathbf{A}_L \mathbf{A}_L^*] \end{aligned} \tag{A.5}$$

Noting that $(\det[\mathbf{\Sigma}_+])^2 = \det[\mathbf{AA}^*]$, the stated SNR gain inequality follows.
Note that a similar inequality would hold in a variety of antenna array configurations, where the array response matrix has the key structure of (A.3). □

A.5 Proof of Lemma 4

Proof: We have the inequalities [24]

$$\rho[\mathbf{A}] + \rho[\mathbf{B}] - k \;\leq\; \rho[\mathbf{AB}] \;\leq\; \min\{\rho[\mathbf{A}], \rho[\mathbf{B}]\}$$

where k is the number of columns in $\mathbf{A}$.

Recognizing that

$$\begin{aligned} \mathbf{R}_{\mathbf{z}}(i) &= \mathbf{H}\mathbf{R}_{\mathbf{x}}(i)\mathbf{H}^{*}, \\ \rho[\mathbf{R}_{\mathbf{z}}(i)] &= \rho[\mathbf{R}_{\mathbf{z}}^{1/2}(i)], \\ \rho[\mathbf{R}_{\mathbf{x}}(i)] &= \rho[\mathbf{R}_{\mathbf{x}}^{1/2}(i)] \end{aligned}$$

we can identify $\mathbf{A} = \mathbf{H},\ \mathbf{B} = \mathbf{R}_{\mathbf{x}}^{1/2}(i),\ \mathbf{AB} = \mathbf{R}_{\mathbf{z}}^{1/2}(i)$, to get

$$\rho[\mathbf{H}] + \rho[\mathbf{R}_{\mathbf{x}}(i)] - (N+\nu) \;\leq\; \rho[\mathbf{R}_{\mathbf{z}}(i)] \;\leq\; \min\{\rho[\mathbf{H}], \rho[\mathbf{R}_{\mathbf{x}}(i)]\}$$

Noticing that each repeated element in $\mathbf{x}_i$ (due to the cyclic prefix) causes loss of rank by one in $\mathbf{R}_{\mathbf{x}}(i)$, and also, because the non-repeated elements (the data) are non-degenerate, we obtain the bounding functions $g_l(i)$ and $g_u(i)$ as specified in Lemma 4. □

A.6 Proof of Theorem 4

Proof: Case 1: Consider $i \in \{(\nu - L_0), \ldots, (\nu - 1)\}$

Because of the cyclic prefix (that repeats elements), we can write (7.1) as

$$\mathbf{z}_i = \mathcal{H}_i \mathbf{w}_i$$

where $\mathbf{w}_i = (x_\nu, \ldots, x_N, \ldots, x_{N+\nu-1}, \ldots, x_{N+\nu-1+i})^t$.

We can write $\mathcal{H}_i$ in terms of its columns as $\mathcal{H}_i = (\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{N+i-1})$. Then, since $N > L_0$ is assumed, we can relate the $\{\mathbf{v}_k\}$ to $\{\mathbf{h}_k\}$ as

$$\begin{aligned} \mathbf{v}_k &= \mathbf{h}_{k+\nu-i} \qquad k = 0, \ldots, N-\nu+i-1, N, \ldots, N+i-1 \\ \mathbf{v}_k &= \mathbf{h}_{k+\nu-i} + \mathbf{h}_{k+\nu-i-N} \qquad k = N-\nu+i, \ldots, N-1 \end{aligned} \tag{A.6}$$

Similarly, we can write

$$\mathbf{z}_{i+1} = \mathcal{H}_{i+1} \mathbf{w}_{i+1}$$

with the appropriate definitions.

Again, writing $\mathcal{H}_{i+1} = (\tilde{\mathbf{v}}_0, \tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_{N+i})$ we can relate $\{\tilde{\mathbf{v}}_k\}$ to $\{\mathbf{h}_k\}$ as

$$\begin{aligned} \tilde{\mathbf{v}}_k &= \mathbf{h}_{k+\nu-i-1} \qquad k = 0, \ldots, N-\nu+i, N, \ldots, N+i \\ \tilde{\mathbf{v}}_k &= \mathbf{h}_{k+\nu-i-1} + \mathbf{h}_{k+\nu-i-N-1} \qquad k = N-\nu+i+1, \ldots, N-1 \end{aligned} \tag{A.7}$$

From (A.6),(A.7), it is clear that

$$\begin{aligned} \mathbf{v}_k &= \tilde{\mathbf{v}}_{k+1} \qquad k = 0, \ldots, N-2, N, \ldots, N+i-1 \\ \mathbf{v}_{N-1} &= \tilde{\mathbf{v}}_N + \tilde{\mathbf{v}}_0 \end{aligned}$$

Since the rank of a matrix is the maximum number of linearly independent columns, and we are losing at most one degree of freedom in $\mathcal{H}_i$ as compared to $\mathcal{H}_{i+1}$, hence,

$$\rho[\mathcal{H}_i] = \rho[\mathcal{H}_{i+1}] - \mu_i \qquad \mu_i \in \{0, 1\} \text{ is an unknown constant}$$

i.e. Rank decreases by at most one in going from $\mathcal{H}_{i+1}$ to $\mathcal{H}_i$. By the assumption of non-degeneracy of the input, $\mathbf{R}_{\mathbf{w}}(i)$ and $\mathbf{R}_{\mathbf{w}}(i+1)$ are full rank, and so we can write

$$\begin{aligned} \rho[\mathbf{R}_{\mathbf{z}}(i)] &= \rho[\mathcal{H}_i], \\ \rho[\mathbf{R}_{\mathbf{z}}(i+1)] &= \rho[\mathcal{H}_{i+1}] \end{aligned}$$

thus giving the result stated in Theorem 4 for this case.

Case 2: Consider $i \in \{-(\nu-1), \ldots, (\nu - L_0)\}$

If we flip the column vectors $\mathbf{z}_i$ and $\mathbf{x}_i$ upside down (re-index their elements as $\cdots 3, 2, 1$ instead of $1, 2, 3, \cdots$), then (7.1) will change to a new convolution equation, but with the channel taps $h_0, \ldots, h_{L_0}$ in reverse order. The same analysis as carried out above will apply, showing that

$$\rho[\mathbf{R}_{\mathbf{z}}(i)] = \rho[\mathbf{R}_{\mathbf{z}}(i-1)] - \mu_i \quad i = -(\nu-1), \ldots, -(\nu - L_0)$$

Case 3: Consider $i \in \{\nu, \ldots, N\}$

In this case, $\mathbf{w}_i = \mathbf{x}_i$, and $\mathbf{R}_{\mathbf{w}}(i)$ is full rank. So, $\rho[\mathbf{R}_{\mathbf{z}}(i)] = \rho[\mathbf{H}] = N + \nu - L_0$.

Case 4: Consider $i \in \{0, \ldots, (\nu - L_0 - 1)\}$

As in case 1, we can write

$$\begin{aligned} \mathbf{z}_i &= \mathcal{H}_i \mathbf{w}_i, \\ \mathbf{z}_{i+1} &= \mathcal{H}_{i+1} \mathbf{w}_{i+1} \end{aligned}$$

Call the first row of $\mathcal{H}_i$ as $(\mathcal{H}_i)_1$, so that $\mathcal{H}_i = ((\mathcal{H}_i)_1^t, \widetilde{\mathcal{H}}_i^t)^t$. Due to the special structure of $\mathcal{H}_i$, its first row is identical to its Nth row.

By observing the structure of $\mathcal{H}_{i+1}$ we can write

$$\mathcal{H}_{i+1} = \left(\begin{array}{ccccc|c} & & \widetilde{\mathcal{H}}_i & & & 0 \\ \hline \cdots & 0 & h_{L_0} & \cdots & h_1 & h_0 \end{array}\right)$$

Now, the two observations made earlier allow us to conclude that

$$\begin{aligned} \rho[\mathcal{H}_i] &= \rho[\widetilde{\mathcal{H}}_i] \\ &= \rho[(\widetilde{\mathcal{H}}_i | 0)] \\ &= \rho[\mathcal{H}_{i+1}] - 1 \\ \rho[\mathbf{R}_{\mathbf{z}}(i)] &= \rho[\mathbf{R}_{\mathbf{z}}(i+1)] - 1 \end{aligned}$$

Case 5: Consider $i \in \{-(\nu - L_0 - 1), \ldots, 0\}$

This can be analyzed by flipping column vectors $\mathbf{z}_i$ and $\mathbf{x}_i$ upside down (re-indexing their elements as $\cdots 3, 2, 1$ instead of $1, 2, 3, \cdots$) and noting that this case reduces to

case 4.

Combining the results for the five cases proves Theorem 4. □

A.7 Proof of Corollary 1

Proof: Since $L \leq L_0$, hence (7.1) can be modified as

$$\mathbf{z}_i = \begin{pmatrix} h_L & \cdots & h_0 & 0 & 0 \\ 0 & \ddots & & \ddots & 0 \\ 0 & 0 & h_L & \cdots & h_0 \end{pmatrix} \begin{pmatrix} x_{i+\Delta} \\ \cdots \\ x_{i+N+\nu-1} \end{pmatrix}$$

Notice that if we replace $\{L_0, \nu\}$ by $\{L, \nu - \Delta\}$, then the case $i \in \{0, 1, \ldots, N\}$ can be analyzed in exactly the same manner as in Theorem 4. The result found there for this case, will apply here, with the above replacements. Note that due to the short (L taps) channel here, at most $\nu - \Delta$ prefix elements can occur in any vector $(x_{i+\Delta}, \ldots, x_{i+N+\nu-1})^t$.

For the case, $i \in \{-\Delta, \ldots, 0\}$, we conclude that $\rho[\mathbf{R}_\mathbf{z}(i)] = \rho[\mathbf{R}_\mathbf{z}(0)]$ because each of these has exactly $\nu - \Delta$ prefix elements in the corresponding vector $(x_{i+\Delta}, \ldots, x_{i+N+\nu-1})^t$.

The analysis of the case $i \in \{-\nu, \ldots, -\Delta\}$ is similar to the analysis of the case $i \in \{-\nu, \ldots, 0\}$ in Theorem 4.

Combining these results, we get Corollary 1. □

A.8 Rank Estimation using MDL

The MDL criterion can be used to estimate the ranks of autocorrelation matrices [59]. Here we derive the criterion for the case of known σ_n^2. According to the MDL criterion, the rank estimate $\widehat{\rho}[\mathbf{R_z}(i)]$ can be obtained as

$$\begin{aligned}\widehat{\rho}[\mathbf{R_z}(i)] &= \text{argmin}_{k,\mathbf{R}} \{-\log(f(\{\mathbf{y}_i\} \mid \mathbf{R})) + \\ &\qquad \frac{\log(P)}{2} \cdot (\text{number of free parameters in } \mathbf{R})\Big\}\end{aligned}$$

where the matrix $\mathbf{R} - \sigma_n^2\mathbf{I}$ has rank k. $f(\cdot)$ is the likelihood function for the received signal $\{\mathbf{y}_i\}$, given underlying autocorrelation model $\mathbf{R}$. Now, the number of free parameters in $\mathbf{R}$ is $2k(N+\nu-L_0) - k^2 + k$, obtained by noting that the rank of the size $(N+\nu-L_0)$ matrix $\mathbf{R} - \sigma_n^2\mathbf{I}$ is k, and that it is complex orthonormal. Further, due to the Gaussian assumption of $\{x_i\}$, $\{y_i\}$ are also Gaussian. For ease of notation, define $N+\nu-L_0 = M$. Define the SVD $\widehat{\mathbf{R}}_\mathbf{y}(i) = \mathbf{U}\widehat{\mathbf{\Sigma}}\mathbf{U}^*$. Then,

$$f(\{\mathbf{y}_i\}|\mathbf{R}) = \frac{1}{\pi^M \det(\mathbf{R})} \exp(-P \cdot \mathbf{Trace}[\widehat{\mathbf{R}}_\mathbf{y}(i)\mathbf{R}^{-1}]) \tag{A.8}$$

Therefore,

$$\begin{aligned}\widehat{\rho}[\mathbf{R_z}(i)] &= \text{argmin}_{k,\mathbf{R}} \Big\{P \cdot \mathbf{Trace}[\widehat{\mathbf{R}}_\mathbf{y}(i)\mathbf{R}^{-1}] + \log\det(\mathbf{R}) + \\ &\qquad \frac{\log(P)}{2} \cdot (\text{number of free parameters in } \mathbf{R})\Big\}\end{aligned}$$

Using the Hadamard inequality ([24]), it is seen that the minimization over $\mathbf{R}$ is partly achieved by choosing $\mathbf{R} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^*$. Minimizing further over $\mathbf{\Sigma}$, enforcing the k rank condition of $\mathbf{R} - \sigma_n^2\mathbf{I}$, and simplifying the expression by removing constant

terms, we obtain the MDL criterion specified in Section 7.4.1. It is sufficient to check the minimization for $k \geq N - L_0$ because of the lower bound on the ranks, as shown in Lemma 4.

Bibliography

[1] R. Ahlswede, "Arbitrarily varying channels with state sequence known to the sender", *IEEE Trans. Inform. Theory*, vol. 32, pp. 621-629, Sept. 1986.

[2] A. Armada, and M. Ramon,"Rapid prototyping of a test modem for terrestrial broadcasting of digital television", *IEEE Trans. Consum. Electr.*, vol. 43, pp. 1100-1109, Nov. 1997.

[3] I. Bettesh, and S, Shamai, "A low delay algorithm for the multiple access channel with Rayleigh fading", *Proc. of Int. Symp. on Personal, Indoor, and Mobile Radio Comm.*, Boston, MA, Sept. 8-11, 1998.

[4] E. Biglieri, J. Proakis, and S. Shamai, "Fading channel: Information theoretic and communication aspects", *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619-2692, Oct. 1998.

[5] J. Boutros, E. Viterbo, C. Rastello, and J. Belfiore, "Good lattice constellations for both Rayleigh fading and Gaussian channels", *IEEE Trans. Inform. Theory*, vol. 42, pp. 502-518, Mar. 1996.

[6] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels" *IEEE Trans. on Inform. Theory*, vol.45, pp. 1468-1489, July 1999.

[7] G. Caire, G. Taricco, and E. Biglieri, "Optimal power control for minimum outage rate in wireless communications", in *Proc. Int. Conf. Communications (ICC'98)*, Atlanta, GA, June 7-11, 1998, pp. 58-62.

[8] A. Calderbank, N. Seshadri, and V. Tarokh, "Space-time codes for wireless communication", *Proc. of IEEE Intern. Symp. on Inform. Theory*, pp. 146, 1997.

[9] W. Choi, R. Negi, and J. Cioffi, "Combined ML and DFE decoding for the V-BLAST system", *Proc, Intern. Conf. Commun.*, New Orleans, LA, June 18-22, 2000.

[10] B. Collins, and R. Cruz, "Transmission policies for time varying channels with average delay constraints", *Proc. Allerton Conf. on Commun., Control and Comp.*, Monticello, IL, 1999.

[11] T. Cover, "Broadcast channels", *IEEE Trans. on Inform. Theory*, vol.18, pp. 2-14, Jan. 1972.

[12] T. Cover, and J. Thomas, *Elements of Information Theory*, New York:Wiley, 1991

[13] C. Cozzo and B. Hughes, "Space diversity in the presence of discrete multipath", *Proc. IEEE Vehicular Tech. Conf.*, vol. 2, pp. 1286-1290, 1999.

[14] I. Csiszar, and P. Narayan, " Arbitrarily varying channels with constrained inputs and states", *IEEE Trans. Inform. Theory*, vol. 34, pp. 27-34, Jan. 1988.

[15] D. Divsalar, and M. Simon, "The design of trellis coded MPSK for fading channels: performance criteria", *IEEE Trans. on Communications*, vol.36, pp. 1004-1012, Sept. 1988.

[16] G.J. Foschini, "Layered space-time architecture for wireless communication in a fading environments using multi-element antennas", *Bell Labs Tech. Jour.*, vol. 1, pp. 41-59, Aut. 1996.

[17] G.J. Foschini, and M. Gans, "On limits of wireless communication in a fading environment when using multiple antennas", *Wireless Personal Commun.*, vol. 6, pp. 311-335, 1998.

[18] R. Gallager, *Information Theory and Reliable Communication*, New York:John Wiley, 1968

[19] E. Gilbert, "Capacity of a burst noise channel", *Bell Labs Tech. Journal*, vol. 39, pp. 1253-1265, 1960.

[20] A.J. Goldsmith, and P.P. Varaiya, "Capacity of fading channels with channel side information", *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986-1992, Nov. 1997.

[21] G. Golub, and C. Loan,*Matrix computations*, Baltimore:John's Hopkins University Press, 1996.

[22] S. Hanly, and D. Tse, "The multi-access fading channel: Shannon and delay-limited capacities", in *Proc. of 33rd Annual Allerton Conf.*, Monticello, IL, Oct. 4-6, 1995, pp.786-795.

[23] O. Hernandez-Lerma, *Adaptive Markov Control Processes*, New York:Springer-Verlag, 1989

[24] R.A. Horn, and C.R. Johnson,*Matrix analysis*, New York:Cambridge University Press, 1996.

[25] R. Knopp, and P.A. Humblet, "Information capacity and power control in single-cell multiuser communications", in *Proc. Int. Conf. Communications (ICC'95)*, Seattle, WA, June 18-22, 1995, pp. 331-335.

[26] D. Landstrom, J. Arenas, J. van de Beek, P. Borjesson, M. Boucheret, P. Odling,"Time and frequency offset estimation in OFDM systems employing pulse shaping", *Proc. ICUPC*, pp. 279-283, San Diego, Oct. 1997.

[27] A. Lapidoth, and P. Narayan, "Relaible communication under channel uncertainty", *IEEE Trans. Inform. Theory*, vol.44, pp. 2148-2176, Oct. 1998.

[28] A. Lapidoth, and S. Shamai, "Fading channels: How perfect need 'perfect side information' be?", *Proc. IEEE Inform. Theory Workshop*, pp. 36-38, South Africa, June 20-25, 1999.

[29] D. Lee, and K. Cheun,"A new symbol timing recovery algorithm for OFDM systems",*IEEE Trans. Consum. Electr.*, vol. 43,pp. 767-775, Aug. 1997.

[30] P. Ligdas, and N. Farvardin, "Power control schemes for fading additive white Gaussian noise channels", *Proc. Conf. on Information Science and Systems*, Baltimore, MD, March 22-24, 1995.

[31] P. Ligdas, and N. Farvardin, "Optimizing the transmit power for slow fading channels", *IEEE Trans. Inform. Theory*, vol.46, pp. 565-576, March 2000.

[32] E. Malkamaki, "Coded diversity on block-fading channels", *IEEE Trans. Inform. Theory*, vol. 45, pp. 771-781, March 1999.

[33] T. Marzetta, and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading", *IEEE Trans. Inform. Theory*, Vol. 45, pp 139-157, Jan 1999.

[34] R.J. McEliece, and W.E. Stark, "Channels with block interference", *IEEE Trans. Inform. Theory*, vol. 30, pp. 44-53, Jan. 1984.

[35] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders", *IEEE Trans. Inform. Theory*, vol.40, pp. 1953-1967, Nov. 1994.

[36] S. Muller-Weinfurtner,"On the optimality of metrics for coarse frame synchronization in OFDM: A comparison", *Proc. of Symp. on PIMRC*, pp. 533-537, Boston, Sept. 1998.

[37] A.F. Naguib, V. Tarokh, N. Seshadri, A. Calderbank, "A space-time coding modem for high-data-rate wireless communications", *IEEE Journ. Selected areas in Commun.*, vol. 16, pp. 1459-1478, 1998.

[38] A. Narula, M. Lopez, and G. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays", *IEEE Trans. Inform. Theory*, vol.45, pp. 2418-2433, Nov. 1999.

[39] R. Negi, M. Charikar, and J. Cioffi, "Minimum outage transmission over fading channels with delay constraint", in *Proc. Int. Conf. Communications (ICC'00)*, New Orleans, LA, June 18-22, 2000.

[40] R. Negi, and J. Cioffi, "Transmission over fading channels with channel side information and delay constraint", *Proc. Globecom*, Rio de Janiero, Brazil, Dec. 5-8, 1999.

[41] R. Negi, and J. Cioffi, "Blind OFDM symbol synchronization in ISI channels", *Proc. IEEE Globecom*, pp. 2812-2817, Sydney, Nov. 1998.

[42] R. Negi, and J. Cioffi, "Stationary Schemes for Optimal Transmission over Fading Channels with Delay Constraint", *Proc. IEEE Vehicular Tech. Conf.*, Boston, MA, Sept. 24-28, 2000.

[43] R. Negi, A. Tehrani, and J. Cioffi, "Adaptive Antennas for Space-Time Coding", *Proc. IEEE Vehicular Technology Conf.*, pp. 70-74, 1999.

[44] B. Ng, J. Chen, and A. Paulraj, "Space time processing for fast fading channels with co-channel interference", *Proc. IEEE Vehicular Technology Conf.*, vol. 3, pp. 1491-1495, 1996.

[45] B. Oksendal, *Stochastic Differential Equations*, New York:Springer-Verlag, 1998

[46] L. Ozarow, S. Shamai, and A. Wyner, " Information theoretic considerations for cellular mobile radio", *IEEE Trans. Vehicular Technology*, vol. 43, pp. 359-378, May 1994.

[47] A. Palin, and J. Rinne, "Enhanced symbol synchronization method for OFDM system in SFN channel", *Proc. IEEE Globecom*, pp. 2788-2793, Sydney, Nov. 1998.

[48] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, New York:McGraw-Hill Inc., 1991

[49] J.G. Proakis, *Digital Communications*, New York:McGraw-Hill Inc., 1995

[50] A. Peled, and A. Ruiz,"Frequency domain data transmission using reduced computational complexity algorithms", *Proc. IEEE ICASSP*, pp. 964-967, Denver, CO, 1980.

[51] M. Puterman, *Markov decision processes*, New York:John Wiley, 1994

[52] G.G. Raleigh, and J.M. Cioffi, "Spatio-temporal coding for wireless communications", *Proc. IEEE Globecom*, vol, 3, pp. 1809-1814, 1996.

[53] C. Shannon, "Channels with side information at the transmitter", *IBM J. Research Devop.*, no. 2, pp. 289-293, 1958

[54] N. Seshadri, V. Tarokh, and A. Calderbank, "Space-time codes for wireless communication: code construction", *Proc. IEEE Vehicular Tech. Conf.*, vol. 2, pp. 637-641, 1997.

[55] V. Tarokh, A. Naguib, N. Seshadri, and A. Calderbank, "Low-rate multidimensional space-time codes for both slow and rapid fading channels", *Proc. of Intern. Symp. on Personal, Indoor and Mobile Radio Commun. - PIMRC*, vol. 3, pp. 1206-1210, 1997.

[56] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction", *IEEE Trans. Inform. Theory*, vol.44, no.2, pp. 744-765, March 1998.

[57] J. van de Beek, M. Sandell, and P. Ola Borjesson,"ML estimation of time and frequency offset in OFDM systems", *IEEE Trans. Signal Proc.*, vol. 45, pp. 1800-1805, July 1997.

[58] S. Vishwanath, W. Yu, R. Negi, and A. Goldsmith, "Decorrelation of the constituent codes in space-time turbo codes and it applications", *Proc. IEEE Globecom*, San Francisco, CA, Nov. 27 - Dec 1, 2000.

[59] M. Wax, and T. Kailath,"Detection of signals by information theoretic criteria", *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 33,pp. 387-392, Apr. 1985.

[60] A. Wittneben, " A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation", *Proc. IEEE Intern. Conf. Commun.*, pp. 1630-1634, 1993.

[61] G.W. Wornell, "Signal processing techniques for efficient use of transmit diversity in wireless communications", *Proc. IEEE Intern. Conf. on Acoustics, Speech, and Signal Proc.*, vol. 2, pp. 1057–60, 1996.

[62] G. Wornell, and M. Trott, " Signal processing techniques for efficient use of transmit diversity in wireless communications", in *Proc. of International Conf. on Acoustics, Speech, and Signal Proc.*, Atlanta, GA, May 7-10, 1996, pp. 1057-60.

[63] J. Yang, and S. Roy, "On joint transmitter and receiver optimization for multiple input-multiple-output (MIMO) transmission systems", *IEEE Transactions on Communications*, vol. 42, pp. 3221-3231, Dec. 1993.

[64] J. Yang, and S. Roy, "Joint transmitter and receiver optimization for multiple input-multiple-output (MIMO) systems with decision feedback", *proc. IEEE Trans. on Inform. Theory*, vol. 40, pp. 1334-1347, Sept. 1994.

[65] W.Y. Zou, and Y. Wu, "COFDM:An overview", *IEEE Trans. Broadcasting*, vol. 41, pp. 1-8, Mar. 1995.